

RESEARCH

Open Access



Crystal structure and biochemical characterization of the recombinant ThBgl, a GH1 β -glucosidase overexpressed in *Trichoderma harzianum* under biomass degradation conditions

Clelton A. Santos^{1†}, Letícia M. Zanphorlin^{2†}, Aline Crucello¹, Celisa C. C. Tonoli³, Roberto Ruller², Maria A. C. Horta¹, Mario T. Murakami³ and Anete Pereira de Souza^{1,4*} 

Abstract

Background: The conversion of biomass-derived sugars via enzymatic hydrolysis for biofuel production is a challenge. Therefore, the search for microorganisms and key enzymes that increase the efficiency of the saccharification of cellulosic substrates remains an important and high-priority area of study. *Trichoderma harzianum* is an important fungus known for producing high levels of cellulolytic enzymes that can be used for cellulosic ethanol production. In this context, β -glucosidases, which act synergistically with cellobiohydrolases and endo- β -1,4-glucanases in the saccharification process, are potential biocatalysts for the conversion of plant biomass to free glucose residues.

Results: In the present study, we used RNA-Seq and genomic data to identify the major β -glucosidase expressed by *T. harzianum* under biomass degradation conditions. We mapped and quantified the expression of all of the β -glucosidases from glycoside hydrolase families 1 and 3, and we identified the enzyme with the highest expression under these conditions. The target gene was cloned and heterologously expressed in *Escherichia coli*, and the recombinant protein (rThBgl) was purified with high yields. rThBgl was characterized using a comprehensive set of biochemical, spectroscopic, and hydrodynamic techniques. Finally, we determined the crystallographic structure of the recombinant protein at a resolution of 2.6 Å.

Conclusions: Using a rational approach, we investigated the biochemical characteristics and determined the three-dimensional protein structure of a β -glucosidase that is highly expressed by *T. harzianum* under biomass degradation conditions. The methodology described in this manuscript will be useful for the bio-prospection of key enzymes, including cellulases and other accessory enzymes, for the development and/or improvement of enzymatic cocktails designed to produce ethanol from plant biomass.

Keywords: *Trichoderma harzianum*, β -Glucosidase, Overexpression, Enzyme prospection, Biomass degradation

*Correspondence: anete@unicamp.br

[†]Clelton A. Santos and Letícia M. Zanphorlin contributed equally to this work

¹ Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, SP, Brazil

Full list of author information is available at the end of the article

Background

β -Glucosidases (EC 3.2.1.21) catalyze the conversion of cellobiose to glucose monomers, which can then be fermented to produce ethanol [1–3]. Together with cellobiohydrolases (EC 3.2.1.91) and endo- β -glucanases (EC 3.2.1.4), β -glucosidases form a powerful cellulolytic system present in all microorganisms that use cellulose as a substrate [1]. The cellobiohydrolases, depending on their specificity, can generate cellobiose by attacking either the reducing or non-reducing ends of cellulose chains, while endo- β -glucanases hydrolyze the internal β -1,4-glucosyl linkages. The β -glucosidases are key enzymes that act at the final stage of plant biomass hydrolysis and are potential candidates for biotechnological applications [1–3].

β -glucosidases play an important role in the saccharification of cellulosic substrates because these enzymes decrease the inhibitory effect of cellobiose on the enzymatic activity of the cellobiohydrolases and endo- β -glucanases [4–6]. However, many known β -glucosidases are sensitive to the glucose product or inhibited by their cello-oligosaccharide substrates [3, 6, 7]. Thus, the enzymatic degradation of cellulosic biomass is a synergistic process, and each enzyme catalyzes an important step in the continuous and complete breakdown of cellulose [5, 8, 9]. Therefore, a better understanding of the complete hydrolytic process can be achieved by studies that dissect the structure and function of the enzymes in this pathway.

The Carbohydrate-Active enZymes (CAZy) database (<http://www.cazy.org>) mainly groups the β -glucosidases into glycosyl hydrolase (GH) families 1 and 3 [10]. This grouping is based on structural characteristics, especially those related to the mechanism of enzymatic catalysis. Although, both act on similar substrates, the GH1 β -glucosidases use a Glu residue as the catalytic nucleophile, whereas the GH3 β -glucosidases use an Asp residue as the nucleophile [11, 12]. The protein data bank (PDB; <http://www.rcsb.org/pdb/home/home.do>) contains many β -glucosidase structures derived from prokaryotic and eukaryotic organisms. The elucidation of the structure of β -glucosidase enzymes has led to a better understanding of how to improve the enzymatic degradation of biomass and constitutes an important field of study.

To increase the efficiency of the saccharification of cellulosic substrates, enzymes with promising features for industrial applications are continually sought [13–16]. In particular, enzymes that are heavily used by microorganisms under specific conditions are promising targets for practical applications. With advances in RNA sequencing (RNA-Seq) technology [17], the transcriptional profiles of *Trichoderma harzianum*, an important fungus with an efficient cellulase machinery, and other cellulolytic microorganisms are currently available [18–22]. This

technology can facilitate the discovery of the differentially expressed genes under specific metabolic conditions.

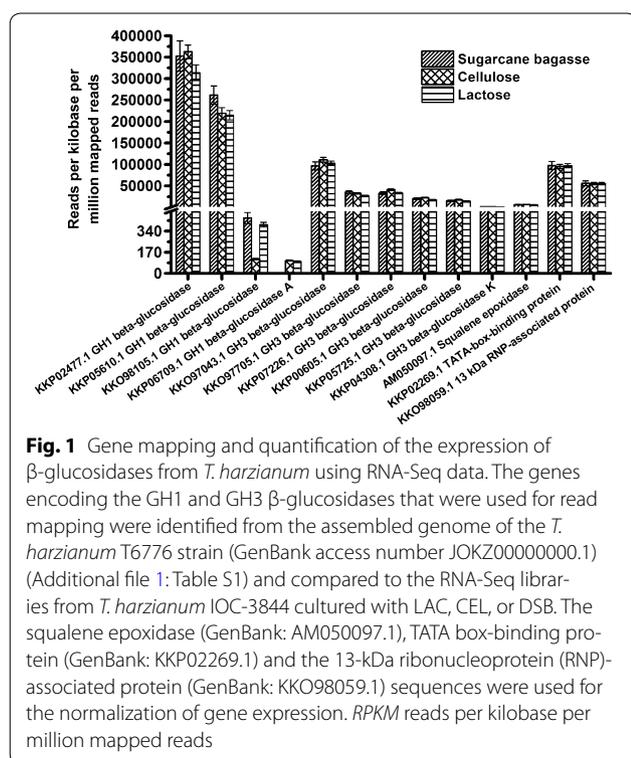
In this study, by screening RNA-Seq libraries, we identified a GH1 β -glucosidase that was highly expressed by the *T. harzianum* strain IOC-3844 under biomass degradation conditions. The target gene was cloned and heterologously expressed in *Escherichia coli*. The crystallographic structure of the purified recombinant protein was determined at a resolution of 2.6 Å. Spectroscopic, hydrodynamic, and biochemical studies were conducted with the recombinant enzyme. Our work sheds new light on the strategies for enzyme bio-prospection and on the potential use of key proteins involved in the enzymatic hydrolysis of cellulose.

Results and discussion

Mapping and quantification of the expression of β -glucosidases from *T. harzianum* using RNA-Seq data

We used RNA-Seq libraries to access the reads per kilo base per million of mapped reads (RPKM) for the multiple β -glucosidases expressed by *T. harzianum* under biomass degradation conditions. Once the transcriptional profile data from an organism under a certain condition is known, the levels of any gene can be determined (mapping and quantifying) within the limitations/conditions of the original experiments [17, 23]. The transcriptome profile of *T. harzianum* IOC-3844 cultured in the presence of lactose (LAC), crystalline cellulose (CEL), or delignified sugarcane bagasse (DSB) was reported by Horta et al. [18]. These data enable the identification of the set of genes involved in biomass degradation and thus provide a powerful tool for future studies.

In this study, we initially identified all of the GH1 and GH3 β -glucosidases using the assembled genome of the *T. harzianum* T6776 strain (GenBank access number JOKZ00000000.1) [24]. We then used these sequences to map and quantify the expression of these genes using *T. harzianum* IOC-3844 transcriptome data. Four GH1 and six GH3 β -glucosidase sequences were found in the *T. harzianum* T6776 genome (Additional file 1: Table S1) and used in the mapping experiments. The mapping results revealed that *T. harzianum* IOC-3844 had higher levels of the GH1 β -glucosidases than the GH3 β -glucosidases (Fig. 1). Among the four GH1 β -glucosidases, two enzymes (KKP02477.1 and KKP05610.1) had the highest overall expression, while the other two enzymes (KKO98105.1 and KKP06709.1) exhibited basal levels of expression. Analyses of the amino acid sequences of these proteins showed sequence identities ranging from 27.46 to 53.78 % (Additional file 2: Table S2). Notably, the sequence for KKP02477.1, the enzyme with the highest expression among all the β -glucosidases studied (approximately 400,000 RPKM),



was the target of our study. However, this protein in *T. harzianum* IOC-3844 (GenBank: KU201604.1) contains a change of a glutamine residue by a glutamic acid at position 335, and a methionine is replaced by an isoleucine residue at position 462, sharing 99.57 % identity.

Although six GH3 β -glucosidases were found in the genome of *T. harzianum*, these enzymes are expressed at similar levels regardless of treatment conditions (LAC, CEL, or DSB) (Fig. 1). The variation in the GH3 β -glucosidase amino acid sequences was investigated, and the sequence identities ranged from 23.93 to 50.00 % (Additional file 3: Table S3). The squalene epoxidase (GenBank: AM050097.1), TATA box-binding protein (GenBank: code: KKP02269.1), and 13-kDa ribonucleoprotein (RNP)-associated protein (GenBank: KKO98059.1) sequences were used as house-keeping genes for the mapping experiments (Fig. 1).

In general, approaches utilizing RNA-Seq libraries may enable the discovery of promising enzymes for industrial applications. Thus, this may be a useful tool to search for highly expressed targets and other accessory enzymes that can improve cellulose hydrolysis during biomass degradation.

Spectroscopic and hydrodynamic characterization of purified rThBgl

After the β -glucosidase with the highest level of expression in *T. harzianum* under biomass degradation

conditions was identified, gene cloning and heterologous protein expression and purification were carried out. rThBgl (468 amino acid residues, 53.2 kDa and theoretical isoelectric point of 5.1) was successfully expressed using *E. coli* as a host. The recombinant protein was purified using two chromatographic steps: nickel affinity (using the N-terminal His₆-tag added by the pET28a expression vector) and size-exclusion chromatography (SEC). Approximately 35.5 mg of protein with a purity greater than 95 % based on 12 % SDS-PAGE was obtained per liter of bacterial culture.

Studies on the protein expression and purification of β -glucosidases from different *Trichoderma* fungus strains have been performed [25–34]; however, most of these studies used eukaryotic expression systems, in particular *Trichoderma reesei* and *Pichia pastoris*. In the current study, we overexpressed and purified rThBgl using an *E. coli* host. This method produces high yields of the recombinant protein, which suggests that it can potentially be used in industrial processes for cellulosic ethanol production [35–38].

The hydrodynamic and spectroscopic properties of the purified rThBgl were assessed (Table 1, Fig. 2). For the hydrodynamic analysis, analytical SEC and analytical ultracentrifugation (AUC) techniques were employed.

Table 1 Hydrodynamic and spectroscopic properties of purified rThBgl

Technique	Property
Predicted hydrodynamic data ^a	$MM_{pred} = 53.24$ kDa $R_0 = 25.15$ Å
Analytical SEC	$MM_{app} = 54.72 \pm 3$ kDa $R_s = 31.39$ Å $f/f_0^b = 1.25 \pm 0.2$
AUC ^c	$S_{20,w}^0 = 4.53 \pm 0.03$ S $MM_{exp} = 58.26 \pm 7$ kDa $f/f_0 = 1.23 \pm 0.05$
CD ^d	α -Helix = 35 ± 4 % β -Sheet = 10 ± 2 %
CD thermal-induced unfolding	$T_m = 49 \pm 1$ °C
Fluorescence	$\lambda_{max}^{rThBgl-folded} = 334 \pm 1$ nm $\langle \lambda \rangle^{rThBgl-folded} = 353 \pm 1$ nm $\lambda_{max}^{rThBgl-denatured} = 354 \pm 2$ nm $\langle \lambda \rangle^{rThBgl-denatured} = 363 \pm 1$ nm

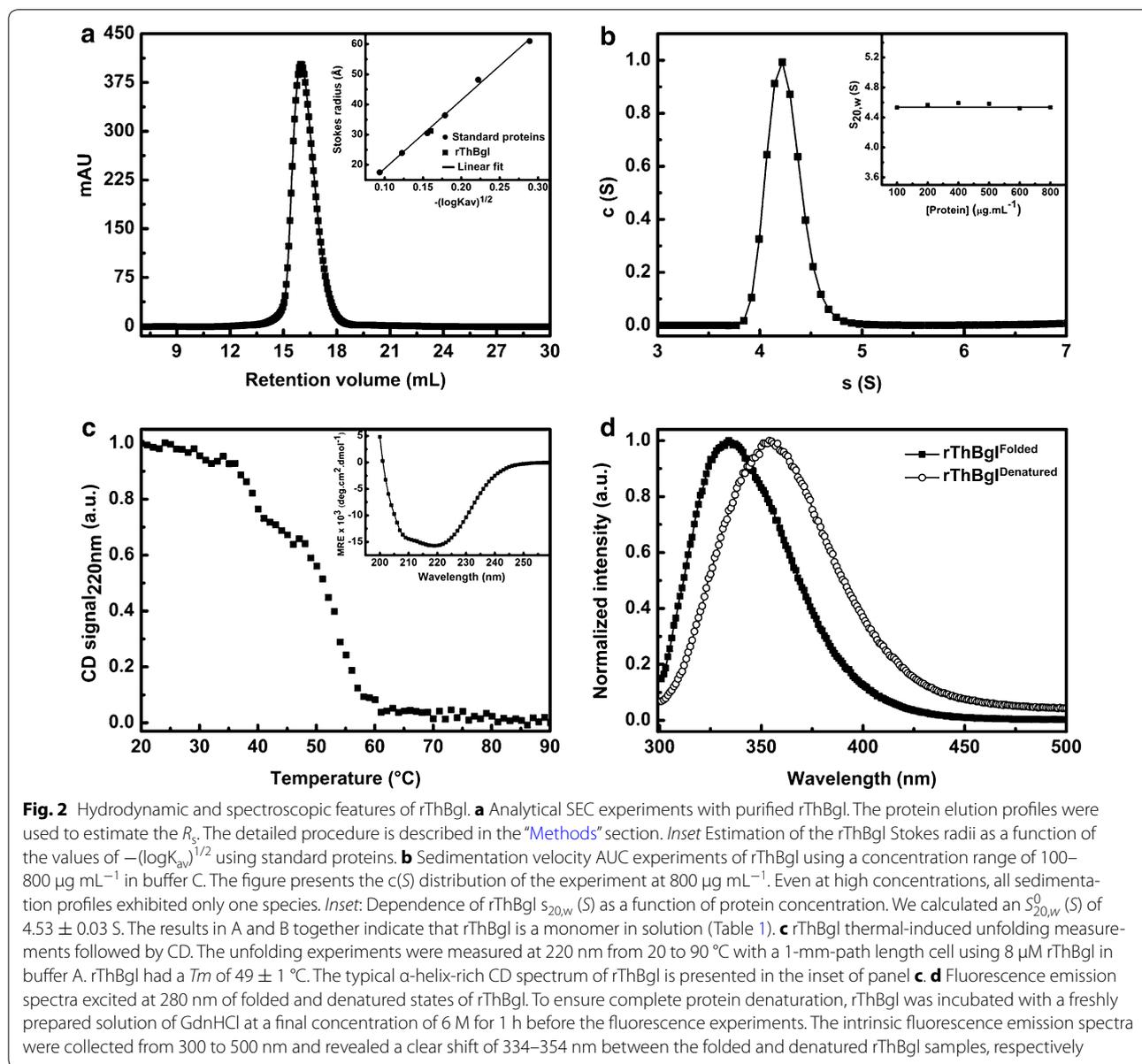
MM_{prep} predicted molecular mass, MM_{exp} experimental molecular mass, MM_{app} apparent molecular mass

^a Predicted data from the amino acid sequence of rThBgl using the Sednterp server (<http://sednterp.unh.edu/>)

^b From the ratio of $R_s:R_0$

^c Obtained from a SedFit analysis

^d Secondary structure generated by deconvolution of the experimental CD spectra using the CDNN Deconvolution program



The results of the analytical SEC experiments revealed that rThBgl was eluted as a unique peak with a retention time that corresponded to an apparent molecular mass (MM_{app}) of 54.72 ± 3 kDa (Fig. 2a). Based on the analytical SEC elution profile and standard proteins with known Stokes radii (see “Methods” section), the Stokes radius (R_s) and frictional ratio (f/f_0) of rThBgl were estimated to be 31.39 Å and 1.25 ± 0.2 , respectively (Table 1, Fig. 2a). The results from the AUC experiments corroborated the analytical SEC data; during sedimentation, rThBgl appeared as a single species with an $S_{20,w}^0$ and experimental molecular mass (MM_{exp}) of 4.53 ± 0.03 S and 58.26 ± 7 kDa, respectively. The f/f_0 value was

1.23 ± 0.05 (Table 1, Fig. 2b). The analytical SEC and AUC data suggest that purified rThBgl behaves as a monodisperse monomer in solution. A similar finding was previously reported for other β -glucosidases from *Trichoderma* spp. [25, 30]. In addition, based on a f/f_0 of 1.2, the protein is expected to have a globular shape [39].

The secondary structure composition and tertiary folding of rThBgl were assessed by circular dichroism (CD) and fluorescence spectroscopy, respectively. rThBgl had a typical CD spectrum for an α/β -folded protein (Fig. 2c, inset) as previously described for the GH1 β -glucosidase family [40, 41]. The α -helix and β -sheet contents estimated from the deconvolution of the rThBgl spectrum

were 35 ± 4 and 10 ± 2 %, respectively (Table 1), which are similar to those observed in the crystallographic structure. We also used the CD technique to evaluate the thermal-induced unfolding properties of rThBgl. These data revealed that rThBgl underwent conformational changes at approximately 35 °C and was completely unfolded at ~61 °C, resulting in a melting temperature (T_m) of 49 ± 1 °C (Table 1, Fig. 2c), which is consistent with the effect of temperature on the enzymatic activity. The results of thermal-induced unfolding were similar to those obtained by differential scanning calorimetry (data not shown) and confirmed the thermal stability properties of rThBgl.

Using the intrinsic fluorescence characteristics of tryptophan residues, the local tertiary structure of rThBgl was investigated in its folded and denatured states. A clear difference in the fluorescence emission spectra between the folded and denatured forms was observed (Fig. 2d). The maximum emission wavelength (λ_{max}) and spectral center of mass ($\langle \lambda \rangle$) for the folded rThBgl were 334 ± 1 and 353 ± 1 nm, respectively, while the denatured form of rThBgl exhibited a λ_{max} and ($\langle \lambda \rangle$) of 354 ± 2 nm and 363 ± 1 nm, respectively (Table 1). rThBgl has 12 tryptophan residues distributed along the protein amino acid sequence; therefore, the changes observed with fluorescence spectroscopy indicate that the recombinant protein was produced in a folded state. The use of a denaturing agent, such as GdnHCl, exposed some of these tryptophan residues to the solvent, thus disrupting the local tertiary structure.

The kinetic and biochemical properties of rThBgl

The biochemical characterization of the purified rThBgl was performed using the general artificial substrate for β -glucosidase activity, 4-nitrophenyl β -D-glucopyranoside (*p*NPG). For all enzymatic assays, the N-terminal 6 \times His-tag was removed using thrombin to prevent any interference with the enzymatic activity. We first investigated the optimal temperature and pH dependence of the enzymatic activity. Maximal rThBgl activity was observed at 40 °C; however, considerable enzymatic activity was observed in assays performed at temperatures above 50 °C (Fig. 3a). The optimum temperature for β -glucosidase activity has been reported as 25–30 °C for cold-adapted enzymes [41, 42] and 90 °C for thermostable enzymes derived from a metagenomic library of the termite gut [43]. The pH-dependent enzyme activity showed that the rThBgl retains its relative highest activity (>60 %) between pH 5.0 and 7.0, with a catalytic optimum at 6.0 (Fig. 3b). Similar findings were reported for other characterized β -glucosidases [44–46]. Knowledge of the physicochemical characteristics of cellulytic enzymes is an important step for the development of

commercial cocktails designed to improve the enzymatic hydrolysis of lignocellulosic compounds [37, 47].

Kinetic parameters obtained with *p*NPG and cellobiose as the substrates under the optimized enzymatic conditions (pH 6.0 and 40 °C) revealed that rThBgl demonstrated typical Michaelis–Menten behavior with a half-saturation constant (K_m) and maximum velocity (V_{max}) values of 0.97 mM and 29.3 ± 0.5 U mg_{enz}^{-1} , respectively, for *p*NPG (Fig. 3c) and 1.22 mM and 10.4 ± 0.6 U mg_{enz}^{-1} , respectively, for cellobiose (Fig. 3d). A high affinity for *p*NPG is a common characteristic of many β -glucosidases, particularly the aryl- β -glucosidases [3]. The effect of glucose on rThBgl was also investigated (Fig. 4). Although low concentrations of glucose (25–50 mM) had a positive effect and improved the enzymatic activity of rThBgl, a decline in activity was observed at concentrations exceeding 50 mM glucose, and 50 % of the enzymatic activity was observed at 300 mM glucose (Fig. 4). These results indicate that rThBgl is tolerant of product inhibition, but its tolerance is lower compared to other highly glucose-tolerant enzymes, such as HiBG [48]. In sum, its high expression in *T. harzianum* under biomass degradation conditions and its enzymatic properties indicate that rThBgl could be used for supplementation of commercial cocktails with glucose-dependent activity, such as Celluclast, Novozyme N188, and Cellic® CTec2 [41].

Overall structure of rThBgl

Crystallographic refinement at a resolution of 2.6 Å converged to an R_{factor} and R_{free} of 0.18 and 0.21, respectively, and resulted in excellent stereochemistry according to Ramachandran and RMSD analyses (Additional file 4: Table S4). rThBgl crystals belonged to the enantiomorphic space group $P6_1$ with a dimer in the asymmetric unit and a high solvent content (77.5 %). The two molecules in the asymmetric unit were very similar, with an RMSD for the C α atoms of 0.15 Å, and each chain comprising the residues Met¹ to Ala⁴⁶³.

rThBgl has the classical (α/β)₈-barrel fold observed in other structurally characterized GH1 β -glucosidases, with the active-site pocket located at the C-terminal region of the barrel (Fig. 5a). rThBgl shares a 90 % sequence identity with TrBgl2 (PDB code 3AHY), and their structural alignment resulted in an RMSD of 0.26 Å with the main differences in the flexible regions, including the N- and C-termini. The active-site architecture was fully conserved, and the high structural similarity between rThBgl and TrBgl2 [49] was reflected in the surface charge distribution, with a canonical negatively charged active-site pocket (Fig. 5b). By comparison, the Glu³⁶⁶ residue is the nucleophile, and the Glu¹⁶⁵ residue is the acid–base [50]. Based on the superposition

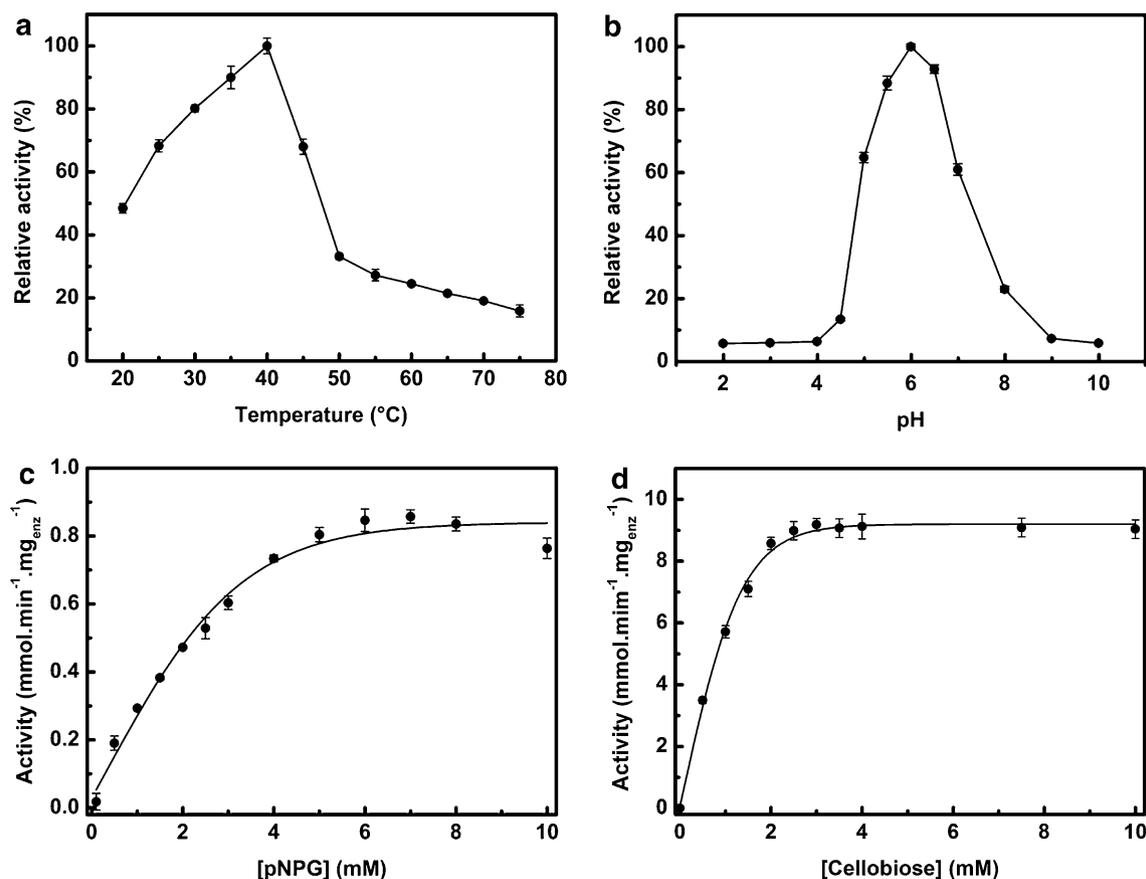


Fig. 3 Biochemical properties of purified rThBgl. **a** Optimal temperature screening for rThBgl activity. **b** pH dependence of rThBgl activity in a 100 mM citrate/phosphate/glycine buffer with a pH range from 2 to 10. **c** The kinetic curves of pNPG (0–10 mM). **d** The kinetic curves of cellobiose (0–10 mM). For kinetic analysis, the reaction mixture contained 100 mM sodium phosphate buffer pH 6.0, and all reactions were incubated for 10 min at 40 °C

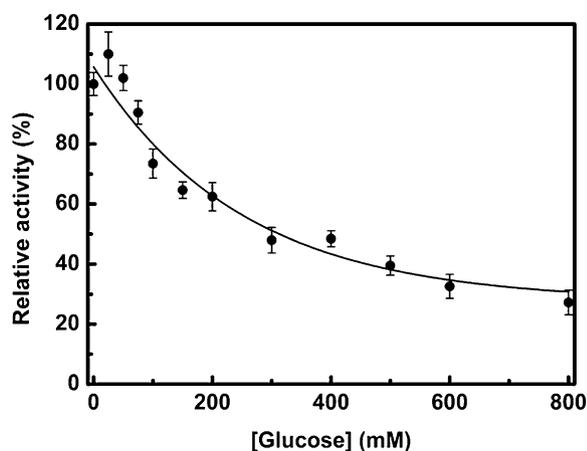
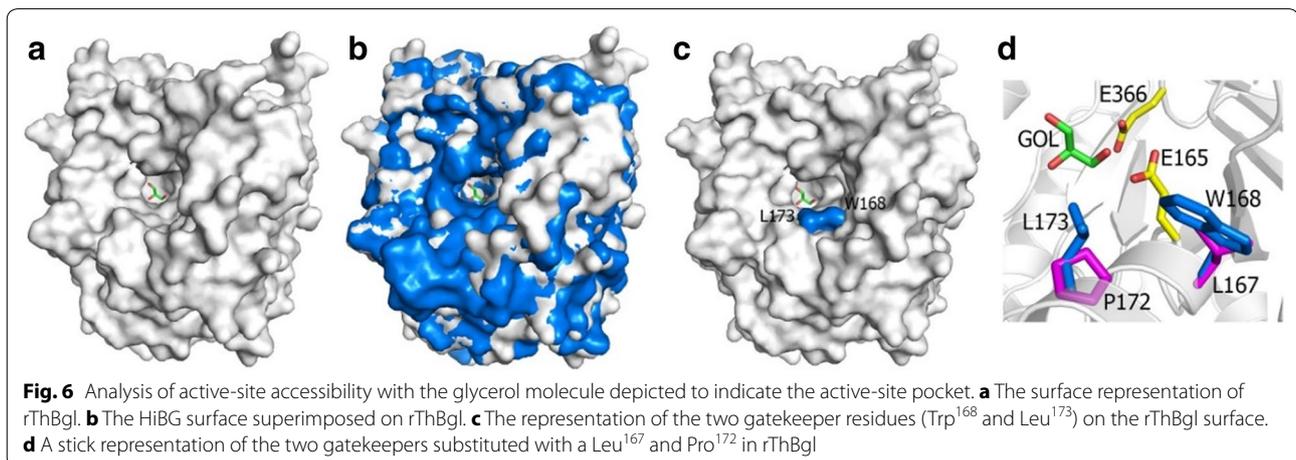
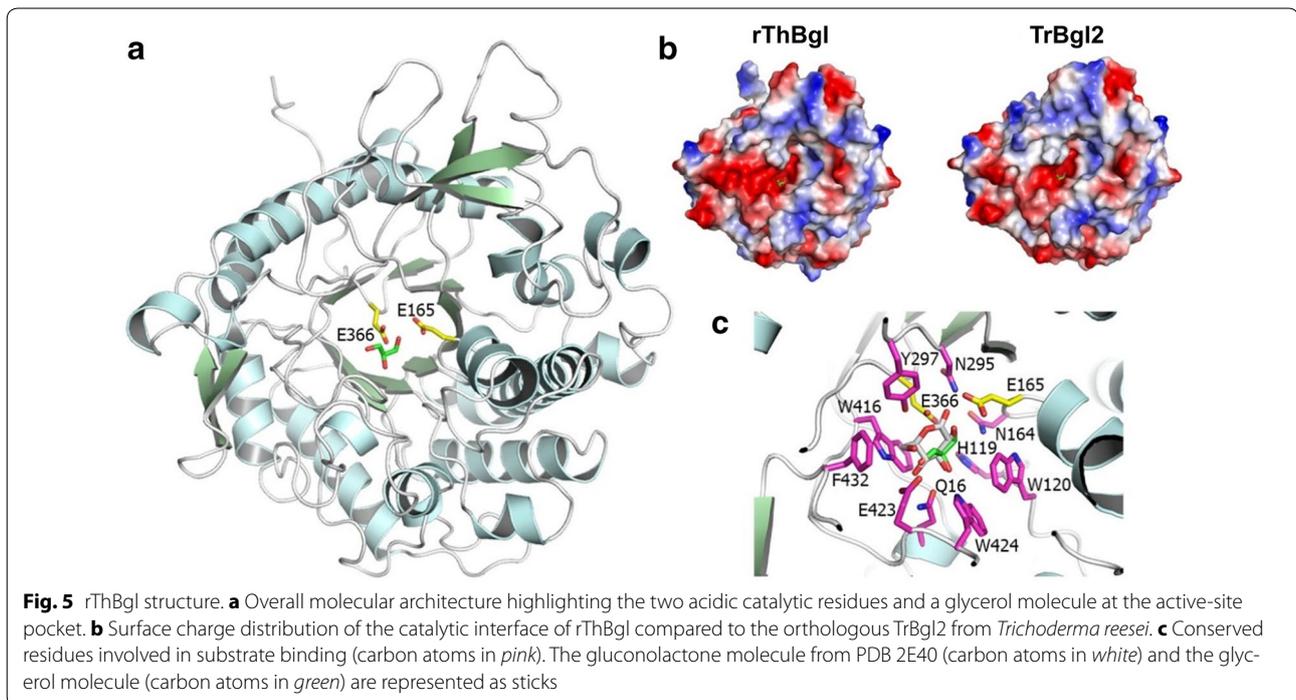


Fig. 4 The effect of glucose on rThBgl activity. The glucose tolerance of rThBgl was investigated using the optimal temperature (40 °C) and pH (pH 6.0) for rThBgl activity with glucose concentrations ranging from 0 to 800 mM. The experiments were performed in triplicate

with a β -glucosidase from *Phanerochaete chrysosporium* (BGL1A) in a complex with gluconolactone (PDB code 2E40; [51]), other residues involved in substrate binding were also conserved, including Gln¹⁶, His¹¹⁹, Trp¹²⁰, Asn¹⁶⁴, Asn²⁹⁵, Tyr²⁹⁷, Trp⁴¹⁶, Glu⁴²³, Trp⁴²⁴, and Phe⁴³² (Fig. 5c). The rThBgl structure also contained a glycerol molecule that mimicked a carbohydrate moiety bound to the active site and interacted with several of the residues considered essential for substrate recognition (Fig. 5c).

In comparison with the structure of HiBG, a highly glucose-tolerant GH1 β -glucosidase from *Humicola insolens*, rThBgl has a broader active-site entrance, which may explain the lower tolerance of this enzyme to glucose inhibition compared to HiBG (Fig. 6a–c). According to de Giuseppe et al. [48], restricted access to the active-site pocket is associated with the high glucose tolerance of some GH1 β -glucosidases, such as HiBG [48]. Moreover, the two HiBG residues Trp¹⁶⁸ and Leu¹⁷³ are considered to be gatekeepers involved in glucose tolerance. These



two residues were not conserved in rThBgl and were replaced by Leu¹⁶⁷ and Pro¹⁷², respectively (Fig. 6c–d).

Conclusions

In the present study, we used RNA-Seq data mining to identify the β -glucosidases highly expressed by *T. harzianum* under biomass degradation conditions. We used a rational approach to investigate the biochemical characteristics of the β -glucosidase and to obtain a detailed three-dimensional structure. Although high mRNA

expression is not necessarily coupled to a high level of the translated protein, up-regulated genes involved in cellulose metabolism suggest that these gene products effectively participate in metabolic pathways. Thus, these enzymes are excellent targets for further investigation into novel biotechnological applications. The strategy described in this study provides a model for the bio-prospection of key enzymes for the development or improvement of enzymatic cocktails designed for ethanol production from plant biomass.

Methods

Data mining using RNA-Seq libraries

The RNA-Seq libraries from *T. harzianum* IOC-3844 cultured in the presence of LAC, CEL, or DSB [18] were used to identify the major β -glucosidase used by *T. harzianum* under biomass degradation conditions. These data can be found in the NCBI's Sequence Read Archive (SRA) under accession numbers SRR579379, SRR631745, and SRR631746 for the DSB, CEL, and LAC libraries, respectively.

The data mining was conducted using the CLC Genomics Workbench (v4.0; Finlandsgade, Dk). The genes encoding the GH1 and GH3 β -glucosidases (Additional file 1: Table S1) used for performing the read mapping were identified in the assembled genome of the *T. harzianum* T6776 strain (GenBank access number JOKZ00000000.1; [24]). For read mapping, the parameters were set to a similarity = 0.8, length fraction = 0.9, and maximum number of hits for a read = 10.

To compare the relative gene expression between the different β -glucosidase sequences under three different conditions (LAC, CEL, and DSB), we \log_2 -transformed the normalized reads per million mapped values. The squalene epoxidase (GenBank: AM050097.1), TATA box-binding protein (GenBank code: KKP02269.1), and 13-kDa ribonucleoprotein (RNP)-associated protein (GenBank: KKO98059.1) sequences were used for the normalization of gene expression.

DNA manipulation

The DNA amplification and recombinant plasmid construction were carried out using standard molecular biology procedures [52].

The *bgl* gene (1398 bp) encoding the GH1 β -glucosidase protein was amplified from *T. harzianum* IOC-3844 cDNA using PCR. Following RNA extraction, the cDNA was synthesized using a SuperScript II kit (Invitrogen, USA) according to the manufacturer's instructions. The primers 5'-TATCATATGTTGCCCAAGGACTTT-3' and 5'-TATGAATTCTTACTCCTTCGCAATC-3' contained *Nde*I and *Eco*RI restriction sites (underlined), respectively, and were designed using the sequence information available in GenBank (access number KU201604.1). The PCR amplification product was cloned into a pET28a(+) (Novagen, Madison, WI, USA) expression vector, which added an N-terminal six-histidine tag and a thrombin protease site to the coding sequence. Nucleotide substitutions in the constructed plasmid were analyzed with DNA sequencing.

Heterologous protein expression and purification

rThBgl was expressed in the *E. coli* Rosetta strain (Novagen, Darmstadt, Germany). The cells were cultured at

37 °C with shaking at 300 rpm in 1 L of LB broth containing chloramphenicol (34 $\mu\text{g mL}^{-1}$) and kanamycin (30 $\mu\text{g mL}^{-1}$) until an OD_{600} of 0.8 was reached. Recombinant protein expression was induced by 0.4 mM IPTG, followed by cultivation for 20 h at 16 °C and 180 rpm. The culture was then centrifuged (3000g, 15 min, 4 °C), and the cells were resuspended in 25 mL of buffer A (40 mM HEPES pH 7.5 and 150 mM NaCl) containing 1 mg mL^{-1} lysozyme, 1 mM PMSF (phenylmethanesulfonyl fluoride), and 0.1 % (v/v) Tween 20. The cells were disrupted by sonication, and the soluble fraction was collected by centrifugation (20,000g, 40 min, 4 °C). The purification of rThBgl was completed using nickel affinity chromatography with a prepacked Ni Sepharose High Performance HisTrap column (GE Life Sciences) previously equilibrated with buffer A. A polishing purification step was performed using gel filtration with a HiPrep 16/60 Sephacryl S-100 HR column (GE Life Sciences) previously equilibrated with buffer A. All chromatographic steps were carried out with columns coupled to an ÄKTA FPLC device (GE Life Sciences), and the protein elution profiles were monitored at an absorbance of 280 nm. The His₆-tag of rThBgl was cleaved by treating 10 mg of the target protein with 1 U of thrombin (Novagen, Darmstadt, Germany) for 4 h at 25 °C. The concentrations of the purified proteins were determined spectroscopically using the molar extinction coefficient (ϵ) predicted by the amino acid sequence. The purity of the rThBgl protein was estimated with SDS-PAGE.

Hydrodynamic and spectroscopic studies

Analytical SEC was performed using a Superdex 200 10/300 GL prepacked column (GE Healthcare, Pittsburgh, PA, USA). Approximately 9 μM protein in buffer A was loaded onto the column at a flow rate of 0.5 mL min^{-1} , and the elution profile was monitored at an absorbance of 280 nm. The R_s of the purified rThBgl protein was estimated using a mix of protein standards with known R_s values, including carbonic anhydrase (MW = 29 kDa, 23.9 Å), ovalbumin (MW = 44 kDa, 30.5 Å), conalbumin (MW = 75 kDa, 36.4 Å), aldolase (MW = 158 kDa, 48.1 Å), and ferritin (MW = 440 kDa, 61 Å). All of the protein standards (GE Healthcare) were prepared and analyzed under the same conditions used for rThBgl. The Blue Dextran 2000 (GE Healthcare) polymer was used to determine the void volume of the column. The analytical SEC data were calculated according to the manufacturer's instructions. The estimated R_s values obtained from the analytical SEC data were used to estimate the f/f_0 as the ratio of the experimental R_s to the predicted radius of a sphere (R_0) of the same molecular mass.

Sedimentation velocity experiments with the rThBgl protein were performed using a Beckman Optima XL-A analytical ultracentrifuge. The data acquisition during the AUC was performed at 280 nm, 20 °C, and 35,000 rpm using an AN-50Ti rotor with a protein sample ranging from 100 to 800 $\mu\text{g mL}^{-1}$ in buffer A. The buffer viscosity ($\eta = 1.0513 \times 10^{-2}$ poise), buffer density ($\rho = 1.0163 \text{ g mL}^{-1}$), and partial-specific rThBgl volume ($V_{\text{bar: rThBgl}} = 0.733194 \text{ mL g}^{-1}$) were estimated using the Sednterp server (<http://sednterp.unh.edu/>). The R_s , MM_{exp} , $S_{20,w}^0$, S , and f/f_0 were obtained from the AUC data using SedFit software (Version 12.1).

The rThBgl secondary structure was analyzed via CD with a Jasco model J-815 CD spectropolarimeter (Japan Spectroscopic; Tokyo, Japan) coupled to a thermoelectric sample temperature controller (Peltier Type Control System PFD 425S-Jasco) to record the CD spectra. The far-UV CD spectra were generated using the rThBgl protein at a concentration of approximately 4 μM in buffer A at 25 °C. The assays were performed using a quartz cuvette with a path length of 1 mm. A total of 18 determinations within the range of 260–200 nm at a rate of 20 nm min^{-1} were recorded and averaged. The statistical analysis of the CD spectra was performed using CDNN deconvolution software. The rThBgl thermal-induced unfolding experiments followed by CD were measured at 220 nm from 20 to 90 °C with a 1-mm-path length cell using approximately 8 μM rThBgl in buffer A.

For the fluorescence spectroscopy analyses, a Varian Cary Eclipse fluorescence spectrophotometer (Agilent Technologies; Santa Clara, USA) was used. rThBgl samples (2.5 μM) containing 0 or 6 M GdnHCl were prepared in buffer A. Using a 10 \times 2-mm-path-length cell, the samples were excited at 280 nm, and the intrinsic fluorescence emission spectra were collected from 300 to 500 nm. The λ_{max} and $\langle\lambda\rangle$ were calculated from the intrinsic fluorescence emission data.

Biochemical characterization

The β -glucosidase activity of the purified rThBgl was initially evaluated using *p*NPG (Sigma-Aldrich) as the substrate. The initial experiments were performed in triplicate with 100 μL reactions containing 25 nM purified enzyme, 100 mM sodium phosphate buffer, pH 7.0, and 0.5 mM *p*NPG. All reactions were incubated for 10 min and stopped with the addition of 100 μL of 1 M Na_2CO_3 . The *p*-nitrophenol released during the reaction was measured at 405 nm using an Infinite[®] 200 PRO microplate reader (TECAN). One unit of enzyme activity was defined as 1 μM of *p*-nitrophenol released per minute.

The optimal temperature was evaluated in assays ranging from 20 to 75 °C. The pH dependence of the enzymatic activity was determined in a pH range from 2.0 to 10.0 using the following buffers: citrate–phosphate (pH 2, 3, 4, 4.5, 5, and 5.5), phosphate (pH 6, 6.5, 7, and 8), and glycine (pH 9 and 10) at a final concentration of 100 mM.

Subsequent to the determination of the optimal temperature and pH, kinetic experiments were performed in 100 mM sodium phosphate buffer (pH 6.0) at 40 °C using the rate of hydrolysis of *p*NPG and cellobiose at various concentrations ranging from 0 to 10 mM. The kinetic parameters (K_m and V_{max}) were obtained using GraphPad Prism (GraphPad Software, San Diego, CA, USA) to adjust for the non-linear fit of the Michaelis–Menten equation.

Activity inhibition by glucose was investigated with glucose concentrations ranging from 0 to 800 mM.

Crystallization, data collection, structure determination, refinement, and validation

The protein was concentrated to 10 mg mL^{-1} in 50 mM phosphate buffer (pH 7.0) for the crystallization experiments. Sitting drops were prepared at 18 °C using a Cartesian HoneyBee 963 system (Genomic Solutions), and 544 conditions were screened based on the commercially available crystallization kits from Hampton Research (SaltRx, Crystal Screen I and II), Emerald BioSystems (Precipitant Synergy and Wizard I and II), and Qiagen/Nextal (PACT and JCSG+). Suitable crystals for X-ray diffraction experiments were obtained with 2.0 M ammonium sulfate, 0.1 M sodium acetate, pH 5.5, and 2 % (v/v) PEG400. The diffraction data were acquired using the W01B-MX2 beamline (LNLS, Campinas, Brazil). A single crystal was soaked in the aforementioned crystallization conditions with 30 % (v/v) glycerol as a cryoprotectant and then directly flash-cooled in a nitrogen gas stream at 100 K. The sample-to-detector distance was set to a maximum resolution of 2.5 Å, and 180° were collected using the fine-slicing method (0.1° per image) and a Pilatus 2 M detector (Dectris). The data were indexed, integrated, and scaled using the XDS package [53, 54]. Molecular replacement calculations were performed using the program MOLREP [55], and the structure of β -glucosidase 2 from the fungus *T. reesei* (TrBgl2, PDB code 3AHY; [49]) was used as a template. Restrained refinement was completed with the phenix.refine program from the PHENIX package [56], and manual inspection and building was performed with COOT [57]. The model quality was assessed using MOLPROBITY [58], and the refinement statistics are presented in Additional file 4: Table S4.

The atomic coordinates and structure factors have been added to the PDB under the accession code 5BWF.

Additional files

Additional file 1: Table S1. GH1 and GH3 β -glucosidases sequences identified on *T. harzianum* T6776 genome (GenBank access number JOKZ00000000.1) and used for mapping experiments via RNA-Seq data.

Additional file 2: Table S2. Percent identity matrix between the GH1 β -glucosidases amino acid sequences. The multiple sequence alignment was performed using the Clustal Omega server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

Additional file 3: Table S3. Percent identity matrix between the GH3 β -glucosidase. The multiple sequence alignment was performed using the Clustal Omega server (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

Additional file 4: Table S4. Data collection and refinement statistics.

Abbreviations

rThBgl: recombinant β -glucosidase of *T. harzianum*; GH: glycosyl hydrolase; PDB: protein data bank; RNA-Seq: RNA sequencing; RPKM: reads per kilo base per million of mapped reads; LAC: lactose; CEL: crystalline cellulose; DSB: delignified sugarcane bagasse; SEC: size-exclusion chromatography; AUC: analytical ultracentrifugation; MM_{prep} : predicted molecular mass; MM_{app} : apparent molecular mass; MM_{exp} : experimental molecular mass; R_0 : predicted radius of a sphere; R_s : Stokes radius; f/f_0 : frictional ratio; CD: circular dichroism; λ_{max} : maximum emission wavelength; $\langle \lambda \rangle$: spectral center of mass; ρ NPG: 4-nitrophenyl β -D-glucopyranoside; K_m : half-saturation constant; V_{max} : maximum velocity.

Authors' contributions

CAS carried out all of the experiments and drafted the manuscript. AC and MACH assisted with the gene cloning and protein expression. LMZ and RR contributed to the biochemical characterization. LMZ, CCCT, and MTM designed and performed the crystallographic studies. APS directed the overall study and drafted the manuscript. All the authors read and approved the final manuscript.

Author details

¹ Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, SP, Brazil. ² Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil. ³ Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil. ⁴ Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Acknowledgements

This study was supported by grants from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2008/52197-4, 2013/13309-0, and 2014/09720-9), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; MCTI/CNPq/Universal, grant no. 457360/2014-3). CAS received a post-doctoral fellowship from CNPq, MACH received a post-doctoral fellowship from FAPESP (2014/18856-1), AC received a PhD fellowship from CNPq and APS and MTM received research fellowships from the CNPq. We thank the Biosciences National Laboratory (LNBio), the Brazilian Synchrotron Light Laboratory (LNLS), and the Bioethanol Science and Technology Laboratory (CTBE) for the allotted time in the ROBOLAB, MX2 beamline, and LAM facilities, respectively.

Competing interests

The authors declare that they have no competing interests.

Received: 26 December 2015 Accepted: 14 March 2016

Published online: 22 March 2016

References

- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002;66(3):506–77. doi:10.1128/mmb.66.3.506-577.2002.
- Singhania RR, Patel AK, Sukumaran RK, Larroche C, Pandey A. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour Technol.* 2013;127:500–7. doi:10.1016/j.biortech.2012.09.012.
- Tiwari P, Misra BN, Sangwan NS. β -Glucosidases from the fungus *Trichoderma*: an efficient cellulase machinery in biotechnological applications. *Biomed Res Int.* 2013;2013:10. doi:10.1155/2013/203735.
- Holtzapfel M, Cognata M, Shu Y, Hendrickson C. Inhibition of *Trichoderma reesei* cellulase by sugars and solvents. *Biotechnol Bioeng.* 1990;36(3):275–87. doi:10.1002/bit.260360310.
- Medve J, Karlsson J, Lee D, Tjerneld F. Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnol Bioeng.* 1998;59(5):621–34. doi:10.1002/(SICI)1097-0290(19980905)59:5<621::AID-BT13>3.0.CO;2-C.
- Xiao Z, Zhang X, Gregg DJ, Saddler JN. Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. *Appl Biochem Biotechnol.* 2004;113–116:1115–26. doi:10.1385/ABAB:115:1-3:1115.
- Teugjas H, Valjamae P. Selecting beta-glucosidases to support cellulases in cellulose saccharification. *Biotechnol Biofuels.* 2013;6(1):105. doi:10.1186/1754-6834-6-105.
- Dashtban M, Maki M, Leung KT, Mao C, Qin W. Cellulase activities in biomass conversion: measurement methods and comparison. *Crit Rev Biotechnol.* 2010;30(4):302–9. doi:10.3109/07388551.2010.490938.
- Zhang M, Su R, Qi W, He Z. Enhanced enzymatic hydrolysis of lignocellulose by optimizing enzyme complexes. *Appl Biochem Biotechnol.* 2010;160(5):1407–14. doi:10.1007/s12010-009-8602-3.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 2009;37:D233–8. doi:10.1093/nar/gkn663.
- Vocadlo DJ, Davies GJ. Mechanistic insights into glycosidase chemistry. *Curr Opin Chem Biol.* 2008;12(5):539–55. doi:10.1016/j.cbpa.2008.05.010.
- Cairns KJ, Esen A. β -Glucosidases. *Cell Mol Life Sci.* 2010;67(20):3389–405. doi:10.1007/s00018-010-0399-2.
- Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. *Adv Biochem Eng Biotechnol.* 2007;108:95–120. doi:10.1007/10_2007_066.
- Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science.* 2007;315(5813):804–7. doi:10.1126/science.1137016.
- Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol.* 2008;26(5):553–60. doi:10.1038/nbt1403.
- Häkkinen M, Valkonen MJ, Westerholm-Parvinen A, Aro N, Arvas M, Vitikainen M, et al. Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production. *Biotechnol Biofuels.* 2014;7(1):1–21. doi:10.1186/1754-6834-7-14.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57–63. doi:10.1038/nrg2484.
- Horta MAC, Vicentini R, Delabona PdS, Laborda P, Crucello A, Freitas S, et al. Transcriptome profile of *Trichoderma harzianum* IOC-3844 induced by sugarcane bagasse. *PLoS ONE.* 2014;9(2):e88689. doi:10.1371/journal.pone.0088689.
- Wei H, Fu Y, Magnusson L, Baker JO, Maness P-C, Xu Q, et al. Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. *Front Microbiol.* 2014;5:142. doi:10.3389/fmicb.2014.00142.
- Coradetti ST, Xiong Y, Glass NL. Analysis of a conserved cellulase transcriptional regulator reveals inducer-independent production of cellulolytic enzymes in *Neurospora crassa*. *Microbiol Open.* 2013;2(4):595–609. doi:10.1002/mbo3.94.
- Ries L, Pullan ST, Delmas S, Malla S, Blythe MJ, Archer DB. Genome-wide transcriptional response of *Trichoderma reesei* to lignocellulose using

- RNA sequencing and comparison with *Aspergillus niger*. *BMC Genom.* 2013;14:541. doi:10.1186/1471-2164-14-541.
22. Bischof R, Fourtis L, Limbeck A, Gamauf C, Seiboth B, Kubicek CP. Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose. *Biotechnol Biofuels.* 2013;6(1):1. doi:10.1186/1754-6834-6-127.
 23. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5:621–8. doi:10.1038/nmeth.1226.
 24. Baroncelli R, Piaggieschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, et al. Draft whole-genome sequence of the biocontrol agent *Trichoderma harzianum* T6776. *Genome Announc.* 2015;3(3):e00647. doi:10.1128/genomeA.00647-15.
 25. Chandra M, Kalra A, Sangwan N, Sangwan R. Biochemical and proteomic characterization of a novel extracellular beta-glucosidase from *Trichoderma citrinoviride*. *Mol Biotechnol.* 2013;53(3):289–99. doi:10.1007/s12033-012-9526-7.
 26. Dashtban M, Qin W. Overexpression of an exotic thermotolerant beta-glucosidase in *Trichoderma reesei* and its significant increase in cellulolytic activity and saccharification of barley straw. *Microb Cell Fact.* 2012;11:63. doi:10.1186/1475-2859-11-63.
 27. Chen P, Fu X, Ng TB, Ye XY. Expression of a secretory beta-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. *Biotechnol Lett.* 2011;33(12):2475–9. doi:10.1007/s10529-011-0724-3.
 28. Chauve M, Mathis H, Huc D, Casanave D, Monot F, Lopes Ferreira N. Comparative kinetic analysis of two fungal β -glucosidases. *Biotechnol Biofuels.* 2010;3:3. doi:10.1186/1754-6834-3-3.
 29. Murray P, Aro N, Collins C, Grassick A, Penttilä M, Saloheimo M, et al. Expression in *Trichoderma reesei* and characterisation of a thermostable family 3 beta-glucosidase from the moderately thermophilic fungus *Talaromyces emersonii*. *Protein Expr Purif.* 2004;38(2):248–57. doi:10.1016/j.pep.2004.08.006.
 30. Saloheimo M, Kuja-Panula J, Ylosmaki E, Ward M, Penttilä M. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* beta-glucosidase BGLII (cel1A). *Appl Environ Microbiol.* 2002;68(9):4546–53. doi:10.1128/AEM.68.9.4546-4553.2002.
 31. Yun SI, Jeong CS, Chung DK, Choi HS. Purification and some properties of a beta-glucosidase from *Trichoderma harzianum* type C-4. *Biosci Biotechnol Biochem.* 2001;65(9):2028–32. doi:10.1271/bbb.65.2028.
 32. Takashima S, Nakamura A, Hidaka M, Masaki H, Uozumi T. Molecular cloning and expression of the novel fungal beta-glucosidase genes from *Humicola grisea* and *Trichoderma reesei*. *J Biochem.* 1999;125(4):728–36.
 33. Chirico WJ, Brown RD Jr. Purification and characterization of a beta-glucosidase from *Trichoderma reesei*. *Eur J Biochem.* 1987;165(2):333–41.
 34. Sternberg D, Vijayakumar P, Reese ET. Beta-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. *Can J Microbiol.* 1977;23(2):139–47. doi:10.1111/j.1432-1033.1987.tb11446.x.
 35. Gray KA, Zhao L, Emptage M. *Bioethanol.* Curr Opin Chem Biol. 2006;10(2):141–6. doi:10.1016/j.cbpa.2006.02.035.
 36. Stephanopoulos G. Challenges in engineering microbes for biofuels production. *Science.* 2007;315(5813):801–4. doi:10.1126/science.1139612.
 37. Brethauer S, Wyman CE. Review: continuous hydrolysis and fermentation for cellulosic ethanol production. *Bioresour Technology.* 2010;101(13):4862–74. doi:10.1016/j.biortech.2009.11.009.
 38. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 2014;5:172. doi:10.3389/fmicb.2014.00172.
 39. Erickson HP. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online.* 2009;11:32–51. doi:10.1007/s12575-009-9008-x.
 40. Cota J, Corrêa TLR, Damásio ARL, Diogo JA, Hoffmam ZB, Garcia W, et al. Comparative analysis of three hyperthermophilic GH1 and GH3 family members with industrial potential. *N Biotechnol.* 2015;32(1):13–20. doi:10.1016/j.nbt.2014.07.009.
 41. Crespim E, Zanphorlin LM, de Souza FHM, Diogo JA, Gazolla AC, Machado CB, et al. A novel cold-adapted and glucose-tolerant GH1 β -glucosidase from *Exiguobacterium antarcticum* B7. *Int J Biol Macromol.* 2016;82:375–80. doi:10.1016/j.ijbiomac.2015.09.018.
 42. Fan H-X, Miao L-L, Liu Y, Liu H-C, Liu Z-P. Gene cloning and characterization of a cold-adapted β -glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium *Micrococcus antarcticus*. *Enzyme Microb Technol.* 2011;49(1):94–9. doi:10.1016/j.enzmictec.2011.03.001.
 43. Wang Q, Qian C, Zhang XZ, Liu N, Yan X, Zhou Z. Characterization of a novel thermostable beta-glucosidase from a metagenomic library of termite gut. *Enzyme Microb Technol.* 2012;51(6–7):319–24. doi:10.1016/j.enzmictec.2012.07.015.
 44. Han YW, Srinivasan VR. Purification and characterization of β -glucosidase of *Alcaligenes faecalis*. *J Bacteriol.* 1969;100(3):1355–63.
 45. Lin J, Pillay B, Singh S. Purification and biochemical characteristics of beta-D-glucosidase from a thermophilic fungus, *Thermomyces lanuginosus*-SSBP. *Biotechnol Appl Biochem.* 1999;30(Pt 1):81–7. doi:10.1111/j.1470-8744.1999.tb01163.x.
 46. Liu D, Zhang R, Yang X, Zhang Z, Song S, Miao Y, et al. Characterization of a thermostable beta-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. *Microb Cell Fact.* 2012;11:25. doi:10.1186/1475-2859-11-25.
 47. Yang B, Dai Z, Ding S-Y, Wyman CE. Enzymatic hydrolysis of cellulosic biomass. *Biofuels.* 2011;2(4):421–49. doi:10.4155/bfs.1.11.116.
 48. de Giuseppe PO, Souza Tde A, Souza FH, Zanphorlin LM, Machado CB, Ward RJ, et al. Structural basis for glucose tolerance in GH1 beta-glucosidases. *Acta Crystallogr D Biol Crystallogr.* 2014;70(Pt 6):1631–9. doi:10.1107/s1399004714006920.
 49. Jeng WY, Wang NC, Lin MH, Lin CT, Liaw YC, Chang WJ, et al. Structural and functional analysis of three β -glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J Struct Biol.* 2011;173(1):46–56. doi:10.1016/j.jsb.2010.07.008.
 50. Withers SG, Warren RAJ, Street IP, Rupitz K, Kempton JB, Aebersold R. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a retaining glycosidase. *J Am Chem Soc.* 1990;112(15):5887–9. doi:10.1021/ja00171a043.
 51. Nijikken Y, Tsukada T, Igarashi K, Samejima M, Wakagi T, Shoun H, et al. Crystal structure of intracellular family 1 beta-glucosidase BGL1A from the basidiomycete *Phanerochaete chrysosporium*. *FEBS Lett.* 2007;581(7):1514–20. doi:10.1016/j.febslet.2007.03.009.
 52. Sambrook J, Maniatis T, Fritsch EF. *Molecular cloning: a laboratory manual.* 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989.
 53. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):125–32. doi:10.1107/s0907444909047337.
 54. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):133–44. doi:10.1107/s0907444909047374.
 55. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr.* 1997;30(6):1022–5. doi:10.1107/S0021889897006766.
 56. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 2):213–21. doi:10.1107/s0907444909052925.
 57. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 4):486–501. doi:10.1107/s0907444910007493.
 58. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* 2004;32:W615–9. doi:10.1093/nar/gkh398.