



Article

Crystal Structure of a GH3 β -Glucosidase from the Thermophilic Fungus *Chaetomium thermophilum*

Imran Mohsin ¹, Nirmal Poudel ¹, Duo-Chuan Li ^{2,*} and Anastassios C. Papageorgiou ^{1,*}

¹ Turku Bioscience, University of Turku and Åbo Akademi University, 20520 Turku, Finland; dr.mohsinimran@hotmail.com (I.M.); nirpou@utu.fi (N.P.)

² Department of Mycology, Shandong Agricultural University, Taian 271018, China

* Correspondence: lidc20@sdau.edu.cn (D.-C.L.); anapap@utu.fi (A.C.P.); Tel.: +358-29-4503752 (A.C.P.)

Received: 29 October 2019; Accepted: 25 November 2019; Published: 27 November 2019



Abstract: Beta-glucosidases (β -glucosidases) have attracted considerable attention in recent years for use in various biotechnological applications. They are also essential enzymes for lignocellulose degradation in biofuel production. However, cost-effective biomass conversion requires the use of highly efficient enzymes. Thus, the search for new enzymes as better alternatives of the currently available enzyme preparations is highly important. Thermophilic fungi are nowadays considered as a promising source of enzymes with improved stability. Here, the crystal structure of a family GH3 β -glucosidase from the thermophilic fungus *Chaetomium thermophilum* (*CtBGL*) was determined at a resolution of 2.99 Å. The structure showed the three-domain architecture found in other β -glucosidases with variations in loops and linker regions. The active site catalytic residues in *CtBGL* were identified as Asp287 (nucleophile) and Glu517 (acid/base). Structural comparison of *CtBGL* with Protein Data Bank (PDB)-deposited structures revealed variations among glycosylated Asn residues. The enzyme displayed moderate glycosylation compared to other GH3 family β -glucosidases with similar structure. A new glycosylation site at position Asn504 was identified in *CtBGL*. Moreover, comparison with respect to several thermostability parameters suggested that glycosylation and charged residues involved in electrostatic interactions may contribute to the stability of the enzyme at elevated temperatures. The reported *CtBGL* structure provides additional insights into the family GH3 enzymes and could offer new ideas for further improvements in β -glucosidases for more efficient use in biotechnological applications regarding cellulose degradation.

Keywords: glycoside hydrolase; cellulose degradation; thermophilic fungus; β -glucosidases; *Chaetomium thermophilum*; protein structure; fungal enzymes

1. Introduction

Beta-glucosidases (β -glucosidases) are key cellulolytic enzymes that catalyze the hydrolysis of cellobiose to glucose. They are terminal enzymes of the cellulase system as they act during the final step of the cellulose degradation. Although beta-glucosidases do not act directly on cellulose, their activity is important to circumvent the inhibitory effect of cellobiose on endoglucanases and cellobiohydrolases. Consequently, inclusion of β -glucosidases in cellulase preparations can synergistically increase the hydrolytic efficiency of other cellulolytic enzymes [1].

Beta-glucosidases (β -glucoside glucohydrolases, EC 3.2.1.21; BGL) have been found in the glycoside hydrolase (GH) families GH1, GH3, GH5, GH9, GH30, and GH116 of the CAZY database [2]. They are anomeric configuration-retaining enzymes that operate through the canonical double-displacement glycosidase mechanism, except those of the GH9 family.

Despite their extensive use in biotechnological application, structural data on fungal β -glucosidases are still scarce. Many fungal BGLs are classified as part of the GH3 family, one of the largest CAZY

families with over 13,600 annotated protein sequences. Two fungal β -glucosidase structures currently available are produced by the mesophilic fungus *Aspergillus aculeatus* [3] and the moderate thermophilic *Rasamsonia emersonii* [4], respectively. A third structure produced by the mesophilic fungus *Trichoderma reesei* of BGL1 has been deposited in the Protein Data Bank (PDB), but no paper has been published yet. Recently, the structure of the filamentous fungus *Neurospora crassa* Cel3A (NcCel3A) was reported at 2.25 Å resolution [5], as well as structures of GH3 β -glucosidases from *Hypocrea jecorina* (a teleomorph of *Trichoderma reesei*) [6], *Aspergillus niger* [7], *Aspergillus fumigatus*, and *Aspergillus oryzae* [8].

The first structure of a β -glucosidase, that of barley (*Hordeum vulgare*) exo-1,3-1,4-glucanase ExoI (HvExoI) [9] showed a two-domain architecture: An N-terminal $(\alpha/\beta)_8$ TIM-barrel that hosts the catalytic nucleophile aspartate and an $(\alpha/\beta)_6$ sandwich domain where the acid/base glutamate resides. Subsequent studies, however, revealed variations in the modular architecture of GH3 β -glucosidases with the first domain to adopt a collapsed/incomplete triosephosphate isomerase (TIM)-barrel domain and the presence of a third domain with a fibronectin FnIII-like fold. The function of the third domain is still unknown, although it has been suggested to stabilize the incomplete TIM-barrel domain [4]. Notably, a fourth domain has also been identified in *Kluyveromyces marxianus* and *Pseudoalteromonas* sp. BB1 β -glucosidases [10,11].

Thermophilic fungi have attracted considerable attention in recent years as an alternative reservoir of thermostable cellulases for cellulose degradation [12,13]. Importantly, thermophilic fungi can produce thermostable enzymes that can be used at temperatures up to 70 °C, whereas enzymes from mesophilic organisms are typically active up to 50 °C. The importance of using thermophilic cellulases in cellulose degradation stems from the fact that, at higher temperatures, cellulose swells and becomes more susceptible to breaking. Various thermophilic fungi have been studied in recent years and their β -glucosidases have been characterized [12]. Understanding the structure–function–stability relationships in fungal β -glucosidases is therefore important for finding new and better alternatives for industrial biocatalysts. Hydrolysis rate, inhibitors, and stability are considered critical factors for the efficient use of β -glucosidases in complex biomass hydrolysis [14]. Moreover, β -glucosidases have been suggested for use in the synthesis of various glycoconjugates, and a GH3 β -glucosidase from the thermophilic fungus *Myceliophthora thermophila* has been found to act as an efficient biocatalyst in alkyl glycoside synthesis [15].

Here, we report the crystal structure of a β -glucosidase from the thermophilic filamentous fungus *Chaetomium thermophilum* (CtBGL) at a resolution of 2.99 Å, and compare it with other β -glucosidases of the GH3 family. This is the second structure of a β -glucosidase produced by a thermophilic fungus, and therefore, it is expected to provide further insights into the structure–function relationship of this family of enzymes in high-temperature settings.

2. Results

2.1. Quality of the Structure

The crystal structure of CtBGL was refined using data up to 2.99 Å resolution. The final R_{cryst} and R_{free} (5% of the reflections excluded from refinement) were 0.201 and 0.252, respectively (Table 1). The structure exhibits good stereochemistry despite the limited resolution with root mean square deviation (rmsd) in bond lengths and bond angles of 0.007 Å and 0.94°, respectively. The Ramachandran plot shows 92.8% of the residues in the most favorable regions and 1.0% in disallowed regions. Residues that fall into the disallowed regions belong to flexible parts of the structure with weak electron density. Structure quality statistics for CtBGL fall within the distribution found in other crystal structures of similar resolution as analyzed and displayed by POLYGON [15]. The refined structure contains 836 residues of the mature protein, starting at residue Trp32. The full-length cDNA encodes for a 867-residue enzyme and the first 16 residues have been suggested to act as a secretion signal peptide [16]. Residues 17–31 at the N-terminal were not modeled, owing to lack of adequate electron density, possibly due to high flexibility. The enzyme crystallized with one molecule in the

crystallographic asymmetric unit in contrast to other β -glucosidases that crystallize either with two molecules such as the β -glucosidases from *Aspergillus oryzae* ($Ao\beta G$) and *A. fumigatus* ($Af\beta G$) [8], *A. aculeatus* ($Aa\beta G$) [8] and *NcCel3A* [5], or with four molecules in the asymmetric unit such as *ReCel3A* [4]. The solvent content is unusually high (~77%) for a protein crystal, which could explain the low resolution of the data and the disintegration of the crystals during manipulation prior to mounting. The NCS dimer found in other β -glucosidases is seen in *CtBGL* as a crystallographic dimer that involves the $-Y, -X, -Z + \frac{1}{2}$ symmetry-related molecule.

Table 1. Data collection and refinement statistics.

Data Collection	
Beamline	EMBL-DESY X13
Wavelength (Å)	0.8123
Data Processing	
Space group	$P4_12_12$
Unit cell dimensions a, b, c (Å)	121.9, 121.9, 264.9
No. of molecules/asymmetric units	1
Mosaicity (°)	0.070
Resolution range (Å)	20.0–2.99 (3.11–2.99) *
Total measurements/unique reflections	222508 (22974)/41051 (4513)
Completeness (%)	99.3 (98.6)
$\langle I/\sigma(I) \rangle$	7.8 (1.4)
R_{meas}	0.287 (1.452)
R_{pim}	0.122 (0.846)
$CC_{1/2}$	0.978 (0.492)
Wilson B-factor (Å ²)	46.2
Refinement	
Reflections (working/test)	38861/2047
R_{cryst}/R_{free}	0.201/0.249
No. of protein atoms	6457
Ligand atoms	12
No. of Sugar Atoms	355
No. of Water Molecules	6
Geometry	
rmsd in bond lengths (Å)	0.007
rmsd in bond angles (°)	0.94
Most favorable regions (%)	92.8
Additional allowed regions (%)	6.2
Outliers (%)	1.0
Clashscore	10.4
Average B-Factors (Å²)	
Protein	50.4
Ligand	63.8
Sugars	72.4
Water molecules	32.8
Pyranose Conformations (Total/Percentage)	
Lowest energy conformations	23/85.2
Highest energy conformations	4/14.8

* Values for the outermost resolution shell are shown in parentheses.

2.2. Overall Structure Comparisons

Structure-based sequence alignment (Figure 1) showed that *CtBGL* has the highest structural similarity with *NcCel3A* (PDB id 5nbs), as indicated by the root mean square deviation (rmsd) of 0.60 Å for 831 aligned residues (sequence identity 72%), followed by *Aspergillus fumigatus AfβG* (PDB id 5fji; rmsd 0.81 Å; seq. identity 62%), *ReCel3A* (PDB id 5ju6; rmsd 0.84 Å; 62.0%) and *Aspergillus aculeatus AaBgl1* (4iig; 0.83 Å; 62.0%). Lower similarities were found with *Hypocrea jecorina HjCel3A* (3zyz; 1.21 Å; 47%).



Figure 1. Structure-based sequence alignment of *CtBGL* with members of GH3 family. Secondary structure elements are shown on the top of the alignment, while the red triangles below the alignment indicate the catalytic conserved residues (Asp287 and Glu517 in *CtBGL*). Disulfide bonds are depicted with green numbers. The three domains are indicated. A column is framed if more than 70% of its residues are similar according to physicochemical properties. Frames in red background with white letters depict strict identity. The figure was constructed with ESPript [16].

2.3. Description of the Structure

The structure of CtBGL (Figure 2) consists of three distinct domains, similar to other GH3 family members: A catalytic triose phosphate isomerase (TIM) barrel-like domain (Leu51–Ser354), an α/β sandwich domain (His396–Gly596) and a FnIII (fibronectin type III) domain (Thr663–Gln867) with a prominent insertion region (677–766). A linker region (residues 355–394) connects domain 1 to domain 2 and a second linker region (600–662) connects domain 2 to domain 3. The 39-residue linker between domain 1 and domain 2 is comparable to that in *NcCel3A* and significantly longer than the linker in other GH3 β -glucosidases such as *HjCel3A* and *HvExoI* with only 18 and 16 residues in length, respectively. In *NcCel3A*, the linker comprises 42 residues (residues 341–383) with a 25-residue (residues 351–376) insertion previously described as a hydrophobic linker responsible for the activation of *ReCel3A* and *AaBgl* in organic solvents [5]. This insertion has been coined as loop II and contains several aromatic residues, of which CtBGL Phe364 (Phe352 in *NcCel3A*) and CtBGL Trp367 (Trp355 in *NcCel3A*) are the most conserved. Moreover, loop II contains residues Phe366, Trp367, and Trp377 that line up one side of the substrate-binding site, and they are also found in *NcCel3A* (Phe354, Trp355, and Trp365, respectively) with variations in other β -glucosidases, such as *ReCel3A* and *AaBgl1*. It has been suggested that this loop is stabilized by interactions with the N-glycans from a neighboring glycosylated Asn residue (Asn57 in *NcCel3A*) [5]. In CtBGL, the equivalent residue is Asn72, which is also glycosylated (see below).

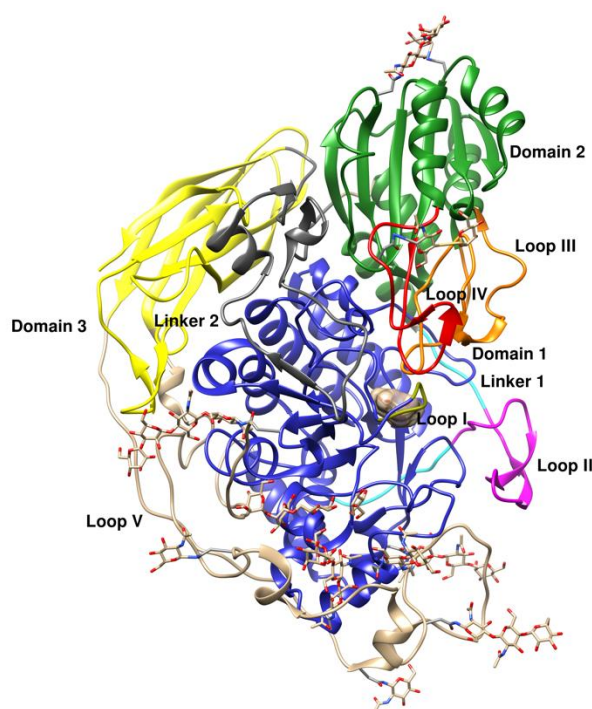


Figure 2. Overall crystal structure of CtBGL in ribbon representation. The domains, linkers, and loops are shown in different colors and labeled: Catalytic TIM barrel-like domain (blue), α/β sandwich domain (green), FnIII domain (yellow), loop I (gold), linker 1 (cyan), loop II or insertion region (magenta), loop III (orange), loop IV (orange–red), linker 2 (gray), loop V (brown). The N-glycans and glucose in the active site are shown as gray sticks. The figure was created with UCSF Chimera [17].

CtBGL second domain, an $(\alpha/\beta)_6$ sandwich fold, is structurally well conserved among all the GH3 enzymes. It consists of residues His396 to Gly596 and includes loops III and IV that encompass residues Gly433–Val467 and Ala513–Asn536, respectively (Figure 3). Loop III houses the conserved cysteine residues Cys442 and Cys447, which form a disulfide bridge that stabilizes the folded loop III architecture. Ser458 and Asp444 are also highly conserved in all compared structures. This domain

hosts the catalytic acid Glu517 in loop IV, which is found to be conserved in all of the compared β -glucosidases, *CtBGL*, *NcCel3A*, *ReCel3A*, *HjCel3A*, *AaBgl1*, and *Thermotoga neapolitana* β -glucosidase 3B (*TnBgl3B*). Moreover, the architecture and location of loops III and IV, which constitute one site of the active site cavity in between loops I and II, are also found conserved. Ser458, Asp444, and Glu517 are found pointing downwards from the loop regions and towards the active site.

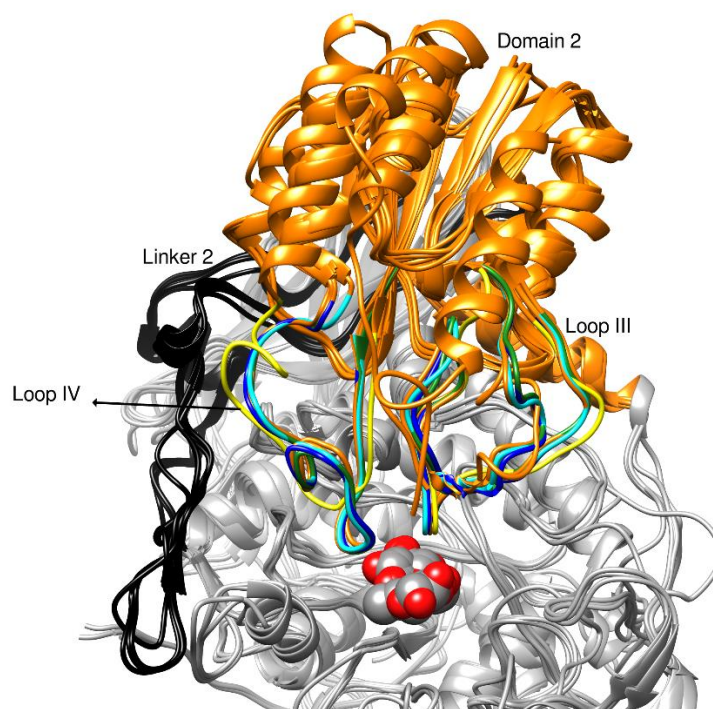


Figure 3. Overview of domain 2 in ribbon representation *CtBGL* superimposed with the compared structures. Domain 2 is colored orange; loops III and IV in *CtBGL*, *NcCel3A*, *ReCel3A*, and *TnBgl3B* are colored in green, blue, cyan, and yellow, respectively, for structural comparisons. Linker 2 loop is colored in black, and the active site is depicted by the bound BGC in sphere representation.

Domain 2 of the *CtBGL* structure is followed by a second linker region (Linker 2; residues Lys597–Ser660). This linker region exhibits an extended structure, which almost resembles a boundary that separates domain I and domain II from domain III. The extended linker region probably plays a role in stabilizing loops I and IV.

The FnIII-like domain 3 was first observed in the GH3 family in the *TnBgl3B* structure [18]. It consists of residues Tyr661–Asn867 that form a beta sandwich composed of a total of nine β -strands arranged in two layers of β -sheets with three and four β -strands, respectively. Loop V takes an extended structure and encompasses domain I. Importantly, loop V is present in *NcCel3A*, *ReCel3A*, and *AaBgl1*, but it is absent in *TnBgl3B* and *HjCel3A* (Figure 4). Several conserved aromatic residues, namely, Tyr716, Tyr718, Tyr733, and Phe740, are found in loop V. Tyr716 and Tyr733 are found to be conserved in all structures except *HjCel3A* and *TnBgl3B*, where loop V is absent, while Tyr718 is replaced by Trp in *ReCel3A* and *AaBgl1*. Moreover, Phe740 is replaced by a Tyr residue in *ReCel3A* and a His residue in *AaBgl1*. These conserved aromatic residues form π -stack interactions with *N*-acetyl- β -D-glucosamine (GlcNAc) residues. Also, conserved in loop V is Asn720, which was found to be N-glycosylated.

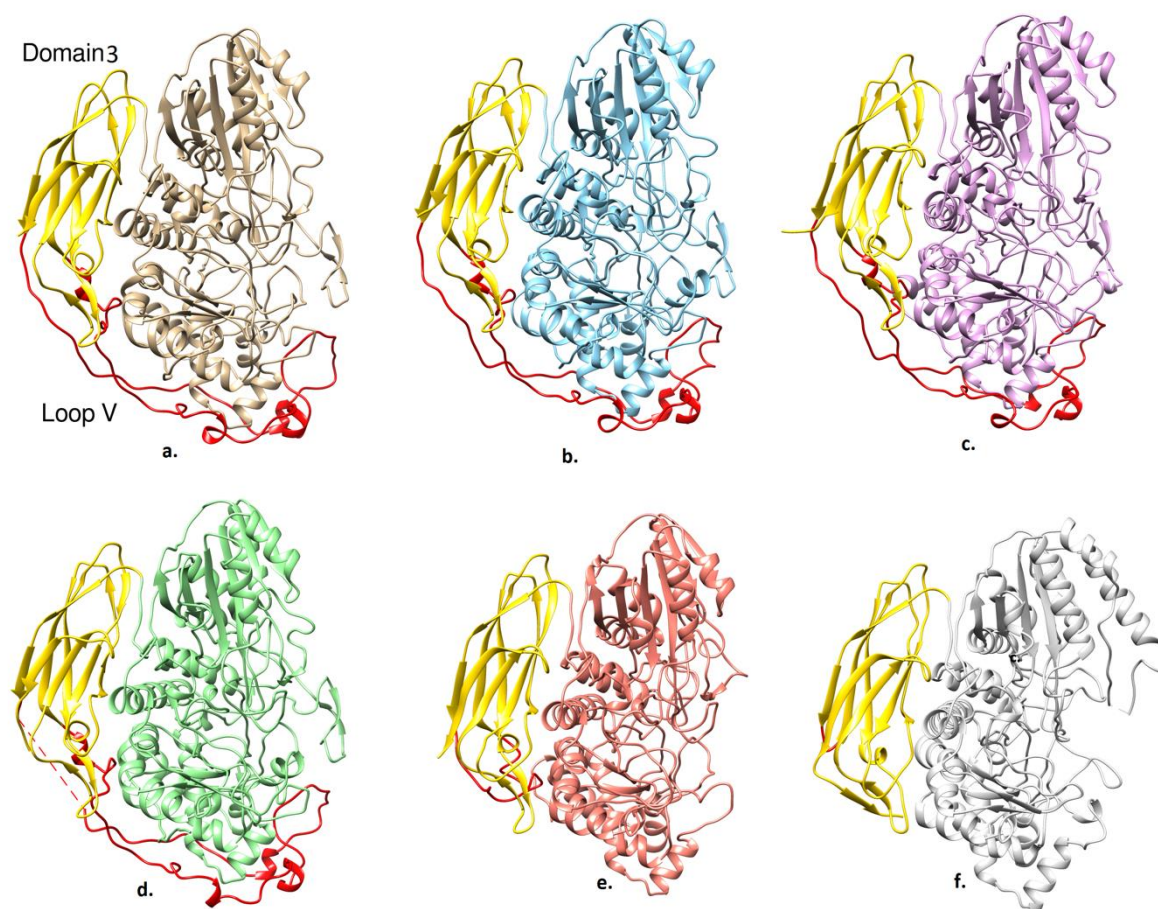


Figure 4. Loop V comparisons in (a) *CtBGL*, (b) *NcCel3A*, (c) *ReCel3A*, (d) *AaBgl1*, (e) *HjCel3A*, and (f) *TnBgl3B*. Domain 3 is colored in yellow and the loop V is represented by red coloration. Domains 1 and 2 are colored in the same color in each enzyme and differently amongst the enzymes.

2.4. Glycosylation Sites

Potential glycosylation sites were observed in the *CtBGL* structure in 10 positions, all in Asn residues (Table 2). The glycosylation sites found in *CtBGL* when compared with other structures showed some variations. A unique glycosylation site was observed in *CtBGL* at Asn504 on the surface of the molecule, where the corresponding aligned structures had either a different residue or a gap in the structure-based alignment (Table 3). Asn259 was found highly conserved and glycosylated in all structures. Moreover, *ReCel3A* possess 16 glycosylation sites [4], similarly to *Aspergillus* β -glucosidases [8]. On the other hand, *HjCel3A* showed the lowest number of glycosylation sites, with only two glycosylated sites (Asn208 and Asn310 in chains A and B, respectively). As observed before in this class of β -glucosidases, all glycosylation sites in *CtBGL* were also located on one face of the molecule (Figure 5).

A total of 27 glycan moieties were found in *CtBGL*. The glycans ranged in length from single GlcNAc to longer chain. The longest glycosylation chain was composed of eight residues. The overall degree of glycosylation in *CtBGL* can be considered as moderate when compared against other glycosylation chains, such as those found in *ReCel3A* and *AaBGL1*, with the longest chain composed of 10 residues and 45–50 glycosylation residues in total per chain in the crystallographic asymmetric unit. Interestingly, the glycosylation pattern in *CtBGL* showed high GlcNAc-type N-glycans, while *ReCel3A* and *AaBGL1* displayed high mannose-type N-glycans. A total of 15 GlcNAc residues were identified in *CtBGL*. Out of the total 10 glycosylation sites in *CtBGL*, three of them were found to consist of a single GlcNAc monosaccharide. Single GlcNAc monosaccharides in other enzymes,

such as *AaBGL1*, were obtained possibly as a result of a treatment with endoglycosidase H prior to crystallization. In the case of *CtBGL*, no endoglycosidase treatment was employed. Similarly, single GlcNAc molecules were found in *AfβG* (in Asn543 and Asn715) without any attempt of enzymatic cleavage. It is possible that, in these cases, the enzymes were subjected to endoglycosidase activity during the expression stage. The other seven N-glycosylation sites in *CtBGL* ranged from two to eight monosaccharides. The N-glycan at *CtBGL* Asn329 was located in domain I of the enzyme and consisted of eight monosaccharides (two GlcNAc, five α -D-Man, and one β -D-Man). Extra densities that could accommodate another two monosaccharide molecules were present in the electron-density map, but they were not modeled owing to lack of clarity. It was, nevertheless, the largest N-glycan structure in *CtBGL* and was involved in forming multiple H-bonds with the enzyme residues at Thr293, Asp730, Val295, Gly734, and Tyr733 and a pi-sigma interaction with Tyr733 residue. The structural equivalent Asn residue in *AaBgl1* (Asn322) had a long-length glycan with eight Man and two GlcNAc that were also involved in extensive stabilizing interactions with domain I. N-glycans at positions Asn72 and Asn720 were located in domains I and III, respectively, with loop V residues in between them. They both participated in interactions with protein residues. The N-glycan at Asn72 was found to interact in conventional H-bonding interactions with Ala90, Tyr716, Tyr718, and in pi-sigma interaction with Tyr715 from domain III. The N-glycan at Asn720 interacted with Arg710 via H-bonding and Tyr704 via weak van der Waals interactions. A long chain glycan moiety at Asn259 at the outer surface of the enzyme was found to interact via H-bonding with Ser35, Glu36, and Asp229.

Notably, the protein glycans were also shown to exhibit a potential binding affinity for polysaccharides such as cellulose [19] and aromatic compounds [20,21]. This suggests a possible role of N-glycans in promoting cellulose binding and also in protein–glycan interactions. Glycans in β -glucosidases can stabilize the crystal packing contacts and also participate in hydrogen-bonding interactions at the dimer interface [8]. There are two glycosylated Asn residues at the dimer interface of β -glucosidases that provide protein–glycan interactions between the two chains. One of those Asn residues corresponds to Asn531 in *CtBGL*. This Asn bears only two monosaccharide molecules, whereas, for comparison, the equivalent Asn residue in *AaGL1* has seven monosaccharides with the terminal ones able to reach the adjacent subunit. The second Asn residue that participates in dimer interface contacts is not present in *CtBGL*. The lack of these contacts could therefore contribute to reduced strength of intermolecular interactions in the crystal lattice, resulting in further instability of the *CtBGL* crystals.

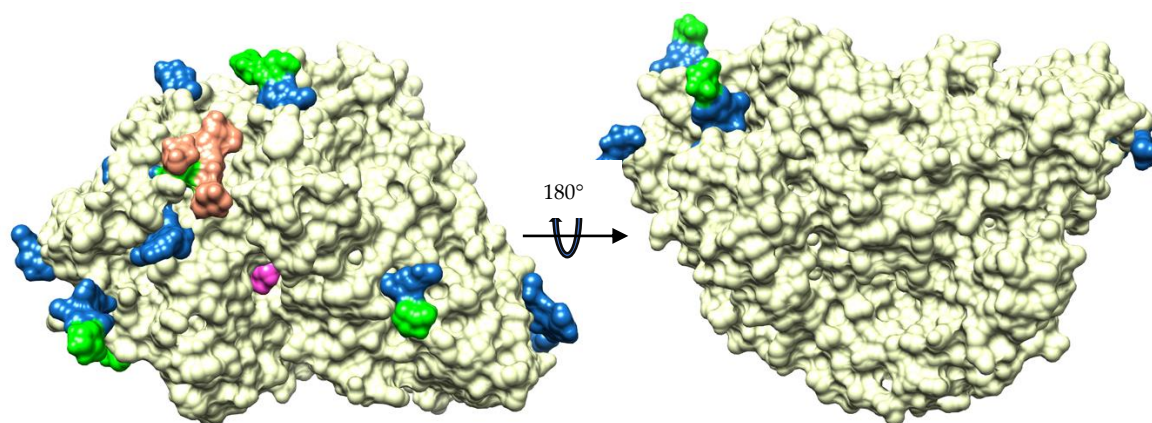


Figure 5. Distribution of glycosylation sites in the *CtBGL* structure. GlcNAc is shown in blue, β -D-mannose in green, α -D-mannose in orange, and BGC in magenta.

Table 2. Glycosylation sites and glycan descriptions in *CtBGL*.

Residue	Corresponding Glycan Structure
Asn72	Man- β 1,4-GlcNAc- β 1,4-GlcNAc
Asn259	Man- β 1,4-Man- β 1,4-GlcNAc- β 1,4-GlcNAc
Asn322	GlcNAc- β 1,4-GlcNAc
Asn329	Man- α 1,2-Man- α 1,6-(α 1,3-Man)- α 1,6-Man-(α 1,3-Man)Man- β 1,4-GlcNAc- β 1,4-GlcNAc
Asn504	GlcNAc
Asn531	Man- β 1,3-GlcNAc
Asn572	GlcNAc- β 1,4-GlcNAc
Asn698	GlcNAc
Asn720	Man- β 1,4-GlcNAc- β 1,4-GlcNAc
Asn744	GlcNAc

Table 3. Comparison of glycosylation sites %.

<i>CtBGL</i>	<i>NcCel3A</i> (5nbs)	<i>AaBgl1</i> (4iif)	<i>ReCel3A</i> (5ju6)	<i>HjCel3A</i> (3zyz) #
Asn72 (+)	Asn57 (+)	Asn61 (+)	Asn61 (+)	Gly30 (-)
Asn259 (+)	Asn248 (+)	Asn252 (+)	Asn249 (+)	Asn208 (+)
Asn322 (+)	Asn311 (+)	Asn315 (+)	Asn312 (+)	Ala270 (-)
Asn329 (+)	Asn318 (+)	Asn322 (+)	Asn319 (+)	Ser277 (-)
Asn504 (+)	Asp492 (-)	Ala496 (-)	Ala492(-)	Lys428 (-)
Asn531 (+)	Asn519 (+)	Asn523 (+)	Asn519 (+)	Asn455 (-)
Asn572 (+)	Asn560 (+)	Asn564 (+)	Asn560 (+)	Gln496 (-)
Asn698 (+)	Ser687 (-)	Asn690 (+)	Asn685 (+)	
Asn720 (+)	Asn710 (+)	Asn712 (+)	Asn707 (+)	
Asn744(+)	Lys734 (-)	Gly736 (-)	Asn731 (+)	

% (+) depicts glycosylation, (-) absence of glycosylation # In subunit A.

2.5. Active Site

The active site is located in a shallow pocket near the interface of the first and second domain. A molecule of β -D-glucose (BGC) was fitted at the active site based on residual electron density observed in electron density $F_o - F_c$ difference maps. The source of the β -D-glucose is most likely the growth medium used as no β -D-glucose was used during crystallizations or soaking. Two catalytic residues were identified, Asp287 (nucleophile) at the N-terminal TIM-barrel domain and Glu517 (acid/base) at the sandwich α/β domain II. Both catalytic residues were found conserved in the GH3 family members (Figure 6). The corresponding catalytic residues were Asp276/Glu505 in *NcCel3A*, Asp277/Glu505 in *ReCel3A*, Asp280/Glu509 in *AaBgl1*, Asp236/Glu441 in *HjCel3A*, and Asp242/Glu458 in *TnBgl3B*. The collapsed TIM-barrel model of domain I is vital for the proper accession of the active site. It was found that near the active site, the second barrel β strand (Gly87 to Thr89) was much shorter and antiparallel, which creates an active site much wider and accessible when compared with GH3 enzyme structures with complete TIM-barrel fold [6]. An additional electron density present at the active site next to BGC was not interpretable and may suggest a bound buffer molecule, MPD from the crystallization mother liquor, or a partially bound glucose molecule.

2.6. *CtBGL* Thermostability

CtBGL has been found to be thermostable at 50 °C and to retain half of its activity after incubation at 65 °C for 55 min [22]. The enzyme also retains 29.7% of its activity after incubation at 70 °C for 10 min. In regard to most other β -glucosidases from thermophilic fungi, *CtBGL* exhibits comparable thermostability, as previously reported [22]. In contrast, *A. fumigatus* β -glucosidase has been found highly thermostable and able to retain most of its activity for at least 19 h at 65 °C [23].

Protein thermostability is usually hard to predict and there is no a common mechanism yet available [24,25]. Several factors of protein thermostability have been proposed that could provide some clues (Table 4). Solvent-accessible surface (SAS), charged residues, and glycosylation patterns are

some of the key indicators. The structure of *HjCel3A* from the mesophilic fungus *Hypocrea jecorina* had the lowest SAS (22812 Å²), owing to the smaller number of residues and the lack of loop V. Similarly, loop V was absent in *TnBgl3A* and the SAS value was reduced. The SAS values for the other β -glucosidases, including *CtBGL*, were quite similar, i.e., around 28100 Å².

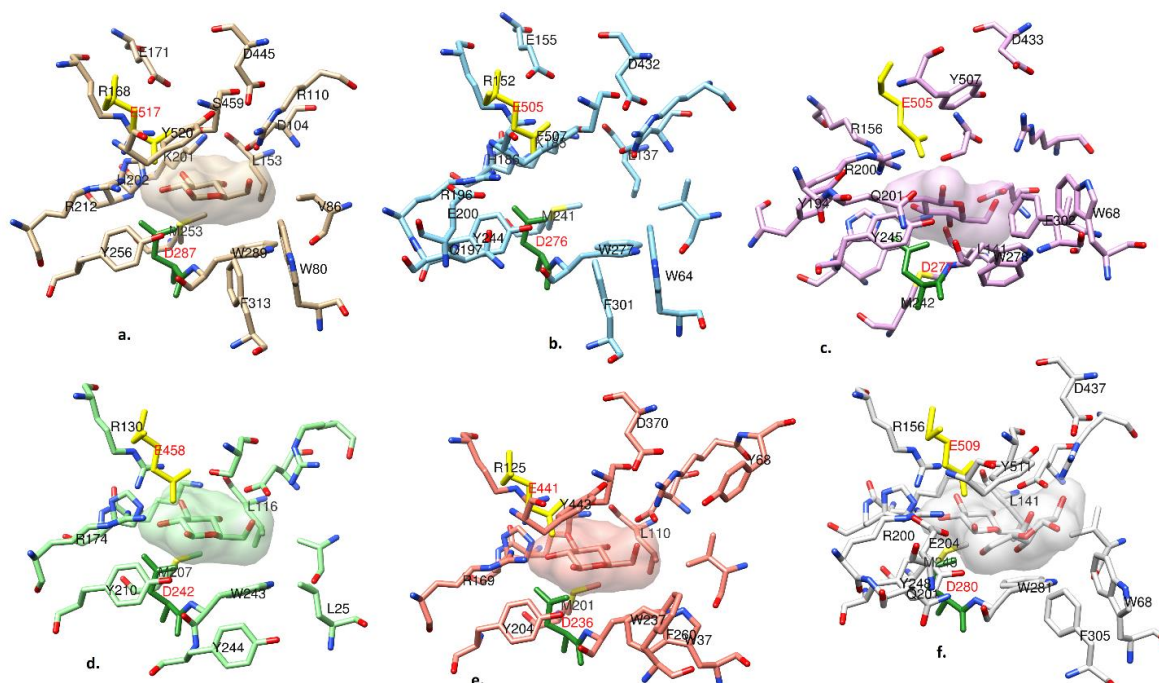


Figure 6. Active site structure comparisons of (a) *CtBGL*, (b) *NcCel3A*, (c) *ReCel3A*, (d) *TnBgl3B*, (e) *HjCel3A*, and (f) *AaBGL1*. The corresponding conserved nucleophile Asp is colored in dark green and the acid/base Glu is colored in yellow. Both residues are labeled in red. Carbon atoms are colored differently in each enzyme. The active site BGC is displayed with its solvent-excluded surface for visual clarity.

Charged residues in the structures can contribute to structural integrity and, in turn, to thermostability. The negatively (Asp and Glu) and positively (Arg and Lys) charged residues may provide a stability profile of the structure [26]. The numbers showed significant variations when the proteins were compared. In particular, *HjCelA* showed reduced numbers of positively and negatively charged residues, despite its high thermostability with an optimum temperature at 90 °C and an unfolding temperature of ~88 °C depending on the enzyme concentration [27]. The number was increased for the β -glucosidases from other thermophilic and mesophilic β -glucosidases. In addition, *Af β G*, despite its high thermostability, was also characterized by a similar content of charged residues, suggesting that the strength of individual ion-pair interactions may play a key role.

Finally, the glycosylation pattern found may also contribute to the thermostability of the enzyme by promoting interactions with amino acid residues. It has been shown that glycosylation enhances the solubility, reduces the aggregation, and increases the thermal stability of proteins [28,29]. The exact mechanism by which the enzyme glycosylation pattern affects the overall function and structure of proteins is not yet well understood. Analysis of protein structures deposited in the Protein Data Bank has suggested that N-glycosylation causes no significant local or global structural changes; however, it decreases the protein dynamics, thus leading to increased stability [26]. Aglycosylated proteins are, in general, found to be less stable, and therefore aggregate more easily than its glycosylated counterparts at certain temperatures. Thus, glycosylation has been suggested as a potential factor to enhance enzyme solubility, stability, and function [30]. Thermostability measurements in *HjCel3A* samples with different degree of N-glycosylation revealed the same melting temperature (74.0 \pm 0.2 °C),

thus suggesting that the effect of glycosylation may be case-specific and other factors could play a role. Interestingly, *TnBgl3B* that lacks glycosylation exhibited thermostability, most likely as a result of the high numbers of charged residues. Nevertheless, the moderate glycosylation in *CtBGL* could not be ruled out as a contributing factor to the limited stability of the enzyme compared to other more thermostable β -glucosidases characterized by extensive glycosylation. Further studies are, however, required to better understand the role of glycosylation in β -glucosidases and in protein stability.

Table 4. Comparison of thermostability parameters.

	<i>CtBGL</i>	<i>AfβG</i> (5FJI)	<i>AaBgl1</i> (4IIF)	<i>ReCel3A</i> (5JU6)	<i>TnBgl3B</i> (2X40)	<i>HjCel3A</i> (3ZYZ)
Thermal stability characteristics	$T_{1/2}$ 65 °C after 55 min	19 h at 65 °C [23]	$T_{1/2}$ 62 °C after 30 min [31]	T_m 87.3 °C [4]	Unfolding temperature 86–89 °C [27]	T_m 77.6 °C [4]
Asp + Glu (–)	93	95	92	87	113	55
Arg + Lys (+)	66	67	65	59	93	51
Pro: Gly	0.51	0.55	0.51	0.61	0.58	0.52
Val (%)	7.0	7.9	8.0	7.5	9.2	8.8
Amino acid residues	867	844	841	857	721	714
SAS (Å ²)	28,502.9	28,092.1	28,074.9	28,481.7	26,430.0	22,551.0
Intra-chain salt Bridges	46	49	42	44	39	34
H-bonding interactions	12 [∞]	22	25	22	23	8
Interface area, Å ²	1544.2 [∞]	1324.3	1445.3	1322.5	1253.1	699.3
Helical content (%)	21.2	21.8	21.8	20.5	26.5	23.7
Sugars	27	45 (chain A)	44 [%]	50 (chain C)	NA [§]	2

[∞] In the crystallographic dimer. [%] Prior to crystallization, the enzyme was subjected to endoglycosidase H treatment with ~50% of the N-glycans removed [3]. [§] The enzyme was produced in *Escherichia coli*.

3. Materials and Methods

3.1. Protein Expression, Purification, and Crystallization

Protein expression and purification of the β -glucosidase from *Chaetomium thermophilum* CT2 was carried out as previously described [22]. Briefly, the enzyme (Uniprot id A6YRT4) was produced in *Pichia pastoris* GS115 cells and purified by ion-exchange chromatography on a DEAE-Sepharose (Pharmacia, Uppsala, Sweden) to homogeneity as judged by SDS-PAGE. β -Glucosidase activity was assayed with salicin using the Miller's method and also detected in native polyacrylamide gel using 4-methylumbelliferyl- β -D-glucopyranoside [22]. The activity of the enzyme was 1.62 ± 0.20 U/mg (1 U corresponds to the release of 1 μ M of glucose per min) from three independent measurements at optimum conditions of pH and temperature.

3.2. Protein Crystallization

Prior to crystallization, the enzyme solution was concentrated to ~10 mg/mL with Amicon[®] Ultra Centrifugal Filters (10,000 MW cut-off) (Millipore, MA, USA) in 10 mM HEPES–NaOH, pH 7.0 buffer. Crystals were obtained by the hanging-drop vapor diffusion method at 16 °C using a well solution of 35–45% *v/v* MPD (Sigma-Aldrich, St. Louis, MO, USA). The drops were prepared by mixing 2 μ L of protein solution with an equal volume of well solution. The crystals grew as octahedra to a maximum size of approximately $0.06 \times 0.06 \times 0.08$ mm³ within a period of 1 month.

3.3. Data Collection and Processing

Data were collected on the X13 beamline at EMBL-Hamburg (c/o DESY) from a single crystal under cryogenic (100 K) temperature using a MARCCD detector. The presence of MPD in the crystallization was sufficient for cryoprotection, thus no additional cryoprotectant was needed. One hundred and fifty diffraction images were collected from a single crystal with a rotation range of 0.45° per image. Data processing was carried out with XDS [32]. The crystal was found to belong to the tetragonal space group $P4_12_1/P4_32_12$. Assuming one molecule in the asymmetric unit, the Matthews coefficient V_M [33] is 5.4 Å³/Da corresponding to a solvent content of ~77%.

3.4. Structure Determination and Refinement

Initial phases were obtained with molecular replacement using Phaser [34], as implemented in Phenix 1.15.2_3472 [35]. The crystal structure of *Aspergillus aculeatus* β -glucosidase in complex with castanospermine (PDB id 4iif; sequence identity 61.5%) was used as a search model after pruning side chains with Sculptor [36] based on sequence alignment considerations. Initially, the search was carried out assuming two molecules in the asymmetric unit, but no solution was produced. Based, however, on the statistics for one molecule (TFZ = 30.9), the search was limited to one molecule and a single solution was obtained in space group $P4_12_12$. Refinement was carried out using simulated annealing (1000 K) in Phenix with maximum likelihood as target function. The refinement was alternated with model visualization and rebuilding using Coot 0.8.9 [37]. Tight restraints were used to avoid overfitting of the structure to the data, owing to the low resolution. The progress of the refinement was monitored by the R_{free} with 5% of the reflections used for the calculations [38]. High-resolution structures were used to assist the rebuilding in places with poor electron density and ambiguities in atom positions. Structure validation was performed with tools implemented in Phenix and Coot. The stereochemistry and conformation of the sugars were tested with Privateer [39]. Data collection and final refinement statistics are shown in Table 1. The atomic coordinates and the structure factors have been deposited to the Protein Data Bank under the accession code 6SZ6.

3.5. Structure Analysis

PDBeFold [40] was employed to identify structural similarities between CtBGL and deposited PDB structures. Structure-based sequence alignment was performed with UCSF Chimera 1.13.1 [41]. ESPript 3 [16] was used in rendering sequence similarities and secondary structure information from aligned sequences. Salt bridges were calculated using ESBRI (<http://bioinformatica.isa.cnr.it/ESBRI/>). SAS area was calculated using PDBePISA [42]. The values for charged residues were calculated using the Expasy server (<http://web.expasy.org/protparam/>) ProtParam tool [43].

4. Conclusions

The CtBGL structure was determined at 2.99 Å resolution and refined to good stereochemical and refinement numbers, despite the limited resolution. The structure exhibited a three-domain architecture with two linker regions as other β -glucosidase structures. Variations in the length of the first linker regions and the third domain were identified. CtBGL showed the highest structural similarity with NcCel3A. The catalytic residues at the active site were identified as Asp287 (nucleophile) and Glu517 (acid/base). CtBGL showed a low number of glycans at glycosylation sites compared to other heavily glycosylated GH3 β -glucosidases. Charged residues and the glycosylation pattern are suggested as potential contributing factors for the thermostability properties of CtBGL. The analysis presented in this study could offer new ideas towards further improvements in β -glucosidases for better use in biotechnological applications.

Author Contributions: Conceptualization, D.-C.L. and A.C.P.; data analysis, I.M., N.P., and A.C.P.; writing—original draft preparation, I.M. and A.C.P.; writing—review and editing, I.M., N.P., D.-C.L., and A.C.P.

Funding: The work was supported by the Chinese National Nature Science Foundation (31571949).

Acknowledgments: The authors thank Teemu Haikarainen for his initial contribution to the project. The authors are also grateful to Biocenter Finland for infrastructure support. Access to EMBL Hamburg (c/o DESY) was provided by the European Community's Seventh Framework Programme (FP7/2007-2012) under grant agreement no. 226716 and Biostruct-X (283570).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

BGC	β -D-glucose
GH3	Glycoside hydrolase family 3
MPD	2-methyl-2,4-pentanediol
PDB	Protein Data Bank
NCS	Non-crystallographic symmetry
rmsd	Root mean square deviation
SAS	Solvent-accessible surface
$T_{1/2}$	Half-life inactivation
T_m	Melting temperature

References

- Davies, G.J.; Gloster, T.M.; Henrissat, B. Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struc. Biol.* **2005**, *15*, 637–645. [[CrossRef](#)]
- Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P.M.; Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **2013**, *6*, 41. [[CrossRef](#)] [[PubMed](#)]
- Suzuki, K.; Sumitani, J.-I.; Nam, Y.-W.; Nishimaki, T.; Tani, S.; Wakagi, T.; Kawaguchi, T.; Fushinobu, S. Crystal structures of glycoside hydrolase family 3 β -glucosidase 1 from *Aspergillus aculeatus*. *Biochem. J.* **2013**, *452*, 211–221. [[CrossRef](#)]
- Gudmundsson, M.; Hansson, H.; Karkehabadi, S.; Larsson, A.; Stals, I.; Kim, S.; Sunux, S.; Furdala, M.; Larenas, E.; Kaper, T.; et al. Structural and functional studies of the glycoside hydrolase family 3 β -glucosidase Cel3A from the moderately thermophilic fungus *Rasamsonia emersonii*. *Acta Crystallogr. Sect. D Struct. Biol.* **2016**, *72*, 860–870. [[CrossRef](#)]
- Karkehabadi, S.; Hansson, H.; Mikkelsen, N.E.; Kim, S.; Kaper, T.; Sandgren, M.; Gudmundsson, M. Structural studies of a glycoside hydrolase family 3 β -glucosidase from the model fungus *Neurospora crassa*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2018**, *74*, 787–796. [[CrossRef](#)]
- Karkehabadi, S.; Helmich, K.E.; Kaper, T.; Hansson, H.; Mikkelsen, N.-E.; Gudmundsson, M.; Piens, K.; Furdala, M.; Banerjee, G.; Scott-Craig, J.S.; et al. Biochemical Characterization and Crystal Structures of a Fungal Family 3 β -Glucosidase, Cel3A from *Hypocrea jecorina*. *J. Biol. Chem.* **2014**, *289*, 31624–31637. [[CrossRef](#)]
- Lima, M.A.; Oliveira-Neto, M.; Kadowaki, M.S.; Rosseto, F.R.; Prates, E.T.; uina, F.; Leme, A.F.; Skaf, M.S.; Polikarpov, I. *Aspergillus niger* β -glucosidase has a cellulase-like tadpole molecular shape: Insights into glycoside hydrolase family 3 (GH3) β -glucosidase structure and function. *J. Biol. Chem.* **2013**, *288*, 32991–33005. [[CrossRef](#)]
- Agirre, J.; Ariza, A.; Offen, W.A.; Turkenburg, J.P.; Roberts, S.M.; McNicholas, S.; Harris, P.V.; McBrayer, B.; Dohnalek, J.; Cowtan, K.D.; et al. Three-dimensional structures of two heavily N-glycosylated *Aspergillus* sp. family GH3 β -D-glucosidases. *Acta Crystallogr. D Struct. Biol.* **2016**, *72*, 254–265. [[CrossRef](#)]
- Varghese, J.; Hrmova, M.; Fincher, G. Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **1999**, *7*, 179–190. [[CrossRef](#)]
- Nakatani, Y.; Cutfield, S.M.; Cowieson, N.P.; Cutfield, J.F. Structure and activity of exo-1, 3/1, 4- β -glucanase from marine bacterium *Pseudoalteromonas* sp. BB1 showing a novel C-terminal domain. *FEBS J.* **2012**, *279*, 464–478. [[CrossRef](#)]
- Yoshida, E.; Hidaka, M.; Fushinobu, S.; Koyanagi, T.; Minami, H.; Tamaki, H.; Kitaoka, M.; Katayama, T.; Kumagai, H. Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β -glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **2010**, *431*, 39–49. [[CrossRef](#)] [[PubMed](#)]
- Li, D.-C.; Papageorgiou, A.C. Cellulases from Thermophilic Fungi: Recent Insights and Biotechnological Potential. In *Fungi in Extreme Environments: Ecological Role and Biotechnological Significance*; Tiquia-Arashiro, S., Grube, M., Eds.; Springer: Cham, Switzerland, 2019.
- de Pereira, C.J.; Marques, P.N.; Rodrigues, A.; de Oliveira, B.T.; Boscolo, M.; da Silva, R.; Gomes, E.; Martins, B.D. Thermophilic fungi as new sources for production of cellulases and xylanases with potential use in sugarcane bagasse saccharification. *J. Appl. Microbiol.* **2015**, *118*, 928–939. [[CrossRef](#)] [[PubMed](#)]

14. Sørensen, A.; Lübeck, M.; Lübeck, P.; Ahring, B. Fungal Beta-Glucosidases: A Bottleneck in Industrial Use of Lignocellulosic Materials. *Biomolecules* **2013**, *3*, 612–631. [[CrossRef](#)] [[PubMed](#)]
15. Karnaouri, A.; Topakas, E.; Paschos, T.; Taouki, I.; Christakopoulos, P. Cloning, expression and characterization of an ethanol tolerant GH3 β -glucosidase from *Myceliophthora thermophila*. *PeerJ* **2013**, *1*, e46. [[CrossRef](#)]
16. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324. [[CrossRef](#)]
17. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [[CrossRef](#)]
18. Pozzo, T.; Pasten, J.; Karlsson, E.; Logan, D.T. Structural and functional analyses of beta-glucosidase 3B from *Thermotoga neapolitana*: A thermostable three-domain representative of glycoside hydrolase. *J. Mol. Biol.* **2010**, *397*, 724–739. [[CrossRef](#)]
19. Payne, C.M.; Jiang, W.; Shirts, M.R.; Himmel, M.E.; Crowley, M.F.; Beckham, G.T. Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J. Am. Chem. Soc.* **2013**, *135*, 18831–18839. [[CrossRef](#)]
20. Yamaguchi, H.; Nishiyama, T.; Uchida, M. Binding Affinity of N-Glycans for Aromatic Amino Acid Residues: Implications for Novel Interactions between N-Glycans and Proteins. *J. Biochem.* **1999**, *126*, 261–265. [[CrossRef](#)]
21. Hudson, K.L.; Bartlett, G.J.; Diehl, R.C.; Agirre, J.; Gallagher, T.; Kiessling, L.L.; Woolfson, D.N. Carbohydrate–Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 15152–15160. [[CrossRef](#)]
22. Xu, R.; Teng, F.; Zhang, C.; Li, D. Cloning of a Gene Encoding β -Glucosidase from *Chaetomium thermophilum* CT2 and Its Expression in *Pichia pastoris*. *J. Mol. Microb. Biotechnol.* **2011**, *20*, 16–23. [[CrossRef](#)] [[PubMed](#)]
23. Kim, K.-H.; Brown, K.M.; Harris, P.V.; Langston, J.A.; Cherry, J.R. A Proteomics Strategy To Discover β -Glucosidases from *Aspergillus fumigatus* with Two-Dimensional Page In-Gel Activity Assay and Tandem Mass Spectrometry. *J. Proteome Res.* **2007**, *6*, 4749–4757. [[CrossRef](#)] [[PubMed](#)]
24. Sadeghi, M.; Naderi-Manesh, H.; Zarrabi, M.; Ranjbar, B. Effective factors in thermostability of thermophilic proteins. *Biophys. Chem.* **2006**, *119*, 256–270. [[CrossRef](#)] [[PubMed](#)]
25. Unsworth, L.D.; van der Oost, J.; Koutsopoulos, S. Hyperthermophilic enzymes—stability, activity and implementation strategies for high temperature applications. *FEBS J.* **2007**, *274*, 4044–4056. [[CrossRef](#)] [[PubMed](#)]
26. Lee, H.; Qi, Y.; Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* **2015**, *5*, 8926. [[CrossRef](#)] [[PubMed](#)]
27. Turner, P.; Svensson, D.; Adlercreutz, P.; Karlsson, E. A novel variant of *Thermotoga neapolitana* β -glucosidase B is an efficient catalyst for the synthesis of alkyl glucosides by transglycosylation. *J. Biotechnol.* **2007**, *130*, 67–74. [[CrossRef](#)] [[PubMed](#)]
28. Ioannou, Y.; Zeidner, K.; Grace, M.; Desnick, R. Human alpha-galactosidase A: Glycosylation site 3 is essential for enzyme solubility. *Biochem. J.* **1998**, *332*, 789–797. [[CrossRef](#)]
29. Kayser, V.; Chennamsetty, N.; Voynov, V.; Forrer, K.; Helk, B.; Trout, B.L. Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnol. J.* **2011**, *6*, 38–44. [[CrossRef](#)]
30. Meldgaard, M.; Svendsen, I. Different effects of N-glycosylation on the thermostability of highly homologous bacterial (1, 3-1, 4)-beta-glucanases secreted from yeast. *Microbiology* **1994**, *140*, 159–166. [[CrossRef](#)]
31. Baba, Y.; Sumitani, J.-I.; Tani, S.; Kawaguchi, T. Characterization of *Aspergillus aculeatus* β -glucosidase 1 accelerating cellulose hydrolysis with *Trichoderma cellulase* system. *AMB Express* **2015**, *5*, 3. [[CrossRef](#)]
32. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132. [[CrossRef](#)] [[PubMed](#)]
33. Matthews, B. Solvent content of protein crystals. *J. Mol. Biol.* **1968**, *33*, 491–497. [[CrossRef](#)]
34. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [[CrossRef](#)] [[PubMed](#)]
35. Adams, P.D.; Afonine, P.V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.-W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221. [[CrossRef](#)]
36. Bunkoczi, G.; Read, R. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 303–312. [[CrossRef](#)]

37. Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126–2132. [[CrossRef](#)]
38. Brünger, A. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472–475. [[CrossRef](#)]
39. Agirre, J. Strategies for carbohydrate model building, refinement and validation. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 171–186. [[CrossRef](#)]
40. Krissinel, E.; Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2256–2268. [[CrossRef](#)]
41. Meng, E.C.; Pettersen, E.F.; Couch, G.S.; Huang, C.C.; Ferrin, T.E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinform.* **2006**, *7*, 339.
42. Krissinel, E. Macromolecular complexes in crystals and solutions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 376–385. [[CrossRef](#)] [[PubMed](#)]
43. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).