

# Small-angle X-ray scattering and structural modeling of full-length: cellobiohydrolase I from *Trichoderma harzianum*

Leonardo H. F. Lima · Viviane I. Serpa · Flávio R. Rosseto ·  
Geraldo Rodrigues Sartori · Mario de Oliveira Neto ·  
Leandro Martínez · Igor Polikarpov

Received: 29 November 2012 / Accepted: 22 April 2013 / Published online: 30 April 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Cellobiohydrolases hydrolyze cellulose releasing cellobiose units. They are very important for a number of biotechnological applications, such as, for example, production of cellulosic ethanol and cotton fiber processing. The *Trichoderma* cellobiohydrolase I (CBH1 or Cel7A) is an industrially important exocellulase. It exhibits a typical two domain architecture, with a small C-terminal cellulose-binding domain and a large N-terminal catalytic core domain, connected by an *O*-glycosylated linker peptide. The mechanism by which the linker mediates the concerted action of the two domains remains a conundrum. Here, we probe the protein shape and domain organization of the CBH1 of *Trichoderma harzianum* (ThCel7A) by small angle X-ray scattering (SAXS) and structural

modeling. Our SAXS data shows that ThCel7A linker is partially-extended in solution. Structural modeling suggests that this linker conformation is stabilized by inter- and intra-molecular interactions involving the linker peptide and its *O*-glycosylations.

**Keywords** Cellobiohydrolase · Cellulose · CBH1 · *Trichoderma*

## Introduction

Cellulases are important for the conversion of cellulose-containing biomass into glucose, which can be converted into bioethanol by fermentation (Li et al. 2009). One of the most studied cellulase complexes is produced by the filamentous fungus *Trichoderma reesei*, consisting of several different enzymes (Hayn

**Electronic supplementary material** The online version of this article (doi:10.1007/s10570-013-9933-3) contains supplementary material, which is available to authorized users.

L. H. F. Lima · V. I. Serpa · F. R. Rosseto ·  
G. R. Sartori · M. de Oliveira Neto ·  
L. Martínez · I. Polikarpov (✉)  
Grupo de Biotecnologia Molecular, Instituto de Física de  
São Carlos (IFSC), Universidade de São Paulo (USP), Av.  
Trabalhador São-carlense, 400, Centro, São Carlos,  
SP 13566-570, Brazil  
e-mail: ipolikarpov@ifsc.usp.br

**Present Address:**  
G. R. Sartori  
Institute of Chemistry of São Carlos, University of São  
Paulo, São Carlos, SP, Brazil

**Present Address:**  
M. de Oliveira Neto  
Departamento de Física e Biofísica, Instituto de  
Biociências de Botucatu, Unesp, Univ Estadual Paulista,  
Distrito de Rubião Jr. s/n, Botucatu, SP, Brazil

**Present Address:**  
L. Martínez (✉)  
Institute of Chemistry, State University of Campinas  
(UNICAMP), Campinas, SP, Brazil  
e-mail: leandro@iqm.unicamp.br

and Esterbauer 1985). Roughly 60 % of total protein secretion of *Trichoderma* strains is cellobiohydrolase I (Cel7A; CBH1; 1.4- $\beta$ -D-glucan cellobiohydrolase, E.C. 3.2.1.91), which belongs to family 7 of glycoside hydrolases, GH7 (Cantarel et al. 2009). This enzyme is capable of degrading of crystalline cellulose by splitting of cellobiose units from the non-reducing end of the substrate (Horn et al. 2012; Serpa and Polikarpov 2011; Nummi et al. 1983).

Limited proteolysis and structural studies of CBH1 from *T. reesei* show that it has two domains: a C-terminal cellulose binding module (CBM), and an N-terminal core domain (CCD), responsible for the enzymatic cleavage of the substrates (Vantilbeurgh et al. 1986). These domains are joined by a flexible, *O*-glycosylated and proline/serine/threonine-rich linker peptide. It has been observed that binding affinity and enzymatic activity are affected when the linker is shortened or deleted, suggesting that the length of the linker is important to ensure flexibility and the concerted action of the CBM and the CCD (Srisodsuk et al. 1993; Ting et al. 2009; Beckham et al. 2010).

The first molecular shape of cellobiohydrolases in solution was obtained from small-angle X-rays scattering (SAXS) data. Schmuck and colleagues found that the CBH1 from *T. reesei* has a tadpole shape consisting of a rather isotropic head and a long and flexible tail (Schmuck et al. 1986). The comparison of SAXS data of the intact enzyme with that of its CCD permitted the identification of the head of the tadpole as the catalytic core domain (Abuja et al. 1988; Divine et al. 1994). The same two-module molecular shape was described for other cellulases (Pilz et al. 1990) and for *T. reesei* Cel7A using high-resolution electron microscopy (Lee and Brown 1997). Recent SAXS ab initio characterization of *Humicola insolens* Cel45 and SANS study of *T. reesei* Cel7A (Receveur et al. 2002; Pingali et al. 2011) significantly improved our comprehension of the molecular architecture of full-length cellulases.

The cellobiohydrolases are complex multi-domain proteins which have fundamental importance in the hydrolysis of crystalline cellulosic substrates and, thus, have a pivotal role in plant biomass depolymerization by the fungal enzymes and, as a consequence, for second generation bioethanol production (Horn et al. 2012; Serpa and Polikarpov 2011). However, many questions remain about how these enzymes function and a detailed molecular understanding of their catalytic activity is still missing. In order to comprehend how these enzymes

function, we will have not only to gather detailed knowledge of the hydrolysis mechanism, but also to obtain an in-depth view of their molecular structure and to acquire insights about the synergy between the individual domains of the enzyme in the full-length settings, their dynamics and their interactions with the cellulose substrate.

In an effort to extend our knowledge of molecular organization and conformational dynamics of a full-length cellulase in solution, we determined the dimensions and molecular shape of Cel7A from *T. harzianum* (ThCel7A) (Roussos and Raimbault 1982; Kalra and Sandhu 1986), its conformational dynamics and energetics, using a combination of SAXS, 2D-3D modeling based on SAXS data, and computational modeling of the fully glycosylated ThCel7A. We show that jointly, SAXS and structural modeling permit one to generate full-length cellulase atomic model that improves the accuracy with which experimental X-ray scattering data can be interpreted, in particular by providing a range of possible extensions of the linker. Our study reveals that the ThCel7A CCD and CBM domains are attached by a linker which does not assume its maximal possible elongation, and might contract and distend in lengths of about one cellobiose unit, thus providing insights into the possible motions of ThCel7A on the substrate surface.

## Materials and methods

### Sample preparation

The enzyme production was accomplished in a 7.5 L bioreactor (*BIOFLO 110, New Brunswick Scientific*). Temperature, pH and dissolved oxygen were controlled at 28 °C, pH 5.0 by addition of NaOH or HCl and dissolved oxygen at 60 % of saturation by aeration and agitation revolution per minute. After 5 days of fermentation of *T. harzianum* with microcrystalline cellulose (Avicel<sup>®</sup>) as inducer of cellulase production, the extracellular extract was filtered and concentrated via tangential flow filtration system (TFF) using a hollow fiber technology (*GE Healthcare Life Science*). The CBH1 from *T. harzianum* was purified by ion exchange chromatography using DEAE-Sephadex (A-50, Sigma) with 50 mM Tris-HCl buffer, pH 7.0. The CBH1 fraction was eluted with 100 mM NaCl in the same buffer. Fraction of the pure protein was submitted to partial papain (Sigma) proteolysis at a

ratio 50:1 (*ThCel7A*: papain) in 40 mM phosphate buffer, pH 6.0; 5 mM L-cysteine; 2 mM EDTA, 25 °C, for 60 min. Separation of the domains was achieved by loading the digested material onto a Superdex 75 column 10/30 (*GE Healthcare*) with 50 mM Tris–HCl buffer, pH 7.0 and 300 mM NaCl.

Protein concentration was measured by Bradford assay (Bradford 1976) with bovine serum albumin as reference standard. Protein samples were separated and analyzed by SDS-PAGE (Laemmli 1970) and stained with Coomassie brilliant blue R-250 (*Bio-Rad Laboratories*).

#### Small-angle X-ray scattering studies

SAXS data for the full *ThCel7A* and its CCD were collected at the D02A-SAXS2 beamline of the Synchrotron Light National Laboratory (Campinas, Brazil), at 45 and 90 mM, respectively. We used monochromatic X-ray with  $\lambda = 1.488 \text{ \AA}$ . X-ray patterns were recorded using a two-dimensional charge-coupled detector (MarResearch, USA). The sample-detector distance was 1,200 mm, resulting in a scattering vector range of 0.013–0.27  $\text{\AA}^{-1}$ , where  $q = 4\pi/\lambda\sin\theta$  ( $2\theta$  is the scattering angle). The samples, in 50 mM Tris–HCl buffer (pH = 7.0), were centrifuged for 30 min at 23,500 $\times g$ , at 4 °C to remove aggregates and then placed on ice. For SAXS measurements a 1-mm path length cell with mica windows at 10 °C. Two successive frames of 300 s each were recorded for each sample to monitor radiation damage and beam stability. Buffer scattering was recorded prior to sample data collection. SAXS patterns were individually corrected for detector response and scaled by the incident beam intensity and the samples absorption. After buffer scattering subtraction, protein SAXS patterns were integrated using Fit2D software (Hammersley et al. 1996, 1997) and scaled by protein concentration.

#### SAXS analysis and modeling

The radius of gyration,  $R_g$  was computed from Guinier equation (Guinier and Fornet 1995) and by indirect Fourier transform method using Gnom package (Svergun 1991). The distance distribution function  $p(r)$  also was calculated using Gnom, and the maximum diameter,  $D_{max}$  was obtained. Molecular weights were calculated for all constructions using SAXSMoW (Fischer et al. 2010). Dummy atom models (DAMs)

were calculated from the SAXS curves of the catalytic domain of *ThCel7A* and also of *ThCel7A* containing both domains using Dammin (Svergun 1999). Twenty independent ab initio shape determinations with different starting conditions and without symmetry constraints led to consistent results as judged by the structural similarity of the output models, yielding nearly identical scattering patterns and fitting statistics in a stable and self-consistent process. DAMAVER was used for automated analysis and averaging of multiple reconstructions, permitting both to analyze the stability of the reconstruction convergence and to yield the most probable particle model (Kozin and Svergun 2001). The CBM and CCD structures were predicted by Swiss Model Server (Arnold et al. 2006; Bordoli et al. 2009) using as templates the structures of the *T. reesei* CBM (PDB id: 1CBH) (Kraulis et al. 1989) and CCD (PDB id: 2V3I) (Momeni et al. 2013), which have 80.6 and 81.8 % of sequence identity with the *ThCel7A* domains, respectively. The crystal structure of the CCD of *ThCel7A* was later obtained (Textor et al. 2012) and the RMSD between the later crystallographic structure and the one from the *T. reesei* CCD is only 0.5  $\text{\AA}$ , thus confirming that the models are adequate in view of the low resolution of SAXS data. SASREF (Petoukhov and Svergun 2005) was used to obtain a rigid body model (Sasref-RBM) displaying the optimal relative positions of the CCD and CBM that reproduce SAXS data. Crysol (Svergun et al. 1995) was used to generate the simulated scattering curve from structural models, and also to evaluate  $R_g$  and  $D_{max}$ . We used Supcomb (Svergun et al. 2001) to superimpose the CBM1 DAM and the RBM. Figures were generated with PyMOL (PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

#### Structural models for full-length *ThCel7A*

Full-length *ThCel7A* was modeled using the I-TASSER server (Zhang 2007) with the structure of the two domains positioned according to the Sasref-RBM as an user added template, and distance restraints on the CBM  $\beta$ -sheets and disulfide bonds. The following internal databank templates were used by the software: *T. reesei* CBM1 CCD (PDB id: 1celA), *Hypocrea jecorina* Cel7A (PDB id: 2v3iA), *Phanerochaete chrysosporium* cellobiohydrolase (PDB id: 1gpiA), *Talaromyces emersoni* cellobiohydrolase (PDB id: 1q9hA) and *Melanocarpus albomyces* cellobiohydrolase (PDB id: 2rfwA). The

most accurate structure built by I-TASSER had a C-score value of 0.47, which indicates that a model is reliable (Zhang 2007; Roy et al. 2010). After this study was finished, the crystallographic structure of the CCD of ThCel7A was obtained (Textor et al. 2012) and revealed that the CCD model produced by I-TASSER differs from the crystal structure only in fine details which are not significant for the interpretation of SAXS data for its low resolution. The reliability of the full-length model is, therefore, reinforced, and the differences are not significant to affect the analysis of the linker extensions which we report.

### Glycosylations

The catalytic domain N-glycans and the linker O-glycans were added using the GLYCAM server (<http://www.glycam.com>). The length of the glycoside at each site was chosen as suggested by previous mass spectrometry and chromatography studies (Stals et al. 2004; Harrison et al. 2002) of *T. reesei* Cel7A (TrCel7A). The ThCel7A linker has two more potential O-glycosylation sites relative to TrCel7A. However, one of the extra threonines in ThCel7A (Thr 470) belongs to the CBM sequence and interacts with the core of the CBM in our molecular model build based on the experimental structure for *T. reesei* CBM (Mattinen et al. 1997). Therefore we chose to under-glycosylate the ThCel7A linker region rather than risking the addition of non-existing glycans in a site interacting with the CBM. One O-linked  $\beta$ -D-mannosyl per serine and two or three O-linked linear bonded (1  $\rightarrow$  4)- $\beta$ -D-mannosyl per threonine were added. Similarly, we added branched segments consisting of the suggested (Glc $\alpha$ )1-(Man $\alpha$ )6-Man $\beta$ -(Glc $\beta$ )2 (Stals et al. 2004) N-linked to the asparagines 126, 283 and 397, of the CCD. Similar strategies were used in previous computational studies of full-length *T. reesei* cellobiohydrolase (Zhong et al. 2008; Zhong and Xie 2009).

### Structural models with variable linker lengths

The modeled structure described above was used as a starting point for the generation of various conformations of the glycosylated linker and, thus, of the relative orientation of the CCD and CBM domains. Cycles of energy minimization and molecular dynamics simulations were used to generate such an ensemble. AMBER type force fields (Cornell et al. 1996)

were used. The protein and the ions were modeled using ff99SB (Hornak et al. 2006), the carbohydrates with GLYCAM06 (Kirschner et al. 2008) and water with the TIP3P force field (Jorgensen et al. 1983). Topologies were constructed with AmberTools (Case et al. 2010). Thirty-two sodium and sixteen chloride ions were added to render the system neutral and to mimic the  $\sim 50$  mM NaCl concentration of the SAXS experiment. 24 cysteine residues were involved in disulfide bonds. The protein was solvated in a rectangular box of water with initial dimensions 128.1 Å by 100.0 Å by 91.0 Å. The resulting system contained 154,152 atoms. MD simulations were performed with NAMD (Phillips et al. 2005) applying periodic boundary conditions. A time-step of 2 fs was used. Non-bonded interactions were cutoff at 14 Å. Covalent bonds involving hydrogens atoms were kept fixed using SHAKE (Vangunsteren and Berendsen 1977). Temperature and pressure were controlled using Langevin barostats and thermostats with a 2 ps<sup>-1</sup> damping coefficient.

The initial model was equilibrated as follows: all the protein and carbohydrate atoms were harmonically restrained, and the energy of the system was minimized by 60,000 conjugated gradient (CG) steps, followed by 72 ps MD at 1 bar and 300 K. Next, the system was minimized without restraints by 20,000 CG steps, followed by another harmonically restrained 60 ps NPT simulation. The carbohydrates and the side chains of the glycosylated residues were allowed to relax for 300 ps. Next, linker residues restraints (residues 445–472) were removed and a 200 ps MD was performed. Finally, the linker C-terminal extension (residues 445–467) was restrained again, and the linker N-terminal and the CBM (residues 468–505) relaxed for 52 ps.

In order to generate a series of structural models that could account for some of the conformational flexibility of CBH1 linker, we conducted a series of short heating/annealing cycles of the structure, by randomly varying the temperature of the system between 250 and 550 K in approximate 500 ps steps, for a total simulation time of 4 ns. Five structural models were extracted from this variable temperature simulation from simulation frames found in plateaus within the 300–310 K temperature range (Supp. Fig. S1). The interaction energies between different subsets (CCD, linker, CBM and glycosylations) were evaluated for these structures as the average energies in the  $\sim 300$  K

temperature plateau from which they were extracted. These conformations were used to refine the SAXS model. Heating was performing by adjusting the Langevin thermostat temperature to target temperatures, with a coupling constant of  $5 \text{ ps}^{-1}$ . The heating cycles were relatively fast, and given the high disulfide bond content of the globular domains, their fold remained mostly unaltered, rendering variable relative positions between the CCD and CBM, with different carbohydrate conformations. Ramachandran plots did not reveal signs of denaturation.

Root mean square deviations (rmsds) were computed by aligning the CCD residues 16–443 of the annealed trajectory to the average structure of each set. The rmsds were used to quantify the structural deviations of each model relative to the model obtained with I-TASSER, and reflect essentially the variability of linker lengths, as the domain structures remained practically unaltered. We also computed the radius of gyration ( $R_g$ ), the CBM-CCD distance ( $d_{\text{CBM-CCD}}$ ), and energetic parameters as protein–protein and carbohydrate–protein interactions, linker hydration energy, linker hydrogen bonds, and the water density in the vicinities of the linker. The CCD-CBM distance was computed as the distance between the geometrical center of the CCD surface which binds the linker and the center of mass of the CBM. A water molecule was considered to be in the vicinity of the linker if found at less than  $5 \text{ \AA}$  of any linker amino acid atom. Energies were calculated with VMD (Humphrey et al. 1996) and NAMD (Phillips et al. 2005) and figures were generated with VMD and Pymol (PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC). The rmsd,  $R_g$  and backbone hydrogen bonds were computed with Wordom (Seeber et al. 2007). To minimize limited sampling problems, averages of the major structures and distance parameters were also computed using linker non-bonded energies as Boltzmann weighting factors for a 300 K configuration ensemble.

## Results and discussion

### SAXS data analysis reveals a partially-extended *ThCel7A* linker conformation

Enzyme concentration was varied in SAXS experiments from 45 to 90 mM for the full-length *ThCel7A* as well as its catalytic domain, with no signs of aggregation, as the scattering curves were unaltered. A

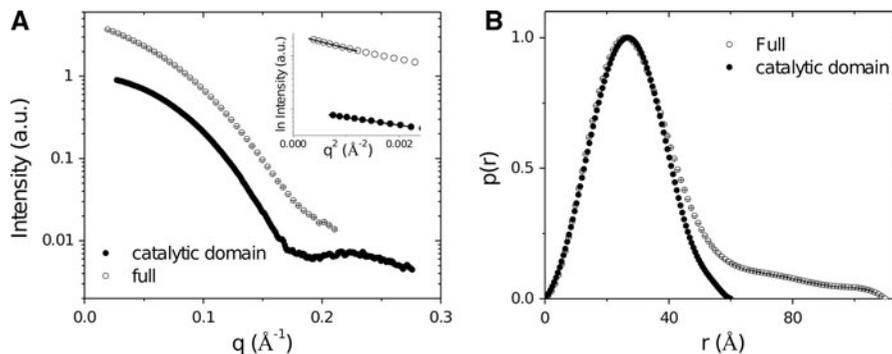
comparison of the scattering curves of full-length *ThCel7A* and its isolated CCD at different concentration are shown in supplementary Material Figure 2 (Supp. Fig. S2). SAXS analyses were performed using scattering data collected with 45 mM concentration for both constructs. (Fig. 1a). From Guinier analysis we computed the radii of gyration ( $R_g$ ) of full-length *ThCel7A* and its isolated catalytic domain as 27.39 and 21.27  $\text{\AA}$ , respectively. The linearity of the Guinier plot in the low  $q$ -region shows that the scattered intensities follow the Guinier law and indicates that *ThCel7A* preparations are monodisperse (Fig. 1a—insert). Computed (Fischer et al. 2010) molecular weights of the isolated CCD and full-length *ThCel7A* are 43 and 46.7 kDa, respectively, whereas the theoretical values are 47.30 and 53.00 kDa. The close match between experimentally determined and theoretically calculated molecular weights supports the fact that both the full-length-*ThCel7A* and the CCD are monomeric in solution.

The pair-distance distribution functions  $p(r)$  indicate that the full-length *ThCel7A* has an elongated form with a  $D_{\text{max}}$  of 110  $\text{\AA}$  and the CCD is globular with  $D_{\text{max}} = 60 \text{ \AA}$  (Fig. 1b). This 50 $\text{\AA}$  difference derives from 58 additional residues (of the linker and CBM), relative to 431 residues of the CCD. The linker has 22 residues, and its maximum possible length is 85  $\text{\AA}$ . Therefore, since the the distance of the CBM and CCD is, at most, 50  $\text{\AA}$ , the linker is only partially extended in solution.

A rigid body model of full-length *ThCel7A* was generated by using SASREF (Sasref-RBM), displaying the optimal relative positions of the CCD and CBM that reproduce SAXS data. Experimental and predicted curves from CCD and full-length *ThCel7A* Sasref-RBM are similar and shown in Fig. 2a and b.

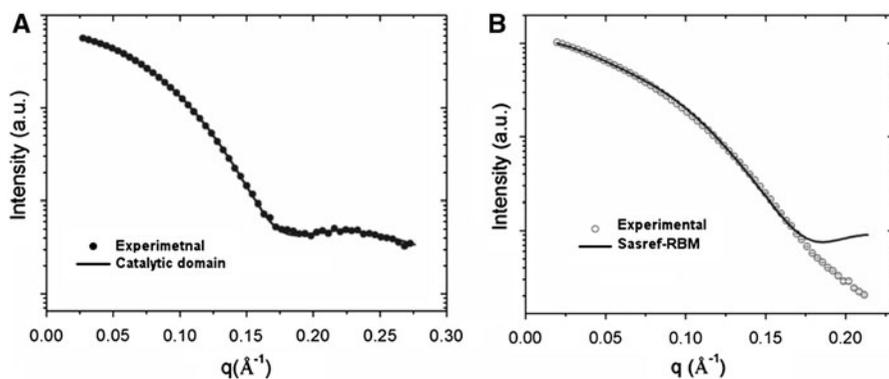
Three-dimensional dummy atom models (DAMs) of the CCD and of the full-length *ThCel7A* were determined from the SAXS curves. The DAM of the full-length *ThCel7A* is elongated, and clearly reveals two domains, one considerably larger than the other. The radii of gyration ( $R_g$ ) and the maximum diameter ( $D_{\text{max}}$ ) for catalytic domain and for full-length *ThCel7A* are given in Table 1. SAXS-derived DAMs of the *ThCBH1* CCD superimposed with the CCD crystal structure is shown in Fig. 3a. The DAM model for the full-length enzyme superimposed to the Sasref-RBM is given in Fig. 3b.

The average dimension of the linker in solution was computed from the relative positions of the CCD and



**Fig. 1** Small-angle X-ray scattering curves for *ThCel7A*. **a** Experimental scattering curves for the full-length *ThCel7A* at 45 mM (open black circles with errors bars) and its core

catalytic domain at 45 mM (black circles with errors bars). An insert shows Guinier plot. **b** Distance distribution function for the full-length enzyme and the catalytic domain of the *ThCel7A*



**Fig. 2** Small-angle X-ray scattering curves and adjustment procedure. **a** Small-angle X-ray scattering curves for catalytic domain of the *ThCel7A* (black circles with errors bars), simulated curves from homology model of the *ThCel7A*

catalytic domain. **b** Small-angle X-ray scattering curves for full-length *ThCel7A* at 45 mM (open black circles with errors bars), simulated curves from the Sasref-RBM (black solid line)

CBM obtained with SASREF, by measuring the distance from the center of mass of the CBM to the geometrical center of the catalytic surface of CCD. The length of the linker is about 49.5 Å, implying an extension of 2.25 Å per residue. This means that the linker is only partially extended. Experimental results obtained for two other cellulases: endoglucanases Cel5G and Cel45 (from *Pseudoalteromonas haloplanktis* (Violot et al. 2005) and *Humicola insolens* (Receveur et al. 2002), respectively) revealed average extensions of 2.0 and 1.4 Å per linker residue, which corresponds to more compact linkers than the one of *ThCel7A*.

#### Structural models for the full-length *ThCel7A*

The model provided by I-TASSER satisfied the CCD orientation and the CCD-CBM distance and best fitted

with the DAM and the Sasref-RBM, confirming the “tadpole” shape for the enzyme. It presents a linker with an N-terminal extension (Gly444 to Ala458) bent over the CCD. This bent portion of the linker contains four *O*-glycosylated residues: Thr453, Ser454, Thr455 and Thr456. The linker’s N-terminal part folds in the partially-extended conformation. Polar interactions between the linker and the surface of the CCD are established, particularly through the mannosyl ring from mono-mannosylated Ser454. This *O*-glycoside chain interacts with the side-chains of residues Asp143, Thr302 and Asn433 from the CCD β-sheets in all structural models obtained.

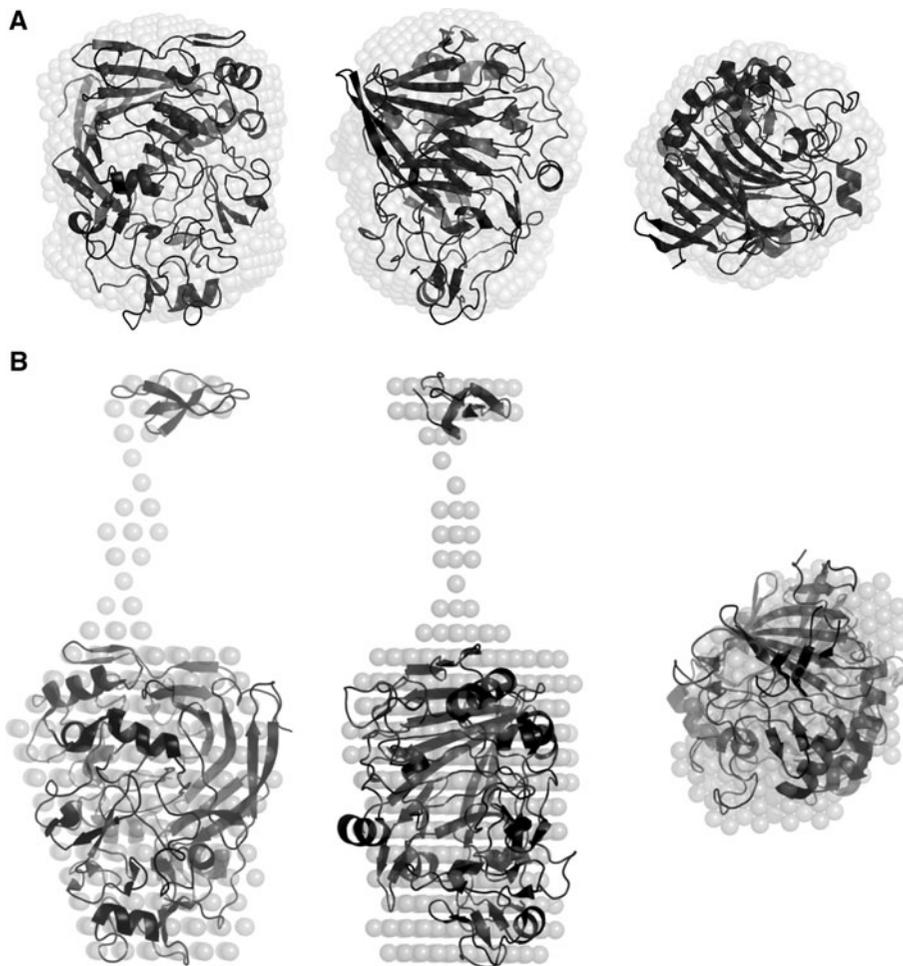
Our short variable-temperature simulation permitted us to obtain a considerable variation of linker lengths. Representative snapshots of five structures are depicted in Fig. 4. The linker extension varied roughly from 48 to 40 Å, while the CBM and CCD domains

**Table 1** SAXS and model derived structural parameters for the catalytic domain and full-length ThCel7A

	Catalytic domain (CCD)			Full-length enzyme		
	Exp 45 mM	Exp 90 mM	Hom model	Exp 45 mM	Exp 90 mM	Sasref-RBM
$D_{\max}/\text{\AA}$	$60.00 \pm 0.50$	$60.00 \pm 0.50$	68.20	$110.00 \pm 0.50$	$110.00 \pm 0.50$	116.80
$R_g/\text{\AA}$	$20.08 \pm 0.05$	$20.04 \pm 0.05$	20.50	$27.56 \pm 0.50$	$27.30 \pm 0.50$	32.67
Resolution/ $\text{\AA}$	22.70	22.70	–	29.39	29.39	–
MW/kDa	43.00	–	–	46.70	–	–
Excluded Vol/ $\text{\AA}^3$	60337	–	–	75699	–	–

Resolution:  $(2\pi/q_{\max})$ 

MW SAXS derived molecular weight

**Fig. 3** SAXS models. **a** Three orthogonal views of the SAXS ab initio envelope for ThCel7A catalytic domain, obtained by Dammin (*shaded spheres*), superimposed to the high-resolution model of the ThCel7A catalytic domain (cartoon). **b** Threeorthogonal views of the SAXS ab initio envelope for the full-length ThCel7A, obtained by Dammin (*shaded spheres*), superimposed to the rigid-body model obtained with Crysol

remained mostly unaltered, as expected (the goal of this short simulation was to provide variable linker conformations only). The rmsd of the structure relative to the initial I-TASSER model are displayed in Fig. 4 and reflect the linker length distribution. The rmsd increases from model I to V, indicating a broader sampling of linker conformations. These conformations will be referred to as CI (Conformation I) to CV (Conformation V) according to the order in which they appeared in the simulation. This order also satisfies a decreasing scale of average linker extensions and gyration radius (Fig. 4).

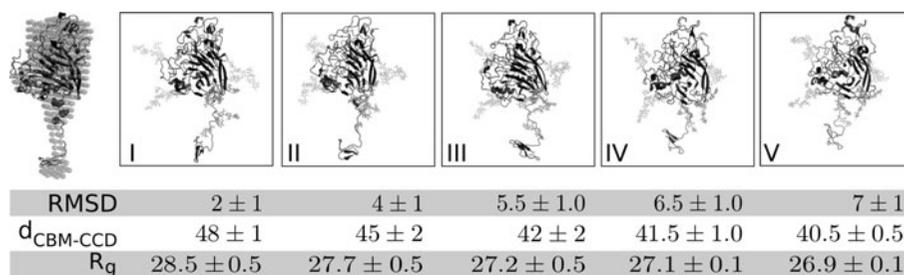
We computed the correlation between the linker extension and its nonbonded intramolecular interactions. As shown in Fig. 5a, CIV and CV display less favorable linker nonbonded interactions. Conformations II and III display the most favorable non-bonded interactions, and span linker extensions from 41 to 47 Å. CI displays the most extended linker and is higher in energy than CII and CIII. Therefore, linker conformations displaying lower intramolecular non-bonded energies are of intermediate length, and the CCD-CBM distance can vary roughly within 42 and 46 Å. The favorable non-bonded linker interactions of CII and CIII result from an intricate energy balance involving protein–protein interactions and hydration of the linker and its glycosylations. The CII energy minimum results from increased solvation (Fig. 5b), associated with minimum linker-protein interactions. Linker-protein interactions are generally favored as the linker is shortened, thus counterbalancing desolvation (Fig. 5c). For CIV and CV, on the other side, favorable linker-protein interactions (Fig. 5c) do not fully counterbalance the penalty of desolvation (Fig. 5b), and these linker conformations display higher non-bonded energies (Fig. 5a).

Boltzmann weighting of gyration radii ( $R_g$ ) and CBM-CCD distances (dCBM-CCD) improved the correspondence between experiment and models. Weighted  $R_g$  and dCBM-CCD were 27.8 and 46.4 Å, respectively. The gyration radius is in agreement with the one recovered from Guinier analysis ( $27.56 \pm 0.50$  Å) and from the DAM ( $30.46$  Å, respectively) and the dCCD-CBM is close to 49.5 Å, the distance experimentally estimated using a *ThCel7A* low-resolution DAM-based model (Table 1).

The maximum model length was that of CI, roughly  $110 \pm 2$  Å, which is in excellent agreement with the  $D_{\max}$  computed from  $p(r)$  analysis and from the DAM ( $110.00 \pm 0.50$  and  $110.70$  Å, respectively). Therefore, the protein dimensions recovered from our MD based models agree well with the experimental SAXS data, supporting the overall structural analysis.

The average structures of CI–CV were aligned to the SAXS envelope (Supp. Fig. S3). The minimum  $\chi$  (1.62) was obtained for CII, which has also the most favorable linker non-bonded interactions (Fig. 5a). CIV and CV, which display less favorable linker interaction energies, fit worse to the SAXS data. Therefore, structural models with partially-extended linkers (I–III) are more consistent with solution data than contracted-linker structures (IV and V). Thus, linker solvation is favored relative to linker-(CBM + CCD) interactions. As interactions with the protein are not maximized, linker function as a dynamic spacer is reinforced.

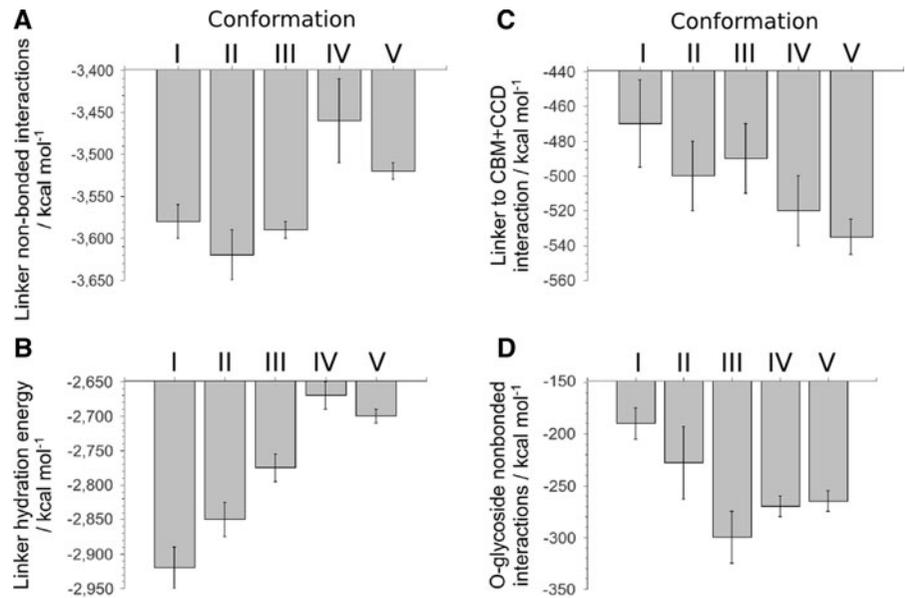
Using the structural models, a maximum likelihood structure for the linker, the CBM and the carbohydrates was built. As CII has smaller linker non-bonded interactions, the proposed model was composed predominantly by conformations representative of



**Fig. 4** Overview of the structural models of full-length *ThCel7A* obtained by computer modeling. The structural model obtained by I-TASSER is superimposed to the DAM model

obtained from SAXS, and has an extended linker. Representative conformations I to V are shown. The RMSDs are computed relative to I-TASSER model

**Fig. 5** Energetic analysis of the conformations extracted from MD simulations: **a** Linker non-bonded interactions; **b** Linker hydration energy; **c** Linker interactions with CBM + CCD (the rest of the structure); **d** Non-bonded interactions of the *O*-glycosylation extensions



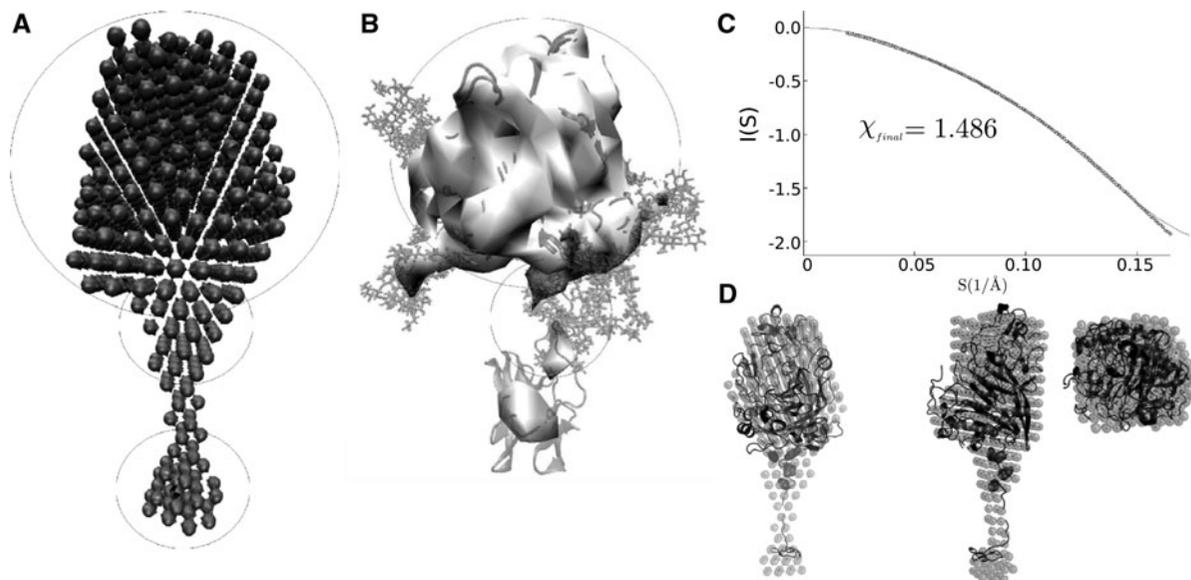
basin II (85 structures of CII for each 10 structures from CII and CIII). Structural averages from each subset were used to reproduce the scattering curve, confirming the greater similarity CII with the experimental model (Supp. Fig. S3). The optimal rigid body model reduces the  $\chi$  curve fitting parameter from 1.62 (for CII only; Supp. Fig. S3) to 1.49 (Fig. 6c). The model also displays improved fitting to the DAM envelope (Fig. 6d). Particularly noteworthy is the fitting of the carbohydrate extensions at residues Ser461, Ser462, Thr463 and Ser465 in the bulky linker region below the CCD. Moreover, the highly mobile CBM averaged over this set of distinct conformations is confined around its center of mass, contributing to the better fit of the DAM density.

The presence of carbohydrate extensions is not clearly reflected in the volume of the SAXS-derived DAM (Fig. 6a, b). This might be a consequence of the high mobility of these side chains in solution. Figure 6a and b compares the SAXS DAM model with a volumetric map derived from the mass-weighted density of *ThCel7A*, averaged from all conformations obtained from MD, using a grid resolution of 5 Å (usually considered an upper limit for the information to be extracted from SAXS data). A superposition of the average structures of the glycosylation carbohydrates is also shown. It is clear that the volume occupied by glycosylations cannot be associated with the volume of the DAM. The DAM represents the atomic density of the core of the protein,

that is, the highly globular and rigid CCD, an average straight linker, and the central region of the CBM. The heterogeneous binding of N-glycosylations also might reduce their contribution to X-ray scattering. Indeed, fully N-glycosylated protein (at three sites) was previously found only in minimal medium, while growth in more rich environments resulted in variable glycosylation number and extension (Stals et al. 2004).

The SAXS and MD data obtained for *ThCel7A* in this work allow us to confirm the tadpole shape of the enzyme and reveal that its maximum dimension is significantly smaller than obtained previously for *TrCel7A* from SAXS (180 Å) (Schmuck et al. 1986) or electron microscopy ( $151 \pm 13$  Å) (Lee and Brown 1997), but close to  $D_{\max}$  obtained from the recent SANS studies of the same enzyme (between  $90 \pm 5$  and  $110 \pm 5$  Å for different pH values (Pingali et al. 2011)). The *ThCel7A*, highly homologous to *TrCel7A* and with a similar size (513 and 505 residues, respectively), has a  $D_{\max}$  of 110 Å. The  $R_g$  value of 42.7 Å obtained in the early studies for the full-length *TrCel7A* is also much larger. Our experimental SAXS data implies a  $R_g$  value equal to 27.6 Å for the full-length *ThCel7A*, which is similar to the  $R_g$  values (between  $26.1 \pm 2.1$  and  $29.7 \pm 1.2$  Å) determined for *TrCel7A* from small-angle neutron scattering (Pingali et al. 2011).

Receveur et al. (2002) demonstrated by SAXS that the cellulase linkers are flexible and extended, and



**Fig. 6** Refinement of the structural model using molecular dynamics simulations: **a** Dummy atom model (DAM) obtained from raw SAXS data. **b** Average volumetric map of conformations obtained in MD simulations, with glycosylations represented as stick models that cannot be fitted into the volume

argued that the cellulases move on crystalline cellulose in a caterpillar-like fashion. However, others suggested that the CCD is the only responsible for sliding on the substrate, the CBM function being the enhancement of the concentration of enzyme at the substrate surface (Igarashi et al. 2009; Stahlberg et al. 1991). Our molecular modeling studies let us to conclude that the most plausible structure of *ThCel7A* has a partially-extended linker conformation, which can display significant linker length variability, supporting the caterpillar-like model.

## Conclusions

Our combined SAXS and molecular modeling analysis indicate that the most energetically favorable and SAXS-compliant conformations (sub-trajectory II) reveal the CCD and CBM relatively distant from each other, but allow for effective interactions of the glycosylations of the linker with the CCD surface. Hence, the linker is long and flexible, enhancing the probability of the CBM to bind to the cellulose surface. Another important detail recovered from MD based structural modeling is that the glycosylations are mobile, and this might

defined by the DAM model. **c** Reproduction of the X-ray scattering curve by a structural model based on the weighted average of conformations I–III. **d** Overlap of the best model obtained with the DAM envelope

have implications to the attachment of the enzyme to cellulose. Matthews et al. (2006) suggested that the first hydration layer of the cellulose surface is structured, blocking direct interactions of enzymes with the crystalline substrate. The glycosylations may accelerate this approximation by a gradual substitution of water-cellulose interactions with sugar-cellulose interactions, which should be entropically favored due to the release of water from the surface. Once the enzyme is attached to the substrate, the concerted movement of the linker and the CCD might facilitate driving the substrate into active site of the catalytic domain.

The small energetic cost in conformational transitions between CI and CIII might account for the caterpillar-like motion proposed by Receveur et al. (2002). The range of most easily accessible linker extensions (Fig. 4) is of roughly 9 Å, similar in length to a cellobiose repeating unit (10.4 Å). This corroborates the idea that the linker movements could span one cellobiose unit for each cycle of the hydrolysis process.

These findings might be important for comprehension of the crystalline cellulose hydrolysis by fungal cellobiohydrolases in general and their processivity, and also provide a starting molecular model for a full-

length fungal Cel7A studies by extensive molecular dynamics simulations.

**Acknowledgments** This work was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) via Thematic Process 08/56255-9, grants 2008/05637-9, 2007/08706-9, 2010/16947-9, CeProBio Project (FAPESP 2009/52840-7 and CNPq 490022/2009-0) and INCT Bioetanol (FAPESP/CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also thank the staff of the National Synchrotron Light Laboratory (LNLS, Brazil) for access to the SAXS beamline and others facilities.

## References

- Abuja P, Pilz I, Claeysens M, Tomme P (1988) Domain-structure of cellobiohydrolase-II as studied by small-angle X-ray-scattering—close resemblance to cellobiohydrolase-I. *Biochem Biophys Res Commun* 156(1):180–185. doi:10.1016/S0006-291X(88)80821-0
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2): 195–201. doi:10.1093/bioinformatics/bti770
- Beckham GT, Bomble YJ, Matthews JF, Taylor CB, Resch MG, Yarbrough JM, Decker SR, Bu L, Zhao X, McCabe C, Wohlert J, Bergenstrahle M, Brady JW, Adney WS, Himmel ME, Crowley MF (2010) The O-glycosylated linker from the *Trichoderma reesei* family 7 cellulase is a flexible. *Disordered Protein Biophys J* 99(11):3773–3781. doi:10.1016/j.bpj.2010.10.032
- Bradford M (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal Biochem* 72(1–2):248–254. doi:10.1006/abio.1976.9999
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37:D233–D238. doi:10.1093/nar/gkn663
- Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM et al (2010) Amber 11, Single edn. University of California Press, Berkeley
- Cornell W, Cieplak P, Bayly C, Gould I, Merz K, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* 118(9):2309. doi:10.1021/ja955032e
- Divine M, Stahlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT, Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265(5171):524–528
- Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J Appl Crystallogr* 43(Part 1):101–109. doi:10.1107/S0021889809043076
- Guinier A, Fournet G (1995) Small angle scattering of X-rays, 1st edn. Wiley, London
- Hammersley A, Svensson S, Hanfland M, Fitch A, Hausermann D (1996) Two-dimensional detector software: from real detector to idealised image or two-theta scan. *High Pressure Res* 14(4–6):235–248. doi:10.1080/08957959608201408
- Hammersley A, Brown K, Burmeister W, Claustre L, Gonzalez A, McSweeney S, Mitchell E, Moy J, Svensson S, Thompson A (1997) Calibration and application of an X-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography. *J Synchrotron Radiat* 4(Part 2):67–77. doi:10.1107/S0909049596015087
- Harrison M, Wathugala I, Tenkanen M, Packer N, Nevalainen K (2002) Glycosylation of acetylxyylan esterase from *Trichoderma reesei*. *Glycobiology* 12(4):291–298. doi:10.1093/glycob/12.4.291
- Hayn M, Esterbauer H (1985) Separation and partial characterization of *Trichoderma-reesei* cellulase by fast chromatofocusing. *J Chromat* 329(3):379–387. doi:10.1016/S0021-9673(01)81944-0
- Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH (2012) Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* 5:45. doi:10.1186/1754-6834-5-45
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–725. doi:10.1002/prot.21123
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–38. doi:10.1016/0263-7855(96)00018-5
- Igarashi K, Koivula A, Wada M, Kimura S, Penttila M, Samejima M (2009) High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase i on crystalline cellulose. *J Biol Chem* 284(52):36186–36190. doi:10.1074/jbc.M109.034611
- Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935. doi:10.1063/1.445869
- Kalra M, Sandhu D (1986) Cellulase production and its localization in *Trichoderma-harzianum*. *Folia Microbiol* 31(4):303–308. doi:10.1007/BF02926955
- Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ (2008) GLY-CAM06: a generalizable biomolecular force field, carbohydrates. *J Comput Chem* 29(4):622–655. doi:10.1002/jcc.20820
- Kozin MB, Svergun DI (2001) Automated matching of high-resolution structural models. *J Appl Crystallogr* 34:33–41
- Kraulis P, Clore G, Nilges M, Jones T, Pettersson G, Knowles J, Gronenborn A (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(18):7241–7257. doi:[10.1021/bi00444a016](https://doi.org/10.1021/bi00444a016)
- Laemmli U (1970) Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* 227(5259):680. doi:[10.1038/227680a0](https://doi.org/10.1038/227680a0)
- Lee H, Brown R (1997) A comparative structural characterization of two cellobiohydrolases from *Trichoderma reesei*: a high resolution electron microscopy study. *J Biotechnol* 57(1–3):127–136. doi:[10.1016/S0168-1656\(97\)00111-9](https://doi.org/10.1016/S0168-1656(97)00111-9)
- Li Xh, Hj Yang, Roy B, Wang D, Wf Yue, Lj Jiang, Park EY, Yg Miao (2009) The most stirring technology in future: cellulase enzyme and biomass utilization. *Afr J Biotechnol* 8(11):2418–2422
- Matthews J, Skopec C, Mason P, Zuccato P, Torget R, Sugiyama J, Himmel M, Brady J (2006) Computer simulation studies of microcrystalline cellulose I beta. *Carbohydr Res* 341(1):138–152. doi:[10.1016/j.carres.2005.09.028](https://doi.org/10.1016/j.carres.2005.09.028)
- Mattinen M, Kontteli M, Kerovuo J, Linder M, Annala A, Lindeberg G, Reinikainen T, Drakenberg T (1997) Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*. *Protein Sci* 6(2):294–303
- Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engstrom A, Sandgren M, Beckham GT, Stahlberg J (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *J Biol Chem* 228(8):5861–5872. doi:[10.1074/jbc.M112.440891](https://doi.org/10.1074/jbc.M112.440891)
- Nummi M, Nikupaavola M, Lappalainen A, Enari T, Raunio V (1983) Cellobiohydrolase from *Trichoderma reesei*. *Biochem J* 215(3):677–683
- Petoukhov M, Svergun D (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89(Part 2):1237–1250. doi:[10.1529/biophysj.105.064154](https://doi.org/10.1529/biophysj.105.064154)
- Phillips J, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802. doi:[10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289)
- Pilz I, Schwarz E, Kilburn D, Miller R, Warren R, Gilkes N (1990) The tertiary structure of a bacterial cellulase determined by small-angle X-ray-scattering analysis. *Biochem J* 271(1):277–280
- Pingali SV, O'Neill HM, McGaughey J, Urban VS, Rempe CS, Petridis L, Smith JC, Evans BR, Heller WT (2011) Small angle neutron scattering reveals pH-dependent conformational changes in *Trichoderma reesei* cellobiohydrolase I implications for enzymatic activity. *J Biol Chem* 286(37):32801–32809. doi:[10.1074/jbc.M111.263004](https://doi.org/10.1074/jbc.M111.263004)
- Receveur V, Czjzek M, Schulein M, Panine P, Henrissat B (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J Biol Chem* 277(43):40887–40892. doi:[10.1074/jbc.M205404200](https://doi.org/10.1074/jbc.M205404200)
- Roussos S, Raimbault M (1982) Cellulose hydrolysis by fungi. 1. Screening of cellulolytic strains. *Ann Microb B* 133(3):455–464
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738. doi:[10.1038/nprot.2010.5](https://doi.org/10.1038/nprot.2010.5)
- Schmuck M, Pilz I, Hayn M, Esterbauer H (1986) Investigation of cellobiohydrolase from *Trichoderma reesei* by small-angle X-ray-scattering. *Biotechnol Lett* 8(6):397–402. doi:[10.1007/BF01026739](https://doi.org/10.1007/BF01026739)
- Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A (2007) Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23(19):2625–2627. doi:[10.1093/bioinformatics/btm378](https://doi.org/10.1093/bioinformatics/btm378)
- Serpa VI, Polikarpov I (2011) Enzymes in bioenergy. In: Buckeridge MS, Goldman GH (eds) *Routes to cellulosic ethanol—part II*. Springer, New York. doi:[10.1007/978-0-387-92740-4\\_7](https://doi.org/10.1007/978-0-387-92740-4_7)
- Srisodsuk M, Reinikainen T, Penttila M, Teeri T (1993) Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase-I in its interaction with crystalline cellulose. *J Biol Chem* 268(28):20756–20761
- Stahlberg J, Johansson G, Pettersson G (1991) A new model for enzymatic-hydrolysis of cellulose based on the 2-domain structure of cellobiohydrolase-I. *Bio-Technology* 9(3):286–290. doi:[10.1038/nbt0391-286](https://doi.org/10.1038/nbt0391-286)
- Stals I, Sandra K, Geysens S, Contreras R, Van Beeumen J, Claeysens M (2004) Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology* 14(8):713–724. doi:[10.1093/glycob/cwh080](https://doi.org/10.1093/glycob/cwh080)
- Svergun D (1991) Mathematical-methods in small-angle scattering data-analysis. *J Appl Crystallogr* 24(Part 5):485–492. doi:[10.1107/S0021889891001280](https://doi.org/10.1107/S0021889891001280)
- Svergun D (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879–2886. doi:[10.1016/S0006-495\(99\)77443-6](https://doi.org/10.1016/S0006-495(99)77443-6)
- Svergun D, Barberato C, Koch M (1995) CRY SOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28(Part 6):768–773. doi:[10.1107/S0021889895007047](https://doi.org/10.1107/S0021889895007047)
- Svergun D, Petoukhov M, Koch M (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953
- Textor LC, Colussi F, Silveira RL, Serpa V, Mello BL, Muniz JRC, Squina FM, Pereira N Jr, Skaf MS, Polikarpov I (2012) *FEBS J* 280(1):56–69. doi:[10.1111/febs.12049](https://doi.org/10.1111/febs.12049)
- Ting CL, Makarov DE, Wang ZG (2009) A kinetic model for the enzymatic action of cellulase. *J Phys Chem B* 113(14):4970–4977. doi:[10.1021/jp810625k](https://doi.org/10.1021/jp810625k)
- Vangunsteren W, Berendsen H (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* 34(5):1311–1327. doi:[10.1080/00268977700102571](https://doi.org/10.1080/00268977700102571)
- Vantilbeurgh H, Tomme P, Claeysens M, Bhikhabhai R, Pettersson G (1986) Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*—separation of functional domains. *FEBS Lett* 204(2):223–227. doi:[10.1016/0014-5793\(86\)80816-X](https://doi.org/10.1016/0014-5793(86)80816-X)
- Violot S, Aghajari N, Czjzek M, Feller G, Sonan G, Guet P, Gerday C, Haser R, Receveur-Brechot V (2005) Structure of a full length psychrophilic cellulase from *Pseudoalteromonas haloplanktis* revealed by X-ray diffraction and

- small angle X-ray scattering. *J Mol Biol* 348(5): 1211–1224. doi:[10.1016/j.jmb.2005.03.026](https://doi.org/10.1016/j.jmb.2005.03.026)
- Zhang Y (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* 69(8):108–117. doi:[10.1002/prot.21702](https://doi.org/10.1002/prot.21702)
- Zhong L, Xie J (2009) Investigation of the effect of glycosylation on human prion protein by molecular dynamics. *J Biomol Struct Dyn* 26(5):525–533
- Zhong L, Matthews JF, Crowley MF, Rignall T, Talon C, Cleary JM, Walker RC, Chukkapalli G, McCabe C, Nimlos MR, Brooks CL III, Himmel ME, Brady JW (2008) Interactions of the complete cellobiohydrolase I from *Trichodera reesei* with microcrystalline cellulose I beta. *Cellulose* 15(2):261–273. doi:[10.1007/s10570-007-9186-0](https://doi.org/10.1007/s10570-007-9186-0)