

# SCIENTIFIC REPORTS

OPEN

## Enhanced features of *Dictyoglomus turgidum* Cellulase A engineered with carbohydrate binding module 11 from *Clostridium thermocellum*

Chiara Cattaneo<sup>1</sup>, Patrizia Cesaro<sup>2</sup>, Stefano Spertino<sup>2</sup>, Sara Icardi<sup>2</sup> & Maria Cavaletto<sup>1,2</sup>

Lignocellulosic biomass (LCB) is a low-cost and abundant source of fermentable sugars. Enzymatic hydrolysis is one of the main ways to obtain sugars from biomass, but most of the polysaccharide-degrading enzymes are poorly efficient on LCB and cellulases with higher performances are required. In this study, we designed a chimeric protein by adding the carbohydrate binding module (CBM) of the cellulosomal enzyme CtLic26A-Cel5E (endoglucanase H or CelH) from *Clostridium (Ruminiclostridium) thermocellum* to the C-terminus of Dtur CelA, an interesting hyperthermostable endoglucanase from *Dictyoglomus turgidum*. The activity and binding rate of both native and chimeric enzyme were evaluated on soluble and insoluble polysaccharides. The addition of a CBM resulted in a cellulase with enhanced stability at extreme pHs, higher affinity and activity on insoluble cellulose.

Lignocellulosic biomass (LCB) represents an invaluable source of “green energy” providing fermentable sugars for different purposes, among which the production of bioethanol. However, the complete exploitation of LCB is still far from being accomplished, due to LCB complexity and recalcitrance, and several efforts are required to improve the efficiency of the process and to increase the product yield. Enzymatic hydrolysis is one of the main ways to obtain sugars from biomass, but often the polysaccharide-degrading enzymes are poorly efficient on LCB. Therefore, the characterization (or engineering) of cellulases with higher performances becomes mandatory to keep LCB sugar production at competitive prices if compared with sugars obtained from readily degradable biomass. *Dictyoglomus turgidum* is an extreme thermophilic anaerobic Gram-negative bacterium that produces a broad range of Carbohydrate Active eNZymes (CAZymes)<sup>1–3</sup>. Currently (January 2018) the CAZy database<sup>4</sup> (<http://www.cazy.org>) reports 57 glycoside hydrolases (GHs) in *D. turgidum* genome. The Dtur\_0670 gene codes for a protein of 312 amino acids belonging to the GH5 family, subfamily 25 (GH5\_25)<sup>5</sup>, named Cellulase A (UniProtKB: B8DZM2). Dtur CelA is a hyperthermostable endoglucanase showing both endo- and exo-activity on soluble and insoluble  $\beta$ -(1,4)-linked glucose-containing substrates as well as endo-activity on soluble and insoluble  $\beta$ -(1,4)-linked mannose containing substrates, but does not recognize  $\beta$ -1,3-linked glucose residues<sup>6</sup>. Most cellulases contain non-catalytic carbohydrate-binding modules (CBMs) connected to the catalytic domain (CD) through a linker peptide that is sometimes highly flexible<sup>7</sup>. The CBMs bring the CD close to the target substrate, increasing the rate of catalysis, although the mechanism by which this interaction occurs is not completely clear<sup>7,8</sup>. It is known that CBMs recognize different polysaccharides in a specific way, thus the CBMs present in cellulases bind cellulose, enhancing enzyme activity on the insoluble polysaccharide<sup>9,10</sup>. According to CAZy, CBMs are classified in 83 families and are grouped into three types, A, B and C, depending upon their ligand-binding site interaction. Type A CBMs bind the flat surface of crystalline polysaccharides, Type B CBMs bind internally on glycan chains and Type C CBMs bind the termini of glycans (“small sugar-binding”)<sup>7,11</sup>. *Clostridium thermocellum* (now reclassified as *Ruminiclostridium thermocellum*) is a thermophilic, anaerobic, Gram-positive bacterium able to grow on crystalline cellulose<sup>12</sup>. In *C. thermocellum*, the degradation of cellulose is carried out by a high molecular weight complex, the cellulosome, consisting of a number of enzymes, each one containing a dockerin domain, which binds to a cohesin domain, enabling the assembly of several enzymes to be brought together into the

<sup>1</sup>University of Piemonte Orientale, Dipartimento di Scienze e Innovazione Tecnologica Complesso Universitario S. Giuseppe, Piazza S. Eusebio 5, Vercelli, 13100, Italy. <sup>2</sup>University of Piemonte Orientale, Dipartimento di Scienze e Innovazione Tecnologica Viale T. Michel 11, Alessandria, 15121, Italy. Chiara Cattaneo and Patrizia Cesaro contributed equally to the work. Correspondence and requests for materials should be addressed to M.C. (email: [maria.cavaletto@uniupo.it](mailto:maria.cavaletto@uniupo.it))

cellulosome<sup>13</sup>. *C. thermocellum* Lic26A-Cel5E (also known as endoglucanase H or CelH; UniprotKB: P16218) is a cellulosomal enzyme containing two catalytic modules (Lic26A and Cel5E), a CBM11 and two C-terminal type I dockerin domains. Lic26A is a GH26 hydrolase with  $\beta$ 1,3–1,4-mixed linked endoglucanase activity, while Cel5E is a bifunctional  $\beta$ -1,4-endoglucanase/xylanase classified as GH5<sub>25</sub><sup>5,14</sup>. CtCBM11 (type B CBM) accommodates a single polysaccharide chain in its cleft<sup>15,16</sup>, and presents a  $\beta$ -sandwich with a classical distorted  $\beta$ -jelly roll fold, where two anti-parallel  $\beta$ -sheets form a convex and a concave side, where the glycan binding takes place<sup>17</sup>. *D. turgidum* CelA has no electronically detectable CBM, as previously reported<sup>6</sup>, thus we designed a chimeric protein by adding the CtCBM11 to the C-terminus of Dtur CelA catalytic domain, connecting them by a linker peptide from CtLic26A-Cel5E. The aim of this study was to evaluate the effects of protein engineering on the activity and binding rate of Dtur CelA on both soluble and insoluble polysaccharides.

## Results and Discussion

**Expression and purification of native and chimeric Dtur CelA.** We designed a chimeric protein by adding the CBM11 and linker peptide from CtCelH to the C-terminus of Dtur CelA. CBM 11 from *Clostridium thermocellum* has been chosen, since it is naturally linked to a catalytic domain (Cel5E) classified as GH5 in the cellulosomal enzyme CelH. The amino acid sequence of chimeric Dtur CelA is reported in Fig. 1. The vector pET20b-CelA codes for a protein with theoretical molecular weight (MW) of 37.8 kDa and isoelectric point (pI) 5.49, while pET20b-CelA-linker-CBM11 codes for a protein with theoretical MW of 62.7 kDa and pI 5.42. Both the recombinant proteins were produced in soluble form. The protein profile after the purification step is shown in Fig. 2 and S1 (supplementary material). The molecular weights estimated by SDS-PAGE were around 37 kDa for Dtur CelA and 63 kDa for chimeric Dtur CelA, in agreement with those predicted from their amino acid sequence. After the removal of imidazole and other buffer components by dialysis, the addition of 20% glycerol was found to be the best preserving of protein function.

**2-DE profile of native and modified Dtur CelA.** Dtur CelA migrates as two spots with the same MW but with different pI (spots 1 and 2, Fig. 3A). These spots were excised from the gel, trypsin digested and submitted to LC-MS/MS analysis. Spots 1 and 2 were identified as Glycoside hydrolase 5 from *D. turgidum* (gi|217967063/YP\_002352569.1), whose sequence coverage is reported in Figures S2 and S3 (supplementary material). Chimeric Dtur CelA migrates as four main spots (3, 4, 5, 6, Fig. 3B), identified as Glycoside hydrolase 5 from *D. turgidum* (gi|217967063/YP\_002352569.1) and Carbohydrate-binding module family 11 protein of CelH from *C. thermocellum* (gi|125973984/YP\_001037894.1, now annotated as endoglucanase WP\_011838089.1 which represents the whole CelH protein) with the sequence coverages reported in Figures S4, S5, S6 and S7 (supplementary material). MS/MS analysis confirmed the presence of the entire CtCBM11 domain in the final protein product. The identified peptides are highlighted in bold red. It is very likely that both proteins Dtur CelA and chimeric Dtur CelA migrate as multiple spots for side reactions due to the iodoacetamide used in sample preparation<sup>18</sup>.

**Biochemical properties and kinetic parameters.** Zymography on 0.4% AZO-CMC revealed endoglucanase activity of the two protein forms as a clear band over a dark background (Fig. 4 and S8 of supplementary material). Since *D. turgidum* and *C. thermocellum* are thermophilic bacteria, their proteins are likely to be thermo-resistant as well. In order to evaluate the effects of CtCBM11 addition on the optimal pH and temperature of Dtur CelA, the two enzyme variants were assayed in the °T range of 40 °C–90 °C and pH range of 3.8–8.8. Optimal assay conditions of both proteins were 70 °C and pH 5.8 (Fig. 5). Chimeric Dtur CelA percentage of activity at 40 °C was 54%, then it increased up to 70 °C, where it reached its top. At 80 °C the activity decreased to 68% and dropped to 15% at 90 °C. Dtur CelA percentage of activity was lower at 40 °C (40%), reached 100% at 70 °C, then it slowly declined to 89% at 80 °C and sharply dropped to 15% at 90 °C. These data are in accordance with those reported previously: Dtur CelA temperature optimum was stated between 70 °C and 80 °C<sup>6</sup>. Concerning CtCBM11, Viegas and collaborators observed that it does not change its structure at 25 °C and 50 °C, confirming its stability in this °T range; moreover, a closer contact between CtCBM11 and cellohexaose was detected at 50 °C than 25 °C<sup>19</sup>. At the most acidic pH tested, chimeric Dtur CelA showed higher relative activity compared to the native form (20% vs 0%). The two enzymes showed a superimposable increase of activity in the pH range 4.8–5.8 (from 30% to 100%). At pH 6.8, relative activity of modified Dtur CelA was slightly lower than native Dtur CelA (64% vs 76%), while it retained higher activity in the pH range 7.8–8.8 when compared to Dtur CelA (60% vs 50%). The detected optimal pH of native Dtur CelA is in accordance with Brumm and collaborators<sup>6</sup>, who found a peak of activity between pH 5.8 and 6.8 for Dtur CelA. The addition of CtCBM11 to Dtur CelA might have enhanced protein stability, as confirmed by the higher relative activity of chimeric Dtur CelA respect with Dtur CelA at extreme pH values, while conserving its optimum at pH 5.8. The enzyme stability over time at 4 °C was also monitored by assaying activity for nine months and at the end of this time period, the chimera did not present any loss of activity (Supplementary material S9, S10).

Native and chimeric Dtur CelA showed  $K_m$  values of  $1.6 \times 10^{-2}$  and  $5.6 \times 10^{-4}$  mM, and  $V_{max}$  values of 19.65 and 1.87 mM min<sup>-1</sup> respectively, using barley  $\beta$ -glucan as substrate. The catalytic efficiency ( $K_{cat}/K_m$ ) of Dtur CelA and the chimera were  $2.6 \times 10^9$  M<sup>-1</sup>s<sup>-1</sup> and  $1.11 \times 10^{10}$  M<sup>-1</sup>s<sup>-1</sup>, respectively (Table 1). The addition of CtCBM11 decreased the  $K_m$  of Dtur CelA of two orders of magnitude and increased its efficiency, with a consequent higher affinity and specificity of chimeric Dtur CelA for barley  $\beta$ -glucan. This is in agreement with previous studies, where it was observed the specificity of CtCBM11 for  $\beta$ -(1,4)- and  $\beta$ -(1,3)- $\beta$ -(1,4)- mixed sugars<sup>17</sup>.

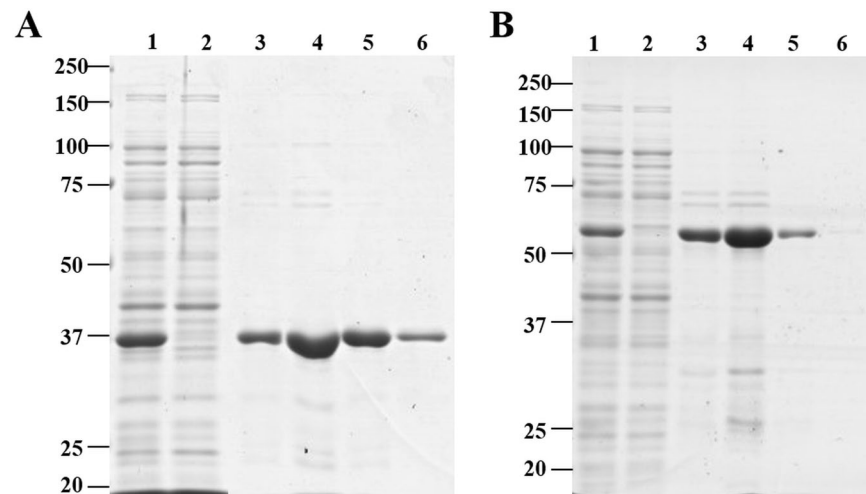
Enzyme activities on soluble and insoluble substrates are listed in Table 2. Dtur CelA engineering resulted in a chimeric enzyme which conserved catalytic activity. Native and chimeric Dtur CelA were more active on barley  $\beta$ -glucan, a highly accessible polysaccharide, than on the insoluble substrates AZCL-HE-Cellulose and Avicel. The activity of native Dtur CelA on barley  $\beta$ -glucan and AZCL-HE-Cellulose was higher than the chimeric form. On the contrary, the chimeric variant showed an increased activity on Avicel respect with native Dtur

atgaacaatcttccctattaagagaggaattaatTTTTGGGGATGCTTTAGAAGCTCCTTAT  
M N N L P I K R G I N F G D A L E A P Y  
gagggggcttggagtggttataataaaggatgaatTTTTAAAATTGTTAAGATGCA  
E G A W S G Y I I K D E Y F K I V K D A  
ggttttgatcatgtaagaattcccataaagtggagtggtttataactcaaaaagaagcccca  
G F D H V R I P I K W S V Y T Q K E A P  
tattctattgagaaaagaatTTTTGACAGAGTGGATCATTTAATAGAAGAAGGACTTAAA  
Y S I E K R I F D R V D H L I E E G L K  
aataactccatggttattataaacattcatcattatgaagagataatggaggatccctta  
N N L H V I I N I H H Y E E I M E D P L  
ggggaaaaagaaagatttctagctatatggagacagatttctgaactacaaggattac  
G E K E R F L A I W R Q I S E H Y K D Y  
cctaataatctttatTTTGAACTTTAAATGAGCCTACTCAAAATCTGAGTAGTGAATTA  
P N N L Y F E L L N E P T Q N L S S E L  
tggaaatcagttTTTAAAGAGGCTATTGAAGTTATAAGGAGAACGAATCCTGAAAGGAAG  
W N Q F L K E A I E V I R R T N P E R K  
attattgtaggaccagataaattggaatagtttatataatcttgagaaattgataattcca  
I I V G P D N W N S L Y N L E K L I I P  
gaaaatgatgagaatcataaattactTTTTCATTATTACAATCCTTTCCCTTTTACTCAT  
E N D E N I I I T F H Y Y N P F P F T H  
cagggggcaggggtggtaaaaattgatttacctgtaggagtaaaatgggttaggtactgag  
Q G A G W V K I D L P V G V K W L G T E  
gaagagaaaagggaaatagaaagagaacttgatatggcagtaagttgggcagaagagcat  
E E K R E I E R E L D M A V S W A E E H  
ggcaatataccctttatatgggggagtttgagcatattcaaaggcagatatggagtca  
G N I P L Y M G E F G A Y S K A D M E S  
agggtaagtgtagacagattttgtggcaagatctgcagaaaaagaggtattgcttggtca  
R V R W T D F V A R S A E K R G I A W S  
tattgggaatTTTATTCTGGATTGGAGTGTGTGATCCTGAGAAAAATGAATGGAGAACG  
Y W E F Y S G F G V F D P E K N E W R T  
cctcttcttagggcattaatacctgaaaggaatataaagcttgacatggctatttacaac  
P L L R A L I P E R N I K L D M A I Y N  
agagatacccgtagcttggacactgaaatcctcaatgcaactatttaatcccggacaat  
R D T R T F D T E I L N A L F N P G T Y  
ccgtcttatttctcgaacacttccaactccaagaccgaccaaaccgcccgtaacaccg  
P S Y S P K P S P T P R P T K P P V T P  
gctgtcggtgaaaaaatgctggatgattttgaggggtgtgtaaatgggggttcatactcc  
**A V G E K M L D D F E G V L N W G S Y S**  
**G E G A K V S T K I V S G K T G N M E**  
gtcagctacaccgggacaacggacggctactggggaacagtatacagtttaccggacggc  
**V S Y T G T T D G Y W G T V Y S L P D G**  
gattggtcaaaatggcttaaaatctcttttgacattaagtccggtgacggttctgccaat  
D W S K W L K I S F D I K S V D G S A N  
gaaatcagatttatgattgctgaaaaagcataaacgggtgtgggagacgggagacactgg  
**E I R F M I A E K S I N G V G D G E H W**  
gtttactcaataactcccgacagttcgtggaaaactatagaaaataccgttctccagcttt  
**V Y S I T P D S S W K T I E I P F S S F**  
agaagaagacttgattatcagccgctggacaggatagagcggtagctttggatcttgac  
**R R R L D Y Q P P G Q D M S G T L D L D**  
aatatagattcaattcacttcatgtatgccaacaacaagtccggaaaaattgtcgtagac  
**N I D S I H F M Y A N N K S G K F V D Y**  
aatatcaagctgattggtgctacttccgatccgactccttcaataaaaacgggacaccat  
**N I K L I G A T S D P T P S I K H G H H**  
caccatcaccattga  
H H H H -

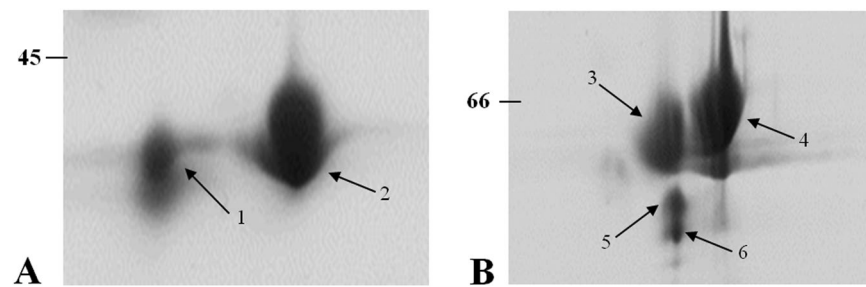
**Figure 1.** Nucleotide and amino acid sequence of chimeric Dtur CelA. The amino acid sequence of Dtur CelA is highlighted in gray, the linker sequence of CtCelH is underlined, and CtCBM11 is in bold.

CelA. A similar trend was reported for *Thermotoga maritima* Cel5A and its chimeric forms CBM1-Cel5A and Cel5A-CBM6, where the native enzyme showed higher activity on soluble cellulose (CMC) and lower activity on insoluble cellulose (Avicel) in comparison with the two chimeric enzymes<sup>20</sup>.

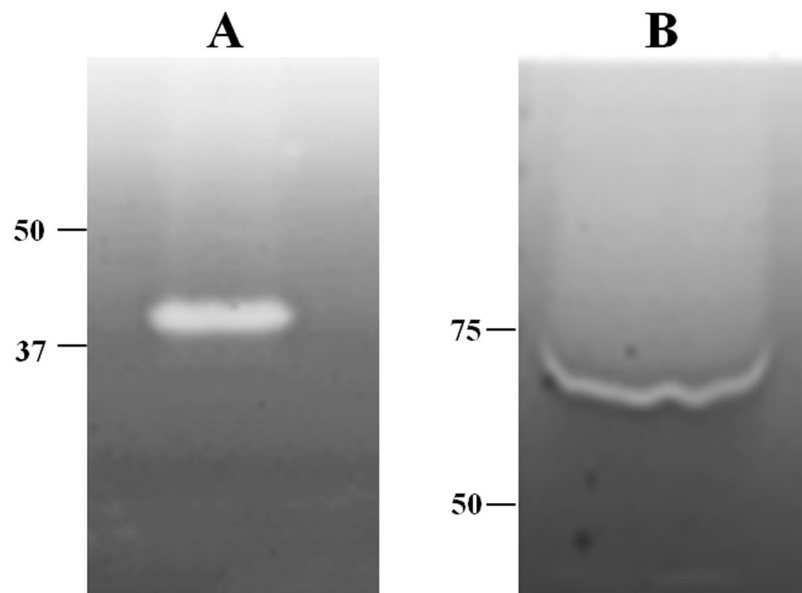
CtCBM11 interacts with  $\beta$ -1,4-linked glucans, but shows a preference for  $\beta$ -1,3-1,4 mixed linked glucose polymers, with the possibility to bind only one  $\beta$ -1,3-linked glucose in its cleft<sup>14</sup>. The structure of barley  $\beta$ -glucan corresponds to the “requirements” of CtCBM11, since its repetitive unit is composed by 2–3 consecutive  $\beta$ -1,4 linkages, separated by a  $\beta$ -1,3 linkage. This preference is confirmed by the lower  $K_m$  observed for chimeric Dtur CelA compared with native Dtur CelA on barley  $\beta$ -glucan, displaying an enhanced affinity for this



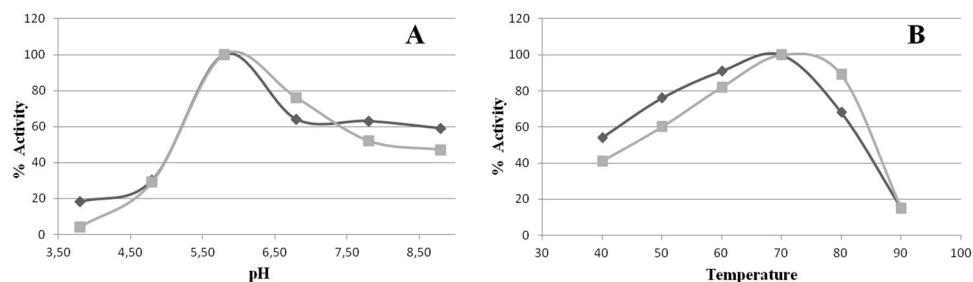
**Figure 2.** SDS-PAGE of native (A) and chimeric (B) Dtur CelA after purification by affinity chromatography. Lane 1: total protein extract; lane 2: flow through; lanes 3–6: elution fractions. The gels were Coomassie stained. Molecular weight markers are indicated on the left.



**Figure 3.** 2-DE separation of Dtur CelA (A) and chimeric Dtur CelA (B). 75  $\mu$ g of protein were loaded and focused on pH 3–10 IPG-strip before separation on a 10% (A) and 12% (B) polyacrylamide gel. Molecular weight markers are indicated on the left.



**Figure 4.** Zymography on 0.4% AZO-CMC of native (A) and chimeric (B) Dtur CelA. Molecular weight markers are indicated on the left.



**Figure 5.** Optimal pH and temperature: percentages of activity observed on  $\beta$ -glucan for native (squares) and chimeric (diamonds) Dtur CelA in Na-acetate 50 mM buffer at different pHs (A) and temperatures (B).

Enzyme	$K_m$ (mM)	$V_{max}$ (mM min <sup>-1</sup> )	$K_{cat}/K_m$ (M <sup>-1</sup> s <sup>-1</sup> )
Dtur CelA	$1.6 \times 10^{-2}$	19.65	$2.6 \times 10^9$
Chimeric Dtur CelA	$5.6 \times 10^{-4}$	1.87	$1.11 \times 10^{10}$

**Table 1.** Michaelis-Menten parameters ( $K_m$ ,  $V_{max}$ ) and catalytic efficiency ( $K_{cat}/K_m$ ) of native and chimeric Dtur CelA using low viscosity barley  $\beta$ -glucan as substrate.

Enzyme	$\beta$ -glucan	HE-cellulose	Avicel
Dtur CelA	$31176 \pm 3507$	$111.5 \pm 0.3$	$1.85 \pm 0.04$
Chimeric Dtur CelA	$9615 \pm 502$	$64.5 \pm 4.7$	$3.22 \pm 0.11$

**Table 2.** Enzyme activity on soluble and insoluble polysaccharidic substrates. Activity is expressed as  $\mu\text{mol} \times \text{min}^{-1}$  per 1  $\mu\text{mol}$  enzyme.

polymer exerted by the addition of CtCBM11. These findings are in accordance with a very recent work<sup>21</sup> on the *Bacillus subtilis* endo- $\beta$ -1,4-glucanase BsCel5A (a GH5 with high activity on  $\beta$ -1,3-1,4-linked glucans), where the exchange of the CBM3, present in the wild type BsCel5A, for CtCBM11 (with high affinity for  $\beta$ -1,3-1,4 glucans) increased the catalytic efficiency of the enzyme on such substrates.

However, despite the higher affinity for barley  $\beta$ -glucan, we did not observe an increase of activity of the chimeric enzyme for this substrate. This evidence is supported by other studies<sup>22,23</sup>, where higher enzyme activities on a specific substrate did not correlate with the binding amounts for that substrate. On the contrary, it seems that the highest activity is correlated with a moderate binding affinity of the CBM.

In our case, a possible explanation could be provided by the inability of Dtur CelA to recognize  $\beta$ -1,3 linkages: this fact may not represent a problem for the native enzyme, that could leave the oligomer when encountering a non-recognized bond but, in the chimeric enzyme, the tighter binding of CtCBM11 to barley  $\beta$ -glucan might impede a free movement of Dtur CelA along the chain, forcing its contact with an unrecognized bond. This is supported by previous observations<sup>23</sup> that tight binding of a CBM may limit the number of productive hydrolytic events and the catalytic domain may not have an easy access to new glycosidic bonds.

The activity of chimeric Dtur CelA was almost half of that measured for native Dtur CelA on AZCL-HE-Cellulose. It is argued that the enhanced activity of CBM-containing cellulases on their substrates may be related to increased enzyme concentration onto the surface and penetration into the fibers of the cellulosic substrate<sup>24,25</sup>. Carvalho and coworkers<sup>14</sup> detected a much weaker binding of CtCBM11 to HE-cellulose when compared to barley  $\beta$ -glucan. Thus, the lower activity of the chimera on HE-cellulose may be due to the poor interaction between enzyme and substrate.

On the contrary, the chimeric enzyme doubled its activity on Avicel compared with the native form. The increase of activity on insoluble rather than soluble substrates is a common trait of many CBMs<sup>11</sup>. Other studies reported the ability of CtCBM11 to increase the activity of GH5 catalytic domain of CtCelH against the insoluble substrate Avicel but not against soluble substrates such as lichenan and carboxymethyl cellulose (CMC)<sup>22,26</sup>, and the absence of CBM reduced the activity of CtCel5E on Avicel, ball-milled cellulose and oat-spelt xylan compared with its CBM containing forms (wild type or CBM3 and CBM6 chimeras)<sup>22</sup>.

Here we observed that the new introduction of a CBM (CtCBM11) to an endoglucanase that does not present a CBM in its native form (DturCelA), could ameliorate its activity on insoluble cellulosic substrates.

**Binding efficiency to insoluble cellulose.** The molar absorption coefficients calculated for native and chimeric Dtur CelA ( $88500 \text{ M}^{-1}\text{cm}^{-1}$  and  $115800 \text{ M}^{-1}\text{cm}^{-1}$ ) are in good agreement with those predicted from the tryptophan and tyrosine contents (Dtur CelA  $88760 \text{ M}^{-1}\text{cm}^{-1}$ , chimeric Dtur CelA  $135700 \text{ M}^{-1}\text{cm}^{-1}$ ). The ability of native and modified Dtur CelA to bind microcrystalline cellulose was tested on Avicel at different time periods after protein purification and the time variable did not produce negative effects on both proteins (data not shown). The binding isotherms were regressed by graphing a double reciprocal plot of bound versus free protein

Construct	Module	Primer name	Nucleotide sequence (5'-3')
Dtur CelA	DturCelA	CelADtur-NdeI-F	ACTGCATATGAACAATCTTCTATTAAAGAGA
		CelADturHis <sub>6</sub> S-XhoI-R	CAGTCTCGAGTCAATGGTGATGGTGATGGTGATATATTCCTTTCAGGTATTA
Chimeric Dtur CelA	DturCelA	CelADtur2-NdeI-F	ACTGCATATGAACAATCTTCTATTAAAGAG
		CelADturNS-HindIII-R	CAGTAAAGCTTATATTCCTTTCAGGTATTAATG
	CtherCelH Carbohydrate binding module	CelHCBM11-HindIII-F	ACTGAAAGCTTGACATGGCTATTTACAACAGAG
		CelHCBM11His <sub>6</sub> S-XhoI-R	CAGTCTCGAGTCAATGGTGATGGTGATGGTGCCGTGTTT TATTGAAGG

**Table 3.** List of primers for the wild-type and chimeric protein used in this study. Restriction sites of NdeI, HindIII and XhoI are underlined.

(native or chimeric variant) (Figures S11, S12 of supplementary material) and the affinity constants were calculated. A linear plot was obtained, indicating a single type of specific interaction. Both the enzymes demonstrated affinity to insoluble cellulose, with higher values registered for chimeric Dtur CelA. Native Dtur CelA showed an association constant ( $K_a$ ) of  $1.01 \mu\text{M}^{-1}$ , while the chimeric variant showed a three times higher  $K_a$  of  $3.28 \mu\text{M}^{-1}$ . The  $K_a$  obtained for chimeric Dtur CelA towards Avicel was comparable to those observed for the *C. thermocellum* CBM3a ( $2.5 \mu\text{M}^{-1}$ ) of the scaffoldin subunit of the cellulosome<sup>27</sup> and *C. cellulovorans* CbpA ( $1.67 \mu\text{M}^{-1}$ )<sup>28</sup>. The binding efficiency of chimeric Dtur CelA towards the long chains of Avicel is consistent with the results of Viegas and collaborators<sup>19</sup>, who observed a higher affinity of CtCBM11 for larger ligands when compared to those with the minimal length to fit the binding cleft. This difference is probably due to the presence of more contact points in larger oligosaccharides.

## Conclusions

This study reports for the first time the production of DturCelA as a chimeric protein. This fusion protein, while being still active on soluble glucans, shows higher activity on crystalline cellulose when compared to the native enzyme.

The addition of a carbohydrate binding module increased the enzyme's stability at extreme pHs, making it usable in a broad range of process conditions. Moreover, its higher affinity to insoluble cellulose presents the economic advantage of reducing the enzyme loading to get effective hydrolysis rates. Despite the amount of literature reporting the fusion of different CBMs with GH5 enzymes, this GH family comprises so many enzymes from different species and with different properties that the resulting fusion proteins could show a different behavior from that predictable a priori. This study revealed important information about the interactions of the enzyme with its possible substrates. For a useful industrial application, the promising traits of this chimeric cellulase make it a good candidate for the degradation of lignocellulosic biomass, by employing it together with other cellulases for the production of enhanced enzymatic cocktails in the effort to improve the hydrolysis step in the biorefinery.

## Materials and Methods

**Gene cloning.** Based on the published sequences of *D. turgidum* CelA and *C. thermocellum* CelH genes (GenBank Accession No. 7083332 and M31903.1, respectively), DNA sequences encoding Dtur CelA and the 48 aa linker peptide plus CBM11 from CtLic26A-Cel5E (CelH) were amplified with primers (Table 3) that allow their insertion into the pET20b vector (Novagen). The DNA fragments were purified using the Nucleo-Spin Extract kit (Macherey-Nagel, Germany). The catalytic module and the linker peptide plus CBM11 of the wild type vectors were combined by PCR amplification to produce a chimeric enzyme. PCR was also used for insertion of restriction sites and preservation of reading frames in the new constructs. The CtCBM11 module together with its adjacent linker was fused to the C-terminus of Dtur CelA catalytic module. All constructs were designed to contain a C-terminal His tag for subsequent purification steps. *Escherichia coli* clones containing the plasmids pET28a-CelA and pET28a-CelH were provided by C5-6 Technologies (WI, USA).

**Protein expression and purification.** *E. coli* BL21(DE3) cells (Stratagene, Canada) were transformed with plasmids pET-CelA-His<sub>6</sub> and pET-CelA-linkerpep-CBM11-His<sub>6</sub> and grown at 37 °C in Luria-Bertani medium containing  $50 \mu\text{g mL}^{-1}$  ampicillin. At  $A_{600}$  of 0.5, IPTG was added to a final concentration of 0.4 mM and cultures were incubated at 37 °C for 3 h. Cells were harvested by centrifugation (4,000 g, 30 min, 4 °C). Proteins were purified under native conditions using the 'Ni-NTA Agarose' resin (QIAGEN, Germany) following the protocol recommended by the manufacturing company. Briefly, cell pellets were resuspended in lysis buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, 300 mM NaCl, 10 mM imidazole) and sonicated. The mixture was centrifuged (20,000 g, 30 min, 4 °C) to remove the cell debris. One mL bed volume Ni-nitrilotriacetic acid (NTA) Agarose resin was poured in a Poly-Prep<sup>®</sup> chromatography column (Bio-Rad, USA) and equilibrated in lysis buffer. Then the supernatant was mixed with the resin for 1 h on a rotator at 4 °C. The column was washed twice with 4 volumes of wash buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, 300 mM NaCl, 20 mM imidazole). Elution was performed by using 50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, 300 mM NaCl, 250 mM imidazole. The purity and molecular mass of recombinant proteins were checked by SDS-PAGE, using Precision Plus Protein standard (Biorad) for molecular weight estimation. The gel images were acquired by the use of a densitometer (Image Scanner III, GE Healthcare, Sweden) at a resolution of 300 dpi. The molecular weights of Dtur CelA and chimeric Dtur CelA were calculated from their primary structure deduced from DNA sequence and compared with the experimental ones. Protein-containing fractions were pooled and dialyzed overnight at 4 °C against storage buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl). Then, 20% glycerol was added to the dialyzed fractions and they were stored at 4 °C until use. Proteins were quantified by the method of Bradford with BSA as the protein standard<sup>29</sup>.

**Two dimensional electrophoresis (2-DE).** The protein samples were mixed with rehydration buffer (7 M Urea, 2 M Thiourea, 4% w/v CHAPS, 50 mM DTT, 5% Triton X100, 5% 4–7 IPG Buffer (GE Healthcare, Sweden), and traces of bromophenol blue (BBF)). First dimension isoelectric focusing (IEF) was performed using 7 cm immobilized linear pH range 4–7 strips on an IPG-Phor unit (GE Healthcare, Sweden). After 7 h of passive rehydration, IEF was performed at 20 °C under the following conditions: 30 V for 7 h, 300 V for 4 h, a gradient step to 1000 V for 30 min, a gradient step to 5000 V for 1 h 30 min and a holding step of 5000 V for a total of 8000 V hours. Immediately before the second dimension, the focused strips were first reduced for 15 min in SDS equilibration buffer (50 mM Tris-HCl pH 8.8, 6 M Urea, 30% v/v Glycerol, 2% w/v SDS) containing DTT (65 mM), and then alkylated for 15 min in SDS equilibration buffer with iodoacetamide (243 mM). Equilibrated strips were placed onto 10 × 8 cm vertical 12% (Dtur CelA) or 10% (chimeric Dtur CelA) polyacrylamide gel. SDS-PAGE was performed with a Mini Protean II Xi System (Bio-Rad, USA). The samples were run at 10 mA per gel for 30 min and then at 30 mA per gel until the tracking dye front reached the bottom of the gel. After SDS-separation, 2DE gels were fixed in a solution containing 40% methanol and 10% acetic acid and stained overnight with Colloidal Coomassie Brilliant Blue G250 (Bio-Rad, CA, USA)<sup>30</sup>. The images of the gels were acquired as indicated above.

**In-gel digestion.** The protein spots were manually excised from 2DE gels, placed into 1.5 mL tubes previously cleaned with 50% acetonitrile (ACN) in HPLC-water, and destained overnight with 50% methanol and 5% acetic acid. The gel pieces were shrunk with 100% ACN, dried in a SpeedVac concentrator 5301 (Eppendorf, Germany) and rehydrated with 100 mM ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>) for 15 min. Then an equal volume of ACN was added for 10 min, followed by drying. The gel pieces were rehydrated with 100 mM NH<sub>4</sub>HCO<sub>3</sub> for 10 min and dried again. Trypsin (Sequencing Grade, Roche, Germany) was reconstituted in 50 mM NH<sub>4</sub>HCO<sub>3</sub> at a concentration of 25 ng/μL and added to the gel pieces. The digestion was performed overnight at 37 °C. The supernatant was collected in a new vial and peptide extraction was carried out twice in 50% ACN/0.1% formic acid (FA) for 10 min with ultrasonication. The supernatants were pooled, dried and stored at –20 °C until mass spectrometry (MS) analysis.

**Protein identification by nanoLC coupled with Q TOF MS/MS.** All nano-HPLC MS/MS experiments were performed on a Q-Star XL (Applied Biosystems, USA) connected to an Ultimate 3000 system equipped with a WPS-3000 autosampler and two low-pressure gradient micropumps LPG-3600 (LC Packings, NL), as described by Lingua *et al.*<sup>31</sup>. The Q-STAR XL operated in positive mode and information-dependent acquisition (IDA) mode, the dynamic exclusion feature of the Analyst QS 1.1 software (Applied Biosystems, CA, USA) was enabled, with an exclusion mass width of 3 m/z for 60 s. LC-MS/MS files obtained from each protein sample were merged into a single MASCOT generic format file (.mgf) and searched against the NCBI non-redundant database. Carbamidomethylation of cysteine residues, oxidation of methionine, deamidation of asparagine and glutamine were set as variable modifications for all Mascot searches. One missed trypsin cleavage site was allowed; tolerance for precursor and fragment masses was 0.25 Da.

**Zymography.** In-gel endo-β-1,4-glucanase activity of Dtur CelA and chimeric Dtur CelA was detected on 10% polyacrylamide gels containing 0.4% AZO-carboxymethyl cellulose (AZO-CMC, Sigma-Aldrich, MO, USA). Proteins were mixed with an equal volume of 2 × non-reducing Laemmli buffer (2% w/v SDS, 20% Glycerol, 125 mM Tris-HCl pH 6.8). The gels were run at 30 mA and washed four times with 50 mM sodium-acetate buffer pH 5.8 for 10 min at room temperature. Enzymatic activity was observed as a clear zone contrasting on a dark blue background. The images were acquired as described before.

**Enzyme characteristics.** The optimal pH and temperature of native and chimeric Dtur CelA were determined on 1% low viscosity barley β-glucan (Megazyme, Ireland) using the Nelson-Somogyi method for reducing sugars<sup>32,33</sup>. For determining the optimal pH, 50 mM sodium-acetate (pH 3.8–5.8) or 50 mM Tris-HCl (pH 6.8–8.8) buffers were used. All subsequent enzyme assays were performed at the optimal parameters unless otherwise noted. Different concentrations of barley β-glucan, ranging from 0.0015 mM to 0.025 mM in 50 mM sodium acetate, were used to determine the Michaelis-Menten constant (K<sub>m</sub>) and the maximum reaction velocity (V<sub>max</sub>). The Hanes-Woolf plot was used to infer the K<sub>m</sub> and V<sub>max</sub> of the two enzymes. The catalytic (K<sub>cat</sub>) and specificity (K<sub>cat</sub>/K<sub>m</sub>) constants were calculated as well.

**Enzyme assays.** Enzyme activity on β-linked soluble polysaccharides was tested on 1% barley β-glucan using the reducing sugar assay mentioned above. The endo-glucanase specificity was determined on insoluble azurine-cross-linked-hydroxyethyl cellulose (AZCL-HE-Cellulose) (Cellazyme C tablets, Megazyme, Ireland) and the absorbance of the soluble dyed fragments was read at 590 nm. Enzymatic hydrolysis of insoluble cellulose was assayed using Avicel PH-101 (Sigma-Aldrich, USA) as substrate, following the method described by Reyes-Ortiz *et al.*<sup>25</sup> with some modifications. The reaction mixture contained 10 mg of Avicel mixed with an aliquot of the enzyme plus β-glucosidase (C5–6 Technologies, USA) in a 1:1 ratio in a final volume of 500 μL of 50 mM sodium acetate. After incubation at 50 °C for 24 h in a thermo shaker, the substrate was removed by centrifugation and reducing sugar formation was determined using the dinitrosalicylic acid reagent (DNS) with a glucose standard curve<sup>34</sup>. Blanks without the enzyme were also included following the same procedure. Enzyme stability at 4 °C was also assessed by measuring activity over time (nine months). The samples were analyzed at least in triplicate and mean values were calculated. Enzyme activity is expressed as μmol of reducing sugars equivalent produced per minute per 1 μmol enzyme. Statistical analysis was performed with StatView 4.5 (Abacus Concepts, NJ, USA). Differences between the values were assessed using the F-test: P < 0.05 was considered to be significant.

**Measurement of native and chimeric Dtur CelA affinity to cellulose.** The ability of Dtur CelA and its chimeric variant to bind micro-crystalline cellulose was tested on Avicel PH-101 (20 mg in 50 mM sodium-phosphate, pH 7.0), as described by Lee *et al.*<sup>35</sup>. The unbound proteins in the supernatant were recovered after 1 h of incubation and removal of the insoluble cellulose together with adsorbed proteins by centrifugation at 20,000 g for 5 min. Centrifugation of supernatants was repeated twice to ensure removal of particulate material. The concentration of unbound protein was determined by reading the absorbance at 280 nm using a DU800 spectrophotometer (Beckman Coulter, CA, USA). The amount of bound protein was deduced subtracting the amount of unbound protein from initial protein concentration. Multiple replicates of samples (at least six) were analyzed and mean values were calculated. The Langmuir isotherm was used to evaluate the binding of the two enzymes to Avicel. An isotherm of bound ( $\mu\text{mol/g}$  Avicel) vs free ( $\mu\text{M}$ ) protein was generated and binding parameters were determined by graphing a double-reciprocal equation (1) (see below) as previously described<sup>36</sup>:

$$\frac{1}{B} = \frac{1}{N_0 K_a} \times \frac{1}{F} + \frac{1}{N_0} \quad (1)$$

where F is the equilibrium concentration of unbound protein, B is the equilibrium concentration of bound protein,  $K_a$  is the equilibrium binding constant ( $\mu\text{M}^{-1}$ ) and  $N_0$  is the equilibrium concentration of available binding sites on Avicel ( $\mu\text{mol/g}$  Avicel).

Molar absorption coefficients were experimentally calculated for each fusion protein using the Lambert-Beer law and compared with the theoretical values obtained with the web tool Peptide Property Calculator (<http://www.basic.northwestern.edu/biotools/proteincalc.html>), which assumes that the spectral contributions of the tyrosine, tryptophan and cystine at 280 nm do not differ significantly in the native form of the protein, relatively to the denatured form, and that the protein contains no other chromophores<sup>37</sup>.

**Data availability statement.** The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

- Svetlichnii, V. A. & Svetlichnaya, T. P. *Dictyoglomus turgidus* sp. nov., a new extreme thermophilic eubacterium isolated from hot springs in the Uzon volcano crater. *Mikrobiologiya Engl Tr.* **57**, 364–370 (1988).
- Nishida, H., Beppu, T. & Ueda, K. Whole-genome comparison clarifies close phylogenetic relationships between the phyla Dictyoglomi and Thermotogae. *Genomics*. **98**, 370–375 (2011).
- Brumm, P. *et al.* Mining *Dictyoglomus turgidum* for enzymatically active carbohydrases. *Appl Biochem Biotechnol.* **163**, 205–214 (2011).
- Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, H. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
- Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. III & Henrissat, B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol.* **12**, 186 (2012).
- Brumm, P., Hermanson, S., Luedtke, J. & Mead, D. Identification, cloning and characterization of *Dictyoglomus turgidum* CelA, an endoglucanase with cellulose and mannanase activity. *J Life Sci.* **5**, 488–496 (2011).
- Gilbert, H. J., Knox, J. P. & Boraston, A. B. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol.* **23**, 669–677 (2013).
- Bolam, D. N. *et al.* *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J.* **331**, 775–781 (1998).
- Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol.* **8**, 548–557 (1998).
- Gill, J. *et al.* The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochem J.* **342**, 473–480 (1999).
- Boraston, A. B., Bolam, D. N. & Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J.* **382**, 769–781 (2004).
- Demain, A. L., Newcomb, M. & Wu, J. H. Cellulase, clostridia and ethanol. *Microbiol Mol Biol Rev.* **69**, 124–154 (2005).
- Raghothama, S. *et al.* Characterization of a cellulosome dockerin domain from the anaerobic fungus *Piromyces equi*. *Nat Struct Biol.* **8**, 775–778 (2001).
- Carvalho, A. L. *et al.* The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates  $\beta$ -1,3-1,4-mixed linked glucans at a single binding site. *J Biol Chem.* **279**, 34785–34793 (2004).
- Viegas, A. *et al.* Molecular determinants of ligand specificity in family 11 carbohydrate binding modules – an NMR, X-ray crystallography and computational chemistry approach. *FEBS J.* **275**, 2524–2535 (2008).
- Bras, N. F., Cerqueira, N. M., Fernandes, P. A. & Ramos, M. J. Carbohydrate-binding modules from family 11: understanding the binding mode of polysaccharides. *Int J Quantum Chem.* **108**, 2030–2040 (2008).
- Hashimoto, H. *Recent structural studies of carbohydrate-binding modules.* *Cell Mol Life Sci.* **63**, 2954–2967 (2006).
- Müller, T. & Winter, D. Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Mol. Cell. Proteomics.* **16**, 1173–1187 (2017).
- Viegas, A. *et al.* Solution structure, dynamics and binding studies of a family 11 carbohydrate-binding module from *Clostridium thermocellum* (CtCBM11). *Biochem J.* **451**, 289–300 (2013).
- Mahadevan, S. A., Wi, S. G., Lee, D.-S. & Bae, H.-J. Site-directed mutagenesis and CBM engineering of Cel5A (*Thermotoga maritima*). *FEMS Microbiol Lett.* **287**, 205–211 (2008).
- Fonseca-Maldonado, R. *et al.* Lignocellulose binding of a Cel5A-RtCBM11 chimera with enhanced  $\beta$ -glucanase activity monitored by electron paramagnetic resonance. *Biotechnol. Biofuels.* **10**, 269 (2017).
- Ichikawa, S., Yoshida, M., Karita, S., Kondo, M. & Goto, M. Carbohydrate-binding modules influence substrate specificity of an endoglucanase from *Clostridium thermocellum*. *Bioscience Biotechnol Biochem.* **80**, 188–192 (2016).
- Walker, J. A. *et al.* Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol. Biofuels.* **8**, 220 (2015).
- Hervé, C. *et al.* Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *PNAS.* **107**, 15293–15298 (2010).
- Reyes-Ortiz, V. *et al.* Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnol Biofuels.* **6**, 93 (2013).



26. Taylor, E. J. *et al.* How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. *J Biol Chem* **280**, 32761–32767 (2005).
27. Morag, E. *et al.* Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of *Clostridium thermocellum*. *Appl Environ Microbiol.* **61**, 1980–1986 (1995).
28. Goldstein, M. A. *et al.* Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein A. *J Bacteriol.* **175**, 5762–5768 (1993).
29. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* **72**, 248–254 (1976).
30. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis.* **9**, 255–262 (1988).
31. Lingua, G. *et al.* Effects of Heavy Metals and Arbuscular Mycorrhiza on the Leaf Proteome of a Selected Poplar Clone: A Time Course Analysis. *PLoS One* **7**, e38662 (2012).
32. Nelson, N. A photometric adaption of the Somogyi method for the determination of glucose. *J Biol Chem.* **153**, 375–380 (1944).
33. Somogyi, M. Notes on sugar determination. *J Biol Chem.* **195**, 19–23 (1952).
34. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem.* **31**, 426–428 (1959).
35. Lee, H. L., Lee, S., Ko, H. & Choi, I. G. An expansin-like protein from *Hahella chejuensis* binds cellulose and enhances cellulose activity. *Mol Cells.* **29**, 379–385 (2010).
36. Kim, I. J., Ko, H.-J., Kim, T.-W., Choi, I.-G. & Kim, K. H. Characteristics of the binding of a bacterial expansin (BsEXLX1) to microcrystalline cellulose. *Biotechnol Bioeng.* **110**, 401–407 (2013).
37. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem.* **182**, 319–326 (1989).

## Acknowledgements

This study was supported by a grant of the Regione Piemonte (Contratto d'insediamento L.R. 34/2004, Misura INT2 n. 53/2008). The authors would also like to thank Dr Jessica Mignone for technical assistance.

## Author Contributions

C.C. carried out protein purification and enzyme assays, collected data, helped in the design and interpretation of experiments, and drafted the manuscript. P.C. helped to design and construct the vectors and chimeras, helped in the design and interpretation of experiments. S.S. and S.I. assisted with electrophoresis experiments. M.C. conceived of the project, assisted in the design and interpretation of experiments, and helped to draft and edit the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-22769-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018