

# SCIENTIFIC REPORTS



OPEN

## An engineered GH1 $\beta$ -glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials

Clelton A. Santos<sup>1,2</sup>, Mariana A. B. Morais<sup>3</sup>, Oliver M. Terrett<sup>2</sup>, Jan J. Lyczakowski<sup>2,4,5</sup>, Leticia M. Zanphorlin<sup>3</sup>, Jaire A. Ferreira-Filho<sup>1</sup>, Celisa C. C. Tonoli<sup>6</sup>, Mario T. Murakami<sup>3</sup>, Paul Dupree<sup>2,4,5</sup> & Anete P. Souza<sup>1,7</sup>

$\beta$ -glucosidases play a critical role among the enzymes in enzymatic cocktails designed for plant biomass deconstruction. By catalysing the breakdown of  $\beta$ -1, 4-glycosidic linkages,  $\beta$ -glucosidases produce free fermentable glucose and alleviate the inhibition of other cellulases by cellobiose during saccharification. Despite this benefit, most characterised fungal  $\beta$ -glucosidases show weak activity at high glucose concentrations, limiting enzymatic hydrolysis of plant biomass in industrial settings. In this study, structural analyses combined with site-directed mutagenesis efficiently improved the functional properties of a GH1  $\beta$ -glucosidase highly expressed by *Trichoderma harzianum* (ThBgl) under biomass degradation conditions. The tailored enzyme displayed high glucose tolerance levels, confirming that glucose tolerance can be achieved by the substitution of two amino acids that act as gatekeepers, changing active-site accessibility and preventing product inhibition. Furthermore, the enhanced efficiency of the engineered enzyme in terms of the amount of glucose released and ethanol yield was confirmed by saccharification and simultaneous saccharification and fermentation experiments using a wide range of plant biomass feedstocks. Our results not only experimentally confirm the structural basis of glucose tolerance in GH1  $\beta$ -glucosidases but also demonstrate a strategy to improve technologies for bioethanol production based on enzymatic hydrolysis.

The recalcitrance of lignocellulose materials is the major problem for the effective utilisation of woody feedstocks to produce biofuels and high-value plant-derived chemicals<sup>1</sup>. The remarkable recalcitrant nature of biomass has a strong impact on its enzymatic saccharification, preventing enzymes from accessing the lignocellulosic matrix<sup>2,3</sup>. Overcoming this physical barrier is the biggest challenge currently facing the development of second generation biofuels<sup>4</sup>. Although studies focusing on the engineering of plant cell walls for enhanced biofuel production<sup>5-7</sup> have shed some light on the matter, the bioprospecting of microorganisms and the improvement and application of novel biocatalysts that can increase the efficiency of the saccharification of cellulosic substrates are still vast fields to be explored<sup>1,8-10</sup>.

Our laboratory has previously reported the crystal structure and biochemical characterisation of a glycoside hydrolase family 1 (GH1)  $\beta$ -glucosidase highly expressed by *Trichoderma harzianum* (ThBgl) under biomass degradation conditions<sup>11</sup>.  $\beta$ -glucosidase (EC 3.2.1.21) catalyses the hydrolysis of  $\beta$ -glucosidic linkages in aryl-

<sup>1</sup>Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, SP, Brazil.

<sup>2</sup>University of Cambridge, Department of Biochemistry, Cambridge, UK. <sup>3</sup>Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil. <sup>4</sup>Natural Material Innovation Centre, University of Cambridge, Cambridge, UK. <sup>5</sup>OpenPlant Synthetic Biology Research Centre, Department of Plant Sciences, University of Cambridge, Cambridge, UK. <sup>6</sup>Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil. <sup>7</sup>Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil. Correspondence and requests for materials should be addressed to A.P.S. (email: [anete@unicamp.br](mailto:anete@unicamp.br))

amino-, or alkyl- $\beta$ -D-glucosides and di- or oligosaccharides at the nonreducing ends, releasing  $\beta$ -D-glucose monomers<sup>12</sup>. This class of enzymes has a critical role within the set of enzymes that compose the enzymatic cocktails designed for plant biomass hydrolysis, thereby alleviating the inhibition of other cellulases by cellobiose during the reaction<sup>13,14</sup>. Notwithstanding their biotechnological potential,  $\beta$ -glucosidases have shown a different degree of sensitivity to high glucose concentrations and may be inhibited by cello-oligosaccharides<sup>15</sup>. Improved  $\beta$ -glucosidase activity was one of the ways in which one of the most efficient commercial enzymatic cocktails (Cellic<sup>®</sup> CTec3, Novozymes A/S, Denmark) available for bioethanol production, was recently boosted<sup>16</sup>.

Based on the effect of glucose on  $\beta$ -glucosidases activity, a classification attempt was recently proposed in which  $\beta$ -glucosidases were grouped as follows: glucose tolerant; inhibited by low concentrations of glucose; stimulated by low glucose concentrations and inhibited by high glucose concentrations; and not inhibited by high glucose concentrations<sup>17</sup>. ThBgl is easily inhibited by glucose<sup>11</sup>. During large-scale biomass conversion, high concentrations of glucose must be reached, which requires enzymes that maintain high catalytic rates under these conditions. The molecular basis of glucose tolerance in  $\beta$ -glucosidases is not completely known<sup>17</sup>; however, it has been correlated with structural features at the entrance channel to the active site rather than the active site itself<sup>18</sup>. A comparative structural analysis of GH1 and GH3  $\beta$ -glucosidases revealed that glucose-tolerant GH1  $\beta$ -glucosidases showed a deeper and narrower substrate channel than other  $\beta$ -glucosidases, and this channel restricts glucose access to the active site, which is inconsistent with the shallow pocket observed in GH3  $\beta$ -glucosidases<sup>18</sup>.

In the present study, structural data combined with site-directed mutagenesis were employed to improve ThBgl properties. The ThBgl structure (PDB ID 5BWF) was used to engineer a high glucose-tolerant protein by substitution with two amino acid residues. The tailored ThBgl (ThBgl-Mut) crystal structure revealed that the selected mutations mimicked the active-site entrance of the highly glucose-tolerant GH1  $\beta$ -glucosidase from *Humicola insolens* (HiBG), acting as gatekeepers, as postulated for the HiBG glucose tolerance mechanism<sup>18</sup>. In addition, the performance of ThBgl-WT and ThBgl-Mut in saccharification and simultaneous saccharification and fermentation using a wide range of plant biomass feedstocks was evaluated. Our structural and biochemical data revealed remarkable differences between the wild-type and tailored proteins, confirming the efficiency of the active-site redesign approach. Taken together, our findings experimentally demonstrate the importance of the active-site entrance for glucose tolerance in GH1  $\beta$ -glucosidases and show a strategy to optimize the efficiency of an enzyme, making it more suitable for industrial applications, especially those related to cellulosic ethanol production.

## Methods

**Protein expression and purification.** The gene encoding the wild-type GH1  $\beta$ -glucosidase from *T. harzianum* (ThBgl-WT) was cloned into pET-28a(+) as described in a previous protocol<sup>11</sup>. The L167W/P172L ThBgl double mutant (ThBgl-Mut) was synthesized by Epoch Life Science, Inc. (Sugar Land, TX, USA) and subcloned into pET-28a(+). The recombinant plasmids for ThBgl-WT and ThBgl-Mut were transformed into the *Escherichia coli* Rosetta strain (Novagen, Darmstadt, Germany) for protein overexpression. The protein expression was achieved by culturing each transformed cell individually in 1L of LB broth supplemented with kanamycin (30  $\mu$ g mL<sup>-1</sup>) and chloramphenicol (34  $\mu$ g mL<sup>-1</sup>) at 37 °C and with shaking at 300 rpm. Once the bacterial cultures reached an OD<sub>600</sub> of 0.8, the recombinant protein expression was induced by 0.4 mM IPTG, followed by cultivation for 18 h at 16 °C and 180 rpm. The induced cells were harvested by centrifugation (3000 g, 15 min, 4 °C), resuspended in 50 mL of buffer A (40 mM HEPES pH 7.5 and 150 mM NaCl) containing 1 mg mL<sup>-1</sup> lysozyme and 1 mM PMSF (phenylmethanesulfonyl fluoride) and incubated for 30 min on ice. The lysed cells were disrupted by ultrasonication, and the soluble fraction was collected by centrifugation (27,000 g, 45 min, 4 °C). Protein purification was achieved using nickel affinity chromatography and gel filtration using a Ni Sepharose High Performance His-Trap column (GE Life Sciences) and a HiPrep 16/60 Sephacryl S-100 HR column (GE Life Sciences), respectively. All chromatographic steps were performed with columns coupled to an ÄKTA FPLC device (GE Life Sciences) or its equivalent, and the absorbance of eluted proteins was monitored at 280 nm. The concentrations of the purified proteins were determined using a Bradford protein assay (Sigma-Aldrich, UK), and the protein purity was estimated by SDS-PAGE.

**Biochemical studies.** Initially, the optimal temperature and pH for ThBgl-WT and ThBgl-Mut proteins were evaluated using *p*-nitrophenyl- $\beta$ -glucopyranoside (*p*NPG, Sigma-Aldrich; St. Louis, MO USA) as the substrate and 100  $\mu$ L reactions containing 3  $\mu$ g of purified enzyme, 100 mM buffer, pH 7.0, and 0.5 mM *p*NPG. All reactions were incubated for 10 min and stopped with the addition of 100  $\mu$ L of 1M Na<sub>2</sub>CO<sub>3</sub>. The absorbance of the *p*-nitrophenol released during the reaction was measured at 405 nm using an Infinite<sup>®</sup> 200 PRO microplate reader (TECAN, Männedorf, Switzerland). 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 3.0 to 10.0 using the buffers citrate-phosphate (pH 3.0, 4.0, 4.5, 5.0, and 5.5), phosphate (pH 6.0, 6.5, 7.0, and 8.0), and glycine (pH 9.0 and 10.0) 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 the optimum temperature of each enzyme by monitoring the rate of hydrolysis of *p*NPG at concentrations ranging from 0.0 to 12.0 mM. The kinetic parameters ( $K_M$  and  $V_{max}$ ) were obtained using GraphPad Prism (GraphPad Software, San Diego, CA, USA) to nonlinearly fit the data with the Michaelis–Menten equation.

The inhibitory effect of glucose on both ThBgl-WT and ThBgl-Mut was investigated by measuring the enzymatic activity on the hydrolysis of *p*NPG at glucose concentrations ranging from 0.0 to 1.0M. The relative activity was defined as the enzymatic activity at different glucose concentrations relative to the enzymatic activity without glucose. All measurements were performed in triplicate and repeated at least twice.

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	91.61, 96.21, 96.45
Molecules per AU <sup>a</sup>	2
Resolution (Å)	48.22 – 2.2
Observed reflections	253,768 (25,032)
Unique reflections	43,877 (4,310)
CC <sub>1/2</sub> <sup>b</sup>	0.99 (0.45)
I/σI	9.74 (1.17)
Completeness (%)	99.87 (99.61)
R-meas	0.164
R-merge	0.149 (1.491)
Multiplicity	5.8 (5.8)
Refinement	
Resolution (Å)	48.22 – 2.2
No. reflections	43,852
R <sub>work</sub> /R <sub>free</sub>	0.178/0.221
No. protein residues	929
B-factor (Å <sup>2</sup> )	
Average	44.80
Macromolecules	44.90
Ligands	59.00
Solvent	38.80
Root mean square deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.79
Ramachandran Plot	
Favored (%)	96
Outliers (%)	0.11
MolProbity clashscore	2.79
PDB code	6EFU

**Table 1.** Data collection and refinement statistics of ThBgl-Mut (Leu167Trp/Pro172Leu). Values in parentheses are for the highest-resolution shell. <sup>a</sup>AU: asymmetric unit; <sup>b</sup>CC<sub>1/2</sub>: correlation between intensities from random half-datasets<sup>37</sup>.

**Crystallization, data collection and processing, and structure determination.** ThBgl-Mut was concentrated to 11.5 mgmL<sup>-1</sup> for crystallization trials. Crystals of ThBgl-Mut were grown by the sitting-drop vapor-diffusion method at 18 °C in drops containing 0.5 μL of the concentrated protein and the same volume of reservoir solution (20% (w/v) PEG 3350, 0.2M sodium nitrate, 0.1M Bis-tris propane pH 6.5).

X-ray diffraction data at a wavelength of 1.459 Å were collected at the W01B-MX2 beamline at the Brazilian Synchrotron Light Laboratory, Campinas, Brazil using a PILATUS2M detector (Dectris, Baden-Dattwil, CHE). Data processing was performed using the XDS package<sup>19</sup>. The ThBgl-Mut structure was solved by the molecular replacement method MOLREP<sup>20</sup> using the wild-type enzyme ThBgl (PDB ID: 5BWF<sup>11</sup>) as a template. Refinement cycles were performed using PHENIX Refine<sup>21</sup> alternated with manual inspection using the program COOT<sup>22</sup>. TLS groups used during refinement, were generated by the TLSMD server<sup>23</sup>.

The final model was validated using the program MolProbity<sup>24</sup>. Data collection, processing and refinement statistics are summarized in Table 1. Representations of the structures were generated in PyMOL<sup>25</sup>, and the structure superposition r.m.s.d. was calculated with PDBE-FOLD<sup>26</sup>. The ThBgl-Mut structure was deposited in the RCSB Protein Data Bank (PDB) under the accession code 6EFU.

**Plant material and AIR preparation.** Sugarcane (steam-exploded bagasse and pith), *Pinus sylvestris* (pine softwood) and birch wood chips (Sigma-Aldrich, UK) were used as feedstocks for the saccharification experiments. The sugarcane consisted of approximately 49.8% (w/v) cellulose, 22.1% (w/v) hemicellulose, and 23.5% (w/v) lignin<sup>27</sup>. Alcohol insoluble residue (AIR) preparation was performed using a ball-milling procedure and no additional chemical pretreatments as described previously<sup>28</sup>.

**Laboratory-scale enzymatic saccharification experiments.** Saccharification reactions were performed using the combination of cellobiohydrolase I (CBHI) from *Trichoderma* sp. and endo-1, 4-β-D-glucanase (endocellulase) from *Aspergillus niger* (Megazyme, Wicklow, Ireland) supplemented with approximately 10 μg of ThBgl-WT or ThBgl-Mut proteins. For all plant biomass materials, enzymatic hydrolysis was carried out in sodium citrate buffer (100 mM, pH 6.0) in a final reaction volume of 0.5 mL using 0.5 mgmL<sup>-1</sup> AIR material for

24 h at 45 °C at 1400 rpm of shaking applied for 30 s every 4 min using a ThermoMixer® C (Eppendorf, UK). The reaction was terminated by heat-treating the suspension at 100 °C for 10 min. Glucose release from the biomass samples was quantified using a commercial kit (K-GLUHK-220A; Megazyme, Wicklow, Ireland). The glucose concentration for each experiment was standardised with readings obtained using biomass and enzyme alone as controls. For each biological sample, at least 3 triplicates were analysed. Average sugar release for each assayed sample was used for the statistical analysis with Student's t-test.

**Simultaneous saccharification and fermentation (SSF) experiments.** For SSF experiments, 50 mg of ball-milled sugarcane bagasse without chemical pretreatments was used for each reaction and prepared as previously described by Lyczakowski *et al.*<sup>6</sup>. The fermentation reactions were amended with approximately 120 µg of ThBgl-WT or ThBgl-Mut in the presence of CBHI and endocellulase or only endocellulase, and 250 µL of a 0.6 OD<sub>600</sub> TOP10 *E. coli* culture carrying the BBa\_K1122676 BioBrick, which encodes a pyruvate decarboxylase and an alcohol dehydrogenase from *Zymomonas mobilis* that allow ethanol production in *E. coli*<sup>29</sup>. Control reactions using only biomass or biomass supplemented with CBHI and endocellulase or endocellulase alone were used to quantify the ethanol production in the presence of ThBgl enzymes. SSF reactions were carried out for 96 h at 37 °C and 200 rpm. Ethanol levels were analysed using a commercial kit (K-ETOH; Megazyme, Wicklow, Ireland). All reactions were performed in triplicate, and the average ethanol production for each assayed sample was used for the statistical analysis with Student's t-test.

**Polysaccharide analysis by carbohydrate gel electrophoresis (PACE).** The products of the sugarcane steam-exploded bagasse saccharification reactions (0.5 mg) digested with ThBgl-WT or ThBgl-Mut in the presence of endocellulase enzyme were analysed by polysaccharide analysis using carbohydrate gel electrophoresis (PACE)<sup>30</sup>. Control hydrolysis reactions using endocellulase alone and no enzyme were also analysed. Released oligosaccharides were dried and reductively aminated with 8-aminonaphthalene-1, 3, 6-trisulphonic acid (ANTS; Invitrogen) and separated by acrylamide gel electrophoresis using a Hoefer SE660 vertical slab gel electrophoresis apparatus (Amersham, Buckinghamshire, UK). Electrophoresis was performed at 10 °C using a 10% (w/v) polyacrylamide gel containing 0.5% (w/v) N,N'-9-methylenebisacrylamide in a 0.1M Tris-borate pH 8.2 buffer system. PACE gels were run at 1000 V for 47 min and then visualized with a G:BOX (Syngene, Cambridge, UK) equipped with a short-pass detection filter (500–600 nm) and long-wave UV tubes (365 nm emission).

## Results

**Biochemical characterisation of ThBgl and its mutant reveal broad enzymatic differences.** The double amino acid replacements L167W/P172L in the ThBgl-WT sequence were selected after comparative structural analyses (PDB ID 5BWF) with HiBG (PDB ID 4MDO), where these two amino acid residues were suggested to be gatekeepers involved in glucose tolerance<sup>18</sup>. The ThBgl double mutant L167W/P172L was overexpressed in and successfully purified from *E. coli* (Supplementary Fig. S1). Assays to find the optimum temperature and pH values were carried out using pNPG (Supplementary Fig. S2). A remarkable increase in the optimum temperature of ThBgl-Mut (50 °C) was observed when compared to ThBgl-WT (40 °C) (Supplementary Fig. S2a). Similarly, ThBgl-Mut showed an expanded range of pH-dependent activity and exhibited high activity from pH 4.0–9.0, which significantly differs from the pH range observed for ThBgl-WT (5.0–7.0) (Supplementary Fig. S2b).

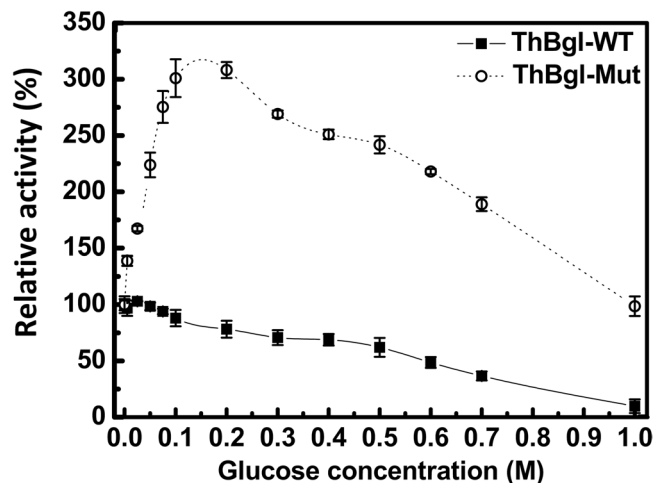
Kinetic data revealed that ThBgl-Mut has a  $K_M$  of  $0.41 \pm 0.30$  mM, which is less than one-half of the  $K_M$  of ThBgl-WT ( $1.17 \pm 0.66$  mM) (Supplementary Fig. S3). Moreover, the catalytic constant ( $k_{cat}$ ) was significantly higher in ThBgl-Mut ( $7.85 \pm 0.22$  s<sup>-1</sup>) than in ThBgl-WT ( $4.92 \pm 0.54$  s<sup>-1</sup>). In fact, the catalytic efficiency parameter ( $k_{cat}/K_M$ ) of ThBgl-Mut is approximately 5-fold higher than that of ThBgl-WT under the conditions assayed.

The inhibitory effect of glucose on the enzymatic activity of both ThBgl-WT and ThBgl-Mut was investigated at glucose concentrations up to 1.0M. The L167W/P172L double mutation had a significant impact on the glucose tolerance and stimulation of ThBgl (Fig. 1). In the presence of 0.1–0.25M glucose, although ThBgl-WT showed an accelerated decline in activity, ThBgl-Mut reached its highest levels, which represented an approximately 300% increase over the value measured under the no glucose conditions (Fig. 1). At the highest glucose concentration assayed (1.0M), the relative activity of ThBgl-Mut was still higher than that without the addition of glucose or than that at those concentrations where ThBgl-WT activity was stimulated (up to 0.05M glucose) (Fig. 1).

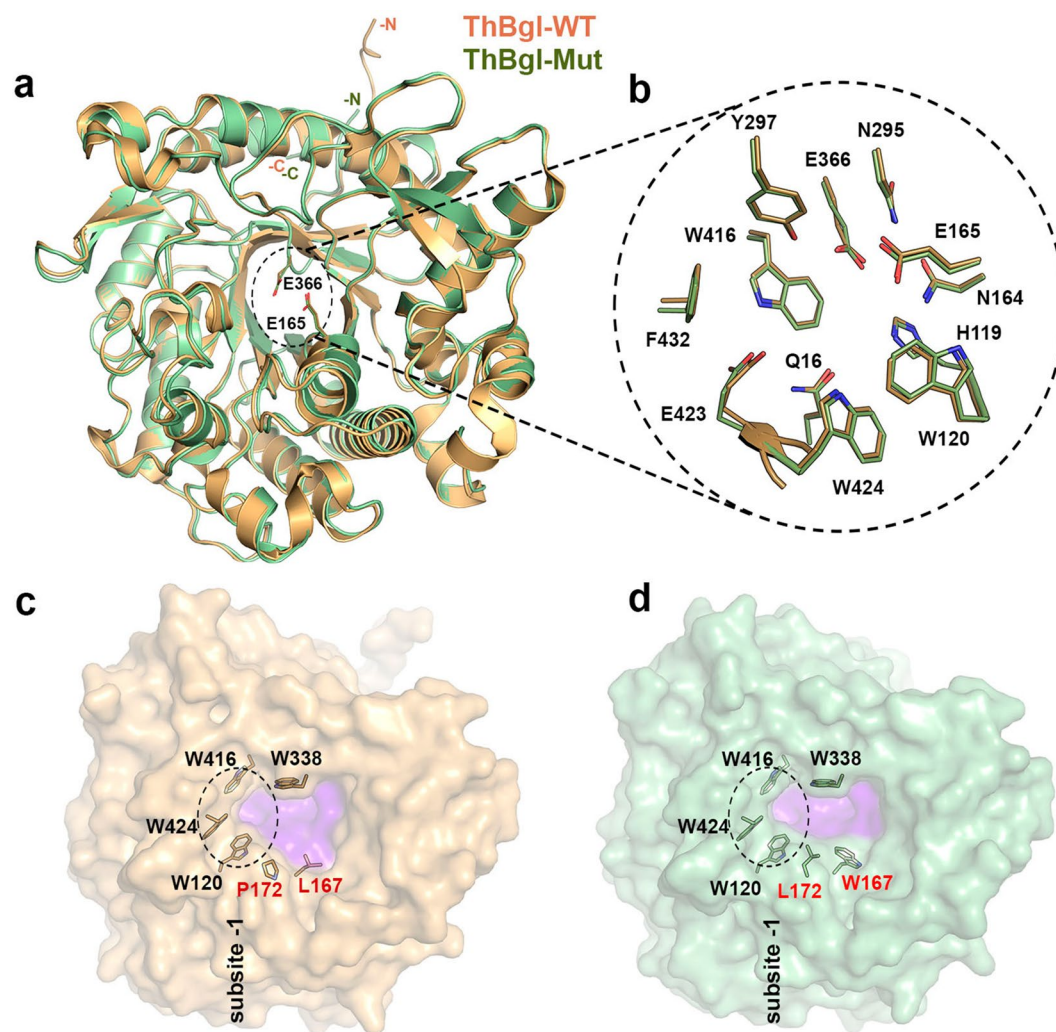
The kinetic and biochemical parameters observed for ThBgl-WT and its mutant corroborate those from a previous study in which a *Trichoderma reesei* β-glucosidase (TrBgl2) mutant L167W/P172L, which has 90% identity with ThBgl, was investigated<sup>31</sup>. However, great differences in the inhibitory effect of glucose were observed between ThBgl-Mut (Fig. 1) and the mutant TrBgl2 because ThBgl-Mut maintained high enzymatic activity at approximately 0.5M glucose, whereas a sharp decline in the activity of TrBgl2 at glucose concentrations higher than 0.1M was reported<sup>31</sup>. This finding demonstrates that ThBgl-Mut may be more suitable than other GH1 enzymes for biomass saccharification at high glucose concentrations.

## Engineered ThBgl-Mut mimics the active-site topology of GH1 glucose-tolerant β-glucosidases.

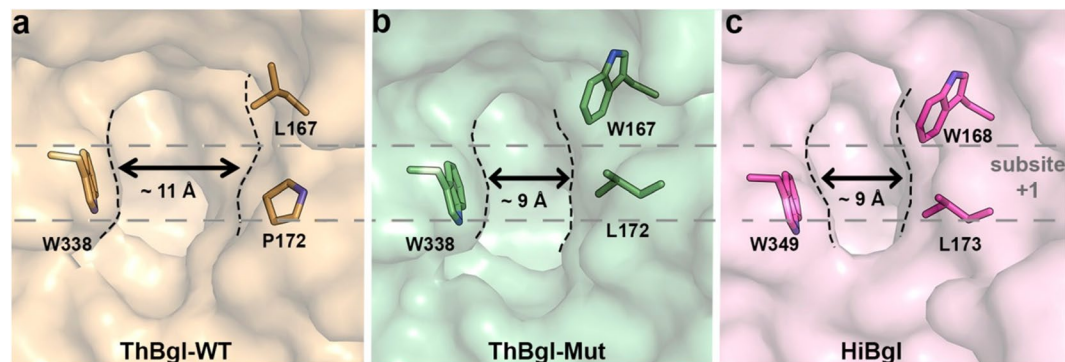
A ThBgl-Mut orthorhombic crystal diffracted to 2.2 Å resolution, and two molecules were found in the asymmetric unit, as observed for ThBgl-WT<sup>11</sup>. As expected, the mutations did not affect the overall tertiary structure, thereby preserving the conserved GH1 (α/β)<sub>8</sub>-TIM barrel fold<sup>32,33</sup>. The ThBgl-Mut and ThBgl-WT structures are very similar (Fig. 2a) and display a r.m.s.d. value of 0.28 Å over 464 Cα atoms, with each monomer being considered. Moreover, the predicted catalytic residues E165 (acid-base) and E366 (nucleophile) are spatially conserved (Fig. 2a) and thus not affected by the rational modifications. Other putative catalytically important residues located in the glycone-binding site of ThBgl also have the same orientation in the WT and mutant enzymes (Fig. 2b).



**Figure 1.** Effect of glucose on ThBgl-WT and ThBgl-Mut activities. Plots are the mean average of the relative activity of independent triplicates, and the error bars represent the standard deviations.



**Figure 2.** Structural comparison between ThBgl-WT and ThBgl-Mut. (a) Cartoon representation of ThBgl-WT (orange) and ThBgl-Mut (green), highlighting the predicted catalytic residues (sticks). N- and C-termini are indicated. (b) Enlarged view of amino acid residues from subsite -1, predicted to be important for catalysis<sup>11</sup>. (c) Surface representation of ThBgl-WT and ThBgl-Mut (d) highlighting the subsite -1 region represented in (b) with dashed lines and the active-site topology in light purple. The mutated residues (P172L and L167W) are labelled in red. Additionally, the positive-subsite delimiting residue W338 is represented as sticks.



**Figure 3.** The engineered ThBgl-Mut active-site topology is similar to that of the glucose-tolerant HiBgl. (a) Surface representation of the ThBgl-WT (a) active-site topology, highlighting the conserved residue W338 and the mutated P172 and L167 residues. In ThBgl-Mut (b), these last two residues were replaced by L172 and W167, resulting in an active-site topology similar to that of HiBgl (c). The +1 subsite region is indicated by gray dashed lines. The estimated measures are an approximated average distance between the active-site extremities represented here.

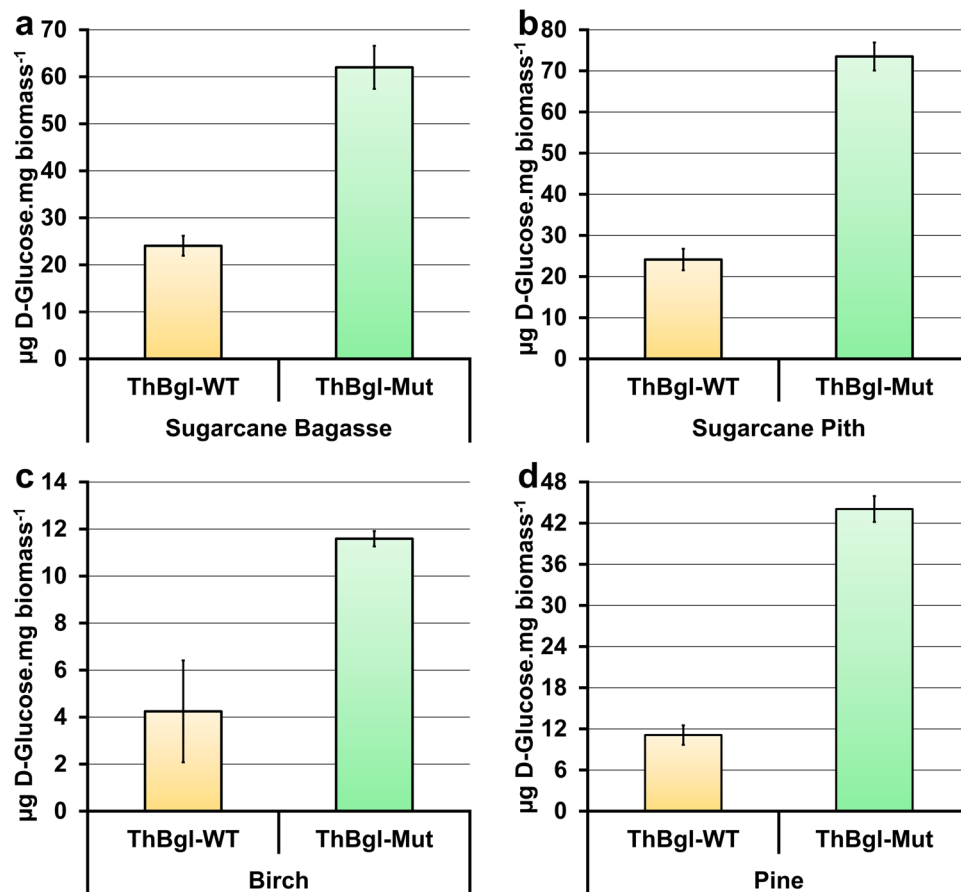
As predicted by our structural analysis, the mutated residues W167 and L172 function as gatekeepers and affect the positive-subsite region topology, narrowing the ThBgl active-site pocket, while the  $-1$  subsite is fully conserved (Figs 2c,d and 3). At the other extremities of these residues, the conserved W338 (corresponding to the residues W349 in HiBgl and W339 in TrBgl) delineates the new, narrowed entrance to the active site. The distance between the two sides of the active site entrance channel is clearly diminished by the ThBgl mutations, which explains its superb tolerance to high glucose concentrations (Fig. 3) and is fully in accordance with the proposed structural determinants for the glucose tolerance mechanism described for HiBgl<sup>18</sup>. Moreover, the same distances between the conserved Trp and the corresponding mutated residues were found in ThBgl-Mut and HiBgl active sites, explaining the similar glucose tolerance levels in these two enzymes (up to 450 mM in HiBgl and 500 mM in ThBgl-Mut) (Fig. 3).

**ThBgl-Mut increases glucose release from lignocellulosic materials.** A wide range of plant biomass materials, including sugarcane, steam-exploded bagasse and raw material, together with pine and birch wood, were employed to compare the efficiency of ThBgl-WT and its mutant protein in laboratory-scale enzymatic saccharification experiments. All assays were performed using AIR without further chemical pretreatments under conditions of limited saccharification and no hemicellulases, such as xylanases, to reveal differences in the rate or extent of saccharification of cellulose with different enzyme cocktails. The  $\beta$ -glucosidases were used in conjunction with a CBHI and an endo-1, 4- $\beta$ -D-glucanase, thus allowing the experiments to specifically focus on  $\beta$ -glucosidase activity.

A remarkable improvement in the glucose release from all lignocellulosic materials used was observed in saccharification reactions when ThBgl-Mut was used (Fig. 4). Initially, the time, temperature and protein loading for the enzymatic hydrolysis were optimized using sugarcane bagasse as a substrate (Supplementary Fig. S4). Despite a time dependence on the sugar release during the saccharification reactions, within 48 h, ThBgl-WT reached the same glucose release observed for ThBgl-Mut at 24 h (Supplementary Fig. S4a). This observation confirms the kinetic data, which show that ThBgl-Mut is a more efficient enzyme (Supplementary Fig. S3). ThBgl-WT showed no significant differences in glucose release at 40 °C and 45 °C and a reduced release at 50 °C, whereas ThBgl-Mut showed greater glucose release at 45 °C (Supplementary Fig. S4b). Concerning the dose effect of ThBgl-WT and ThBgl-Mut concentrations, varying the protein load from 10 to 100  $\mu$ g did not produced significant increase in glucose release (Supplementary Fig. S4c). Therefore, all saccharification assays were performed at 45 °C for 24 h using 10  $\mu$ g of ThBgl enzymes as established in the optimization experiments.

As expected, differences in the amount of glucose released varied with the sort of biomass tested (Fig. 4). For example, higher glucose yields were obtained for hydrolysis reactions using steam-exploded sugarcane bagasse and sugarcane pith compared with the use of other plant feedstocks. In the reactions using the sugarcane material the use of ThBgl-Mut, compared with ThBgl-WT, resulted in an approximately 300% increase in glucose release (Fig. 4a,b). The improvement in the ThBgl-Mut efficiency was also confirmed with the hardwood and softwood materials (Fig. 4c,d). Birch saccharification reactions released less glucose than other plant biomass materials; however, ThBgl-Mut still resulted in increased glucose release (Fig. 4c).

**ThBgl-Mut can boost bioethanol production.** The influence of ThBgl-WT and ThBgl-Mut on bioethanol production was tested in SSF experiments using a synthetic bacterial system able to produce ethanol<sup>29</sup>. In a finding of special importance to biorefineries that use sugarcane bagasse for bioethanol production, our results showed that compared with ThBgl-WT, ThBgl-Mut boosted ethanol production two-fold (Fig. 5a). The initial PACE analysis of the cello-oligosaccharide products from sugarcane bagasse saccharification indicated notable differences between ThBgl-WT and ThBgl-Mut in the reactions supplemented only with endocellulase (Supplementary Fig. S5). SSF assays under these conditions were also performed. Our findings showed that



**Figure 4.** Comparative analysis of the performances of ThBgl-WT and ThBgl-Mut enzymes in laboratory-scale biomass saccharification assays. Saccharification assays were performed using ball-milled biomass material derived from (a) sugarcane steam-exploded bagasse, (b) sugarcane pith, (c) pine and (d) birch. Glucose release was quantified using a commercial kit. The detailed procedure is described in the ‘Methods’ section. Error bars indicate standard errors of the mean from triplicate experiments. All differences in sugar release between ThBgl-WT and ThBgl-Mut were statistically significant at a  $p$ -value  $\leq 0.01$ .

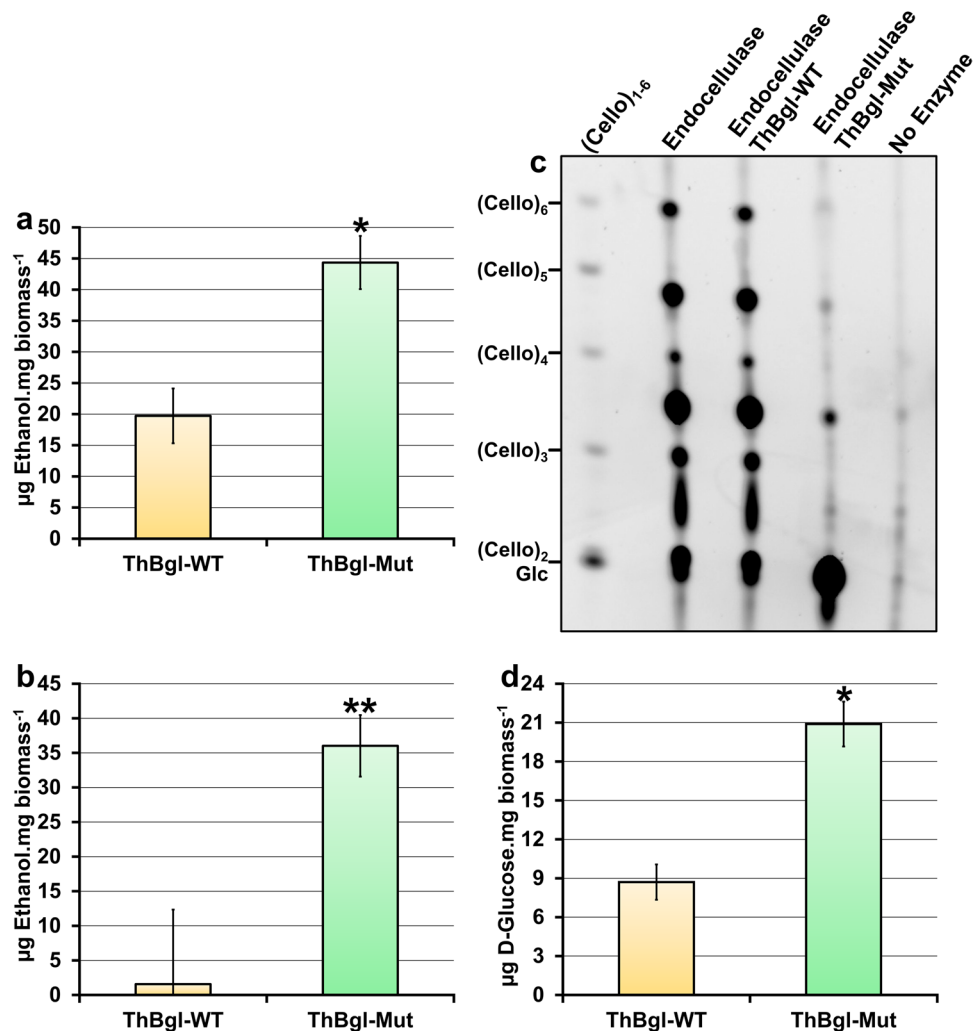
ThBgl-Mut activity greatly differs from that of ThBgl-WT, enabling higher bioethanol yields, just as in the absence of CBHI (Fig. 5b).

To further clarify the increase in bioethanol production, especially in those SSF reactions exploring only the endo-1, 4- $\beta$ -D-glucanase and  $\beta$ -glucosidase activities, the glucose release and the products of sugarcane bagasse saccharification under these conditions were analysed in detail via PACE<sup>30</sup> (Fig. 5c). Our findings strongly support the hypothesis that ThBgl-Mut hydrolyses short cello-oligosaccharides more efficiently, thereby generating more glucose than the WT, as experimentally confirmed (Fig. 5d). Because of the observed difference between ThBgl-Mut substrate specificity and that of the WT enzyme, a possible synergistic action between ThBgl-WT and ThBgl-Mut to boost the saccharification yield was also investigated; however, no enhancement was observed (Supplementary Fig. S6). The ThBgl-Mut-dependent improvement of both sugar release and bioethanol yield seems to be coupled to the better utilization of biomass for enhancing enzyme activity and significantly differs from the behaviour of ThBgl-WT.

## Discussion

$\beta$ -glucosidases that are highly tolerant to glucose inhibition and exhibit efficient hydrolytic properties are continually sought after for industrial applications<sup>1,9</sup>. Herein, we present new insights into glucose tolerance and demonstrate that the stimulation of GH1  $\beta$ -glucosidase activity can be achieved by rationally converting a low glucose-tolerant enzyme into a high glucose-tolerant enzyme with improved catalytic activity on both pure carbohydrates and lignocellulosic materials. Furthermore, the differential enzymatic properties of ThBgl-WT and ThBgl-Mut in the laboratory-scale saccharification of a wide range of plant biomass materials and the impact of ThBgl-Mut on bioethanol production are presented for the first time.

Few studies have engineered fungal  $\beta$ -glucosidases to improve enzyme activity and thermostability; in addition, most of these studies were performed on *T. reesei*  $\beta$ -glucosidase<sup>31,34</sup>. Nevertheless, *T. reesei* shows weak  $\beta$ -glucosidase activity, and the use of its cellulases for biomass saccharification requires supplementation with additional  $\beta$ -glucosidases<sup>34,35</sup>.



**Figure 5.** Bioethanol production and analysis of the cello-oligosaccharides generated by ThBgl-WT and ThBgl-Mut activities. The ethanol concentration after 96 h of SSF was measured (a) for ThBgl-WT or ThBgl-Mut in the presence of CBHI and endocellulase or (b) of only endocellulase. Ethanol yields were standardised using readings from a control fermentation reaction with only CBHI and endocellulase enzymes (Supplementary Table S1). (c) 10% PACE gel of sugarcane bagasse saccharification products generated by ThBgl-WT and ThBgl-Mut in the presence of endocellulase. Sugarcane bagasse digested using only endocellulase or in the absence of enzymes is also shown. A cello-oligosaccharide (Cello)<sub>1-6</sub> ladder is shown on the left. Under these standard conditions, glucose (Glc) and cellobiose (Cello)<sub>2</sub> closely comigrate in the PACE. High- and low-contrast (exposure) gels are shown in Supplementary Fig. S7. (d) Glucose released from sugarcane bagasse saccharification performed using ThBgl-WT or ThBgl-Mut and endoglucanase. Error bars represent the standard deviation of triplicate experiments, \* $p$ -value  $\leq 0.05$ ; \*\* $p$ -value  $\leq 0.01$ .

Notwithstanding, consistent improvement of  $\beta$ -glucosidase activity has been achieved using site-directed mutagenesis based on high-resolution protein structures available in the PDB ([www.rcsb.org](http://www.rcsb.org)). For example, after solving a protein structure, Lee and collaborators<sup>34</sup> characterised a wide range of potential mutations in the substrate entrance of *T. reesei*  $\beta$ -glucosidase (TrBgl2; PDB ID 3AHY)<sup>33</sup>. Interestingly, the potential of L167W and P172L single mutations for improving the enzyme activity was highlighted; although L167W increased the optimum temperature by 10 °C and both constructs showed reduced  $K_M$  and the best enzymatic properties compared to wild-type TrBgl2, no glucose stimulation was observed; moreover, the double mutant L167W/P172L was not analysed<sup>34</sup>.

Subsequently, mutations in TrBgl2 were also generated by replacing the amino acid residues L167, P172 and P388<sup>31</sup> with the corresponding residues W168, L173 and F348 in the HiBgl structure (PDB ID 4MDO)<sup>18</sup>. In this study, the L167W/P172L TrBgl2 double mutant was characterised<sup>31</sup>; however, the current study is the first to determine the crystal structure of both WT and mutant proteins and demonstrate how these mutations alter the active-site channel to mimic the entrance of high glucose-tolerant enzymes. Although ThBgl and TrBgl share high sequence identity and structural similarity, some residues surrounding the active-site are not conserved, such as P338 (replaced by F348 in HiBgl and S337 in ThBgl) (Supplementary Fig. S8). Because this residue is closely located to a conserved tryptophan that participates in the enclosure of the active site (W338 in ThBgl, W349 in



HiBgl and W339 in TrBgl) and due to the influence of a proline to phenylalanine substitution in TrBgl catalysis, it is likely that these neighbouring residues also play a role in glucose tolerance, explaining the differences observed in glucose stimulation levels of TrBgl<sup>31</sup> and ThBgl mutants.

In contrast to TrBgl-Mut, ThBgl-Mut showed continuous stimulation even at high glucose concentrations (1.0M), and the results were very similar to those observed in HiBgl<sup>36</sup>. The designed ThBgl mutations rendered its active-site topology very similar to that of HiBgl, in total accordance with the experimental observations regarding glucose tolerance and with the previously proposed glucose tolerance mechanism<sup>18</sup>. Parallel SSF experiments also demonstrated the improved performance of the ThBgl-Mut over the WT protein. Our findings show that the rational mutation approach used herein was able to enhance the glucose tolerance of ThBgl; however, the glucose tolerance itself does not seem to be the unique property that improves the ThBgl-Mut performance for biomass saccharification. ThBgl-Mut may more efficiently handle short cello-oligosaccharides and/or may remain active at high catalytic rates, thus reflecting the increased glucose release upon saccharification assays. Further experiments, including those in which the catalytic activity of enzymes against complex substrates can be evaluated at high concentrations of glucose, are still required to address these issues.

Together, our results not only experimentally demonstrate the mechanisms of glucose tolerance and stimulation in GH1  $\beta$ -glucosidases but also shed light on the strategies for improving technologies based on enzymatic hydrolysis for bioethanol production.

## Data Availability

The coordinates and structure factors of ThBgl-Mut have been deposited in the RCSB Protein Data Bank under the accession code 6EFU.

## References

- Himmel, M. E. *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804–807, <https://doi.org/10.1126/science.1137016> (2007).
- Chang, V. S. & Holtzapfel, M. T. Fundamental factors affecting biomass enzymatic reactivity. *Appl. Biochem. Biotechnol.* **84–86**, 5–37 (2000).
- Alvira, P., Tomas-Pejo, E., Ballesteros, M. & Negro, M. J. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresour. Technol.* **101**, 4851–4861, <https://doi.org/10.1016/j.biortech.2009.11.093> (2010).
- Hassan, S. S., Williams, G. A. & Jaiswal, A. K. Emerging technologies for the pretreatment of lignocellulosic biomass. *Bioresour. Technol.* **262**, 310–318, <https://doi.org/10.1016/j.biortech.2018.04.099> (2018).
- Loque, D., Scheller, H. V. & Pauly, M. Engineering of plant cell walls for enhanced biofuel production. *Curr. Opin. Plant. Biol.* **25**, 151–161, <https://doi.org/10.1016/j.pbi.2015.05.018> (2015).
- Lyczakowski, J. J. *et al.* Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy. *Biotechnol. Biofuels* **10**, 224, <https://doi.org/10.1186/s13068-017-0902-1> (2017).
- Cai, Y. *et al.* Enhancing digestibility and ethanol yield of *Populus* wood via expression of an engineered monoglignol 4-O-methyltransferase. *Nat. Commun.* **7**, 11989, <https://doi.org/10.1038/ncomms11989> (2016).
- Morais, S. *et al.* Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. *mBio* **3**, <https://doi.org/10.1128/mBio.00508-12> (2012).
- Adsul, M. G., Singhvi, M. S., Gaikawai, S. A. & Gokhale, D. V. Development of biocatalysts for production of commodity chemicals from lignocellulosic biomass. *Bioresour. Technol.* **102**, 4304–4312, <https://doi.org/10.1016/j.biortech.2011.01.002> (2011).
- Horta, M. A. C. *et al.* Network of proteins, enzymes and genes linked to biomass degradation shared by *Trichoderma* species. *Sci. Rep.* **8**, 1341, <https://doi.org/10.1038/s41598-018-19671-w> (2018).
- Santos, C. A. *et al.* Crystal structure and biochemical characterization of the recombinant ThBgl, a GH1  $\beta$ -glucosidase overexpressed in *Trichoderma harzianum* under biomass degradation conditions. *Biotechnol. Biofuels* **9**, 71, <https://doi.org/10.1186/s13068-016-0487-0> (2016).
- Ketudat Cairns, J. R. & Esen, A. beta-Glucosidases. *Cell. Mol. Life Sci.* **67**, 3389–3405, <https://doi.org/10.1007/s00018-010-0399-2> (2010).
- Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C. & Pandey, A. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* **127**, 500–507, <https://doi.org/10.1016/j.biortech.2012.09.012> (2013).
- 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. *Biotech. Bioengine.* **59**, 621–634, [https://doi.org/10.1002/\(sici\)1097-0290\(19980905\)59:5<621::Aid-bit13>3.0.Co;2-c](https://doi.org/10.1002/(sici)1097-0290(19980905)59:5<621::Aid-bit13>3.0.Co;2-c) (1998).
- Xiao, Z., Zhang, X., Gregg, D. J. & Saddler, J. N. Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates. *Appl. Biochem. Biotechnol.* **113–116**, 1115–1126, <https://doi.org/10.1385/ABAB:115-1-3:1115> (2004).
- Cellulosic Ethanol Novozymes Cellic® CTec3 - secure your plant's lowest total cost. *Novozymes A/S · Luna 2012-01394-01 Application Sheet*, 6. [http://s3.amazonaws.com/zanran\\_storage/bioenergy.novozymes.com/ContentPages/2546502386.pdf](http://s3.amazonaws.com/zanran_storage/bioenergy.novozymes.com/ContentPages/2546502386.pdf).
- Salgado, J. C. S., Meleiro, L. P., Carli, S. & Ward, R. J. Glucose tolerant and glucose stimulated beta-glucosidases - a review. *Bioresour. Technol.* **267**, 704–713, <https://doi.org/10.1016/j.biortech.2018.07.137> (2018).
- de Giuseppe, P. O. *et al.* Structural basis for glucose tolerance in GH1  $\beta$ -glucosidases. *Acta. Crystallogr. D. Biol. Crystallogr.* **70**, 1631–1639, <https://doi.org/10.1107/S1399004714006920> (2014).
- Kabsch, W. XDS. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 125–132, <https://doi.org/10.1107/S0907444909047337> (2010).
- Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025, <https://doi.org/10.1107/s0021889897006766> (1997).
- Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta. Crystallogr. D. Biol. Crystallogr.* **68**, 352–367, <https://doi.org/10.1107/s0907444912001308> (2012).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132, <https://doi.org/10.1107/S0907444904019158> (2004).
- Painter, J. & Merritt, E. A. TLSMDweb server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111, <https://doi.org/10.1107/s0021889805038987> (2006).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 12–21, <https://doi.org/10.1107/S0907444909042073> (2010).
- DeLano, W. L. The PyMOL molecular graphics system v. 1.3r1 (Schrodinger, LLC, New York, 2010).
- Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta. Crystallogr. D. Biol. Crystallogr.* **60**, 2256–2268, <https://doi.org/10.1107/S0907444904026460> (2004).

27. Camassola, M. & Dillon, A. J. P. Steam-exploded sugar cane bagasse for on-site production of cellulases and xylanases by *Penicillium echinulatum*. *Energy & Fuels* **26**, 5316–5320, <https://doi.org/10.1021/ef3009162> (2012).
28. Mortimer, J. C. *et al.* Absence of branches from xylan in *Arabidopsis gux* mutants reveals potential for simplification of lignocellulosic biomass. *Proc. Natl. Acad. Sci. USA* **107**, 17409–17414, <https://doi.org/10.1073/pnas.1005456107> (2010).
29. Lewicka, A. J. *et al.* Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*. *ACS Synth Biol* **3**, 976–978, <https://doi.org/10.1021/sb500020g> (2014).
30. Goubet, F., Jackson, P., Deery, M. J. & Dupree, P. Polysaccharide analysis using carbohydrate gel electrophoresis: a method to study plant cell wall polysaccharides and polysaccharide hydrolases. *Anal. Biochem.* **300**, 53–68, <https://doi.org/10.1006/abio.2001.5444> (2002).
31. Guo, B., Amano, Y. & Nozaki, K. Improvements in Glucose Sensitivity and Stability of *Trichoderma reesei* beta-Glucosidase Using Site-Directed Mutagenesis. *PLoS One* **11**, e0147301, <https://doi.org/10.1371/journal.pone.0147301> (2016).
32. Henrissat, B. *et al.* Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA* **92**, 7090–7094 (1995).
33. Jeng, W. Y. *et al.* Structural and functional analysis of three beta-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J. Struct. Biol.* **173**, 46–56, <https://doi.org/10.1016/j.jsb.2010.07.008> (2011).
34. Lee, H. L., Chang, C. K., Jeng, W. Y., Wang, A. H. & Liang, P. H. Mutations in the substrate entrance region of beta-glucosidase from *Trichoderma reesei* improve enzyme activity and thermostability. *Protein Eng. Des. Sel.* **25**, 733–740, <https://doi.org/10.1093/protein/gz073> (2012).
35. 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.* **125**, 728–736 (1999).
36. Souza, F. H. M. *et al.* Purification and biochemical characterization of a mycelial glucose- and xylose-stimulated  $\beta$ -glucosidase from the thermophilic fungus *Humicola insolens*. *Process Biochem.* **45**, 272–278, <https://doi.org/10.1016/j.procbio.2009.09.018> (2010).
37. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033, <https://doi.org/10.1126/science.1218231> (2012).

## Acknowledgements

This study was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, #2015/09202–0 and #2015/26982–0), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; MCTI/CNPQ/Universal, grant no. #430350/2018–0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Computational Biology Program and by the Leverhulme Trust Centre for Natural Material Innovation and the OpenPlant Synthetic Biology Research Centre. CAS is the recipient of a BEPE Postdoctoral fellowship from FAPESP (#2017/17390–7). MABM is the recipient of a Postdoctoral fellowship from FAPESP (#2016/19995–0). J.J.L. was in receipt of a studentship from the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK as part of the Cambridge BBSRC–DTP Programme (Reference BB/J014540/1). OMT was a recipient of an iCASE studentship from the BBSRC (Reference BB/M015432/1). JAFF is the recipient of a PhD fellowship from CNPq. APS and MTM are recipients of 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.

## Author Contributions

C.A.S. carried out all of the experiments and drafted the manuscript. O.M.T. and J.J.L. helped with the saccharification and S.S.F. experiments and drafted the manuscript. J.A.F.F. helped with the protein expression. M.A.B.M., C.C.C.T. and M.T.M. designed and performed the crystallization trials, structure solution and refinement, and structural analyses and drafted the manuscript. L.M.Z. contributed to the biochemical characterization. P.D. designed the experiments, analyzed the data and drafted the manuscript. A.P.S. directed the overall study and drafted the manuscript. All the authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-41300-3>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019