

# Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker

Svetlana Petkun<sup>1</sup>, Inna Rozman Grinberg<sup>1</sup>, Raphael Lamed<sup>1</sup>, Sadanari Jindou<sup>2</sup>, Tal Burstein<sup>1</sup>, Oren Yaniv<sup>1</sup>, Yuval Shoham<sup>3</sup>, Linda J.W. Shimon<sup>4</sup>, Edward A. Bayer<sup>5</sup> and Felix Frolow<sup>1</sup>

<sup>1</sup> Department of Molecular Microbiology and Biotechnology, The Daniella Rich Institute for Structural Biology, Tel Aviv University, Ramat Aviv, Israel

<sup>2</sup> Department of Life Sciences, Meijo University, Nagoya, Japan

<sup>3</sup> Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa, Israel

<sup>4</sup> Department of Chemical Research Support, The Weizmann Institute of Science, Rehovot, Israel

<sup>5</sup> Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel

## ABSTRACT

Non-cellulosomal processive endoglucanase 9I (Cel9I) from *Clostridium thermocellum* is a modular protein, consisting of a family-9 glycoside hydrolase (GH9) catalytic module and two family-3 carbohydrate-binding modules (CBM3c and CBM3b), separated by linker regions. GH9 does not show cellulase activity when expressed without CBM3c and CBM3b and the presence of the CBM3c was previously shown to be essential for endoglucanase activity. Physical reassociation of independently expressed GH9 and CBM3c modules (containing linker sequences) restored 60–70% of the intact Cel9I endocellulase activity. However, the mechanism responsible for recovery of activity remained unclear. In this work we independently expressed recombinant GH9 and CBM3c with and without their interconnecting linker in *Escherichia coli*. We crystallized and determined the molecular structure of the GH9/linker-CBM3c heterodimer at a resolution of 1.68 Å to understand the functional and structural importance of the mutual spatial orientation of the modules and the role of the interconnecting linker during their re-association. Enzyme activity assays and isothermal titration calorimetry were performed to study and compare the effect of the linker on the re-association. The results indicated that reassembly of the modules could also occur without the linker, albeit with only very low recovery of endoglucanase activity. We propose that the linker regions in the GH9/CBM3c endoglucanases are important for spatial organization and fixation of the modules into functional enzymes.

Submitted 26 May 2015  
Accepted 4 July 2015  
Published 15 September 2015

Corresponding author  
Edward A. Bayer,  
ed.bayer@weizmann.ac.il

Academic editor  
Guo-Qiang Chen

Additional Information and  
Declarations can be found on  
page 20

DOI 10.7717/peerj.1126

© Copyright  
2015 Petkun et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Biotechnology

**Keywords** Family-9 glycoside hydrolase, *Clostridium thermocellum*, Protein–protein interaction, X-ray structure, Carbohydrate-binding module (CBM)

## INTRODUCTION

Cellulose is a major component of the plant cell wall, lending structural stability and resilience to an otherwise flaccid material. The propensity of cellulose to form ordered, tightly packed, para-crystalline fibrils hinders its enzymatic degradation. Indeed, the recalcitrant properties of cellulose are such that numerous enzymes are required to act synergistically in achieving its efficient degradation. Many types of bacteria and fungi are capable of degrading cellulose and other plant cell wall polysaccharides in an effective manner, producing a variety of various cellulases and related enzymes, either existing in the free state, or associated with a multi-enzyme complex known as the cellulosome (Bayer *et al.*, 2004; Bayer *et al.*, 2008; Demain, Newcomb & Wu, 2005; Doi & Kosugi, 2004; Fontes & Gilbert, 2010). *Clostridium thermocellum* is an anaerobic thermophilic bacterium, known for its efficient degradation of cellulose and other plant cell wall polysaccharides (Béguin, Millet & Aubert, 1992; Freier, Mothershed & Wiegel, 1988; Garcia-Martinez *et al.*, 1980; Ng, Weimer & Zeikus, 1977; Wiegel, Mothershed & Puls, 1985). The cellulase system of this bacterium includes a remarkable variety of enzymes, some existing in the free state but most associated with a highly efficient multi-enzyme complex, termed cellulosome (Bayer *et al.*, 2004; Lamed, Setter & Bayer, 1983; Lamed *et al.*, 1983; Shoham, Lamed & Bayer, 1999), capable of converting a wide variety of plant-derived polysaccharides directly into soluble sugars and fermentation products (Béguin & Alzari, 1998; Felix & Ljungdahl, 1993; Schwarz, 2001; Schwarz, Zverlov & Bahl, 2004). These capabilities render *C. thermocellum* a high utility candidate for use in consolidated bioprocessing (CBP) applications (reviewed in Akinosho *et al.*, 2014).

Cellulases are a class of modular enzymes with a catalytic glycoside hydrolase (GH) module that hydrolyzes the  $\beta$ -1,4-glucosidic bond of the cellulose chain (Cantarel *et al.*, 2009; Davies & Henrissat, 1995; Gilbert & Hazlewood, 1993; Henrissat, 1991; Henrissat & Davies, 1997; Wilson & Irwin, 1999). The catalytic module is usually associated with various numbers of accessory modules that serve to modulate the enzyme activity, and the enzymes have been categorized into families according to the amino-acid sequence of the GH domain (Cantarel *et al.*, 2009; Gilkes *et al.*, 1991; Henrissat & Davies, 1997; Henrissat & Davies, 2000; Henrissat & Romeu, 1995). Cellulases have been broadly divided into two types: endoglucanases that can hydrolyze bonds internally in cellulose chain, and exoglucanases that act preferentially on chain ends, progressively cleaving off cellobiose as the main product. The distinction between endo- and exo-acting enzymes is also reflected by the architecture of the respective class of active site, whereby endoglucanases, for example, are commonly characterized by a groove or open binding cleft, into which any part of the linear cellulose chain can fit. On the other hand, the exoglucanases bear tunnel-like active sites, which can only accept a substrate chain via its terminus (either the reducing or non-reducing end, depending on the enzyme), thereby cleaving cellulose in a sequential manner. The sequential hydrolysis of a cellulose chain has earned the term “processivity” (Beckham *et al.*, 2014; Davies & Henrissat, 1995; Wilson & Kostylev, 2012), and processive enzymes are considered to be key components which contribute to the overall efficiency of a given cellulase system. Some endoglucanases, notably from GH family 9, have also

been shown to sequentially hydrolyze cellulose chains and are thus referred to as processive endoglucanases (Gal et al., 1997; Gilad et al., 2003; Irwin et al., 1998; Jeon et al., 2012; Kuusk, Sorlie & Valjamae, 2015; Zverlov, Velikodvorskaya & Schwarz, 2003). Such enzymes appear to possess extended catalytic clefts and the observed processivity appears to require highly coordinated substrate-binding affinities from opposite sides of the cleavage site (Bu et al., 2012; Li, Irwin & Wilson, 2010; Payne et al., 2011).

Cellulase 9I (Cel9I), is a non-cellulosomal family 9 processive endoglucanase from *Clostridium thermocellum*, which degrades crystalline cellulose (Avicel and filter paper) as well as phosphoric acid-swollen cellulose (PASC) and carboxymethyl cellulose (CMC) (Gilad et al., 2003). This enzyme consists of a catalytic GH9 module at its N terminus, followed by two family 3 carbohydrate-binding modules (CBMs): CBM3c and CBM3b. The three modules are separated by distinctive linker sequences. Such intermodular linker segments were proposed to be important for the physical association of the modules in the space, and to promote intermodular and/or intersubunit protein–protein interactions (Bayer et al., 1998; Bayer et al., 2009; Noach et al., 2008).

The C-terminal CBM3b module, as a classic CBM3, is responsible for targeting the Cel9I enzyme to the planar surface of the crystalline cellulose substrate (Gilad et al., 2003; Su, Mackie & Cann, 2012; Tormo et al., 1996). It has also been proposed to disrupt the crystalline regions of cellulose, rendering it more accessible to the GH9 catalytic module (Yi et al., 2013) and to contribute to enzyme processivity by preventing the desorption of the catalytic module from cellulose (Telke et al., 2012). The function of the CBM3c is less straightforward. Removal of CBM3c from *C. thermocellum* Cel9I, *C. cellulolyticum* Cel9G and *P. Barcinonensis* Cel9B significantly reduces the enzyme activity (Burstein et al., 2009; Chiriac et al., 2010; Gal et al., 1997). CBM3c modules have been shown to alter the normal function of the GH9 catalytic module of *Thermobifida fusca* Cel9A from the standard endo-acting mode into a processive endoglucanase (Bayer et al., 1998; Irwin et al., 1998). Thus, Gilad et al., showed in 2003 that the endoglucanase activity of Cel9I is dependent upon the presence of the CBM3c module and suggested that the fused CBM3c serves an important accessory role for the catalytic domain by altering its character to facilitate processive cleavage of recalcitrant cellulose substrates.

In addition to the Cel9 CBM3c, several other examples of CBMs that are considered to modulate catalytic specificity and act cooperatively with the catalytic domain have recently been discovered. These include CBM66 that directs the cognate enzyme towards highly branched glucans rather than linear fructose polymers (Cuskin et al., 2012), CBM48 that contributes to substrate binding at the active site of a glucan phosphatase (Meekins et al., 2014), family-43  $\beta$ -xylosidases where the GH43 is complemented by an additional module that confers hydrolytic activity to the mature enzyme (Moraïs et al., 2012), and CBM46, that constitutes part of the catalytic cleft required for the hydrolysis of  $\beta$ -1,3-1,4-glucans (Venditto et al., 2015). The carbohydrate-binding PA14 domain is also known to affect substrate binding of the catalytic domain by contributing to the formation of its active site (Gruninger et al., 2014; Zmudka, Thoden & Holden, 2013).

We have previously shown that independently expressed GH9 and linker-containing CBM3c modules of Cel9I readily re-associate *in vitro* and that this physical reassociation recovers 60–70% of the intact Cel9I endoglucanase activity (Burstein *et al.*, 2009).

We have examined in this work the interaction of the CBM3c with the catalytic module either with or without the intermodular linker in order to better understand the function of the CBM3c in the family-9 enzymes and the role of the linker regions. The effect of the re-association of the CBM3c with linker (CBM3cL) and the CBM3c without linker (CBM3cNL) on the enzymatic activity of GH9 has been studied by the crystallization and structure determination of the reassembled GH9-CBM3cL complex at a resolution of 1.68 Å. The results of this study will help us to understand the contribution of ancillary modules in the action of multi-modular glycoside hydrolases.

## MATERIALS AND METHODS

### Cloning of the GH9, CBM3cL and CBM3cNL proteins

Cloning of the DNA fragments encoding the C-terminally His-tagged CBM3c with the linker and the untagged GH9 module from Cel9I of *C. thermocellum* (GenBank accession code L04735) was described earlier (Burstein *et al.*, 2009; Gilad *et al.*, 2003). C-terminally His-tagged CBM3c without the linker connecting it to the GH9 was amplified using the same procedure and the following primers: F′-5′CCATGGGCGAAGTTCGGAGGATGAAATA and R′-5′CTCGAGCGGTTCCCTTCCAAATACCAG. The PCR products were purified and cleaved with restriction enzymes *NcoI* and *XhoI* and inserted into the pET-28a(+) expression vector (Novagen, Madison, WI, USA).

### Expression and purification of recombinant proteins

The GH9 and CBM3c modules both with (GH9L, CBM3cL) and without (GH9NL and CBM3cNL) the linker regions were expressed independently by the identical expression procedure. *Escherichia coli* strain BL21(DE3)RIL harboring the plasmids was aerated at 310 K in 3-liters Terrific Broth supplemented with 25 mg ml<sup>-1</sup> kanamycin. After 3 h, the culture reached an A<sub>600</sub> of 0.6; 0.1 mM isopropyl-β-D-1-thiogalactopyranoside was added to induce gene expression, and cultivation was continued at 310 K for an additional 12 h. Cells were harvested by centrifugation (5,000 × g for 15 min) at 277 K and were subsequently re-suspended in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, containing 300 mM NaCl at a ratio of 1 g wet pellet to 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 20 min), and the supernatant was collected.

The soluble expressed His-tagged CBM3c modules with or without the linker, according to the type of the experiment, were applied batchwise to Ni-IDA resin during 1-h incubation with gentle stirring at 4 °C. Non-specifically bound proteins were washed with a buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 6, 300 mM NaCl, 10% glycerol and 10 mM imidazole. Crude extract supernatant fluids, containing the expressed GH9 module, were added to the CBM3c-bound Ni-IDA resin, and the mixture was incubated overnight with gentle stirring at 4 °C. The adsorbed protein complexes were eluted with 300 mM imidazole and

subjected to further purification by size-exclusion chromatography. Fast protein liquid chromatography (FPLC) was performed using a Superdex 75pg column and ÄKTA Prime system (GE Healthcare, Piscataway, New Jersey, USA) to further purify the complex. One peak, corresponding approximately to 70 kDa, matching the predicted molecular weight of the GH9-CBM3c complex, was observed in the chromatogram. The 15 amino-acid linker sequence (about 1.5 kDa) did not significantly affect the elution volume, compared to that of the complex without the linker, presumably due to the limited resolution of the column. The relevant fractions (the purified complexed proteins) were analyzed by 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie brilliant blue staining. Two clear bands, of about 52 and 19.5 kDa were observed. The rearranged modules were concentrated to  $6 \text{ mg ml}^{-1}$  using Centriprep YM-3 centrifugal filter devices YM-3 (Amicon Bioseparation, Millipore Corporation, Bedford, Massachusetts, USA). Protein concentration was determined by measuring UV absorbance at 280 nm.

The full-length Cel9I was purified by affinity chromatography on Avicel as reported earlier ([Burstein et al., 2009](#); [Gilad et al., 2003](#)).

### Microcalorimetric analysis

Isothermal titration calorimetry (ITC) experiments were carried out using a VP-ITC MicroCalorimeter (MicroCal, LLC, Northampton, Massachusetts, USA) at 298 K. About  $300 \mu\text{M}$  solution of CBM3cNL was injected into a  $65 \mu\text{M}$  solution of GH9. The reaction was performed in a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% sodium azide. Heats of dilution of the titrants were subtracted from the titration data, and the corrected data were analyzed using the Origin ITC analysis software package supplied by MicroCal. Thermal titration data were fit to the one binding site model, and enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ), association constant ( $K_a$ ) and stoichiometry of binding ( $N$ ) were determined. In all cases, the calculated stoichiometry ( $N$ ) was lower than one, most likely due to the fact that the CBM3 proteins lost their native functionality with time. For the analysis, the CBM3 protein concentrations were corrected as to provide a stoichiometry of one. Two titrations were performed to evaluate reproducibility.

### Enzyme activity assay

Reactions were performed at 333 K, in 50 mM citrate buffer (pH 6.0). The soluble cellulosytic substrate was carboxymethyl cellulose (CMC, Sigma Chem. Co. St. Louis, Missouri, USA). The amount of reducing sugars released from the substrate was determined with the 3,5-dinitrosalicylic acid (DNS) reagent as described by Miller et al. ([Miller, 1959](#)). Activity was defined as the amount (micromole) of reducing sugar released after 10 min of reaction.

### Crystallization

Initially the protein samples containing 6 mg/ml protein solution in 1.2 mM Tris-HCl pH 7.5, 1.5 mM sodium chloride, 0.025% sodium azide, were screened, using the microbatch crystallization method under 1:1 mixture of silicon and paraffin oil ([Chayen et al., 1990](#)), using 288 conditions from the Hampton Research HT screens (SaltRx, Index HT, and Crystal Screen HT; Hampton Research, Aliso Viejo, California, USA) and 96 conditions



of the Wizard Crystallization kit from Emerald BioSystems (Rigaku Reagents, Bainbridge Island, Washington, USA). The dyad of GH9 and CBM3cNL did not yield any crystals. Screening of the GH9-CBM3cL resulted in plate-like crystals that appeared after several days under several conditions, all of which contained PEG 3350 and 4000. The best crystals were obtained in 30% PEG (both 3350 and 4000), 0.2 M magnesium chloride, and 0.1 M HEPES, pH 7.5. Attempts to optimize this condition using microbatch, hanging-drop, and sitting drop methods were unsuccessful, as the crystals remained very thin and fragile. The superfine Eyelash (Ted Pella, Inc, Redding, California, USA) was used to touch these crystals and consequently to streak the sitting drops, composed of 5  $\mu$ l of the protein solution and 5  $\mu$ l of the precipitating solution (24% PEG 3350, 0.2 M magnesium chloride, 0.1 M HEPES, pH 8.0). After one day, crystals of different morphology, with maximum size of about 0.05 mm, appeared in the drop.

### Data collection and crystallographic analysis

The crystals of the GH9-CBM3cL complex were harvested from the crystallization drop using a nylon cryo-loop (Hampton Research, Aliso Viejo, California, USA). For data collection, crystals were mounted on the MiTeGen stiff micro-mount (MiTeGen, Ithaca, New York, USA) made of polyimide and flash-cooled in a nitrogen stream produced by Oxford Cryostream low temperature generator (Cosier & Glazer, 1986) at a temperature of 100 K. Mother-liquor of the crystals served for cryo-protection during the cooling in liquid nitrogen.

Diffraction data from the GH9-CBM3cL crystals were measured using the ID23-2 beam line at ESRF, Grenoble, France. A MAR CCD 225 area detector and X-ray radiation of 0.873 Å wavelength were used. Diffraction data of 480 images with 0.5° oscillation per image were collected. Data were processed with DENZO and scaled with SCALEPACK as implemented in HKL2000 (Otwinowski & Minor, 1997). The crystals diffracted to 1.68 Å resolution and belong to the orthorhombic space group  $P2_12_12_1$ , with unit cell parameters  $a = 70.4$ ,  $b = 88.5$ ,  $c = 106.5$  Å. There is one GH9-CBM3cL complex per asymmetric unit with a Matthews density  $V_M$  of 2.37 Å<sup>3</sup> Da<sup>-1</sup>, corresponding to a solvent content of 48.15% (Matthews, 1968). The X-ray data analysis statistics are presented in Table 1 (Stout & Jensen, 1968).

Molecular replacement was carried out with MOLREP (Vagin & Teplyakov, 1997), using the coordinates of the GH9 and CBM3c modules of endoglucanase 9G from *Clostridium cellulolyticum* (PDB code 1G87, 66 and 51% sequence identity, respectively), as a search model. The MOLREP calculations with the GH9 domain converged into a clear solution with 1 molecule in the asymmetric unit with an R-factor of 0.533 and correlation coefficient of 0.567. This solution was inserted into MOLREP calculations as a fixed molecule and the coordinates of CBM3c module were used for the search producing a solution with an  $R_{\text{cryst}}$  of 0.505, and correlation coefficient of 0.582. The resulting model with 5% of reflections forming test set (Brünger, 1992) was subjected to 10 cycles of restrained refinement using anisotropic B-factors, yielding the  $R_{\text{cryst}}$  and  $R_{\text{free}}$  0.329 and 0.359, respectively (REFMAC5) (Murshudov, Vagin & Dodson, 1997). Automated model

**Table 1** Diffraction data of the GH9-CBM3c *in vitro* reassembled complex from Cel9I from *C. thermocellum*. Values shown in parentheses are for the highest resolution cell.

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Number of crystals	1
Total rotation range (°)	240
<i>a</i> (Å)	70.39
<i>b</i> (Å)	88.54
<i>c</i> (Å)	106.49
<i>V</i> (Å <sup>3</sup> )	663,743.40
Resolution range (Å)	30–1.68 (1.71–1.68)
Total number of reflections	676,571
Unique reflections	76,727
Mosaicity range (°)	0.18–0.46
Average redundancy	9.0
Completeness, overall (%)	97.9 (74.8)
Mean <i>I</i> /σ( <i>I</i> )	34.72 (2.08)
<i>R</i> <sub>merge</sub> <sup>a</sup> (%)	7.4 (49.8)

**Notes.**

<sup>a</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $\sum_{hkl}$  denotes the sum over all reflections and  $\sum_i$  the sum over all equivalent and symmetry-related reflections.

building by ARP/wARP (Perrakis, Morris & Lamzin, 1999) produced a complete structure with  $R_{\text{cryst}}$  and  $R_{\text{free}}$  of 0.218 and 0.243, respectively. The model was manually corrected using COOT (Emsley & Cowtan, 2004) and refined using REFMAC5 (Murshudov, Vagin & Dodson, 1997). The  $R_{\text{cryst}}$  and  $R_{\text{free}}$  improved to 0.184 and 0.228, respectively. Solvent atoms were built using ARP/warp (Perrakis, Morris & Lamzin, 1999). Refinement of TLS (rigid body translation/libration/screw motions) parameters was performed (Winn, Isupov & Murshudov, 2001; Winn, Murshudov & Papiz, 2003). The model was subjected to several additional cycles of manual rebuilding and refinement. The model converged to final  $R_{\text{cryst}}$  and  $R_{\text{free}}$  factors of 0.144 and 0.176, respectively.

The refinement statistics of the structure are summarized in Table 2. The structure was validated using MolProbity (Davis et al., 2007).

**Protein sequence analysis**

Sequence alignments were performed using CLUSTALW (Larkin et al., 2007) and the coloring of residues (representing degree of conservation) using ProtSkin (Deprez et al., 2005). Sources of the sequences used in this work are as follows: *Clostridium thermocellum* Cel9I GH9 module, CBM3c and CBM3b (AAA20892.1); *Clostridium cellulolyticum* Cel9G GH9 module, CBM3c (AAA73868.1); *Thermobifida fusca* Cel9A GH9 module and CBM3c (AAB42155.1); *Cellulomonas fimi* Ce9A CBM3c (AAA23086.1); *Clostridium cellulovorans* EngH CBM3c (AAC38572.2) and CbpA CBM3a (AAA23218.1); *Clostridium stercorarium* CelZ CBM3c and CBM3b (CAA39010.1) and CelY CBM3b (CAA93280.1); *Clostridium thermocellum* CipA CBM3a (CAA48312.1), CelQ CBM3c (BAB33148.1), Cel9V CBM3c' and CBM3b' (CAK22315.1), Cel9U CBM3c' and CBM3b' (CAK22317.1) and Cbh9A CBM3b (CAA56918.1); *Clostridium cellulolyticum* CipC CBM3a (AAC28899.2) and CelJ

**Table 2** Refinement statistics and results of *MolProbity* validation for reassembled GH9-CBM3c complex.

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range	30–1.68
No. of reflections in working set	71,559
No. of reflections in test set	3,580
No. of protein atoms	5,071
No. of solvent atoms	835
No. of Cl ion atoms	3
No. of Ca ion atoms	2
Overall B factor from Wilson plot (Å <sup>2</sup> )	16.06
Averaged B factor (Å <sup>2</sup> )	21.12
<i>R</i> <sub>cryst</sub>	0.1441
<i>R</i> <sub>free</sub>	0.1759
<b>Geometry</b>	
RMS bonds (Å)	0.014
RMS bond angles (°)	1.371
<b>MolProbity validation</b>	
Ramachandran favored (%) (goal > 98%)	96.7
Ramachandran outliers (%) (goal < 0.2%)	0.5
<i>C</i> <sub>β</sub> deviations > 0.25 Å (goal 0)	1
Clash score <sup>a</sup> (all atoms)	2.88
Rotamer outliers (%) (goal < 1%)	0.8
Residues with bad bonds (%) (goal < 1%)	0.00
Residues with bad angles (%) (goal < 0.5)	0.33

**Notes.**

<sup>a</sup> Clash score is the number of serious steric overlaps (>0.4 Å) per 1,000 atoms.

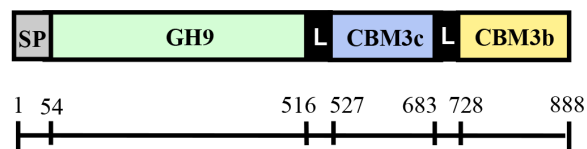
CBM3c (AAG45158.1); *Acetivibrio cellulolyticus* Cel9B CBM3c' and CBM3b' (CAI94607.1) and CipV (ScaA) CBM3b (AAF06064.1); *Clostridium josui* CipA (CipJ) CBM3a (BAA32429.1); *Bacteroides cellulosolvens* ScaA CBM3b (AAG01230.2); *Bacillus subtilis* CelA CBM3b (AAA22307.1); *Pectobacterium atrosepticum* CelVI CBM3b (X79241.2); *Bacillus licheniformis* CelA CBM3b (CAJ70714.1).

**RESULTS****Cloning, expression and purification of Cel9I and its modular components**

The full-length *C. thermocellum* Cel9I enzyme and its individual component parts were over-expressed in *Escherichia coli*, according to *Burstein et al. (2009)*, in order to investigate the contribution of the ancillary modules and their linkers to the catalytic activity of the enzyme. These include the isolated GH9 module with and without a His tag, the His-tagged CBM3c module together with its adjacent N-terminal linker that connects it to the GH9 module (CBM3cL) and His-tagged CBM3c module without the N-terminal linker (CBM3cNL). For details, see [Fig. 1](#). Following purification procedures, all recombinant proteins showed a single band in SDS-PAGE of the anticipated molecular masses.



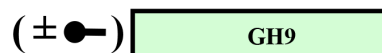
Cel9I gene product



(a) Cel9I (full-length)



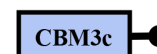
(b) GH9



(c) CBM3cL (with linker)



(d) CBM3cNL (no linker)



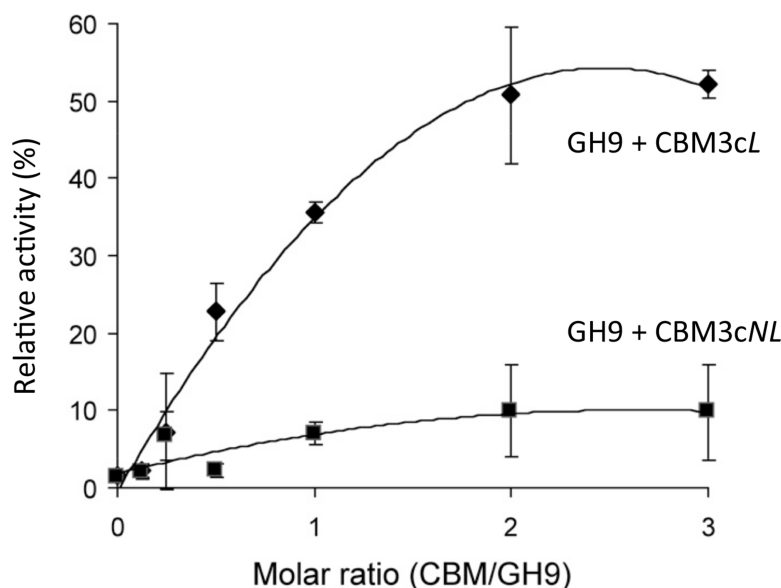
**Figure 1** Schematic diagram of the Cel9I gene product and the recombinant proteins (A–D) prepared for this study. The GH9 module alone (B) was prepared with and without an N-terminal His tag (shown schematically in the figure), and the CBM3c's were prepared with C-terminal His tags. Scale shows the number of amino acid residues and the boundaries of the different regions of the protein.

### Recovery of endoglucanase activity upon association of CBM3cNL and GH9 compared to CBM3cL and GH9

Previous works (*Burstein et al., 2009*; *Gilad et al., 2003*) demonstrated that the Cel9I catalytic module alone has no detectable activity on CMC (carboxymethyl cellulose) and that adding the CBM3cL to form the Cel9I-CBM3cL dyad serves to recover up to 70% of the lost activity. To further examine the importance of the linker connecting the GH9 and the CBM3c modules, we tested the ability of CBM3cNL to recover the CMCase activity of GH9. A fixed amount of the catalytic module (70 pmol in 400  $\mu$ l) was mixed with increasing amounts of CBM3cL or CBM3cNL. The activity of the intact Cel9I enzyme was defined as 100%, and the activity of the reconstituted complexes was measured relative to that of Cel9I. The results indicated that GH9-CBM3cNL exhibit only about 10% of the intact Cel9I activity towards CMC, whereas the reassembled GH9-CBM3cL provided up to 50% of the activity (*Fig. 2*). The fact that a higher than one molar ratio was required to obtain maximum activity can be explained by the fact that the CBM protein was only partly functional as was also observed in the ITC experiments described below. Overall, the results suggest that the linker is required for better fitting of the reconstituted CBM3c which results in better recovered activity.

### Overall structure of the reassembled GH9-CBM3c

Initially, we tried to overexpress and purify the covalently linked GH9-CBM3c, however the full-length protein was unstable and proteolysis occurred during the overexpression and purification stages, partially resulting in separate GH9 and CBM3c modules. Therefore, the obtained protein samples were not homogenous and were not suitable for crystallization trials. Instead, we employed an alternative approach where we expressed the



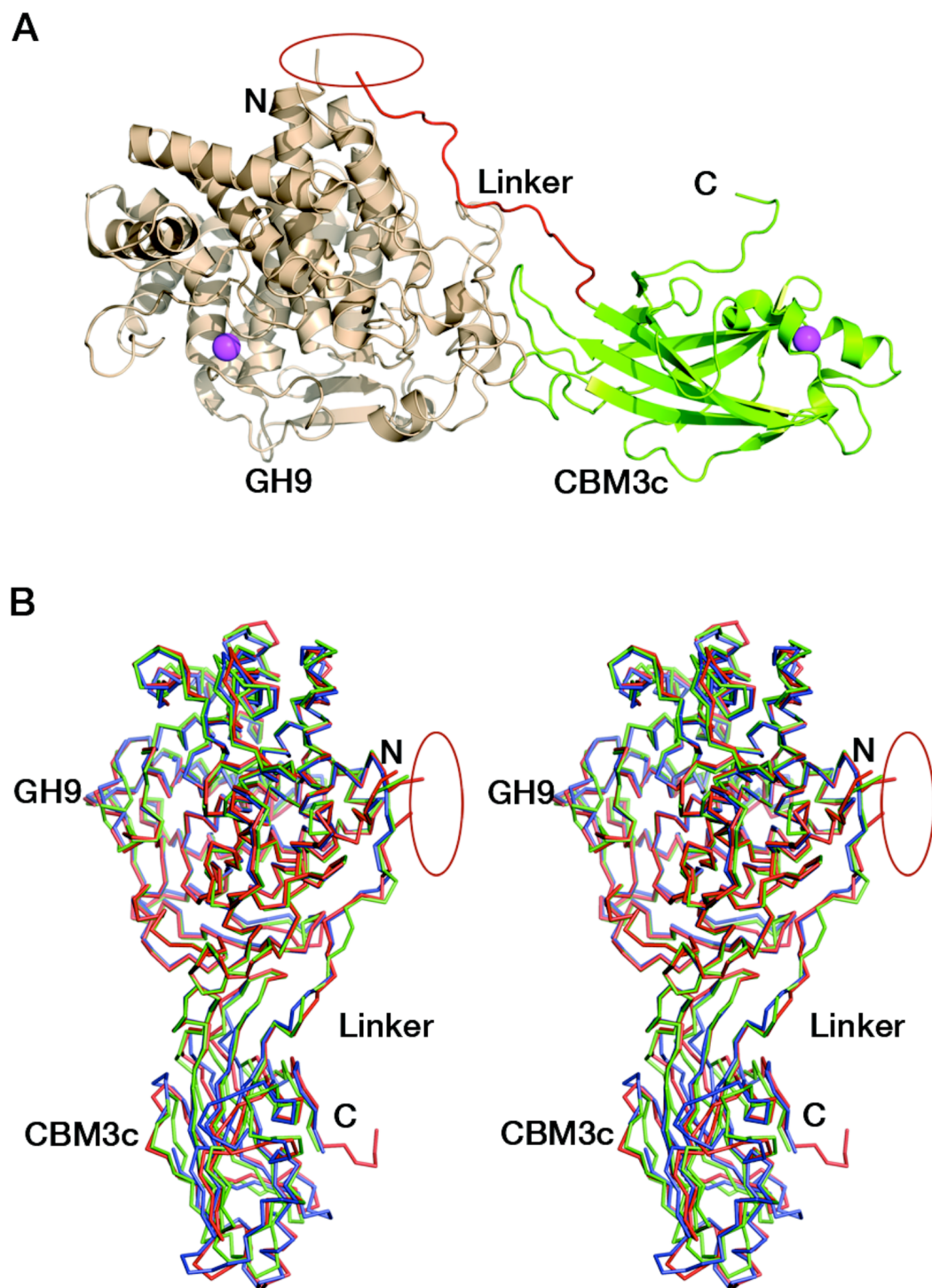
**Figure 2** Recovery of activity upon association of CBM3c (with and without linker) and GH9. CMCase activity ( $\mu\text{mol}$  reducing sugar released in a 10-min reaction) of His-tagged GH9, mixed either with CBM3cL (diamonds) or CBM3cNL (squares), was examined. A fixed amount (70 pmol) of the GH9 catalytic module was mixed with increasing amounts of the indicated helper module, and their respective activities were compared to that of the intact Cel9I (GH9-CBM3c-CBM3b, set as 100%).

two domains separately and combined them *in vitro*. Surprisingly, the combined modules crystallized and formed a structure similar to those of the known GH9 cellulases.

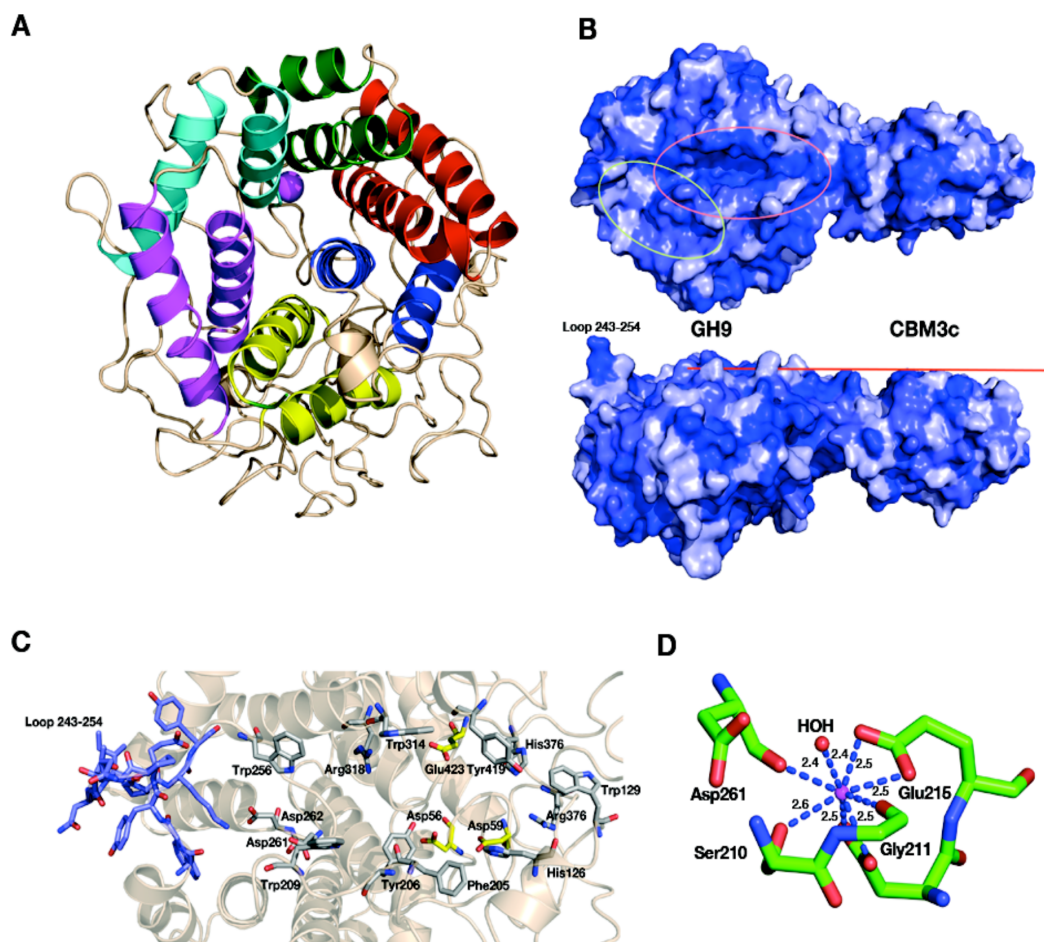
The crystal structure of the reassembled *C. thermocellum* Cel9I GH9-CBM3cL dyad was determined by molecular replacement and the coordinates are deposited in Protein Data Bank with code 2XFG. Data collection and refinement statistics are given in Tables 1 and 2. The catalytic GH9 and the ancillary CBM3c modules reassembled *in vitro* to form a dyad (Fig. 3A) similar in structure to the intact tandem GH9-CBM3c modules of the orthologous endoglucanases: Cel9G from *C. cellulolyticum* (1G87) and Cel9A (previously termed cellulase E4) from *Thermobifida fusca* (1TF4), with an RMS deviation of 0.783 Å over 468 C $\alpha$  atoms with Cel9G and 0.757 Å with Cel9A (Fig. 3B).

### Structure of the GH9 module

The catalytic module of the Cel9I enzyme consists of residues 1–446, comprising 15  $\alpha$ -helices, whereby the twelve longest ones form the ( $\alpha/\alpha$ )<sub>6</sub>-barrel (Fig. 4A). The hydrophobic core of the GH9 module is formed by 118 hydrophobic and aromatic amino acids, the vast majority of which are also conserved in the GH9 modules from *C. cellulolyticum* Cel9G and *T. fusca* Cel9A. Hydrophobic and aromatic cores have been proposed to play an important role in the formation of ( $\alpha/\alpha$ )<sub>6</sub>-barrels (Mandelman *et al.*, 2003). The GH9 module of Cel9I thus shows high structural similarity with the two latter GH9 structures: *C. cellulolyticum* Cel9G (0.367 Å RMS deviation over 349 C-alpha atoms) and *T. fusca* Cel9A (0.532 Å RMS deviation over 359 C-alpha atoms).



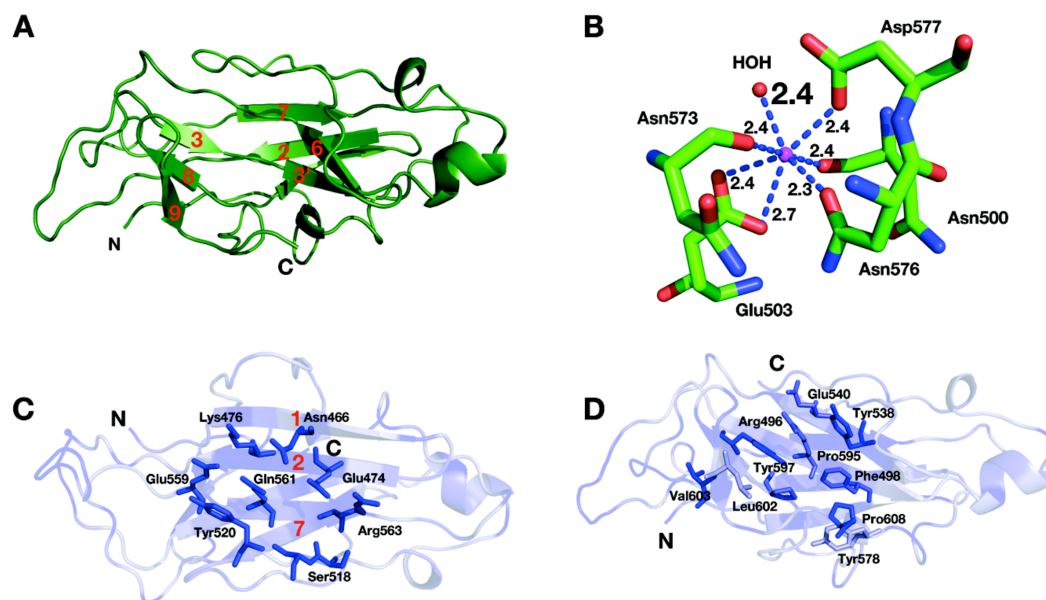
**Figure 3** Reassembled GH9-CBM3c from Cel9I. C and N termini are indicated, and the break between the GH9 and CBM3c modules is marked by a red ellipse. (A) The *in vitro* reassembled complex of the catalytic (GH9, wheat) and carbohydrate-binding (CBM3c, green) modules of Cel9I from *C. thermocellum*, cartoon representation. Calcium atoms are shown as magenta-colored spheres. (B) Stereo-view (cross-eyed) of the superposition of the reassembled GH9-CBM3c structure of *C. thermocellum* Cel9I (red) with the bimodular structures of *C. cellulolyticum* Cel9G (blue) and *T. fusca* Cel9A (green).



**Figure 4** Structural components of the reassembled *C. thermocellum* GH9-CBM3c. (A) Structure of the GH9 catalytic module, cartoon representation. Twelve  $\alpha$ -helices form an  $(\alpha/\alpha)_6$ -barrel fold. Pairs of helices, comprising the fold, are emphasized by red, blue, yellow, magenta, cyan and green. (B) Surface representation of the reassembled GH9-CBM3c complex. The residues are shaded according to the extent of their conservation with Cel9G from *C. cellulolyticum* and Cel9A from *T. fusca*. Darker blue indicates higher conservation. Top, birds-eye view of the catalytic cleft. Bottom, lateral view, showing the flat surface (red bar). Pink ellipse indicates the catalytic cleft, and green ellipse designates terminal portion of the catalytic site. (C) Close-up (same orientation as in B, top) of the catalytic cleft of the Cel9I GH9 module showing functional residues. Carbohydrate-binding residue carbons are colored gray, catalytic residue carbons are colored yellow. Loop 243–254 carbons are colored in light blue. (D) Calcium-binding site of the *C. thermocellum* Cel9I GH9 module. Coordinating residues are shown in stick representation. The calcium ion is colored magenta, and distances to the coordinating atoms are indicated.

The catalytic site of the GH9 module is located at the depression in the flat surface, formed by the loops connecting the N termini of the barrel helices (Fig. 4B). The flat face is rich in charged and polar residues (Fig. 4B), highly conserved also in Cel9G (1G87) and Cel9A (1TF4). The GH9 modules of these cellulases (Mandelman et al., 2003; Sakon et al., 1997; Zhou et al., 2004) exhibit similar flat faces and clefts, and these conserved residues (His 126, Trp 129, Phe 205, Tyr 206, Trp 209, Trp 256, Asp 261, Asp 262, Trp 314, Arg 318, His 376, Arg 378, and Tyr 419) have been shown to bind natural and synthetic oligosaccharides (Fig. 4C). In the present structure, as in the other known GH9-CBM3c





**Figure 5** Structure of the CBM3c of Cel9I from *C. thermocellum*. C and N termini are indicated (A). Cartoon representation,  $\beta$ -strands are numbered according to the alignment with Cel9G from *C. cellulolyticum*, and Cel9A from *T. fusca*. (B) Calcium-binding site of the CBM3c. (C) Birds-eye view of the flat surface. Residues are shaded according to their degree of conservation with *C. cellulolyticum* Cel9G and *T. fusca* CEL9A. Surface-exposed conserved residues are shown in stick representation. (D) Shallow groove of the CBM3c. Conserved surface residues are shown in stick representation. The residues are colored according to the degree of conservation in CBM3a, CBM3b and CBM3c modules derived from the sequences listed in the Methods section.

bimodular structures, one end of this cleft is blocked by a loop formed by residues 243–254 and the other end is fused with the flat surface of the CBM3c module (Fig. 4B). Details of the catalytic cleft are presented in Fig. 4C.

One calcium ion is found near the catalytic cleft of the GH9 module of Cel9I and is coordinated by a Ser 210 (OG) 2.6 Å, Gly 211 (O) 2.4 Å, Asp 261 (O) 2.4 Å, Asp 214 bifurcated (OD1, OD2) 2.5 Å, and Glu 215 bifurcated (OD1, OD2) 2.5 Å (Fig. 4D). Despite some minor changes in the residues of coordination this ion seems to be structurally equivalent to those of *T. fusca* Cel9A (RMS deviation 0.160 Å over 5 C $\alpha$  atoms of the coordinating residues), and *C. cellulolyticum* Cel9G (RMS deviation 0.503 Å over 4 C $\alpha$  atoms). In all three cases the calcium ion draws together the N-terminal ends of  $\alpha$ -helices 8 and 10.

### Structure of the CBM3c module

The CBM3c module consists of 150 amino acids arranged in an eight  $\beta$ -stranded sandwich motif homologous to other known family 3 CBM structures (Gilbert, Knox & Boraston, 2013; Mandelman et al., 2003; Petkun et al., 2010b; Sakon et al., 1997; Shimon et al., 2000; Tormo et al., 1996; Yaniv et al., 2014; Yaniv et al., 2012b; Yaniv et al., 2011). The “lower” face of the sandwich is formed by  $\beta$ -strands 1, 2, and 7; the “upper” face is formed by  $\beta$ -strands 3, 3', 6, 8, and 9 (Fig. 5A). The structure of Cel9I CBM3c is particularly similar to the structures of the other two previously described CBM3c structures (RMS deviation 0.734 Å over 116 C-alpha atoms with CBM3c from *C. cellulolyticum* Cel9G; RMS deviation 0.829 Å

over 113 atoms with CBM3c from *T. fusca* Cel9A). Only 31% of amino acids are located in  $\beta$ -strands of the CBM3c module from Cel9I; others are found in the loop regions.

One calcium ion was found in the upper  $\beta$ -sheet of the CBM3c molecule (Fig. 5B) and is coordinated by a water molecule and five residues from the upper  $\beta$ -sheet: Asn 500 (O), Glu 503 bifurcated (OE1, OE2), Asn 573 (O), Asn 576 (OD1), Asp 577 (OD1). This calcium atom is in a similar location as in Cel9A and Cel9G, and probably plays a structural role for most CBM3 modules, as was suggested previously (Tormo *et al.*, 1996).

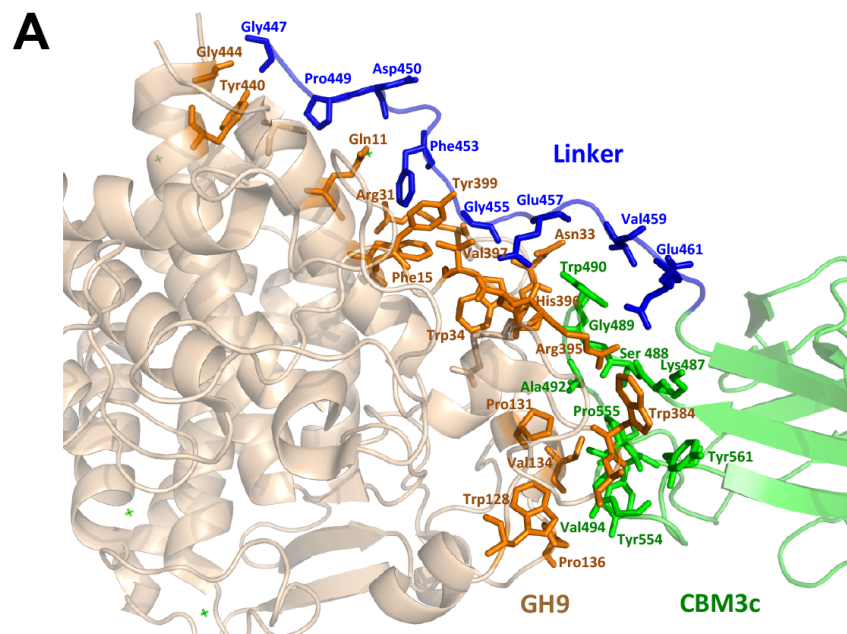
The lower sheet forms a flat platform conserved between the CBM3c modules and the other two molecular structures. This flat surface is rich in charged and polar conserved surface residues: Asn 466, Glu 474, Lys 476, Ser 518, Tyr 520, Glu 559, Gln 561, and Arg 563 (Fig. 5C). The planar region of the CBM3c modules in all three enzymes is particularly aligned in continuation of the catalytic cleft of the catalytic modules, and has been proposed to bind single chains of cellulose and guide them to the cleft (Mandelman *et al.*, 2003; Sakon *et al.*, 1997).

The CBM3c possesses a very interesting surface structure, formed by the  $\beta$ -strands on the opposite side of the flat surface, called the “shallow groove” (Shimon *et al.*, 2000; Tormo *et al.*, 1996). The “shallow groove” is lined by four aromatic rings (Phe 498, Tyr 538, Tyr 578 and Tyr 597), two charged or polar residues (Arg 496, and Glu 540), Leu 602, Pro 595 and Pro 608. These residues are also conserved in other CBM3 modules regardless of their subgroup relation (a, b, or c), their cellulose-binding ability and their effect on the activity of the catalytic module. Figure 5D shows the shallow groove of the CBM3c module from the Cel9I enzyme colored according to the extent of the conservation of the residues in other CBM3a, b and c modules (darker blue represents more conservation). The alignment was performed over 25 CBM3 sequences (11 CBM3c, 12 CBM3b and CBM3b', and 4 CBM3a). Conservation of this surface structure, regardless of the particular known function of the CBMs, implies that this site has some kind of “generic” function. This shallow groove may serve to bind to single oligosaccharide chains or to peptide chains, such as the intermodular linkers common to cellulases or cellulosomal scaffoldin subunits. There is evidence that the shallow groove interacts with a linker region (Petkun *et al.*, 2010a; Shimon *et al.*, 2000; Yaniv *et al.*, 2012a).

### Contact residues between the GH9, linker and CBM3c

The *in vitro* reassembled GH9-CBM3cL complex has a large intermodular interface, the contact area of which is 1,108.3 Å<sup>2</sup>, corresponding to 12.3% of the total surface-exposed area of the CBM3c module and 6.2% of the exposed GH9 module (Krissinel & Henrick, 2007). The GH9 and the CBM3cL modules of Cel9I are assembled into the reconstituted GH9-CBM3c complex by 31 hydrogen bonds (4 main chain-main chain, 19 main chain-side chain, and 8 side chain-side chain), 14 hydrophobic, 3 aromatic interactions, and 3 ionic bonds (<http://pic.mbu.iisc.ernet.in/index.html>) (Tina, Bhadra & Srinivasan, 2007). Sixteen residues from the GH9 module and seventeen residues from the CBM3c participate in these interactions (contact residues are shown in Fig. 6A). The vast majority of the contact residues and contacts are similar to those of *C. cellulolyticum* Cel9G and of *T. fusca* Cel9A (Fig. 6B).





**B**

**GH9 module**

```

CelI  TGAFNYGEALQKAIFFYECQRSGKLDSSSTLRLLNWRGDSGLDDGKDAGIDLTTGGWYDAGDH  60
CelG  AGTYNYGEALQKSIMFYEFQRSGDLPAD-KRDNRDDSGMKDGSVDVGLTTGGWYDAGDH  59
E4    EPAFNYAEALQKSMFFYEAQRSGKLPEN-NRVSWRGDSGLNDGADVGLDLTTGGWYDAGDH  59

```

```

CelI  DGHADHAWWGFPAEVMPEMERPSYKVDRSSPGSTVVAETSAAALAIASIFKVKDGEYSKECL  180
CelG  DGGKDHSWWGFPAEVMQMERPSFKVDASKPGSAVCASTAASLASAAVVFKSSDPTYAEKCI  179
E4    DGDADHKWWGFPAEVMPEMERPSFKVDSPCGSDVAAETAAMAASSIVFADDDPAYAATLV  179

```

```

CelI  G--RSFVVVGFGENPPKRPHPHRTAHGSWADSQMEPPEHRHVLVY GALVGGPDST-DNYTDDI  416
CelG  G--RSFVVVGVGNPPQHPHPHRTAHGSWTDQMTSPTYHRHTIY GALVGGPDNA-DGYTDEI  413
E4    PRNSSVYVGFGNPPRNPHPHRTAHGSWTDSIASPAENRHVLVY GALVGGPGSPNDAYTDDR  417

```

```

CelI  SNYTCNEVACDYNAGFVLLAKMYKLYGEL  446
CelG  NNYVNNEIACDYNAGFTGALAKMYKHS  441
E4    QDYVANEVATDYNAGFSSALAMLVEEY  445

```

**CBM3c module**

**Linker**

```

CelI  GSPDPKFNGLIEVPEDEIFVEAGVNASGNFIEIKAIVNNKSGWPARVCENLSFRYFINI  506
CelG  GDPIPNFKAIKINTDEVIIKAGLNSTGPNYTEIKAVVYNTGWPAPRVTDKISFKYFMDL  501
E4    GTPLADFPPTEPDGPEIFVEAQINTPGTTTFTEIKAMIRNQSWPARMRLDKGTFRYWFTL  505

```

```

CelI  EEIVNAGKSASDLQVSSSYNQAKLS--DV--KHYKDNIIYYVEVDLSGTKIYPGGQSA  562
CelG  SEIVAAGIDPLSLVTSSNYSEGKNTKVSGLVPWDVSNVYVNVDLTGENIYPGGQSACR  561
E4    DE----GVDPADITVSSAYNCATPE--D--VHHVSGDLYVEIDTGEKIFPGGQSEHR  557

```

```

CelI  KEVQFRISAPEGTV-FNPENDYSYQGLSAGTV-VKSEYIPVYDAGVLVFGREPLE  615
CelG  REVQFRIAAPQGTTYWNPKNDFSVDGLPTTSTVNTVTNIPVYDNGVKVFGNEP--  614
E4    REVQFRIAGGPG---WDPSNDWSFQIGNE--LAPAPYIVLYDDGVPVWGTAP--  605

```

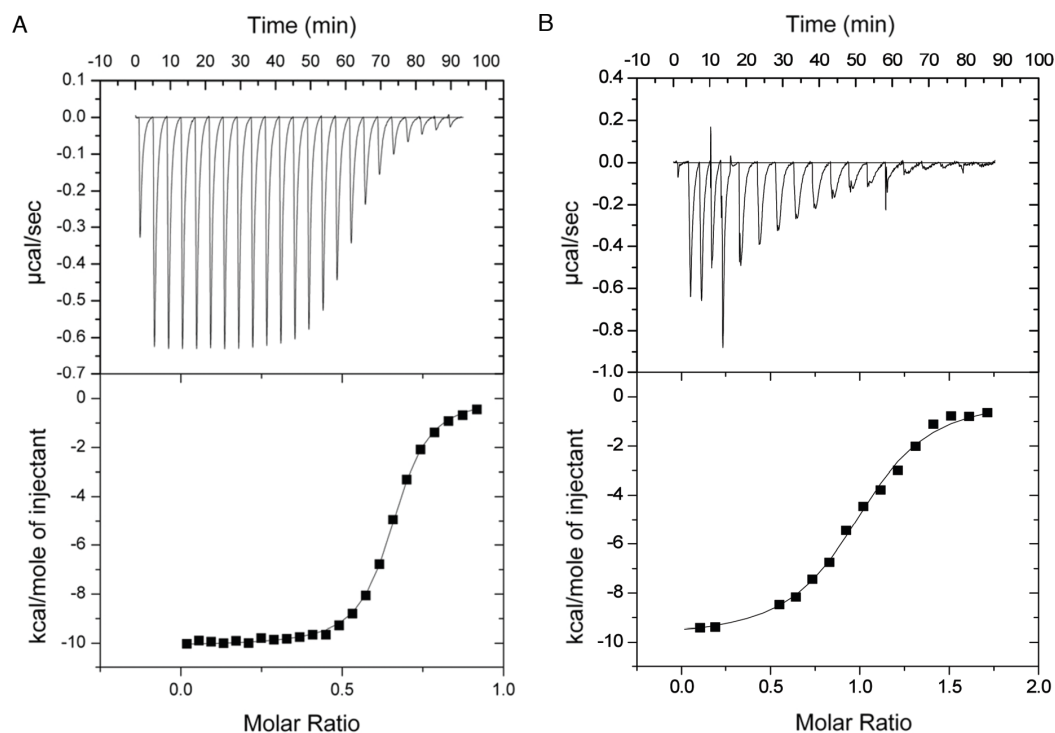
**Figure 6** Contact residues of the reassembled GH9-CBM3c complex. (A) The GH9 module is colored in brown, CBM3c in green and the linker in blue. Contact residues of the GH9, CBM3c and linker are colored orange, green and blue, respectively. The contact residues between the linker and the domains are described in the text. (B) Alignment of the GH9 and CBM3c modules of *C. thermocellum* Cel9I, *C. cellulolyticum* Cel9G, and *T. fusca* Cel9A (E4) cellulases. Contact residues are highlighted in yellow. Only the relevant regions of the alignment are shown. Residues of linker sequences are shown blue font.

Conserved residues of the linker (which spans from Gly447 to Asp462) make numerous contacts with conserved residues of the GH9 module, emphasizing the importance of the linker in this interaction (Fig. 6). A conserved Gly447 of the linker interacts via hydrogen bonds with Gly444 and Tyr440 of the GH9 module. Pro449 forms hydrophobic interactions with Tyr440, and Asp450 forms hydrogen bonds with Gln11. Another linker residue, Phe453, forms hydrophobic and aromatic interactions with two aromatic residues of the GH9 module, i.e., Phe15 and Tyr399. Gly455 forms hydrogen bonds with Asn33 and Arg31. Glu457 forms intricate interactions with a variety of residues of the GH9 module, which include hydrogen bonding with Asn33, Val397, Arg395, as well as hydrogen and ionic interactions with His396. Glu461 exhibits hydrogen and ionic interactions with Arg395. Additionally, linker residues Glu457 and Glu461 form hydrogen bonds and salt bridges with Trp490 and Lys487, respectively. The latter belong to the CBM3c module and are part of a loop (486–498), which protrudes towards and forms interactions with several amino acids of the GH9 module. There are also many hydrogen bonds formed between the neighboring amino acids of the linker, thus contributing to its defined conformation. Altogether the multiple, well-conserved interactions between the linker, the GH9 and the CBM3c modules stabilize the spatial orientation of the modules towards one another and contribute to the structural rigidity of the entire molecule, resulting in an active enzyme structure.

As mentioned above, the mutual spatial orientation of the GH9 and CBM3c modules is very similar to that in the native, intact bimodular pairs from Cel9G and Cel9A leading to the overall similarity in structures. The remarkable conservation of the overall architecture in the reassembled *in vitro* complex together with the striking conservation of the contact residues implies its high functional importance. In all of these structures (Cel9G, Cel9A, and the reassembled GH9-CBM3cL from Cel9I), the flat surface of the CBM3c module is aligned in continuation with the catalytic cleft of the GH9 module, making an extended platform (Fig. 4B). This platform is rich in charged and polar surface residues that are highly conserved throughout the family 3 CBM3c's.

### Microcalorimetric analysis of the GH9-CBM3c complex formation

The binding constants of GH9 and the CBM3c were obtained by performing isothermal titration calorimetry (ITC) experiments in which a solution of GH9 was titrated with a solution of CBM3c with or without the linker (Fig. 7). Control experiments for each of the components alone were conducted and subtracted from the titration data. In both cases the titration curve could be fitted to a one-site binding model although the calculated stoichiometry was less than one. The low stoichiometry is probably a result of the fact that the soluble CBM module lost its functionality with time and its true active concentration was less than the measured protein concentration. To estimate the binding constants for the two CBM3c forms, the CBM3c concentrations were corrected to provide a stoichiometry of one. In all cases, the binding reactions were enthalpy driven with a negative entropy contribution. CBM3cL provided binding constants ( $K_d$ ) between  $1.3\text{--}2.0 \times 10^{-6}$  M, whereas CBM3cNL exhibited stronger binding constants,  $K_d$  between



**Figure 7** Representative ITC titration of (A) GH9 and CBM3cNL (B) GH9 and CBM3cL. The top panel shows the calorimetric titration and the bottom panel displays the integrated injection heats corrected for control dilution heat. The solid line is the curve of the best fit used to derive the binding parameters, and the fitted data describe an interaction of a one binding site model.

$2.9\text{--}4.3 \times 10^{-7}$  M. Thus, the linker may serve as a mitigating factor for the binding process, ensuring specific binding orientation. This is consistent with the structural data and the activity assays, which emphasizes the important role of the linker in enzyme functioning. In the case of CBM3cNL, the binding process may occur faster in the absence of linker, but may also lead to unspecific binding and aggregation of the modules.

## DISCUSSION

A striking feature of the family 9 glycoside hydrolases is their subdivision into architectural themes, which are defined by their conserved modular composition (Bayer, Shoham & Lamed, 2006). In this context, the Theme B1 endoglucanases contain a GH9 catalytic module followed by a purportedly fused family 3c CBM. Biochemical studies of some of the members of this group (Arai et al., 2001; Chiriac et al., 2010; Gal et al., 1997; Irwin et al., 1998; Li, Irwin & Wilson, 2007) have shown that the CBM3c acts as a modulator of the function of the catalytic module. However, the exact manner in which the CBM3c functions is still unclear. It has been shown (Gal et al., 1997; Gilad et al., 2003; Irwin et al., 1998) that family 3c CBMs (including the CBM3c from *C. thermocellum* Cel9I) fail to bind insoluble cellulosic substrates, implying that they do not act as targeting agents for such substrates. The targeting of the enzyme to crystalline cellulose is achieved either

through an additional CBM ([Kostylev et al., 2012](#)) or by attachment of the enzyme to a CBM-containing scaffoldin via a cohesin-dockerin interaction ([Mingardon et al., 2011](#)).

The CBM3c module of Cel9A from the *T. fusca* has been proposed to loosely anchor the enzyme to cellulose, to disrupt the hydrogen bonds in crystalline cellulose and to guide a single cellulose strand towards the active site of the GH9 catalytic module ([Bayer, Shoham & Lamed, 2006](#); [Li, Irwin & Wilson, 2007](#)). This hypothesis has been supported by molecular docking and molecular dynamics simulation studies ([Oliveira et al., 2009](#)). Moreover, double point mutations indicated that high coordination between the substrate affinities of the catalytic module and CBM needs to be precisely controlled ([Li, Irwin & Wilson, 2010](#)). Enzyme thermostability was reported to be affected by the presence of the CBM3c probably due to the formation of a compact structure ([Chiriac et al., 2010](#); [Su, Mackie & Cann, 2012](#); [Yi et al., 2013](#)).

The previously reported structures of Cel9A from *T. fusca* ([Sakon et al., 1997](#)) and Cel9G from *C. cellulolyticum* ([Mandelman et al., 2003](#)) revealed that the catalytic module and the CBM3c are separated by a ~20-residue linker that forms multiple polar and hydrophobic interactions mainly with the GH9 module. In an earlier report, we demonstrated that separately expressed GH9 and CBM3cL from Cel9I of *C. thermocellum* interact with one another to form an enzymatically active complex ([Burstein et al., 2009](#)). In the current article, we showed further that the GH9 and CBM3c can also be reassembled without the linker, albeit at the expense of catalytic activity, thus emphasizing the importance of the linker in positioning correctly the CBM relative to the GH9 catalytic module.

There is evidence that linkers in multi-modular proteins may serve communication roles between the modules via allosteric mechanisms and variation in their sequences affect enzyme activity ([Ma et al., 2011](#)). Linker length and rigidity was shown to play a critical role in the cooperative action of the catalytic module of a cellulase and a CBM ([Ting, Makarov & Wang, 2009](#)). Computational studies of *T. fusca* Cel9A suggested that thermal contributions to enzyme plasticity and molecular motion at high temperatures may play a role in enhancing CBM and catalytic domain synergy, and the linker may have an important role in this process ([Batista et al., 2011](#)). The length of the linkers may also play an important role in protein function and adaptation to the environment ([Sonan et al., 2007](#)). Studies in cellulolytic fungi revealed that linkers undergo modifications such as glycosylation and have also been shown to directly bind to the cellulose substrate ([Beckham et al., 2012](#); [Payne et al., 2013](#); [Sammond et al., 2012](#); [Srisodsuk et al., 1993](#)). Point mutations in different fungal GH-CBM linkers have also been shown to significantly affect the activity of the enzymes and their stability ([Couturier et al., 2013](#); [Lu et al., 2014](#)).

The characteristics of the reassembled linker-containing complex are corroborated by the X-ray crystallographic data. Indeed, it is quite surprising that the two separately expressed entities recombined in such a way that the complex could in fact be crystallized. Moreover, the resultant structure was remarkably similar to the known structures of the intact bimodular GH9-CBM3c pairs from *C. cellulolyticum* Cel9G and *T. fusca* Cel9A. Accordingly, the vast majority of the contact residues are similar among the three structures. Conserved residues of the linker make contacts with conserved residues of

the GH9 module, highlighting the importance of the linker in this interaction. Multiple hydrogen, ionic and hydrophobic bonds between the linker and the functional GH9 and CBM3c stabilize the spatial organization of the modules. The similarity of the reassembled and native intact structures is particularly intriguing, as it suggests that folding of the modular structures and emplacement of the linker during biosynthesis and intermodular recognition during complex formation are governed by the same interactions, which may have distinct functional consequences. In contrast to the GH9-CBM3cL, the re-associated GH9-CBM3cNL complex never crystallized, suggesting that the reassembly of the two modules in the absence of linker was somewhat heterogeneous in character.

Single proteins commonly fold into defined structures, wherein their N- and C-terminal ends are in relatively close proximity to one another. If we view the structures of the Theme B1 enzymes, it is evident that their individual modules, the GH9 catalytic module and the CBM3c, are consistent with this rule. The positions of the N- and C-termini of the Theme B1 catalytic module are similar to those of the other GH9 thematic members, including those of Theme A, which lack additional modules. Likewise, the N- and C-termini of CBM3c are essentially the same as all other members of the family 3 CBMs, regardless of their source (i.e., parent cellulase, scaffoldin, etc.). The evolutionary significance of this observation is that, originally, the functional relationship between the two modules was likely a more conventional one, whereby an ancestral CBM3 played a standard targeting role to deliver the GH9 catalytic module to its substrate. During the course of evolution, this relationship changed, and the precise positioning and fusion of a mutated CBM3 with a GH9 catalytic module served to modulate the activity characteristics of the latter. For this purpose, the flat surface of the CBM3c is aligned with the flat surface of the catalytic module, and the appropriate residues that interact with the single cellulose chain are thus aligned with the active site of the GH9 module. As a consequence, the two closely juxtaposed modules can be considered as a single functional entity. The functional positioning and fusion of the two modules, however, are at odds with the inherent locations of the termini of the two modules, such that the C-terminus of the catalytic module is very distant from the N-terminus of the CBM3c. Consequently, nature has provided a very distinctive type of conserved linker, which both connects the two modules and helps secure their required orientation for processive endoglucanase activity.

## CONCLUSIONS

Cellulase 9I (Cel9I), a non-cellulosomal family 9 processive endoglucanase from *Clostridium thermocellum*, which degrades crystalline cellulose phosphoric acid-swollen cellulose (PASC) and carboxymethyl cellulose (CMC), consists of a catalytic GH9 module followed by two family 3 carbohydrate-binding modules (CBMs): CBM3c and CBM3b, separated by linker regions. C-terminal CBM3b module, as a classic CBM3, is responsible for targeting the Cel9I enzyme to the planar surface of the crystalline cellulose. The CBM3c is crucial for the GH9 enzymatic activity. In this work, we investigated the interaction of separately expressed catalytic module and CBM3c either with or without the intermodular

linker in order to better understand the function of the CBM3c in the family-9 enzymes and the role of the linkers regions.

GH9 catalytic module and CBM3c were able to interact and reassemble both with and without the linker; however, the linker was essential for the endoglucanase catalytic activity. Surprisingly, we were able to crystallize these two separately expressed entities, meaning that their reassembly was very ordered and structurally homogeneous. The molecular structure of the GH9 and CBM3c with the linker region showed that they form a complex similar in structure to the intact tandem GH9-CBM3c modules of the orthologous endoglucanases Cel9G from *C. cellulolyticum* and Cel9A from *Thermobifida fusca*. The flat, conserved surface of the CBM3c module is aligned in continuation with the catalytic cleft of the GH9 module, presumably forming one functional entity, which binds to the planar surface of the cellulose. Conserved residues of the linker make contacts with conserved residues of the GH9 module, highlighting the importance of the linker in this interaction. Overall, our results demonstrate that the linker regions in the GH9/CBM3c endoglucanases are necessary to achieve the right spatial organization of the modules and for the fixation of the modules into functional enzymes.

## ACKNOWLEDGEMENTS

This article is dedicated to the memory of Professor Felix Frolow, who passed away on 29 August 2014. We thankfully acknowledge the ESRF for synchrotron beam time and staff scientists of the ID-29 beam line for their assistance.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This research was supported by the Israel Science Foundation (ISF; Grant nos. 293/08 to FF and 1349/13 to EAB). Additional support was obtained by a grant (No. 24/11) issued to RL by The Sidney E. Frank Foundation through the ISF. A grant was awarded to EAB from the F. Warren Hellman Grant for Alternative Energy Research in Israel in support of alternative energy research in Israel administered by the Israel Strategic Alternative Energy Foundation (I-SAEF). This research was also supported by the establishment of an Israeli Center of Research Excellence (I-CORE Center No. 152/11, EAB) managed by the Israel Science Foundation, from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel, by the Weizmann Institute of Science Alternative Energy Research Initiative (AERI) and the Helmsley Foundation, and a grant to EAB and RL from the Israel Ministry of Science (IMOS). EAB was also supported by the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.



### Grant Disclosures

The following grant information was disclosed by the authors:

Israel Science Foundation: 293/08, 1349/13, 24/11.

Israel Strategic Alternative Energy Foundation (I-SAEF).

Israeli Center of Research Excellence: 152/11.

United States-Israel Binational Science Foundation (BSF).

Institute of Science Alternative Energy Research Initiative (AERI).

Helmsley Foundation.

Israel Ministry of Science (IMOS).

### Competing Interests

Edward Bayer is an Academic Editor for PeerJ.

### Author Contributions

- Svetlana Petkun conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.
- Inna Rozman Grinberg analyzed the data, wrote the paper, reviewed drafts of the paper, prepared figures and/or tables.
- Raphael Lamed conceived and designed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.
- Sadanari Jindou performed the experiments, contributed reagents/materials/analysis tools.
- Tal Burstein conceived and designed the experiments, performed the experiments.
- Oren Yaniv conceived and designed the experiments, performed the experiments, analyzed the data.
- Yuval Shoham conceived and designed the experiments, contributed reagents/materials/analysis tools.
- Linda J.W. Shimon analyzed the data, wrote the paper, reviewed drafts of the paper.
- Edward A. Bayer conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Felix Frolow conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.

### Data Availability

The following information was supplied regarding data availability:

Protein Data Bank, accession code [2XFG](#).

## REFERENCES

- Akinosho H, Yee K, Close D, Ragauskas A. 2014. The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Frontiers in Chemistry* 2:Article 66 DOI [10.3389/fchem.2014.00066](#).

- Arai T, Ohara H, Karita S, Kimura T, Sakka K, Ohmiya K. 2001. Sequence of celQ and properties of celQ, a component of the *Clostridium thermocellum* cellulosome. *Applied Microbiology and Biotechnology* 57:660–666 DOI 10.1007/s00253-001-0832-4.
- Batista PR, De Souza Costa MG, Pascutti PG, Bisch PM, De Souza W. 2011. High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics. *Physical Chemistry Chemical Physics* 13:13709–13720 DOI 10.1039/c0cp02697b.
- Bayer EA, Belaich J-P, Shoham Y, Lamed R. 2004. The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* 58:521–554 DOI 10.1146/annurev.micro.57.030502.091022.
- Bayer EA, Lamed R, White BA, Flint HJ. 2008. From cellulosomes to cellulosomics. *The Chemical Record* 8:364–377 DOI 10.1002/tcr.20160.
- Bayer EA, Morag E, Lamed R, Yaron S, Shoham Y. 1998. Cellulosome structure: four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics. In: Claeysens M, Nerinckx W, Piens K, eds. *Carbohydrases from trichoderma reesei and other microorganisms*. London: The Royal Society of Chemistry, 39–65.
- Bayer EA, Shoham Y, Lamed R. 2006. Cellulose-decomposing prokaryotes and their enzyme systems. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, eds. *The prokaryotes*. Third Edition. New York: Springer-Verlag, 578–617.
- Bayer EA, Smith SP, Noach I, Alber O, Adams JJ, Lamed R, Shimon LJW, Frolow F. 2009. Can we crystallize a cellulosome? In: Sakka K, Karita S, Kimura T, Sakka M, Matsui H, Miyake H, Tanaka A, eds. *Biotechnology of lignocellulose degradation and biomass utilization*. Tokyo: Ito Print Publishing Division, 183–205.
- Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, Baker SE, Himmel ME. 2012. Harnessing glycosylation to improve cellulase activity. *Current Opinion in Biotechnology* 23:338–345 DOI 10.1016/j.copbio.2011.11.030.
- Beckham GT, Stahlberg J, Knott BC, Himmel ME, Crowley MF, Sandgren M, Sorlie M, Payne CM. 2014. Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Current Opinion in Biotechnology* 27:96–106 DOI 10.1016/j.copbio.2013.12.002.
- Béguin P, Alzari PM. 1998. The cellulosome of *Clostridium thermocellum*. *Biochemical Society Transactions* 26:178–185 DOI 10.1042/bst0260178.
- Béguin P, Millet J, Aubert J-P. 1992. Cellulose degradation by *Clostridium thermocellum*: from manure to molecular biology. *FEMS Microbiology Letters* 100:523–528 DOI 10.1111/j.1574-6968.1992.tb05750.x.
- Brünger TA. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–475 DOI 10.1038/355472a0.
- Bu L, Nimlos MR, Shirts MR, Stahlberg J, Himmel ME, Crowley MF, Beckham GT. 2012. Product binding varies dramatically between processive and nonprocessive cellulase enzymes. *Journal of Biological Chemistry* 287:24807–24813 DOI 10.1074/jbc.M112.365510.
- Burstein T, Shulman M, Jindou S, Petkun S, Frolow F, Shoham Y, Bayer EA, Lamed R. 2009. Physical association of the catalytic and helper modules of a processive family-9 glycoside hydrolase is essential for activity. *FEBS Letters* 583:879–884 DOI 10.1016/j.febslet.2009.02.013.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active Enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37:D233–D238 DOI 10.1093/nar/gkn663.

- Chayen NE, Shaw Stewart PD, Maeder DL, Blow DM. 1990. An automated system for micro-batch protein crystallization and screening. *Journal of Applied Crystallography* 23:297–302 DOI 10.1107/S0021889890003260.
- Chiriac AI, Cadena EM, Vidal T, Torres AL, Diaz P, Pastor FI. 2010. Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* displaying a novel architecture. *Applied Microbiology and Biotechnology* 86:1125–1134 DOI 10.1007/s00253-009-2350-8.
- Cosier J, Glazer AM. 1986. A nitrogen-gas-stream cryostat for general X-ray-diffraction studies. *Journal of Applied Crystallography* 19:105–107 DOI 10.1107/S0021889886089835.
- Couturier M, Feliu J, Bozonnet S, Roussel A, Berrin JG. 2013. Molecular engineering of fungal GH5 and GH26 beta-(1,4)-mannanases toward improvement of enzyme activity. *PLoS ONE* 8:e79800 DOI 10.1371/journal.pone.0079800.
- Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B, Coutinho PM, Strazzulli A, Solovyova AS, Davies GJ, Gilbert HJ. 2012. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proceedings of the National Academy of Sciences of the United States of America* 109:20889–20894 DOI 10.1073/pnas.1212034109.
- Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3:853–859 DOI 10.1016/S0969-2126(01)00220-9.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall 3rd WB, Snoeyink J, Richardson JS, Richardson DC. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35(Web Server issue):W375–W383 DOI 10.1093/nar/gkm216.
- Demain AL, Newcomb M, Wu JH. 2005. Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews* 69:124–154 DOI 10.1128/MMBR.69.1.124-154.2005.
- Deprez C, Lloubes R, Gavioli M, Marion D, Guerlesquin F, Blanchard L. 2005. Solution structure of the *E. coli* TolA C-terminal domain reveals conformational changes upon binding to the phage g3p N-terminal domain. *Journal of Molecular Biology* 346:1047–1057 DOI 10.1016/j.jmb.2004.12.028.
- Doi RH, Kosugi A. 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology* 2:541–551 DOI 10.1038/nrmicro925.
- Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *ACTA Crystallographica. Section D: Biological Crystallography* 60:2126–2132 DOI 10.1107/S0907444904019158.
- Felix CR, Ljungdahl LG. 1993. The cellulosome—the exocellular organelle of *Clostridium*. *Annual Review of Microbiology* 47:791–819 DOI 10.1146/annurev.mi.47.100193.004043.
- Fontes CM, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual Review of Biochemistry* 79:655–681 DOI 10.1146/annurev-biochem-091208-085603.
- Freier D, Mothershed CP, Wiegel J. 1988. Characterization of *Clostridium thermocellum* JW20. *Applied and Environmental Microbiology* 54:204–211.
- Gal L, Gaudin C, Belaich A, Pagès S, Tardif C, Belaich J-P. 1997. CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. *Journal of Bacteriology* 179:6595–6601.
- Garcia-Martinez DV, Shinmyo A, Madia A, Demain AL. 1980. Studies on cellulase production by *Clostridium thermocellum*. *European Journal of Applied Microbiology and Biotechnology* 9:189–197 DOI 10.1007/BF00504485.

- Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, Gilbert HJ, Shoham Y. 2003.** Cell, a non-cellulosomal family-9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. *Journal of Bacteriology* **185**:391–398 DOI [10.1128/JB.185.2.391-398.2003](https://doi.org/10.1128/JB.185.2.391-398.2003).
- Gilbert HJ, Hazlewood GP. 1993.** Bacterial cellulases and xylanases. *Journal of General Microbiology* **139**:187–194 DOI [10.1099/00221287-139-2-187](https://doi.org/10.1099/00221287-139-2-187).
- Gilbert HJ, Knox JP, Boraston AB. 2013.** Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology* **23**:669–677 DOI [10.1016/j.sbi.2013.05.005](https://doi.org/10.1016/j.sbi.2013.05.005).
- Gilkes NR, Henrissat B, Kilburn DG, Miller RCJ, Warren RAJ. 1991.** Domains in microbial b-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews* **55**:303–315.
- Gruninger RJ, Gong X, Forster RJ, McAllister TA. 2014.** Biochemical and kinetic characterization of the multifunctional beta-glucosidase/beta-xylosidase/alpha-arabinosidase, Bgxa1. *Applied Microbiology and Biotechnology* **98**:3003–3012 DOI [10.1007/s00253-013-5191-4](https://doi.org/10.1007/s00253-013-5191-4).
- Henrissat B. 1991.** A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **280**:309–316 DOI [10.1042/bj2800309](https://doi.org/10.1042/bj2800309).
- Henrissat B, Davies G. 1997.** Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7**:637–644 DOI [10.1016/S0959-440X\(97\)80072-3](https://doi.org/10.1016/S0959-440X(97)80072-3).
- Henrissat B, Davies GJ. 2000.** Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiology* **124**:1515–1519 DOI [10.1104/pp.124.4.1515](https://doi.org/10.1104/pp.124.4.1515).
- Henrissat B, Romeu A. 1995.** Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochemical Journal* **311**:350–351 DOI [10.1042/bj3110350](https://doi.org/10.1042/bj3110350).
- Irwin D, Shin D-H, Zhang S, Barr BK, Sakon J, Karplus PA, Wilson DB. 1998.** Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *Journal of Bacteriology* **180**:1709–1714.
- Jeon SD, Yu KO, Kim SW, Han SO. 2012.** The processive endoglucanase EngZ is active in crystalline cellulose degradation as a cellulosomal subunit of *Clostridium cellulovorans*. *New Biotechnology* **29**:365–371 DOI [10.1016/j.nbt.2011.06.008](https://doi.org/10.1016/j.nbt.2011.06.008).
- Kostylev M, Moran-Mirabal JM, Walker LP, Wilson DB. 2012.** Determination of the molecular states of the processive endocellulase *Thermobifida fusca* Cel9A during crystalline cellulose depolymerization. *Biotechnology and Bioengineering* **109**:295–299 DOI [10.1002/bit.23299](https://doi.org/10.1002/bit.23299).
- Krissinel E, Henrick K. 2007.** Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**:774–797 DOI [10.1016/j.jmb.2007.05.022](https://doi.org/10.1016/j.jmb.2007.05.022).
- Kuusk S, Sorlie M, Valjamae P. 2015.** The predominant molecular state of bound enzyme determines the strength and type of product inhibition in the hydrolysis of recalcitrant polysaccharides by processive enzymes. *Journal of Biological Chemistry* **290**:11678–11691 DOI [10.1074/jbc.M114.635631](https://doi.org/10.1074/jbc.M114.635631).
- Lamed R, Setter E, Bayer EA. 1983.** Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology* **156**:828–836.
- Lamed R, Setter E, Kenig R, Bayer EA. 1983.** The cellulosome—a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnology and Bioengineering Symposium* **13**:163–181.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007.** ClustalW and ClustalX version 2. *Bioinformatics* **23**:2947–2948 DOI [10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404).

- Li Y, Irwin DC, Wilson DB. 2007. Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Applied and Environmental Microbiology* 73:3165–3172 DOI 10.1128/AEM.02960-06.
- Li Y, Irwin DC, Wilson DB. 2010. Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of *Thermobifida fusca* Cel9A. *Applied and Environmental Microbiology* 76:2582–2588 DOI 10.1128/AEM.02735-09.
- Lu H, Luo H, Shi P, Huang H, Meng K, Yang P, Yao B. 2014. A novel thermophilic endo-beta-1,4-mannanase from *Aspergillus nidulans* XZ3: functional roles of carbohydrate-binding module and Thr/Ser-rich linker region. *Applied Microbiology and Biotechnology* 98:2155–2163 DOI 10.1007/s00253-013-5112-6.
- Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R. 2003. X-ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *Journal of Bacteriology* 185:4127–4135 DOI 10.1128/JB.185.14.4127-4135.2003.
- Ma B, Tsai CJ, Haliloglu T, Nussinov R. 2011. Dynamic allostery: linkers are not merely flexible. *Structure* 19:907–917 DOI 10.1016/j.str.2011.06.002.
- Matthews BW. 1968. Solvent content of protein crystals. *Journal of Molecular Biology* 33:491–497 DOI 10.1016/0022-2836(68)90205-2.
- Meekins DA, Raththagala M, Husodo S, White CJ, Guo HF, Kotting O, Vander Kooi CW, Gentry MS. 2014. Phosphoglucan-bound structure of starch phosphatase Starch Excess4 reveals the mechanism for C6 specificity. *Proceedings of the National Academy of Sciences of the United States of America* 111:7272–7277 DOI 10.1073/pnas.1400757111.
- Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Biochemistry* 31:426–428.
- Mingardon F, Bagert JD, Maisonnier C, Trudeau DL, Arnold FH. 2011. Comparison of family 9 cellulases from mesophilic and thermophilic bacteria. *Applied and Environmental Microbiology* 77:1436–1442 DOI 10.1128/AEM.01802-10.
- Moraís S, Alber O, Barak Y, Hadar Y, Wilson DB, Lamed R, Shoham Y, Bayer EA. 2012. Functional association of the catalytic and ancillary modules dictates enzymatic activity in glycoside hydrolase family 43  $\beta$ -xylosidase. *Journal of Biological Chemistry* 287:9213–9221 DOI 10.1074/jbc.M111.314286.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *ACTA Crystallographica. Section D: Biological Crystallography* 53:240–255 DOI 10.1107/S0907444996012255.
- Ng TK, Weimer TK, Zeikus JG. 1977. Cellulolytic and physiological properties of *Clostridium thermocellum*. *Archives of Microbiology* 114:1–7 DOI 10.1007/BF00429622.
- Noach I, Alber O, Bayer EA, Lamed R, Levy-Assaraf M, Shimon LJW, Frolow F. 2008. Crystallization and preliminary X-ray analysis of *Acetivibrio cellulolyticus* cellulosomal type II cohesin module: two versions having different linker lengths. *Acta Crystallographica F* 64:58–61 DOI 10.1107/S1744309107066821.
- Oliveira OV, Freitas LCG, Straatsma TP, Lins RD. 2009. Interaction between the CBM of Cel9A from *Thermobifida fusca* and cellulose fibers. *Journal of Molecular Recognition* 22:38–45 DOI 10.1002/jmr.925.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* 307–326.



- Payne CM, Bomble YJ, Taylor CB, McCabe C, Himmel ME, Crowley MF, Beckham GT. 2011. Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. *Journal of Biological Chemistry* **286**:41028–41035 DOI [10.1074/jbc.M111.297713](https://doi.org/10.1074/jbc.M111.297713).
- Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor 2nd LE, Sandgren M, Stahlberg J, Stals I, Tan Z, Beckham GT. 2013. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceedings of the National Academy of Sciences of the United States of America* **110**:14646–14651 DOI [10.1073/pnas.1309106110](https://doi.org/10.1073/pnas.1309106110).
- Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with iterative structure refinement. *Nature Structural Biology* **6**:458–463 DOI [10.1038/8263](https://doi.org/10.1038/8263).
- Petkun S, Jindou S, Shimon LJW, Rosenheck S, Bayer EA, Lamed R, Frolow F. 2010b. Structure of a family 3b' carbohydrate-binding module from the Cel9V glycoside hydrolase from *Clostridium thermocellum*. Structural diversity and implications for carbohydrate binding. *Acta Crystallogr D* **66**:33–43.
- Petkun S, Jindou S, Shimon LJW, Rosenheck S, Bayer EA, Lamed R, Frolow F. 2010a. Structure of a family 3b' carbohydrate-binding module from the Cel9V glycoside hydrolase from *Clostridium thermocellum*: structural diversity and implications for carbohydrate binding. *ACTA Crystallographica. Section D: Biological Crystallography* **66**:33–43 DOI [10.1107/S0907444909043030](https://doi.org/10.1107/S0907444909043030).
- Sakon J, Irwin D, Wilson DB, Karplus PA. 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology* **4**:810–818 DOI [10.1038/nsb1097-810](https://doi.org/10.1038/nsb1097-810).
- Sammond DW, Payne CM, Brunecky R, Himmel ME, Crowley MF, Beckham GT. 2012. Cellulase linkers are optimized based on domain type and function: insights from sequence analysis, biophysical measurements, and molecular simulation. *PLoS ONE* **7**:e48615 DOI [10.1371/journal.pone.0048615](https://doi.org/10.1371/journal.pone.0048615).
- Schwarz WH. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* **56**:634–649 DOI [10.1007/s002530100710](https://doi.org/10.1007/s002530100710).
- Schwarz WH, Zverlov VV, Bahl H. 2004. Extracellular glycosyl hydrolases from Clostridia. *Advances in Applied Microbiology* **56**:215–261.
- Shimon LJW, Pagès S, Belaich A, Belaich JP, Bayer EA, Lamed R, Shoham Y, Frolow F. 2000. Structure of a family IIIa scaffoldin CBD from the cellulosome of *Clostridium cellulolyticum* at 2.2 Å resolution. *ACTA Crystallographica. Section D: Biological Crystallography* **56**:1560–1568 DOI [10.1107/S0907444900012889](https://doi.org/10.1107/S0907444900012889).
- Shoham Y, Lamed R, Bayer EA. 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiology* **7**:275–281 DOI [10.1016/S0966-842X\(99\)01533-4](https://doi.org/10.1016/S0966-842X(99)01533-4).
- Sonan GK, Receveur-Brechot V, Duez C, Aghajari N, Czjzek M, Haser R, Gerday C. 2007. The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. *Biochemical Journal* **407**:293–302 DOI [10.1042/BJ20070640](https://doi.org/10.1042/BJ20070640).
- Srisodsuk M, Reinikainen T, Penttilä M, Teeri TT. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *Journal of Biological Chemistry* **268**:20756–20761.
- Stout GH, Jensen LH. 1968. *X-ray structure determination. A practical guide*. London: MacMillan.



- Su X, Mackie RI, Cann IK. 2012. Biochemical and mutational analyses of a multidomain cellulase/mannanase from *Caldicellulosiruptor bescii*. *Applied and Environmental Microbiology* 78:2230–2240 DOI 10.1128/AEM.06814-11.
- Telke AA, Ghatge SS, Kang SH, Thangapandian S, Lee KW, Shin HD, Um Y, Kim SW. 2012. Construction and characterization of chimeric cellulases with enhanced catalytic activity towards insoluble cellulosic substrates. *Bioresource Technology* 112:10–17 DOI 10.1016/j.biortech.2012.02.066.
- Tina KG, Bhadra R, Srinivasan N. 2007. PIC: protein interactions calculator. *Nucleic Acids Research* 35:W473–W476 DOI 10.1093/nar/gkm423.
- Ting CL, Makarov DE, Wang ZG. 2009. A kinetic model for the enzymatic action of cellulase. *The Journal of Physical Chemistry B* 113:4970–4977 DOI 10.1021/jp810625k.
- Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA. 1996. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO Journal* 15:5739–5751.
- Vagin A, Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. *The Journal of Physical Chemistry* 30:1022–1025.
- Venditto I, Najmudin S, Luis AS, Ferreira LM, Sakka K, Knox JP, Gilbert HJ, Fontes CM. 2015. Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and beta-1,3-1,4-glucans through distinct mechanisms. *Journal of Biological Chemistry* 290:10572–10586 DOI 10.1074/jbc.M115.637827.
- Wiegel J, Mothershed CP, Puls J. 1985. Differences in xylan degradation by various noncellulolytic thermophilic anaerobes and *Clostridium thermocellum*. *Applied and Environmental Microbiology* 49:656–659.
- Wilson DB, Irwin DC. 1999. Genetics and properties of cellulases. *Advances in Biochemical Engineering* 65:1–21 DOI 10.1007/3-540-49194-5.1.
- Wilson DB, Kostylev M. 2012. Cellulase processivity. *Methods in Molecular Biology* 908:93–99.
- Winn MD, Isupov MN, Murshudov GN. 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *ACTA Crystallographica. Section D: Biological Crystallography* 57:122–133 DOI 10.1107/S0907444900014736.
- Winn MD, Murshudov GN, Papiz MZ. 2003. Macromolecular TLS refinement in REFMAC at moderate resolution. *Methods in Enzymology* 374:300–321.
- Yaniv O, Fichman G, Borovok I, Shoham Y, Bayer EA, Lamed R, Shimon LJ, Frolow F. 2014. Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-sigmaI factors. *ACTA Crystallographica. Section D: Biological Crystallography* 70:522–534 DOI 10.1107/S139900471302926X.
- Yaniv O, Frolow F, Levy-Assraf M, Lamed R, Bayer EA. 2012a. Interactions between family 3 carbohydrate binding modules (CBMs) and cellulosomal linker peptides. *Methods in Enzymology* 510:247–259 DOI 10.1016/B978-0-12-415931-0.00013-6.
- Yaniv O, Petkun S, Shimon LJ, Bayer EA, Lamed R, Frolow F. 2012b. A single mutation reforms the binding activity of an adhesion-deficient family 3 carbohydrate-binding module. *ACTA Crystallographica. Section D: Biological Crystallography* 68:819–828 DOI 10.1107/S0907444912013133.
- Yaniv O, Shimon LJ, Bayer EA, Lamed R, Frolow F. 2011. Scaffoldin-borne family 3b carbohydrate-binding module from the cellulosome of *Bacteroides cellulosolvens*: structural diversity and significance of calcium for carbohydrate binding. *ACTA Crystallographica. Section D: Biological Crystallography* 67:506–515 DOI 10.1107/S0907444911011322.

- Yi Z, Su X, Revindran V, Mackie RI, Cann I. 2013.** Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by *Caldicellulosiruptor bescii* during growth on crystalline cellulose. *PLoS ONE* **8**:e84172 DOI [10.1371/journal.pone.0084172](https://doi.org/10.1371/journal.pone.0084172).
- Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB. 2004.** Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochemistry* **43**:9655–9663 DOI [10.1021/bi049394n](https://doi.org/10.1021/bi049394n).
- Zmudka MW, Thoden JB, Holden HM. 2013.** The structure of DesR from *Streptomyces venezuelae*, a beta-glucosidase involved in macrolide activation. *Protein Science* **22**:883–892 DOI [10.1002/pro.2204](https://doi.org/10.1002/pro.2204).
- Zverlov VV, Velikodvorskaya GA, Schwarz WH. 2003.** Two new cellulosome components encoded downstream of *celI* in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP. *Microbiology* **149**:515–524 DOI [10.1099/mic.0.25959-0](https://doi.org/10.1099/mic.0.25959-0).