A thermostable GH45 endoglucanase from yeast: impact of its atypical multimodularity on activity

Couturier et al.
A thermostable GH45 endoglucanase from yeast: impact of its atypical multimodularity on activity

Marie Couturier, Julia Feliu, Mireille Haon, David Navarro, Laurence Lesage-Meessen, Pedro M Coutinho, and Jean-Guy Berrin*

Abstract

Background: The gene encoding an atypical multi-modular glycoside hydrolase family 45 endoglucanase bearing five different family 1 carbohydrate binding modules (CBM1), designated PpCel45A, was identified in the Pichia pastoris GS115 genome.

Results: PpCel45A (full-length open reading frame), and three derived constructs comprising (i) the catalytic module with its proximal CBM1, (ii) the catalytic module only, and (iii) the five CBM1 modules without catalytic module, were successfully expressed to high yields (up to 2 grams per litre of culture) in P. pastoris X33. Although the constructs containing the catalytic module displayed similar activities towards a range of glucans, comparison of their biochemical characteristics revealed striking differences. We observed a high thermostability of PpCel45A (Half life time of 6 h at 80°C), which decreased with the removal of CBMs and glycosylated linkers. However, both binding to crystalline cellulose and hydrolysis of crystalline cellulose and cellohexaose were substantially boosted by the presence of one CBM rather than five.

Conclusions: The present study has revealed the specific features of the first characterized endo β-1,4 glucanase from yeast, whose thermostability is promising for biotechnological applications related to the saccharification of lignocellulosic biomass such as consolidated bioprocessing.

Keywords: biomass, crystalline cellulose, cellulase, Pichia pastoris, thermostable GH45 endoglucanase

Background

Lignocellulosic biomass is the largest renewable source of carbohydrates for the production of biofuels, biomaterials, and high-value products but its recalcitrance to enzymatic degradation makes industrial processes complex and costly [1,2]. The main component of plant cell wall is cellulose, a linear polymer of β-1,4 linked glucose units. Cellulose chains are arranged in linear microfibrils that form very recalcitrant crystalline-like structures, and form a tight complex with varying proportions of hemicellulose and lignin [3]. The hydrolysis of cellulose into glucose monomers requires the coordinated action of several types of complementary enzymes: endoglucanases (endo-β-1,4-glucanases, EC 3.2.1.4) randomly cleaving glycosidic bonds on cellulose polymers, cellobiohydrolases (cellulose β-1,4-cellobiosidases, EC 3.2.1.94) sequentially releasing cellobiose from cellulose chain ends, and β-glucosidases (β-1,4-glucosidases, EC 3.2.1.21) converting cellobiose into glucose monomers [4]. More recently, other types of enzymes, e.g. GH61, have been reported as able to enhance the conversion of cellulose when used in conjunction with a cellulase or a mixture of cellulases [5]. They may contribute to cellulose degradation using a different mode of action [6].

Among the 128 glycoside hydrolases (GHs) families (CAZy - [7,8]), fungal endoglucanases are grouped, along with other enzymes, into 8 GH families, including GH5, GH6, GH7, GH9, GH12, GH45, GH48, and GH74. Family GH45 endoglucanases are distantly related to plant expansins and found in a broad range of organisms including bacteria, plants, animals, and fungi. They are characterized by a low molecular weight and an inverting stereochemical mechanism [9], and usually produce cello-oligosaccharides as end-products from cellulose substrates without any glucose release as reported for endoglucanases belonging to other families [10].
Many glycoside hydrolases modular enzymes, where catalytic and non catalytic modules may be separated by linkers, are often highly glycosylated in fungi [11]. Carbohydrate binding modules (CBM), able to bind one or several type of polysaccharides, are found among the non-catalytic modules. The sugar-binding activity of CBMs increases the enzyme concentration in the vicinity of the substrates, leading to a more effective hydrolysis [12]. It has also been suggested that CBMs could play a role in the degradation of polysaccharides via a destructurative action on the substrate fibrils [13,14]. Amongst the 64 CBMs families classified in CAZy, CBM1s are almost exclusively found in fungi and specifically bind crystalline cellulose. They are characterized by the prevalence of aromatic amino acid residues in the binding surface that forms a platform-like architecture [15]. This planar organization of the binding site is thought to be complementary to the flat surfaces presented by cellulose crystals. CBMs can be localized at the N- or C-terminal end of the catalytic module alone or in multiple organizations [16]. The presence of multiple CBMs in a glycoside hydrolase is usually found in bacteria, such as in the α-amylase from *Lactobacillus amylovorus* that contains five CBM26 arranged in tandem [17]. To our knowledge, the only example of a characterized eukaryotic GH carrying multiple CBM1s modules is a GH45 endoglucanase from *Mucor circinelloides* which bears two CBM1 [18].

Upon analysis of the recently sequenced *Pichia pastoris* GS115 genome [19], we have identified an intriguing modular family 45 endoglucanase containing five different CBM1 modules arranged in tandem [20]. To investigate the role played by these multiple CBMs in the hydrolysis of cellulose, we homologously expressed the full length endoglucanase with three truncated derivatives, leading to a more effective hydrolysis [12]. It has also been suggested that CBMs could play a role in the degradation of polysaccharides via a destructurative action on the substrate fibrils [13,14]. Amongst the 64 CBMs families classified in CAZy, CBM1s are almost exclusively found in fungi and specifically bind crystalline cellulose. They are characterized by the prevalence of aromatic amino acid residues in the binding surface that forms a platform-like architecture [15]. This planar organization of the binding site is thought to be complementary to the flat surfaces presented by cellulose crystals. CBMs can be localized at the N- or C-terminal end of the catalytic module alone or in multiple organizations [16]. The presence of multiple CBMs in a glycoside hydrolase is usually found in bacteria, such as in the α-amylase from *Lactobacillus amylovorus* that contains five CBM26 arranged in tandem [17]. To our knowledge, the only example of a characterized eukaryotic GH carrying multiple CBM1s modules is a GH45 endoglucanase from *Mucor circinelloides* which bears two CBM1 [18].

Upon analysis of the recently sequenced *Pichia pastoris* GS115 genome [19], we have identified an intriguing modular family 45 endoglucanase containing five different CBM1 modules arranged in tandem at the N-terminus. To investigate the role played by these multiple CBMs in the hydrolysis of cellulose, we homologously expressed the full length endoglucanase and three truncated derivatives, and further characterized them biochemically to evaluate their potential as tools for cellulose conversion.

**Results**

**Bioinformatic analysis of *PpCel45A***

*PpCel45A* is a 1,845 bp gene (defined as [Genbank: PAS_chr4_0643]) containing no identified introns that encodes five N-terminal CBM1 modules and one C-terminal family 45 endoglucanase module (Figure 1). The six modules are separated by linkers carrying numerous potential O-glycosylation sites (Ser and Thr). Comparison of the five amino-acid CBM1 sequences revealed that they were all different with a maximum identity of 91% between CBM1-1 and CBM1-2 and a minimum identity of 41% between CBM1-2 and CBM1-3 (Figure 2). Interestingly, the result of BLAST searches showed that *PpCel45A* is the first endo β-1,4 glucanase identified in ascomycete yeast. The X33 strain is derived from the GS115 strain and we suggest that it probably also contains *PpCel45A*, as does the *P. pastoris* CBS 7435 strain that was also sequenced recently. However, to our knowledge other *Pichia* strains (e.g. *P. stipitis* CBS 6054) do not have any family 45 endoglucanase. The *P. pastoris* GS115 GH45 module sequence was distantly related to *T. reesei* family 45 endoglucanase V/Cel45a (17% amino-acid identity) but had high identity with endo-β-1,4-glucanases from the Mucorales order. Indeed, a phylogenetic analysis of 55 fungal GH45 was performed and revealed that *PpCel45A* and Mucorales GH45 are grouped together with *P. equi* Cel45a in GH45 subfamily A (Figure 3), whereas *T. reesei* Cel45a belongs to subfamily B. The high similarity between *PpCel45A*, *P. equi* Cel45a and Mucorales GH45 suggests that *PpCel45A* as well as *P. equi* Cel45a may have been acquired by horizontal gene transfer events. Horizontal gene transfer events have been described as common among the anaerobic fungi of the rumen [20], but to our knowledge such events have not been described between fungi and yeasts (Saccharomycetales).

**Cloning, expression, and purification of *PpCel45A* and derived constructs in *P. pastoris***

The DNA sequences encoding either the entire *P. pastoris* *PpCel45A* (full-length open reading frame) and three truncated derivatives, *PpCBM1×GH* (catalytic module with its proximal CBM1), *PpGH* (catalytic module), and *PpCBM1×GH* (five CBM1 modules without catalytic module) (Figure 1) were amplified by PCR using genomic DNA from *P. pastoris* GS115. Each sequence was inserted into the *P. pastoris* expression vector pPICZaA, in frame with the yeast α-factor secretion peptide and a (His)6 tag located at the C-terminus, and under the control of the AOX1 promoter as described in Materials and Methods. Full-length protein and truncated proteins were all successfully expressed in *P. pastoris* X33, i.e. they were visualized in the supernatant after 3-days induction using SDS-PAGE, indicating correct processing of the α-factor signal sequence. SDS-PAGE analysis showed that expression yield reached up to 2 g per litre of culture (data not shown). Only traces of endogenous proteins were detected in the culture supernatants of transformants (data not shown). Each selected transformant was scaled-up to increase the production of each recombinant protein. Purification was performed by affinity chromatography, taking advantage of C-term (His)6 tags. High production and purification yields were obtained with up to 250 mg of purified protein per litre of induced culture after three days induction (Table 1).

**Biochemical characterization**

The four purified recombinant proteins were analyzed on SDS-PAGE for assessment of apparent molecular weight (Figure 4, Table 1). They displayed apparent molecular weights higher than the theoretical ones,
probably resulting from post-translational modifications [21]. **PpCBM1** displayed no enzymatic activity under all the conditions tested. Apparent optimum pH and temperature were determined on **PpCel45A**, **PpGH**, and **PpCBM1**:GH. All three exhibited the same pH and temperature profile with an optimum pH at 4.8 and more than 50% activity between pH 3 and 8. Their optimum temperature was 65°C, with more than 50% activity between 30°C and 70°C (Table 1). Thermal stability was also evaluated on the three constructions at 65°C and 80°C. (Table 1 Figure 5). It revealed that **PpCel45A** was stable up to 80°C with a residual activity of 70% and 60% after 48 hours at 65°C and 4 hours at 80°C, respectively. **PpGH** was less stable with 9% activity retained after 48 hours at 65°C and 6% after 4 hours at 80°C. **PpCBM1**:GH was more stable than **PpGH** with respectively 51% and 45% residual activity after 48 hours at 65°C and 4 hours at 80°C, respectively.

**Catalytic properties of PpCel45A**

The four constructions were evaluated for their ability to degrade a wide range of cellulosic and hemicellulosic substrates (Table 2). Although **PpCBM1** had no detectable activity towards the substrates tested, the three other constructions (**PpGH**, **PpCBM1**:GH, **PpCel45A**) displayed similar activities towards soluble cellulose derivatives (CMC and HEC), β-1,3/β-1,4 glucans (β-glucan and lichenan), and one hemicellulolytic substrate (konjac glucomannan), the highest activity being obtained using lichenan. However, no activity was detected using pNP-Cel, pNP-Lac, pNP-Glu, filter paper, mann, and chitin, even after extensive hydrolysis. Avicel hydrolysis was performed for 16 hours at 45°C or 65°C (Figure 6). **PpCBM1**: displayed no activity towards Avicel. **PpGH**, **PpCBM1**:GH, and **PpCel45A** hydrolyzed Avicel significantly and more efficiently at 65°C, yielding specific activities of 3.8, 8.4 and 5.5 U/μmol. There was no improvement of Avicel hydrolysis using equimolar amount of **PpGH** and **PpCBM1**: compared to **PpGH** alone. Although no difference was observed between **PpCel45A** and **PpGH** at 45°C, **PpCel45A** was 1.5-fold more active than **PpGH** at 65°C. Surprisingly, **PpCBM1**:GH was significantly more active than **PpGH** and **PpCel45A** on Avicel both at 45°C and 65°C.

**Figure 1** Modular representation of **PpCel45A** and derived constructs used in the study.

**Figure 2** Alignment of the 5 CBMs amino-acid sequences. The three conserved aromatic amino acids (by referring to T. reesei Cel7a CBM1 sequence) forming a hydrophobic platform for interaction with cellulose are shown in grey.
Analysis of end products

To evaluate the mode of action of *P. pastoris* endoglucanase, the soluble sugars generated upon hydrolysis of either crystalline cellulose, soluble cellulose or cello-oligosaccharides by the three active recombinant constructions were determined by anion exchange chromatography. The products formed upon hydrolysis of polymeric substrates (Avicel and CMC) were analyzed, showing similar product patterns for the three enzymes, i.e. the oligo-saccharides G4, G3, and G2 were produced in the first minutes of the reaction. Toward the end of the reaction, cellobiose (G2) and cellotriose (G3) were the main products detected (data not shown). Upon hydrolysis of soluble cellulose (CMC), the three enzymes yielded similar amounts of oligo-saccharides at the end of the reaction. However, when crystalline cellulose (as in Avicel) was used as a substrate, *PpCel45A* released two times more end-products than *PpGH*, and *CBM15:GH* almost three times more than *PpGH*. The products formed upon hydrolysis of cellohexaose (G6), cellopentaose (G5), and cellotetraose (G4) were also analyzed. The three enzymes showed similar hydrolytic patterns towards cello-oligosaccharides. G6 was

![Figure 3 Phylogenetic representation of family GH45 from fungal origin](http://www.microbialcellfactories.com/content/10/1/103)

**Figure 3** Phylogenetic representation of family GH45 from fungal origin. The cladogram highlights the relative position of the proteins labelled with the abbreviation of the species name, the reference public database accession numbers (Genbank) and the identified activity for characterized enzymes (EC numbers). Subfamilies are indicated based on [28].

Table 1 Biochemical characterization of *PpCel45A* derived constructs

|                 | Purification yield (mg/l) | Molecular mass (kDa) | Apparent optimum pH | Apparent optimum temperature (°C) |
|-----------------|---------------------------|----------------------|---------------------|-----------------------------------|
|                 | a                         | b                    |                     |                                   |
| *PpCel45A*      | 172                       | 61.9                 | 4.8                 | 37-75 (65)                        |
| *PpCBM15:GH*    | 250                       | 31.5                 | 4.8                 | 37-75 (65)                        |
| *PpGH*          | 215                       | 22.6                 | 4.8                 | 37-75 (65)                        |
| *PpCBM11-5*     | 68                        | 35.2                 | n.d.                | n.d.                              |

*a* expression yield is expressed in mg of recombinant protein purified per litre of culture

*b* values are estimated from SDS-PAGE

*c* values indicate that relative enzyme activity remaining was above 50%. Values in brackets indicate apparent optimum temperature under the conditions used

n.d. : not determined since *PpCBM11-5* displays no catalytic activity.
hydrolyzed to produce G4 and G2 with traces amounts of G3, and G5 was degraded to form exclusively G3 and G2. G4 was hydrolyzed to produce exclusively G2 but with a lower rate that circumvented the determination of the catalytic efficiency. Neither PpCel45A nor their truncated derivatives showed detectable hydrolysis on cellotriose. Initial rate data from hydrolyses of G6 and G5 at 40°C and pH 4.8 indicated that the catalytic efficiency (kcat/Km) increased with increasing chain length (Table 3). It is interesting to note that PpCBM15:GH catalytic efficiency towards G6 was significantly higher compared to PpGH and PpCel45A.

Substrate binding capacities of PpCel45A and derived constructs
The ability of PpCel45A and its respective constructs to bind crystalline cellulose was investigated. Binding to Avicel was observed for PpCel45A, PpCBM1-5, PpCBM15:GH, but not for PpGH. Based on typical double-reciprocal plots, binding parameters were determined for the three constructions, yielding Kd values of 0.45, 0.78, and 0.95 for PpCBM1-5, PpCel45A, PpCBM15:GH, respectively (Table 4). The Bmax value of PpCBM15:GH (0.80 μmol per g of substrate) was 4 times higher compared to PpCBM1-5 and PpCel45A (0.19 and 0.20 μmol per g of substrate, respectively).

Discussion
Several family 45 endoglucanases have been characterized in the last years from fungi [9,22], bacteria [23], or animals [24]. However, to our knowledge the present study is the first report on the characterization of a GH45 from yeast. Since P. pastoris GS115 growth on cellulose as a sole carbon source was very limited, the

---

Figure 4 SDS-PAGE analysis of PpCel45A and its derived constructs. 10 μg of each purified recombinant protein was loaded onto a 12% Tris glycine SDS-PAGE and proteins were stained with Coomassie blue. PpCel45A (lane 1), PpCBM1-5:GH (lane 2), PpGH (lane 3) and PpCBM15-5 (lane 4). Molecular masses (kDa) of marker proteins are shown on the right.

Figure 5 Thermostability of PpCel45A and its derived constructs. Thermostability profiles at 65°C (A) and 80°C (B). Light gray: PpGH, dark gray: PpCBM1-5:GH, black: PpCel45A. Half-lives of the three constructions are indicated on the right.
Table 2 Specific activities (U/μmol) of PpCel45A derived constructs on pNP substrates and polysaccharides from different sources

|                  | Lichenan | β-glucan | CMC     | Glucomannan | HEC | Avicel |
|------------------|----------|----------|---------|-------------|-----|--------|
| PpCel45A        | 833 ± 34 | 203 ± 17 | 196 ± 16| 190 ± 23    | 54 ± 2| 2.74 ± 0.53 |
| PpCBM1<sup>15</sup>GH | 792 ± 27 | 123 ± 9  | 177 ± 13| 188 ± 19    | 53 ± 5| 5.43 ± 0.41 |
| PpCel45A        | 810 ± 29 | 124 ± 21 | 160 ± 2 | 161 ± 9     | 43 ± 4| 2.77 ± 0.35 |
| PpCBM1<sup>15</sup> | ND      | ND      | ND     | ND          | ND  | ND     |

Specific activities were determined at a substrate concentration of 10 mg/ml at 45°C in 50 mM citrate phosphate buffer pH 4.8 and expressed as units per μmol of enzyme in order to ensure comparison independently of molecular mass. Values are means of triplicate independent measures and standard errors of the mean are shown next to each value. N.D.: no activity detected.

function and origin of such an endoglucanase in yeast remains enigmatic and could be the result of a horizontal transfer from Mucorales. Since P. pastoris X33 has been proven to be an ideal host for the production of eukaryotic glycosidases [25-27], we successfully used it for homologous expression of PpCel45A and three truncated derivatives. The four proteins were successfully produced with high yields and purified in one step thus allowing their in depth characterization.

The highest specific activity displayed by the three active derivatives of PpCel45A was obtained toward lichenan, which has also been described as the best substrate for other GH45, as PcCel45A from Phanerochaete chrysosporium [28] or Bx-ENG-1, -2 and -3 from the pine wood nematode Bursaphelenchus xylophilus [29]. Although chitin and cellulose structures are similar, no activity of PpCel45A was detected toward chitin. Hydrolysis experiments using mannan and glucomannan revealed that PpCel45A was not capable to hydrolyze the β-1,4 linkages between mannosyl residues. However it hydrolyzed efficiently glucomannan suggesting that part of β-1,4 linkages found between mannose and glucose residues were efficiently hydrolysed. PpCel45A revealed typical classical endoglucanase features, i.e. efficient hydrolysis of soluble celluloses derivatives (CMC and HEC), slow hydrolysis of crystalline cellulose forms (Avicel) and no hydrolysis of pNP-Cel, rather than processive features. Although it has been suggested that carbohydrate binding modules could play a role in the destructuration of cellulose [14], PpCBM1<sup>15</sup> displayed no identified activity toward any of the polysaccharides tested. Furthermore, the number of CBMs of P. pastoris endoglucanase did not influence the activity of PpCel45A on soluble substrates, in good agreement with Baba et al. [18].

The patterns of CMC, Avicel, and cello-oligosaccharides hydrolysis end products were similar for the three constructions. The major CMC end products were cellobiose and cellobiose, with trace amounts of glucose produced at the end of the reaction, as reported for the endoglucanase from Penicillium decumbens [22] but different from T. reesei Cel45A (EG V) that produced only trace of cellobiose from CMC [10]. The major Avicel end products were also cellobiose, cellobiose, and trace amounts of glucose, but the amounts of end products released varied depending on the enzyme component. PpCBM1<sup>15</sup> was therefore more efficient than PpCel45A, which was more efficient than PpGH. No glucose was released from any of the oligosaccharides tested, confirming the endo mode of action of PpCel45A. End products profiles from hydrolyzed cello-oligosaccharides were in good agreement with previous studies [30,31] and suggest an organization of the catalytic site in six subsites. This hypothesis is strengthened by the
crystal structure of *Humicola insolens* GH45 [9] that revealed six subsites in the active site, with an organization from -4 to +2 [32].

Thermostable hemicellulases and cellulosases are gaining interest because they are well suited for harsh industrial process conditions [33]. Several thermostable endoglucanases have already been isolated from thermotolerant or thermophilic organisms like the GH45 endoglucanase from *Syncphalastrum racemosum* [34] and the Cel5A endoglucanase from *Thermoanaerobacter tengcongensis* [35]. Compared to other thermostable endoglucanases, *Pp* Cel45A thermal properties were remarkable since *P. pastoris* is a mesophilic organism. We observed that the high thermostability of *Pp* Cel45A (t1/2 of 6 h at 80°C) decreased with the removal of CBMs and glycosylated linkers (t1/2 of 3.8 h for *Pp* CBM15:GH and t1/2 of 3.2 h for *Pp* GH). Controversial data are reported in the literature about factors affecting GH thermostability. The comparison of catalytic module alone with catalytic module fused to CBM showed either improvement of thermostability [36], or on the contrary appeared to have a negative impact [37]. In our study, results strongly suggest that the presence of linkers and CBMs is responsible for the protection of *Pp* Cel45A catalytic module against thermal denaturation.

Removal of CBM moieties from cellulates or hemicellulases generally reduces their hydrolytic activity on insoluble substrates, whereas their activity remains unchanged on soluble substrates [38,39]. In this study, we have shown that the presence of five CBMs in *Pp* Cel45A was not an advantage for crystalline cellulose binding and hydrolysis. Instead of diminishing catalytic module capacities, the deletion of four CBMs favoured endoglucanase activity, as shown by the measurement of (i) specific activities, (ii) end-products formation, and (iii) binding parameters. The deletion of the five CBMs diminished the catalytic properties of *P. pastoris* endoglucanase only at 65°C, which is probably the consequence of a lower thermostability of *Pp* Cel45A. Molecular modeling of individual CBMs showed that all of them share similar conformation with a flat surface containing the three key aromatic residues involved in crystalline cellulose binding (not shown). Thus, the presence of these five CBMs appended to the catalytic module of *P. pastoris* endoglucanase remains enigmatic. Our hypothesis is that the steric hindrance of five CBMs could prevent the exhibition of the flat surface required for interaction with cellulose chains, limiting enzyme action. Upon deletion of four CBMs, the enzyme probably recovers an extended conformation due to the remaining glycosylated linker, thus more adapted to efficient degradation of crystalline cellulose.

The present study has revealed the specific features of the first characterized endo β-1,4 glucanase natively issued from yeast (Saccharomycetales order). The thermostability properties of this endoglucanase could be of great interest for lignocellulosic biomass applications related to the saccharification and fermentation processes. In these industrial processes aiming at decreasing conversion cost, a high dry matter content is necessary. Integration of a high temperature pre-hydrolysis step in the process sequence leading to rapid liquefaction is an attractive way to overcome the problems associated with high initial solids, i.e. poor mixing and mass transfer, high viscosity [33]. In these processes, high temperature enzymatic liquefaction is necessary as a pre-hydrolysis step to decrease viscosity and requires thermostable (hemi)cellulases. Endoglucanases were shown to have a superior ability to rapidly reduce the viscosity of pretreated wheat straw [40]. The high thermostability of *Pp* Cel45A would be an appropriate tool for an application in such process. Consolidated bioprocessing is also an approach developed for competitive production of soluble sugars from lignocellulosic biomass in which the hydrolytic enzymes are simultaneously produced in situ by the fermentative microorganism [41]. The dominant strategy for engineering an efficient biocatalyst is to express cellulosytic and hemicellulosytic enzymes in *Saccharomyces cerevisae*, but expression of fungal and bacterial genes in *S. cerevisiae* has some limitations [42]. The utilization of enzymes from yeast such as *P. pastoris* endoglucanase, could also be a promising way to optimize the ethanol production process without the addition of exogenous enzymes.

### Materials and methods

#### Materials and strains

*P. pastoris* X-33 and GS115 strains, *Escherichia coli* DH5α competent cells were purchased from Invitrogen (Cergy-Pontoise, France). Glucose, cellobiose, sodium...
carboxymethyl cellulose (CMC-Na, medium viscosity), hydroxyethyl cellulose (HEC), Avicel PH101, chitin, 4-nitrophenyl β-D-glucopyranose (pNP-glu), 4-nitrophenyl β-D-cello-bioside (pNP-cel), and 4-nitrophenyl β-D-lactopyranose (pNP-lac) were purchased from Sigma (Lyon, France). Barley β-glucan, lichenan, Konjac gluco-manann (KGM), mannan, and cello-oligosaccharides ranging from 3 to 6 residues (G3-G6) were purchased from Megazyme (Wicklow, Ireland). Walseth cellulose was provided by N. Lopes-Ferreira (IFPEN, Paris, France).

Bioinformatics analysis
The catalytic domain of 55 modularly annotated fungal members of family GH45 was extracted from the CAZy database (September 2011, [8]). All sequence alignments were performed with the multiple sequence alignment program MUSCLE v3.7 [43] under default conditions. Using an in-house modified version of Jalview [44], we calculated a maximum-likelihood distance matrix using the JTT matrix [45] and submitted it to hierarchical clustering using the Ward method [46]. The figure was edited using Dendroscope 2.6.1. [47].

Culture of P. pastoris GS115 on Walseth cellulose plates
P. pastoris GS115 was grown overnight at 30°C on agar plates containing minimal agar medium supplemented with either 2% dextrose or 6% walseth cellulose.

Construction of expression plasmids
Genomic DNA was purified from P. pastoris GS115 using NucleoSpin® Plant II kit (Macherey-Nagel, Düren, Germany). The DNA sequence [Genbank:PAS_chr4_0643] corresponding to [Genbank:CAV71902] was amplified by polymerase chain reaction (PCR) using specific primers (Table 5). EcoRI and XbaI restrictions sites were introduced respectively at the 5’ and 3’ end of the amplified fragments. DNA amplification was carried out with Expand High Fidelity polymerase (Roche Applied Science, Meylan, France) through 35 cycles of denaturation (20 seconds at 94°C), annealing (40 seconds at 55°C), and elongation (2 minutes at 72°C). DNA fragments were sub-cloned into pCRII-TOPO-TA vector (Invitrogen, Cergy-Pontoise, France) and transformed into E. coli DH5α competent cells (One shot chemically competent DH5α, Invitrogen) according to manufacturer’s instructions. Recombinant plasmids were purified (Wizard plus SV miniprep kit, Promega Charbonnières, France) and sequenced. Amplified fragments were digested by EcoRI and XbaI and ligated to the EcoRI-XbaI digested pPIC-ZαA vector (Invitrogen), in frame with both the α-factor and C-term-(His)6-tag encoding sequences. The resulting recombinant expression plasmids were linearized with Pmel and transformed into P. pastoris X-33 by electroporation as described in Couturier et al, 2011. The transformants were selected on YPDS plates (1% yeast extract, 2% peptone, 2% glucose, 1 M sorbitol, 2% agar) containing zeocin concentrations varying from 100 to 1000 μg/ml. After three days at 30°C, zeocin-resistant P. pastoris transformants were screened for methanol utilization (Mut+ phenotype) on MD (3.4% YNB, 1% ammonium sulphate, 2% glucose, 0.4 μg/ml D-biotin, 2% agar) and MM (3.4% YNB, 1% ammonium sulphate, 0.4 μg/ml D-biotin, 0.5% agar) plates. 10 Mut+ colonies were used for protein expression as follows: cultures were carried out in 10 ml BMGY (phosphate buffer 100 mM pH6, 1% yeast extract, 2% peptone, 3.4% YNB, 1% ammonium sulphate, 1% glycerol, 0.4 μg/ml D-biotin) in 50 ml loosely capped tubes at 30°C and shaken at 200 rpm, and expression was induced by transferring the cells into 2 ml BMMY (phosphate buffer 100 mM pH6, 1% yeast extract, 2% peptone, 3.4% YNB, 1% ammonium sulphate, 0.4 μg/ml D-biotin, 3% v/v methanol). The medium was supplemented daily with 3% methanol, and after 4 days the cultures were centrifuged and supernatants were analyzed by SDS-PAGE to select the best transformant for each enzyme.

Recombinant enzyme production and purification
For each enzyme, large-scale production (2 litres) was performed in 500 ml baffled flasks, each containing 100 ml of BMGY medium. Cells were grown overnight at 30°C and 200 rpm, and recovered by centrifugation the following day. Pellets were pooled and resuspended in 400 ml (100 ml of BMMY in 500 ml flasks), and induction was carried out for 4 days. Supernatant was then collected and concentrated as described in [27]. It was filtered through 0.2 μm membranes (Durapore GV membrane filters, 0.22 μm, Millipore, Molsheim, France), and concentrated (Vivaspin 10 kDa, PES, Sartorius, Palaiseau, France). A nickel chelate (GE Healthcare, Buc, France) column (His trap HP column, GE Healthcare, 0.7 cm × 5 cm) was connected to a FPLC Äkta (GE Healthcare) and equilibrated with the equilibration buffer (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole 10 mM) before purification. The concentrated supernatant was diluted in the equilibration

Table 5: Oligonucleotide sequences of the primers used in this study

| Orientation and primer | Sequence (5'-3') |
|------------------------|-----------------|
| Forward                |                 |
| GH45CBM1EcoF           | GAATTCGCTCAAGCTGAAACTGCATCCC |
| GH45SpiEcoF            | GAATTCGCTCAAGCTGAAACTGCATCCC |
| GH45CBM1EcoF           | GAATTCGCTCAAGCTGAAACTGCATCCC |
| Reverse                |                 |
| CBMP/xbAR              | TCTAGACCTGTATGCAATGCGTCTTC |
buffer and loaded onto the column at 4°C. The enzyme was eluted with Tris- HCl 50 mM pH 7.8, NaCl 150 mM, imidazole 150 mM, and the eluate was concentrated and dialysed with 50 mM sodium acetate buffer pH 5 as described in [27].

Biochemical characterization

SDS-PAGE was performed in 10% (w/v) polyacrylamide gels (Bio-Rad, Marne-la- coquette, France) using a Pharmacia LMW electrophoresis calibration kit (GE Healthcare). Proteins were detected by staining the gel with Coomassie blue. Protein concentrations were determined by the Folin-Lowry method [48] using Folin Ciocalteu’s phenol reagent (Sigma-Aldrich) and bovine serum albumin as a standard.

Enzymatic activity assays

Unless otherwise indicated, assay mixtures were prepared in citrate phosphate buffer pH 4.8. The activities of enzymes towards different cellulose derivatives (CMC, HEC, Avicel, Walseth cellulose) and complex polysaccharides (barley β-glucan, lichenan, chitin, Konjac glucomannan and mannan) were determined. All substrates excepting chitin were purchased as powders. To determine the activity towards chitin, dry chips of substrate were ground using a cutting mill (A11 cutting mill, IKA-Werke, Staufe, Germany) and further comminuted using a ball-mill grinder (MM 400, Retsch, Haan, Germany). A 0.2 mm selection screen was used to collect the finest fraction.

To determine specific activity towards the various substrates, the DNS method was used as described in [49]. Briefly, 100 μl of suitably diluted enzyme (2 μg of PpCBM11−5 or equimolar amounts of PpGH, PpCBM1:GH, and PpCel45A) were mixed with 100 μl of 2% substrate in 50 mM citrate phosphate buffer pH 4.8, or 100 μl buffer containing one disk of filter paper Whatman n°1. The reactions mixtures were incubated at 45°C for various time lengths. The reaction was terminated by the addition of 300 μl 1% w/v dinitrosalicylic reagent, and samples were heated at 95°C for 10 minutes. The reaction mixtures were cooled at room temperature, and 80 μl were transferred to a microtiter plate. Reducing sugar release was determined by measuring the absorbance at 540 nm, and a glucose standard curve (0 to 10 mM) was used to calculate the amount of protein releasing 1 μmol of sugar per minute. To take the different size of the four constructions into account, we chose to express our specific activities as U per μmol enzyme or U per nmol enzyme. Activity towards PNPCel, PnPLac, and PNPGluc was determined by measuring the release of p-nitrophenol in 50 mM citrate phosphate buffer, pH 4.8 after 2 hours incubation at 45°C. Either 2 μg of PpCBM11−5 or equimolar amounts of PpGH, PpCBM1:GH, and PpCel45A were used in a final reaction volume of 110 μl. Following the incubation, the same volume of 1 M sodium carbonate, pH 9 was added to terminate the reaction, and the release of p-nitrophenol was quantified at 410 nm using the molar extinction coefficient for pNP of 18300 M−1.cm−1. One unit of enzyme activity was defined as the amount of protein that released 1 μmol p-nitrophenol per minute.

Effect of pH and temperature

The apparent optimal temperature was estimated in the range 4-85°C (10 temperature tested) using CMC at 1% in 50 mM citrate phosphate buffer, pH 4.8 hydrolyzed for 30 minutes at 45°C. The apparent optimal pH was estimated in the range 3-10 (9 pH tested: pH 3-6 citrate phosphate buffer, pH 7-8 phosphate buffer, pH 9-10 carbonate-bicarbonate buffer) under the conditions used. For determination of thermal stability, the purified proteins were incubated at 65 or 80°C for various time lengths (30 minutes to 48 hours), and the residual activity towards CMC was subsequently determined as described above.

Determination of dissociation constant and carbohydrate binding capacity

For the binding assay, the reaction mixture contained 5 mg of Avicel PH101 and 40 to 200 μg of protein in a final volume of 200 μl of citrate phosphate buffer (pH 4.8). All assays were carried out in triplicates. For each assay, the mixture was incubated at 4°C for 1 h with vertical rotation (24 rpm). Substrate was removed by filtration and the concentration of unbound protein ([UB], μM) in the supernatant was measured by the Folin-Lowry method. The bound protein concentration ([B], μmol per g of substrate) was determined by subtracting [UB] from the total protein concentration. Adsorption parameters were determined based on typical double-reciprocal plots using the equation \[ B = UB \times \frac{B_{\text{max}}}{(K_d + [UB])} \], where \( K_d \) (μM) and \( B_{\text{max}} \) (μmol per g of substrate) are the equilibrium dissociation constant and the maximum amount of protein bound, respectively [50].

Analysis of sugar release by HPAEC-PAD

Monosaccharides and oligosaccharides generated after hydrolysis of cello-oligosaccharides (glucotetraose G4, glucopentaose G5, and glucohexaose G6) and cellulose derivatives (CMC and Avicel) were analyzed by high performance anion exchange chromatography (HPAEC) coupled with pulsed amperometric detection (PAD) (ICS 3000; Dionex, Sunnyvale, USA) equipped with a carbo-Pac PA-1 analytical column (250 x 4 mm). 20 μl of suitably-diluted enzyme were incubated at 40°C for various time lengths with 180 μl of 1 mM substrate in 50 mM acetate buffer pH 4.8. 10 μl of enzymatic reactions were stopped by the addition of 90 μl of 100 mM NaOH before injection (5 μl) into the HPAEC system. Elution was carried out in
130 mM NaOH using a 25-min linear gradient program from 100% A (130 mM NaOH) to 60% A and 40% B (NaOAc, 500 mM; NaOH, 130 mM). Calibration curves were plotted using β-1,4-cello-oligosaccharides as standards from which response factors were calculated (Chromelone program, Dionex) and used to estimate the amount of products released. The specificity constants were calculated using the Matsui equation [51,52].

Abbreviation list
GH: glycoside hydrolase; CBM: carbohydrate-binding module; CMC: carboxymethyl cellulose; DNS: dinitrosalicylic acid; HEC: Hydroxyethyl cellulose; pNP: paranitrophenol; *PpCel45A*: full-length *P. pastoris* GH45 endoglucanase; *PpCBM1*-GH: catalytic module with its proximal CBM1; *PpGH*: catalytic module; *PpCBM1*-5: five CBM1 modules without catalytic modules.

Acknowledgements
This study was funded by the Futurol project. We thank N. Lopes-Ferreira for providing the Walthell cellulose and A. Blanchard for Dionex analyses. P.M. Coutinho thanks the French National Research Agency (ANR, program E-TRICEL ANR-07-BIOE-006) for financial support.

Author details
1 INRA, UMR 1163 BCF, 13288 Marseille, France. 2 CNRS, AFBMB, 13288 Marseille, France. 3 Aix-Marseille Universités, 13288 Marseille, France.

Authors’ contributions
MC participated in the design of the study, carried out most of the bench work and drafted the manuscript. JF and MH helped to carry out experimental work. PM carried out the phylogenetic analysis. LDM, DN and PMC participated in the coordination of the study and proofread the manuscript. JGB conceived of the study, participated in its design and coordination and helped to draft the manuscript and submitted the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 10 October 2011 Accepted: 6 December 2011 Published: 6 December 2011

References
1. Marguet A, Hahn-Hagerdal B, Edlund M, Slade R, Monot F: New improvements for lignocellulosic ethanol. Cur Opin Biotechnol 2009, 20:372-380.
2. Himmel M: Deconstructing the plant cell wall for bioenergy Blackwell Pub 2008.
3. Sánchez C: Deconstructing the plant cell wall for bioenergy.
4. Henrissat B, Driguez H, Viet C, Schülein M: New carbohydrate-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J 1998, 337:77-81.
5. Wang L, Zhang Y, Gao P: A novel function for the cellulose-binding module of cellobiohydrolase I. Sci China C Life Sci 2008, 51:620-629.
6. Arantes V, Saddler JN: Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnol Biofuels 2010, 3:3-4.
7. Bolam D, Cruela A, Michel-Durand, Miller RC, Warren RAJ: Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 1991, 55:303-315.
8. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The Carbohydrate-Active Enzymes database (CAzy): an expert resource for glycoconjugomics, Nucleic Acids Res 2009, 37:233-238.
9. Davies GJ, Dobson GG, Hubbard RE, Tolley SP, Dauter Z, Hjort C, Mikkelsen JN, Rasmussen G, Schülein M: Structure and function of endoglucanase V. Nature 1993, 365:362-364.
10. Karlsson J, Siika-Aho M, Tenkanen M, Tjerneld F: Enzymatic properties of the low molecular mass endoglucanases Cell12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. J Biotechno 2002, 99:63-78.
11. Gilkes NR, Hennissat B, McFarland KC, Hjort C, Mikkelsen JN, Rasmussen G, Schülein M: Structure and function of endoglucanase V. Nature 1993, 365:362-364.
12. Bolam D, Cruela A, McQueen-Mason S, Simpson S, Williamson MP, Rixon JE, Bolan A, Hazlewood GP, Gilbert H: Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J 1998, 337:77-81.
13. Wang L, Zhang Y, Gao P: A novel function for the cellulose-binding module of cellobiohydrolase I. Sci China C Life Sci 2008, 51:620-629.
14. Arantes V, Saddler JN: Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnol Biofuels 2010, 3:3-4.
15. Bolan A, Bolam D, Gilbert H, Davies GJ: Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 2004, 382:769-781.
16. Guillén D, Sánchez S, Rodríguez-Sandoja R: Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol 2010, 85:1241-1249.
17. Graud E, Cuny T: Molecular characterization of the alpha-amylase genes of *Lactobacillus plantarum* A6 and *Lactobacillus amylovorus* reveals an unusual 3’ end structure with direct tandem repeats and suggests a common evolutionary origin. Gene 1997, 198:149-157.
18. Baba Y, Shimokura A, Koga J, Kubota H, Kono T: Alternative splicing produces two endoglucanases with one or two CBM in *Mucor circinelloides*. J Bacteriol 2005, 187:3045-3051.
19. De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Couturier M, Haon M, Coutinho PM, Henrissat B, Lesage-Meessen L, Mikkelsen JM, Rasmussen G, Schülein M: Structure and function of endoglucanase V. Nature 1993, 365:362-364.
20. Garcia-Valvé S, Romeu A, Palau J: Horizontal gene transfer in bacterial and archaea complete genomes, Genome Res 2000, 10:1719-1725.
21. Qin Y, Wei X, Liu X, Wang T, Qu Y: Purification and characterization of recombinant endoglucanase of *Trichoderma reesei* expressed in *Saccharomyces cerevisiae* with higher glycosylation and stability. Protein Expr Purif 2008, 58(1):162-167.
22. Liu G, Wei X, Qin Y, Qu Y: Characterization of the endoglucanase and glucomannanase activities of a glycoside hydrolase family 45 protein from *Penicillium decumbens*. J Gen Appl Microbiol 2010, 56:233-239.
23. Seon Park J, Russell JR, Wilson DR: Characterization of a family 45 glycosyl hydrolase from *Fibrobacter succinogenes* S58. Anaerobe 2007, 13(2):83-88.
24. Lee SJ, Lee KS, Kim SR, Guu ZZ, Kim YS, Yoon HJ, Kim I, Kang PD, Sohn HD, Jin BR: A novel cellulase gene from the mulberry longicorn beetle, *Aproia german*: gene structure, expression, and enzymatic activity. Comp Biochem Physiol B Biochem Mol Biol 2005, 140(4):551-560.
25. Bauer S, Vasu P, Persson S, Mart AJ, Somerville CD: Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant wall cells. Proc Natl Acad Sci USA 2006, 103(30):11417-11422.
26. Do BC, Dang TT, Berrin JG, Halrich D, To KA, Sigoillot JC, Yamabhai M: Cloning, expression in *Pichia pastoris* and characterization of a thermostable GH5 mannan endo-1,4-beta-mannosidase from *Aspergillus niger* BKO1. Microb Cell Fact 2009, 8:59.
27. Couturier M, Haon M, Coutinho PM, Hennissat B, Legeay-Meessen L, Berrin JG: *Podospora anserina* hemcellulases potentiate the *Trichoderma reesei* secretome for saccharification of lignocellulosic biomass. Appl Environ Microbiol 2011, 77(1):237-246.
28. Igarashi K, Ishida T, Hori C, Samejima M: Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete *Phanerochaete chrysosporium*. Appl Environ Microbiol 2008, 74:5628-5634.
29. Shibuya H, Kikuchi T: Purification and characterization of recombinant endoglucanases from the pine wood nematode Bursaphelenchus xylophilus. Basic Biotechnol Biochem 2008, 72:1325-1332.

30. Morita T, Murashima K, Nakane A, Yana K, Sumida K, Koga M, Murakami T, Kono T: Molecular cloning of endoglucanase genes rce1, rce2, rce3 from Rhizopus oryzae. J Bacteriol 2003, 185:1749-1756.

31. Schou C, Rasmussen G, Kaltoft MB, Henriksen B, Schlaen M: Stereochemistry, specificity and kinetics of the hydrolysis of reduced cellobiose by nine cellulases. Eur J Biochem 1995, 217:947-953.

32. Davies GJ, Tolley SP, Henriksen B, Hjort C, Schlaen M: Structures of oligosaccharide-bound forms of the endoglucanase V from Humicola insolens at 1.9 Å resolution. Biochemistry 1995, 34:16210-16220.

33. Wikari L, Alapuranen M, Puranen T, Veheanenpa J, Silka-Aho M: Thermostable enzymes in lignocellulose hydrolysis. Adv Biochem Eng Biotechnol 2007, 108:121-145.

34. Wonganu P, Pootaranakit K, Boonyapakron K, Champradee V, Tanaponpipat S, Euwilaichitr L: Cloning, expression and characterization of a thermotolerant endoglucanase from Bacillus subtilis strain 6D23 (BBC18080) in Pichia pastoris. Biochem Eng J 2008, 58:78-86.

35. Lang C, Xue Y, Fioroni M, Rodriguez-Ropero F, Zhou C, Schwanenberg U, Ma Y: Cloning and characterization of a thermostable and halo-tolerant endoglucanase from Thermotoga maritima MB-4. Appl Microbiol Biotechnol 2011, 89:315-326.

36. Pham TA, Berrin JG, Record E, To KA, Sigillott JC: Hydrolysis of softwood by Aspergillus mannanase: role of a carbohydrate-binding module. J Biotechnol 2010, 148:163-170.

37. Selvaraj T, Kim SK, Kim YH, Jeong YS, Kim YJ, Phuong ND, Jung KH, Kim J, Yun HD, Kim H: The role of carbohydrate-binding module (CBM) repeat of a multimodular xylanase (XynM) from Clostridium thermocellum in cellulose and xylan binding. J Microbiol 2010, 48:656-661.

38. Irwin DC, Spezio M, Walker LP, Wilson DB: Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol Bioeng 1993, 42:1002-1013.

39. Sumitani J, Tottori T, Kawaguchi T, Arai M: New type of starch-binding domain: the direct repeat motif in the C-terminal region of Bacillus sp. no. 195 alpha-amylase contributes to starch binding and raw starch degrading. Biochem J 2000, 350:477-484.

40. Szijártó N, Horan E, Zhang J, Puranen T, Silka-Aho M, Wikari L: Thermostable endoglucanases in the liquefaction of hydrothermally pretreated wheat straw. Biotechnol Biofuels 2011, 4(1):2.

41. Elkins JG, Raman B, Keller M: Engineered microbial systems for enhanced conversion of lignocellulosic biomass. Curr Opin Biotechnol 2010, 21:657-662.

42. van Zyl WH, Lynd LR, den Haan R, McBride JE: Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae. Adv Biochem Eng Biotechnol 2007, 108:205-235.

43. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32:1792-1797.

44. Waterhouse AM, Procter JB, Martin DMA, Clamp ML, Barton GJ: Jalview Version 2: a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009, 25:1189-1191.

45. Jones DT, Taylor WR, Thornton JM: The rapid generation of mutation data matrices from protein sequences. Comput Appl Bios 1992, 8:275-282.

46. Ward JH: Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963, 58:236-244.

47. Dendroscope 2.6.1. (http://ab.inf.uni-tuebingen.de/software/dendroscope/).

48. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ: Protein measurement with the Folin phenol reagent. J Biol Chem 1951, 193:265-275.

49. Navarro D, Couturier M, da Silva GC, Berrin JG, Rouau X, Esther M, Bigmon C: Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. Microb Cell Fact 2010, 9:58.

50. Okazaki F, Tamaru Y, Hashikawa S, Li YT, Araki T: Novel carbohydrate-binding module of beta-1,3-xylanase from a marine bacterium, Alcaligenes sp. strain SY-X34. J Bacteriol 2002, 184(19):5299-5303.

51. Matsui I, Ishikawa K, Matsui E, Miyairi S, Fukui S, Honda K: Subsite structure of Saccharomycopsis alpha-amylase secreted from Saccharomyces cerevisiae. J Biochem 1991, 109:566-569.

52. Berrin JG, Ajandouz El H, Georis J, Arnau F, Juge N: Substrate and product hydrolysis specificity in family 11 glycoside hydrolase: an analysis of