Exploring family 6 glycoside hydrolases from *Podospora anserina*: insights into exo- and endo-glucanase activities

Laetitia Poidevin¹²†, Julia Feliu², Annick Doan², Jean-Guy Berrin³, Mathieu Bey², Pedro M. Coutinho³, Bernard Henrissat¹, Eric Record², Senta Heiss-Blanquet¹#

¹IFP Energies nouvelles, 1 et 4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France
²INRA, Aix Marseille Université, UMR 1163 de Biotechnologie des Champignons Filamenteux, Polytech, CP 925, 13228 Marseille cedex 09, France
³Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, CNRS UMR 7257, 163 avenue de Luminy, 13228 Marseille, France

†present address: Department of Plant Pathology and Microbiology, University of California, Riverside, CA 92521, USA

#Corresponding author. Mailing address: IFP Energies nouvelles, 1 et 4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France. Phone: +33 147 527 256. Fax: +33 147 527 058. E-Mail: senta.blanquet@ifpen.fr
Abstract

The ascomycete *Podospora anserina* is a coprophilous fungus that grows at late stages on droppings of herbivores. Its genome encodes a large diversity of carbohydrate-active enzymes. Among them, four genes encode glycoside hydrolases from family GH6, the members of which comprise putative endoglucanases and exoglucanases, some of them exerting important functions for biomass degradation in fungi. Therefore, this family was selected for functional analysis. Three of the enzymes, *PaCel6A*, *PaCel6B* and *PaCel6C*, were functionally expressed in the yeast *Pichia pastoris*. All three GH6 enzymes hydrolysed crystalline and amorphous cellulose, but were inactive on hydroxyethyl cellulose, mannan, galactomannan, xyloglucan, arabinoxylan, arabinan, xylan and pectin. *PaCel6A* had a catalytic efficiency on cellotetraose comparable to *TrCel6A*, but *PaCel6B* and *PaCel6C* were clearly less efficient. *PaCel6A* was the enzyme with the highest stability at 45°C, while *PaCel6C* was the least stable enzyme, loosing more than 50% of its activity after incubation at temperatures above 30°C for 24h. In contrast to *TrCel6A*, all three studied *P. anserina* GH6 cellulases were stable over a wide range of pH and conserved high activity at pH values up to 9. Each enzyme displayed a distinct substrate and product profile highlighting different modes of action, with *PaCel6A* being the most similar enzyme to *TrCel6A*. *PaCel6B* was the only enzyme with higher specific activity on CMC than on Avicel and showed lower processivity than the others. Structural modelling predicts an open catalytic cleft, suggesting that *PaCel6B* is an endoglucanase.
**Introduction**

Cellulose, a polysaccharide of β-1,4-linked D-glucose units, is the most abundant biopolymer on earth. It is the main constituent of plant cell walls where it forms a tight complex together with hemicelluloses and embedded in the lignin matrix. Cellulose is a recalcitrant material organised in microfibrils and composed of highly ordered glucan chains interlinked by hydrogen bonds. Depending on the source of cellulose, these fibrils have more or less crystalline character and for deconstruction of these complex structures microorganisms have developed specialized enzymatic systems. All cellulolytic organisms produce multiple enzymes for cellulose degradation, but three main catalytic activities are necessary for complete hydrolysis: exo-glucanases (or cellobiohydrolases) which attack cellulose chains from the chain ends, endo-glucanases that cleave the cellulose chain randomly, while β-glucosidases hydrolyse cellobiose, the reaction product of cellobiohydrolases. For efficient degradation, all three enzymatic activities have to be present, and synergistic interactions have been shown to be essential for a rapid degradation process (1-4). While exo-endo synergy can be easily explained by endoglucanase creating new chain ends for exoglucanases, exo-exo synergy still lacks a fully satisfactory explanation.

In contrast to synergistic interactions, the reaction mechanism of isolated enzymes is better understood since structure-function studies have led to the identification of catalytic residues for cellobiohydrolases and endoglucanases (5-11). Sequence similarity, and therefore structural properties, are at the origin of the classification of these enzymes in the CAZy (Carbohydrate Active enZymes) database, facilitating the assignment of function (12). Fungal endo-glucanases are classified in glycoside hydrolase families GH5, GH6, GH7, GH9, GH12, GH45 and GH74, while fungal exo-glucanases can be found in families GH6, GH7, and GH48.
At the three-dimensional level, cellobiohydrolases are characterized by a typical tunnel-shaped active site, with the roof of the tunnel formed by long flexible loops (5,13-15). By contrast, their homologous endoglucanases bear shorter loops converting their active site into a classical cleft. This difference in active site topology has consequences in the action pattern of the enzymes. The cellulose chain end is believed to enter the tunnel active site of cellobiohydrolases and to be cut processively into cellobiose units as it threads through the tunnel (5,16). Instead the open cleft active site of endoglucanases is thought to allow random binding and cleavage along the cellulose chain.

An increasing number of genomic and proteomic studies on fungi reveal a large variety of lignocellulose-degrading enzymes present in these organisms. A very interesting ascomycete is *Podospora anserina* which grows on herbivore dung. Its mycelium develops at a late stage, after the most easily utilisable biomass components, such as hemicellulose and pectin, have already been degraded by other species. *P. anserina* is therefore hypothesized to be able to specifically use the recalcitrant parts of lignocellulose. The analysis of its genome has indeed revealed an impressive number of genes encoding CAZymes and enzymes putatively involved in lignin degradation, including oxidoreductases, cellobiose dehydrogenase, copper radical oxidases or laccases, which is a rather atypical feature for ascomycetes (17). The total number of putative glycoside hydrolases is similar to other sequenced ascomycetes, but it has the largest panel of cellulose-degrading enzymes and carbohydrate binding modules (CBMs). For instance, *P. anserina* has 30 genes encoding putative cellulases belonging to the families cited above, against only 8 for *T. reesei* which is particularly low compared to most other carbohydrate-degrading ascomycetes (18).

The specific activity of cellobiohydrolases (CBH) on soluble and insoluble substrates has been shown to be lower than that of endoglucanases (19) and the CBH activity is probably rate-limiting for cellulose hydrolysis (20). For efficient hydrolysis, a large amount of CBH
enzyme is therefore needed in cellulolytic complexes, such as the one of *T. reesei*. In the latter
organism, CBH are the most abundant enzymes, i.e. Cel7A (CBH1) makes up 40-60 % and
Cel6A (CBH2) about 20-30 % of the total amount of secreted proteins (21). An enhancement
of Cel6A was shown to be beneficial for hydrolytic activity of the complex (22) and Cel6A
deletion mutants showed a 33 % decrease in saccharification efficiency (23). Intriguingly, *T.
reesei* only contains one enzyme of the GH6 family, in contrast to *H. insolens* and
*Myceliophthora thermophila*, two other fungi that produce efficient cellulolytic cocktails
containing three family GH6 enzymes. *H. insolens* Cel6A and Cel6B have been shown to
display endoglucanase and cellobiohydrolase activity, respectively (24-26). *P. anserina* has
four genes encoding putative GH6 enzymes, but to date nothing is known about the enzymatic
properties of this important family of cellulose-degrading enzymes in this organism. The
reason for the existence of multiple enzymes is not known and the question arises as to
whether there is redundancy or if each enzyme has a distinct role.

To answer this question, we undertook the cloning of the four *P. anserina* genes encoding
putative GH6 cellulases to express them heterologously. We successfully produced three of
the enzymes, termed *PaCel6A*, *PaCel6B* and *PaCel6C*. Substrate and product profiles of the
purified enzymes revealed significant differences concerning the specificities and modes of
action on cellulose model substrates. The biochemical characteristics of *P. anserina* GH6
cellulases were compared to those of *T. reesei* Cel6A, revealing differences in their mode of
action.

**Materials and methods**

**Culture media**
P. anserina was grown at 27°C on M2 medium composed of: KH2PO4 0.25 g l⁻¹, K2HPO4 0.3 g l⁻¹, MgSO4 7H2O 0.25 g l⁻¹, urea 0.5 g l⁻¹, thiamin 0.05 g l⁻¹, biotin 0.25 g l⁻¹, citric acid 2.5 mg l⁻¹, ZnSO4 2.5 mg l⁻¹, CuSO4 0.5 mg l⁻¹, MnSO4 125 µg l⁻¹, boric acid 25 µg l⁻¹, Na2MoO4 25 µg l⁻¹, iron alum 25 µg l⁻¹, dextrin 5 g l⁻¹, yeast extract 10 g l⁻¹, and adjusted to pH 7.0 with KH2PO4. For P. pastoris, three media were used, MM composed of yeast nitrogen base (YNB) (Difco) 3.4 g l⁻¹, ammonium sulphate 10 g l⁻¹, agar 20 g l⁻¹, 2 ml of biotin 200 g l⁻¹, 5 ml of pure methanol; BMGY containing YNB 3.4 g l⁻¹, ammonium sulphate 10 g l⁻¹, glycerol 10 g l⁻¹, yeast extract 10 g l⁻¹, peptone 10 g l⁻¹, 100 ml of phosphate buffer 1M pH 6.0, 2 ml of biotin 200 g l⁻¹; and BMMY which is identical to BMGY except that it contains 30 ml l⁻¹ of pure methanol instead of glycerol.

Cloning procedures

The P. anserina strain S mat⁺ used in this study was kindly provided by P. Silar (CNRS, Paris, France). P. anserina was grown in baffled flasks at 120 rpm and 27°C on M2 medium (KH2PO4 0.25 g l⁻¹, K2HPO4 0.3 g l⁻¹, MgSO4 7H2O 0.25 g l⁻¹, urea 0.5 g l⁻¹, thiamin 0.05 mg l⁻¹, biotin 0.25 µg l⁻¹, citric acid 2.5 mg l⁻¹, ZnSO4 2.5 mg l⁻¹, CuSO4 0.5 mg l⁻¹, MnSO4 125 µg l⁻¹, boric acid 25 µg l⁻¹, Na2MoO4 25 µg l⁻¹, iron alum 25 µg l⁻¹, dextrin 5 g l⁻¹, pH 7) supplemented with 1% Avicel cellulose; with or without induction by 0.1% sophorose 1 hour prior to harvesting the mycelia. Total RNA was extracted from 3 or 5 days-old cultures, with the RNeasy plant kit (Qiagen, Courtaboeuf, France), and cDNA were synthesized using SuperScript reverse transcriptase (Life Technologies, NY, USA), according to the manufacturer’s instructions. PaCel6B and PaCel6D were amplified by PCR using Pfu DNA Polymerase (Promega, WI, USA) and the following primers: PaCel6B-F: 5'-TAG AAT TCG CCC CTT CCC CGA CCA CC-3' and PaCel6B-R: 5'-GAT CTA GAC CGA GAA GGG AAG GGT TAG A-3', PaCel6D-F: 5'-TAG AAT TCT CTC CCC TTG AGG CAC GC-3' and PaCel6D-R: 5'-GAT CTA GAC CGA GAA GGG AAG GGT TAG A-3'.
PaCel6D-R: 5'-GAT CTA GAC CCA AGC ACT GCG AAT ACC A- 3'. PaCel6A and PaCel6C could not be obtained by amplification and their coding sequences were synthesized after codon optimization for expression in *P. pastoris* (Eurogentec, Belgium). The TrCel6A gene was amplified from cDNA obtained from the *T. reesei* CL847 strain. Coding sequences for PaCel6A, PaCel6B, PaCel6C and TrCel6A were cloned in the pPICZαA vector (Life Technologies), in frame with the yeast α-secretion factor, and with C-terminal HA- and His-tags. For an unknown reason and despite multiple trials, the PaCel6D coding sequence, which was obtained after RT-PCR, could not be cloned into the expression vector, and therefore not be produced. The enzyme is therefore termed PaGH6D in the following.

Competent X33 yeast cells were prepared and transformed by electroporation following the EasySelect Pichia Expression Kit protocol (Life Technologies). Expression vectors were linearized by *Pme*I (New England BioLabs, MA, USA) prior to transformation. After electroporation, cells were spread on YPDS-Zeocine plates. Transformants were selected by their lower growth rates on MM plates (containing methanol as inducer) compared to growth on MD plates.

**Enzyme production and purification in *Pichia pastoris***

Culture supernatants of 10 clones were first analyzed by small-scale protein productions in a total volume of 10 ml BMGY as described by (27). When the OD$_{600}$ reached between 2 and 6, cells were transferred to 2 ml BMMY and grown at 30°C, with daily addition of 3% methanol (v/v). After SDS-PAGE analysis, clones with the highest secretion levels were selected and cultured in 200 ml BMGY starting culture and transferred to 40 ml BMMY. Recombinant proteins were recovered after 5 days of methanol induction. After 10 min centrifugation at 4000g, supernatants were passed through a 0.2 µm filter. Samples were then concentrated 10 times in binding buffer (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole...
10 mM) by ultrafiltration in a Vivaspin 20 column (polyethersulfone membrane, 10 kDa cut-off, Sartorius, France). Subsequently, recombinant proteins were purified on a 5 ml HisTrapTM column (GE Healthcare) connected to a FPLC Äkta apparatus (GE Healthcare), following the manufacturer’s instructions. Enzymes were eluted with binding buffer supplemented with 150 mM imidazole. A final ultrafiltration step was used to concentrate proteins in acetate buffer 50 mM, pH 5. Protein homogeneity was checked on a 12 % SDS polyacrylamide gel, followed by Coomassie staining.

**Protein analysis**

Deglycosylation of 2 µg recombinant protein was performed with 2 µl of either Endo-α-N-acetylgalactosaminidase and 2 µl Neuraminidase or 2 µl Endoglycosidase H (Endo H) (all from New England Biolabs), for 2 h at 37°C following the manufacturer’s instructions.

N-terminal sequences were determined by Edman degradation with a Procise® cLC Sequencing System, Model 494cLC (Applied Biosystems), from purified protein samples electroblotted onto a PVDF membrane (Life Technologies).

**Biochemical characterization**

**Activity assays**

Activity profiles on solid substrates were performed in microtiter plates, by measuring the release of reducing sugars with 3,5-dinitrosalicylic acid (DNS), as published earlier (28). The total volume was 100 µl, containing 6.4 mg l⁻¹ of purified enzymes in 50 mM acetate buffer pH 5 for *TrCel6A* or citrate/phosphate buffer pH 7 for the three *PaCel6* and 10 g l⁻¹ substrate.

Substrates used were Avicel PH-101 cellulose, carboxymethylcellulose (CMC), hydroxyethyl cellulose (HEC), birchwood xylan, pectin (all from Sigma-Aldrich), barley β-glucan, 1,4 β-D mannana, carob galactomannan, sugar beet arabinan, wheat arabinoxylan, arabinogalactan,
xyloglucan and konjak glucomannan (all from Megazyme, Wicklow, Ireland). Reactions were incubated at 35°C or 45°C, depending on enzyme stability for 30 min. Reactions were stopped by boiling and centrifuged at 4000 x g for 3 min.

Activities on soluble cellobioisacharides were measured at optimal pH and temperature with 20 mg/L cellobioisacharides and 10 mg L⁻¹ enzyme. Reactions were stopped after 30 minutes by placing them into boiling water for 5 minutes. Reaction products were analysed with a HPLC ISC300 Dionex system as described earlier (29). For determination of kinetic constants, enzymes were incubated at an appropriate concentration (0.5 – 2 nM) in a 50 mM phosphate buffer pH 7 and at optimal temperature (PaCel6A : 45°C, PaCel6B and PaCel6C : 35°C) with cellotetraose (0.5 – 30 µM) for up to 20 minutes. Reactions were inactivated by placing them in a boiling water bath for 5 minutes and reaction products were analysed by HPEAC-PAD (high-performance anion exchange chromatography coupled to a pulsed amperometric detector). Kinetic constants were determined using the least square method.

Product profiles were established in duplicates on Avicel and CMC at 1% dry matter in 10 mM citrate/phosphate buffers. The substrates were incubated with 10 mg g⁻¹ enzyme for 15 minutes and 24h under optimal conditions (TrCel6A : 45°C, pH 5; PaCel6A : 45°C, pH 7; PaCel6B : 35°C, pH 7 and PaCel6C : 25°C, pH 6). Reactions were stopped by boiling for 5 min and reaction products (glucose, cellobiose and cellotriose) analysed by HPEAC-PAD. Processivity was determined using oligosaccharide ratios as described by (30) and (31).

**Effect of pH and Temperature on enzymatic activity**

Apparent optimal pH was estimated using 1% Avicel in a total volume of 100 µl of 50 mM citrate/phosphate buffer (pH 3 to 7), 50 mM Tris-HCl (pH 7 to 9) and 50 mM Tris-Maleate (pH 9-10). Optimal reaction temperature was studied by incubating the enzymes with 1%
Avicel for 15 min at temperatures between 30 and 70°C, at pH5 for *TrCel6A* and pH7 for *PaCel6s*. For enzymatic stabilities, enzymes were incubated in different buffers or at different temperatures for 24 h, and residual activity measured as described above. Reducing sugars were determined according to a protocol adapted from (32). Briefly, 40 µl of supernatant or glucose standard were added to 120 µl of reagent 1, made extemporarily from 2 volumes of reagent 1A (K₃Fe(CN)₆ 0.5 g l⁻¹, K₂HPO₄ 34.8 g l⁻¹, pH 10.6) and 1 volume of reagent 1B (Na₂CO₃ 5.3 g l⁻¹ and KCN 0.65 g l⁻¹). The plates were sealed and heated to 96°C. After 10 min at room temperature, 80 µl of the reaction mix were added to 40 µl of reagent C (FeCl₃ 2.5 g l⁻¹, polyvinyl pyrrolidon K₂5 20 g l⁻¹, H₂SO₄ 2N). The optical density was measured at 520 nm after 15 min incubation at room temperature in the dark. Each point was the mean of three independent measurements with a standard deviation of 2 - 10 %.

Cellulose binding assay

The adsorption capacity of enzymes to Avicel were determined after incubating 60 µg of protein with 300 µl 1% Avicel PH-101 (Sigma-Aldrich) in 50 mM acetate buffer pH 5 for *TrCel6A* or citrate/phosphate buffer pH 7 for the three *PaCel6*, for 4 hours at 4°C, under agitation. After centrifugation, the amount of cellulases left in the supernatant was quantified using the Bio-Rad protein assay kit with bovine serum albumin as a standard (Bio-Rad, France). Controls without protein or with BSA were analyzed in parallel.

Structure modeling and bioinformatic analysis

Sequence alignments were done using ClustalW on the EMBL-EBI server ([http://www.ebi.ac.uk/Tools/msa/clustalw2/](http://www.ebi.ac.uk/Tools/msa/clustalw2/)). *PaCel6s* structures were predicted using the PHYRE2 server ([http://www.sbg.bio.ic.ac.uk/phyre2/](http://www.sbg.bio.ic.ac.uk/phyre2/)) (33). Models were visualized with PyMOL.
(The PyMOL Molecular Graphics System, Version 1.1, DeLano Scientific, USA). All PaCel6 models were registered in the Protein model database [http://www.caspur.it/PMDB].

The NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc/) and NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) servers were used for glycosylation sites prediction.

Results

Sequence analysis and expression of *P. anserina* GH6 enzymes

The annotated *P. anserina* genome has four genes encoding family GH6 glycoside hydrolases (*PaCel6A*: XM_001903135, *PaCel6B*: XM_001903174, *PaCel6C*: XM_001903858, *PaGH6D*: XM_001903610). *PaCel6A* and *PaGH6D* both harbour a cellulose-binding module of family CBM1 at their N-terminus, explaining their slightly higher molecular weight (Table 1). *PaCel6B*, *PaCel6C* and *PaGH6D* catalytic domains have very limited identity to the *T. reesei* cellobiohydrolase 2 enzyme (*TrCel6A*).

At the time of writing, the most similar proteins that have been biochemically characterized are the exoglucanase A from *Humicola insolens* (Q9C1S9) displaying 72% and 50% identity to *PaCel6A* and *PaCel6C* respectively, and the *H. insolens* endoglucanase 6B (Q7SIG5) with 54 and 82% identity to *PaCel6B* and *PaGH6D*. Identities between the four *PaCel6* enzymes ranged between 25 and 45%, *PaCel6B* and *PaGH6D* displaying the highest similarity to each other and *PaCel6A* and *PaGH6D* the least.

Sequence analysis of the four *PaCel6* enzymes was realized by a multiple alignment including *TrCel6A* (Fig. 1). All amino acids which are known to be involved in catalysis, namely Asp221 (following the *TrCel6A* numbering), and Asp175 (7), are conserved in all four *PaCel6* sequences. Tyr169, proposed to participate in catalysis by causing distortion of the glucose ring undergoing catalysis (34), was also found to be conserved in all four proteins.
Two other aspartic acid residues, Asp401 and Asp412, not directly involved in catalysis, but contributing to full enzymatic activity (11), are conserved in PaCel6A and PaCel6C, but Asp412 is absent in PaCel6B and PaGH6D. The most important amino acids involved in binding of the glucose moiety at the -2, +1, +2 and +4 subsites (throughout this paper we follow the subsite nomenclature proposed by (35)) are also present in all four PaCel6s (Trp135, Trp269, Trp272, Trp367) (9), but Glu107 (-3 subsite) is not conserved in PaCel6B and PaGH6D (11).

In order to induce the expression of PaCel6 genes, P. anserina was grown on Avicel as sole carbon source. Under this condition, only cDNAs encoding PaCel6B and PaGH6D were obtained, whereas synthetic genes had to be generated for cloning of PaCel6A and PaCel6C in P. pastoris. For an unknown reason, cloning of the PaGH6D coding sequence failed despite several trials and the corresponding protein could therefore not be produced in P. pastoris. The three other proteins, PaCel6A, PaCel6B and PaCel6C, finally yielded 30 to 130 mg l⁻¹ of protein after induction of P. pastoris transformants by methanol. In parallel, TrCel6A was also expressed in P. pastoris for comparative studies. After purification on a Ni-column, proteins were analysed by SDS-PAGE (Fig. 2). All proteins displayed higher apparent molecular weights than the predicted ones, suggesting glycosylation by the host. Whereas PaCel6B and PaCel6C seemed only slightly glycosylated, PaCel6A and TrCel6A had an apparent molecular weight of about 70 kDa (theoretical weight of TrCel6A: 47.9 kDa).

For all proteins, both N- and O-glycosylation sites are predicted (table 1). Deglycosylation by Endo H led to reduction of molecular weights of all proteins. The most important reduction was observed for TrCel6A and PaCel6B which also have the highest number of predicted N-glycosylation sites (fig. 2). A high number of O-glycosylation sites are predicted in the linker domain of TrCel6A and PaCel6A (table 1). Treatment with Endo-α-N-acetylgalactosaminidase reduced the molecular weight of all recombinant proteins. Several
bands appeared after deglycosylation of PaCel6B and PaCel6C, which is supposed to be due
to incomplete deglycosylation. The molecular weight of PaCel6A could not be reduced
significantly, neither by N- nor by O-deglycosylation. Lower incorporation of SDS or
inefficient deglycosylation could explain the migration at a higher apparent molecular weight.
The N-terminal sequence was determined for all proteins and confirmed that all of them were
correctly processed, except PaCel6A which, for an unknown reason, lacked the first five
amino acids. The N-terminal sequence of the mature protein started with ERQN. However,
this should not impact the function of the CBM which is situated at the N-terminus, but is
predicted to start only 8 amino acids further.

**Enzymatic properties of P. anserina GH6 enzymes**

PaCel6 enzymatic activities were assayed on cellooligosaccharides and structurally different
polysaccharides. Table 2 summarizes the product profiles on G3 to G6 oligosaccharides. The
observed cleavage pattern for TrCel6A confirmed earlier results with the native enzyme (36).
PaCel6 enzymes had all similar cleavage patterns with the exception that glucose was never
detected after cleavage by PaCel6A, in contrast to the three other enzymes. All enzymes were
active on Avicel and CMC. In addition, PaCel6B displayed activity on barley β-(1,3;1,4)
glucan and glucomannan. No activity was detected on mannan, galactomannan, arabinoxylan,
arabinan, xylan, xyloglucan, HEC and pectin. As all enzymes were active on Avicel, this
substrate was chosen to determine the apparent temperature and pH optima, as well as
stability. PaCel6 enzymes had rather narrow temperature optima, which were generally lower
than TrCel6A. PaCel6A presented the highest activity at 55°C but had a 30% lower activity at
65°C, in contrast to TrCel6A which displayed the same activity at both temperatures (Fig. 3a).
Temperature stability profiles showed that PaCel6A was the most stable of the *P. anserina*
enzymes (up to 45°C, as \(TrCel6A\)), whereas \(PaCel6B\) and \(PaCel6C\) lost nearly all activity after 24h at 45°C and 35°C, respectively (Fig. 3b).

The Cel6 enzymes from \(P. anserina\) showed interesting activity profiles when varying the pH (Fig. 3c). They maintained nearly 100% of their activity from pH 5 up to 9, in contrast to \(TrCel6A\), which lost about 50% of its activity at pH \(\geq 6\). Concerning stability, \(PaCel6A\) was also fairly stable for 24h at these pH values. The activities of \(PaCel6B\) and \(PaCel6C\) declined upon incubation for 24h at pH \(\geq 7\). The following experiments including the determination of kinetic parameters were conducted under optimal conditions for each enzyme.

Michaelis-Menten constants were determined on the soluble substrate cellotetraose. This substrate was cleaved by all four enzymes exclusively into two cellobiose units. HPAEC-PAD analyses revealed only traces of glucose and cellotriose. The kinetic behaviour was very different for each \(P. anserina\) enzyme (Table 3). Kinetic constants for \(PaCel6A\) and \(TrCel6A\) were rather similar. The \(k_{cat}\) value of \(PaCel6A\) was two-fold higher than that of \(TrCel6A\), but it was determined at the optimal temperature for this enzyme (45°C), in contrast to \(TrCel6A\) (27°C). Interestingly, \(PaCel6B\) had a very high turnover number (\(k_{cat} \approx 27.7 \text{ s}^{-1}\)), but the lowest affinity for cellotetraose (\(K_m = 43 \text{ µM}\)). Compared to \(PaCel6B\), the \(k_{cat}\) of \(PaCel6C\) was about 100 fold lower, but the affinity for cellotetraose was much higher. The catalytic efficiencies (\(k_{cat}/K_m\)) of \(PaCel6B\) and \(PaCel6C\) were two-fold lower than those of \(PaCel6A\) and \(TrCel6A\).

**Product profiles analyses**

Because glycoside hydrolase family GH6 groups together both exo- and endoglucanases, we have determined the product profiles of the three \(P. anserina\) GH6 enzymes, using CMC and Avicel as substrates (table 4). The main hydrolysis product from these cellulose substrates was cellobiose (G2) and to a lower extent glucose (G1) and cellotriose (G3). The fact that
TrCel6A slowly cleaves G3 leads to a higher amount of G3 and lower G2/G3 ratios at 15 min than at 24h. G2/G1 decreased slightly or remained the same at 24h, which is in accordance with formation of G1 following the cleavage of G3.

For PaCel6A, G3 increased from 6 to 49 µM between 15 min and 24h, but the G2/G3 ratio increased at the same time. A similar trend was observed for G1, i.e. the glucose amounts and the G2/G1 ratio increased with time. These data suggest that G3 and G1 might be formed mainly in the first minutes of the reaction, by an initial attack liberating G1 or G3 instead of the G2 produced in a processive action ("false initial attack", (37)). With time, the ongoing processive hydrolysis might then lead to proportionally less G3 and G1 formation. Glucose thus seems to originate only from false initial attack, which is confirmed by the fact that PaCel6A, in contrast to TrCel6A, cannot cleave G3.

PaCel6B and PaCel6C produced small amounts of G3 and G1 after 15 min. G2/G3 ratios increased after 24h for both enzymes, whereas G2/G1 ratios remained similar. As for TrCel6A, this is probably due to a slow hydrolysis of G3. The ratio of G2/(G1+G3) or, for enzymes degrading G3, the ratio (G2-G1)/(G3+G1) can be used to estimate processivity (30).

The last column in table 3 shows that processivity was lower for PaCel6B and PaCel6C than for PaCel6A and TrCel6A.

All four tested enzymes were also able to hydrolyse CMC with the highest activity observed for TrCel6A, followed by PaCel6B. TrCel6A released more glucose than G3, which is in contrast to the three P. anserina enzymes. A possible reason could be that TrCel6A slowly cleaves G3 into glucose and G2. Alternatively, more glucose than G3 could be produced by false initial attacks. The second hypothesis would be consistent with the fact that the bulkier side chains of CMC might have more difficulty to enter the tunnel shaped active site.

PaCel6A, PaCel6B and PaCel6C produced more G3 than glucose which is in contrast to the results obtained with Avicel at 24h. CM-cellotriose might not be substrate for these enzymes.
On the other hand, the rather high production of G3 by *PaCel6A* and *PaCel6B* could indicate a higher frequency of false initial attacks due to reduced processivity of these enzymes on CMC.

**Endo- or exoglucanase activity?**

To gain information on the active site topology of *P. anserina* Cel6 enzymes, homology models of all four *PaCel6* enzymes were built based on known template 3D structures (Figure 4). *PaCel6A* and *PaCel6C* display active site tunnels suggesting that they are processive enzymes, in good agreement with the biochemical data that demonstrate highest activity on crystalline cellulose. The tunnel structure could also explain the lower activity on CMC, the substituted glucose units having more difficulty to enter the narrow active site. The processivity of *PaCel6A* on Avicel is similar to that of *TrCel6A*, suggesting that the type of action of the two enzymes could be similar. However, more work is needed to elucidate the reason for the lower catalytic efficiency and specific activity of *PaCel6A*. In contrast to *PaCel6A* and *PaCel6C*, the *PaCel6B* and *PaGH6D* structures are predicted to contain a binding cleft rather than a tunnel structure. These two enzymes indeed lack 15 amino acids of the C-terminal loop (residues 406-420, *TrCel6A* numbering) generating an open topography of the active site. A similar short loop was also found for *H. insolens* Cel6B, an endoglucanase, which represents with 54% and 82% identity the closest characterized enzyme to *PaCel6B* and *PaGH6D*, respectively. The Asn182 of the N-terminal loop which is in direct contact with the Arg410 of the C-terminal loop in exoglucanases (38), is also absent in the latter two enzymes (Fig. 2). The homology model thus supports the hypothesis that these enzymes have endo-type activity. In the case of for *PaCel6B* this was already suggested by the biochemical data obtained: a higher activity on CMC than on Avicel and a lower processivity on Avicel were found. In addition, this enzyme was shown to cleave...
Discussion

Since the 1960s, *P. anserina* has been used as a model for studying fundamental biological processes, such as life cycle and fungal sexual reproduction. Because this coprophilous fungus develops after zygomycetes and (hemi)cellulose-degrading ascomycetes, it is believed to have to cope with the most recalcitrant plant residues. With the current intense research for efficient biomass conversion enzymes, *P. anserina* is beginning to be investigated for its biomass degrading capacities. The publication of its genome indeed revealed a large portfolio of lignocellulose-degrading enzymes (17) which only start to be characterized (27). The present study is, to our knowledge, the first one describing cellulolytic enzymes of this organism.

*P. pastoris* has already been used successfully to express several *P. anserina* polysaccharide-degrading enzymes (27,39) and numerous other CAZymes (40-42). In some cases, the heterologous proteins expressed in *P. pastoris* displayed a higher degree of glycosylation than their native counterparts (43-46), and this was also observed for TrCel6A and PaCel6A in the present study. Deglycosylation evidenced the presence of both *N*- and *O*-glycans in TrCel6A. However, PaCel6A could be only marginally deglycosylated. It is possible that the high number of *O*-glycosylation sites in the linker domain leads to difficultly accessible glycan chains which might have prevented *O*-deglycosylation. The question arises whether enzymatic properties are affected by this high glycosylation degree. Whereas glycosylation of CBM and catalytic domains can alter enzyme properties, such as activity and stability, glycosylation of the linker is generally considered to protect the enzyme from protease degradation (47). Concerning TrCel6A, glycosylation does not seem to alter its functioning:
the specific activity of the recombinant TrCel6A enzyme on Avicel was similar to what was already reported in the literature, even if the experimental conditions were not strictly identical: 0.016 U/mg in this study, compared to 0.027 U/mg in the study of Tomme et al. (48) and Billard et al. (49) which were obtained with the native enzyme purified from T. reesei. The pH and temperature profiles of recombinant TrCel6A were also close to those previously obtained (50). It cannot be ruled out, however, that glycosylation or the presence of tags affects the biochemical properties of one of the heterologously produced P. anserina enzymes which could not be compared to their wild type counterparts.

The pH and temperature profiles for the three PaCel6 enzymes were different from those for the recently described P. anserina hemicellulase enzymes (27). The GH6 enzymes studied here were generally less thermostable than the hemicellulases. In contrast, PaCel6 enzymes were active and stable in a much larger pH range, and especially at alkaline pH, which was not observed for hemicellulases. Other fungal GH6 cellulases with alkaline pH optima have been described: H. insolens CBH2 (Cel6A) and M. oryzae Cel6A show highest activity at pH 9, but H. insolens Cel6B, an endoglucanase, has a much narrower pH optimum centred at pH 7 (26,51).

In contrast to PaCel6B and PaCel6C, PaCel6A harbours a CBM 1 module at its N-terminus. When adsorption was measured on Avicel, about 40% of added enzyme were bound to cellulose at 4°C after 4h, while PaCel6B and PaCel6C did not adsorb to the substrate (not shown), suggesting that PaCel6A-CBM1 is a functional CBM. The missing CBM in the latter two enzymes might also be the reason for the lower specific activities observed on Avicel. However, the CBM does not seem to be important for hydrolysis of CMC (a soluble substrate) as the specific activities are not correlated to its presence.
Modelling of the three-dimensional structures corroborated experimental data suggesting an endo- or exo-type of hydrolytic attack of \textit{PaCell6} enzymes. It has been known for quite some time, however, that the distinction between endo- and exoglucanases is not absolute and sometimes hard to establish experimentally (16,52-54). \textit{H. insolens} Cell6A, a cellobiohydrolase, is acting on crystalline cellulose ribbons in an endo-like fashion and hydrolyses amorphous cellulose and CMC (26). However, its structure was shown to be very similar to that of \textit{TrCell6A} with two surface loops forming a substrate-binding tunnel, characteristic of exoglucanases (38). Due to their ability to hydrolyse substituted glucose chains, such as CMC, both enzymes are therefore considered to be endo-processive cellobiohydrolases (24,55). \textit{PaCell6C} has a lower processivity on crystalline cellulose compared to \textit{TrCell6A} and hydrolyses CMC nearly as well as crystalline cellulose. This suggests that this enzyme is better characterized as a processive glucanase with an even more pronounced endo-character than the Cell6A enzymes of \textit{T. reesei} or \textit{H. insolens}.

Other examples of processive endoglucanases are known: \textit{Agaricus bisporus} CEL3 possesses loops enclosing the active site tunnel, but its activity profile was shown to be intermediate between typical cellobiohydrolases and endoglucanases (56). In prokaryotes, processive endoglucanases have been identified in \textit{Saccharophagus degradans} (57) and a GH9 cellulase with both endo- and exoglucanase activities was described in \textit{Cellulomonas fimi} (58).

The degree of endo-type activity can be related to different structural features, such as the structure of the catalytic core which determines its binding strength (53). Binding of the substrate to the active site is known to involve several tryptophan residues (Trp135, Trp 272, Trp 367, Trp269) (34). The most important one, Trp135, binds to the glucose unit at the -2 subsite and its binding strength was demonstrated to drive the movement of the cellulose chain leading to processive cleavage (11). All of these residues are conserved in the \textit{P.}
\textit{P. anserina} Cel6s, but different electrostatic environments at this subsite could lead to different degrees of processivity.

Another factor that may affect processivity is the active site loops. These are thought to display some flexibility and undergo conformational changes in response to ligand binding, which could lead to occasional endo cleavage even if the enzyme presents a tunnel-like structure (38,59,60). Cleavage of fluorescein-labelled cellobextrins by CBHs supports the hypothesis that such conformational changes occur (61). Therefore, the degree of flexibility and length of the active site loops may also explain the different extent of endo-action observed for the \textit{Pa}Cel6 enzymes.

In a recent study, \textit{P. anserina} secretomes that had been generated by induction with different substrates were tested for their capacity to supplement a \textit{T. reesei} enzyme cocktail. By replacing half of the enzyme content with an equivalent protein amount of \textit{P. anserina} enzymes the hydrolysis yield of steam-exploded wheat straw could be increased up to 17\% upon hydrolysis at 37°C. Proteomic analysis of the two secretomes that resulted in the highest gain of hydrolysis yield, namely those obtained after induction by sugar beet pulp and Avicel, revealed the presence of the four \textit{Pa}Cel6 enzymes: \textit{Pa}Cel6A, \textit{Pa}Cel6B were induced by both substrates, whereas \textit{Pa}Cel6C and \textit{Pa}GH6D were present in only one of the two secretomes (L. Poidevin, manuscript in preparation). It cannot be confirmed at this point that the \textit{Pa}Cel6 enzymes are responsible for the improvement in hydrolysis yield observed. But this finding indicates that even though a \textit{P. anserina} secretome hydrolyses micronized or steam pretreated wheat straw less efficiently than a \textit{T. reesei} enzyme cocktail (L. Poidevin, unpublished results), \textit{P. anserina} produces enzymes which can effectively increase hydrolysis yields when added to a classical \textit{Trichoderma} cellulase cocktail. We are only at the beginning of the characterization of the enzymatic complex of \textit{P. anserina}, which includes about 180 genes encoding plant cell wall degrading enzymes (17). The presence of 15 GH5, 6 GH7, encoding
putative cellobiohydrolases, and 33 GH61 is particularly intriguing. Transcriptomic and proteomic studies are now underway to determine which enzymes are produced in the presence of lignocelluloses and gain a deeper mechanistic understanding of the cell wall degrading arsenal of this fungus.

Acknowledgements

We thank M. Haon for her assistance on the expression of targets in P. pastoris. This study was funded by the French National Research Agency (ANR, program E-TRICEL ANR-07-BIOE-006).
Figure Legends

Figure 1
Amino acid sequence alignment of catalytic domains of T. reesei Cel6A and the four P. anserina family 6 glycoside hydrolases. Conserved residues are marked by an asterisk. AA involved in catalysis are shaded. AA implicated in glucose binding are boxed. Residues N182 and R410 which establish a direct contact between the two surface loops in TrCel6A are indicated by an arrow. The sequence stretch corresponding to the second surface loop present in TrCel6A, PaCel6A and PaCel6C, but absent in PaCel6B and PaGH6D, is boxed.

Figure 2

Figure 3
Activity of P. anserina Cel6 and T. reesei Cel6A on Avicel as a function of temperature and pH. (A) Optimal temperature: 1 % Avicel was hydrolysed for 15 min at the indicated temperature. (B) Temperature stability. Activity was measured after incubation at the indicated temperature for 24h. Hydrolysis reactions were conducted at 35°C. (C) Optimal pH. Hydrolysis reactions were conducted at 35°C for 15 min. (D) pH stability. Activity was...
measured at 35°C after incubation at indicated pH for 24h. Symbols represent: filled diamonds: TrCel6A, filled squares: PaCel6A, open triangles: PaCel6B, open circles: PaCel6C. Values are means of triplicate measurements.

**Figure 4**

Structural analysis of PaCel6. **Panel A** shows surface representations of PaCel6A (left; green), PaCel6B (middle left; dark blue), PaCel6C (middle right; light green) and PaGH6D (right, light blue). PaCel6A and PaCel6C (green) exhibit a tunnel formation whereas PaCel6B and PaGH6D show a cleft shaped active site. **Panel B** represents a structural alignment of the catalytic sites of PaCel6A (green) with PaCel6B (blue). The loop corresponding to amino acids 415-429 in PaCel6A leading to the formation of the tunnel shape is absent in PaCel6B. Models of PaGH6 were built using PHYRE2 and visualized with PyMOL. PaCel6A and PaCel6C models exhibited 100% confidence with *H. insolens* Cel6A (PDB ID: 1OC7 or 1BVW) and TrCel6A ((PDB ID: 1QJW), PaCel6B and PaGH6D models showed 100% confidence to *H. insolens* Cel6B (PDB ID: 1DYS).
References


Cel6A from *Trichoderma reesei*: The roles of aspartic acids D221 and D175. J. Am. Chem. Soc. **124**:10015-10024.


Table 1: Molecular characteristics of the four *P. anserina* GH6 and identity with known enzymes. Amino acid ranges of functional domains of mature proteins were predicted by InterproScan.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>MW (pI)</th>
<th>CBM</th>
<th>Linker</th>
<th>Catalytic domain</th>
<th>No of predicted glycosylation sites</th>
<th>% identity to <em>TrCel6A</em> (CBH2)*</th>
<th>Database entry with highest similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PaCel6A</td>
<td>6187194</td>
<td>49.3 (5.8)</td>
<td>9-45</td>
<td>46-102</td>
<td>103-467</td>
<td>1-29</td>
<td>67.8 Chaetomium globosum hsp. protein EAO82944 (77%)</td>
</tr>
<tr>
<td>PaCel6B</td>
<td>6187311</td>
<td>41.3 (6.2)</td>
<td>-</td>
<td>-</td>
<td>1-364</td>
<td>3-2</td>
<td>35.4 Myceliophthora thermophila GH6 AEO57190 (73%)</td>
</tr>
<tr>
<td>PaCel6C</td>
<td>6188027</td>
<td>40.9 (5.1)</td>
<td>-</td>
<td>-</td>
<td>1-380</td>
<td>1-3</td>
<td>45.8 Sordaria macrospora CBI58887 (79%)</td>
</tr>
<tr>
<td>PaGH6D</td>
<td>6187939</td>
<td>44.0 (7.5)</td>
<td>377-412</td>
<td>349-376</td>
<td>1-348</td>
<td>1-9</td>
<td>35.0 Humicola insolens endoglucanase 6B Q7S935 (82%)</td>
</tr>
</tbody>
</table>

*Identity calculated with catalytic domains only
Table 2: Product profile of recombinant *PaCel6* enzymes and *TrCel6A* on soluble oligosaccharides.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>TrCel6A</em></td>
<td>G3</td>
</tr>
<tr>
<td></td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td>G4</td>
<td>G2</td>
</tr>
<tr>
<td>G5</td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td>G6</td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td><em>PaCel6A</em></td>
<td>G3</td>
</tr>
<tr>
<td></td>
<td>G3</td>
</tr>
<tr>
<td></td>
<td>G2</td>
</tr>
<tr>
<td></td>
<td>G2, G3</td>
</tr>
<tr>
<td>G6</td>
<td>G2, G3</td>
</tr>
<tr>
<td><em>PaCel6B</em></td>
<td>G3</td>
</tr>
<tr>
<td></td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td>G4</td>
<td>G2</td>
</tr>
<tr>
<td>G5</td>
<td>G2, G3</td>
</tr>
<tr>
<td>G6</td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td><em>PaCel6C</em></td>
<td>G3</td>
</tr>
<tr>
<td></td>
<td>G1, G2, G3</td>
</tr>
<tr>
<td>G4</td>
<td>G2</td>
</tr>
<tr>
<td>G5</td>
<td>G2, G3</td>
</tr>
<tr>
<td>G6</td>
<td>G2, G3</td>
</tr>
</tbody>
</table>
Table 3: Specific activities of recombinant enzymes on cellulose substrates (nmol min\(^{-1}\) mg\(^{-1}\)) and kinetic constants for PaCel6A, 6B and 6C, compared to the constants for TrCel6A, measured on cellohexaose in optimal conditions for each enzyme. CMC: carboxymethylcellulose

<table>
<thead>
<tr>
<th></th>
<th>Avicel</th>
<th>CMC</th>
<th>(K_m) (µM)</th>
<th>(k_{cat}) (s(^{-1}))</th>
<th>(k_{cat}/K_m) (µM(^{-1}) s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TrCel6A*</td>
<td>16.4</td>
<td>6.7</td>
<td>2.6</td>
<td>3.1</td>
<td>1.2</td>
</tr>
<tr>
<td>PaCel6A</td>
<td>7.6</td>
<td>3.0</td>
<td>4.7</td>
<td>6.6</td>
<td>1.4</td>
</tr>
<tr>
<td>PaCel6B</td>
<td>2.3</td>
<td>5.5</td>
<td>43.0</td>
<td>27.7</td>
<td>0.65</td>
</tr>
<tr>
<td>PaCel6C</td>
<td>1.9</td>
<td>1.1</td>
<td>0.59</td>
<td>0.3</td>
<td>0.54</td>
</tr>
</tbody>
</table>

*Kinetic constants from (36) measured at 27°C and pH5
Table 4: Hydrolysis products from 1% Avicel or CMC with *P. anserina* GH6 enzymes or *TrCel6A*, measured after 15 min (Avicel) and 24h (Avicel and CMC) hydrolysis under optimal conditions. Values are µM of respective sugars, determined by HPLC.

<table>
<thead>
<tr>
<th></th>
<th>Glucose</th>
<th>Celllobiose</th>
<th>Cellotriose</th>
<th>Total</th>
<th>Processivity*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Avicel 15 min</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>TrCel6A</em></td>
<td>9.2</td>
<td>112</td>
<td>5.7</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td><em>PaCel6A</em></td>
<td>7.5</td>
<td>57</td>
<td>5.9</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td><em>PaCel6B</em></td>
<td>3.3</td>
<td>32</td>
<td>4.5</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td><em>PaCel6C</em></td>
<td>3.6</td>
<td>23</td>
<td>2.7</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td><strong>Avicel 24 h</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>TrCel6A</em></td>
<td>246</td>
<td>2119</td>
<td>0</td>
<td>2365</td>
<td>7.6</td>
</tr>
<tr>
<td><em>PaCel6A</em></td>
<td>75</td>
<td>977</td>
<td>49</td>
<td>1101</td>
<td>7.9</td>
</tr>
<tr>
<td><em>PaCel6B</em></td>
<td>43</td>
<td>283</td>
<td>8.3</td>
<td>335</td>
<td>4.7</td>
</tr>
<tr>
<td><em>PaCel6C</em></td>
<td>32</td>
<td>238</td>
<td>4.6</td>
<td>275</td>
<td>5.7</td>
</tr>
<tr>
<td><strong>CMC 24 h</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>TrCel6A</em></td>
<td>129</td>
<td>809</td>
<td>24</td>
<td>961</td>
<td>5.3</td>
</tr>
<tr>
<td><em>PaCel6A</em></td>
<td>12</td>
<td>355</td>
<td>69</td>
<td>436</td>
<td>4.4</td>
</tr>
<tr>
<td><em>PaCel6B</em></td>
<td>19</td>
<td>650</td>
<td>123</td>
<td>793</td>
<td>4.6</td>
</tr>
<tr>
<td><em>PaCel6C</em></td>
<td>4.4</td>
<td>142</td>
<td>16</td>
<td>162</td>
<td>7.1</td>
</tr>
</tbody>
</table>

*For Avicel: calculated as molar ratio (G2-G1)/(G3+G1), except for *PaCel6A* where it is G2/(G1+G3). For CMC: calculated as G2/(G3+G1).*