Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the catalytic domain of the *Clostridium thermocellum* cellulase CelT

Muppuru M. Kesavulu,[a]‡ Jia-Yin Tsai,[a]‡ Hsiao-Lin Lee,[b]‡ Po-Huang Liang[b]* and Chwan-Deng Hsiao[a]*

[a]Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan, and [b]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan

‡ These authors contributed equally to this work.

Correspondence e-mail: phliang@gate.sinica.edu.tw, hsiao@gate.sinica.edu.tw

Received 21 July 2011
Accepted 16 January 2012

PDB Reference: CelTΔdoc, 2yik.

Cellulases hydrolyze cellulose, a major component of plant cell walls, to oligosaccharides and monosaccharides. Several *Clostridium* species secrete multi-enzyme complexes (cellulosomes) containing cellulases. *C. thermocellum* CelT, a family 9 cellulase, lacks the accessory module(s) necessary for activity, unlike most other family 9 cellulases. Therefore, characterization of the CelT structure is essential in order to understand its catalytic mechanism. Here, the crystal structure of free CelTΔdoc, the catalytic domain of CelT, is reported at 2.1 Å resolution. Its structure differs in several aspects from those of other family 9 cellulases. CelTΔdoc contains an additional α-helix, α-helices of increased length and two additional surface-exposed β-strands. It also contains three calcium ions instead of one as found in *C. cellulolyticum* Cel9M. CelTΔdoc also has two flexible loops at the open end of its active-site cleft. Movement of these loops probably allows the substrate to access the active site. CelT is stable over a wide range of pH and temperature conditions, suggesting that CelT could be used to convert cellulose biomass into biofuel.

1. Introduction

Cellulose is an energy source rich in carbon and therefore potentially a naturally available biofuel. Cellulose fibres, the major component of plant cell walls, consist of linear β-1,4-glucose polymers. Cellulases are glycoside hydrolases that hydrolyze β-1,4-glycosidic bonds and convert cellulose into short-chain glucosides. On the basis of their sequence similarities, glycoside hydrolases have been classified into 129 families (Cantarel *et al.*, 2009). Various microorganisms, *e.g.* fungi and bacteria, secrete cellulases as independent units or in the form of multi-enzyme complexes known as cellulosomes (Lamed *et al.*, 1983). The cellulases within a cellulosome act synergistically to completely break down cellulose to glucose (Doi & Kosugi, 2004; Gilbert, 2007; Demain *et al.*, 2005; Béguin & Lemaire, 1996; Shoham *et al.*, 1999; Bayer *et al.*, 2004).

Various *Clostridium* species, *e.g. C. cellulolyticum* (Bélaich *et al.*, 1997), *C. cellulovorans* (Doi *et al.*, 1994), *C. josui* (Kakiuchi *et al.*, 1998), *C. papyrosolvens* (Pohlschröder *et al.*, 1994) and *C. thermocellum* (Bayer *et al.*, 1998), produce cellulosomes. Among these species, the *C. cellulolyticum* and *C. thermocellum* cellulases have been the most extensively studied. *C. thermocellum* CelT (EC 3.2.1.4), which was identified in 1997, is a family 9 glycoside hydrolase (Kurokawa *et al.*, 2002). With the notable exception of CelT, which contains

Table 1
List of oligonucleotide PCR primers used in this study.

Primer	Sequence (5′–3′)†
P1	G**GAATTC**GCGGGCATTGTG
P2	CCG**CTCGAG**TTACACGTTATGGCTCGGCATG
P3	CCAGAACACCAACATGCCGGATGGCGATGGCAAAGTGG
P4	CCACTTTGCCATCGCCATCCGGCATGTTGGTGTTCTGG
P5	GATTGCCATGTGTATGATGCGGGCGATGGCAAAGTG-GATG
P6	CATCCACTTTGCCATCGCCCGCATCATACACATGGCAATC

† The sequences shown in bold are the restriction-enzyme cutting sites: *Eco*RI, GAATTC; *Xho*I, CTCGAG.

only a dockerin domain in addition to its catalytic domain, most other *C. thermocellum* family 9 cellulases also contain accessory module(s), *i.e.* a family 3 or family 4 carbohydrate-binding module (CBM) and/or an immunoglobulin-like module (Kurokawa *et al.*, 2002). Family 9 cellulases use acid–base catalysis mediated by a conserved glutamate and two aspartates to cleave the β-1,4 bond with inversion of the anomeric carbon (Davies & Henrissat, 1995) and are classified according to four themes, A, B, C and D, on the basis of their modular arrangements (Bayer *et al.*, 2006). Theme A cellulases are plant enzymes and possess only a catalytic domain. In addition to their catalytic domains, theme B cellulases have a family 3c CBM, theme C cellulases have an N-terminal immunoglobulin-like module and theme D cellulases have a family 4 CBM. Because it contains only a catalytic domain, according to this classification scheme CelT can be considered to be a theme A cellulase even though it is not a plant cellulase.

The activity of family 9 glycoside hydrolases is abolished or reduced when the family 3c CBM accessory module is removed (Sakon *et al.*, 1997; Gal *et al.*, 1997). Additionally, although its functional role is unknown, the immunoglobulin-like domain in family 9 glycoside hydrolases also seems to be required for catalytic activity (Kataeva *et al.*, 2004; Béguin & Alzari, 1998).

Only nine crystal structures have been solved for the 388 bacterial cellulases that have been sequenced. All *C. thermocellum* cellulases for which crystal structures are available have accessory modules in addition to the catalytic domain. The rationale for the study reported here was to delineate how CelT functions in the absence of the accessory module(s) and to identify residues involved in binding, catalysis and stability by structurally characterizing a truncated form of CelT (CelTΔdoc; residues 22–532) that contained only the catalytic domain.

2. Materials and methods

2.1. Cloning, expression and purification of CelTΔdoc and its truncated forms

The gene encoding CelTΔdoc was amplified from *C. thermocellum* strain ATCC 27405 genomic DNA with a pair of primers, P1 and P2 (Table 1), to generate an *Eco*RI site at the 5′-terminus and an *Xho*I site at the 3′-terminus. For the deletion of extra amino acids from CelTΔdoc, the truncated gene was prepared by overlap PCR. The primer pairs P1/P4 and P3/P2 (Table 1) were used to create the gene encoding CelTΔdoc without the H4 helix (amino acids 89–101): CelTΔ89–101Δdoc. Another truncated gene encoding CelTΔ78–103Δdoc was created using the primer pairs P1/P6 and P5/P2 (Table 1). These PCR products were cloned into the *Eco*RI and *Xho*I sites of pHTPP15 (Wang *et al.*, 2009), resulting in the addition of an N-terminal hexahistidine tag and a Trx tag. The pHTPP15 constructs were transformed into *Escherichia coli* JM109 and screened for the correct colonies.

The constructs were transformed into *E. coli* BL21 (DE3) for protein expression. A 10 ml overnight culture grown with a single colony was inoculated with 1 l autoclaved LB medium containing 50 mg ml^{-1} kanamycin and the cells were grown at 310 K to an optical density (OD_{600}) of 0.6. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M* to induce protein expression and the cells were grown at 310 K for a further 4 h and then harvested by centrifugation (5000 rev min^{-1} for 20 min at 277 K). Protein purification was performed at 277 K. The cell pellets were suspended in nickel column binding buffer (20 m*M* Tris–HCl pH 7.9, 5 m*M* imidazole, 500 m*M* NaCl) and lysed using a French press (Constant Cell Disruption Systems) operated at 138 MPa. The cell lysate was centrifuged (15 000 rev min^{-1} for 30 min at 277 K) and the supernatant was loaded onto a 20 ml Ni–NTA column equilibrated in binding buffer. The column was washed with three column volumes of binding buffer and protein bound to the column was eluted with a linear gradient (5–500 m*M*) of imidazole in binding buffer. Peak fractions were analyzed by SDS–PAGE. Fractions containing pure protein were pooled and exchanged into reaction buffer (100 m*M* NaCl, 50 m*M* Tris–HCl, 5 m*M* $CaCl_2$ pH 8.0) to cleave extra tags using factor Xa (Novagen). The protein without tags was purified on a Superdex 75 Fast Flow column (GE Healthcare) equilibrated in 50 m*M* bis-tris pH 9.0, 100 m*M* NaCl. Fractions containing pure protein were pooled and concentrated using 10 kDa molecular-weight cutoff Millipore Amicon ultracentrifugal filter devices. The final concentration of CelTΔdoc after gel-filtration column purification was 6.0 mg ml^{-1}. All other family 9 cellulases employed in this study for activity measurements were recombinantly expressed and purified.

2.2. Crystallization, data collection, phasing and refinement

Crystallization trials used 3.0 mg ml^{-1} CelTΔdoc in 50 m*M* bis-tris pH 9.0 and 100 m*M* NaCl, reagents from Hampton Research screening kits (Hampton Research, Aliso Viejo, California, USA) and the hanging-drop vapour-diffusion method. Each crystallization drop consisted of 1 µl protein solution and 1 µl reservoir solution. Diffraction-quality crystals appeared after 2 d using PEG/Ion 2 screen condition No. 9 (4% Tacsimate, 12% polyethylene glycol 3350 pH 4.0). Crystals grew in rhombus shapes to dimensions of 0.07 × 0.07 × 0.04 mm.

Table 2
Data-collection and refinement statistics for CelTΔdoc.

Values in parentheses are for the last shell.

Data collection	
Wavelength (Å)	1.000
Unit-cell parameters (Å)	$a = b = 96.86$, $c = 159.92$
Space group	$P4_32_12$
Resolution (Å)	30–2.1 (2.23–2.10)
No. of reflections	578399 (42562)
Completeness (%)	94.2 (98.8)
Multiplicity	13.6 (13.5)
R_{merge} (%)	6.1 (36.4)
$\langle I/\sigma(I)\rangle$	42.71 (6.63)
Refinement	
No. of reflections	38800
R_{work} (%)	18.2
R_{free} (%)	19.9
Geometry deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.5
Mean B values (Å^2)	
Protein atoms	29.70
Zinc	31.23
Calcium	30.70
Water molecules	36.8
No. of atoms	
Protein	934
Zinc	1
Calcium	3
Water molecules	343
Ramachandran plot (%)	
Most favoured	89.8
Additionally allowed	9.5
Generously allowed	0.7

Crystals were cryoprotected with 25% glycerol, mounted in a 0.1 mm nylon CryoLoop and flash-cooled in liquid nitrogen. Diffraction data were collected using an Area Detector Systems Corp. (ADSC) Quantum 210 charge-coupled device (CCD) area detector using a synchrotron-radiation source (beamline BL13C1, National Synchrotron Radiation Research Center, Taiwan). Data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The crystal diffracted to a resolution of 2.1 Å and belonged to the tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 96.86$, $c = 159.92$. Assuming one molecule per asymmetric unit, the Matthews coefficient (V_M; Matthews, 1968) was 3.27 Å^3 Da^{-1}, which corresponds to a solvent content of 62%. Data-collection details are summarized in Table 2.

Phases were solved by molecular replacement using the coordinates of Cel9M (Parsiegla *et al.*, 2002; PDB entry 1ia6) obtained from the Protein Data Bank (http://www.pdb.org; Berman *et al.*, 2000). For molecular replacement, *MOLREP* (Vagin & Teplyakov, 2010) from the *CCP*4 suite of programs (Winn *et al.*, 2011) was used. Data between 8.0 and 4.0 Å resolution and a Patterson radius of 20.0 Å were used for the rotation and translation calculations. After initial rigid-body refinement, the correlation coefficient and R factor were 56.5% and 25.7%, respectively. The model was built and residue geometries were adjusted using *XtalView* and *ARP*/*wARP* (McRee, 1999l Perrakis *et al.*, 1999). Data refinement was performed using *CNS* (Brünger *et al.*, 1998), with 5% of the data set aside and used for cross-validation. A solvent mask was calculated with the bulk-solvent program to improve the fit of the calculated structure factors to the reflection data. The final CelTΔdoc structure contained 493 residues (38–531), 343 water molecules, three calcium ions and a zinc ion, with R_{work} = 18.2 and R_{free} = 19.9% at 2.1 Å resolution. The φ/ψ angles of the residues were plotted as a Ramachandran plot (Ramakrishnan & Ramachandran, 1965) using *PROCHECK* (Laskowski *et al.*, 1993) and the angles for 89.8% of the residues were found to be in the most favoured regions, with none in the disallowed region. Refinement statistics are shown in Table 2. The CelTΔdoc atomic coordinates have been deposited in the Protein Data Bank with accession code 2yik.

2.3. Activity assays

Reducing-sugar concentration was determined using the 3,5-dinitrosalicylic acid assay (Miller, 1959) and a D-glucose standard curve. Endoglucanase activity was measured with carboxymethyl cellulose (CMC) as the substrate. The assay mixtures contained 1%(*w*/*v*) CMC and enzyme in 20 m*M* 2-(*N*-morpholino)ethanesulfonic acid pH 6.0 and were incubated for 10 min at an optimum temperature. Reactions were stopped by adding 3,5-dinitrosalicylic acid and were then heated for 5 min at 373 K. The absorbance was measured at 540 nm using a spectrophotometer. One unit of activity was defined as the amount of enzyme that released 1 µmol of glucose-reducing equivalent per minute.

2.4. Effects of temperature and pH on activity

The effects of temperature and pH on activity were determined using the assay presented in §2.3 with 0.1 µ*M* enzyme and 1%(*w*/*v*) CMC. Assays were performed at different temperatures (303–353 K) in 20 m*M* 2-(*N*-morpholino)-ethanesulfonic acid pH 6.0 or in buffers of different pH values (pH 3.0–8.0) at 323 K.

2.5. Effects of metal ions and reducing agents on activity

Assays were carried out for 10 min at 323 K in 20 m*M* 2-(*N*-morpholino)ethanesulfonic acid pH 6.0 containing 5 m*M* LiCl, $NiCl_2$, $CuCl_2$, $HgCl_2$, $CoCl_2$, $ZnCl_2$, $CaCl_2$, $MgCl_2$, $MnCl_2$, dithiothreitol or β-mercaptoethanol. The enzyme and substrate concentrations were 0.1 µ*M* and 1%(*w*/*v*), respectively. The activity of a sample that did not contain an additive was normalized to 100%.

2.6. Docking and multiple sequence alignment

As efforts to crystallize a CelTΔdoc–substrate complex were unsuccessful, molecular docking was performed to find the most likely substrate-binding site. The online server Hex at http://www.csd.abdn.ac.uk/hex_server/ was used for docking (Macindoe *et al.*, 2010). The coordinates of the substrate cellobiose were taken from the Cel9M complex (Parsiegla *et al.*, 2002) and those of cellotetraose from the CbhA complex (Schubot *et al.*, 2004). Docking was carried out for ten runs.

Multiple sequence alignment was performed using *ClustalW* (Thompson *et al.*, 1994) and *BoxShade* (http://sourceforge.net/projects/boxshade/) from the SDSC Biology Workbench (http://workbench.sdsc.edu; Subramaniam, 1998). All figures containing structures were prepared with *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Overall structure of CelTΔdoc

C. thermocellum CelT contains a catalytic domain and a dockerin domain. Because the N-terminal region of CelT is rich in hydrophobic residues and the C-terminal dockerin domain is mostly disordered, we prepared CelTΔdoc, which was missing the first 21 residues and the final 79 residues of CelT. In the CelTΔdoc crystal structure, the positions of the first 16 residues and the C-terminal residue could not be determined because their electron densities were poor.

CelTΔdoc has a similar topology to those of other family 9 glycoside hydrolases; it has a flat cone shape with the presumed substrate-binding site at its base. The overall fold of CelTΔdoc and other family 9 glycoside hydrolases is an $(\alpha/\alpha)_6$-barrel. CelTΔdoc contains 14 long α-helices and five short α-helices (Figs. 1 and 2*a*). 12 of the α-helices form the central $(\alpha/\alpha)_6$-barrel (Figs. 2*a* and 2*b*). The structure also contains five antiparallel β-strands (Figs. 1 and 2*a*).

Several bacterial family 9 glycoside hydrolase structures have been determined (Parsiegla *et al.*, 2002; Schubot *et al.*, 2004; Sakon *et al.*, 1997; Juy *et al.*, 1992). CelT is unlike other *C. thermocellum* family 9 glycoside hydrolases as it lacks the family 3c CBM that binds crystalline cellulose. Although CelT does not have a CBM or an immunoglobulin-like domain, its activity is comparable to those of other family 9 glycoside hydrolases (Kurokawa *et al.*, 2002). The activity data obtained in our study agreed with these results (Table 3). The overall structure of the CelTΔdoc catalytic domain is similar to those of other family 9 glycoside hydrolases, *e.g.* Cel9D and Cel9M, in the absence of substrate (Juy *et al.*, 1992; Parsiegla *et al.*, 2002). The presence or absence of a Trp near the substrate-binding site has been suggested to determine whether a glycoside hydrolase acts processively or nonprocessively, respectively (Parsiegla *et al.*, 2002). Trp209 is found near the active site in Cel9A, a processive cellulase, whereas the corresponding residue is Thr207 in the nonprocessive cellulase Cel9M and Ser276 in CelTΔdoc (Fig. 1). Thus, the absence of

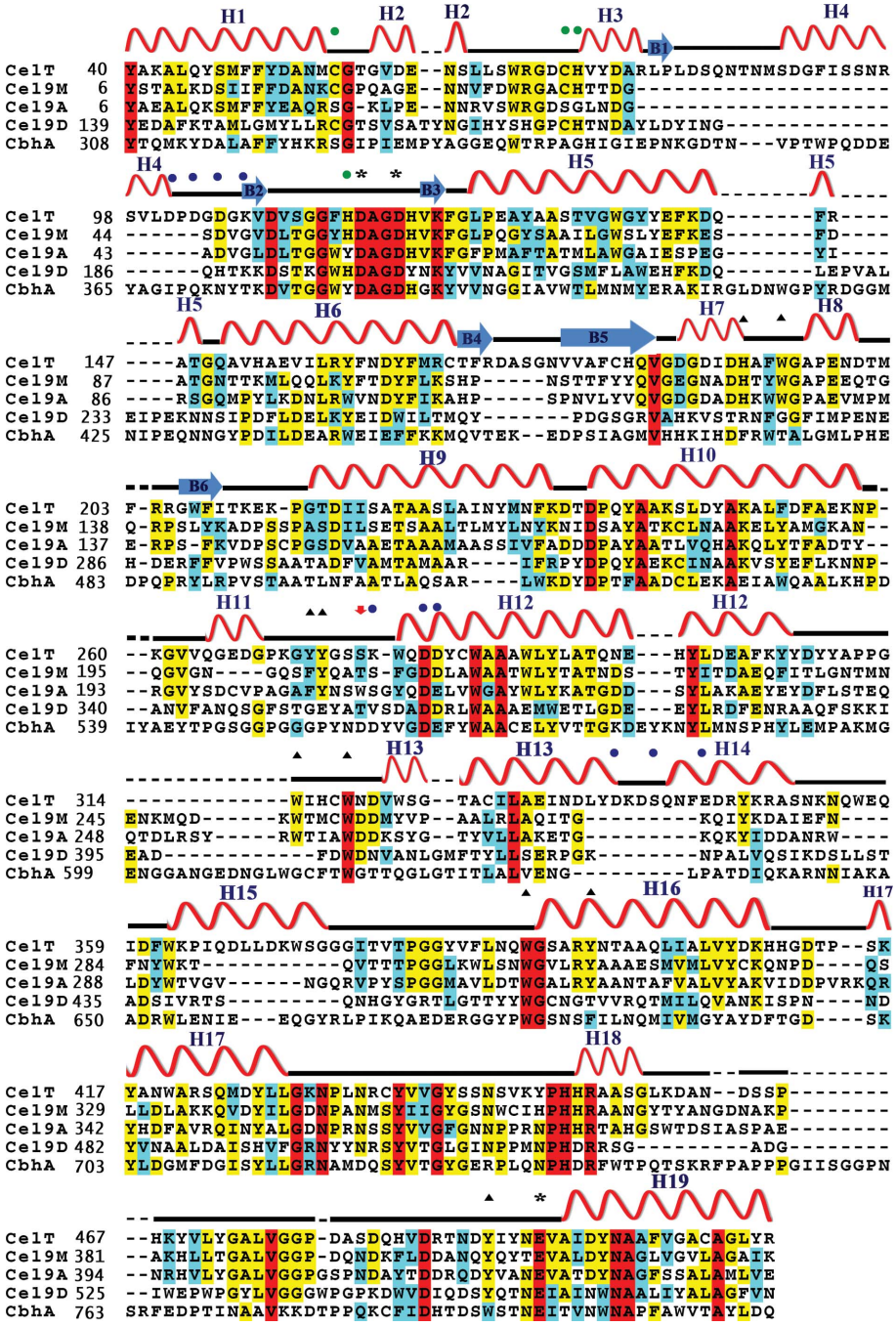

Figure 1
Multiple sequence alignment of the catalytic domains of five family 9 cellulases: *C. thermocellum* CelT, Cel9D and CbhA, *C. cellulolyticum* Cel9M and *T. fusca* Cel9A. α-Helices, β-strands and loops in CelTΔdoc are indicated in red, blue and black, respectively, above the sequences. The catalytic residues are marked by asterisks and the conserved aromatic residues in the active site are marked by triangles. The cysteines and histidines that bind Zn^{+} are indicated by green dots and the residues that interact with the Ca^{2+} ions are indicated by blue dots. The position of the residue that differentiates processive and nonprocessive glycoside hydrolases is identified by an arrow.

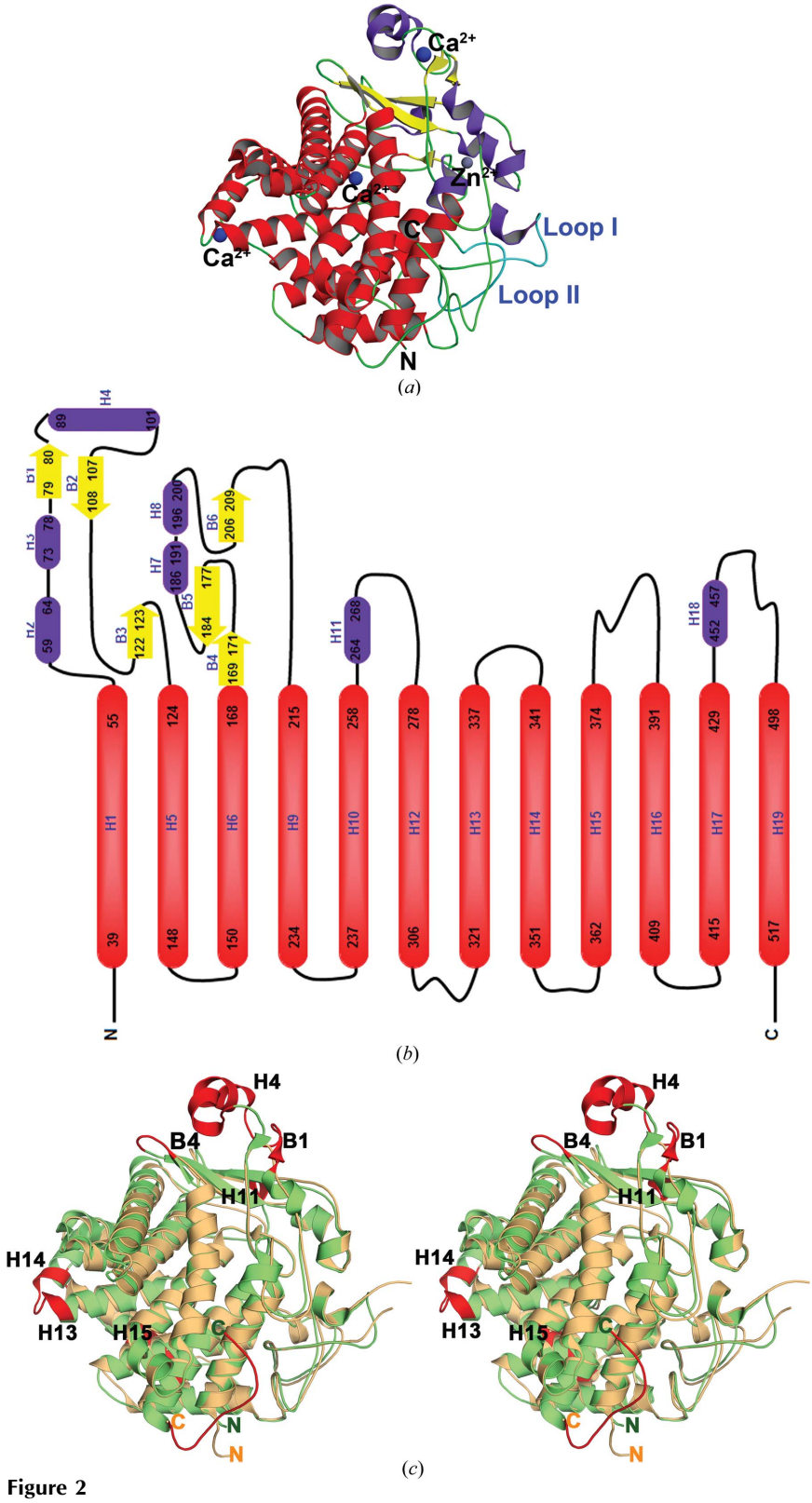

Figure 2
(*a*) The CelTΔdoc structure. The α-helices forming the $(\alpha/\alpha)_6$ barrel, the rest of the helices, the β-strands, the loops and the two flexible loops are shown in red, purple, yellow, green and cyan, respectively. The Zn^{2+} and Ca^{2+} ions are shown as grey and blue spheres, respectively. The flexible loops and N- and C-termini are labelled. (*b*) Schematic representation of the topology of the *C. thermocellum* family 9 glycoside hydrolase CelTΔdoc. The helices and sheets are represented in the same colours as in (*a*). (*c*) Stereoview of the superpositioned CelTΔdoc (pale green) and apo Cel9M (pale orange; PDB entry 1ia6) catalytic domains. Additional CelTΔdoc residues involved in α-helices and β-strands are coloured red.

this tryptophan in CelTΔdoc suggests that this enzyme might also be a nonprocessive cellulase.

3.2. Structural comparison of CelTΔdoc and the family 9 cellulase Cel9M

Although the overall structure of CelTΔdoc resembles those of several other family 9 glycoside hydrolases, we will only compare it with the catalytic domain of Cel9M because these two cellulases have the greatest sequence identity (37%). CelTΔdoc and Cel9M share extensive structural similarities (Table 4), with a few notable exceptions as described below (Fig. 2*c*). The N-terminal region of CelTΔdoc contains an additional α-helix (H4), β-strand (B1) and loop (residues 78–103). An addition to B4 and a loop composed of residues 171–175 are also only found in CelTΔdoc (Fig. 2*c*). The additional residues (171–175) extend the downstream loop and this loop is proximal to H4. Four residues (265–268) form the short H11 helix, which is not found in other cellulases, and residues 335–343 extend H13 and H14 by three residues each (335–337 and 341–343) and a short loop (338–340) connects these two helices (Fig. 2*c*). Residues 355–358 form an additional short loop and residues 365–374 form H15, which is not found in Cel9M. Two additional residues contribute to the loop that is downstream of H17. CelTΔdoc also has an additional loop formed by residues 517–529 at the C-terminus. As most of these additional residues are on the surface and not near the active-site cleft, they are probably not involved in catalysis, but may substitute functionally for a CBM or contribute to the stability of CelTΔdoc. As the H4 helix is found only in CelTΔdoc, and not in Cel9A, Cel9D and CbhA, we speculate that the H4 helix may play some other role if it is not involved in substrate binding. The possible role of the H4 helix will be discussed later.

The three catalytic residues found in Cel9M (Asp56, Asp59 and Glu410) and other family 9 glycoside hydrolases are also found in CelTΔdoc (Asp116, Asp119 and Glu496; Fig. 1). The orientations of Asp116 and Asp119 in CelTΔdoc are identical to those of Asp56 and Asp59 in Cel9M, whereas the orientation of Glu496 in CelTΔdoc differs slightly from that of Glu410 in free Cel9M (Fig. 3*a*). In the

Table 3
Enzyme activities of family 9 glycoside hydrolases.

Enzyme	Optimum pH	Optimum temperature (K)	Specific activity (IU mg^{-1})
C. thermocellum CelT (CtCelT)	6	343	35.3 ± 0.6
C. thermocellum Cel9I (CtCel9I)	6	333	21.5 ± 0.5
C. thermocellum CelD (CtCelD)	3–5	333	51.0 ± 3.6
Nasutitermes takasagoensis endoglucanase (NtEG)	6–8	333	14.4 ± 0.2

Table 4
Results of the *DALI* search (Holm & Rosenström, 2010).

No.	PDB code	*Z*	R.m.s.d. (Å)	Identity (%)	Molecule
1	1ia6, chain *A*	53.3	1.6	42	Cellulase Cel9M
2	1ksd, chain *A*	52.1	1.5	36	Endo-β-1,4-glucanase
3	1kfg, chain *B*	51.7	1.6	41	Endoglucanase G
4	2xfg, chain *A*	51.0	1.7	42	Endoglucanase 1
5	1js4, chain *A*	50.7	1.8	39	Endo/exocellulase E4

Cel9M–substrate complex Glu410 assumes a slightly different orientation which is closer to the substrate.

3.3. Water molecules

Several water molecules are scattered around the active-site cleft. Among these, a water molecule which forms hydrogen bonds to both catalytic aspartates is present in CelTΔdoc, Cel9M and CbhA (Fig. 3*b*). The catalytic residue Glu496 is also in close proximity to the water molecule. Therefore, we suggest that this water molecule may assist in the hydrolysis of substrate in CelTΔdoc and other family 9 glycoside hydrolases. The location of this water in CelTΔdoc is similar to that found in Cel9M. During hydrolysis, Glu496 may be the proton donor and the two aspartates (Asp116 and Asp119) forming hydrogen bonds (2.8 and 3.0 Å) may deprotonate the water *via* nucleophilic attack. Additional waters found near Asp119 might supply water to the active site. A similar water was also observed in the *Alicyclobacillus acidocaldarius* Cel9A structure and it has been suggested that this water helps in catalysis (Pereira *et al.*, 2009).

3.4. Flexible loops at the entry of the substrate-binding cleft

CelTΔdoc has two loops, loop I and loop II, near the substrate-binding cleft. Loop I (residues 457–469) occludes the substrate-binding cleft (Fig. 3*c*) and appears to be flexible because the average *B* factor for the residues in this loop (44.0 Å^2) is larger than that for the remainder of the protein (29.7 Å^2). Flexibility of this loop may be important during substrate binding. A second loop, loop II (residues 480–495), is situated next to loop I. The average *B* factor (39.0 Å^2) for the loop II residues is also greater than the average *B* factor for other regions of the protein. In substrate-free Cel9M loop I (residues 239–249) and loop II (residues 370–378) are considered to be flexible, as their associated electron densities were not clear and their *B* factors were larger than those found when substrate was present (Parsiegla *et al.*, 2002). These loops are present at the ends of the substrate-binding cleft. We were able to build loops I and II in CelTΔdoc without much difficulty (Fig. 3*c*), even though their *B* factors were slightly greater than the average *B* factor for the remainder of the protein. There are two possible reasons for this. Firstly, the sequences of the loops in CelTΔdoc and Cel9M are not identical. Secondly, several hydrogen bonds were observed

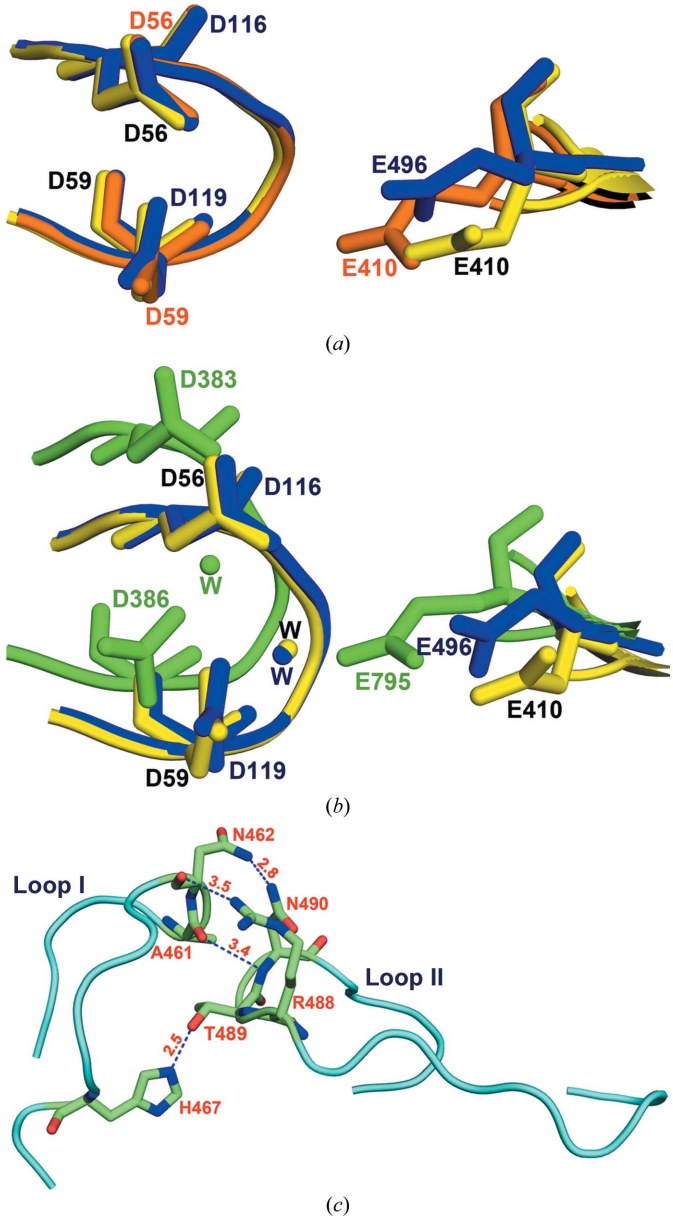

Figure 3
(*a*) The CelTΔdoc and Cel9M active sites. The three catalytic residues in CelTΔdoc (blue), Cel9M (PDB entry 1ia6, yellow) and the Cel9M–substrate complex (PDB entry 1ia7, orange) are shown as stick models and labelled. (*b*) The water within hydrogen-bonding distance of the two catalytic aspartates is shown as a blue, a yellow and a green sphere in CelTΔdoc, Cel9M and CbhA, respectively. The two aspartates and the glutamate involved in catalysis are shown as stick models and are labelled. (*c*) An enlargement of the flexible-loop region (shown in Fig. 2*a*). The loop residues that interact with each other are shown as stick models. Hydrogen bonds are shown as dotted lines and the hydrogen-bond distances are shown above the dotted lines.

between these two loops, including ND2(Asn462)···ND2(Asn490), CO(Asn462)···NH2(Arg488), CO(Ala461)···NH(Asn490) and NH(His467)···OG1(Thr489), with distances of 2.8, 3.5, 3.4 and 2.5 Å, respectively (Fig. 3*c*). These interactions are responsible for keeping these two loops stable.

3.5. Metal ion-binding sites

In all known family 9 glycoside hydrolase structures, a Ca^{2+} ion is present in a cavity near the active site. Ca^{2+} probably plays a structural role and is not involved in catalysis (Parsiegla *et al.*, 2002). Three Ca^{2+} ions were found in CelTΔdoc (Figs. 4*a*, 4*b* and 4*c*). No other family 9 glycoside hydrolase apart from *C. thermocellum* Cel9D contains more than two Ca^{2+} ions. Although three Ca^{2+} ions were also observed in Cel9D, their binding sites are not all the same as those in CelTΔdoc. Site *B* (Fig. 4*b*) is conserved in all family 9 glycoside hydrolases. Site *A* is located in a helix(H13)–loop–helix(H14) region (EF-hand motif), which is a motif that is commonly found in Ca^{2+}-binding proteins. The Ca^{2+} ion in site *C* (Fig. 4*c*) is coordinated to residues from a helix (H4), a β-strand (B2) and a loop, with a coordination number of five. All three Ca^{2+} ions interact with loop residues, which suggests that they may be necessary for stability. In addition to Ca^{2+} ions, a Zn^{2+} ion is found in most family 9 glycoside hydrolases. A Zn^{2+} ion was found in CelTΔdoc (Fig. 4*d*) at a location similar to those in other family 9 glycoside hydrolases, *e.g.* Cel9M and Cel9D. As found in other family 9 glycoside hydrolases, four residues, Cys56, Cys72, His73 and His115, coordinate the Zn^{2+} ion in CelTΔdoc, with coordination distances of 2.3, 2.3, 2.1 and 2.3 Å, respectively (Fig. 4*d*). Apparently, Zn^{2+} stabilizes glycoside hydrolases against heat (Chauvaux *et al.*, 1995). Through its interactions with the residues mentioned above, the Zn^{2+} ion may also stabilize the loops that connect helices H1 and H2, helices H2 and H3 and β-strands B2 and B3 and consequently secondary-structural elements in this region.

3.6. The active site contains many aromatic amino acids

Stacking interactions between the substrate and conserved aromatic residues proximal to the active-site cleft have been noticed in Cel9M and other glycoside hydrolases (Parsiegla *et al.*, 2002; Pereira *et al.*, 2009). CelTΔdoc contains the conserved aromatic residues Trp194, His191, Tyr272, Tyr273, Trp314 and Trp318, which are arranged in the same manner as in free Cel9M, whereas the side chains of Tyr492 and Trp390 occupy the empty substrate-binding site (Fig. 5*a*).

3.7. Docking of cellobiose and cellotetraose into the substrate-binding cleft

Attempts to cocrystallize CelTΔdoc with cellobiose or cellobiose analogues were unsuccessful. Because we could only solve the structure of free CelTΔdoc, we docked cellobiose to CelTΔdoc *in silico* in order to determine whether the binding mode was the same as for cellobiose in Cel9M. The docked complexes revealed that cellobiose bound in the site that corresponded to that in Cel9M (Fig. 5*b*). Model 1 had the lowest score (−217) and the same orientation as in Cel9M. The cellobiose positions in models 3, 4 and 6 were also very similar to the cellobiose-binding mode found in Cel9M and were clustered around the model 1 position. In models 8 and 9 the docked cellobiose was displaced compared with the other

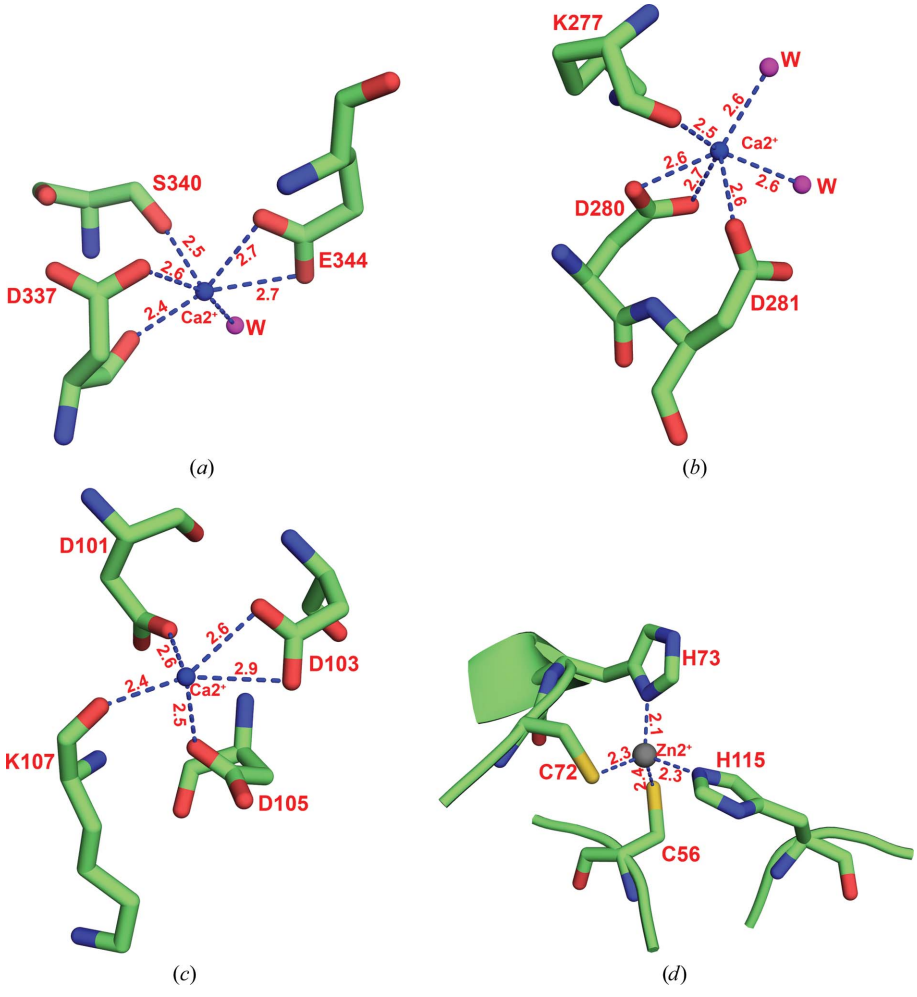

Figure 4
(*a*, *b*, *c*) The three Ca^{2+} ions in CelTΔdoc are shown. Residues that coordinate a Ca^{2+} ion (blue spheres) are shown as stick models and are labelled. Water molecules are shown as small magenta spheres. (*d*) The Zn^{+}-binding site in CelTΔdoc. Residues that coordinate the Zn^{2+} ion (grey sphere) are shown as stick models and are labelled.

models. The different positions found for cellobiose in the substrate-binding cleft strongly suggested that the binding site in CelTΔdoc can accommodate oligosaccharides containing more than two glucose units. Models 2, 5, 7 and 10 contained cellobiose that was docked outside the binding cleft. When cellotetraose was docked (models 4, 5, 8, 9 and 10) two of the glucose units were found in the same position that had been found for cellobiose in Cel9M and the other two units were positioned towards the interior of the active-site cleft (data not shown). Cellotetraose in model 7 was found in a slightly different orientation and was displaced from the major cluster. However, the three best models were not found in the active-site cleft. The energies associated with all of the models were negative, although they were all of the same order of magnitude (−315 for the best model and −239 for the worst model).

Kurokawa *et al.* (2002) assayed the activity of CelTΔdoc and found that oligosaccharides, *e.g.* cellotetraose, cellopentose and cellohexose, were better substrates than cellotriose and cellobiose. In the CelTΔdoc structure, the substrate-binding site appears to be able to accommodate an oligosaccharide of chain length greater than two. Our docking results also suggested that additional glucose units could be accommodated in the presence of cellobiose.

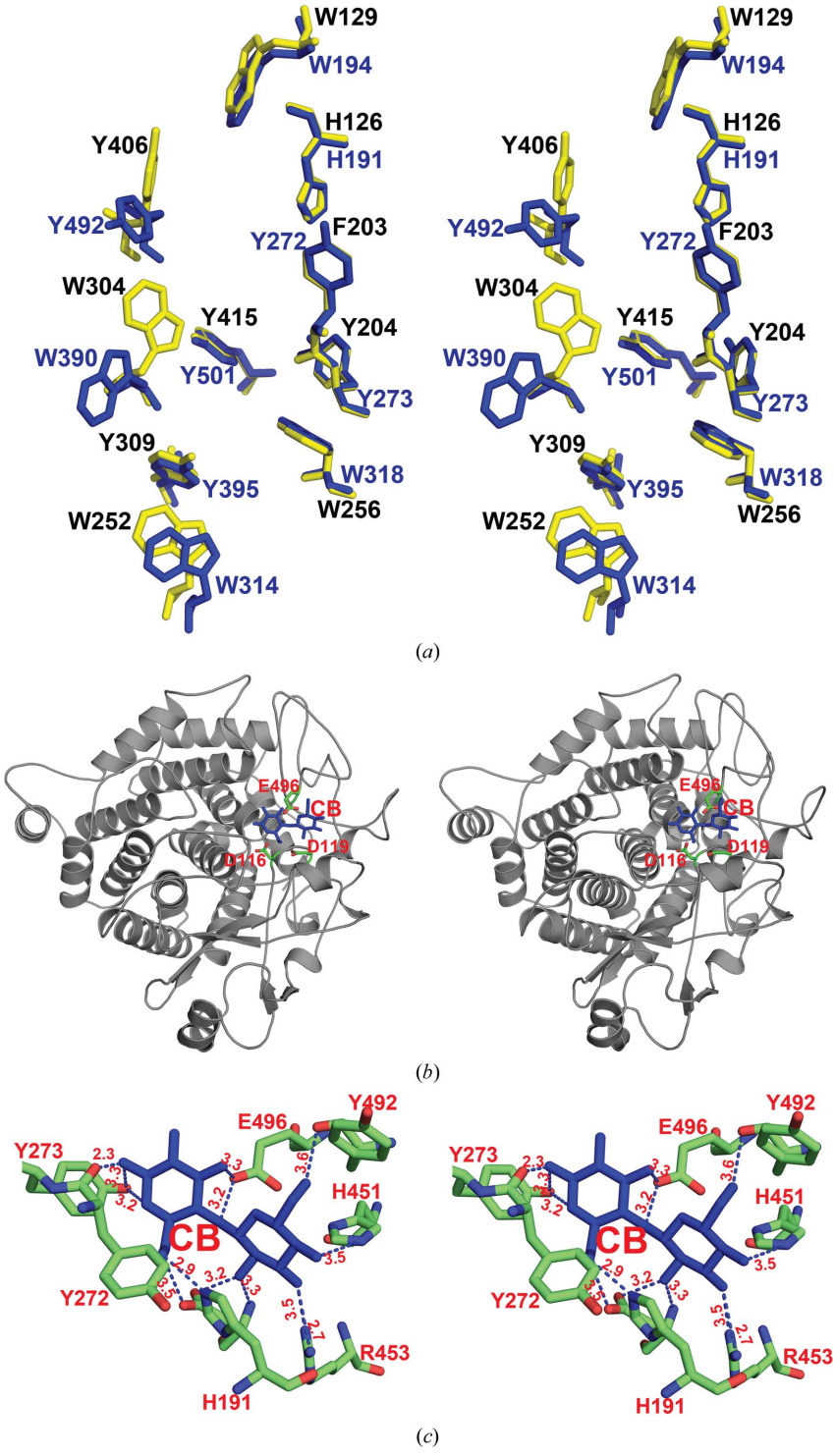

Figure 5
(*a*) Stereoview of the CelTΔdoc substrate-binding cleft. The aromatic residues that line the CelTΔdoc and Cel9M active-site clefts are shown as blue and yellow stick models, respectively, and are labelled. (*b*) Stereoview of the best model of docking; docked cellobiose and the three catalytic residues are shown as blue and green stick models, respectively, and are labelled. (*c*) Stereoview of the enlarged active-site cleft of the docking model (shown in *a*). The residues that interact with the docked substrate are shown as stick models.

3.8. Characteristics of CelTΔdoc and its truncated forms

In support of our hypothesis that the H4 helix and/or the loop on the surface may function like CBM, the following activity studies were performed with deletion mutants of CelTΔdoc. The glycoside hydrolase (CMCase) activities of CelTΔdoc and two deletion mutants (CelTΔ78–103Δdoc and CelTΔ89–101Δdoc) were assayed at different pH values and temperatures. When the CMCase activities of CelTΔdoc and its truncated forms were measured at different temperatures (Fig. 6*a*), CelTΔdoc showed optimal activity at 343 K, but the deletion mutants (CelTΔ89–101Δdoc and CelTΔ78–103Δdoc) failed to show any activity at this temperature, indicating their low thermal stability. These two truncated proteins, CelTΔ89–101Δdoc and CelTΔ78–103Δdoc, showed optimum activity at 323 and 313 K, respectively. A significant change in thermal stability was observed between CelTΔdoc and its truncated mutants. CelTΔdoc and CelTΔ89–101Δdoc showed optimum CMCase activity at pH 6.0 and CelTΔ78–103Δdoc at pH 7.0

(Fig. 6*b*). Under extreme pH conditions, the truncated proteins (CelTΔ78–103Δdoc and CelTΔ89–101Δdoc) were not as stable as CelTΔdoc.

Under the optimal conditions, CelTΔdoc (pH 6, 343 K), CelTΔ89–101Δdoc (pH 6, 323 K) and CelTΔ78–103Δdoc (pH 7, 323 K) showed activities of 35.3 ± 0.6, 33.8 ± 0.2 and 35.3 ± 1.4 IU, respectively. Little significant difference in activities was observed between the three proteins under their optimal conditions, suggesting that the additional H4 helix and the helix with the downstream loop are in fact only essential for stability, not for substrate binding and catalysis.

3.9. Effect of temperature on activity

For glycoside hydrolases to find use in the preparation of cellulose as a biofuel, they must be highly active and stable at a variety of temperatures and pH values. To investigate the thermal stability of CelTΔdoc, activity assays were performed between 303 and 353 K. CelTΔdoc was active over a wide temperature range, with optimal activity at 343 K (Fig. 7*a*). Its thermal stability was therefore similar to those of other *C. thermocellum* family 9 glycoside hydrolases (Fig. 7*a*). Recently, Mingardon *et al.* (2011) suggested that *C. thermocellum* Cel9I and *Thermobifida fusca* Cel9A from thermophilic bacteria are more thermally stable than the mesophilic cellulase Cel9D from *C. cellulolyticum*, which is in agreement with our results, as we also found that all family 9 glycoside hydrolases from *C. thermocellum*, which is a thermophilic bacterium, were more active at higher temperatures (Fig. 7*a*).

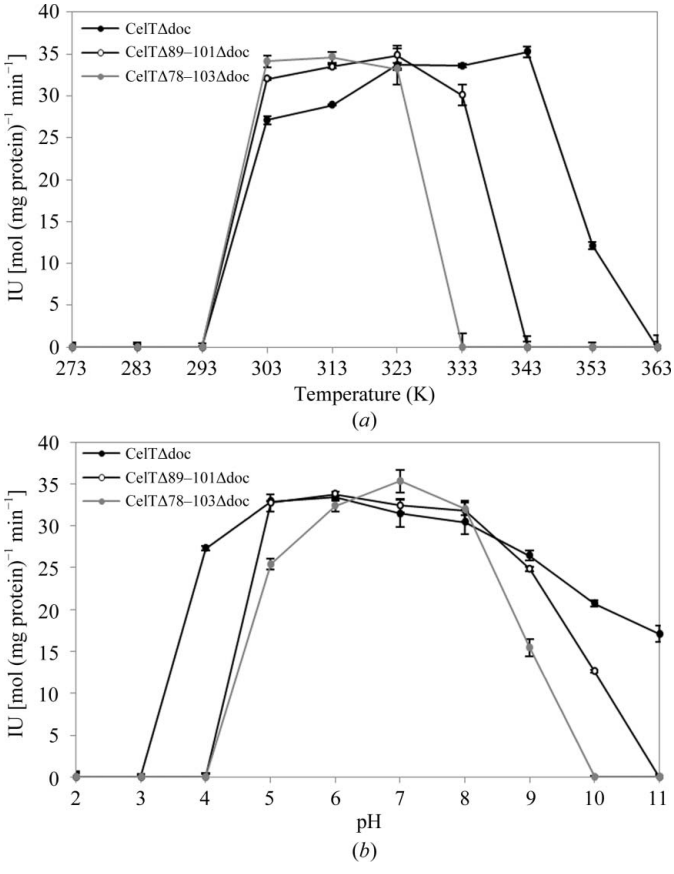

Figure 6
The effect of (*a*) temperature and (*b*) pH on the enzyme activities of CelTΔdoc, CelTΔ78–103Δdoc and CelTΔ89–101Δdoc.

3.10. Effect of pH on activity

The effect of pH on CelTΔdoc was assessed between pH 3 and pH 8. Acidic environments (pH 3.0 and 4.0) diminished the activity, whereas between pH 5 and pH 8 the activity was

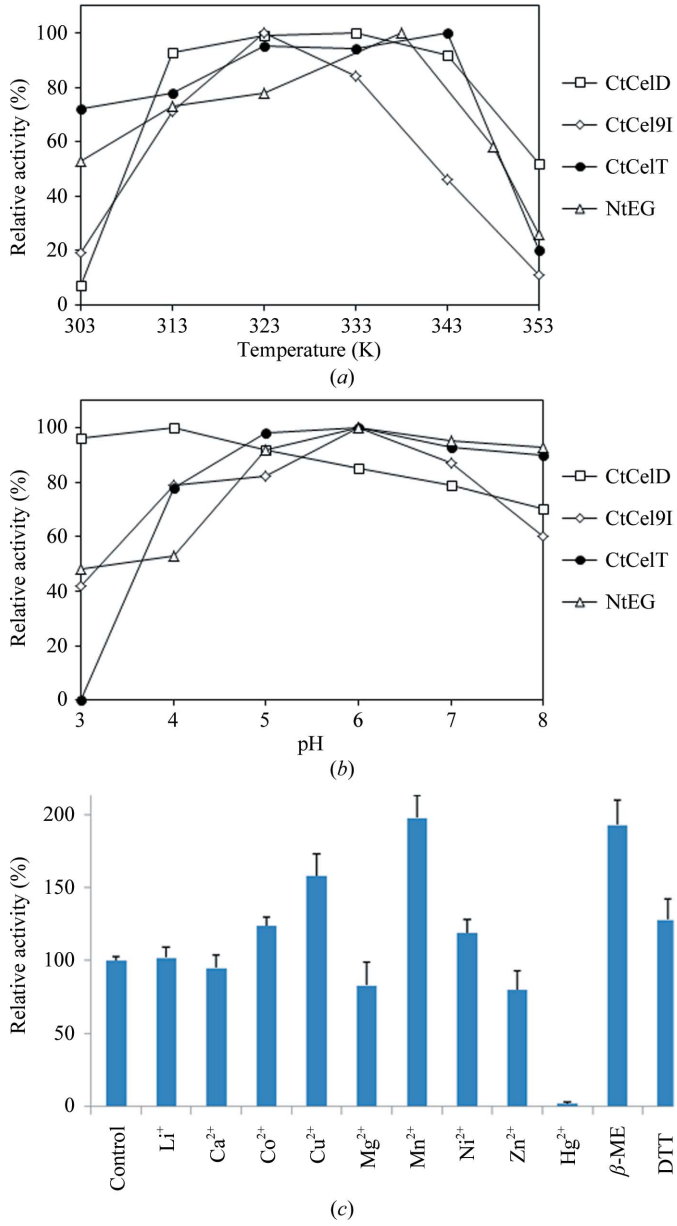

Figure 7
The effect of temperature, pH, metal ions and reducing reagents on cellulase activities. (*a*) The activities of *C. thermocellum* family 9 glycoside hydrolases were measured at various temperatures. The activities of the different glycoside hydrolases are shown with different symbols. (*b*) The activities of *C. thermocellum* family 9 glycoside hydrolases were measured under various pH conditions. (*c*) Effect of metal ions and reducing reagents on CelTΔdoc activity. The activity of the control sample (no additive) was set to 100% and the activities of the other samples were normalized to that of the control. β-ME, β-mercaptoethanol.

between 70 and 100%; the enzyme was most active between pH 6.0 and pH 7.0 (Fig. 7*b*). Most other *C. thermocellum* family 9 glycoside hydrolases have a maximum activity at pH 7.0, with the exception being Cel9D, which has an optimum activity at pH 4.0 (Fig. 7*b*).

The specific activity of CelTΔdoc at its optimum pH and temperature was 35.3 IU, which makes it the second most active *C. thermocellum* family 9 glycoside hydrolase, with only Cel9D being more active at its optimum temperature and pH (Table 3).

3.11. Effect of metals and reducing agents on activity

Enzyme assays were carried out at the pH (6.0) and temperature optima (343 K) of CelTΔdoc in the presence of a metal ion or a reducing agent. Addition of an excess of Ca^{2+} or Zn^{2+} did not increase CelTΔdoc activity even though these two ions were found in the CelTΔdoc crystal structure (Fig. 7*c*). Surprisingly, the addition of Mn^{2+} ions almost doubled the activity. The addition of β-mercaptoethanol also significantly increased the activity. However, the addition of Hg^{2+} ions resulted in inhibition of enzyme activity. Obviously, further structural studies are needed to verify the effect on enzyme activity of different metal ions.

The abovementioned biochemical data suggest that CelTΔdoc is very stable over a wide temperature range and has optimal activity at temperatures higher than most other *C. thermocellum* family 9 glycoside hydrolases.

4. Conclusions

CelT is an unusual *C. thermocellum* family 9 glycoside hydrolase as it does not need accessory module(s) for activity. With the exception of the additional surface residues, the crystal structure of CelTΔdoc is similar to those of other family 9 glycoside hydrolases. Because the H4 helix is near to the active-site cleft and is on the surface, we hypothesized that it might have a function similar to that of a CBM and will thereby help to maintain the catalytic activity of CelT. To our surprise, biochemical studies performed on deletion mutants suggested that the additional H4 helix and/or loop were only needed for stability, not for enzyme activity. Activity assays carried out in our study with different family 9 glycoside hydrolases indicated that CelT is relatively stable at various pH values and temperatures, indicating that it has the qualities that are required for use as a catalyst in biofuel production.

We are grateful for access to beamline BL13C1 at the National Synchrotron Radiation Research Center in Taiwan. This work was supported by Academia Sinica (C-DH) and by a grant from the National Science Council, Taiwan, Republic of China (grant NSC98-2311-B-001-009-MY3 to C-DH).

References

Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Bayer, E. A., Shimon, L. J., Shoham, Y. & Lamed, R. (1998). *J. Struct. Biol.* **124**, 221–234.
Bayer, E. A., Shoham, Y. & Lamed, R. (2006). *The Prokaryotes*, 3rd ed., Vol. 2, edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt, pp. 578–617. New York: Springer-Verlag.
Béguin, P. & Alzari, P. M. (1998). *Biochem. Soc. Trans.* **26**, 178–185.
Béguin, P. & Lemaire, M. (1996). *Crit. Rev. Biochem. Mol. Biol.* **31**, 201–236.
Bélaich, J.-P., Tardif, C., Bélaich, A. & Gaudin, C. (1997). *J. Biotechnol.* **57**, 3–14.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Chauvaux, S., Souchon, H., Alzari, P. M., Chariot, P. & Béguin, P. (1995). *J. Biol. Chem.* **270**, 9757–9762.
Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Demain, A. L., Newcomb, M. & Wu, J. H. D. (2005). *Microbiol. Mol. Biol. Rev.* **69**, 124–154.
Doi, R. H., Goldstein, M., Hashida, S., Park, J.-S. & Takagi, M. (1994). *Crit. Rev. Microbiol.* **20**, 87–93.
Doi, R. H. & Kosugi, A. (2004). *Nature Rev. Microbiol.* **2**, 541–551.
Gal, L., Gaudin, C., Belaich, A., Pages, S., Tardif, C. & Belaich, J.-P. (1997). *J. Bacteriol.* **179**, 6595–6601.
Gilbert, H. J. (2007). *Mol. Microbiol.* **63**, 1568–1576.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Béguin, P. & Aubert, J.-P. (1992). *Nature (London)*, **357**, 89–91.
Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1998). *J. Bacteriol.* **180**, 4303–4308.
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.
Kurokawa, J., Hemjinda, E., Arai, T., Kimura, T., Sakka, K. & Ohmiya, K. (2002). *Appl. Microbiol. Biotechnol.* **59**, 455–461.
Lamed, R., Setter, E. & Bayer, E. A. (1983). *J. Bacteriol.* **156**, 828–836.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D. & Ritchie, D. W. (2010). *Nucleic Acids Res.* **38**, W445–W449.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Miller, G. L. (1959). *Anal. Chem.* **31**, 426–428.
Mingardon, F., Bagert, J. D., Maisonnier, C., Trudeau, D. L. & Arnold, F. H. (2011). *Appl. Environ. Microbiol.* **77**, 1436–1442.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parsiegla, G., Belaïch, A., Belaïch, J.-P. & Haser, R. (2002). *Biochemistry*, **41**, 11134–11142.
Pereira, J. H., Sapra, R., Volponi, J. V., Kozina, C. L., Simmons, B. & Adams, P. D. (2009). *Acta Cryst.* D**65**, 744–750.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pohlschröder, M., Leschine, S. B. & Canale-Parola, E. (1994). *J. Bacteriol.* **176**, 70–76.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.

Schubot, F. D., Kataeva, I. A., Chang, J., Shah, A. K., Ljungdahl, L. G., Rose, J. P. & Wang, B.-C. (2004). *Biochemistry*, **43**, 1163–1170.
Shoham, Y., Lamed, R. & Bayer, E. A. (1999). *Trends Microbiol.* **7**, 275–281.
Subramaniam, S. (1998). *Proteins*, **32**, 1–2.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, H.-M., Shih, Y.-P., Hu, S.-M., Lo, W.-T., Lin, H.-M., Ding, S.-S., Liao, H.-C. & Liang, P.-H. (2009). *Biotechnol. Prog.* **25**, 1582–1586.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.