the FEBS Journal

# Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity

Larissa C. Textor[1,†], Francieli Colussi[1,†], Rodrigo L. Silveira[2,†], Viviane Serpa[1], Bruno L. de Mello[1], João Renato C. Muniz[1], Fabio M. Squina[3], Nei Pereira Jr[4], Munir S. Skaf [2] and Igor Polikarpov[1]

1 Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil
2 Institute of Chemistry, State University of Campinas – UNICAMP, Campinas, SP, Brazil
3 Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, SP, Brazil
4 Centro de Tecnologia, Escola de Química, Laboratório de Desenvolvimento de Bioprocessos (LaDeBio), Universidade Federal do Rio de Janeiro, RJ, Brazil

**Keywords**
biomass; cellobiohydrolase; enzymatic hydrolysis; molecular dynamics; X-ray structure

**Correspondence**
I. Polikarpov, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador Saocarlense 400, São Carlos, SP, 13560-970, Brazil
Fax: +55 16 3373 9881
Tel: +55 16 3373 8088
E-mail: ipolikarpov@ifsc.usp.br

†Equally contributing authors.

(Received 21 August 2012, revised 18 October 2012, accepted 19 October 2012)

doi:10.1111/febs.12049

Aiming to contribute toward the characterization of new, biotechnologically relevant cellulolytic enzymes, we report here the first crystal structure of the catalytic core domain of Cel7A (cellobiohydrolase I) from the filamentous fungus *Trichoderma harzianum* IOC 3844. Our structural studies and molecular dynamics simulations show that the flexibility of Tyr260, in comparison with Tyr247 from the homologous *Trichoderma reesei* Cel7A, is enhanced as a result of the short side-chains of adjacent Val216 and Ala384 residues and creates an additional gap at the side face of the catalytic tunnel. *T. harzianum* cellobiohydrolase I also has a shortened loop at the entrance of the cellulose-binding tunnel, which has been described to interact with the substrate in *T. reesei* Cel7A. These structural features might explain why *T. harzianum* Cel7A displays higher $k_{cat}$ and $K_m$ values, and lower product inhibition on both glucoside and lactoside substrates, compared with *T. reesei* Cel7A.

## Introduction

It is now generally accepted that environmental changes have become a serious problem worldwide and might drastically affect the Earth's climate, with severe consequences for humankind if not addressed properly. Several studies have associated global warming with increasing levels of atmospheric $CO_2$, generated, among other sources, by the burning of fossil fuels [1–3]. Hence, extensive scientific efforts have been invested toward utilization and recovery of natural renewable resources for bioenergy production [4].

**Abbreviations**
BMCC, bacterial microcrystalline cellulose; CBM, cellulose binding module; CCD, catalytic core domain; CMC, carboxymethyl cellulose; cNPL, 2-chloro-4-nitrophenyl-β-lactoside; DP, degree of polymerization; GH, glycosyl hydrolase; Ma_Cel7B, *Melanocarpus albomyces* Cel7B; MD, molecular dynamics; Pc_Cel7D, GH family 7 cellobiohydrolase from *Phanerochaete chrysosporium*; PDB, Protein Data Bank; pNPC, *p*-nitrophenyl-D-cellobioside; Te_Cel7A, GH family 7 cellobiohydrolase from *Talaromyces emersonii*; Th_Cel7A, *Trichoderma harzianum* cellobiohydrolase I; Tr_Cel7A, catalytic core domain of Cel7A (CBHI) from *Trichoderma reesei*.

 

Environment-friendly alternatives, such as wind, solar power and biomass, represent important sources of renewable energy. As lignocellulosic material comprises most of the biomass on Earth, biodegradation of cellulose and hemicellulose has become one focus of research on renewable liquid biofuels. Hydrolysis of cellulose represents a considerable challenge because the substrate is insoluble and recalcitrant against depolymerization and enzymatic action.

The biotechnological use of filamentous fungi is a growing field with promising practical applications in bioenergy and biotechnology because these microorganisms are known to express several hemicellulose- and cellulose-degrading enzymes at high yields [4]. A wide range of industrial technologies use cellulases, mainly in starch processing, grain alcohol fermentation, malting and brewing, extraction of fruit and vegetable juices, pulp, paper and textile [5,6]. Future promising applications for cellulases include the production of fuel ethanol from the abundant lignocellulosic biomass through the enzymatic breakdown of cellulose [7].

The well-studied filamentous fungus *Trichoderma reesei* expresses a number of cellulases that can be broadly divided into three types: endoglucanases (endo-1,4-β-D-glucanases); cellobiohydrolases or exoglucanases (exo-1,4-β-D-glucanases); and β-glucosidases (1,4-β-D-glucosidases) [5,8]. The classical view of cellulose hydrolysis is described as a simultaneous and synergistic action of these three types of enzymes. The hydrolysis process starts with endoglucanases that randomly hydrolyze internal β-1,4-glycosidic linkages in the amorphous region of cellulose microfibrils, reducing significantly the degree of substrate polymerization and creating new terminal ends. Simultaneously, the accessible reducing and nonreducing ends of cellulose chains are attacked by types I and II cellobiohydrolases, respectively. The cellobiose units released are then hydrolyzed into glucose units by β-glucosidases [8].

Cellulases can be classified into glycosyl hydrolase (GH) families based on their amino-acid sequence and folding similarities [9]. Together with some endoglucanases and chitosanases, type I cellobiohydrolases belong to the GH family 7 [10]. The exoglucanase, Cel7A, is the most important single enzyme component for cellulose depolymerization and conversion into cellobiose in fungi [11]. It is considered to be the key enzyme in the hydrolysis process because it is able to hydrolyze crystalline cellulose extensively, albeit at a slow rate [12]. Moreover, Cel7A is the predominant enzyme secreted by the fungus *Trichoderma* sp. when grown under conditions suitable for the production of cellulases, reaching up to 60% of the total secreted protein [13].

In general, the cellulases produced by *T. reesei* are glycoproteins that frequently contain both O- and N-linked glycosylation sites [14]. Most cellulases display a similar structural organization, consisting of two distinct functional and structural domains (Fig. 1A): a cellulose-binding domain and an N-glycosylated catalytic core domain (CCD) [15]. The domains are joined together by a long and highly O-glycosylated linker region. Structural studies have shown that N-glycosylation of Cel7A from *T. reesei* strains QM9414 and ALKO2877 (derived from QM9414) is present as single *N*-acetylglucosamine modification on three (Asn45, Asn270 and Asn384) out of four potential sites located in the protein catalytic domain [16]. The occurrence of such modifications depends on several factors, including strain and

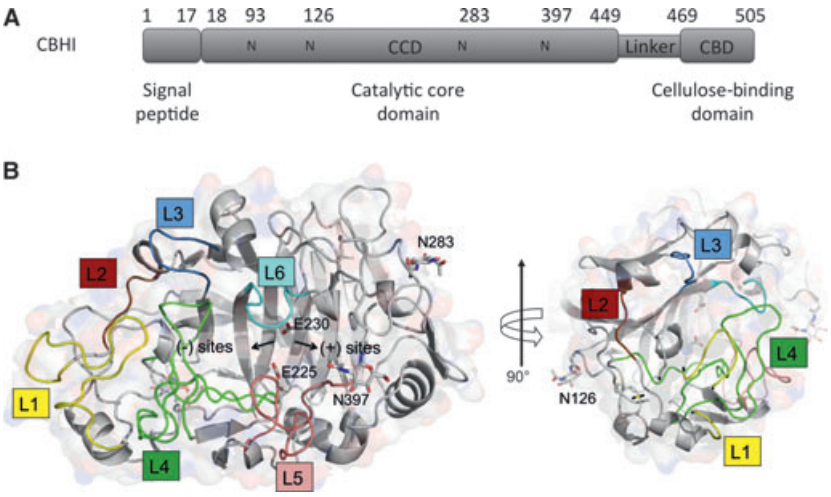


**Fig. 1.** Structure of cellobiohydrolase I from *Trichoderma harzianum*. (A) Structural organization. Predicted N-glycosylation sites are identified (N) and the O-glycosylated linker is shown. (B) Ribbon diagram of the crystallographic structure of the CCD from Th_Cel7A highlighting the loops of the tunnel. Observed glycosylation sites at N126, N283 and N397 are displayed.

culture conditions, and generates several enzyme isoforms with similar catalytic and adsorption properties [17,18].

X-ray structures of four GH family 7 cellobiohydrolases have been published to date: two correspond to the CCDs of *Trichoderma reesi* Cel7A [Protein Data Bank (PDB) entry 1CEL) [19] and of *Phanerochaete chrysosporium* Cel7D (PDB entry 1GPI) [20], whereas two other structures represent the single-domain thermostable filamentous fungi cellobiohydrolases Cel7A from *Talaromyces emersonii* (PDB id code 1Q9H) [21] and Cel7B from *Melanocarpus albomyces* (PDB id code 2RFW) [22]. In addition, the crystallographic structure of cellobiohydrolase Cel7A from *Heterobasidion annosum* is available at the PDB under the code 2XSP, but has not yet been published. Furthermore, solution structures of several cellulose binding modules (CBMs) have been determined, including that of *T. reesei* Cel7A (PDB entry code 1CBH [23]).

Crystallographic structures of the full-length proteins have not yet been determined, presumably because of the presence of the highly mobile linker region, which hampers crystallization. At the same time, CCD structures of *T. reesei* cellobiohydrolase I (Tr_Cel7A) have been determined in complex with a number of oligosaccharides and ligands [24]. The CCD fold consists of two large antiparallel β-sheets, which form a β-sandwich tertiary structure. An elongated cellulose-binding tunnel surrounded by disulfide-bridged loops is formed along the concave side of the β-sandwich. The tunnel encloses the active site at its far end where the catalytic residues aspartate and both glutamates are located [19]. The presence of the elongated tunnel mediates the enzyme specificity that is achieved by a substantial number of interactions between the cellulose chain ends and the residues of the tunnel [25]. The polysaccharide chain slides through the substrate-binding tunnel and every second glycosidic bond is correctly presented to the catalytic apparatus, which explains why these enzymes progressively liberate disaccharide units (cellobiose) from the reducing end of a cellulose chain [19]. After enzymatic cleavage, the cellobiose is found bound to sites +1 and +2 near the exit of the tunnel [19]. The catalytic triad of carboxylate groups is strictly conserved among the enzymes of GH family 7 [26].

*Trichoderma harzianum* is a filamentous fungus that exhibits both cellulolytic and mycoparasitic activities, and, unlike *T. reesei*, its cellulolytic activity and biotechnological potential are still largely unexplored. Enzymatic studies of several strains of *T. harzianum* have shown that these fungi are able to produce cellulolytic complexes with higher β-glucosidase activity compared with those produced by *T. reesei*. Moreover, *T. harzianum* secretes an effective and well-balanced enzymatic system that is able to completely hydrolyze cellulosic substrates into monomeric glucose [27]. *T. harzianum* has a higher efficiency of hydrolysis compared with *T. reesei*, owing to the higher β-glucosidase and xylanase activities; consequently this represents a potential advantage for *T. harzianum* in terms of more efficient biomass deconstruction.

In order to investigate, in greater detail, the enzymes present in the cellulolytic complex secreted by *T. harzianum*, we studied the enzymatic kinetics and inhibition by cellobiose of Cel7A from *T. harzianum* and compared these data with enzymatic parameters of the homolog *T. reesei* Cel7A (81% amino acid sequence identity with *T. harzianum* Cel7A). We also determined the crystallographic structure of *T. harzianum* cellobiohydrolase I (Th_Cel7A), and performed molecular dynamics (MD) computer simulations on these structures to obtain more detailed information on the structural features of the enzymes and their dynamics. Our analyses provide insights into the molecular basis of the observed differences in the enzymatic activity and kinetics between Tr_Cel7A and Th_Cel7A.

## Results

### Kinetic constants and enzyme specificity for different soluble and insoluble substrates

Assuming Michaelis–Menten kinetics of the full-length *T. harzianum* Cel7A, the experimentally determined $K_m$ and $k_{cat}$ values were, respectively, 3.4 mM and 23.2 $min^{-1}$ for *p*-nitrophenyl-D-cellobioside (pNPC) and 3.7 mM and 44.6 $min^{-1}$ for 2-chloro-4-nitrophenyl-β-lactoside (cNPL) (Table 1). In comparison with the reported kinetic parameters of Tr_Cel7A, which ranged from 0.4 to 1.23 mM [4,26,28–30], our results showed that Th_Cel7A exhibits significantly lower affinity for these substrates, whereas the turnover value, $k_{cat}$, was approximately 3.5 times larger than that of Tr_Cel7A on the lactoside substrate ($\sim$12 $min^{-1}$) [26,28]. Moreover, our inhibition assays showed that Th_Cel7A was competitively inhibited by cellobiose, although with a much higher inhibition constant of 7.2 ± 1.2 mM compared with the $K_i$ of 0.02 mM for Tr_Cel7A [31]. Low inhibition by the end-product and by the competitive inhibitor, cellobiose, might be potentially important for biotechnological applications of the enzyme.

Specific activity and substrate-hydrolysis assays of the full-length Th_Cel7A and its CCD are shown in Fig. 2. Our experiments revealed that Th_Cel7A is

**Table 1.** Kinetic parameters of Th_Cel7A. Activity measurements were performed using 700 nM enzyme in 50 mM sodium acetate buffer, pH 5.0, at 50 °C

| | pNPC | | | cNPL | | | Cellobiose |
|---|---|---|---|---|---|---|---|
| | $K_m$ (mM) | $V_{max}$ (mM·min$^{-1}$) | $k_{cat}$ (min$^{-1}$) | $K_m$ (mM) | $V_{max}$ (mM·min$^{-1}$) | $K_{cat}$ (min$^{-1}$) | $K_i$ (mM) |
| Th_Cel7A | 3.4 ± 0.5 | 0.016 ± 0.001 | 23.2 ± 1.1 | 3.7 ± 1.2 | 0.031 ± 0.004 | 44.6 ± 4.7 | 7.2 |

active and very efficient in degrading bacterial microcrystalline cellulose (BMCC), achieving almost 90% conversion of the latter substrate in 24 h under the conditions of our experiment. As expected, the CCD is significantly less efficient in the hydrolysis of crystalline cellulose (Avicel, Sigmacell 20 and BMCC) compared with the full-length enzyme containing both CCD and CBM domains (Fig. 2A,B). Furthermore, Th_Cel7A is capable of the enzymatic cleavage of amorphous carboxymethyl cellulose (CMC); however, in this case, the enzymatic efficiency was low and there was no significant difference in specific activity between the CCD and the full-length enzyme (Fig. 2A, B). Furthermore, neither full-length enzyme nor the CCD was able to hydrolyze more than 10% of this substrate after 24 h of reaction. Conversion of CMC rapidly reached a plateau and the reaction did not proceed any further (Fig. 2C,D). Both specific activities

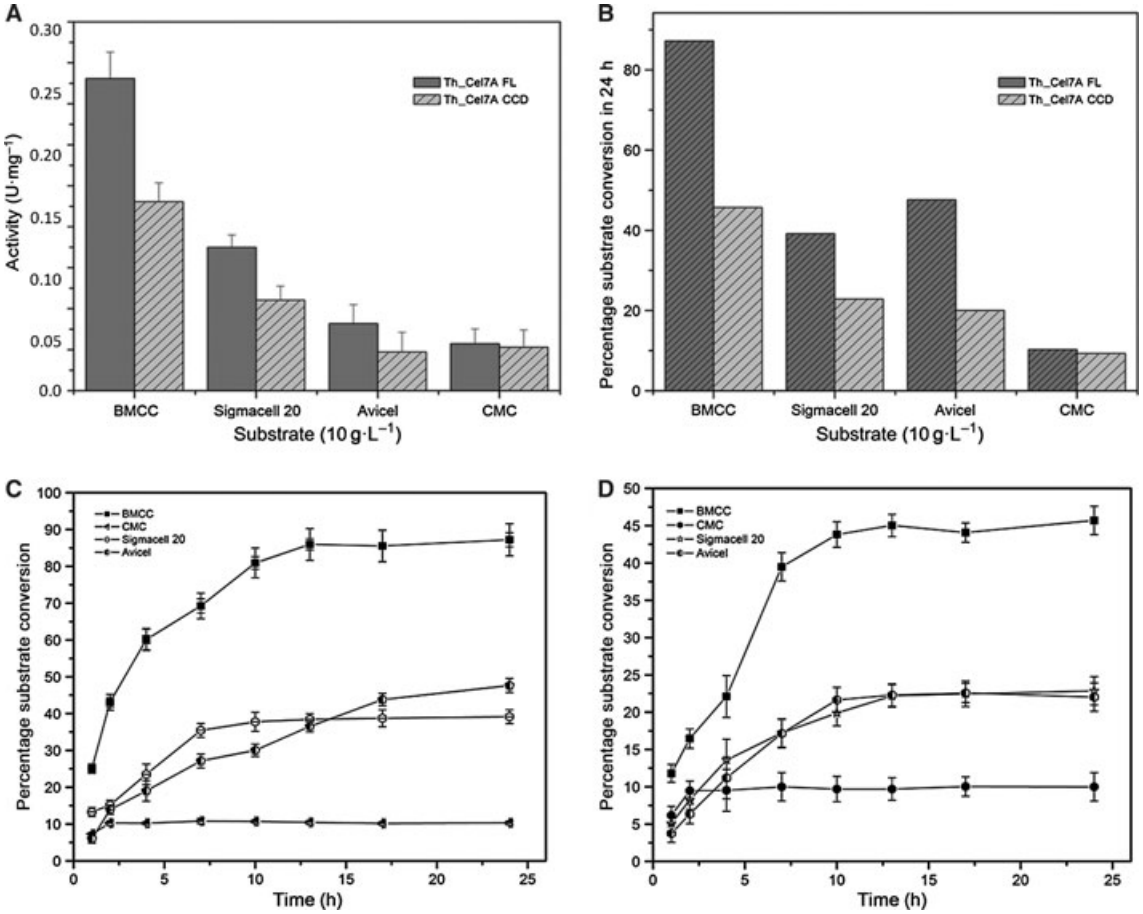


**Fig. 2.** Activities of purified cellobiohydrolase I from *Trichoderma harzianum* (Th_Cel7A FL) and its CCD (Th_Cel7A CCD) on different cellulosic substrates. Assays were performed in 50 mM sodium citrate buffer, pH 5.0, at 50 °C. (A) Specific activities of Th_Cel7A and (B) percentage substrate conversion in 24 h. (C) Time course of substrate conversion by Th_Cel7A and (D) by its CCD. [Note the change in scale between (C) and (D)].

and hydrolytic yields of Th_Cel7A on Avicel and Sigmacell 20 were at intermediate levels (between those observed for BMCC and CMC) and there was a considerable difference in activity and hydrolysis yields between complete enzyme and its CCD (Fig. 2). The measured specific activities of Th_Cel7A were comparable with that of Tr_Cel7A (Fig. 2A). According to the reported data, the specific activities of Tr_Cel7A against Avicel or Sigmacell 20 were 0.014–0.3 $U \cdot mg^{-1}$ and 0.04–0.07 $U \cdot mg^{-1}$, respectively, for CMC under similar experimental conditions [15,32]. These results show that the exoglucanase, Cel7A, from *T. harzianum* is efficient in hydrolyzing crystalline cellulosic substrates, such as BMCC, Avicel and Sigmacell 20, but is not effective in the hydrolysis of substituted substrates, such as CMC.

## Overall structure of *T. harzianum* Cel7A CCD

The refined Cel7A CCD model (Table 2), comprises residues 18–445 and assumes the characteristic fold of Cel7A enzymes [19–22,24,26], with the β-strands of the β-sandwich structure connected by the loops (L1 to L6) that enclose the long cellulose-binding tunnel (Fig. 1B).

The CCD contains a total of 20 cysteine residues, all of which are involved in the disulfide bridges known to stabilize the loops around the cellulose-binding cleft. As commonly found in enzymes secreted by fungi [19–22,24,26], the N-terminal glutamine is cyclized into a pyroglutamate residue. Three of the four glycosylation sites predicted according to the sequence motif N-X-S/T (where X is any amino acid residue, except proline) could be identified in the electron density maps, and single *N*-acetyl D-glucosamine residues were modeled covalently bound to Asn126, Asn283 and Asn397. All *N*-acetyl D-glucosamine residues make contact with symmetry-related protein molecules.

**Table 2.** X-ray refinement statistics

| | |
|---|---|
| Diffraction data | |
| Space group | $P2_12_12_1$ |
| Unit cell parameters (Å) | $a = 56.7$, $b = 69.2$, $c = 90.6$ |
| Resolution limits (Å)[a] | 45.3–1.67 (1.71–1.67) |
| Number of unique reflections | 42070 (4077) |
| Redundancy | 6.1 (5.5) |
| Completeness (%) | 99.8 (98.8) |
| $R_{merge}$ (%)[b] | 9.1 (32.9) |
| I/sigI (I)[b] | 24.4 (4.7) |
| Refinement | |
| $R_{factor}/R_{free}$ (%) | 13.1/16.2 |
| Number of atoms | |
| Protein | 3281 |
| *N*-acetyl-D-glucosamine | 42 |
| Poly(ethylene glycol) | 24 |
| Water | 694 |
| B factors (Å$^2$) | |
| Protein | 10.5 |
| *N*-acetyl-D-glucosamine | 28.3 |
| PEG | 33.0 |
| Water | 27.2 |
| rmsd | |
| Bond length (Å) | 0.011 |
| Bond angle (°) | 1. 494 |
| Ramachandran (preferred/allowed)% | 98.1/1.9 |

[a] Values in parentheses are for the highest resolution shell. [b] $R_{merge} = \sum|I_j - \langle I\rangle|/\sum \langle I\rangle$, where $I_j$ is the observed intensity of an individual reflection and $\langle I\rangle$ is the average intensity of that reflection.

## Overall comparison of the *T. harzianum* Cel7A CCD with other Cel7A CCD structures

Consistent with the high sequence homology between the Cel7A from *T. harzianum* and the Cel7A from *T. reesei*, structural superposition showed that both molecules have a very similar global fold, with an rmsd of 0.5 Å for 426 Cα aligned residue pairs. As expected, the largest structural deviations were associated with the loop regions that display high temperature factors and therefore are mobile or structurally disordered. The first important structural difference between Th_Cel7A and Tr_Cel7A was associated with loop 3 at the entrance of the tunnel (residues Ser99 to Gln101 in the Tr_Cel7A sequence). Loop 3 is critical in defining the overall opening profile of the ligand-binding site. Some residues present in this loop in Tr_Cel7A are absent in Th_Cel7A and consequently the enzyme has a more open entrance to the binding tunnel (Figs 3 and 4). Gln101 from Tr_Cel7A loop 3 is a key residue in the substrate recognition at subsites −6 and −5 [24] and such interactions are thus not available in Th_Cel7A.

A similar shortening of loop 3 was also described for GH family 7 cellobiohydrolases from *T. emersonii* (Te_Cel7A) and *P. chrysosporium* (Pc_Cel7D). The CCDs of Te_Cel7A and Pc_Cel7D display a sequence similarity of 70% and 60%, respectively, with the CCD of Th_Cel7A, and a structural rmsd of 0.8 Å for 417 Cα Te_Cel7A aligned residues and 0.9 Å for 415 Cα Pc_Cel7D aligned residues, respectively. The equivalent loop 3 in both structures was similar to that of Th_Cel7A and consequently they also exhibit a more open tunnel entrance when compared with Tr_Cel7A [20,24].

A second important feature of the Th_Cel7A CCD structure is related to the loops that form the sidewalls

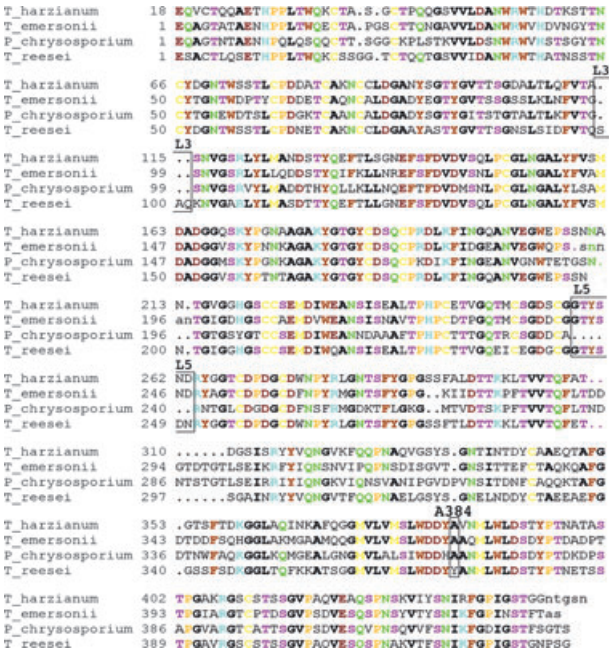


**Fig. 3.** Structure-based alignment of cellobiohydrolase I from *Trichoderma harzianum* (Uniprot Q9P8P3), *Talaromyces emersonii* (Uniprot Q8TFL9, PDB 1Q9H), *Phanerochaete chrysosporium* (Uniprot Q7LIJ0, PDB 1GPI) and *Trichoderma reesei* (Uniprot P07981.1, PDB 1CEL). The alignment was carried out using Strap (http://3d-alignment.eu/). Numbering of protein sequences is according to the deposited PDB structures. The position of Ala384 and gaps in loops 3 and 5 are depicted in rectangular boxes.

of the ligand-binding tunnel (loops 4–6), in which the conformation and sequence composition of the loops are crucial for determining the shape of the channel side-faces. Loops 5 and 6 of Th_Cel7A enclose part of the side tunnel surface and create a gap that gives access to the binding tunnel at the substrate site (Fig. 5A). Whereas the equivalent loop 5 in Pc_Cel7D was significantly shorter, resulting in a totally exposed tunnel (Fig. 5B), the side face of the cellulose channel from Tr_Cel7A was almost completely covered (Fig. 5C). The entrance to the catalytic tunnel of Th_Cel7A was also more exposed compared with that of Th_Cel7A (although not as much as in Pc_Cel7D) because Th_Cel7A residues Ile203 and Tyr371 are substituted by shorter side-chains of Val216 and Ala384 in Th_Cel7A (Fig. 5A). As discussed later, MD simulations demonstrate that the presence of Ala384 in loop 6 of Th_Cel7A in place of Tyr371 in Tr_Cel7A is responsible for notable effects on the dynamics of the loops enclosing the substrate-binding site. Our crystal structure and MD simulations revealed that Tyr260 can adopt different conformations in the Th_Cel7A CCD. Conformations in which the Tyr260 side chain is more exposed to the solvent can partially obstruct the enhanced access to the cellulose-binding channel in Th_Cel7A. Regarding the catalytic residues at the cellulose-binding cleft, the only observed conformational difference was associated with Glu217, which is H-bonded to the glucosyl unit at the product +1 site in Tr_Cel7A, whereas the equivalent Glu230 of Th_Cel7A assumed an alternative conformation, forming an H-bond to Asp154.

## MD simulations of *T. harzianum* and *T. reesei* Cel7A CCDs

To gain further insight into the mobilities of Th_Cel7A and Tr_Cel7A CCDs, we performed MD simulations of these enzymes starting from their *apo* structures. According to the results of our simulations,

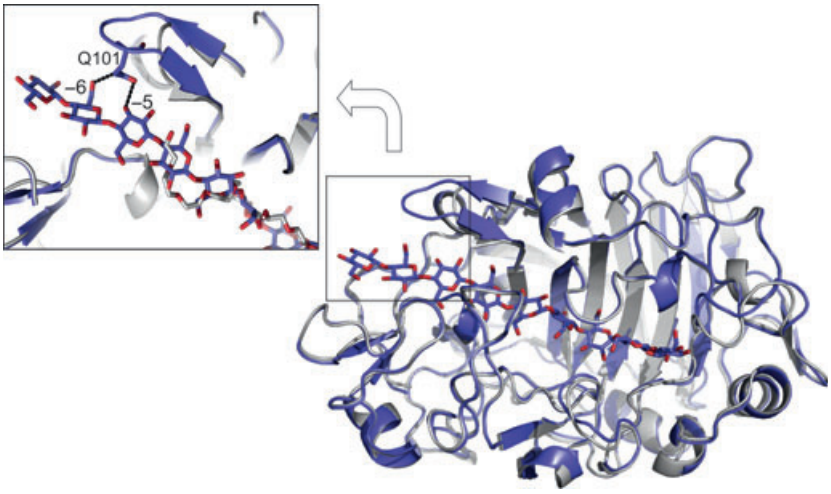


**Fig. 4.** Structural superposition of Cel7A from *Trichoderma harzianum* (gray) and Cel7A from *Trichoderma reesei* (blue, PDB code 8CEL). The larger boxed area represents a closer view of the loop absent in *T. harzianum*, highlighting Gln101 and the glucosyl residues from Tr_Cel7A (blue) at the entrance of the cellulose-binding tunnel. H-bonds are shown as dashed lines.

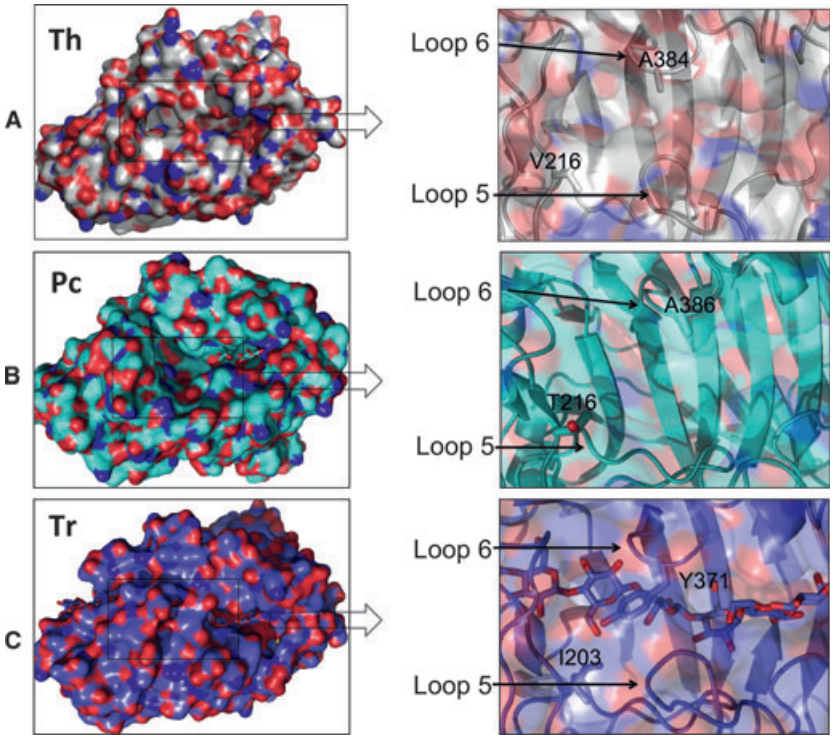


**Fig. 5.** Molecular surface representation of Cel7A from (A) *Trichoderma harzianum* (Th; gray), (B) *Phanerochaete chrysosporium* (Pc; cyan) and (C) *Trichoderma reesei* (Tr; slate). The side face of the tunnel is depicted in a rectangular box. Residues and loops composing the tunnel side-face are indicated. Regions in red and dark blue represent negatively and positively charged residues, respectively.

the presence of Ala384 in loop 6 of Th_Cel7A in place of Tyr371 in Tr_Cel7A was responsible for considerable differences in the dynamics of the loops enclosing the substrate-binding site. In particular, Tyr260 (loop 5) experienced significant conformational changes in Th_Cel7A, consistent with the higher values of B-factors obtained from the crystal structures. Even more interestingly, the MD simulations showed that the side-chain conformation of Tyr260 plays an important role in modulating the degree of exposure of the substrate-binding site in the unliganded Th_Cel7A CCD. Figure 6A displays the time evolution of the rmsd values of Tyr247 (in loop 5 of Tr_Cel7A) and Tyr260 (in loop 5 of Th_Cel7A), where we observed two states. State 1 denotes protein conformations whose rmsd values relative to the average structure are below 3.0 Å, whereas State 2 refers to those protein conformations whose rmsd values are above 3.0 Å. Figure 6B shows representative snapshots from States 1 and 2 in the region of loops 5 and 6. The results showed that the side chain of Th_Cel7A Tyr260 is substantially more mobile than the corresponding Tyr247 in Tr_Cel7A. In addition, Tyr260 interactions with the residues of loop 6, mainly Ala384, are frequently broken. Whenever such interactions are broken, the binding site becomes completely exposed and Tyr260 adopts a conformation which very closely resembles that in our 2.9-Å resolution structure (Structure 1). By contrast, the side chains of Tyr247 (loop 5) and Tyr371 (loop 6) in Tr_Cel7A remain constantly in contact and the loops enclosing the binding site are never found open. Even though Tyr247 in Tr_Cel7A is also capable of undergoing conformational changes, as shown by the peak in the Tyr247 rmsd near 60 ns, the fluctuations are not sufficient to break the contact between loops 5 and 6, thus preventing the binding site from being completely exposed (opened). Thus, the MD simulations suggest that side-chain conformations of Tyr260 in loop 5 may break contacts with Ala384 in loop 6 and expose the binding site in Th_Cel7A, whereas its counterpart residues Tyr247 (loop 5) and Tyr371 (loop 6) in Tr_Cel7A are never observed to be opened except in the course of our simulations, thus preventing exposure of the binding site.

Analysis of root-mean-square-fluctuations of the α-carbons depicted in Figure 7 also provided valuable information on the dynamics of these enzymes. The data showed that the flexibilities of most loops of Th_Cel7A are greater than those of Tr_Cel7A. In particular, the data showed that loop 4 is substantially

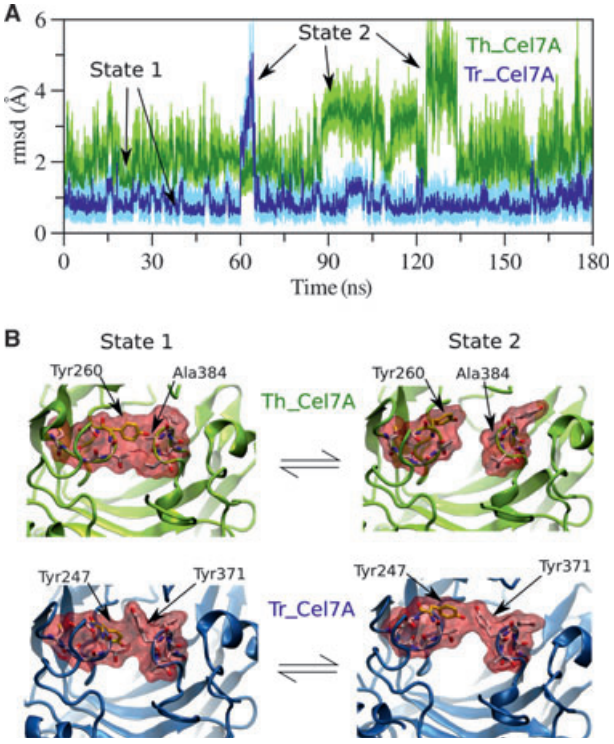


**Fig. 6.** (A) Time evolution of the rmsd for Tyr260 (Th_Cel7A) and Tyr247 (Tr_Cel7A), showing State 1 (rmsd below 3.0 Å) and State 2 (rmsd above 3.0 Å). (B) Snapshots of the region between loops 5 and 6, representative of States 1 and 2. Molecular surfaces are depicted light red. Tyr260 is more mobile than Tyr247, and the conformational changes of Tyr260 in Th_Cel7A create a completely exposed binding site. Although Tyr247 (Tr_Cel7A) can also undergo significant conformational changes, these are not sufficient to create an opened binding site owing to the presence of Tyr371 at loop 6, as opposed to Ala384 at loop 6 in Th_Cel7A.

more mobile in Th_Cel7A. This is interesting because loop 4 is the main loop that covers the tunnel and is located at the entrance of the binding site. A larger flexibility in this region might facilitate the entry of the substrate into the tight binding tunnel of the enzyme.

## Discussion

Our structural characterization of the *T. harzianum* Cel7A CCD reveals that this secreted enzyme has the canonical fold of the GH family 7, with the loops of the β-sandwich domain forming a typical cellobiohydrolase cellulose-binding tunnel instead of the far more open groove found in endoglucanases [19,24]. Despite the high sequence homology of this enzyme with the Cel7A from *T. reesei*, Th_Cel7A, from a structural point of view, is intermediate between Tr_Cel7A and Pc_Cel7D, with significant similarities to the structure of Te_Cel7A in terms of tunnel entrance and shape (Figs 3 and 5).

The identified tunnel side cleft, which gives direct access to the enzyme's active site and exposes the catalytic glutamates of Th_Cel7A to the solvent, is directly dependent on the conformation of the Tyr260 side chain. The cleft is partially obstructed when this residue is solvent-exposed. The observed flexibility of Th_Cel7A Tyr260 in the absence of ligands is not unexpected. A similar effect has been previously reported for the equivalent Tyr247 in unliganded Tr_Cel7A [19,26]. Nevertheless, our MD simulations in the absence of substrate indicate that the presence of the short side-chain from Ala384 allows almost free motion of Tyr260 toward the inner and solvent-exposed conformations in Th_Cel7A. The movements of the Tyr247 side chain in Tr_Cel7A are restrained by the interactions with Tyr371 (Ala384 in Th_Cel7A) and are accomplished simultaneously with conformational changes of Tyr371, which points towards the inner part of the tunnel when Tyr247 is in the solvent-exposed conformation, and vice versa. Despite the relative conformational freedom of Tyr247, our MD results indicate that this residue makes persistent contacts with Tyr371 at loop 6 of Tr_Cel7A. Thus, the side opening that exists in Th_Cel7A structure, does not exist in Tr_Cel7A since it is mostly covered either by Tyr247 or Tyr371 (Figs 3 and 6).

Sequence analyses of cellobiohydrolases I from different organisms show that the position occupied by Ala384, although moderately conserved, displays predominantly certain types of residues, amongst which alanine or tyrosine are the most frequent. Interestingly, most of the Cel7A enzymes that have a gap at loop 3, and thus exhibit a wider tunnel entrance, also contain an alanine residue, which in *T. harzianum* is observed to guarantee additional access to the cellulose-binding tunnel. The dynamically enhanced side opening to the substrate-binding tunnel, alluded to above, is structurally novel. While the crystallographic structure of Te_Cel7A lacks part of the side tunnel face owing to the presence of a flexible loop [21], the Pc_Cel7D structure shows a completely opened tunnel [20]. This is a result of a shorter loop 5 that completely exposes the tunnel and may be associated with higher rates of hydrolysis, reduced processivity and considerable endoglucanase activity of Pc_Cel7D [20,21].

Considering the overall tunnel shape, length and residue conservation, it is expected that the sugar contacts described for Tr_Cel7A [24] are maintained in Th_Cel7A. Except for Gln101, all the residues interacting with the glucosyl units in *T. reesei* are conserved in *T. harzianum*, including the four tryptophan residues

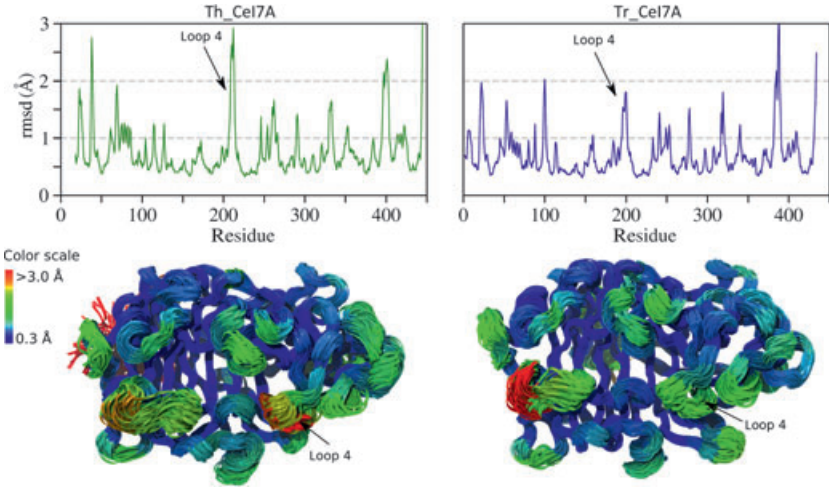


**Fig. 7.** Root-mean-square-fluctuation (rmsf) of the α-carbons of Th_Cel7A and Tr_Cel7A. The upper part of the figure shows the rmsf as a function of the residue numbers and the lower part of the figure shows superposed snapshots of the MD simulations colored according to the rmsf of each α-carbon (see the color scale). Loop 4, which covers the tunnel at the entrance of the binding site and is substantially more mobile in Th_Cel7A, is indicated.

that account for the formation of sites −7, −4, −2 and +1 in Tr_Cel7A, which, together with tyrosine residues, interact with the hydrophobic glucosyl faces of the sugar units and presumably act as lubricating agents facilitating the processive action of the enzyme [24]. It is important to mention that the corresponding Tr_Cel7A residue, Q101, is not present in Th_Cel7A, and the absence of specific H-bonds at sites −6 and −5 lowers the binding affinity of Th_Cel7A to the substrate, which is consistent with the higher $K_{max}$ experimentally observed for Th_Cel7A (compared with Tr_Cel7A). However, the same amino-acid substitution might be favorable for the sliding of the cellulose chain into the active site. Moreover, as a result of the less restricted conformational mobility of Th_Cel7A Tyr260, which is equivalent to the more conformationally constrained Tr_Cel7A Tyr247, the specific H-bond interactions with the substrate at site −2 might be weaker in Th_Cel7A, contributing to the sliding of the sugar units within the Th_Cel7A catalytic tunnel. This is consistent with the higher $k_{cat}$ value, as observed in our enzymatic studies of the enzyme.

Our activity studies showed that the cellobiohydrolase from *T. harzianum* is active against cellulosic substrates with different degrees of crystallinity and polymerization. Th_Cel7A was efficient in depolymerization of highly crystalline BMCC, less efficient for hydrolysis of crystalline and insoluble Avicel and Sigmacell 20, and poorly efficient in CMC hydrolysis (Fig. 2). The specific hydrolysis profile obtained for Th_Cel7A in general correlates well with the published data for Tr_Cel7A [32]. The smaller conversion yields of Avicel and Sigmacell 20 might be related to several factors, such as lower accessibility of these substrates compared with BMCC, their dehydration/rehydration before experimental studies, significantly different degrees of polymerization, a different ratio between Iα and Iβ types of crystalline cellulose between these substrates, higher amorphous content and a presence of the minor hemicellulose fraction in their composition [33,34]. BMCC was used in our experiments in a never-dried form, whereas both Avicel and Sigmacell 20 were purchased in a dried form and rehydrated before use as substrates for enzymatic hydrolysis. It is known that dehydration/rehydration cycles of cellulosic materials have a significant effect on their porosity and accessibility, both of which tend to show rather a strong decrease after dehydration/rehydration [34,35]. This renders substrates that are less susceptible to enzymatic cleavage [36,37]. Furthermore, aggregation of cellulose microcrystals, entangled structures, noncellulosic components attached to the cellulose, as well as amorphous regions of the cellulose were recently shown to represent obstacles for processive cellobiohydrolase action and to slow down Tr_Cel7A cellulose hydrolysis significantly [38,39]. It is known that native crystalline cellulose is composed of two different crystalline components: cellulose Iα and cellulose Iβ [40]. Bacterial-type celluloses are mostly composed of the Iα (triclinic) form, whereas the Iβ (monoclinic) form of cellulose is dominant in the cotton/wood type of biomass [40–42]. The less thermodynamically stable Iα form of cellulose was shown to be more susceptible to enzymatic degradation by *Trichoderma* cellulases than was Iβ cellulose, and in the mixed substrates the Iα form of cellulose was degraded first by cellulases [42]. Finally, xylose and xylooligosaccharides were found to be strong inhibitors of cellulose hydrolysis by cellulases and they might also have some detrimental effect on the Th_Cel7A conversion of Avicel and Sigmacell 20 [43]. The presence of hemicellulose was demonstrated to be detrimental for biomass

hydrolysis using commercial enzymatic preparations, a problem that can be alleviated by the addition of xylanase [44,45].

For hydrolysis of all three crystalline substrates (BMCC, Avicel and Sigmacell 20), the presence of CBM is beneficial and leads to higher specific activity and conversion yields. Th_Cel7A, as a processive enzyme, is ineffective in CMC hydrolysis, a functional feature shared by Tr_Cel7A which is known to have low CMCase activity [46,47]. The presence of CBM has no effect on the hydrolysis of CMC (Fig. 2), and the tunnel-like shape of processive Th_Cel7A makes it rather ineffective for hydrolysis of this soluble and substituted noncrystalline substrate.

In agreement with our kinetic analysis, the Th_Cel7A catalytic rates are consistently higher than the reported activities for Tr_Cel7A [15,32] using both glucoside and/or lactoside substrates. This seems to be consistent with the more open catalytic active cleft of the enzyme. Te_Cel7A and Pc_Cel7D, both presenting shorter active-site loops and thus a more open and accessible binding site, have higher $k_{cat}$ values compared with Tr_Cel7A, particularly for short-chain oligosaccharides [20,21]. A more open substrate-binding site also seems to be positively correlated with the decreased processivity and increased endoglucanase activity of the enzyme [20,48]. This allows us to speculate that Th_Cel7A will probably have lower processivity but higher additional endoglucanase activity compared with Tr_Cel7A. However, its endoglucanase activity will probably be smaller than that of Pc_Cel7D [48].

To date, *Melanocarpus albomyces* Cel7B (Ma_Cel7B) is the only enzyme that was structurally characterized in its native form in the complex with oligosaccharides [22]. (The Tr_Cel7A–cellooligosaccharide complexes were determined using catalytically deficient mutants [19]). The Ma_Cel7B is strongly inhibited by the product of the enzymatic reaction, cellobiose ($K_d$ = 6 mM), which probably was an important factor for success in obtaining the enzyme crystal structures in the complex with oligosaccharides [22]. The Th_Cel7A enzyme shows very weak inhibition by the end product of the reaction, cellobiose. Unfortunately, all our attempts to solve its native X-ray structure in complex with oligosaccharides failed and production of the catalytically deficient mutants was impeded by the necessity to produce recombinant enzymes expressed in its soluble form using an appropriate expression system. Although the molecular mechanism underlying the low inhibition of Th_Cel7A by cellobiose still remains elusive, this feature of the enzyme might be important for its applications in industrial settings.

## Materials and methods

### Expression and purification of Th_Cel7A

Expression and purification of the wild-type Cel7A from *T. harzianum* IOC 3844 was carried out following the protocol described previously [49]. In short, the cells were cultured for 48 h at 25 °C, then used to inoculate a 7.5-L vessel BioFlo 115 Fermenter (New Brunswick Scientific Co., Edison, NJ, USA) and cultured for a further 5 days at 28 °C, pH 5.0 and 200 rpm. The aeration rate was adjusted to keep the dissolved $O_2$ levels below 60% in the culture medium. The enzyme was purified by chromatography though columns containing DEAE–Sephadex (A-50; Sigma-Aldrich, St. Louis, MO, USA) and phenyl sepharose (GE Healthcare Biosciences, Pittsburgh, PA, USA). pH and temperature optima for Th_Cel7A enzymatic activity are pH 5.0 and 50 °C, respectively [49]. Possible endoglucanase contamination in the protein preparations was investigated using activity assays against hydroxyethyl cellulose (Sigma) [26], and no significant activity was detected. The purified protein was dialyzed overnight against 50 mM Tris/HCl, pH 7.0, and concentrated to 8 $mg \cdot mL^{-1}$ before crystallization. The protein concentration was monitored by measuring the absorbance at 280 nm.

### Enzyme activity assays

One per cent (weight by volume) suspensions of the substrates Avicel® PH101 (Sigma), Sigmacell 20 (Sigma) and a solution of CMC (Sigma) were prepared according to the manufacturer's instructions and used for analysis of enzyme activity on insoluble and soluble substrates, respectively. BMCC was prepared from commercially available *Acetobacter xylinum* cellulose, as described previously [50].

Avicel® and Sigmacel 20 have a degree of polymerization (DP) between 150 and 500, and are mostly composed of crystalline cellulose (85–95%), with minor fractions of amorphous cellulose and up to 5% of xylose, mannose, galactose and arabinose. The main difference between Avicel® and Sigmacel 20 is their crystallinity, which is approximately 92% for Avicel and 85% for Sigmacell 20 [33,51].

Soluble ionic-substituted noncrystalline CMC is often used for determining the activity of endoglucanases, which randomly cleave intramolecular β-1,4-glucosidic bonds, resulting in a dramatic reduction in the DP and the specific viscosity of CMC [33]. CMC is rather a poor substrate for cellobiohydrolases because owing to the steric confinement imposed by the active-site tunnel, only nonsubstituted glucose units are accessible for enzymatic hydrolysis [19,33,47]. Therefore, an exoglucanase will trim off the unsubstituted residues until further progress along the CMC is prevented by the carboxymethyl groups. Finally, BMCC has a low DP (100–150) and higher crystallinity, accessibility and purity compared with Avicel and Sigmacel 20 [33]. This highly

crystalline cellulose, with a simple morphology, has been used as a model substrate to study the mechanism of crystal erosion [42] and cellobiohydrolase processivity [48]. The crystallinity index of BMCC, measured using X-ray diffraction, was reported to be 95.2% [51].

Activity assays for both full-length Th_Cel7A and its CCD were carried out in 50 mM citrate buffer at the optimum pH and temperature of the enzyme (pH 5.0 at 50 °C) [49]. The reaction mixture of 15 μM of either full-length Th_Cel7A or its CCD was incubated for 1 h for both the soluble and the insoluble substrates. The conversion yields were measured for 24 h under the same conditions described above. No agitation was applied and to avoid inhibition of Th_Cel7A by the cellobiose over time, 10 μM of *Aspergillus niger* β-glucosidase (Novozymes SP188; Novo Nordisk A/S, Bagsvaerd, Denmark) was added to the reaction mixtures [38]. To ensure that Th_Cel7A was totally active during our experiments, we monitored its activity during 24 h under the same experimental conditions. No changes in the specific activity were observed (data not shown). The quantification of released reducing sugars was performed using the 3,5-dinitrosalicylic acid [15]. All enzyme activities were expressed in International Units [i.e. one unit of activity corresponded to the quantity of enzyme hydrolyzing 1 μmol of substrate or releasing 1 μmol of reducing sugars (in glucose equivalents) per minute].

The kinetic assays were carried out in 50 mM sodium acetate buffer (pH 5.0) containing 700 nM enzyme and pNPC and cNPL (Sigma) as soluble substrates at 12 different final concentrations (from 0 to 15 mM). The reaction was started by the addition of pNPC or cNPL at 50 °C for 10 min and stopped with 1.0 M sodium carbonate. The amount of p- or c-nitrophenol released was measured at $\lambda = 405$ nm (spectrophotometer U-2001; Hitachi) and quantified according to the respective standard curves. All experiments were carried out in triplicate.

The effects of Cel7A inhibition were analyzed, as described above, using pNPC as the substrate and different concentrations of cellobiose (0–40 mM) at pH 5.0 and 50 °C. The kinetic parameters were obtained from a Michaelis–Menten plot $V_0$ versus pNPC or cNPL by fitting to a rectangular hyperbolic curve. Both linear and nonlinear fits were carried out using GRAPHPAD PRISM version 5.04 for Windows (GraphPad Software, La Jolla, CA, USA; www.graphpad.com).

## X-ray structure determination

Collection of X-ray data was carried out on beamline MX2 at the Synchrotron Light National Laboratory (LNLS, Campinas, Brazil) [52]. The first X-ray diffraction data set was collected at 2.9 Å resolution from a needle-shaped crystal obtained by the hanging-drop vapor-diffusion technique in the presence of 25% poly(ethylene glycol) 4000 grown using 35% poly(ethylene glycol) 4000 as a precipitant at pH 7.0 [53] (data not shown). The second data set, collected at 1.67 Å resolution, was obtained from optimized three-dimensional crystals grown in 25% poly(ethylene glycol) 4000 and 5 mM $CaCl_2$. The diffraction data were processed using the program HKL2000 [54]. The crystal structures of Cel7A CCD (residues 18–449) were determined by molecular replacement with the program Phaser [55] using the CCD of Cel7A from *T. reesei* (1CEL) as template (82% sequence identity over 494 amino acids of the CCD). The $2F_o - F_c$ map calculated from the unique solution of the top rotation and translation searches showed a clear and contiguous electron density for the protein and active-site molecules not included in the search model. Model building and refinement was carried out using the program Coot [56]. Cycles of restrained refinement with the program Refmac5 [57] were carried out initially using overall temperature factors and then by using isotropic temperature factors. One cycle of simulated annealing was applied using the program Phenix [58]. Water molecules were added using the program Arp/wArp [59], and the stereochemical quality of the protein complex was validated using the program MolProbity [60]. Statistics of the refined structures are presented in Table 2. The visualization and all representations of the structures were carried out using the program PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC; Cambridge, MA, USA).

## MD simulations

The initial coordinates for MD simulations of the CCD of *T. harzianum* and *T. reesei* were taken from the obtained crystallographic structure and from the PDB (PDB entry 8CEL), respectively. Hydrogen atoms were added and special attention was paid to the ionizable residues, whose protonation states were estimated using the server H++ at pH 6 [61,62]. All substrate molecules and glucosamine groups were removed. The structures were solvated using Packmol [63] by insertion of water molecules within a rectangular box such that the solvent layer was at least 16 Å thick. Sodium and chloride ions were added to a total concentration of 0.10 M in order to render the systems electrically neutral. The constructed systems consisted of approximately 60 000 atoms.

The simulations were performed in the ensemble NpT at 1 atm and 310 K using the Langevin/Nosé-Hoover barostat and the Langevin thermostat to control pressure and temperature, respectively. The CHARMM force field [64] was used for protein and the TIP3P model was used for water molecules [65]. The van der Waals interactions were truncated at a cut-off radius of 12 Å and electrostatics were evaluated using the particle mesh Ewald method [66]. All bonds involving apolar hydrogen atoms were constrained at their equilibrium lengths using the SHAKE algorithm [67]. The time step for the numerical integration of the equations of motion was 2.0 fs.

The following steps were performed to prepare the systems for the production runs: step 1, 1000 steps of

conjugated gradient (CG) energy minimization followed by a 200-ps MD run with all protein atoms harmonically restrained; step 2, 1000 steps of CG followed by a 200-ps MD with only the protein alpha carbons harmonically restrained; and step 3, 600 ps MD without any restraints. After these initial steps, 60-ns production trajectories were generated and considered for analyses. For each system, three independent simulations were performed. All simulations were performed using NAMD [68].

### Accession numbers

The coordinates and structure factors of Cel7A CCD from *T. harzianum* have been deposited at the PDB (PDB codes 2YOK and R2YOKSF, respectively).

## Acknowledgements

We thank the support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grants 2008/56255-9, 2009/54035-4 and 2010/08680-2; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via grants 490022/2009-0, 471834/2009-2, 550985/2010-7 and 151951/2008-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPESP).

## References

1 Cao L, Bala G, Caldeira K, Nemani R & Ban-Weiss G (2010) Importance of carbon dioxide physiological forcing to future climate change. *Proc Natl Acad Sci USA* **107**, 9513–9518.
2 Idso SB (1989) Carbon Dioxide and Global Change: Earth in Transition. IBR Press, Tempe, AZ (631 E. Laguna Dr., Tempe 85282).
3 Koerner C & Bazzaz FA (1996) Carbon Dioxide, Populations, and Communities. Academic Press, San Diego.
4 on Ossowski I, Stahlberg J, Koivula A, Piens K, Becker D, Boer H, Harle R, Harris M, Divne C, Mahdi S *et al.* (2003) Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Ce17A. A comparison with Phanerochaete chrysosporium Cel7D. *J Mol Biol* **333**, 817–829.
5 Gao JM, Weng HB, Zhu DH, Yuan MX, Guan FX & Xi Y (2008) Production and characterization of cellulolytic enzymes from the thermoacidophilic fungal *Aspergillus terreus* M11 under solid-state cultivation of corn stover. *Bioresour Technol* **99**, 7623–7629.
6 Zhou J, Wang YH, Chu J, Zhuang YP, Zhang SL & Yin P (2008) Identification and purification of the main components of cellulases from a mutant strain of *Trichoderma viride* T 100–14. *Bioresour Technol* **99**, 6826–6833.
7 Ahamed A & Vermette P (2008) Culture-based strategies to enhance cellulase enzyme production from *Trichoderma reesei* RUT-C30 in bioreactor culture conditions. *Biochem Eng J* **40**, 399–407.
8 Kleywegt GJ, Zou JY, Divne C, Davies GJ, Sinning I, Stahlberg J, Reinikainen T, Srisodsuk M, Teeri TT & Jones TA (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 angstrom resolution, and a comparison with related enzymes. *J Mol Biol* **272**, 383–397.
9 Coutinho PM & Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In Recent Advances in Carbohydrate Bioengineering (Gilbert HJ, Davies GJ, Henrissat B & Svensson B, eds), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.
10 Davies GJ, Wilson KS & Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321**, 557–559.
11 Lynd LR, Weimer PJ, van Zyl WH & Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology (vol 66, pg 506, 2002). *Microbiol Mol Biol Rev* **66**, 739–739.
12 Fagerstam LG & Pettersson LG (1980) The 1.4-beta-glucan cellobiohydrolases of *Trichoderma reesei* Qm-9414 – a new type of cellulolytic synergism. *FEBS Lett* **119**, 97–100.
13 Nummi M, Niku-Paavola ML, Lappalainen A, Enari TM & Raunio V (1983) Cellobiohydrolase from *Trichoderma reesei*. *Biochem J* **215**, 677–683.
14 Stals I, Sandra K, Geysens S, Contreras R, Van Beeumen J & Claeyssens M (2004) Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology* **14**, 713–724.
15 Tomme P, Vantilbeurgh H, Pettersson G, Vandamme J, Vandekerckhove J, Knowles J, Teeri T & Claeyssens M (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem* **170**, 575–581.
16 Harrison MJ, Nouwens AS, Jardine DR, Zachara NE, Gooley AA, Nevalainen H & Packer NH (1998) Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *Eur J Biochem* **256**, 119–127.
17 Medve J, Lee D & Tjerneld F (1998) Ion-exchange chromatographic purification and quantitative analysis of *Trichoderma reesei* cellulases cellobiohydrolase I. II and endoglucanase II by fast protein liquid chromatography. *J Chromatogr A* **808**, 153–165.
18 Teeri TT, Lehtovaara P, Kauppinen S, Salovuori I & Knowles J (1987) Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase I. *Gene* **51**, 43–52.
19 Divne C, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT & Jones TA

(1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524–528.
20 Munoz IG, Ubhayasekera W, Henriksson H, Szabo I, Pettersson G, Johansson G, Mowbray SL & Stahlberg J (2001) Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 A resolution and homology models of the isozymes. *J Mol Biol* **314**, 1097–1111.
21 Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM & Tuohy MG (2004) Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *Eur J Biochem* **271**, 4495–450622.
22 Parkkinen T, Koivula A & Rouvinen J (2008) Crystal structures of *Melanocarpus albomyces* cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. *Protein Sci* **17**, 1383–1394.
23 Kraulis PJ, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J & Gronenborn AM (1989) Determination of the 3-dimensional solution structure of the C-terminal domain of cellobiohydrolase-I from *Trichoderma reesei* – a study using nuclear magnetic-resonance and hybrid distance geometry dynamical simulated annealing. *Biochemistry* **28**, 7241–7257.
24 Divne C, Stahlberg J, Teeri TT & Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* **275**, 309–325.
25 Barr BK, Hsieh YL, Ganem B & Wilson DB (1996) Identification of two functionally different classes of exocellulases. *Biochemistry* **35**, 586–592.
26 Stahlberg J, Divne C, Koivula A, Piens K, Claeyssens M, Teeri TT & Jones TA (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* **264**, 337–349.
27 Castro AM, Pedro KC, Cruz JC, Ferreira MC, Leite SG & Pereira N Jr (2010) *Trichoderma harzianum* IOC-4038: a promising strain for the production of a cellulolytic complex with significant beta-glucosidase activity from sugarcane bagasse cellulignin. *Appl Biochem Biotechnol* **162**, 2111–2122.
28 Claeyssens M, Van Tilbeurgh H, Tomme P, Wood TM & McRae SI (1989) Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. *Biochem J* **261**, 819–825.
29 Voutilainen SP, Puranen T, Siika-Aho M, Lappalainen A, Alapuranen M, Kallio J, Hooman S, Viikari L, Vehmaanpera J & Koivula A (2008) Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol Bioeng* **101**, 515–528.
30 Deshpande MV, Eriksson K-E & Gran Pettersson L (1984) An assay for selective determination of exo-1,4,-β-glucanases in a mixture of cellulolytic enzymes. *Anal Biochem* **138**, 481–487.
31 Vonhoff S, Piens K, Pipelier M, Braet C, Claeyssens M & Vasella A (1999) Inhibition of cellobiohydrolases from *Trichoderma reesei*. Synthesis and evaluation of some glucose-, cellobiose-, and cellotriose-derived hydroximolactams and imidazoles. *Helv Chim Acta* **82**, 963–980.
32 Jager G, Wu Z, Garschhammer K, Engel P, Klement T, Rinaldi R, Spiess AC & Buchs J (2010) Practical screening of purified cellobiohydrolases and endoglucanases with alpha-cellulose and specification of hydrodynamics. *Biotechnol Biofuels* **3**, 3–18.
33 Zhang YHP, Himmel ME & Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv* **24**, 452–481.
34 Bragatto J, Segato F, Cota J, Mello DB, Oliveira MM, Buckeridge MS, Squina FM & Driemeier C (2012) Insights on how the activity of an endoglucanase is affected by physical properties of insoluble celluloses. *J Phys Chem B* **116**, 6128–6136.
35 Esteghlalian AR, Bilodeau M, Mansfield SD & Saddler JN (2001) Do enzymatic hydrolyzability and simons' stain reflect the changes in the accessibility of lignocellulosic substrates to cellulase enzymes? *Biotechnol Prog* **17**, 1049–1054.
36 Arantes V & Saddler JN (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol Biofuels* **3**, 4.
37 Arantes V & Saddler JN (2011) Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. *Biotechnol Biofuels* **4**, 3.
38 Jalak J & Vaeljamaee P (2010) Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnol Bioeng* **106**, 871–883.
39 Jalak J, Kurašin M, Teugjas H & Väljamäe P (2012) Endo-exo synergism in cellulose hydrolysis revisited. *J Biol Chem* **287**, 28802–28815.
40 Atalla RH & VanderHart DL (1984) Native cellulose: a composite of two distinct crystalline forms. *Science* **223**, 283–285.
41 Wiley JH & Atalla RH (1987) Band assignments in the Raman spectra of celluloses. *Carbohydr Res* **160**, 113–129.
42 Hayashi N, Sugiyama J, Okano T & Ishihara M (1998) The enzymatic susceptibility of cellulose microfibrils of the algal-bacterial type and the cotton-ramie type. *Carbohydr Res* **305**, 261–269.
43 Qing Q, Yang B & Wyman CE (2010) Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes. *Bioresour Technol* **101**, 9624–9630.

44 Berlin A, Gilkes N, Kilburn D, Bura R, Markov A, Skomarovsky A, Okunev O, Gusakov A, Maximenko V, Gregg D *et al.* (2005) Evaluation of novel fungal cellulase preparations for ability to hydrolyze softwood substrates – evidence for the role of accessory enzymes. *Enzyme Microb Technol* **37**, 175–184.
45 Varnai A, Siika-aho M & Viikari L (2010) Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose. *Enzyme Microb Technol* **46**, 185–193.
46 Irwin DC, Spezio M, Walker LP & Wilson DB (1993) Activity studies of 8 purified cellulases – specificity, synergism, and binding domain effects. *Biotechnol Bioeng* **42**, 1002–1013.
47 Penttila ME, Andre L, Lehtovaara P, Bailey MJ, Teeri TT & Knowles JKC (1988) Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene* **63**, 103–112.
48 Kurasin M & Valjamae P (2011) Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* **286**, 169–177.
49 Colussi F, Serpa V, Delabona PS, Manzine LR, Voltatodio ML, Alves R, Mello BL, Pereira N, Farinas CS, Golubev AM *et al.* (2011) Purification, and biochemical and biophysical characterization of cellobiohydrolase I from *Trichoderma harzianum* IOC 3844. *J Microbiol Biotechnol* **21**, 808–817.
50 Valjamae P, Sild V, Nutt A, Pettersson G & Johansson G (1999) Acid hydrolosis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. *Eur J Biochem* **266**, 327–334.
51 Park S, Johnson DK, Ishizawa CI, Parilla PA & Davis MF (2009) Measuring the crystallinity index of cellulose by solid state C-13 nuclear magnetic resonance. *Cellulose* **16**, 641–647.
52 Guimaraes BG, Sanfelici L, Neuenschwander RT, Rodrigues F, Grizolli WC, Raulik MA, Piton JR, Meyer BC, Nascimento AS & Polikarpov I (2009) The MX2 macromolecular crystallography beamline: a wiggler X-ray source at the LNLS. *J Synchrotron Radiat* **16**, 69–75.
53 Colussi F, Textor LC, Serpa V, Maeda RN, Pereira N & Polikarpov I (2010) Purification, crystallization and preliminary crystallographic analysis of the catalytic domain of the extracellular cellulase CBHI from *Trichoderma harzianum*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 1041–1044.
54 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromol Crystallogr Pt A* **276**, 307–326.
55 Storoni LC, McCoy AJ & Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **60**, 432–438.
56 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
57 Skubak P, Murshudov GN & Pannu NS (2004) Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr D Biol Crystallogr* **60**, 2196–2201.
58 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
59 Perrakis A, Harkiolaki M, Wilson KS & Lamzin VS (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* **57**, 1445–1450.
60 Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375–W383.
61 Gordon JC, Myers JB, Folta T, Shoja V, Heath LS & Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, W368–W371.
62 Myers J, Grothaus G, Narayanan S & Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **63**, 928–938.
63 Martinez L, Andrade R, Birgin EG & Martinez JM (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* **30**, 2157–2164.
64 MacKerell AD, Bashford DM, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586–3616.
65 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935.
66 Darden T, York D & Pedersen L (1993) Particle mesh Ewald: an N. log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089–10092.
67 Ryckaert JP, Ciccotti G & Berendsen HJC (1997) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Chem* **23**, 327–341.
68 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802.