BBA - Proteins and Proteomics 1866 (2018) 569–579

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

# A novel β-glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil)

Danyelle Toyama[a,1], Mariana Abrahão Bueno de Morais[b,1], Felipe Cardoso Ramos[b], Letícia Maria Zanphorlin[b], Celisa Caldana Costa Tonoli[c], Augusto Furio Balula[a], Fernando Pellon de Miranda[d], Vitor Medeiros Almeida[e], Sandro Roberto Marana[e], Roberto Ruller[b], Mario Tyago Murakami[b,*], Flavio Henrique-Silva[a,**]

[a] Laboratory of Molecular Biology, Department of Genetics and Evolution, Federal University of São Carlos, SP, Brazil
[b] Brazilian Bioethanol Science and Technology Laboratory, National Center for Research in Energy and Materials, Campinas, SP, Brazil
[c] Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, Campinas, SP, Brazil
[d] Petróleo Brasileiro S.A. (Petrobras), Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello, Rio de Janeiro, RJ, Brazil
[e] Departamento de Bioquímica, Instituto de Química, SP, Brazil

ARTICLE INFO

*Keywords:*
Amazon
Metagenome
Cellulose
β-Glucosidase
GH1-family

ABSTRACT

The Amazon region holds most of the biological richness of Brazil. Despite their ecological and biotechnological importance, studies related to microorganisms from this region are limited. Metagenomics leads to exciting discoveries, mainly regarding non-cultivable microorganisms. Herein, we report the discovery of a novel β-glucosidase (glycoside hydrolase family 1) gene from a metagenome from Lake Poraquê in the Amazon region. The gene encodes a protein of 52.9 kDa, named AmBgl-LP, which was recombinantly expressed in *Escherichia coli* and biochemically and structurally characterized. Although AmBgl-LP hydrolyzed the synthetic substrate *p*-nitrophenyl-β-D-glucopyranoside (pNPβG) and the natural substrate cellobiose, it showed higher specificity for pNPβG ($k_{cat}/K_m = 6\,s^{-1}{\cdot}mM^{-1}$) than cellobiose ($k_{cat}/K_m = 0.6\,s^{-1}{\cdot}mM^{-1}$). AmBgl-LP showed maximum activity at 40 °C and pH 6.0 when pNPβG was used as the substrate. Glucose is a competitive inhibitor of AmBgl-LP, presenting a $K_i$ of 14 mM. X-ray crystallography and Small Angle X-ray Scattering were used to determine the AmBgl-LP three-dimensional structure and its oligomeric state. Interestingly, despite sharing similar active site architecture with other structurally characterized GH1 family members which are monomeric, AmBgl-LP forms stable dimers in solution. The identification of new GH1 members by metagenomics might extend our understanding of the molecular mechanisms and diversity of these enzymes, besides enabling us to survey their industrial applications.

## 1. Introduction

The Amazon comprises the largest hydrographic basin in the world. Despite its enormous biodiversity, the microorganisms present in their rivers and lakes remain unexplored. The great genetic and metabolic diversity of both cultivable and non-cultivable organisms present in this important region may result in the discovery of new enzymes of biotechnological interest, such as enzymes involved in the degradation of the plant cell walls.

Cellulose is the most abundant component of plant cell wall [1], and the soils, rivers, and lakes of the Amazon have large amounts of organic matter derived from plants degraded by microorganisms. In this way, microorganisms present in this region must produce enzymes capable of degrading cellulose and also converting it into fermentable sugars (*e.g.*, cellulases, cellobiosidases and β-glucosidases).

Cellulases (EC 3.2.1.x) are enzymes synthesized by several organisms and are involved in cellulose degradation [2,3]. They are glycoside hydrolases (GH) and act by hydrolyzing β-1,4 glycosidic bonds in the cellulose chain. They are classified as endoglucanases (EC 3.2.1.4) and exoglucanases (EC 3.2.1.91) that act on cellulose to produce cellobiose

* Correspondence to: M. T. Murakami, Brazilian Bioethanol Science and Technology Laboratory, National Center for Research in Energy and Materials, Giuseppe Maximo Scolfaro 10000, Campinas, SP, Brazil.
** Correspondence to: F. Henrique-Silva, Federal University of São Carlos, Department of Genetics and Evolution, Laboratory of Molecular Biology, Rodovia Washington Luiz, km 235, São Carlos, SP, Brazil.
*E-mail addresses:* mario.murakami@ctbe.cnpem.br (M.T. Murakami), dfhs@ufscar.br (F. Henrique-Silva).
[1] Both authors have contributed equally.

https://doi.org/10.1016/j.bbapap.2018.02.001
Received 19 September 2017; Received in revised form 9 February 2018; Accepted 14 February 2018
Available online 16 February 2018

and together with the β-glucosidases (EC 3.2.1.21), which act on cellobiose to produce glucose, must work synergistically for an efficient cellulose degradation [1,4].

Although they do not act directly on cellulose, β-glucosidases play a fundamental role in its degradation, since the activities of endoglucanase and exoglucanase are inhibited by cellobiose. In addition to the production of glucose by hydrolysis of cellobiose, they also mitigate the inhibitory effect on the enzymatic complex [5,6]. However, glucose may act by inhibiting the activity of β-glucosidases; thus, there is an increasing demand for enzymes insensitive to inhibition by its product and with high thermal stability to improve the saccharification process of lignocellulosic materials [6]. Currently, enzymes of the cellulolytic complex, such as β-glucosidases, are of great biotechnological interest, mainly due to their application in the enzymatic saccharification of lignocellulosic materials, such as sugarcane bagasse, for the production of biofuels [2].

In this study, a novel β-glucosidase from a freshwater metagenome of Lake Poraquê from the Amazon region was recombinantly expressed and biochemically and structurally characterized. Our data showed that AmBgl-LP preserves all the molecular characteristics of a typical GH1 member and reveals that AmBgl-LP dimerizes in solution, which is an unusual feature among structurally characterized GH1 members. X-ray crystallography studies elucidated the molecular basis for dimerization, which was not found in other previously characterized members. However, the role of the quaternary structure in the function is not yet fully understood. Moreover, our data indicate that AmBgl-LP can be used as component of cocktails for biomass degradation, reinforcing the importance of metagenomic approach for both purposes.

## 2. Materials and methods

### 2.1. Sampling and DNA extraction

Water samples were collected during September 2008 (dry season) from the Lake Poraquê, adjacent to the upper Solimões River, in the Amazon region, Brazil (03°57′36.36″S; 63°09′48.17″W). The samples were collected at 1-meter depth (total depth: 1.8 m) in a 10-liter Van Dorn bottle and filtered using a peristaltic pump (Millipore, Darmstadt, DE) and a 142-mm diameter standing stainless steel filter holder (Millipore) with an AP 20 glass microfiber membrane and a 5 μm isopore polycarbonate membrane (142-mm diameter, Millipore). Finally, bacteria and archaea in the water were retained in a 0.22 μm PVDF filter (Sterivex GV, Millipore). Sterivex filters were filled with extraction buffer (40 mM EDTA, 50 mM Tris-HCl pH 8.3, and 0.75 M sucrose) and DNA was extracted using the Water Metagenomic DNA Isolation Kit (Epicentre, Madison, WI, USA) following the manufacturer's instructions. The DNA integrity was confirmed by electrophoresis on a 1% agarose gel and quantified using the NanoDrop ND-1000 system (Thermo Scientific, Waltham, MA, USA).

### 2.2. Sequencing and annotation

The metagenomic DNA was used to construct the library using the Nextera XT SamplePrep kit (Illumina, Inc., San Diego, CA, USA) following the manufacturer's instructions. Sequencing was performed using the HiSeq 1000 (Illumina) platform with pair-end 2 × 100-bp reads. The quality of the sequences in FASTQ format was analyzed using FastQC v0.10.1 and the low-quality bases (Phred < 20) were trimmed using Trimming Reads.pl of the package NGSQCToolkit_v2.3.3.

A *de novo* assembly for metagenome reads was performed with the program IDB-UD v0.19. To predict ORFs, we used the MetaGeneMark v2.8 program, and the annotation was performed using BLAST against Uniprot/Swissprot databases. Databases like PFAM, Gene Ontology, and CAZY were also used. The annotated sequences were organized in a local database of our laboratory based on their functionality. The sequences obtained in the metagenome of the Lake Poraquê have been deposited in the NCBI Short Read Archive (GenBank accession number SRX659579).

### 2.3. Identification and isolation of a novel β-glucosidase

A search in the database created for this metagenome was performed by sequence similarity and resulted in the identification of an annotated sequence as a β-glucosidase, named AmBgl-PL. The β-glucosidase protein sequence was aligned with proteins from other organisms with the aid of the MultiAlin software [7]. The theoretical molecular mass was calculated with the tool available at the ExPASy website (http://web.expasy.org/protparam/).

The ORF encoding the β-glucosidase was amplified by PCR using total metagenomic DNA from Lake Poraquê as a template. The forward primer used was 5′-AAACATATGTTCCCTGACGGATTTGTCTG -3′ (the underlined region indicates *Nde*I restriction site), and the reverse primer was 5′- AAAGGATCCCTAGTTCAAAGTGTCGAGGTGG -3′ (underlined region indicates *Bam*HI restriction site).

The PCR was performed with approximately 5 ng of metagenomic DNA; 200 μM of each dNTP (Promega, Madison, WI, USA); 1.25 U of ACTaq High Fidelity DNA polymerase (ACTGene); 1 × PCR buffer with $MgCl_2$ (ACTGene); 10 pmol of each oligonucleotide; and sterile water to achieve a final volume of 50 μL. The amplification was performed in a T100 thermocycler (BIO-RAD, Hercules, CA, USA), with the following program: 1 cycle at 95 °C for 5 min; 35 cycles at 95 °C for 1 min, 50 °C for 1 min, and 72 °C for 3 min; and a final extension at 72 °C for 20 min.

The amplification product was analyzed by electrophoresis on a 1% agarose gel and visualized under UV light after staining with ethidium bromide. The amplicon of the expected size (1422 bp) was purified from the gel with the *QIAquick* gel extraction kit (Qiagen, Hilden, DE) according to the manufacturer's instructions. The purified product was then visualized on a 1% agarose gel and quantified.

After amplification, the AmBgl-PL ORF was digested with *Nde*I and *Bam*HI restriction enzymes (Thermo Scientific) and cloned into a pET-28a vector (Novagen, Madison, WI, USA). *E. coli* DH5α strain was used to transform the recombinant plasmid. The recombinant plasmid, named pET28a_AmBgl-LP, was sequenced in a MegaBACE 1000 Flexible system using the DYEnamic ET Dye Terminator Kit for MegaBACE™ (GE Healthcare, Little Chalfont, UK). The AmBgl-PL sequence was deposited in the European Nucleotide Archive under accession code LT965074.

### 2.4. Protein expression and purification

For protein expression, *E. coli* Rosetta (DE3) strain cells were transformed with the plasmid pET28a_AmBgl-PL and cultured in Luria-Bertani (LB) Agar (Invitrogen, Carlsbad, CA, USA) containing 25 μg/mL of kanamycin (Thermo Scientific) and chloramphenicol (Thermo Scientific). One colony was picked for growth in 5 mL of LB broth containing 25 μg/mL of kanamycin and chloramphenicol with shaking at 250 rpm for 16 h at 37 °C. After this period, 5 mL of the culture was inoculated into 500 mL of LB Broth containing 25 μg/mL of kanamycin and chloramphenicol for 2 h at 37 °C and 200 rpm until it reached an $OD_{600}$ of approximately 0.6. Protein expression was induced by isopropyl-β-D-thiogalactopyranoside (IPTG) (Invitrogen) at a final concentration of 0.4 mM and incubated at 20 °C for 20 h and 200 rpm. Then, cells were harvested by centrifugation in Sorvall RC-5C Plus (4 °C, 15000 × *g*, 5 min), resuspended in lysis buffer pH 8,0 (100 mM NaCl, 50 mM $NaH_2PO_4$, 10 mM Tris) and lysed by sonication in Sonic Dismembrator 500 (6 pulses of 60 s with 30 s intervals, 20% amplitude). After cell lysis, the soluble fraction of proteins was obtained by centrifugation (4 °C, 25000 × *g*, 10 min). To analyze the solubility of the protein, the supernatant (soluble protein) and cell pellet (insoluble protein) were resolved by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [8].

The soluble recombinant β-glucosidase was purified from the

supernatant by affinity chromatography using an agarose column containing nickel (Ni-NTA Agarose resin, Qiagen). The column was washed with lysis buffer, and the recombinant β-glucosidase was eluted with the same buffer with increasing concentrations of imidazole until 250 mM. The purified protein was analyzed in 12% SDS-PAGE. The fractions containing the purified protein were pooled and dialyzed three times to the membranes (MEMBRA-CEL – Viskase 14000 MWCO) against lysis buffer pH 8.0 at 4 °C for 20 h, filtered in 0.22 μm filters (Millipore) and quantified using the BCA Protein Assay Kit (Thermo Scientific). Glycerol was added to 15% (v/v) and the enzyme, named AmBgl-PL, was stored at −20 °C.

### 2.5. Enzymatic characterization

For functional characterization, we used the synthetic substrate *p*-nitrophenyl-β-D-glucopyranoside (pNPβG) (Sigma Aldrich, St. Louis, MO, USA) prepared in 100 mM citric acid-sodium phosphate buffer pH 6.0. Briefly, the assays were performed with 2 mM pNPβG and 26 nM enzyme in a final volume of 100 μL. Reactions were terminated by adding 100 μL of 0.5 M $Na_2CO_3$ and absorbance was read at 405 nm in a VICTOR$^3$ spectrophotometer (PerkinElmer, Waltham, MA, USA). All reactions were performed in triplicates and incubated for 15 min in a Veriti thermal cycler (Applied Biosystems, Foster City, CA, USA). Optimum temperature was determined by incubating the reactions at temperatures ranging from 20 to 65 °C with increments of 5 °C. The optimum pH for AmBgl-PL activity was determined using 80 μL of Davies universal buffer [9], within a range from 3 to 9 with increments of 0.5, pNPβG as substrate at 40 °C.

Thermostability was evaluated by pre-incubating the purified enzyme for 0, 5, 10, 15, 20, 30, 45, and 60 min, in the absence of substrate, at temperatures of 0, 20, 25, 30, 35, 40, 45, 50, and 60 °C. After each incubation time of the enzyme at different temperatures, the reactions were immediately carried out with 2 mM pNPβG and 26 nM enzyme in a final volume of 100 μL (4 reactions for each incubation time and temperature). The reactions were then incubated in a water bath for 20 min under optimum temperature conditions (40 °C) and pH (pH 6.0) previously established. The reactions were stopped by adding 100 μL of 0.5 M $Na_2CO_3$ every 5 min, and absorbance was read at 405 nm in a VICTOR$^3$ spectrophotometer.

Kinetic parameters including Michaelis-Menten constant ($K_m$), maximum velocity ($V_{max}$), and catalytic rate ($k_{cat}$) were determined by fitting the data obtained in initial rates ($v_0/[S]$) plots to the Michaelis-Menten equation. Enzymatic reactions were performed under standard conditions with 29 different concentrations of pNPβG (from 0.05 to 20 mM final concentrations) and 26 nM AmBgl-PL in a final volume of 100 μL. The reactions were incubated in a water bath (40 °C) for 20 min and stopped by adding 100 μL of 0.5 M $Na_2CO_3$ every 5 min. The reaction was stopped, and absorbance was read at 405 nm in a VICTOR$^3$ spectrophotometer [10].

The enzyme activity upon the synthetic substrate *p*-nitrophenyl-β-D-galactopyranoside (pNPβGal) (Sigma Aldrich) was also evaluated. This substrate was prepared in 100 mM citric acid-sodium phosphate buffer pH 6.0, and the assay conditions were similar to described for pNPβG above. As the ($v_0/[S]$) plots were linear, data were submitted to linear regression fitting, and the $k_{cat}/K_m$ was determined based on the line slope.

The enzyme was incubated with nine different pNPβG concentrations in five concentrations of glucose (ranging from 50 to 250 mM) to evaluate the inhibitory effect of glucose on AmBgl-PL. $K_i$ value was determined from the replot of the slope of Lineweaver-Burk lines *versus* inhibitor concentration [11]. Substrate and inhibitor were prepared in 100 mM citric acid-sodium phosphate buffer pH 6.0.

Kinetic parameters were also evaluated for cellobiose substrate. For this, 25 different concentrations of cellobiose (Sigma Aldrich) (0.1 to 24 mM final concentrations) were prepared in a 100 mM citric acid-sodium phosphate buffer pH 6.0. The reactions were performed with 260 nM AmBgl-PL, and substrate at different concentrations in a final volume of 100 μL. The reactions were incubated in water bath at the optimum temperature (40 °C) for 4, 8, 12, and 16 min. For this assay, the reactions were stopped by heating (90 °C) in a Veriti thermal cycler for 5 min. For determination of glucose released, the enzymatic kit Glucose-PP (Gold Analisa Diagnóstica, Belo Horizonte, Brazil) was used, adding 200 μL of the commercial reagent and 50 μL of the enzyme/substrate reaction. The reaction was incubated at 37 °C for 10 min, and absorbance was read at 490 nm on a Biochrom Asys UVM 340 spectrophotometer [10].

All statistical analyses were performed at the software Prism 6 (GraphPad Prism, La Jolla, CA, USA).

### 2.6. Circular dichroism spectroscopy and thermal unfolding studies

Circular dichroism (CD) analyses were performed at a J-815 Spectropolarimeter (JASCO, Oklahoma City, OK, USA). The final CD spectrum was an average of 20 accumulations measured over wavelengths ranging from 190 to 260 nm at 20 °C at a 0.1 cm optical path cuvette and scanning speed of 100 nm·min$^{-1}$. The melting temperature ($T_M$) was determined by measuring the residual molar ellipticity at 210 nm over a temperature range from 20 to 90 °C, at a heating rate of 1 °C·min$^{-1}$. The protein concentration in all experiments was 0.2 mg·mL$^{-1}$ diluted in 25 mM phosphate buffer (50 mM NaCl, pH 7.4).

### 2.7. Protein crystallization

The recombinant AmBgl-PL was concentrated to 12 mg·mL$^{-1}$ in buffer pH 8.0 (100 mM NaCl, 50 mM $NaH_2PO_4$, 10 mM Tris and 15% (v/v) glycerol), in an Amicon-4/30 kDa centrifugal filter (Millipore), at 3220 ×*g* and 4 °C. Crystallization trials were carried out in 96-well sitting-drop plates using a Honeybee 963 automated system (Digilab, Marlborough, MA, USA). Drops consisting of 0.5 μL of protein sample mixed with 0.5 μL of reservoir solution were equilibrated against 80 μL of the same reservoir solution. Crystallization conditions formulated based on the commercial kits Crystal Screen HT (Hampton Research, Aliso Viejo, CA, USA), Wizard I and II (Emerald BioSystems, Bainbridge Island, WA, USA), The PACT Suite (Qiagen), JCSG-plus (Molecular Dimensions, Newmarket, UK), SaltRX HT (Hampton Research) and Precipitant Synergy Primary (Emerald BioSystems), were tested. The crystal, used for data collection, was obtained in the condition from the PACT Suite (Qiagen), containing 0.1 M MES pH 6.0, PEG 6000 20% (w/v) and 0.2 M sodium chloride.

### 2.8. Data collection, structure determination and refinement

X-ray diffraction data were collected at the W01BMX2 beamline (LNLS, Campinas, Brazil). Data were indexed, integrated, merged, and scaled using XDS [12]. The structure was solved by molecular replacement using the program MOLREP [13] with the atomic coordinates of the β-glucosidase from *Streptomyces* sp. (54% sequence identity; PDB entry 1GNX) as template. Model refinement was carried out alternating cycles of phenix.refine [14] or Refmac5 [15], with visual inspection of the electron density maps and manual rebuilding with COOT [16]. A detailed statistics of Data collection and refinement is provided in Table 1. The AmBgl-PL structure was deposited in the RCSB Protein Data Bank under accession code 5WKA.

### 2.9. Small Angle X-ray Scattering (SAXS)

SAXS data were collected at the SAXS1 beamline (Brazilian Synchrotron Light Laboratory, Campinas, Brazil). The radiation wavelength was set to 1.54 Å, and a Pilatus 300 K (Dectris, Baden-Dättwil, CHE) detector was used to record the scattering patterns. The sample-to-detector distance was set to 1 m to give a scattering vector-range from 0.1–5.0 nm$^{-1}$. Protein samples at 6, 3 and 1 mg·mL$^{-1}$ were

**Table 1**
Data collection and refinement statistics.

| Data collection | |
|---|---|
| Space group | C121 |
| Cell dimensions | |
| a, b, c (Å) | 165.47, 56.65, 209.55 |
| α, β, γ (°) | 90.00, 96.66, 90.00 |
| Molecules per AU[a] | 4 |
| Resolution | 46.46–2.75 (2.84–2.75)[b] |
| Observed reflections | 148,890 (11,159) |
| Unique reflections | 49,452 (4550) |
| $CC_{1/2}$[c] (%) | 98.0 (53.8) |
| I/σI | 6.0 (1.04) |
| Completeness | 97.14 (90.28) |
| $R_{meas}$ (%) | 23.9 (122.0) |
| $R_{merge}$ (%) | 19.7 (97.0) |
| Multiplicity | 3.0 (2.5) |
| Refinement | |
| Resolution (Å) | 46.46–2.75 |
| No. reflections | 49,403 |
| $R_{work}/R_{free}$ | 0.23/0.27 |
| No. protein residues | 1859 |
| Average B factor (Å$^2$) | 57.0 |
| Macromolecules (Å$^2$) | 56.9 |
| Ligands (Å$^2$) | 71.8 |
| Solvent (Å$^2$) | 52.5 |
| Root mean square deviations | |
| Bond lengths (Å) | 0.003 |
| Bond angles (°) | 0.68 |
| Ramachandran plot | |
| Favored (%) | 95.1 |
| Allowed (%) | 4.7 |
| Disallowed (%) | 0.2 |
| Molprobity clashscore | 3.54 |

[a] AU, asymmetric unit.
[b] Values in parentheses are for highest-resolution shell.
[c] $CC_{1/2}$, correlation between intensities from random half-datasets [17].

prepared in buffer pH 8.0 (100 mM NaCl, 50 mM $NaH_2PO_4$, 10 mM Tris and 15% (v/v) glycerol). 10 frames with exposure time of 30 s were recorded, and buffer baselines were collected under identical conditions. Background scattering was subtracted from the protein scattering pattern, which was then normalized and corrected. Experimental data fitting and evaluation of the pair-distance distribution function *P*(r) were performed using the program GNOM [18]. The low-resolution envelopes were determined using *ab initio* modeling as implemented in the program DAMMIN [19]. An averaged model was generated using the package DAMAVER [20]. The low-resolution model and the crystal structure were superimposed using the program SUPCOMB [21].

### 2.10. Dynamic light-scattering (DLS)

The hydrodynamic behavior of AmBgl-PL was assessed by dynamic light scattering (DLS). DLS experiment was performed in a Dynapro Molecular Sizing instrument (Wyatt Technology, CA, USA) at 20 °C. The protein sample was prepared at a final concentration of 12 mg·mL$^{-1}$ in buffer pH 8.0 (100 mM NaCl, 50 mM $NaH_2PO_4$, 10 mM Tris and 15% (v/v) glycerol). Samples were previously centrifuged for 20 min at 20,000 × *g*. Data were collected with at least 100 acquisitions. The diffusion coefficient ($D_T$) was determined from the analysis of measured time-dependent fluctuations in the scattering intensity and used to calculate the hydrodynamic radius ($R_h$) of the protein according to the Stokes-Einstein equation. Data analysis was performed at the software Dynamics V6.3.40.

## 3. Results

### 3.1. Sequence analysis

The β-glucosidase AmBgl-PL gene contains an ORF of 1422 bp, which encodes a protein with 473 amino acids and predicted molecular mass of 52.9 kDa. Blastx showed that AmBgl-PL is closely related to β-glucosidases (GH1) from *Acidithrix ferrooxidans*, *Rubrobacter* spp. and *Streptomyces* spp. with maximum identity of 57% for *A. ferrooxidans*. The amino acid sequence of AmBgl-LP was aligned with other β-glucosidases and showed many identical regions including conserved regions of the GH1 family proteins. The residue of glutamate Glu369, probably is the catalytic nucleophile, whereas the residue of glutamate Glu161, represents the general acid/base (Fig. 1).

### 3.2. Heterologous expression and purification of AmBgl-LP

The AmBgl-LP protein was expressed with a molecular mass of 55 kDa, including 1 kDa from 6 His N-terminal tag, as expected. The expression was performed at 20 °C (to improve solubility of the protein) for 20 h with the majority of the enzyme recovered in the soluble fraction (Supplementary Fig. 1). The recombinant AmBgl-LP was eluted in lysis buffer containing 75 mM imidazole, and the final yield of the purified protein was approximately 28 mg per liter of culture.

### 3.3. Biochemical characterization of recombinant AmBgl-LP

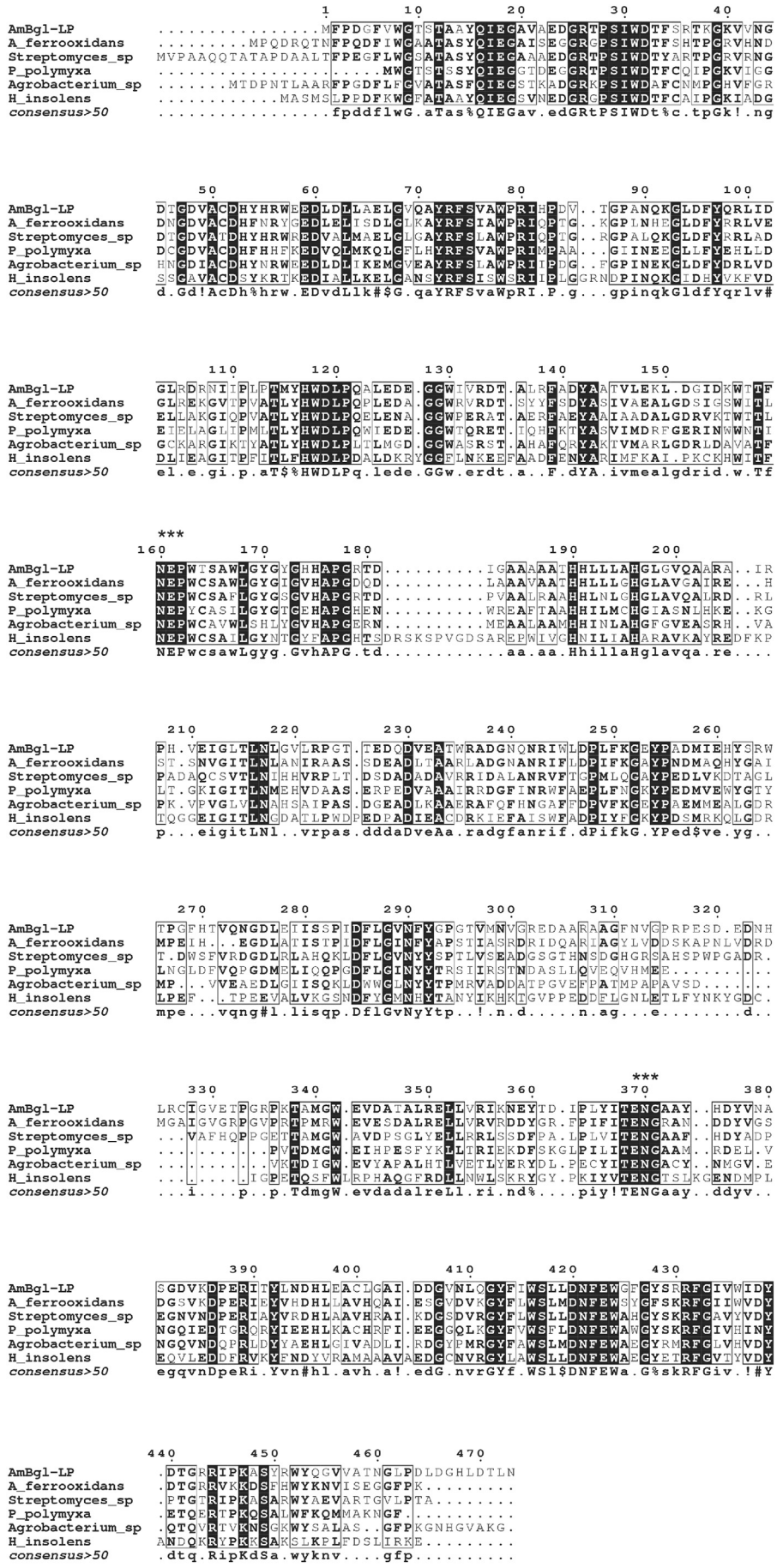
The optimum temperature and pH for AmBgl-LP were tested using pNPβG as substrate. The AmBgl-LP had an optimum temperature of 40 °C with 96% of its activity observed at 35 °C (Fig. 2A). The enzyme was most active in a pH range between 5.5 and 6.5 with maximum activity at pH 6.0 (Fig. 2B). Regarding enzyme thermostability, the enzyme remained highly stable up to an hour of incubation at temperatures of 20, 25, and 30 °C but in the optimum temperature, the AmBgl-LP exhibited about 60% of its activity after 30 min of incubation (Fig. 2C).

The enzyme was able to hydrolyze the synthetic substrate pNPβG and pNPβGal and the natural substrate cellobiose. Enzyme kinetic parameters were obtained by measuring the rate of hydrolysis of pNPβG and cellobiose using a non-linear Michaelis-Menten curve, as shown in Fig. 3A and B, respectively. These results revealed that AmBgl-PL hydrolyzes pNPβG more efficiently than cellobiose (Table 2). A linear $v_0$ *versus* [S] plot was observed for the substrate pNPβGal (Supplementary Fig. 2), suggesting that the $K_m$ for this substrate is very high and also precluding the separate determination of the $k_{cat}$ and $K_m$. Nevertheless, the $k_{cat}/K_m$ ratio (0.1 s$^{-1}$·mM$^{-1}$) was calculated based on the slope of that line, showing that the enzyme activity upon pNPβGal is very low.

The effect of glucose on the AmBgl-LP activity was evaluated by determining the enzyme activity upon increasing concentrations of substrate in the presence of a fixed concentration of glucose. The analysis of the Lineweaver-Burk plots (Fig. 4) showed that glucose is a competitive inhibitor of AmBgl-LP, presenting a $K_i$ of 14 mM.

### 3.4. Biophysical characterization of AmBgl-LP

CD analysis revealed that AmBgl-LP showed a typical spectrum of α/β proteins with a minimum peak at 210 nm and confirmed the proper folding of the recombinant enzyme as expected according to activity assays (Fig. 5A). To assess the structural stability of the enzyme, it was thermally unfolded up to 70 °C, and CD signal was monitored at 210 nm. The unfolding process followed the two-state model, and the enzyme exhibited a $T_M$ of 46.7 °C, being completely unfolded at 50 °C (Fig. 5B). The thermal unfolding process showed to be irreversible since the temperature reduction did not result in the recovery of the original CD profile.




*(caption on next page)*

**Fig. 1.** Alignment between AmBgl-LP and other previously reported β-glucosidases. Motifs involved in catalysis (NEP and ENG) in GH1 β-glucosidases are marked with asterisks. The β-glucosidases used in the alignment are from the following organisms: *Acidithrix ferrooxidans* (WP_052606161.1), *Agrobacterium* sp. (A28673), *Streptomyces* sp. (PDB code: 1GNX), *Paenibacillus polymyxa* (PDB code: 2O9P), *Humicola insolens* (PDB code: 4MDP). Black boxes indicate sequence identity of 100% and empty boxes indicate identity between 50 and 80%.

### 3.5. The AmBgl-LP three-dimensional structure

The crystal of AmBgl-LP diffracted to 2.75 Å, which allowed structure determination by molecular replacement calculations using the β-glucosidase from *Streptomyces* spp. (PDB ID 1GNX, amino acid sequence identity = 54%) as the search model (Table 1). A search for structural relatives using PDBeFold [22] with AmBgl-PL as template revealed a close similarity to other GH1 β-glucosidases, with an r.m.s.d. for pairwise C superpositions of 1.027 and 0.904 Å between 48% and 54% sequence identity. The asymmetric unit contains 4 molecules, and each monomer exhibits the typical $(\beta/\alpha)_8$ TIM-barrel fold of GH1-glucosidases (Fig. 6A). The distance between the two catalytic residues Glu369 and Glu161, predicted by analogy, falls between the average distances found in enzymes that adopt the retaining catalytic mechanism, as GH1 members [23].

The active site located in a deep pocket is characteristic of most GH1 enzymes (Fig. 6B). Presumably, at the −1 subsite, according to structural comparisons, a glycerol molecule, used as cryoprotectant, was found (Fig. 6A and C), as previously observed in other GHs [24]. The glycerol is placed near the putative catalytic residues Glu369 and Glu161 (Fig. 6A). At approximately 7 Å of distance from the cryoprotectant, a PEG molecule was found in a hydrophobic patch, near the inferred +2 subsite (Fig. 6C). The predicted aglycone region is surrounded by hydrophobic residues (Val220, Pro294, and Ala237) that accommodate the reducing end portion of the substrate and restricts the entrance to the active site (Fig. 6C). The residues His174 and Trp342 delimit the access to the −1 subsite and are critical to define the active site architecture (Fig. 6C).

### 3.6. AmBgl-LP is a dimer in solution

Despite the presence of four molecules in the asymmetric unit, energetic analysis by PDBePISA [26] suggested a homodimeric quaternary assembly as the most stable, with a $\Delta G_{int} = -8.1$ kcal/mol, for the dimer between monomers A and B and $\Delta G_{int} = -9.1$ kcal/mol between monomers C and D. This is in agreement with the DLS data, which indicated that AmBgl-LP assumes a dimeric structure in solution, exhibiting a hydrodynamic radius of 4.3 ± 1.0, with 15.1% of polydispersity (Supplementary Fig. 3).

To confirm the biological dimer, SAXS experiments were carried out, and as predicted, the SAXS scattering curve displayed excellent agreement with the theoretical curve calculated from the crystallographic dimer (Fig. 7A). By the Guinier fitting of SAXS scattering curve, it was possible to obtain a radius of gyration of 3.17 ± 0.01 nm (Fig. 7B). Pair-distance distribution curve (Fig. 7C) showed a maximum molecular dimension ($D_{max}$) of approximately 10 nm and the same radius of gyration, in agreement with the Guinier analysis and DLS data. Moreover, three different protein concentrations produced similar results in SAXS experiments, excluding the effect of protein concentration in the oligomerization under the tested conditions (Supplementary Table 1). Moreover, fitting of the crystallographic structure of dimer into the *ab initio* SAXS envelope indicated a good shape complementarity, also supporting the dimeric state (Fig. 7D).

### 3.7. The dimeric arrangement of AmBgl-LP

The calculated interface areas between the monomers that form each dimer present in the AmBgl-LP asymmetric unit (according to

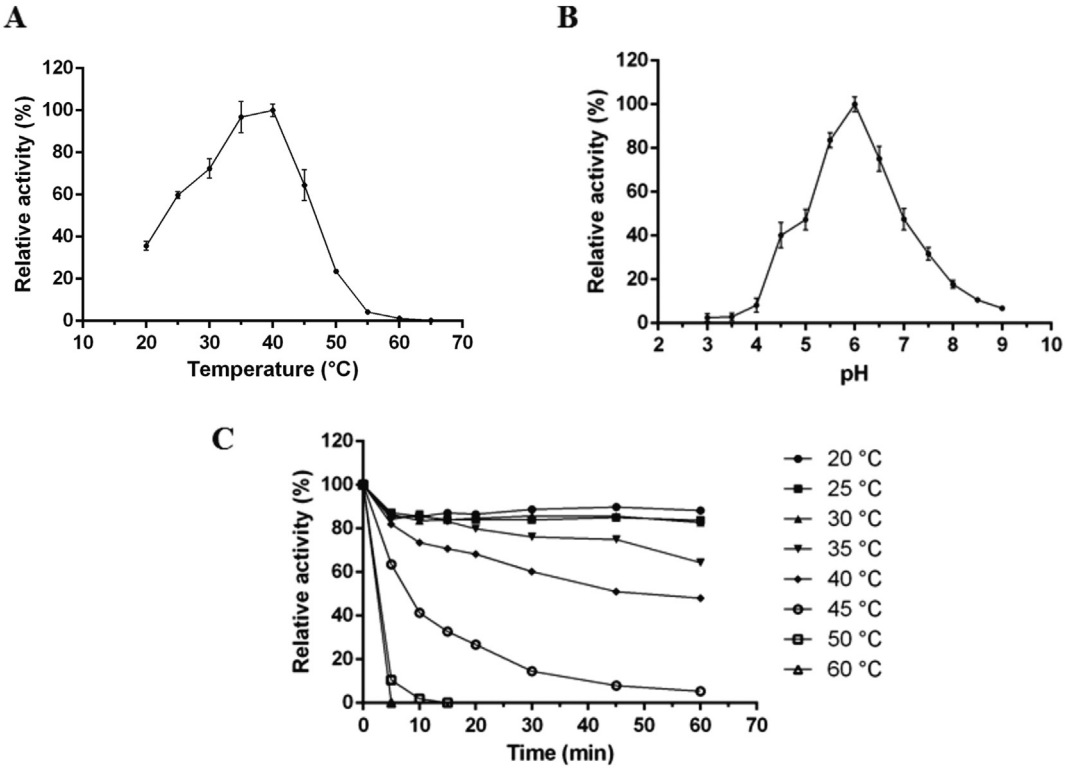



**Fig. 2.** Biochemical characterization of recombinant AmBgl-LP. A. Effect of temperature on the relative enzymatic activity. pNPβG was used as the specific substrate. The optimum temperature of the enzyme was 40 °C. B. Effect of pH on the enzymatic activity. pNPβG was used as the specific substrate. The optimum pH of the enzyme was 6.0. C. Determination of thermostability of AmBgl-LP. Thermal stability was verified by pre-incubation of the enzyme for up to 1 h in temperatures ranging from 20 to 60 °C. pNPβG was used as the specific substrate and the reaction occurred under optimum temperature and pH conditions.

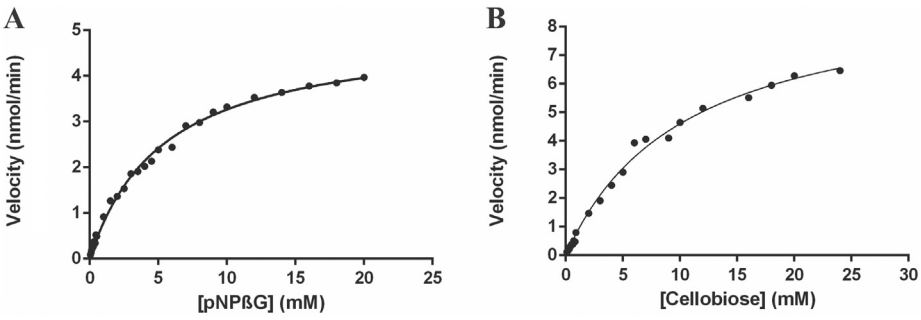



**Fig. 3.** Kinetic characterization of the enzymatic activity of AmBgl-LP. A. Kinetic characterization using pNPβG as substrate. B. Kinetic characterization of AmBgl-LP using cellobiose as substrate. Enzyme activity was determined in 100 mM citric acid-sodium phosphate buffer pH 6.0 at 40 °C. The Michaelis-Menten curve was obtained by plotting the velocity of the reaction *versus* the increasing concentration of the substrate.

PDBePISA [26]) are very similar, with values of 1264 Å$^2$ and 1377 Å$^2$. In the homodimers of AmBgl-LP, the catalytic pockets are located on opposite sides. The AmBgl-LP dimeric interface is mainly stabilized by salt bridges and hydrogen bonds and is composed essentially by three main contact areas (Fig. 8A and B). The region I is stabilized by hydrogen bonds between Arg307, Asn312, and Phe311 (main chain) and the region II involves hydrogen bonds between Tyr 377, Asn43 and Asp45. The region III comprises salt bridges between the residues Arg443 and Asp440 (Fig. 8C).

Structural comparisons with monomeric GH1s (PDB code 4MDP, [27] and 2O9P [28]), confirmed a high level of structural similarity, including the conservation of several residues from the active site (Fig. 8D and E). However, differences were observed at the three regions that compose the dimerization interface (Supplementary Fig. 4). In monomeric proteins, the interactions in region II and III can be partially conserved, while, in the region I, which is stabilized by hydrogen bonds, the structure is very divergent (Fig. 8D), indicating that this region might be crucial for dimer stabilization.

## 4. Discussion

Metagenomic studies have great potential in discovering new genes and enzymes. However, there is a lack of studies in one of the most important regions of the world, the Amazon region. Our group performed the first freshwater metagenome [29] in Amazon, and recently, we published a snapshot of the microbiota of Amazonian rivers and lakes using large-scale sequencing [30,31]. Despite that, none of the enzymes identified in these studies have been characterized. Amazon rivers and lakes can be a rich source of new enzymes such as cellulases that are fundamental in the breakdown and conversion of cellulose into fermentable sugars for the production of biofuels. In 2014, Bergmann et al. [32] constructed and analyzed a soil metagenomic library from the Amazon region, from which GH3 β-glucosidase genes were identified and biochemically characterized. In this study, we identified a novel GH1 β-glucosidase in a metagenome from an Amazonian lake, which was further heterologously produced, characterized and its structure solved by X-ray crystallography.

After sequencing of the metagenomic DNA from the Lake Poraquê, 233,575 contigs and 348,051 ORFs were generated, which resulted in the identification of a sequence of 1422 bp encoding a putative β-glucosidase, named AmBgl-PL. The alignment of AmBgl-PL with other β-glucosidases (Fig. 1) showed many identical regions, including the two catalytic residues (glutamic acids) that are present in conserved regions of the GH1 family proteins, in the motif "NEP" (where E acts as catalytic acid/base) and in the motif "ENG" (where E acts as Nucleophile) [33,34].

The purified AmBgl-PL showed activity on the synthetic substrate (pNPβG) and on the natural substrate (cellobiose), with an optimum temperature close to 40 °C (Fig. 2A). AmBgl-PL exhibits characteristics of mesophilic β-glucosidases, whose optimum temperatures are between 30 and 65 °C, being inactivated at temperatures of 50 to 70 °C [35]. These activity assays corroborate with thermal unfolding states in which the enzyme showed a $T_M$ with degrees above the optimum temperature and total loss of structure above 50 °C. Metagenomics has already allowed isolating some β-glucosidases that presented optimal temperatures around 40 °C [36] to 90 °C [37].

The use of cellulases in the hydrolysis of cellulose for the production of ethanol has been investigated, mainly in the process of simultaneous

**Table 2**
Kinetic parameters of AmBgl-LP and other β-glucosidases isolated from different metagenomes and some specific organisms.

| | *Specific activity* (U/mg) | | $K_m$ (mM) | | $k_{cat}$ ($s^{-1}$) | | $k_{cat}/K_m$ ($s^{-1}$·$mM^{-1}$) | | Reference |
|---|---|---|---|---|---|---|---|---|---|
| | *p*NPβG | Cellobiose | *p*NPβG | Cellobiose | *p*NPβG | Cellobiose | *p*NPβG | Cellobiose | |
| Metagenome environment | | | | | | | | | |
| Lake Poraquê (Amazon) AmBgl-LP – GH1 | 36.7 | 6.57 | 5.38 | 10.43 | 33.7 | 6 | 6.3 | 0.6 | This work |
| South China Sea Bgl1A – GH1 | 50.7 | 15.5 | 0.39 | 20.4 | – | – | – | – | [36] |
| Soil (Amazon) AmBGL17 – GH3 | 85 | – | 0.30 | – | 38.57 | – | 128 | – | [32] |
| Compost microbial (bark and humus) Td2F2 - GH1 | – | – | 0.39 | 4.44 | 12 | 7.13 | 30.6 | 1.61 | [42] |
| Termite gut (*Globitermes sulphureus*) Bgl-gs1 – GH1 | 110.6 | 72.5 | 0.18 | – | – | – | – | – | [37] |
| Organism | | | | | | | | | |
| *Exiguobacterium antarcticum* B7 EaBgl1A- GH1 | 36.7 | 15.4 | 1.07 | 6.18 | 32.98 | 6.92 | 30.8 | 1.12 | [43] |
| *Thermoanaerobacterium aotearoense* GH1 | 103.8 | 740.39 | 0.66 | 25.45 | – | – | – | – | [41] |
| *Aspergillus niger* CCRC 31494 | 198.5 | 5.27 | 21.7 | – | – | – | – | – | [40] |

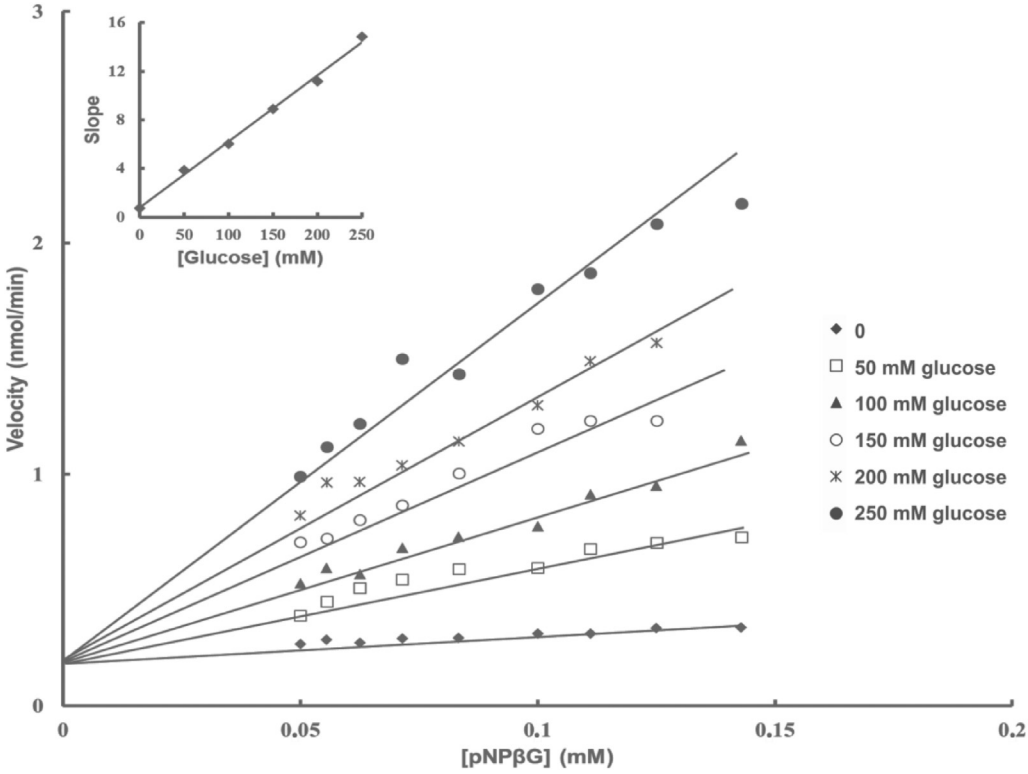



**Fig. 4.** Kinetic characterization of the enzymatic activity of AmBgl-LP in the presence of glucose. The reactions were carried out with increasing concentrations of pNPβG in the absence (control) and presence of fixed concentrations of glucose (ranging from 50 to 250 mM). The $K_i$ was calculated based on the slope of the Lineweaver-Burk lines *versus* [glucose] (insert). Enzyme activity was determined in 100 mM citric acid-sodium phosphate buffer pH 6.0 at 40 °C.

saccharification and fermentation (SSF). In this process, released glucose is concomitantly used by yeast for fermentation of the sugar obtained for the production of ethanol, resulting in reduced cost of production for the industries. Lin et al. [38], studied the thermotolerance and performance of the *Saccharomyces cerevisiae* BY4742 in the fermentation process for the production of ethanol in batch. This strain showed maximum performance for growth and ethanol production in temperatures between 30 and 45 °C. In contrast, fungal cellulases, which are routinely used in cellulose hydrolysis, have optimum temperatures around 50 °C [39]. Despite the optimum temperature for AmBgl-LP being 40 °C, 96% of its activity was observed at 35 °C and 72% at 30 °C, which signifies its application for industrial processes such as SSF.

The optimum pH of most β-glucosidases ranges from 4.0 to 7.5 [35]. The AmBgl-LP showed the same optimum pH value found for an enzyme isolated from a soil metagenome in the same region (pH 6.0) [32]. This pH is slightly higher than that (pH 5.0) observed in Lake Poraquê waters. However, for SSF process, the ideal range for ethanol production by *Saccharomyces cerevisiae* BY4742 was between pH 4.0 and 5.0 [38]. AmBgl-LP had maximum activity at pH 6.0, and 47% of activity at pH 5.0 (Fig. 2B). Despite a lower efficiency at pH 5.0, the enzyme could still be used in the SSF process.

Analyzing the thermostability data, it is evident that AmBgl-LP is highly stable up to 1-hour incubation at temperatures of 20, 25, and 30 °C (Fig. 2C). Therefore, it can be used preferably in the temperature range of 30 to 35 °C, where its activity is still relatively high. This characteristic is also recommended for SSF process.

Kinetic analysis of AmBgl-LP (Table 2) indicates a catalytic

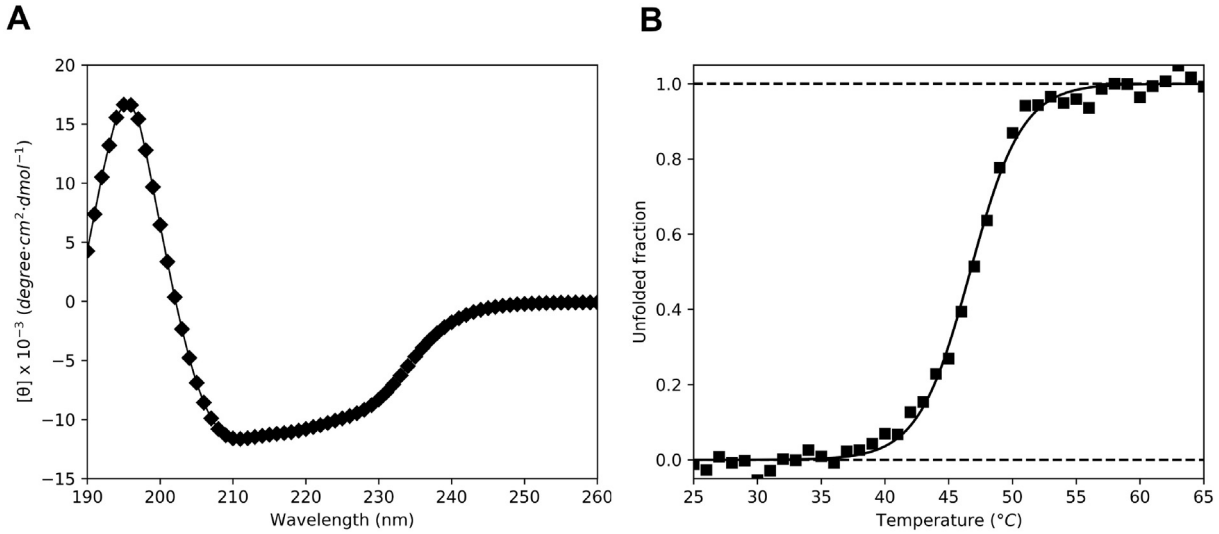



**Fig. 5.** Circular dichroism spectroscopy. (A) The far-UV CD spectrum of amBgl-LP. (B) The thermal denaturation curve obtained by monitoring the 210 nm wavelength. The unfolded fraction is shown as a function of temperature. The calculated value for $T_M$ was 46.69 °C.

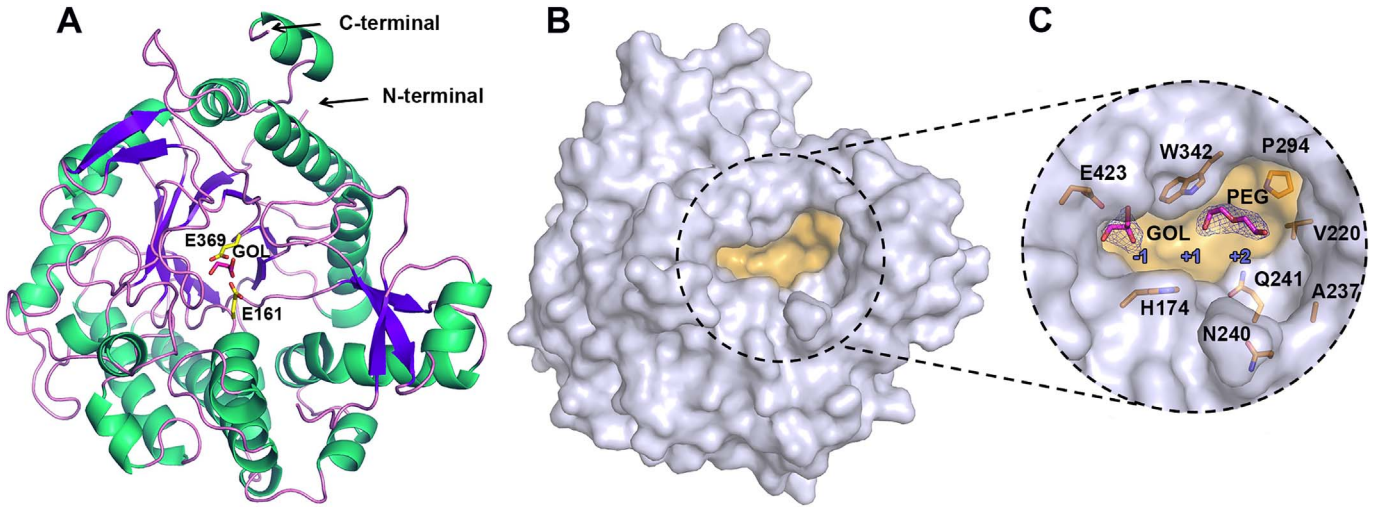



**Fig. 6.** A. Overall $(\beta/\alpha)_8$-barrel structure of AmBgl-LP monomer with the catalytic residues represented as sticks (carbon atoms in yellow). Glycerol (GOL) molecule is drawn as sticks with carbon atoms in pink. B. AmBgl-LP monomer surface with catalytic pocket in orange. C. Expansion of the active-site region with glycerol and PEG molecules shown with 2Fo-Fc electron-density maps contoured at the 1.0 σ level. Residues surrounding the catalytic site (carbon atoms in orange), glycerol (GOL) and PEG molecules (carbon atoms in pink) are shown. Subsites, inferred by structural comparisons, are identified by blue numbers, following the nomenclature described by Davies et al., 1997 [25].

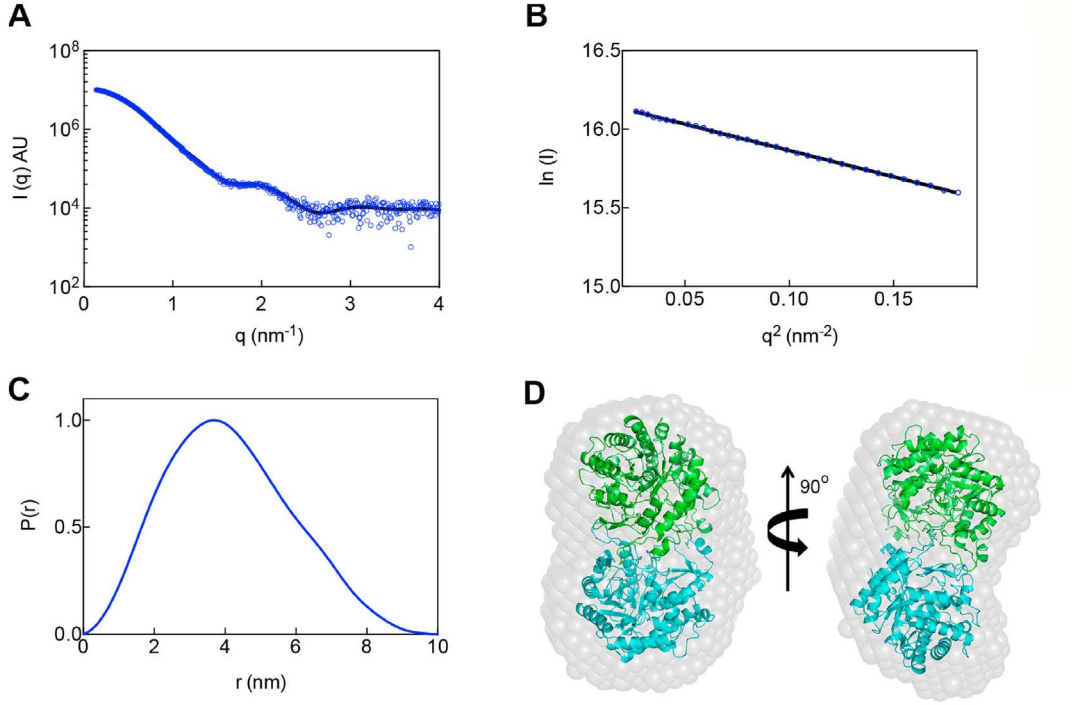



**Fig. 7.** SAXS analysis. A. Experimental SAXS curve (open circles) and theoretical scattering profile (line) of AmBgl-LP. B. Guinier region of SAXS curve showing the linear range (black line) used to estimate the radius of gyration ($R_G$). C. The distance distribution P(r) function of AmBgl-LP. D. Different views of the crystallographic structure of the dimer (green and blue ribbon) fitted into the averaged SAXS envelope (gray transparent surface).

efficiency ($k_{cat}/K_m$) for the hydrolysis of the synthetic substrate (pNPβG) 10-fold higher than that observed for the hydrolysis of cellobiose. We also compared kinetic parameters of AmBgl-LP with β-glucosidases from different metagenomes and with enzymes from different organisms (Table 2). The catalytic rate constant ($k_{cat}$) calculated for AmBgl-LP was compared to the $k_{cat}$ values for other isolated β-glucosidases. The $k_{cat}$ for AmBgl-LP was higher than for β-glucosidases isolated from a microbial compost [42] and that isolated from *Exiguobacterium antarcticum* B7 when pNPβG was used [43] (Table 2).

β-Glucosidases can be inhibited by their product. A search for glucose tolerant β-glucosidases is of great interest to the industries. Some glucose tolerant β-glucosidases have shown $K_i$ values of > 1 M [36,44]. The assays for glucose tolerance of AmBgl-PL revealed that it has a $K_i$ of 14 mM. Although AmBgl-LP is moderately tolerant to high glucose concentrations, it could be applied in the SSF process for the hydrolysis of cellulose, a process in which the released glucose is rapidly consumed by the yeast for ethanol production.

The AmBgl-LP preserves all the structural characteristics of a typical GH1- family member, including the active-site pocket organization and the retaining catalytic mechanism characteristics [23]. Interestingly, AmBgl-LP is stable as dimers in solution. Recently, in a study that involves two authors of the present work, it was observed that a GH1 β-glucosidase from *Exiguobacterium antarcticum* exhibits a tetrameric arrangement [45]. Other GH1-family 6-P-β-glucosidases, such as, from *Lactobacillus plantarum* (LpPbg1) and *Streptococcus mutans* (SmBgl), both sharing 31% of sequence identity with AmBgl-LP, seem to form

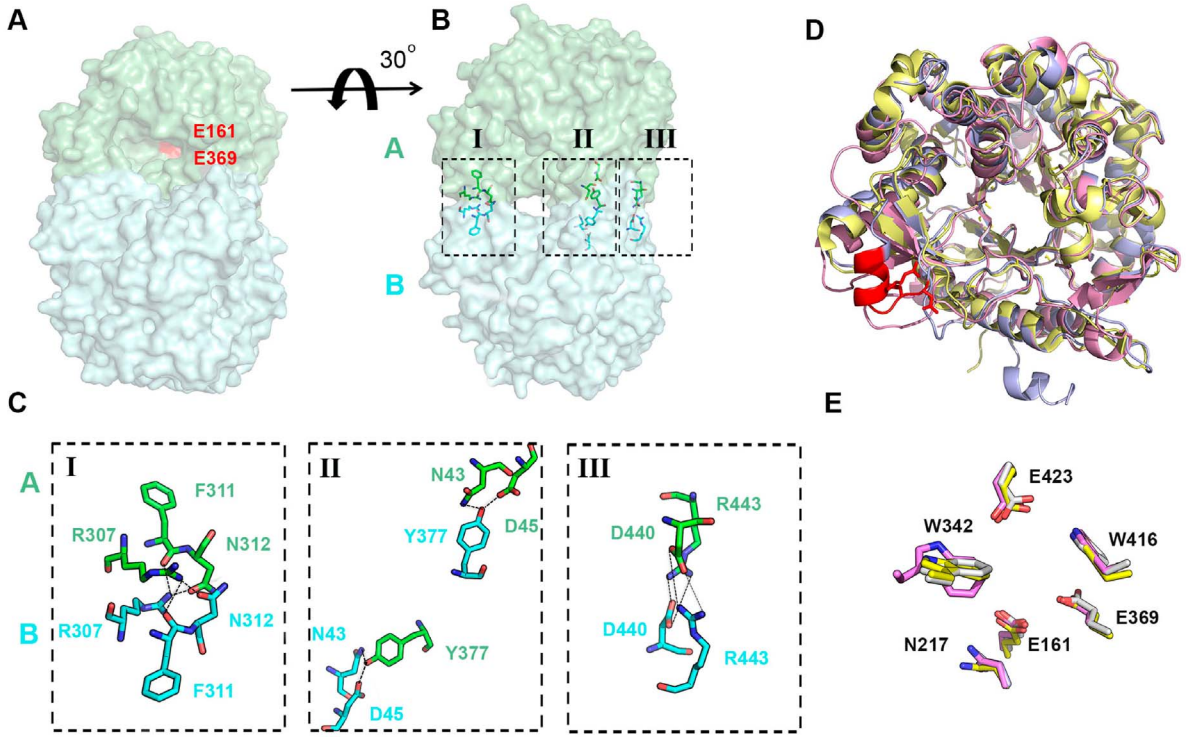



**Fig. 8.** A. Surface of dimeric structure of AmBgl-LP, composed by monomer A (light green) and B (light cyan), highlighting the catalytic residues (red) of monomer A. B. Surface of dimeric structure of AmBgl-LP, highlighting the interfacial regions of monomer A (carbons in green) and B (carbons in cyan). C. Residues of the regions I, II and III of AmBgl-LP dimeric interface. D. Structural alignment of AmBgl-LP (purple gray) with monomeric β-glucosidases from *Paenibacillus polymyxa* (PpBlgB; PDB code: 2O9P) (yellow) and from *Humicola insolens* (HiBgl; PDB code: 4MDP) (pink) The most divergent portion, composed by residues from the Region I, is represented in red. E. Structural alignment of conserved residues in the catalytic site of AmBgl-LP (purple gray) with PpBlgB (PDB code: 2O9P) (yellow) and HiBgl (PDB code: 4MDP) (pink). Residues numbers correspond to AmBgl-LP.

homodimeric assemblies in solution [46].

Due to the similarities found between the catalytic site of monomeric GH1 members and AmBgl-LP and the distance of dimeric interface from the catalytic site, it is unlikely that dimerization could interfere in the AmBgl-LP activity, glucose tolerance or transglycosylation capacity.

## 5. Conclusion

Metagenomes of rivers and lakes of the Amazon region have already been described [29–31]. However, it is the first time that an enzyme (β-glucosidase) from a sample of an Amazonian lake was isolated, recombinantly produced and characterized by this approach. Moreover, the three-dimensional structure of AmBgl-LP was determined. In this work, metagenome has shown to be a valuable approach to understand the structural diversity of GH1 enzymes and also to search for cellulases with desirable characteristics for the process of SSF, since the β-glucosidase isolated in this work has desirable features for this application.

## Conflict of interest

Fernando Pellon de Miranda is employed by Petróleo Brasileiro S.A. (Petrobras, in Rio de Janeiro, Brazil).

## Transparency document

The Transparency document associated with this article can be found, in online version.

## Acknowledgments

We are grateful to the Brazilian Biosciences National Laboratory (LNBio) and the Brazilian Synchrotron Light Laboratory (LNLS) for the provision of time on the MX2 and SAXS1 beamlines, ROBOLAB and LEC.

## Funding

This study was supported by Petróleo Brasileiro S.A. (Petrobras) as part of a research agreement (#0050.0081178.13.9) with the Federal University of São Carlos, SP, Brazil, within the context of the Geochemistry Thematic Network and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2015/26982-0). FHS, MTM and SRM are recipients of a Research Productivity Scholarship from the National Council for Scientific and Technological Development (CNPq #311745/2013-0, #306135/2016-7 and #303654/2015-5, respectively). SRM is a staff member of the Departamento de Bioquímica – Instituto de Química – USP. DT was the recipient of a doctoral scholarship from the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES). VMA is a graduate student supported by CAPES.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2018.02.001.

## References

[1] L.R. Lynd, P.J. Weimer, W.H. Van Zyl, I.S. Pretorius, Microbial cellulose utilization: fundamentals and biotechnology, Microbiol. Mol. Biol. Rev. 66 (2002) 506–577.
[2] R.C. Kuhad, R. Gupta, A. Singh, Microbial cellulases and their industrial applications, Enzym. Res. 2011 (2011).
[3] M.L. Maki, M. Broere, K.T. Leung, W. Qin, Characterization of some efficient cellulase producing bacteria isolated from paper mill sludges and organic fertilizers, Int. J. Biochem. Mol. Biol. 2 (2011) 146–154.
[4] A.A. Klyosov, Trends in biochemistry and enzymology of cellulose degradation, Biochemistry 29 (1990) 10577–10585.
[5] R.F.H. Dekker, Kinetic, inhibition, and stability properties of a commercial β-D-

Glucosidase (Cellobiase) preparation from *Aspergillus niger* and its suitability in the hydrolysis of lignocellulose, Biotechnol. Bioeng. 28 (1986) 1438–1442.
[6] Y. Bhatia, S. Mishra, V.S. Bisaria, Microbial β-glucosidases: cloning, properties, and applications, Crit. Rev. Biotechnol. 22 (2002) 375–402.
[7] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic Acids Res. 16 (1988) 10881–10890.
[8] V.K. Laemmli, Cleavage of structural protein during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[9] M.T. Davies, A universal buffer solution for use in ultra-violet spectrophotometry, Analyst 84 (1959) 248–251.
[10] F.K. Tamaki, E.M. Araujo, R. Rozenberg, S.R. Marana, A mutant β-glucosidase increases the rate of the cellulose enzymatic hydrolysis, Biochem. Biophys. Rep. 7 (2016) 52–55.
[11] H.I. Segel, Enzyme kinetics, Behavior and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems, John Wiley and Sons, NY, 1993.
[12] W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132.
[13] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Crystallogr. 30 (1997) 1022–1025.
[14] P.V. Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards automated crystallographic structure refinement with phenix refine, Acta Crystallogr. D Biol. Crystallogr. 68 (2012) 352–367.
[15] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D Biol. Crystallogr. 53 (1997) 240–255.
[16] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2126–2132.
[17] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality, Science 336 (2012) 1030–1033.
[18] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503.
[19] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886.
[20] V.V. Volkov, D.I. Svergun, Uniqueness of *ab initio* shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864.
[21] M.B. Kozin, D.I. Svergun, Automated matching of high- and low-resolution structural models, J. Appl. Crystallogr. 34 (2001) 33–41.
[22] E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2256–2268.
[23] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, G. Davies, Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases, Proc. Natl. Acad. Sci. 92 (1995) 7090–7094.
[24] T. Matsuzawa, T. Jo, T. Uchiyama, J.A. Manninen, T. Arakawa, K. Miyazaki, S. Fushinobu, K. Yaoi, Crystal structure and identification of a key amino acid for glucose tolerance, substrate specificity and transglycosylation activity of metagenomic beta-glucosidase td2f2, FEBS J. 283 (2016) 2340–2353.
[25] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559.
[26] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.
[27] P.O. Giuseppe, T.A.B. Souza, F.H. Souza, L.M. Zanphorlin, C.B. Machado, R.J. Ward, J.A. Jorge, R.P. Furriel, M.T. Murakami, Structural basis for glucose tolerance in GH1 β-glucosidases, Acta Crystallogr. D Biol. Crystallogr. 70 (2014) 1631–1639.
[28] P. Isorna, J. Polaina, L. Latorre-García, F.J. Cañada, B. González, J. Sanz-Aparicio, Crystal structures of *Paenibacillus polymyxa* beta-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases, J. Mol. Biol. 371 (2007) 1204–1218.
[29] R. Ghai, F. Rodriguez-Valera, K.D. Mcmahon, D. Toyama, R. Rinke, T.C.S. Oliveira, J.W. Garcia, F.P. Miranda, F. Henrique-Silva, Metagenomics of the water column in the pristine upper course of the Amazon river, PLoS One 6 (2011) e23785.
[30] D. Toyama, L.T. Kishi, C.D. Santos-Júnior, A. Soares-Costa, T.C.S. De Oliveira, F.P. De Miranda, F. Henrique-Silva, Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin, Genome Announc. 4 (2016) e01440-16, , http://dx.doi.org/10.1128/genomeA.01440-16.
[31] C.D. Santos-Júnior, L.T. Kishi, D. Toyama, A. Soares-Costa, T.C.S. Oliveira, F.P. De Miranda, F. Henrique-Silva, Metagenome sequencing of prokaryotic microbiota collected from rivers in the Upper Amazon Basin, Genome Announc. 5 (2017) e01450-16, , http://dx.doi.org/10.1128/genomeA.01450-16.
[32] J.C. Bergmann, O.Y.A. Costa, J.M. Gladden, S. Singer, R. Heins, P. D'haeseleer, B.A. Simmons, B.F. Quirino, Discovery of two novel β-glucosidases from an Amazon soil metagenomics library, FEMS Microbiol. Lett. 351 (2014) 147–155.
[33] S.G. Withers, R.A.J. Warren, I.P. Street, K. Rupitz, J.B. Kempton, R. Aebersold, Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase, J. Am. Chem. Soc. 112 (1990) 5887–5889.
[34] Q. Wang, D. Trimbur, R. Graham, R.A.J. Warren, S.G. Withers, Identification of the acid/base catalyst in agrobacterium faecalis β-glucosidase by kinetic analysis of mutants, Biochemistry 34 (1995) 14554–14562.
[35] J.R.K. Cairns, A. Esen, β-Glucosidases, Cell. Mol. Life Sci. 67 (2010) 3389–3405.
[36] Z. Fang, W. Fang, J. Liu, Y. Hong, H. Peng, X. Zhang, B. Sun, Y. Xiao, Cloning and characterization of a β-glucosidase from marine microbial metagenome with excellent glucose tolerance, J. Microbiol. Biotechnol. 20 (2010) 1351–1358.
[37] Q. Wang, C. Qian, X.-Z. Zhang, N. Liu, X. Yan, Z. Zhou, Characterization of a novel thermostable β-glucosidase from a metagenomics library of termite gut, Enzym. Microb. Technol. 51 (2012) 319–324.
[38] Y. Lin, W. Zhang, L. Chunkie, K. Sakakibara, S. Tanaka, H. Kong, Factors affecting ethanol fermentation using *Saccharomyces cerevisiae* BY4742, Biomass Bioenergy 47 (2012) 395–401.
[39] Z. Kádár, Z. Szengyel, K. Réczey, Simultaneous saccharification and fermentation (SSF) of industrial wastes for the production of ethanol, Ind. Crop. Prod. 20 (2004) 103–110.
[40] T.-R. Yan, C.-L. Lin, Purification and characterization of a glucose-tolerant β-glucosidase from *Aspergillus niger* CCRC 31494, Biosci. Biotechnol. Biochem. 61 (1997) 965–970.
[41] F. Yang, X. Yang, Z. Li, C. Du, J. Wang, S. Li, Overexpression and characterization of a glucose-tolerant β-glucosidase from *T. aotearoense* with high specific activity for cellobiose, Appl. Microbiol. Biotechnol. 99 (2015) 8903–8915.
[42] T. Uchiyama, K. Miyazaki, K. Yaoi, Characterization of a novel β-glucosidase from a compost microbial metagenome with strong transglycosylation activity, J. Biol. Chem. 288 (2013) 18325–18334.
[43] E. Crespim, L.M. Zanphorlin, F.H. De Souza, J.A. Diogo, A.C. Gazolla, C.B. Machado, F. Figueiredo, A.S. Sousa, F. Nóbrega, V.H. Pellizari, M.T. Murakami, R. Ruller, A novel cold-adapted and glucose-tolerant GH1 β-glucosidase from *Exiguobacterium antarcticum* B7, Int. J. Biol. Macromol. 82 (2016) 375–380.
[44] C. Riou, J.-M. Salmon, M.-J. Vallier, Z. Gunata, P. Barre, Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β-glucosidase from *Aspergillus oryzae*, Appl. Environ. Microbiol. 64 (1998) 3607–3614.
[45] L.M. Zanphorlin, P.O. De Giuseppe, R.V. Honorato, C.C. Tonoli, J. Fattori, E. Crespim, P.S. De Oliveira, R. Ruller, M.T. Murakami, Oligomerization as a strategy for cold adaptation: structure and dynamics of the GH1 β-glucosidase from *Exiguobacterium antarcticum* B7, Sci. Rep. 31 (2016) (6:23776).
[46] K. Michalska, K. Tan, H. Li, C. Hatzos-Skintges, J. Bearden, G. Babnigg, A. Joachimiak, GH1-family 6-P-β-glucosidases from human microbiome lactic acid bacteria, Acta Crystallogr. D Biol. Crystallogr. 69 (2013) 451–463.