Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structural basis for glucose tolerance in GH1 β-glucosidases

Priscila Oliveira de Giuseppe,[a]‡ Tatiana de Arruda Campos Brasil Souza,[a,b]‡ Flavio Henrique Moreira Souza,[c] Leticia Maria Zanphorlin,[d] Carla Botelho Machado,[c,e] Richard John Ward,[c,e] Joao Atilio Jorge,[f] Rosa dos Prazeres Melo Furriel[c] and Mario Tyago Murakami[a]*

[a]Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, CP 6192, 13083-970 Campinas-SP, Brazil, [b]Carlos Chagas Institute, Fiocruz, 81350-010 Curitiba-PR, Brazil, [c]Department of Chemistry, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, 14040-901 Ribeirão Preto-SP, Brazil, [d]Department of Organic Chemistry, Institute of Chemistry, State University of Campinas, 13083-970 Campinas-SP, Brazil, [e]Brazilian Bioethanol Science and Technology Laboratory, National Center for Research in Energy and Materials, Giuseppe Maximo Scolfaro 10000, 13083-970 Campinas-SP, Brazil, and [f]Department of Biology, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, 14040-901 Ribeirão Preto-SP, Brazil

‡ These authors contributed equally to this work.

Correspondence e-mail: mario.murakami@lnbio.cnpem.br

Received 19 September 2013
Accepted 28 March 2014

PDB references:
β-glucosidase, 4mdo; glucose complex, 4mdp

Product inhibition of β-glucosidases (BGs) by glucose is considered to be a limiting step in enzymatic technologies for plant-biomass saccharification. Remarkably, some β-glucosidases belonging to the GH1 family exhibit unusual properties, being tolerant to, or even stimulated by, high glucose concentrations. However, the structural basis for the glucose tolerance and stimulation of BGs is still elusive. To address this issue, the first crystal structure of a fungal β-glucosidase stimulated by glucose was solved in native and glucose-complexed forms, revealing that the shape and electrostatic properties of the entrance to the active site, including the +2 subsite, determine glucose tolerance. The aromatic Trp168 and the aliphatic Leu173 are conserved in glucose-tolerant GH1 enzymes and contribute to relieving enzyme inhibition by imposing constraints at the +2 subsite that limit the access of glucose to the −1 subsite. The GH1 family β-glucosidases are tenfold to 1000-fold more glucose tolerant than GH3 BGs, and comparative structural analysis shows a clear correlation between active-site accessibility and glucose tolerance. The active site of GH1 BGs is located in a deep and narrow cavity, which is in contrast to the shallow pocket in the GH3 family BGs. These findings shed light on the molecular basis for glucose tolerance and indicate that GH1 BGs are more suitable than GH3 BGs for biotechnological applications involving plant cell-wall saccharification.

1. Introduction

β-Glucosidases work synergistically with endoglucanases and cellobiohydrolases to convert cellulose to fermentable sugars (Tomme *et al.*, 1995) by cleaving cellobiose into two glucose molecules. This activity is a key factor in preventing the inhibition of upstream enzymes by cellobiose (Lamed *et al.*, 1991); however, inhibition of β-glucosidases by glucose has been a limiting factor in enzyme-based saccharification technologies using plant biomass as a feedstock.

The discovery and biochemical characterization of novel microbial β-glucosidases have been the subject of intensive efforts (Bhatia *et al.*, 2002; Singhania *et al.*, 2013). This is in part owing to their essential role in biochemical technologies for the production of biofuels from plant lignocellulosic material (Singhania *et al.*, 2013). Advances in this field include the identification of β-glucosidases with desirable catalytic properties such as high hydrolytic activity (Chen *et al.*, 2012; Pei *et al.*, 2012), thermotolerance and resistance to reduced pH (Hong *et al.*, 2009; Souza *et al.*, 2010).

In recent decades, another attractive attribute has been identified in some β-glucosidases: that of glucose tolerance (Saha & Bothast, 1996; Zanoelo *et al.*, 2004; Fang *et al.*, 2010; Souza *et al.*, 2010; Pei *et al.*, 2012; Uchiyama *et al.*, 2013). This

feature is of considerable economic interest because it can decrease the cellulase and β-glucosidase loads needed to attain acceptable yields during lignocellulose hydrolysis, thereby reducing the costs of biofuel production (Singhania *et al.*, 2013).

Glucose tolerance in β-glucosidases is sometimes coupled to a stimulatory effect of glucose on substrate degradation in a given range of glucose concentrations (up to 1 *M*). This property seems to be exclusive to some GH1 (glycosyl hydrolase family 1; Henrissat, 1991; Cantarel *et al.*, 2009) β-glucosidases, and enzymatic assays have suggested that the stimulation is regulated either through an allosteric effect by glucose binding to a secondary site (Zanoelo *et al.*, 2004; Souza *et al.*, 2010) or by kinetic modulation of enzyme turnover associated with transglycosylation (Uchima *et al.*, 2011, 2013). Furthermore, the family GH3 β-glucosidases known to date are all inhibited by glucose. However, the structural bases for the effects of glucose on GH1 and GH3 β-glucosidases remain poorly understood (Liu *et al.*, 2011).

To address this issue, we have solved the first crystal structure of a fungal β-glucosidase stimulated by glucose in complex with this monosaccharide. Our target was the glucose-stimulated and xylose-stimulated β-glucosidase from the thermophilic fungus *Humicola insolens* (HiBG; Souza *et al.*, 2010). This GH1-family enzyme (Henrissat, 1991; Cantarel *et al.*, 2009) presents approximately twofold activation at 50 and 100 m*M* concentrations of glucose or xylose, respectively, and maintains full or higher activity up to 450 m*M* glucose or 730 m*M* xylose (Souza *et al.*, 2013). Moreover, HiBG shows high catalytic efficiency for cellobiose hydrolysis and good thermal stability, thus being a very attractive biocatalyst for industrial use in plant-biomass saccharification processes (Souza *et al.*, 2010).

Our results shed light on the structural determinants for glucose tolerance in β-glucosidases and highlight the importance of the steric constraints imposed by the unique active-site entrance and the +2 subsite which limit glucose access to the −1 subsite. The geometry of the active-site pocket may support a secondary catalytic transglycosylation reaction using glucose as acceptor, which is in agreement with the stimulatory mechanism previously demonstrated for Td2F2, a GH1 β-glucosidase recently isolated from a microbial metagenomic sample (Uchima *et al.*, 2011, 2013). Since structure-based protein-engineering methods are proven strategies to obtain improved enzymes (Bornscheuer *et al.*, 2012; Lee *et al.*, 2012; Lundemo *et al.*, 2013), our results offer valuable information to guide rational approaches aiming to integrate glucose tolerance/activation with other desirable properties to generate catalytically enhanced β-glucosidases for biofuel production and other industrial applications.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Native	Glucose complex
PDB code	4mdo	4mdp
Data collection		
Space group	$P6_522$	$P6_522$
Unit-cell parameters		
$a = b$ (Å)	113.54	113.54
c (Å)	179.06	178.74
$\alpha = \beta$ (°)	90	90
γ (°)	120	120
Resolution (Å)	30.00–2.60 (2.69–2.60)	30.00–2.05 (2.12–2.05)
$R_{\rm merge}$†	0.11 (0.41)	0.10 (0.43)
$\langle I/\sigma(I)\rangle$	12.7 (3.2)	16.5 (4.0)
Completeness (%)	98.7 (98.6)	99.1 (97.5)
Multiplicity	5.0 (4.7)	7.1 (6.2)
Refinement		
Resolution (Å)	28.95–2.60	28.72–2.05
No. of reflections	20179	40196
$R_{\rm work}/R_{\rm free}$‡	0.17/0.24	0.16/0.21
No. of atoms		
Protein	3833	3859
Ligand/ion	20	49
Water	214	348
Mean B factors (Å^2)		
Protein	30.26	24.55
Ligand/ion	47.46	42.53
Water	26.47	33.09
R.m.s. deviations		
Bond lengths (Å)	0.011	0.015
Bond angles (°)	1.420	1.662
Ramachandran plot§		
Favoured (%)	95.37	96.84
Allowed (%)	4.63	3.16
Disallowed (%)	0.00	0.00

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm work} = \sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$; $R_{\rm free}$ was calculated using a 5% test set. § As calculated using *MolProbity* (Chen *et al.*, 2010).

2. Materials and methods

2.1. Protein expression and purification

Escherichia coli BL21 (DE3) cells transformed with the *bgl4* gene from *H. insolens* cloned into the pET-28a vector as described in Souza *et al.* (2013) were grown at 37°C with 200 rev min^{-1} agitation to an OD_{600} of 0.6 in 1 l shake flasks containing 200 ml HDM medium (25 g l^{-1} yeast extract, 15 g l^{-1} tryptone, 10 m*M* $MgSO_4$) supplemented with 40 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol. After 5 h induction with 1 m*M* IPTG (isopropyl β-D-1-thiogalactopyranoside), the cells were pelleted and transferred to 10 ml lysis buffer (50 m*M* NaH_2PO_4 pH 8.0, 1 m*M* phenylmethylsulfonyl fluoride, 150 m*M* NaCl). After disruption by sonication, the cell debris was removed by centrifugation and the supernatant was applied onto a nickel affinity column (His-Select Nickel Affinity Gel, Sigma–Aldrich) equilibrated with lysis buffer. The protein was purified using 10 and 200 m*M* imidazole in the wash and elution buffers, respectively. The eluted protein was visualized by SDS–PAGE (Laemmli, 1970) and the protein concentration was determined by the method of Read & Northcote (1981).

2.2. Protein crystallization

The recombinant β-glucosidase from *H. insolens* (HiBG) was concentrated to 3 mg ml^{-1} in 50 m*M* phosphate buffer pH 6.0 using an Amicon Ultra-4 10K centrifugal filter (Millipore). The HiBG–glucose complex was prepared by overnight

incubation of the protein with the ligand at 5 m*M*. Crystallization experiments were carried out in 96-well sitting-drop plates using a Honeybee 963 automated system (Digilab). Drops consisting of 1 µl protein sample mixed with an equal volume of reservoir solution were equilibrated against 80 µl of the same reservoir solution. 544 crystallization conditions formulated based on the commercial kits Crystal Screen HT (Hampton Research), Wizard I and II (Emerald Bio), The PACT Suite (Qiagen), JCSG-*plus* (Molecular Dimensions), SaltRX HT (Hampton Research) and Precipitant Synergy Primary (Emerald Bio) were tested. Native crystals were obtained from a solution consisting of 2 *M* lithium sulfate, 2%(*v*/*v*) PEG 400, 0.1 *M* Tris pH 8.5, whereas crystals of HiBG in complex with glucose were grown in 10%(*w*/*v*) PEG 3000, 0.1 *M* sodium acetate pH 4.5, 0.2 *M* zinc acetate. Native and glucose-complexed crystals grew in 5–7 d.

2.3. Data collection, processing and structure determination

Crystals were soaked in a cryoprotectant solution [*i.e.* crystallization solution plus 20%(*v*/*v*) PEG 400 or glycerol] for 30 s and were then flash-cooled in a nitrogen-gas stream at 100 K. X-ray diffraction data were collected on the W01B-MX2 beamline at the Brazilian Synchrotron Light Laboratory, Campinas, Brazil. Data were indexed, integrated, merged and scaled using the *HKL*-2000 package (Otwinowski & Minor, 1997). The structures were solved by the molecular-replacement method using *MOLREP* (Vagin & Teplyakov, 2010) with the atomic coordinates of β-glucosidase 2 from *Trichoderma reesei* (74% sequence identity; PDB entry 3ahy; Jeng *et al.*, 2011) as the template. The atomic coordinates were examined and manually fitted based on the $2F_o - F_c$ and $F_o - F_c$ electron-density maps using *Coot* (Emsley *et al.*, 2010) and were refined using *REFMAC*5 (Murshudov *et al.*, 2011). Data-collection and refinement statistics are detailed in Table 1. The atomic coordinates and structure factors of HiBG in the native state and in complex with glucose have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/) under accession codes 4mdo and 4mdp, respectively. The figures were generated using *PyMOL* (http://www.pymol.org). Active-site volumes were calculated using *POCASA* (Yu *et al.*, 2010).

2.4. Small-angle X-ray scattering (SAXS) data collection and analysis

SAXS data were collected on the D02A/SAXS2 beamline at the Brazilian Synchrotron Light Laboratory (LNLS). The radiation wavelength was set to 1.488 Å and a charge-coupled device area detector (MAR CCD 165 mm) was used to record the scattering patterns. The sample-to-detector distance was set to give a scattering-vector range from 0.15 to 2.43 nm^{-1}. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the radiation wavelength. Protein samples were prepared in 20 m*M* Tris–HCl buffer pH 7.5 at three different concentrations (2, 4 and 6 mg ml^{-1}). Prior to X-ray exposure, the samples were centrifuged at 10 000*g* for 10 min and filtered to remove any aggregates. Buffer baselines were collected under identical conditions before and after sample data collection to guarantee an accurate solvent correction. Frames were recorded with an exposure time of 600 s to avoid radiation-induced protein damage, taking into account the radiation brilliance at the SAXS2 beamline. The

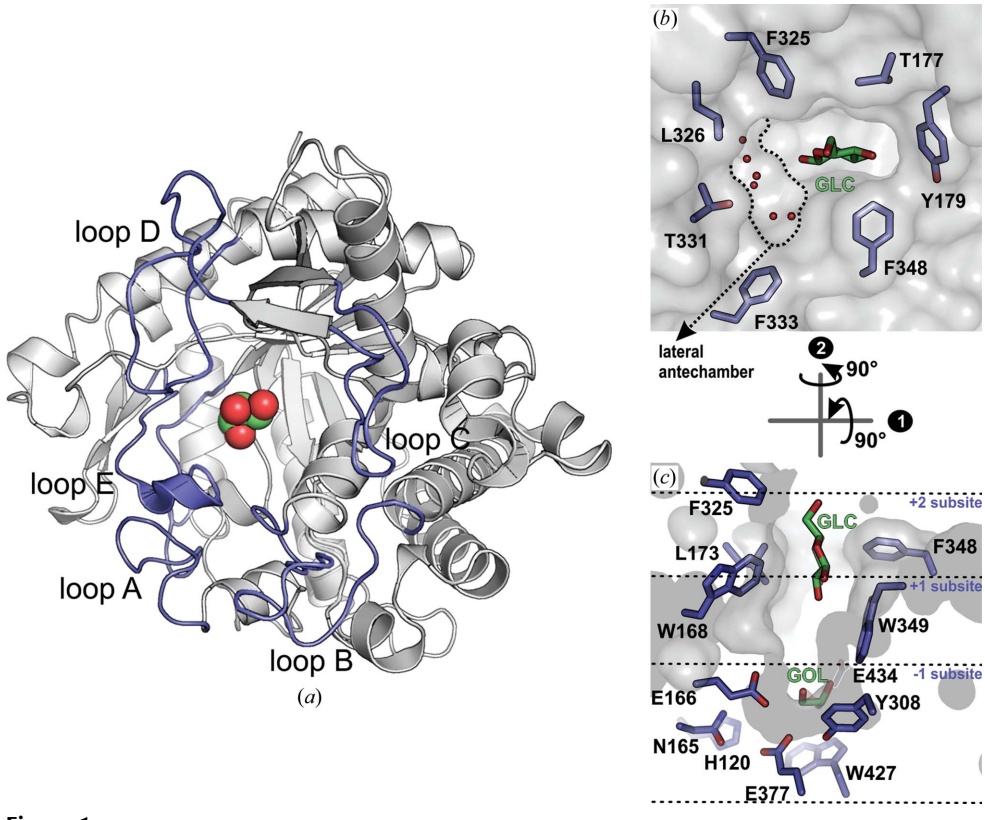


Figure 1
The catalytic interface of HiBG. (*a*) Ribbon representation of HiBG with the loops surrounding the active site coloured light blue. The glycerol molecule occupying the −1 subsite is represented as spheres (green, carbon; red, oxygen). (*b*) The active-site entrance highlighting the gatekeeper residues (light blue C atoms), the water network (red spheres) in the lateral antechamber and the glucose molecule (GLC; green C atoms) bound at the aglycone-binding site. (*c*) Residues forming the glycone-binding site (−1 subsite) where glycerol (GOL) is bound and the aglycone-binding site (+1 and +2 subsites) where glucose is trapped by hydrophobic interactions with Trp168, Leu173 and Phe348. The views in (*c*) and (*b*) are related by a 90° rotation about the horizontal and the vertical axes, respectively, as indicated by the arrows.

experimental intensities were corrected for background, buffer contributions, detector inhomogeneities and sample transmission using the program *FIT2D* v.12.081 (Hammersley *et al.*, 1996). The radius of gyration (R_g) was evaluated using the Guinier approximation (Guinier & Fournet, 1955) as implemented in *PRIMUS* (Konarev *et al.*, 2003). Fitting of the experimental data and evaluation of the pair-distance distribution function $p(r)$ were performed using *GNOM* (Svergun, 1992). The maximum dimension D_{max} was estimated from the real-space $p(r)$ function as the distance r where the $p(r)$ value reaches zero. SAXS dummy-atom models were determined using *ab initio* modelling as implemented in *DAMMIN* (Svergun, 1999). At least ten SAXS envelopes were calculated using the *DAMAVER* suite of programs (Volkov & Svergun, 2003). The low-resolution model and the crystal structure were superimposed using *SUPCOMB* (Kozin & Svergun, 2001). The calculation of the theoretical scattering curve from crystallographic atomic coordinates for subsequent comparison with the experimental scattering curve was performed using *CRYSOL* (Svergun *et al.*, 1995).

2.5. Dynamic light-scattering (DLS) measurements

The hydrodynamic behaviour of HiBG was also assessed by dynamic light scattering (DLS). DLS experiments were performed in a DynaPro 810 (Protein Solutions) system equipped with a Peltier module for temperature control. The protein sample was prepared at a final concentration of 1 mg ml^{-1} in 20 m*M* Tris–HCl buffer pH 7.5. 100 measurements at intervals of 20 s were recorded and averaged for further calculations of hydrodynamic parameters using *DYNAMICS* v.6.3.40. The hydrodynamic radius (R_H) was extrapolated from the translational diffusion coefficient (D_T) using the Stokes–Einstein equation.

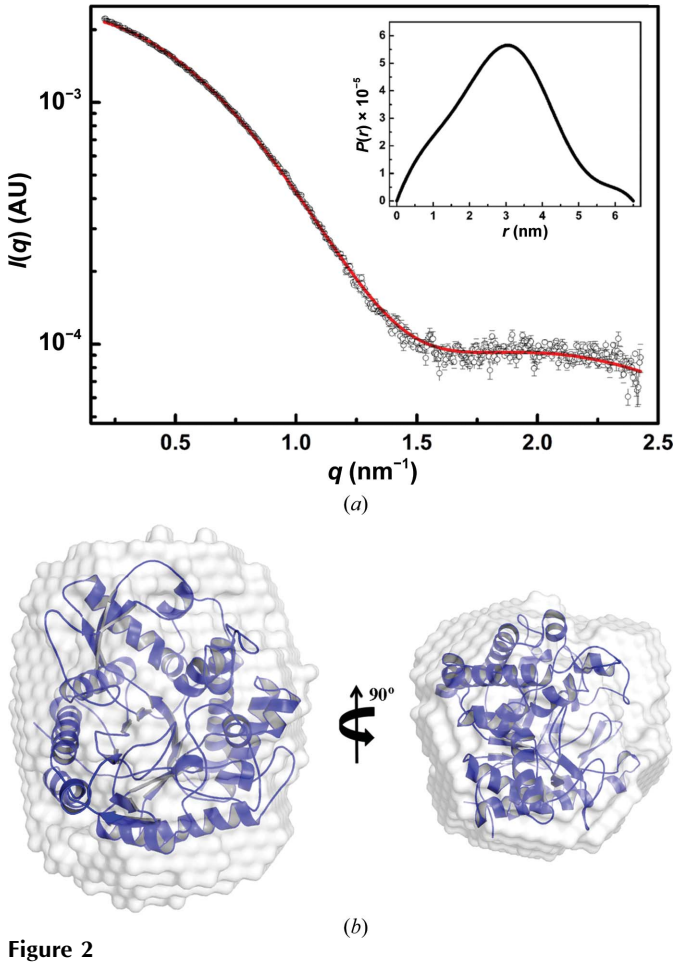
3. Results and discussion

3.1. The monomeric α_8/β_8-barrel structure of the glucose-tolerant β-glucosidase HiBG

The native and glucose-complex crystals were isomorphous ($P6_522$) and diffracted to 2.60 and 2.05 Å resolution, respectively (Table 1). The structures are very similar and superimposed with an r.m.s.d. of 0.12 Å for 477 C^α atoms. HiBG conserves the canonical TIM-barrel fold of GH1 β-glucosidases comprised of eight parallel β-strands surrounded by eight α-helices (Fig. 1*a*). Short helices and strands decorate the conserved structural core and five long loops contribute to a unique design of the catalytic face that surrounds a deep pocket which harbours the active site (Fig. 1). HiBG also contains a 12-residue extended loop (loop D) that contributes to a protein–protein interface formed between HiBG monomers. Despite this quaternary arrangement in the crystalline state (Supplementary Fig. S1[1]), the SAXS and DLS data showed that HiBG is monomeric in solution, as typically observed for GH1 β-glucosidases (Fig. 2). DLS analysis indicated that the enzyme is monodisperse in solution (polydispersity = 12.5%), exhibiting an R_H of 28.1 ± 0.4 Å. From the SAXS scattering and pair-distance distribution curves (Fig. 2*a*), it is possible to observe a maximum molecular dimension (D_{max}) of approximately 65 Å and a radius of gyration of 23.6 Å, which are in agreement with the DLS analysis. Moreover, fitting of the crystallographic monomer into the *ab initio* SAXS envelope indicated a good shape complementarity, also supporting the monomeric state (Fig. 2*b*). These results demonstrate that quaternary-structure differences can be discarded as a molecular event related to glucose tolerance and stimulation in HiBG.

3.2. The singular active site of glucose-tolerant GH1 β-glucosidases

Structure and sequence alignments of HiBG with well characterized GH1 β-glucosidases (cyanogenic BG, PDB entry 1cbg, Barrett *et al.*, 1995; *Agrobacterium* sp. BG, UniProt ID P12614) show that HiBG conserves the retaining catalytic mechanism typical of GH1 enzymes, with Glu377 acting as a



(*a*)

(*b*)

Figure 2
SAXS analysis. (*a*) Experimental (open black circles with errors bars) and regularized (red solid line) scattering curves. Inset: the distance distribution function calculated using *GNOM*. (*b*) Different views of the crystallographic atomic coordinates (blue ribbon) fitted into the averaged SAXS envelope (grey transparent surface).

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: XB5079).

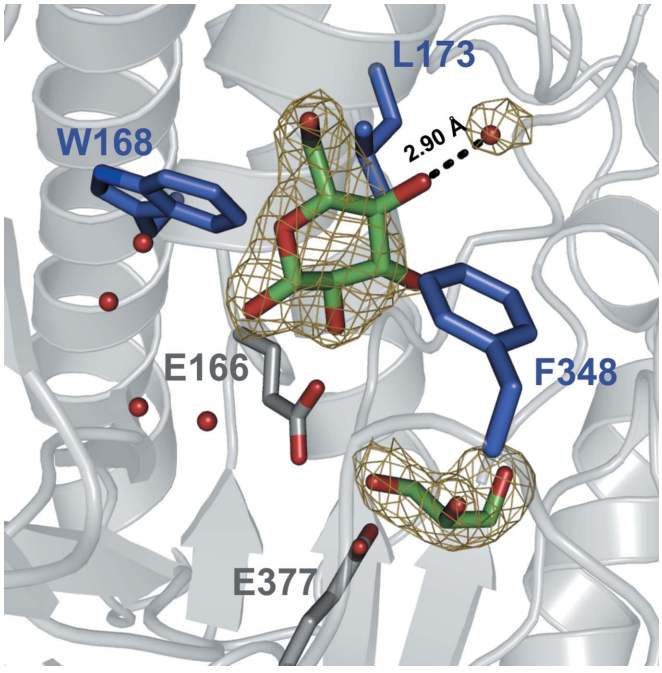


Figure 3
Electron-density map at the 1.0σ level (yellow) of glucose, glycerol and water 312. The panel also shows the acidic catalytic residues (Glu166 and Glu377; grey C atoms), the glucose-binding site (Trp168, Leu173 and Phe348; light blue C atoms) and other solvent molecules in the vicinity of glucose.

nucleophile and Glu166 as an acid/base (Withers *et al.*, 1990; Wang *et al.*, 1995). These residues lie at the bottom of a deep pocket located at the C-terminal side of the barrel and form the glycone-binding site together with Gln17 and the highly conserved His120, Trp121, Asn165, Tyr308, Trp427, Glu434, Trp435 and Phe443 residues.

In contrast to the glycone-binding site, the aglycone-binding site is less conserved. In the HiBG structure this site is narrow and surrounded mainly by aromatic and other neutral residues. Three phenylalanine residues (Phe325, Phe333 and Phe348) together with Thr177, Tyr179, Leu326 and Thr331 function as gatekeepers and demarcate a restricted entrance to the active site (Fig. 1*b*). Furthermore, a network of water molecules interacts with the side chains of residues defining a lateral antechamber, which may play a role in solvent dynamics during the catalytic cycle (Figs. 1*b* and 3).

In the native HiBG structure, a Tris and a PEG molecule are bound to the glycone-binding and aglycone-binding sites, respectively. In the HiBG–glucose complex structure a glycerol molecule occupies the glycone-binding site (−1 subsite), whereas a glucose molecule is trapped in the aglycone-binding site (Fig. 1*c*). Hydrogen bonds to Glu434 and Glu166 as well as hydrophobic interactions with Trp427 and Tyr308 stabilize the binding of glycerol to the catalytic pocket. The glycerol hydroxyl groups adopt similar positions of the glucose OH groups as observed in glucose covalently

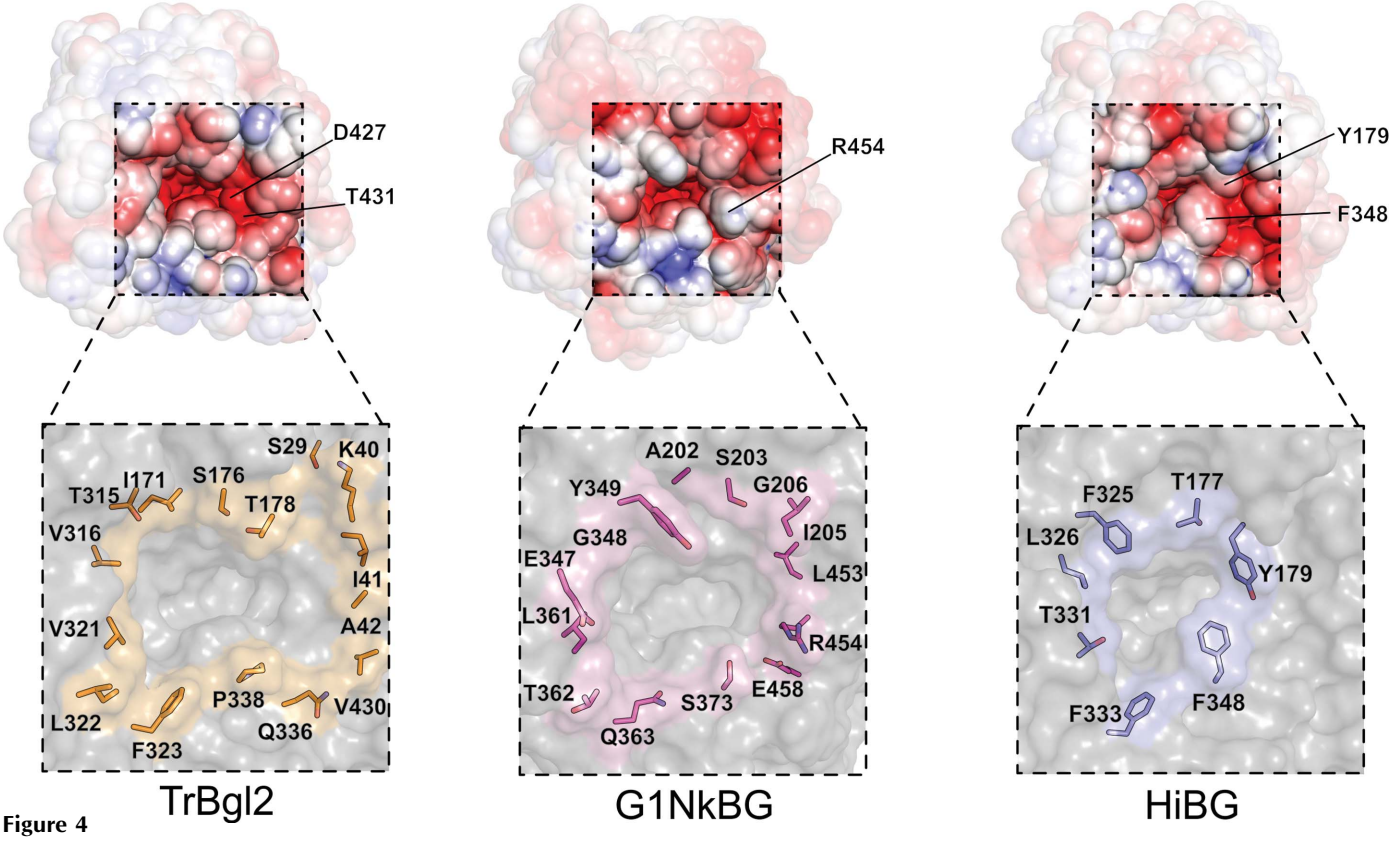


Figure 4
Structural comparison of the active-site entrances of TrBgl2 (PDB entry 3ahy chain *B*), G1NkBG (PDB entry 3vik) and HiBG. The upper row shows the electrostatic surface, whereas the lower row highlights the residues that delineate the respective active-site entrances. Electrostatic potentials were calculated by *PBEQ Solver* (Jo, Vargyas *et al.*, 2008) using the *CHARMM-GUI* interface (Jo, Kim *et al.*, 2008). The residues Arg454 in G1NkBG and Tyr179 and Phe348 in HiBG that overlap the electronegative patch formed by Asp42 and Thr431 in TrBgl2 are indicated in the upper panels.

bound to the nucleophile of a human β-glucosidase (PDB entry 2zox; Noguchi *et al.*, 2008), mimicking the glycosyl-enzyme intermediate state.

The bound glucose molecule was stabilized at the +2 subsite of the aglycone-binding site mainly by hydrophobic interactions with Trp168, Leu173 and Phe348 (Figs. 1*c* and 3). Only the 4-OH group of glucose makes hydrogen-bond contacts with a water molecule. However, this water does not interact with protein residues and has a higher *B* factor (54 Å^2) than average (33 Å^2), suggesting either high mobility or lower occupancy. It is therefore likely that the absence of stable polar interactions permits multiple ways of accommodating glucose at this site. This observation is compatible with the formation of glycosidic linkages at different positions (β1,2 and β1,3) between the anomeric C atom of a glycosyl-enzyme intermediate and a nucleophilic –OH group of the acceptor glucose bound at the +1 subsite.

3.3. The structural basis for glucose tolerance and stimulation

To investigate the structural basis for glucose tolerance and stimulation in GH1 β-glucosidases, we compared the crystal structure of HiBG, which tolerates up to 450 m*M* glucose with maximum stimulation at 50 m*M* glucose, with the structures of *T. reesei* β-glucosidase 2 (TrBgl2), which displays low glucose tolerance and no glucose stimulation (Jeng *et al.*, 2011; Lee *et al.*, 2012), and *Neotermes koshunensis* β-glucosidase (G1NkBG), which tolerates up to 600 m*M* glucose with maximum stimulation at 200 m*M* glucose (Uchima *et al.*, 2011; Jeng *et al.*, 2012).

The crystal structure of HiBG superimposed on those of TrBgl2 (PDB entry 3ahy; 74% sequence identity; Jeng *et al.*, 2011) and G1NkBG (PDB entry 3vik; 38% sequence identity; Jeng *et al.*, 2012) with r.m.s.d.s of 0.524 Å (461 aligned C$^\alpha$ atoms) and 1.161 Å (436 aligned C$^\alpha$ atoms), respectively. Comparative analysis of their active sites indicated that the shape and electrostatic properties of the deep active-site entrance and the +2 subsite determine the degree of tolerance to glucose. TrBgl2, with a glucose inhibition constant of 48.5 m*M* (Lee *et al.*, 2012), has the widest active-site entrance and presents an electronegative patch formed by Asp427 and Thr431 (Fig. 4*a*). Moreover, the active-site volume of TrBgl2 (917 Å^3) is considerably larger than those of G1NkBG (643 Å^3) and HiBG (381 Å^3). These characteristics may facilitate access to the catalytic site, favouring the competitive inhibition of TrBgl2 by glucose.

In contrast, G1NkBG (high glucose tolerance) presents a narrower and more neutral entrance (Fig. 4*b*). The presence of

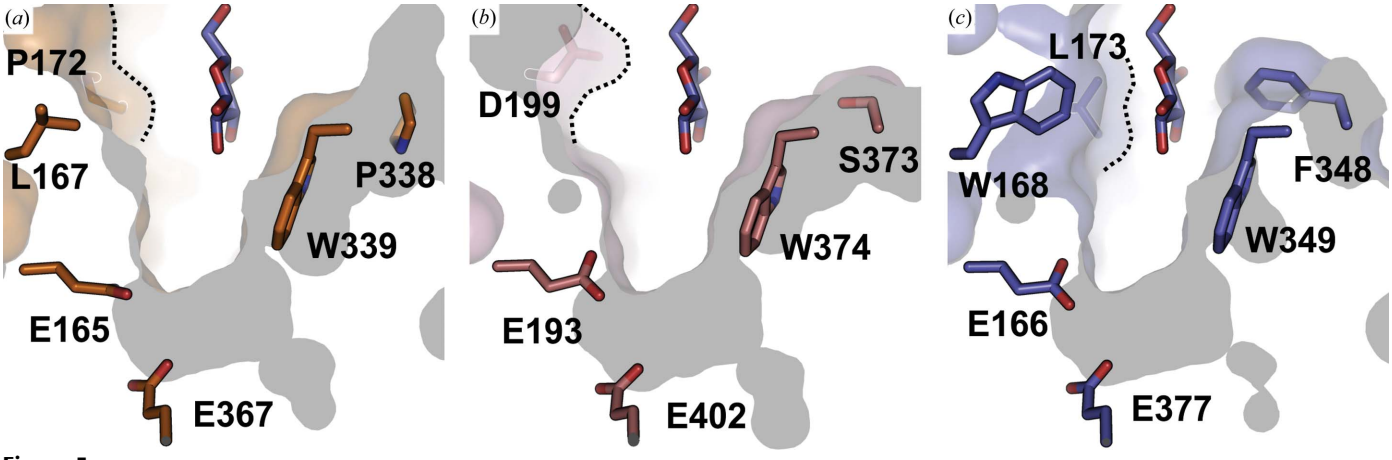


Figure 5
Lateral views of the catalytic pockets of TrBgl2 (PDB entry 3ahy chain *B*; orange C atoms), G1NkBG (PDB entry 3ahz; pink C atoms) and HiBG (light blue C atoms). For purposes of comparison, the glucose (light blue C atoms) observed in the HiBG active site was superposed onto the TrBgl2 and G1NkBG structures. The catalytic pairs (Glu165/Glu367, Glu193/Glu402 and Glu166/Glu377) and the fully conserved tryptophan that serves as a platform to stabilize the aglycone moiety at the +1 subsite (Trp339, Trp374 and Trp349) are shown as sticks together with the nonconserved residues from the aglycone-binding site and from the active-site entrance that may play a role in glucose tolerance.

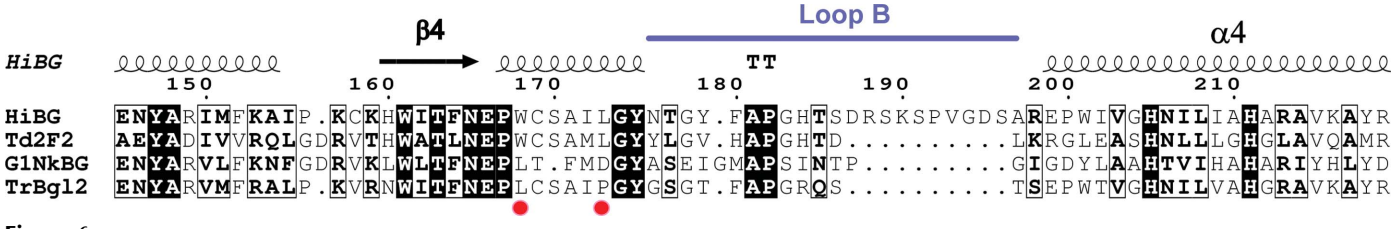


Figure 6
Sequence alignment of GH1 β-glucosidases with different levels of glucose tolerance. HiBG, *H. insolens* β-glucosidase (BG); Td2F2, BG from a compost metagenomic library (GenBank accession No. HV538882); G1NkBG, *N. koshunensis* BG1 (GenBank accession No. AB073638); TrBgl2, *T. reesei* BG2 (NCBI accession No. AB003110). Red spots indicate amino-acid positions that seem to play a role in glucose tolerance. Trp168 and Leu173 of HiBG are conserved in Td2F2; both are glucose-tolerant enzymes ($\geq$450 m*M*). When mutated at these positions (Leu167Trp and Pro172Leu), TrBgl2 (glucose K_i = 48.5 m*M*) showed increased glucose tolerance (3.0-fold and 1.6-fold, respectively; Lee *et al.*, 2012).

the Arg454 residue disrupts the electronegative patch observed in TrBgl2, while the substitution of two prolines by Ser373 and Asp199 confers flexibility and may modulate the entry of glucose to the active site (Figs. 4 and 5). In a similar way to G1NkBG, some gatekeeper residues in HiBG have larger side chains (Fig. 4*c*), such as Tyr179 and Phe348 suppress the electronegative patch observed at the entrance to the TrBgl2 active site (Fig. 4). At the HiBG +2 subsite, a tryptophan (Trp168) replaces a leucine (Leu167 in TrBgl2), narrowing the channel that leads to the catalytic acidic residues (Fig. 5). In addition to restricting access, the HiBG gatekeepers along with Trp168 and Leu173 select a preferential orientation for glucose to enter the active site (Figs. 4*c* and 5).

The hydrophobic interactions of Trp168 and Leu173 with glucose in the HiBG aglycone-binding site suggest a role for these residues in determining glucose tolerance (Fig. 3). Site-directed mutagenesis of TrBgl2 supports this hypothesis, showing that Leu167Trp (corresponding to Trp168 in HiBG) and Pro172Leu (corresponding to Leu173 in HiBG) mutations relieve glucose inhibition (Lee *et al.*, 2012). Furthermore, the highly glucose-tolerant β-glucosidase Td2F2 (Uchiyama *et al.*, 2013) also conserves Trp168 and Leu173 of HiBG, which further supports the role of these residues in glucose tolerance (Fig. 6).

An additional mechanism that is likely to be associated with tolerance to glucose in GH1 BGs is the transglycosylation activity. For instance, transglycosylation has been shown to play a role in glucose tolerance and stimulation in Td2F2 (Uchiyama *et al.*, 2013). According to this mechanism, at sufficiently high glucose concentrations the transglycosylation process is favoured and contributes to maintaining the enzyme turnover in spite of the high glucose content.

Interestingly, the predominant transglycosylation products of Td2F2 at elevated glucose concentration are sophorose (with a β1,2-glycosidic bond), followed by laminaribiose (with a β1,3-glycosidic bond) (Uchiyama *et al.*, 2013). This observation is in agreement with the orientation assumed by the glucose at the HiBG +2 subsite reported here (Fig. 3). The crystal structure of the HiBG–glucose complex shows that on entering the channel leading to the active site, the glucose −OH groups at positions 2 and 3 are oriented towards the catalytic site. On binding to the +1 subsite, these nucleophilic groups might promote the formation of β1,2 and β1,3 bonds *via* transglycosylation, as biochemically demonstrated for Td2F2 (Uchiyam *et al.*, 2013). In this model, the glucose acts as a glycosyl acceptor attacking the anomeric C atom from the glucosyl-enzyme intermediate, restoring the active site for a new catalytic cycle. The glucose –OH groups are protected from the solvent by the deep and narrow active site and are the favoured

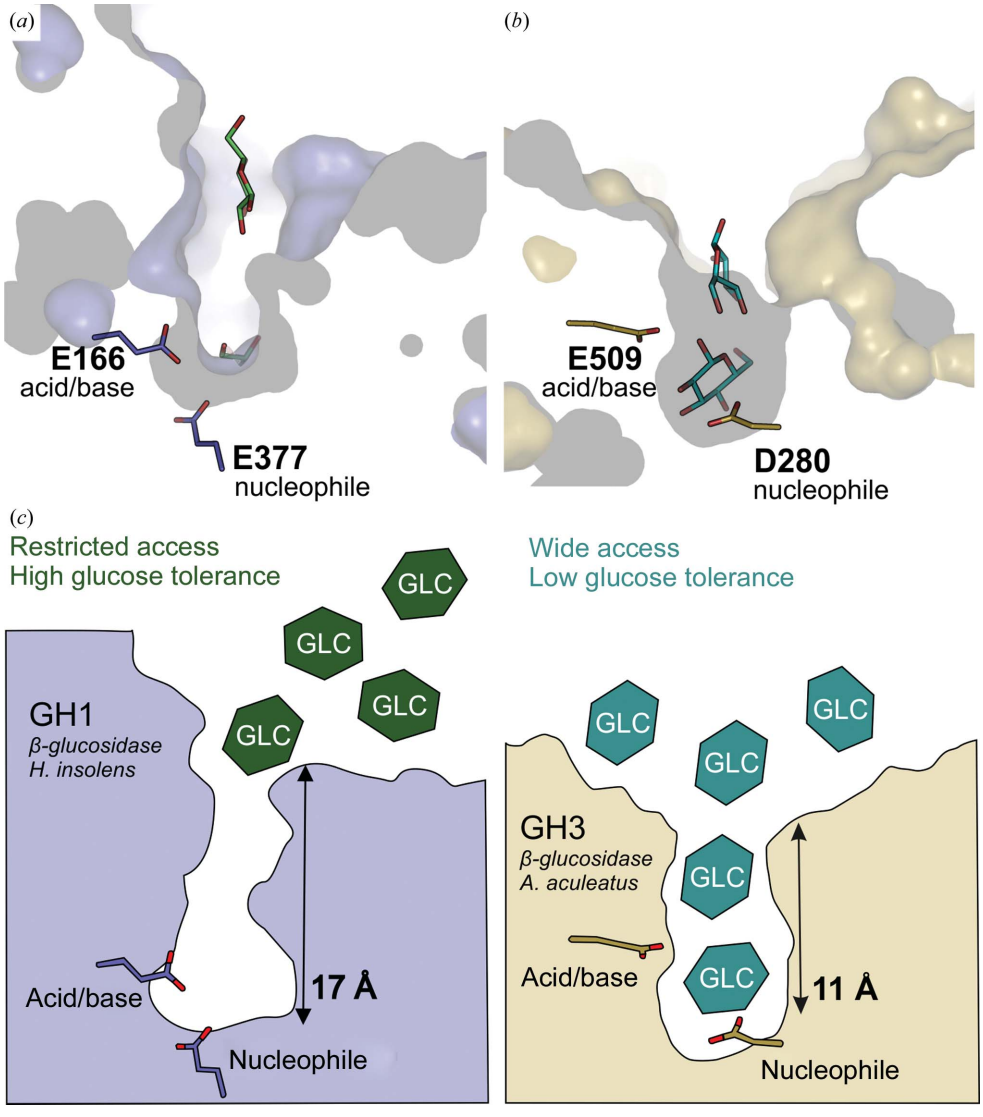


Figure 7
Active-site comparison between GH1 and GH3 β-glucosidases. Lateral views of (*a*) the GH1 HiBG active site occupied by glycerol and glucose (green C atoms) and (*b*) the GH3 *A. aculeatus* BG1 (AaBGL1) active site occupied by two glucose molecules (cyan C atoms; PDB entry 4iig; Suzuki *et al.*, 2013). (*c*) shows a schematic view of the GH1 and GH3 active sites and the mechanism proposed for glucose tolerance in β-glucosidases. Arrows illustrate the depth of the nucleophiles Glu377 (17 Å) and Asp280 (11.3 Å) calculated by *DEPTH* (Tan *et al.*, 2013). GLC represent glucose molecules.

nucleophiles during the catalytic cycle, liberating the enzyme from the intermediate state more rapidly and resulting in higher hydrolytic rates when the release of the aglycone moiety is monitored.

3.4. Why are GH1 β-glucosidases more tolerant to glucose than GH3 β-glucosidases?

GH1 BGs generally are tenfold to 1000-fold more glucose-tolerant than GH3 BGs (Decker *et al.*, 2000; Karnchanatat *et al.*, 2007; Fang *et al.*, 2010; Krogh *et al.*, 2010; Uchima *et al.*, 2011, 2013; Pei *et al.*, 2012; Bohlin *et al.*, 2013) and this seems to be the consequence of the ease of access of glucose to the catalytic site. The GH1 BGs harbour the active site in a deep and narrow cavity (381 Å^3), whereas in GH3 BGs the active site lies in a shallow pocket (297 Å^3). In HiBG, for example, the depth of the nucleophile residue is 17 Å compared with 11.3 Å in the GH3 *A. aculeatus* BG1, whose K_i (inhibition constant) for glucose is 3.7 m*M* (Decker *et al.*, 2000; Fig. 7).

The shallow aglycone-binding site of GH3 BGs enables glucose to easily access the −1 subsite. When bound at this subsite, glucose blocks the catalytic cycle, acting as a competitive inhibitor. In contrast, the deep and narrow aglycone-binding site in the GH1 BGs functions to constrain the path of glucose to the −1 subsite (Fig. 7). Consequently, in GH1 BGs higher concentrations of glucose are needed to promote competitive inhibition between glucose and the substrate.

Although stimulatory concentrations of glucose may shift the catalytic reaction from hydrolysis to transglycosylation in GH1 BGs, this change probably causes less impact on hydrolytic yields than the product inhibition in GH3 BGs (Bohlin *et al.*, 2013). Moreover, the transglycosylation products of GH1 BGs are substrates for the hydrolytic reaction, being fully converted to glucose in conditions favouring hydrolysis (Uchiyama *et al.*, 2013). Here, we have reported the first crystal structure of the glucose-stimulated HiBG family GH1 β-glucosidase in the native state and complexed with glucose, which has revealed key aspects of the structural bases of glucose stimulation in GH1 BGs. This will be useful for the design of catalytically enhanced BGs for integration into enzymatic platforms for cellulosic ethanol production, where glucose tolerance is a highly desirable property.

We are grateful to the Brazilian Biosciences National Laboratory (LNBio) and the Brazilian Synchrotron Light Laboratory (LNLS) for the provision of time on the MX2 beamline, Robolab and LEC. This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant Nos. 09/08312-6 and 13/13309-0 to MTM), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant Nos. 478059/2009-4 and 486841/2012-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J. & Hughes, M. A. (1995). *Structure*, **3**, 951–960.
Bhatia, Y., Mishra, S. & Bisaria, V. S. (2002). *Crit. Rev. Biotechnol.* **22**, 375–407.
Bohlin, C., Praestgaard, E., Baumann, M. J., Borch, K., Praestgaard, J., Monrad, R. N. & Westh, P. (2013). *Appl. Microbiol. Biotechnol.* **97**, 159–169.
Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C. & Robins, K. (2012). *Nature (London)*, **485**, 185–194.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Chen, H.-L., Chen, Y.-C., Lu, M.-Y., Chang, J.-J., Wang, H.-T., Ke, H.-M., Wang, T.-Y., Ruan, S.-K., Wang, T.-Y., Hung, K.-Y., Cho, H.-Y., Lin, W.-T., Shih, M.-C. & Li, W.-H. (2012). *Biotechnol. Biofuels*, **5**, 24.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Decker, C. H., Visser, J. & Schreier, P. (2000). *J. Agric. Food Chem.* **48**, 4929–4936.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fang, Z., Fang, W., Liu, J., Hong, Y., Peng, H., Zhang, X., Sun, B. & Xiao, Y. (2010). *J. Microbiol. Biotechnol.* **20**, 1351–1358.
Guinier, A. & Fournet, G. (1955). *Small-angle Scattering of X-rays*, pp. 17–19. New York: John Wiley & Sons.
Hammersley, A. P., Svensson, S. O., Hanfland, M., Fitch, A. N. & Hausermann, D. (1996). *High Press. Res.* **14**, 235–248.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Hong, M.-R., Kim, Y.-S., Park, C.-S., Lee, J.-K., Kim, Y.-S. & Oh, D.-K. (2009). *J. Biosci. Bioeng.* **108**, 36–40.
Jeng, W.-Y., Wang, N.-C., Lin, C.-T., Chang, W.-J., Liu, C.-I. & Wang, A. H.-J. (2012). *Acta Cryst.* D**68**, 829–838.
Jeng, W.-Y., Wang, N.-C., Lin, M.-H., Lin, C.-T., Liaw, Y.-C., Chang, W.-J., Liu, C.-I., Liang, P.-H. & Wang, A. H.-J. (2011). *J. Struct. Biol.* **173**, 46–56.
Jo, S., Kim, T., Iyer, V. G. & Im, W. (2008). *J. Comput. Chem.* **29**, 1859–1865.
Jo, S., Vargyas, M., Vasko-Szedlar, J., Roux, B. & Im, W. (2008). *Nucleic Acids Res.* **36**, W270–W275.
Karnchanatat, A., Petsom, A., Sangvanich, P., Piaphukiew, J., Whalley, A. J. S., Reynolds, C. D. & Sihanonth, P. (2007). *FEMS Microbiol. Lett.* **270**, 162–170.
Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.
Krogh, K. B. M., Harris, P. V., Olsen, C. L., Johansen, K. S., Hojer-Pedersen, J., Borjesson, J. & Olsson, L. (2010). *Appl. Microbiol. Biotechnol.* **86**, 143–154.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
Lamed, R., Kenig, R., Morgenstern, E., Calzada, J. F., Micheo, F. & Bayer, E. A. (1991). *Appl. Biochem. Biotechnol.* **27**, 173–183.
Lee, H.-L., Chang, C.-K., Jeng, W.-Y., Wang, A. H.-J. & Liang, P.-H. (2012). *Protein Eng. Des. Sel.* **25**, 733–740.
Liu, J., Zhang, X., Fang, Z., Fang, W., Peng, H. & Xiao, Y. (2011). *J. Biosci. Bioeng.* **112**, 447–450.
Lundemo, P., Adlercreutz, P. & Karlsson, E. N. (2013). *Appl. Environ. Microbiol.* **79**, 3400–3405.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (2011). *Acta Cryst.* D**67**, 355–367.
Noguchi, J., Hayashi, Y., Baba, Y., Okino, N., Kimura, M., Ito, M. & Kakuta, Y. (2008). *Biochem. Biophys. Res. Commun.* **374**, 549–552.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pei, J., Pang, Q., Zhao, L., Fan, S. & Shi, H. (2012). *Biotechnol. Biofuels*, **5**, 31.
Read, S. M. & Northcote, D. H. (1981). *Anal. Biochem.* **116**, 53–64.
Saha, B. C. & Bothast, R. J. (1996). *Appl. Environ. Microbiol.* **62**, 3165–3170.
Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C. & Pandey, A. (2013). *Bioresour. Technol.* **127**, 500–507.

Souza, F. H. M., Machado, C. B., Meleiro, L. P., Maldonado, R. F., Souza, T. A. C. B., Masui, D. C., Ward, R. J., Murakami, M. T., Jorge, J. A. & Furriel, R. P. M. (2014). Submitted.
Souza, F. H. M., Nascimento, C. V., Rosa, J. C., Masui, D. C., Leone, F. A., Jorge, J. A. & Furriel, R. P. M. (2010). *Proc. Biochem.* **45**, 272–278.
Suzuki, K., Sumitani, J.-I., Nam, Y.-W., Nishimaki, T., Tani, S., Wakagi, T., Kawaguchi, T. & Fushinobu, S. (2013). *Biochem. J.* **452**, 211–221.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R. & Madhusudhan, M. S. (2013). *Nucleic Acids Res.* **41**, W314–W321.
Tomme, P., Warren, R. A. & Gilkes, N. R. (1995). *Adv. Microb. Physiol.* **37**, 1–81.
Uchima, C. A., Tokuda, G., Watanabe, H., Kitamoto, K. & Arioka, M. (2011). *Appl. Microbiol. Biotechnol.* **89**, 1761–1771.
Uchiyama, T., Miyazaki, K. & Yaoi, K. (2013). *J. Biol. Chem.* **288**, 18325–18334.
Vagin, A. & Teplyakov, A. (2010). *J. Appl. Cryst.* **30**, 1022–1025.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Wang, Q., Trimbur, D., Graham, R., Warren, R. A. & Withers, S. G. (1995). *Biochemistry*, **34**, 14554–14562.
Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B. & Aebersold, R. (1990). *J. Am. Chem. Soc.* **112**, 5887–5889.
Yu, J., Zhou, Y., Tanaka, I. & Yao, M. (2010). *Bioinformatics*, **26**, 46–52.
Zanoelo, F. F., Polizeli, M. L. T. M., Terenzi, H. F. & Jorge, J. A. (2004). *FEMS Microbiol. Lett.* **240**, 137–143.