GMR

Characterization of glucose-tolerant β-glucosidases used in biofuel production under the bioinformatics perspective: a systematic review

D.C.B. Mariano[1], C. Leite[1], L.H.S. Santos[1], L.F. Marins[2], K.S. Machado[3], A.V. Werhli[3], L.H.F. Lima[4] and R.C. de Melo-Minardi[1]

[1]Laboratório de Bioinformática e Sistemas,
Departamento de Ciência da Computação,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]Laboratório de Biologia Molecular, Instituto de Ciências Biológicas,
Universidade Federal de Rio Grande, Rio Grande, RS, Brasil
[3]Grupo de Biologia Computacional, Centro de Ciências Computacionais-C3,
Universidade Federal de Rio Grande, Rio Grande, RS, Brasil
[4]Universidade Federal de São João Del-Rei, Campus Sete Lagoas,
Sete Lagoas, MG, Brasil

Corresponding authors: D.C.B. Mariano / R.C. de Melo-Minardi
E-mail: diegomariano@ufmg.br / raquelcm@dcc.ufmg.br

Genet. Mol. Res. 16 (3): gmr16039740
Received May 26, 2017
Accepted July 28, 2017
Published August 17, 2017
DOI http://dx.doi.org/10.4238/gmr16039740

ABSTRACT. β-glucosidases are enzymes that catalyze the hydrolysis of oligosaccharides and disaccharides, such as cellobiose. These enzymes play a key role in cellulose degrading, such as alleviating product inhibition of cellulases. Consequently, they have been considered essential for the biofuel industry. However, the majority of the characterized β-glucosidases is inhibited by glucose. Hence, glucose-tolerant β-glucosidases have been targeted to improve the production of second-generation biofuels. In this

paper, we proceeded a systematic literature review (SLR), collected protein structures and constructed a database of glucose-tolerant β-glucosidases, called betagdb. SLR was performed at PubMed, ScienceDirect and Scopus Library databases and conducted according to PRISMA framework. It was conducted in five steps: i) analysis of duplications, ii) title reading, iii) abstract reading, iv) diagonal reading, and v) full-text reading. The second, third, fourth, and fifth steps were performed independently by two researchers. Besides, we performed bioinformatics analysis on the collected data, such as structural and multiple alignments to detect the most conserved residues in the catalytic pocket, and molecular docking to characterize essential residues for substrate recognizing, glucose tolerance, and the β-glucosidase activity. We selected 27 papers, 23 sequences, and 5 PDB files of glucose-tolerant β-glucosidases. We characterized 11 highly conserved residues: H121, W122, N166, E167, N297, Y299, E355, W402, E409, W410, and F418. The presence of these residues may be essential for β-glucosidases. We also discussed the importance of residues W169, C170, L174, H181, and T226. Furthermore, we proposed that the number of contacts for each residue in the catalytic pocket might be a metric that could be used to suggest mutations. We believe that the herein propositions, together with the sequence and structural data collection, might be helpful for effective engineering of β-glucosidases for biofuel production and may help to shed some light on the degradation of cellulosic biomass.

Key words: Biofuel; Glucose-tolerant β-glucosidases; Bioinformatics; Systematic literature review

INTRODUCTION

β-glucosidase (E.C. 3.2.1.21) is a class of heterologous enzymes that hydrolyze glycosidic bonds of disaccharides, oligosaccharides, alkyl- and aryl-β-glycosides (Cairns and Esen, 2010). They have been found in metagenomes (Uchiyama et al., 2015) and several organisms, such as animals (Uchima et al., 2011), fungi (Saha and Bothast, 1996), plants (Pentzold et al., 2014), and bacteria (Crespim et al., 2016). In animals, these enzymes help in the metabolism of glycolipids and digestive functions (Cairns and Esen, 2010). In plants, they play several roles, such as defense, the release of flavor compounds, cell wall catabolism, and lignification (Pentzold et al., 2014). In bacteria and fungi, they are essential components of cellulose hydrolysis (Ramani et al., 2015a). β-glucosidases have applications in several areas of biotechnological industries, such as aroma improvement of juices and wine (Swangkeaw et al., 2010), hydrolysis of soybean isoflavone glycosides (Singhania et al., 2013), toxicity reduction of animal feed (Cota et al., 2015), and cellulose degradation for biomass conversion in biofuel production (Singhania et al., 2013).

Based on substrate specificity, these enzymes were classified into three main groups: i) aryl-glucosidases, ii) cellobiases, and iii) broad-specificity β-glucosidases (Singhania et al.,

2013; Ramani et al., 2015a; Yang et al., 2015a). However, the limitations of this classification required a new classification method based on sequence similarities (Henrissat, 1991). Since then, β-glucosidases have been classified into glycoside hydrolase (GH) families 1, 3, 5, 9, and 30 (Cairns and Esen, 2010). However, the majority of the β-glucosidases belongs to families 1 and 3 (Singhania et al., 2013; Crespim et al., 2016). β-glucosidases from the GH3 family present an aspartate as a catalytic nucleophile and a glutamate as a catalytic acid/base. On the other hand, enzymes of the GH1 family belong to the clan GH-A, that group proteins with $(\alpha/\beta)_8$ TIM barrel fold and conserved catalytic amino acids on β-strands 4 and 7 of the barrel (Cairns and Esen, 2010; Jabbour et al., 2012; Crespim et al., 2016). GH1 β-glucosidases hydrolyze through the mechanism of retention of the anomeric carbon, using two glutamates as a catalytic nucleophile and catalytic acid/base. GH1 β-glucosidases have attracted attention due to the high resistance to product inhibition, which has many applications for cellulose degradation (Yang et al., 2015b).

Cellulose is the major source of biomass on the Earth, accounting for around 40 to 50% of the plant biomass weight (Uchima et al., 2012; Uchiyama et al., 2015). Cellulose is constituted of glucose monomers connected by β-1,4 glucosidic bonds (Ramani et al., 2015b). The glucose obtained by cellulose degradation may be fermented to produce bioethanol, a promising green alternative and renewable source for the production of fuels (Teugjas and Väljamäe, 2013). β-glucosidases act synergistically with endoglucanases (E.C. 3.2.1.4) and exoglucanases or cellobiohydrolases (E.C. 3.2.1.91) to compose the enzymatic system for cellulose bioconversion (Kumar et al., 2008). While endoglucanases act in the cellulose chain producing oligosaccharides of variable length, exoglucanases act in the oligosaccharides producing mainly cellobiose. β-glucosidases cleave the link β-1,4 glucosidic bonds with the help of a water molecule, producing two glucose molecules (Béguin and Aubert, 1994). It is well established that β-glucosidases have a pivotal role in this enzymatic system removing cellobiose, which is a strong inhibitor of both endoglucanases and exoglucanases (Murphy et al., 2013; Zhao et al., 2013; Chamoli et al., 2016). However, most of the β-glucosidases reported are inhibited by the increase of glucose concentration (Teugjas and Väljamäe, 2013; Yang et al., 2015b). Hence, there is the growing interest in searching for thermostable and glucose-tolerant β-glucosidases. Their production may help to shed some light on the degradation of cellulosic biomass and may improve the saccharification process for biofuel production (Pei et al., 2012).

In this paper, we conducted a systematic literature review (SLR) to collect, analyze, and summarize the state-of-the-art research about glucose-tolerant β-glucosidases. We aimed to identify research trends about β-glucosidase enzymes used to improve the biofuel production. Although the mechanisms and molecular basis for glucose tolerance have not been completely enlightened, several studies have presented insights about the role of glucose tolerance, as well as structural aspects of interaction with the substrates and products, and also some discoveries on more efficient β-glucosidases in biomass degradation (Singhania et al., 2013). Also, we collected protein structures and constructed a glucose-tolerant β-glucosidase database. The construction of glucose-tolerant β-glucosidase structure database may be useful for bioinformatics analysis, such as the characterization of residues responsible for glucose tolerance in β-glucosidases. We also performed molecular docking and characterized important amino acids near to the cellobiose-binding region for the occurrence of hydrolysis. The provided information may be useful for engineering efficient enzymes for second-generation biofuel production.

MATERIAL AND METHODS

The SLR protocol was created based on the guide for performing SLR in bioinformatics (Mariano et al., 2017), the guideline for systematic reviews of Kitchenham (2004), and the PRISMA statement (Liberati et al., 2009).

Search strategy

Data collection for the SLR was performed from November 2015 until February 2016. Search terms were defined based on interviews with researchers in the area, the study of Pei et al. (2012), and were iteratively improved based on the results obtained from the first queries. PubMed (http://www.ncbi.nlm.nih.gov/pubmed), ScienceDirect (http://www.sciencedirect.com/), and Scopus (http://www.scopus.com/) databases were used to collect the studies. The query used was ("beta-glucosidases" or "beta-glucosidase") and ("glucose-tolerant" or "glucose tolerance" or "glucose insensitive" or "product tolerant" or "product insensitive"). The period of publication for the three databases was "all the time". For ScienceDirect, we applied a filter to return only content type declared as "journal". For Scopus, we applied filters to return manuscripts in English, to search in any part of the document, and with document type declared as "article". No filter was applied for PubMed. The filters were applied to remove undesirable formats, such as book chapters or articles in idioms different from English.

Eligibility criteria and selection

The SLR was performed in five steps: i) exclusion of duplications, ii) analysis of titles, iii) analysis of abstracts, iv) diagonal reading (when only the introduction, titles of the figures and conclusion are read), and v) full-text reading. Second, third, fourth, and fifth steps were performed independently by two reviewers to minimize the risk of bias or mistakes. In steps two, three and four, a paper was kept for the next step if at least one of the evaluators had approved. In these steps, the reviewers evaluate if the paper presented a β-glucosidase applied for biofuel production and if glucose tolerance was cited as crucial for improving glucose production. We prioritized articles containing publicly available protein structural and kinetic data, such as inhibition constant for glucose and affinity for cellobiose. Other criteria, such as culture mediums and reagents used were not evaluated. In the last step, we performed a full-text reading. For each paper, we answered five questions: i) does the study present a β-glucosidase with the enzyme kinetic patterns, such as inhibition constant (K_i) for glucose, Michaelis constant (K_m) for cellobiose, catalytic constant (K_{cat}), characterized? ii) Is the β-glucosidase presented as a possible enzyme for biofuel production and it hydrolyzes cellobiose as a substrate? iii) Does the study show a genetic engineering strategy, such as mutagenesis or fusion, to solve the problem of glucose inhibition? iv) Does the paper seek to explain the mechanisms of inhibition by glucose in β-glucosidases? v) Does the paper present β-glucosidase sequence or three-dimensional structure data publicly available for bioinformatics analysis?

For each question, both evaluators, in consensus, gave scores: (0) if it does not attend requirements of the question; (1) if it attends the requirements of the question partially; and (2) if it attends requirements of the question. A paper was included in the review if it has a score equal or higher than six (60%). We also collected the following information from the papers: i) author names, ii) title, iii) sequence IDs, iv) PDB IDs, and v) enzyme kinetic information (if available).

Biological data collection

The sequences were collected in the databases UniProt (http://www.uniprot.org/) and GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Structural files were obtained in the Protein Data Bank (PDB; Berman et al., 2000). Protein three-dimensional structures are crucial to infer about the glucose tolerance mechanism in β-glucosidases. However, few experimental structures of glucose-tolerant were found.

Bioinformatics analysis

Homology modeling has been used to obtain three-dimensional structures for sequences of β-glucosidases (Yang et al., 2015b; Chamoli et al., 2016; Crespim et al., 2016). To make available as much information as possible to enable studies of glucose-tolerant β-glucosidases, we constructed 18 three-dimensional models to represent the sequences without a structure available. The sequences were modeled by homology using a consolidated pipeline of the literature (Bitar and Franco, 2014). For each protein, 100 models were built using MODELLER (Webb and Sali, 2014). NCBI BLAST web interface (Johnson et al., 2008) was used for template selection. The best model was selected based on the lowest value of MODELLER objective function and by the highest number of residues in favored region calculated by the RAMPAGE software (Lovell et al., 2003). Sequences were compared with a global alignment using the ggsearch36 software (Pearson, 2016) that implements the Needleman-Wunsch algorithm. Results with e-value lower than 0.001 were eliminated. An identity matrix was constructed using in-house scripts. We also built a database and website to make available visualizations of the three-dimensional structures and characteristics of the PDB files collected and modeled using 3Dmol (Rego and Koes, 2015).

We used the β-glucosidase from the termite *Neotermes koshunensis* in complex with cellobiose (Jeng et al., 2012) to characterized the catalytic pocket of GH1 β-glucosidases. We considered the catalytic pocket as the residues present in the channel that guides the substrate to the two catalytic amino acids. We selected 24 residues whose distances from ligand were less than or equal to 6.5 Å from the ligand. The distance was chosen based on previous metrics to determine residues of pockets (Pires et al., 2013). Then, we performed structural alignments among the 21 collected glucose-tolerant GH1 β-glucosidases collected and the *N. koshunensis* β-glucosidase to obtain the corresponding residues of the catalytic pocket using MultiProt (Shatsky et al., 2004). We also performed sequence alignments using Clustal Omega (Sievers et al., 2011) to detect a consensus subsequence and constructed a residue conservation visualization using D3 JavaScript library (https://d3js.org).

Molecular docking has been used to understand cellobiose binding in β-glucosidases (Khairudin and Mazlan, 2013). We performed molecular docking to verify which residues of the 21 glucose-tolerant GH1 β-glucosidases were directly related to the recognizing of the substrate and to determine the number of contacts carried out for each residue with the cellobiose. DOCK 6 (version 6.7) was used to perform molecular docking (Allen et al., 2015). DOCK 6 uses an incremental construction method as a sampling algorithm for flexible ligand docking, the so-called anchor-and-grow algorithm (Ewing and Kuntz, 1997). In this approach, the ligand is separated in layers, and the major rigid substructure of the ligand is primarily recognized as the anchor, the anchor is then rigidly oriented in the binding site. Subsequently, each layer of flexible bonds is grown from each cluster, minimized, ranked, and clustered

again. The latter procedure is repeated until the molecule is fully constructed.

Before molecular docking with DOCK 6, all systems (receptor and ligand molecules) were evenly prepared. The crystal structure, PDB code: 3VIK (Jeng et al., 2012), bound to the cellobiose (CBI) ligand, was chosen as a reference, and all structures were superimposed on it, using the matchmaker tool in Chimera program (Pettersen et al., 2004). The superimposing of structures before molecular docking facilitates comparison between the crystallographic CBI and pose outcomes (**Figure S1**). For the structures extracted from the PDB database, ions, and other co-factors were removed. Hydrogens were added at physiological pH, and standard receptor residues were assigned AMBER ff14sb atomic partial charges (Maier et al., 2015), while CBI was assigned AM1-BCC charges (Jakalian et al., 2002). All receptor-CBI systems were subjected to energy minimization, 500 steepest descent minimization steps, where all residues were unrestrained using Chimera's Minimize Structure module (Pettersen et al., 2004). This minimization step allows the newly added hydrogen atoms of receptor and ligand to adjust in physically moderate positions, while also transporting the protein to a potential energy point inferior to the one from before.

Receptor and CBI were separated so that a specific DOCK preparation could be performed. DOCK's sphgen program (DesJarlais et al., 1988) was used to generate spheres within the binding site. The spheres that were within 8.0 Å of the crystallographic CBI position (PDB code: 3VIK) were kept for docking. A box around the spheres plus a 5.0 Å margin in all directions was used to restrict the receptor space for energy grid calculations. Lastly, energy interactions between a dummy probe atom and all receptor atoms on a 0.3 Å resolution grid within the box were calculated with DOCK's GRID program (Meng et al., 1992). In the calculated grid, van der Waals interactions were modeled through the Lennard-Jones potential with 6-12 attractive and repulsive exponents, respectively, whereas, a distance-dependent dielectric coefficient was used to represent the electrostatic interactions. The grid files are essential for rapid score evaluation using DOCK's native energy-based score GridScore. CBI was treated as flexible in all docking outcomes, according to the "standard flexible docking (FLX) protocol" described in Allen et al. (2015). A maximum of 300 ligand conformations was retained and clustered (RMSD cut-off of 2.0Å) for all receptor-CBI systems to compose our docking analysis. We used all the structures to calculate the number of contacts between residues and substrate. A contact is defined as all kinds of direct interactions: polar, nonpolar, favorable, and unfavorable (including clashes).

RESULTS

The results of SLR were described according to PRISMA workflow and checklist (Figure 1 and **Table S1**). We collected 665 papers from three databases: PubMed, ScienceDirect, and Scopus. We performed five steps to evaluate the quality of the publications according to predefined objectives of the present SLR. After the analysis of two evaluators, 27 papers were included in the SLR (**Table S2**). We grouped the 27 papers into three categories: 1) papers that report a new glucose-tolerant β-glucosidases with proper application for biofuel production and with sequences or three-dimensional structures publicly available; 2) papers that report an engineering genetic technique being used to improve the β-glucosidase activity; 3) papers that report a comprehensive explanation about glucose tolerance mechanism and try to determine the related structural aspects (**Table S2**). We also collected five protein structural files, 23 β-glucosidase sequences, and their kinetic parameters (**Table S3**). Besides, 18 three-

dimensional models of the sequences without a PDB file available in the PDB database were constructed by homology (**Figure S2**). We decided to perform the next steps only with the 21 GH1 β-glucosidases because glucose-tolerant β-glucosidases (from now called GTBGL) have been described only in the GH1 family.

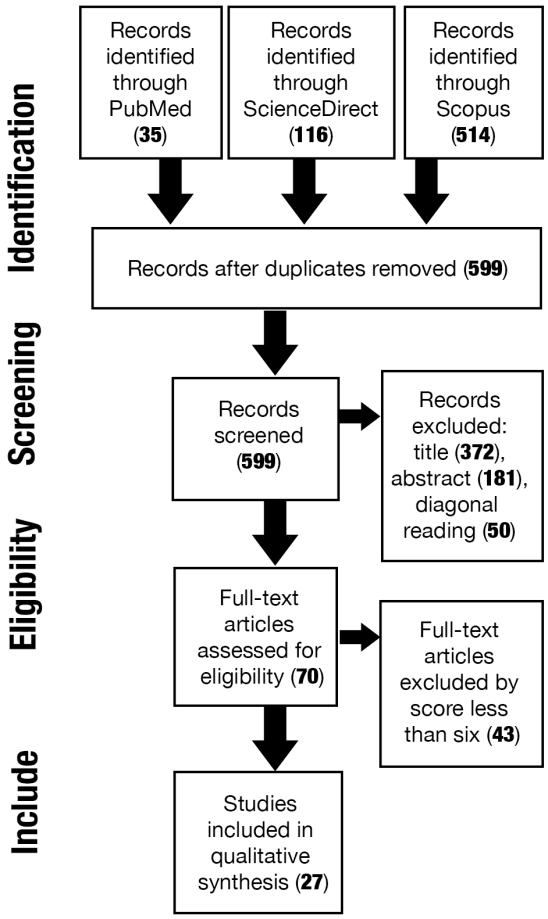

Figure 1. PRISMA workflow describing the number of papers collected during the systematic literature review based on the PRISMA statement.

Conserved residues in the catalytic pocket of GTBGL

Based on the information collected in the SLR, we hypothesized that the key amino acids responsible for regulating the glucose tolerance process should be in the channel that guides to the active site. We called this region as the catalytic pocket (Figure 2A-C). To verify this, we performed alignments among the residues in the pocket and detected a GTBGL consensus subsequence composed of 22 most conserved amino acids near to the likely region of cellobiose binding: "HWNEWCLHNLTANYYTNEWEWF" (Figure 2D). These amino acids corresponded in the *Thermoanaerobacter brockii* β-glucosidase to the residues: H121,

W122, N166, E167, W169, C170, L174, H181, N224, L225, T226, A244, N297, Y298, Y299, T300, S302, E355, W402, E409, W410, and F418 (Figure 2). The catalytic pocket of the *T. brockii* β-glucosidase presented the highest quantity of similarities with the GTBGL consensus subsequence. For this reason, from now on it was used as a reference (the corresponding residues for all GH1 glucose-tolerant β-glucosidases are available in **Table S4**).

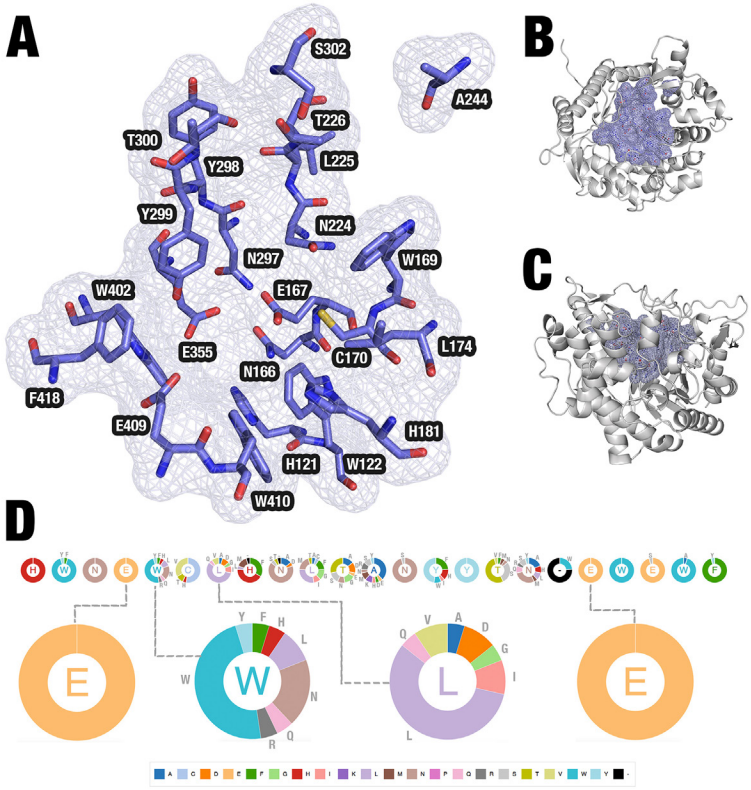

Figure 2. Conserved residues in the catalytic pocket of glucose-tolerant GH1 β-glucosidases. **A.** Residues to 6.5 ångström of the ligand in the GH1 β-glucosidase from *Thermoanaerobacter brockii*. **B.** Enzyme top view. **C.** Enzyme side view. **D.** Conserved residues in the 21 GH1 glucose-tolerant β-glucosidases. The conservation of the amino acid is shown in the pie chart. We highlighted the two catalytic residues, E167 and E355, and the residues W169 and L173 described in the literature as essential for glucose tolerance. The amino acids "HWNEWCLHNLTANYYTNEWEWF" correspond in the β-glucosidase from *T. brockii* to the residues: H121, W122, N166, E167, W169, C170, L174, H181, N224, L225, T226, A244, N297, Y298, Y299, T300, S302 (correspond to an asparagine), E355, W402, E409, W410, and F418. Images generated by PyMOL (http://pymol.org) and D3.js (http://d3js.org).

Six residues were conserved in all glucose-tolerant β-glucosidases: H121, N166, E167, Y299, E355, and W402 (Table 1). E167 and E355 were acid/base catalytic and nucleophile catalytic residues, respectively. The other residues were located near to both catalytic residues. The conserved residues were essential to recognize the substrate. The residues W122, N297, E409, W410, and F418 were conserved in more than 90% of the sequences. Also, W169,

C170, L174, N224, A244, Y299, and T300 were conserved in the majority of the sequences. The low conservation of amino acids L225, T226, and S302 (an asparagine in the consensus subsequence) suggested minor importance to these residues.

Table 1. Residue conservation and contacts with cellobiose.

#	Residue (*T. brockii*)	Residue conservation		Contacts with cellobiose	
		>90% conserved	100% conserved	Corresponding residue makes contact in all GTBGL	No contacts when substituted (reference: residue of *T. brockii*)
1	H121	x	x		
2	W122	x		x	
3	N166	x	x	x	
4	E167	x	x	x	
5	W169				x
6	C170			x	
7	L174			x	
8	H181				x
9	N224				x
10	L225				
11	T226				x
12	A244				
13	N297	x			
14	Y298				
15	Y299	x	x	x	
16	T300				
17	S302				x
18	E355	x	x	x	
19	W402	x	x	x	
20	E409	x		x	
21	W410	x		x	
22	F418	x			

The table corresponds to the 22 corresponding residues in the catalytic pocket of the 21 glucose-tolerant β-glucosidases (GTBGL).

Cellobiose docking

From the retained and clustered docking conformations, we calculated the number of contacts between all receptor residues and cellobiose (**Figure S3**). The corresponding residues to W122, N166, E167, C170, L174, Y299, E355, W402, E409, and W410 formed contacts with the cellobiose in every GTBGL (Table 1 and Figure 3). Moreover, the corresponding residues to H181 and N224 established contacts in almost every GTBGL. However, they are not present in the *Bacillus subtilis* β-glucosidase. Also, some residues established several contacts, except when they were substituted by other amino acids, such as W169 when substituted by a glutamine or T226 when substituted by a glycine. Furthermore, the residue found in the 17th position of the catalytic pocket consensus subsequence, which in the *T. brockii* β-glucosidase was S302, presented many contacts when it was a serine but mostly none when the residue was an alanine. When this residue was an asparagine, the most conserved amino acid in this position, in some β-glucosidases, it performed several contacts while in another no contact at all, such as in *Neurospora crassa* and *Trichoderma reesei* β-glucosidases, respectively. The corresponding residues to L225, A244, N297, Y298, T300, and F418 established few or no contacts (Table 1 and Figure 3). However, the residues in the corresponding position of N297 and F418 established several contacts when they were substituted by serine and tyrosine, respectively.

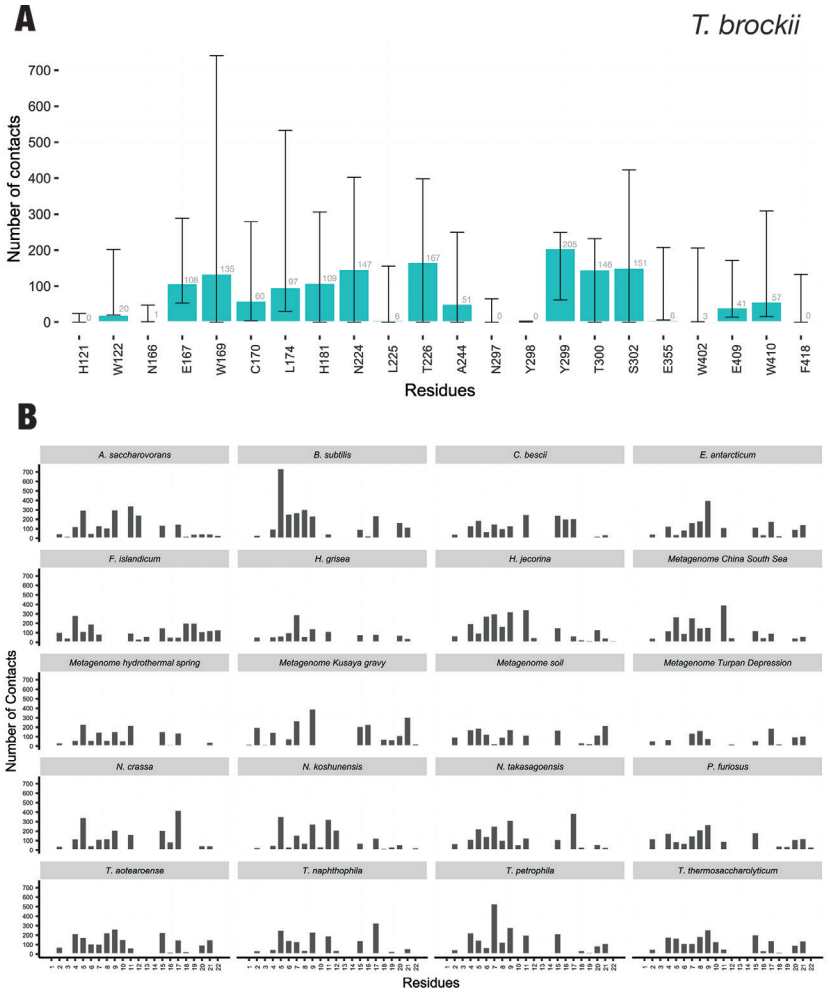

Figure 3. Contacts between residues and the cellobiose docked in glucose-tolerant GH1 β-glucosidases. **A.** Contacts between residues and the cellobiose docked in the GH1 β-glucosidase from *T. brockii*. A black line indicates the minor and maximum value detected in the corresponding position of other glucose-tolerant β-glucosidases. **B.** Contacts with the ligand that occurs in the other glucose-tolerant β-glucosidases. A complete table of the corresponding values can be obtained in **Table S4**.

DISCUSSION

Glucose inhibits the β-glucosidase activity competing with the substrate (cellobiose) to bind in the active site. Hence, glucose-tolerant β-glucosidases may increase the biofuel production capacity and reduce costs (Yang et al., 2015b). An efficient hydrolysis of biomass requires active β-glucosidases at higher glucose concentration (Singhania et al., 2013). For this reason, glucose-tolerant β-glucosidases are important. However, it has been hard to identify a β-glucosidase completely tolerant to product inhibition, due to the similarities between product and substrate. It is observed that when the affinity to the product is reduced,

the substrate affinity is reduced as well. However, β-glucosidases with low sensitivity to inhibition have been found. Moreover, the enzymatic hydrolysis of cellulose that produces glucose concentrations may reach 650-1000 mM (Meleiro et al., 2015). Hence, an ideal β-glucosidase for biofuel production should have elevated levels of glucose tolerance and also a high catalytic efficiency.

In addition to glucose tolerance, a stimulatory effect of glucose for β-glucosidases was reported in several studies (Pei et al., 2012; Yang et al., 2015b; Crespim et al., 2016; Guo et al., 2016). This effect has been reported exclusively in some GH1 β-glucosidases (de Giuseppe et al., 2014; Yang et al., 2015b). However, not every glucose-tolerant β-glucosidases present this stimulatory effect (Meleiro et al., 2015). Glucose stimulation consists in an improvement in the β-glucosidase activity in a particular range of glucose concentrations. For instance, it has been reported that in the presence of 50 mM glucose the activity of a β-glucosidase from *Humicola insolens* RP86 was stimulated about 1.8-fold (Souza et al., 2014). Glucose stimulation occurs due to an allosteric effect by glucose binding in a secondary site or by transglycosylation (Cao et al., 2015). It has been suggested that when glucose concentration increases during the saccharification process, substrate inhibition is gradually prevented (Guo et al., 2016), supporting the idea that glucose stimulation occurs not due to the presence of certain glucose quantity, but due to the reduction of substrate concentration in the reaction environment. Besides glucose inhibition, some β-glucosidases are inhibited by cellobiose.

Conserved patterns in the catalytic pocket

β-glucosidase structures collected during the SLR were used to characterize conserved residues present in the catalytic pocket. Residue conservation is a primary indication of conserved patterns that could be used to propose mutations for non-tolerant β-glucosidases. The residues H121, N166, E167, Y299, E355, W402, W122, N297, E409, W410, and F418 are conserved in more than 90% of the sequences (Table 1). Hence, any mutation of these residues could not be indicated for engineering more efficient β-glucosidases once conservation is an indicative of the residue importance. On the other hand, some residues appear in the majority of the sequences. However, they are not highly conserved, such as W169, C170, L174, N224, A244, Y299, and T300. Other residues are poorly conserved, such as L225, T226, and S302, what can indicate that these positions have little importance.

Also, the docking results (Figure 3) showed that some residues are important for the substrate recognizing. The corresponding residues to W122, N166, E167, C170, L174, Y299, W402, E409, and W410 performed several contacts with the ligand. The residue E167 is one of the catalytic glutamates and performs several contacts with the ligand. Moreover, E355, the other catalytic glutamate, produced few contacts with the ligand. The H121 is another residue that is highly conserved but performs few contacts with the ligand. The H121 probably acts together with E167 and E355 in the catalytic activity. It has been highlighted that H121 is somehow involved in substrate binding or transition state stabilization (Barrett et al., 1995; Sanz-Aparicio et al., 1998).

We also detected a possible co-occurrence of a cysteine (C170) when appearing both a tryptophan (W169) and leucine (L174). Glucose-tolerant GH1 β-glucosidases present a deep and narrow channel that limits the glucose access to the active site. It has been suggested that the channel shape guides the substrate to the active site and it is responsible for reducing the glucose access and, consequently, glucose tolerance in GH1 β-glucosidases (de Giuseppe et

al., 2014). Equally, some amino acids located near to the active site were reported in some studies as essential for the release of glucose (Figure 4; de Giuseppe et al., 2014; Guo et al., 2016). In the β-glucosidase from *Humicola grisea* var. *thermoidea* (PDB: 4MDO), the amino acids W168 and L173 (W169 and L174 of *T. brockii*) were reported as contributors for relieving enzyme inhibition. They apply constraints at the +2 subsite, and consequently, cause to glucose access limitation at the -1 subsite (Figure 4A). In another study, the site-directed mutagenesis L167W and P172L (corresponding residues of W169 and L174) were used to improve the glucose tolerance, pH and thermostability of β-glucosidase from *T. reesei* (Figure 4B; Guo et al., 2016). The W169 and L174 were conserved in the majority of glucose-tolerant β-glucosidases. This conservation highlights the importance of these amino acids for the regulation of entrance and exit of products and substrate, already shown in other articles (de Giuseppe et al., 2014; Guo et al., 2016). Likewise, the residue C170 appears in the majority of the catalytic pockets, which indicates that mutating an amino acid in this corresponding position for a cysteine may be beneficial, as reported in the literature (Figure 4C; Cao et al., 2015). In a recent study, the authors used random mutagenesis and detected three beneficial mutations: V174C, A404V, and L441F (where V174 corresponds to C170 in the *T. brockii* β-glucosidase). These mutations allowed the construction of a new glucose-tolerant and thermostable mutant, which enhanced sugarcane bagasse conversion by 14-35%.

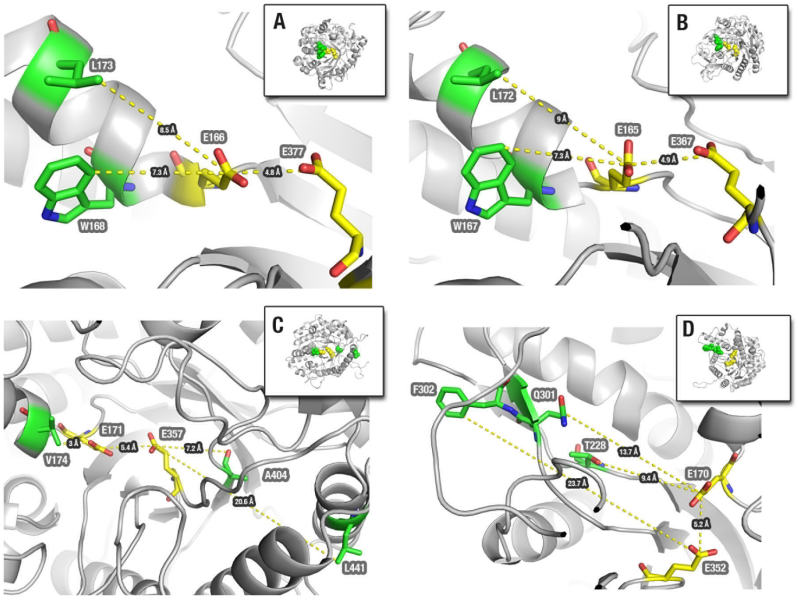

Figure 4. Structural data of β-glucosidases, amino acids related to glucose tolerance, and the active site. **A.** *Humicola grisea* β-glucosidase presents a narrow and deep channel that guide to the active site. The acid/base catalytic and nucleophile amino acids, E166 and E377, are ~5 Å away. The residues W168 and L173 are responsible for contributing to relieving enzyme inhibition. **B.** *Trichoderma reesei* β-glucosidase. The active site residues are E165 and E367. Mutations L167W and P172L improved the activity. **C.** β-glucosidase from metagenome Turpan Depression. The active site residues are E357 and E171. Mutations V174C, A404V, and L441F improved the catalytic activity, although glucose-tolerance reduced of 3.5 M to 3 M. **D.** β-glucosidase from metagenome China South Sea. The active site residues are E170 and E352. Mutations H228T, N301Q, and V302F were performed in the entrance and middle of the channel that guides to the active site; Images generated by PyMOL (http://pymol.org).

Our results suggest that the appearing of a cysteine in this position seems to be correlational with the appearing of tryptophan and leucine in a nearby region, such as W169 and L174. Furthermore, according to the molecular docking results, when the three residues tryptophan, cysteine, and leucine appear together, they perform fewer contacts than when other residues substitute them; this suggests that co-occurrence may be related to the glucose tolerance. For instance, the residues tryptophan, cysteine, and leucine of the *Exiguobacterium antarcticum* β-glucosidase performed 347, 51, and 117 contacts, respectively (**Table S5**). In the same positions, the *Nasutitermes takasagoensis* β-glucosidase presents an arginine, an aspartate, and a threonine, that performed 740, 261, and 275 contacts, respectively (**Table S5**). The *E. antarcticum* β-glucosidase showed IC50 for glucose of 1000 mM (Crespim et al., 2016). On the other hand, the *N. takasagoensis* β-glucosidase showed a glucose tolerance of 600 mM (Uchima et al., 2012). These values are not full evidence that these amino acids are related to a higher glucose tolerance. However, other glucose-tolerant β-glucosidase, such as the extracted of metagenome from China South Sea (Yang et al., 2015b), Turpan Depression (Cao et al., 2015), and Kusaya gravy (Uchiyama et al., 2015), present fewer contacts in these positions and IC_{50} of 1000, 3500, and 750 mM, respectively (**Table S5**). This information might suggest residues that perform fewer contacts with the cellobiose in the middle of the catalytic pocket channel and may be beneficial for glucose tolerance.

Furthermore, mutagenesis of residues at the entrance and in the middle of the same channel of a GH1 β-glucosidases isolated from a metagenome of the China South Sea has been used to characterize other sites that glucose has more preference than the active site (Figure 4D; Yang et al., 2015b). The glucose tolerance on a non-tolerant β-glucosidase (bgl1B) was increased through the mutations H228T and N301Q/V302F (Figure 4D; Yang et al., 2015b). This outcome may suggest that secondary sites, which cellobiose binds the path of the active site, are of great importance in the release of glucose process and the binding inhibition. The corresponding residue to H228 in the *T. brockii* β-glucosidase is the T226. Hence, the mutation H228T appears to be beneficial. Although T226 was not highly conserved (Figure 2), its substitution for another amino acid could change the number of contacts (Table 1). Indeed, a histidine does not appear in this position in any glucose-tolerant β-glucosidase.

Moreover, the importance of the amino acid 184 of a β-glucosidase obtained from a marine microbial metagenomic library had been revealed earlier through the H184F mutation (Liu et al., 2011). The residue H184 (corresponding to H181 in *T. brockii* β-glucosidase) appears in the majority of the tolerant β-glucosidases reported. The mutation H184F showed an increase in the glucose tolerance when the substrate was pNPG (4-nitrophenyl-β-D-glucopyranoside; Liu et al., 2011). Indeed, phenylalanine is the second more common residue in this position based on multiple alignments (**Figure S4**). We detected that a histidine performs more contacts with the cellobiose than a phenylalanine in the same position. However, there is no report of the impact of this mutation on the glucose tolerance when the substrate was cellobiose. Hence, more experiments are necessary to make inferences.

SLR evaluation

Although SLR was used originally to medical research, they have been used to perform unbiased reviews in various areas (Kitchenham, 2004). Recently, several reviews have covered: i) general aspects of β-glucosidases (Cairns and Esen, 2010); ii) efficiency of β-glucosidases from fungus (Tiwari et al., 2013); iii) the role of β-glucosidases in the

hydrolysis of cellulose (Singhania et al., 2013); iv) reduction of product inhibition during enzymatic lignocellulose hydrolysis (Andrić et al., 2010); and v) advances in enzymes for lignocellulosic biomass conversion (Saha and Bothast, 1997). However, to our knowledge, this is the first report of an SLR that analyzes the role of glucose tolerance in β-glucosidases for biofuel production focusing on bioinformatics analysis.

Most of the glucose-tolerant β-glucosidases reported in this SLR belong to the GH1 family (21 of 23). Indeed, β-glucosidases from GH1 family have been reported to be 10- to 1000-fold more glucose tolerant than the ones from the GH3 family (de Giuseppe et al., 2014). The potential for industrial use of β-glucosidase enzymes belonging to the GH1 family has been highlighted due to its broad substrate specificity and weak inhibition by glucose (Cota et al., 2015). Although we have detected two glucose-tolerant GH3 β-glucosidases that showed activity in glucose concentrations of 140 mM (Huang et al., 2014) and 400 mM (Ramani et al., 2015b), this SLR confirms that GH1 β-glucosidases are more promising for the second-generation biofuel production. However, more works with GH3 β-glucosidases in the future may bring new conclusions.

In the SLR, we detected that glucose inhibition constant (K_i) was the main parameter used to measure product inhibition. For example, the β-glucosidase from *Bacillus subtilis* showed the highest glucose K_i value (1.9 M) among the sequences collected (**Table S2**). When the K_i for glucose was not available, the papers reported another tolerance parameter, such as IC_{50}. For instance, the *Mucor circinelloides* β-glucosidase retained 84% activity at glucose concentrations up to 140 mM (Huang et al., 2014), a high value for a β-glucosidase from GH3 family. It has been suggested that an ideal β-glucosidase could be obtained by improving the tolerance of GH3 or improving the specificity constant (k_{cat} / K_m) of the GH1 (Cao et al., 2015). Moreover, it has been reported two β-glucosidases from the GH1 and GH3 families present in the same operon of *Thermoanaerobacter brockii* (Breves et al., 1997). Hence, the diversity of β-glucosidases from different families may be beneficial for the organism to perform the biomass degradation.

Bias risk

The different methods used to measure the glucose inhibition may result in a risk of bias for the SLR. The inhibition constant (K_i) is an effective metric to measure the inhibition. However, it is harder to be determined experimentally. For this reason, the majority of the works has preferred to use IC50 to determine the glucose inhibition in β-glucosidases. In this SLR, we noticed that the IC50 had been determined by different methods, which can be a problem for comparisons among the results. Comparisons among IC50 results would depend on enzymatic assays, different from K_i. Also, IC50 may vary at different enzyme concentrations, even if the parameters considered to measure the enzyme activity were the same. In the various studies accessed in this SLR, enzymatic concentrations used in IC50 assays were not the same and were only slightly related.

Another risk of bias concerns substrates used in glucose inhibition tests. Although several works have presented β-glucosidases with resistance at high glucose concentrations and affinity tests with cellobiose, most studies have used pNPG (4-nitrophenyl-β-D-glucopyranoside) as a substrate for glucose tolerance tests. The pNPG substrate concentration can be measured without interference from glucose added, which is more convenient for bench experiments. However, for biofuel production, the β-glucosidase natural substrate is

cellobiose (Bohlin et al., 2010). The inhibition constant for glucose and substrate specificity constant for cellobiose are the most important metrics to select β-glucosidases (Teugjas and Väljamäe, 2013).

CONCLUSIONS

In summary, we performed a broad search in the literature guide by the restrictions of a systematic literature review. We selected 27 papers that report the state-of-art of glucose-tolerant β-glucosidases used in the production of second-generation biofuel and collected structural data. We constructed a database with five three-dimensional structure files, 23 glucose-tolerant β-glucosidases sequences, and their kinetic information, and also modeled by homology 18 glucose-tolerant β-glucosidase sequences. We identified the most conserved residues in the catalytic pocket and the residues that perform more and fewer contacts with the cellobiose. This information may be substantial for understanding the β-glucosidases mechanisms, identifying sites for mutations and engineering novel and more efficient β-glucosidases for biofuel production. Based on the SLR results, we conclude that site-directed mutagenesis seems to be a great strategy to produce more efficient β-glucosidases for biofuel production. For this reason, it is important to determine the residues related to the glucose tolerance. We detected that the residues H121, N166, E167, Y299, E355, W402, W122, N297, E409, W410, and F418 are highly conserved, and for this reason, they probably are critical for substrate recognizing. We also detected an apparent co-occurrence of the residue C170 when both residues W169 and L174 appear, which may be related to the glucose tolerance. We also verify that the residue T226 may be essential for the glucose tolerance. In addition, we detected that the residue H181, which has been described as important for glucose tolerance, perform more contacts with the cellobiose than when is substituted by phenylalanine. However, the use of experimental methods is necessary to infer about the importance of this residue for GTBGL. We also proposed that the number of contacts of a residue in a determined position could be used as a metric to suggest mutation in non-tolerant β-glucosidases.

To the best of our knowledge, this is the first report of an SLR about glucose tolerance in β-glucosidases and a database of glucose-tolerant β-glucosidases. The results of this work may be useful for *in silico*, *in vitro* and *in vivo* experiments and may help shed light on the production of second-generation biofuel. The sequences and structural data were organized in a database, called betagdb, which is available at: <http://bioinfo.dcc.ufmg.br/betagdb>.

ACKNOWLEDGMENTS

The authors thank the funding agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors thank Larissa Leijôto for the help in the conception of Figure 4.

REFERENCES

Akram F, Haq I ul, Khan MA, Hussain Z, Mukhtar H, et al. (2016). Cloning with kinetic and thermodynamic insight of a novel hyperthermostable β-glucosidase from *Thermotoga naphthophila* RKU-10T with excellent glucose tolerance. *J. Mol. Catal. B Enzym.* 124: 92-104.

Allen WJ, Balius TE, Mukherjee S, Brozell SR, et al. (2015). DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* 36: 1132-1156 https://doi.org/10.1002/jcc.23905.

Andrić P, Meyer AS, Jensen PA and Dam-Johansen K (2010). Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis: I. Significance and mechanism of cellobiose and glucose inhibition on cellulolytic enzymes. *Biotechnol. Adv.* 28: 308-324 https://doi.org/10.1016/j.biotechadv.2010.01.003.

Bai A, Zhao X, Jin Y, Yang G, et al. (2013). A novel thermophilic β-glucosidase from *Caldicellulosiruptor bescii*: Characterization and its synergistic catalysis with other cellulases. *J. Mol. Catal. B Enzym.* 85-86: 248-256.

Badal CS and Bothast RJ (1997). Enzymes in Lignocellulosic Biomass Conversion. In Fuels and Chemicals from Biomass (Vol. 666, pp. 46-56). American Chemical Society.

Barrett T, Suresh CG, Tolley SP, Dodson EJ, et al. (1995). The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. *Struct. Lond. Engl*. 3: 951-960.

Béguin P and Aubert JP (1994). The biological degradation of cellulose. *FEMS Microbiol. Rev.* 13: 25-58. https://doi.org/10.1111/j.1574-6976.1994.tb00033.x

Benoliel B, Poças-Fonseca MJ, Torres FAG and de Moraes LMP (2010). Expression of a Glucose-tolerant β-glucosidase from *Humicola grisea* var. thermoidea in Saccharomyces cerevisiae. *Appl. Biochem. Biotechnol.* 160: 2036-2044.

Berman HM, Westbrook J, Feng Z, Gilliland G, et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28: 235-242. https://doi.org/10.1093/nar/28.1.235

Bitar M and Franco GR (2014). A basic protein comparative three-dimensional modeling methodological workflow theory and practice. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 11: 1052-1065. https://doi.org/10.1109/TCBB.2014.2325018

Bohlin C, Olsen SN, Morant MD, Patkar S, et al. (2010). A comparative study of activity and apparent inhibition of fungal β-glucosidases. *Biotechnol. Bioeng.* 107: 943-952 https://doi.org/10.1002/bit.22885.

Breves R, Bronnenmeier K, Wild N, Lottspeich F, et al. (1997). Genes encoding two different beta-glucosidases of *Thermoanaerobacter brockii* are clustered in a common operon. *Appl. Environ. Microbiol.* 63: 3902-3910.

Cao LC, Wang ZJ, Ren GH, Kong W, et al. (2015). Engineering a novel glucose-tolerant β-glucosidase as supplementation to enhance the hydrolysis of sugarcane bagasse at high glucose concentration. *Biotechnol. Biofuels* 8: 202 https://doi.org/10.1186/s13068-015-0383-z.

Chamoli S, Kumar P, Navani NK and Verma AK (2016). Secretory expression, characterization and docking study of glucose-tolerant β-glucosidase from *B. subtilis*. *Int. J. Biol. Macromol.* 85: 425-433 https://doi.org/10.1016/j.ijbiomac.2016.01.001.

Cota J, Corrêa TLR, Damásio ARL, Diogo JA, et al. (2015). Comparative analysis of three hyperthermophilic GH1 and GH3 family members with industrial potential. *N. Biotechnol.* 32: 13-20 https://doi.org/10.1016/j.nbt.2014.07.009.

Crespim E, Zanphorlin LM, de Souza FHM, Diogo JA, et al. (2016). A novel cold-adapted and glucose-tolerant GH1 β-glucosidase from *Exiguobacterium antarcticum* B7. *Int. J. Biol. Macromol.* 82: 375-380 https://doi.org/10.1016/j.ijbiomac.2015.09.018.

de Giuseppe PO, Souza T de ACB, Souza FHM, Zanphorlin LM, et al. (2014). Structural basis for glucose tolerance in GH1 β-glucosidases. *Acta Crystallogr. D Biol. Crystallogr.* 70: 1631-1639 https://doi.org/10.1107/S1399004714006920.

DesJarlais RL, Sheridan RP, Seibel GL, Dixon JS, et al. (1988). Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* 31: 722-729. https://doi.org/10.1021/jm00399a006

Ewing TJA and Kuntz ID (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* 18: 1175-1189 https://doi.org/10.1002/(SICI)1096-987X(19970715)18:9<1175::AID-JCC6>3.0.CO;2-O.

Fang Z, Fang W, Liu J, Hong Y, et al. (2010). Cloning and characterization of a β-glucosidase from marine microbial metagenome with excellent glucose tolerance. *J. Microbiol. Biotechnol.* 20: 1351-1358.

Gumerov VM, Rakitin AL, Mardanov AV, Ravin NV, et al. (2015). A novel highly thermostable multifunctional beta-glycosidase from crenarchaeon *Acidilobus saccharovorans*, A novel highly thermostable multifunctional beta-glycosidase from crenarchaeon *Acidilobus saccharovorans*. *Archaea Archaea* 2015: e978632.

Guo B, Amano Y and Nozaki K (2016). Improvements in glucose sensitivity and stability of *Trichoderma reesei* β-glucosidase using site-directed mutagenesis. *PLoS One* 11: e0147301 https://doi.org/10.1371/journal.pone.0147301.

Henrissat B (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280: 309-316 https://doi.org/10.1042/bj2800309.

Huang Y, Busk PK, Grell MN, Zhao H, et al. (2014). Identification of a β-glucosidase from the *Mucor circinelloides* genome by peptide pattern recognition. *Enzyme Microb. Technol.* 67: 47-52 https://doi.org/10.1016/j.enzmictec.2014.09.002.

Jabbour D, Klippel B and Antranikian G (2012). A novel thermostable and glucose-tolerant β-glucosidase from

Fervidobacterium islandicum. *Appl. Microbiol. Biotechnol.* 93: 1947-1956 https://doi.org/10.1007/s00253-011-3406-0.

Jakalian A, Jack DB and Bayly CI (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 23: 1623-1641 https://doi.org/10.1002/jcc.10128.

Jeng W-Y, Wang N-C, Lin C-T, Chang W-J, et al. (2012). High-resolution structures of *Neotermes koshunensis* β-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates. *Acta Crystallogr. D Biol. Crystallogr.* 68: 829-838 https://doi.org/10.1107/S0907444912013224.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, et al. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res., 36(Web Server issue), W5-9.

Cairns JR and Esen A (2010). β-Glucosidases. *Cell. Mol. Life Sci.* 67: 3389-3405 https://doi.org/10.1007/s00018-010-0399-2.

Khairudin NBA and Mazlan NSF (2013). Molecular docking study of Beta-glucosidase with cellobiose, cellotetraose and cellotetriose. *Bioinformation* 9: 813-817 https://doi.org/10.6026/97320630009813.

Kitchenham B (2004). Procedures for Performing Systematic Reviews. Keele University, Keele 33: 1-26.

Kumar R, Singh S and Singh OV (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* 35: 377-391 https://doi.org/10.1007/s10295-008-0327-8.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6: e1000100 https://doi.org/10.1371/journal.pmed.1000100.

Liu J, Zhang X, Fang Z, Fang W, et al. (2011). The 184th residue of β-glucosidase Bgl1B plays an important role in glucose tolerance. *J. Biosci. Bioeng.* 112: 447-450 https://doi.org/10.1016/j.jbiosc.2011.07.017.

Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PIW, et al. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50: 437-450 https://doi.org/10.1002/prot.10286.

Lu J, Du L, Wei Y, Hu Y, et al. (2013). Expression and characterization of a novel highly glucose-tolerant β-glucosidase from a soil metagenome. *Acta Biochim. Biophys. Sin.* 45: 664-673.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, et al. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11: 3696-3713 https://doi.org/10.1021/acs.jctc.5b00255.

Mariano D, Leite C, Santos LHS, Rocha REO, et al. (2017). A guide to performing systematic literature reviews in bioinformatics (Technical Report No. RT.DCC.002/2017) (p. 63). Belo Horizonte, MG, Brazil: Universidade Federal de Minas Gerais.

Meleiro LP, Salgado JCS, Maldonado RF, Alponti JS, et al. (2015). A *Neurospora crassa* β-glucosidase with potential for lignocellulose hydrolysis shows strong glucose tolerance and stimulation by glucose and xylose. *J. Mol. Catal., B Enzym.* 122: 131-140 https://doi.org/10.1016/j.molcatb.2015.09.003.

Meng EC, Shoichet BK and Kuntz ID (1992). Automated docking with grid-based energy evaluation. *J. Comput. Chem.* 13: 505-524 https://doi.org/10.1002/jcc.540130412.

Murphy L, Bohlin C, Baumann MJ, Olsen SN, et al. (2013). Product inhibition of five *Hypocrea jecorina* cellulases. *Enzyme Microb. Technol.* 52: 163-169 https://doi.org/10.1016/j.enzmictec.2013.01.002.

Pearson WR (2016). Finding Protein and Nucleotide Similarities with FASTA. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al, 53, 3.9.1-3.9.25.

Pei J, Pang Q, Zhao L, Fan S, et al. (2012). Thermoanaerobacterium thermosaccharolyticum β-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose. *Biotechnol. Biofuels* 5: 31. https://doi.org/10.1186/1754-6834-5-31

Pentzold S, Zagrobelny M, Rook F and Bak S (2014). How insects overcome two-component plant chemical defence: plant β-glucosidases as the main target for herbivore adaptation. *Biol. Rev. Camb. Philos. Soc.* 89: 531-551 https://doi.org/10.1111/brv.12066.

Pettersen EF, Goddard TD, Huang CC, Couch GS, et al. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25: 1605-1612 https://doi.org/10.1002/jcc.20084.

Pires DEV, de Melo-Minardi RC, da Silveira CH, Campos FF, et al. (2013). aCSM: noise-free graph-based signatures to large-scale receptor-based ligand prediction. *Bioinformatics* 29: 855-861 https://doi.org/10.1093/bioinformatics/btt058.

Ramani G, Meera B, Rajendhran J and Gunasekaran P (2015a). Transglycosylating glycoside hydrolase family 1 β-glucosidase from *Penicillium funiculosum* NCL1: Heterologous expression in Escherichia coli and characterization. *Biochem. Eng. J.* 102: 6-13 https://doi.org/10.1016/j.bej.2015.03.018.

Ramani G, Meera B, Vanitha C, Rajendhran J, et al. (2015b). Molecular cloning and expression of thermostable glucose-tolerant β-glucosidase of *Penicillium funiculosum* NCL1 in Pichia pastoris and its characterization. *J. Ind. Microbiol. Biotechnol.* 42: 553-565 https://doi.org/10.1007/s10295-014-1549-6.

Rego N and Koes D (2015). 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 31: 1322-1324 https://doi.org/10.1093/bioinformatics/btu829.

Saha BC and Bothast RJ (1996). Production, purification, and characterization of a highly glucose-tolerant novel beta-glucosidase from *Candida peltata*. *Appl. Environ. Microbiol.* 62: 3165-3170.

Saha BC and Bothast RJ (1997). Enzymes in Lignocellulosic Biomass Conversion. *Fuels and Chemicals from Biomass* 666: 46-56.

Sanz-Aparicio J, Hermoso JA, Martínez-Ripoll M, Lequerica JL, et al. (1998). Crystal structure of beta-glucosidase a from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosyl hydrolases. *J. Mol. Biol.* 275: 491-502 https://doi.org/10.1006/jmbi.1997.1467.

Shatsky M, Nussinov R and Wolfson HJ (2004). A method for simultaneous alignment of multiple protein structures. *Proteins* 56: 143-156 https://doi.org/10.1002/prot.10628.

Sievers F, Wilm A, Dineen D, Gibson TJ, et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7: 539 https://doi.org/10.1038/msb.2011.75.

Singhania RR, Patel AK, Sukumaran RK, Larroche C, et al. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* 127: 500-507 https://doi.org/10.1016/j.biortech.2012.09.012.

Schröder C, Elleuche S, Blank S and Antranikian G (2014). Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. *Enzyme Microb. Technol.* 57: 48-54.

Souza FHM, Meleiro LP, Machado CB, Zimbardi ALRL, et al. (2014). Gene cloning, expression and biochemical characterization of a glucose- and xylose-stimulated β-glucosidase from *Humicola insolens* RP86. *J. Mol. Catal., B Enzym.* 106: 1-10 https://doi.org/10.1016/j.molcatb.2014.04.007.

Swangkeaw J, Vichitphan S, Butzke CE and Vichitphan K (2010). Characterization of β-glucosidases from Hanseniaspora sp. and *Pichia anomala* with potentially aroma-enhancing capabilities in juice and wine. *World J. Microbiol. Biotechnol.* 27: 423-430 https://doi.org/10.1007/s11274-010-0474-8.

Teugjas H and Väljamäe P (2013). Selecting β-glucosidases to support cellulases in cellulose saccharification. *Biotechnol. Biofuels* 6: 105. https://doi.org/10.1186/1754-6834-6-105

Tiwari P, Misra BN and Sangwan NS (2013). β -Glucosidases from the fungus trichoderma: an efficient cellulase machinery in biotechnological applications. *BioMed Res. Int.* 2013: 203735 https://doi.org/10.1155/2013/203735.

Uchima CA, Tokuda G, Watanabe H, Kitamoto K, et al. (2011). Heterologous expression and characterization of a glucose-stimulated β-glucosidase from the termite *Neotermes koshunensis* in *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 89: 1761-1771 https://doi.org/10.1007/s00253-010-2963-y.

Uchima CA, Tokuda G, Watanabe H, Kitamoto K, et al. (2012). Heterologous expression in *Pichia pastoris* and characterization of an endogenous thermostable and high-glucose-tolerant β-glucosidase from the termite *Nasutitermes takasagoensis*. *Appl. Environ. Microbiol.* 78: 4288-4293 https://doi.org/10.1128/AEM.07718-11.

Uchiyama T, Yaoi K and Miyazaki K (2015). Glucose-tolerant β-glucosidase retrieved from a Kusaya gravy metagenome. *Front. Microbiol.* 6: 548 https://doi.org/10.3389/fmicb.2015.00548.

Webb B and Sali A (2014). Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al, 47, 5.6.1-32.

Yang F, Yang X, Li Z, Du C, et al. (2015a). Overexpression and characterization of a glucose-tolerant β-glucosidase from T. aotearoense with high specific activity for cellobiose. *Appl. Microbiol. Biotechnol.* 99: 8903-8915 https://doi.org/10.1007/s00253-015-6619-9.

Yang Y, Zhang X, Yin Q, Fang W, et al. (2015b). A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases. *Sci. Rep.* 5: 17296 https://doi.org/10.1038/srep17296.

Zhao L, Pang Q, Xie J, Pei J, et al. (2013). Enzymatic properties of Thermoanaerobacterium thermosaccharolyticum β-glucosidase fused to *Clostridium cellulovorans* cellulose binding domain and its application in hydrolysis of microcrystalline cellulose. *BMC Biotechnol.* 13: 101. https://doi.org/10.1186/1472-6750-13-101

Supplementary material

Figure S1. Redocking of cellobiose using the flexible ligand docking protocol.

Figure S2. Crystallographic structures and models of β-glucosidases.

Figure S3. Docking of cellobiose using the flexible ligand docking protocol in all 21 structures.

Figure S4. Alignment among the catalytic pocket of glucose-tolerant β-glucosidases.

Table S1. PRISMA checklist.

Table S2. List of references, with main topics, groups, and scores.

Table S3. List of glucose-tolerant β-glucosidases collected by organism.

Table S4. Number of contacts between residues and the cellobiose docked in glucose-tolerant GH1 β-glucosidases.

Table S5. Matrix of the identity percentage among sequences using global alignment (algorithm Needleman-Wunsch).