

Sets of Covariant Residues Modulate the Activity and Thermal Stability of GH1 β -Glucosidases

Fábio K. Tamaki¹, Larissa C. Textor², Igor Polikarpov², Sandro R. Marana^{1*}

1 Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, SP, Brazil, **2** Instituto de Física, IFSC-USP, São Carlos, SP, Brazil

Abstract

The statistical coupling analysis of 768 β -glucosidases from the GH1 family revealed 23 positions in which the amino acid frequencies are coupled. The roles of these covariant positions in terms of the properties of β -glucosidases were investigated by alanine-screening mutagenesis using the fall armyworm *Spodoptera frugiperda* β -glycosidase (Sf β gly) as a model. The effects of the mutations on the Sf β gly kinetic parameters (k_{cat}/K_m) for the hydrolysis of three different *p*-nitrophenyl β -glycosides and structural comparisons of several β -glucosidases showed that eleven covariant positions (54, 98, 143, 188, 195, 196, 203, 398, 451, 452 and 460 in Sf β gly numbering) form a layer surrounding the active site of the β -glucosidases, which modulates their catalytic activity and substrate specificity via direct contact with the active site residues. Moreover, the influence of the mutations on the transition temperature (T_m) of Sf β gly indicated that nine of the coupled positions (49, 62, 143, 188, 223, 278, 309, 452 and 460 in Sf β gly numbering) are related to thermal stability. In addition to being preferentially occupied by prolines, structural comparisons indicated that these positions are concentrated at loop segments of the β -glucosidases. Therefore, due to these common biochemical and structural properties, these nine covariant positions, even without physical contacts among them, seem to jointly modulate the thermal stability of β -glucosidases.

Citation: Tamaki FK, Textor LC, Polikarpov I, Marana SR (2014) Sets of Covariant Residues Modulate the Activity and Thermal Stability of GH1 β -Glucosidases. PLoS ONE 9(5): e96627. doi:10.1371/journal.pone.0096627

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, Russian Federation

Received January 9, 2014; **Accepted** April 9, 2014; **Published** May 7, 2014

Copyright: © 2014 Tamaki et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project was supported by Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo; Grant 08/55914-9), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and Instituto Nacional de Ciências e Tecnologia para o Bioetanol (Grants: 08/57908-1 and 574002/2008-1). F.K.T. and L.C.T. are post-docs supported by Fapesp. I.P. and S.R.M. are CNPq research fellows. I.P. is a staff member of the Departamento de Física e Informática – IFSC USP. S.R.M. is a staff member of the Departamento de Bioquímica – IQUSP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: srmarana@iq.usp.br

Introduction

In recent years, the search for new enzymes using improvements in sequencing technologies has resulted in a large collection of protein sequences. For example, the glycoside hydrolase family 1 (GH1) groups have more than 5,000 β -glucosidase sequences in the CAZY database, only 274 of which have been marked as characterized to date [1]. Structural data - 41 crystallographic structures are available - and biochemical characterization of GH1 β -glucosidases revealed these enzymes share the same fold, the $(\beta/\alpha)_8$ barrel, and that their catalytic activity depends on a pair of glutamate residues, which act as acid/base and nucleophile in a double substitution mechanism [1]. Moreover, their companions, R and Y residues, are involved in the modulation of the pK_a of the catalytic glutamates. These catalytic residues are highly conserved among β -glucosidases, except for the myrosinases [2,3]. Additionally, a network of hydrogen bonds formed by Q, H, W and E residues, which are placed and conserved in the active site of the β -glucosidases, modulates their substrate glycone specificity [4–7]. Finally, a set of variable residues, for which only the structural placement is conserved, forms the aglycone binding region for different β -glucosidases [8–10].

In parallel with the increase in sequence data, new methods to characterize the correlation between functions and structures of proteins have been developed that use special approaches to

globally analyze protein sequences, and have revealed groups of residues that are jointly involved in determining functional properties. One of these methods is statistical coupling analysis (SCA), which, through covariation analysis of large multiple sequence alignments, is capable of identifying sets of residues that are important for protein folding [11], allostery [12], enzymatic activity and thermal stability [13]. Moreover, it was recently demonstrated that sets of covariant residues, termed sectors, are important starting points for protein engineering [14].

Among the characterized GH1 enzymes, the β -glucosidase from the fall armyworm *Spodoptera frugiperda* (Sf β gly) has been extensively studied, including biochemical and site-directed mutagenesis of the active site residues involved in pH optimum modulation and substrate specificity [2,4,10,15,16]. Additionally, residues outside of the active site that affect Sf β gly enzymatic activity via indirect contacts have been identified [17]. These findings make Sf β gly an excellent model for analyzing the function of covariant residues of the GH1 family.

We applied SCA to an alignment containing 768 β -glucosidase sequences and identified 23 covariant positions. Using Sf β gly as an experimental model, alanine residues were introduced at 18 of these covariant positions, and these single mutants were characterized for their thermal stability (T_m) and kinetic parameters (k_{cat}/K_m) for the hydrolysis of three different chromogenic substrates. Based on the results, a set of 11 covariant positions, which are

related to the enzymatic activity and form a layer surrounding the active site of the β -glucosidases, was identified. In addition, a second set of 9 covariant positions related to enzyme thermal stability, consisting of amino acid residues mostly at the loop regions of this $(\beta/\alpha)_8$ barrel structure, was identified.

Materials and Methods

Identification of the covariant positions in β -glucosidases

SCA was performed as described previously [11,18] using a multiple sequence alignment containing 768 non-redundant β -glucosidases from the GH1 family, which were retrieved from the PFAM server (<http://pfam.sanger.ac.uk>). Site conservation (ΔG^{stat}) and positional statistical coupling ($\Delta\Delta G^{stat}$) parameters were calculated as previously described [19] using our own programs written in C/C++ [18]. Perturbations were considered significant when differences were present in at least 25% of the sequences in the alignment. A set of residue positions containing correlated conservations was obtained by clustering the $\Delta\Delta G^{stat}$ matrix using the Matlab (MathWorks) software package.

Site-directed mutagenesis

Site-directed mutants were constructed using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) following the manufacturer's instructions. Wild-type Sf β gly cloned into the pAE vector [20] was used as a template for PCR reactions performed with mutagenic primer pairs, which are presented on the Table S1. The pAE vector coding for wild-type Sf β gly, previously available, contains a T7 promoter, an ampicillin resistance mark and the insert was cloned in the *Nde*I and *Xho*I sites. Mutation incorporation was verified by DNA sequencing.

Expression and purification of recombinant and wild-type Sf β gly

NovaBlue (DE3) competent cells (EMD Millipore, Billerica, MA, USA) were transformed with pAE plasmids encoding wild-type or mutant Sf β gly, plated on LB-agar containing ampicillin (50 μ g/mL) and grown at 37°C for 16 h. Single colonies were grown at 20°C in LB broth containing ampicillin (50 μ g/mL) until they reached an attenuation of 0.500 at 600 nm. Next, 0.4 mM (final concentration) isopropyl β -D-1-thiogalactopyranoside (IPTG) was added for 16 h to induce recombinant protein expression, after which cells were harvested by centrifugation (4,000 $\times g$, 20 min, 4°C) and frozen at -80°C. The pelleted cells were resuspended in 100 mM sodium phosphate pH 7.4 containing 200 mM NaCl, 60 mM imidazole and 10% (v/v) glycerol and were lysed with 3 ultrasound pulses (15 s, output 10 using a Branson Sonifer 250 adapted with a microtip) with 1 min intervals in ice to avoid heating the samples. The supernatants were recovered by centrifugation (13,200 g, 20 min, 4°C), and soluble recombinant proteins were incubated with 200 μ L Ni-NTA Agarose (4°C, 1 h) (Qiagen, Hilden, Germany). The resin was pelleted and washed 5 times with 100 mM sodium citrate-sodium phosphate pH 6.0 containing 200 mM NaCl and 60 mM imidazole. Protein elution was performed using the same buffer containing 500 mM imidazole. Protein purity was verified by SDS-PAGE [21], and purified proteins were desalted using a minitrap G-25 (GE Healthcare, Uppsala, Sweden) and stored at 4°C.

Protein concentration was determined by measuring the absorbance at 280 nm in 20 mM sodium phosphate pH 6.0 containing 6 M guanidinium hydrochloride. Extinction coefficients ($\epsilon_{280\text{ nm}}$) were calculated based on the primary sequences of the wild-type and mutant Sf β gly proteins using the ProtParam

server at ExPaSy (<http://web.expasy.org/protparam/>). The extinction coefficients ranged from 2.149 to 2.053 (Table S2).

Kinetic and thermal characterization of the mutant and wild-type Sf β gly

Enzyme kinetic parameters (k_{cat} , K_m and k_{cat}/K_m) were determined for purified enzymes by measuring the initial rates (v_0) of hydrolysis of at least 10 different concentrations of substrates, including p -nitrophenyl- β -D-glucopyranoside, p -nitrophenyl- β -D-fucopyranoside and p -nitrophenyl- β -D-galactopyranoside (Sigma, St. Louis, MO, USA) prepared in 100 mM sodium citrate – sodium phosphate buffer pH 6.0. Experiments were performed at 30°C. The hydrolysis of these substrates was detected following the formation of p -nitrophenolate by absorbance at 420 nm after addition of 250 mM sodium carbonate – sodium bicarbonate buffer pH 11 to the reaction samples. Kinetic parameters K_m and k_{cat} were determined by fitting the v_0 and [S] data to the Michaelis-Menten equation using the Enzfitter software (Elsevier-Biosoft, Cambridge, UK).

Differential scanning fluorimetry (DSF) experiments of both wild-type and mutant Sf β gly were performed using SYPRO® Orange solution (500-fold dilution) (Sigma, St. Louis, MO, USA). Melting studies were performed in optical tubes using a 7500 Real-Time System (Applied Biosystems, Foster City, CA, USA). The temperature gradient ranged from 25°C to 95°C with a slope of 0.5% per step. The melting data were fitted according to recent literature [22], resulting in a theoretical T_m . Fitting processes were performed using EnzFitter software.

Tridimensional modeling and computational structure comparison

The tridimensional structure of Sf β gly was homology-modeled based on the crystallographic structure of the β -glucosidase from *Neotermes koshunensis* (PDB, 3VIK) using the Phyre2 software [23]. The sequences had 49% identity and 66% similarity. The structural models were visualized using the PyMOL molecular graphics system v1.1 (Schrödinger, LLC). The distances between amino acid pairs were calculated using DeepView/SwissPDB-Viewer v3.7 software [24]. The topology schemes of different β -glucosidases were manually drawn using information regarding the secondary structures visualized in PyMOL.

Results and Discussion

The SCA [19] using 768 sequences of GH1 β -glucosidases revealed 23 covariant positions in their primary sequences (49, 54, 57, 62, 98, 112, 143, 176, 188, 195, 196, 203, 223, 278, 309, 329, 398, 445, 449, 451, 452, 456 and 460 in Sf β gly numbering). Except for position 451, involved in the β -glucosidase's specificity for the substrate glycone [4–7], none had been experimentally correlated with any biochemical or biophysical properties of the β -glucosidases.

Because SCA is based on the concept that covariant positions of a primary protein structure should be occupied by amino acid residues that are jointly involved in determining the same functional property [13], an experimental approach was designed to search the identified set of covariant positions for subgroups linked to the enzymatic activities or thermal stabilities of the GH1 β -glucosidases. Therefore, Sf β gly (GenBank code: AF 052729), a digestive enzyme from the fall armyworm *Spodoptera frugiperda* that has been extensively studied [4,10,15], was chosen as a representative β -glucosidase, and residues at its 23 covariant positions were separately replaced with alanine using a site-directed mutagenesis technique. Eighteen of these mutant enzymes were successfully

Table 1. Enzyme kinetic parameters for the hydrolysis of *p*-nitrophenyl β -glycosides catalyzed by the wild-type and mutant Sf β gly proteins.

Enzyme	NP β fuc			NP β glu			NP β gal		
	K_m (mM)	k_{cat} (min $^{-1}$)	k_{cat}/K_m (min $^{-1} \cdot \text{mM}^{-1}$)	K_m (mM)	k_{cat} (min $^{-1}$)	k_{cat}/K_m (min $^{-1} \cdot \text{mM}^{-1}$)	K_m (mM)	k_{cat} (min $^{-1}$)	k_{cat}/K_m (min $^{-1} \cdot \text{mM}^{-1}$)
wild-type	0.37 \pm 0.02	0.408 \pm 0.006	1.10 \pm 0.01	4.1 \pm 0.2	0.70 \pm 0.01	0.17 \pm 0.01	4.2 \pm 0.2	0.025 \pm 0.003	0.0060 \pm 0.008
K49A	0.81 \pm 0.07	0.65 \pm 0.01	0.80 \pm 0.06	3.1 \pm 0.2	0.283 \pm 0.008	0.092 \pm 0.008	7.7 \pm 1	0.031 \pm 0.002	0.0040 \pm 0.006
W51A	1.97 \pm 0.08	0.0530 \pm 0.0008	0.027 \pm 0.001	7.3 \pm 0.2	0.0279 \pm 0.0004	0.0038 \pm 0.0001	2.2 \pm 0.1	0.00073 \pm 0.00001	0.00033 \pm 0.00002
M57A	1.08 \pm 0.06	1.05 \pm 0.01	0.91 \pm 0.05	2.8 \pm 0.1	0.37 \pm 0.04	0.13 \pm 0.01	3.4 \pm 0.2	0.0266 \pm 0.007	0.0079 \pm 0.006
P62A	0.26 \pm 0.03	0.054 \pm 0.001	0.209 \pm 0.008	2.4 \pm 0.2	0.039 \pm 0.001	0.016 \pm 0.001	0.9 \pm 0.1	0.00195 \pm 0.00007	0.0021 \pm 0.0002
F98A	-	-	-	-	-	-	-	-	-
N112A	4.1 \pm 0.5	1.82 \pm 0.07	0.44 \pm 0.06	1.16 \pm 0.08	0.305 \pm 0.003	0.263 \pm 0.018	2.8 \pm 0.2	0.035 \pm 0.001	0.0125 \pm 0.0013
W143A	35 \pm 4	0.020 \pm 0.002	0.00057 \pm 0.0009	20 \pm 2	0.00124 \pm 0.00008	0.00062 \pm 0.00007	12 \pm 1	0.00026 \pm 0.00002	0.000022 \pm 0.00003
P188A	0.17 \pm 0.01	0.000239 \pm 0.000003	0.0014 \pm 0.0001	-	-	0.0000364 \pm 0.000004	3.3 \pm 0.3	0.000237 \pm 0.000008	0.000073 \pm 0.00007
G195L	1.7 \pm 0.2	0.070 \pm 0.002	0.041 \pm 0.005	2.1 \pm 0.1	0.0404 \pm 0.0007	0.019 \pm 0.001	12 \pm 1	0.0113 \pm 0.0004	0.000094 \pm 0.00008
Y196A	2.7 \pm 0.2	0.205 \pm 0.006	0.076 \pm 0.006	10.2 \pm 0.6	0.115 \pm 0.003	0.0113 \pm 0.0007	13 \pm 1	0.0103 \pm 0.0004	0.00079 \pm 0.0006
P203A	0.87 \pm 0.08	0.215 \pm 0.007	0.25 \pm 0.02	50 \pm 5	0.52 \pm 0.04	0.010 \pm 0.001	-	-	0.00287 \pm 0.00001
H223A	0.90 \pm 0.06	0.23 \pm 0.01	0.26 \pm 0.02	5.7 \pm 0.5	0.025 \pm 0.0096	0.0044 \pm 0.0004	6.5 \pm 1	0.021 \pm 0.002	0.00322 \pm 0.0007
P278A	1.3 \pm 0.1	0.034 \pm 0.001	0.026 \pm 0.003	2.7 \pm 0.1	0.00250 \pm 0.00005	0.00093 \pm 0.00006	0.4 \pm 0.1	0.00058 \pm 0.00003	0.0145 \pm 0.0005
P309A	0.580 \pm 0.004	0.015 \pm 0.002	0.026 \pm 0.001	1.9 \pm 0.1	0.0499 \pm 0.0009	0.026 \pm 0.002	4.4 \pm 0.2	0.0057 \pm 0.0001	0.00130 \pm 0.00008
T398A	-	-	-	-	-	-	-	-	-
S445A	1.30 \pm 0.09	0.39 \pm 0.01	0.30 \pm 0.02	1.58 \pm 0.07	0.317 \pm 0.003	0.200 \pm 0.009	5.0 \pm 0.3	0.46 \pm 0.01	0.002 \pm 0.007
W452A	-	-	-	-	-	-	-	-	-
F460A	9.1 \pm 0.5	0.046 \pm 0.001	0.0051 \pm 0.0003	58 \pm 11	0.013 \pm 0.001	0.00022 \pm 0.00005	11.6 \pm 0.7	0.00268 \pm 0.00008	0.00023 \pm 0.00001

- No activity.

Table 2. Transition temperatures (T_m) for thermal denaturation of the wild-type and mutant Sf β gly proteins.

Enzyme	T_m (K)
wt	319
K49A	331.6
W54A	318.8
M57A	317.9
P62A	313.5
F98A	316.9
N112A	319.3
W143A	311.8
P188A	322.2
G195A	319
Y196A	319
P203A	319.8
H223A	325
P278A	324.7
P309A	311.9
S445A	317.1
W452A	316.2
F460A	306.8

Standard deviations were lower than 0.5 K.
doi:10.1371/journal.pone.0096627.t002

expressed as recombinant proteins in NovaBlue (DE3) bacteria and purified (Figure S1). The E451A mutant of Sf β gly has been previously studied [4]. Mutant enzymes with replacements at positions 176, 329, 449 and 456 were not studied due to poor solubility.

The enzyme kinetic parameters for the hydrolysis of three different substrates catalyzed by wild-type and the 18 mutant Sf β gly proteins were determined (Table 1 and Figure S2). Additionally, the transition temperatures (T_m) for the denaturation of the mutant and wild-type Sf β gly proteins were evaluated in Differential Scanning Fluorimetry (DSF) experiments (Table 2 and Figure S3).

Mutations resulting in at least a 4-fold change in the k_{cat}/K_m for the hydrolysis of at least two different substrates were considered replacements of residues relevant for the enzymatic activity because such variation corresponds to a $\Delta\Delta G^\ddagger$ higher than 3 kJ/mol, which is equivalent to the disruption or formation of one hydrogen bond [25]. Based on these criteria, only 4 mutations had no significant effects on the enzymatic activity, namely K49A, M57A, N112A and S445A (Table 3).

Similarly, the mutational effect on the thermal stability of Sf β gly was considered significant for T_m changes higher than 2.5 K, given that in the DSF experiments, this variation range corresponds to four-fold changes in the ratio of native to denatured wild-type Sf β gly. Based on that threshold, the mutations K49A, P62A, W143A, P188A, H223A, P278A, P309A, W452A and F460A were identified as replacements at positions involved in the thermal stability of Sf β gly (Table 3).

Table 3. Mutational effects on the catalytic activity (k_{cat}/K_m ratio) and thermal stability (ΔT_m) of Sf β gly proteins.

Mutation position	k_{cat}/K_m ratio (mutant/wt)			ΔT_m (K)
	NP β glc	NP β gal	NP β fuc	
49	0.5	0.06	0.7	-7.2
54	0.02	0.005	0.02	-0.2
57	0.77	0.12	0.8	-1.0
62	0.09	0.03	0.18	-5.5
98	-	-	-	-2.1
112	1.5	0.19	0.38	-0.3
143	0.0004	0.0003	0.0005	-7.2
188	0.0002	0.0011	0.0012	3.2
195	0.11	0.01	0.03	0.1
196	0.06	0.01	0.06	0.0
203	0.06	0.04	0.21	0.8
223	0.30	0.05	0.075	6.0
278	0.005	0.02	0.02	5.8
309	0.15	0.02	0.02	-7.1
398	-	-	-	nd
445	1.1	1.5	0.27	-1.9
452	-	-	-	-2.8
460	0.0012	0.0037	0.0044	-12.2

Only mutational effects higher than a 4-fold change in the k_{cat}/K_m ratio ($0.25 > k_{cat}/K_m$ ratio > 4) were considered significant for the enzymatic activity. Mutational effects on the thermal stability were considered relevant only for $\Delta T_m > 2.5$ K. -, no activity; nd, not determined; wt, wild-type.
doi:10.1371/journal.pone.0096627.t003

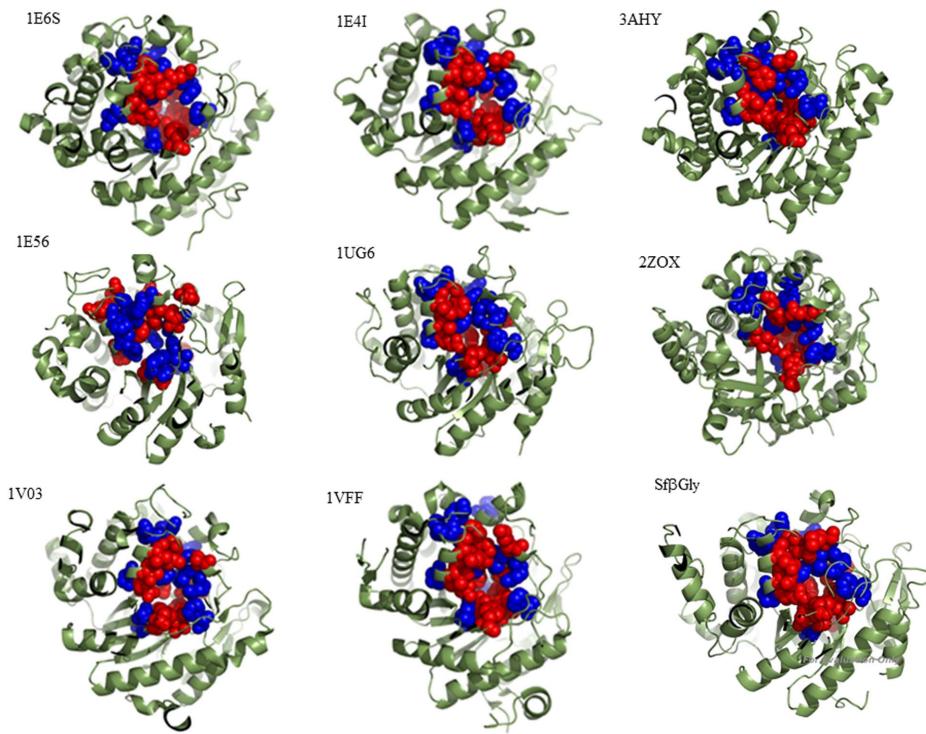


Figure 1. Structural comparison of β -glucosidases showing the active site residues (red) and sector A positions (blue). Myrosinase from *Sinapis alba* (1E6S); β -glucosidase A from *Paenibacillus polymyxa* (1E4I); β -glucosidase from *Trichoderma reesei* (3AHY); β -glucosidase Zmglu from *Zea mays* (1E56); β -glucosidase from *Thermus thermophilus* (1UG6); Human cytosolic β -glucosidase (2ZOX); SbDhr from *Sorghum bicolor* (1V03); β -glucosidase from *Pyrococcus horikoshii* (1VFF); β -glucosidase from *Spodoptera frugiperda* Sf β Gly. The distances between sector A and the active site residues are shorter than 4.5 Å. The structures were visualized using PyMOL software.

doi:10.1371/journal.pone.0096627.g001

SC positions related to the catalytic activity in β -glucosidases

Although no structural information was used in the SCA, visual inspection of the Sf β Gly structural model revealed that residues at 11 positions related to the enzymatic activity are directly connected to active site residues through covalent or non-covalent bonds. Therefore, those 11 residues form a single group indirectly connected to each other through the active site residues and their non-covalent interactions with the substrate. This group was labeled sector A (Figure 1; Table 4). This observation is similar to previous SCAs of different protein families, which also showed groups of residues forming chains of interactions that connect distant points of their structures [19,26]. In the particular case of serine proteases, two sets of covariant positions were identified. The first set contained residues involved in the catalytic mechanism (including the catalytic triad), and the second set contained residues present in the S1 pocket and was involved in substrate specificity [13].

Structural comparison showed that as in Sf β Gly, the placement of sector A positions in close contact to the active site is also observed for different β -glucosidases from the GH1 family (Figure 1; Table 4). In brief, residues at sector A positions contact two glutamate residues (essential for catalysis), a pair of residues (tyrosine and arginine) involved in the modulation of the pK_a of the catalytic glutamate, and a group of residues that bind the substrates glycone and aglycone (Table 4). Thus, sector A residues might modulate several β -glucosidase properties, from catalytic activity to substrate specificity by affecting the positioning and properties of their active site residues. Moreover, data presented here suggest that the active site of the β -glucosidases is formed by a

“layer” of highly conserved residues that interacts directly with the substrate and promotes its hydrolysis and is surrounded by a second “layer” formed by residues of the sector A positions (Table 4; Figure 1).

Evidence that joint variation of residues at sector A positions is involved in substrate binding and specificity is found when comparing β -glucosidases and 6-phospho- β -glucosidases, both groups belonging to the GH1 family. Indeed, 99% of the sequences exhibiting Y at position 460 also present S at position 451, and among these sequences, several were previously characterized as 6-phospho- β -glucosidases. Similarly, 99% of the sequences presenting E at position 451 present F at position 460, and several have β -glucosidase activity. Thus, the identities of residues at positions 451 and 460 are linked and directly connected to enzyme specificity. In 6-phospho- β -glucosidases, the replacement of E by S creates space in their active sites for the binding of substrates containing a phosphate group linked to the 6-OH of the glycone and also favors the formation of a hydrogen bond with this substrate. An additional hydrogen bond with the 6-phosphate group is formed by Y460 [31,32] (Figure S4). Therefore, changing the substrate specificity of these two groups of β -glucosidases depends on replacements at both positions 451 and 460. A similar result was observed for the joint mutations C42A, C58A/V and S195T at trypsin covariant residues, which converted this enzyme from a serine to threonine protease [13].

Finally, five residues, P62, G195, Y196, P278 and P309, (Table 3) were not included in sector A because no direct or even indirect interactions connecting them to the active site were identified. Thus, the replacements of those residues affected the Sf β Gly activity through currently unknown mechanisms.

Table 4. Residues from the active sites of β -glucosidases in direct contact with sector A positions.

Sector A position	Active site residues	1U66	1E41	1E56	1V03	1E65	2ZOX	3AHY
54	K201 ^{abr} , M453 ^{abr}	H178 ^{ab} , A394 ^{ab}	H179, A407	F205 ^{ab} , F466 ^{ab}	L203 ^a ^{br} , S462 ^{ab}	D201, N466	F179 ^a ^{br} , N426	F179, A426
98	R97 ^p	R75	R77	R96	R97	R95	R75	R73
143	H142 ^{bg} , E187 [*] , E190 ^{ab} , K201 ^{ab}	H119 ^{bg} , E164 [*] , C167, H178 ^{ab}	H121 ^{bg} , E166 [*] , C168, H179	H142 ^{bg} , D191 [*] , T194, F205 ^{ab}	H143 ^{bg} , D189 [*] , T192, L203 ^{ab}	H141 ^{bg} , Q187, S190 ^{ab} , D201	H120 ^{bg} , Q165 [*] , F179 ^{ab}	H119 ^{bg} , E165 [*] , C168, F179
188	E187 [*] , E190 ^{ab}	E164 [*] , C167	C167, L171 ^{ab} , H178	D191 [*] , T194	D189 [*] , T192	Q187, S190 ^{ab}	Q165 [*] , V168 ^{ab}	E165 [*] , C168
195	E194 ^{ab}	C168, L172, H179	C168, L172, H179	T194, F198 ^{ab} , F205 ^{ab}	T192, V196 ^{ab} , L203 ^{ab}	S190 ^{ab} , R194 ^{ab} , D201	M172 ^{ab} , F179 ^{ab}	C168, P172, F179
196	E194 ^{ab}	L171 ^{ab}	L172	F198 ^{ab}	V196 ^{ab}	R194 ^{ab}	M172 ^{ab}	P172
203	E194 ^{ab} , K201 ^{ab}	H178 ^{ab}	H179	F205 ^{ab}	L203 ^a ^{br}	D201	F179 ^{ab}	F179
398	R97 ^p , E399 [*]	R75, E338 [*]	R77, E352 [*]	R96, E406 [*]	R97, E404 [*]	R95, E409 [*]	E373 [*]	R73, E367 [*]
451	Q39 ^{bg} , W444 ^{bg} , M453 ^{ab}	Q18 ^{bg} , W385 ^{bg} , A394 ^{ab}	Q20 ^{bg} , W398 ^{bg} , A407	Q38 ^{bg} , W457 ^{bg} , F466 ^{ab}	Q39 ^{bg} , W453, S462 ^{ab}	Q39 ^{bg} , W457 ^{bg} , N466	Q17, W417 ^{bg} , N426	Q16 ^{bg} , W417 ^{bg} , N426
452	Q39 ^{bg} , H142 ^{bg} , K201 ^{ab} , M453 ^{ab}	Q18 ^{bg} , H119 ^{bg} , H178 ^{ab} , A394 ^{ab}	Q20 ^{bg} , H121 ^{bg} , H180, A407	Q38 ^{bg} , H142 ^{bg} , F205 ^{ab} , F466 ^{ab}	Q39 ^{bg} , H143 ^{bg} , L203 ^{ab} , S462 ^{ab}	Q39 ^{bg} , H141 ^{bg} , N466	Q17, H120 ^{bg} , F179 ^{ab} , N426	Q17, H120 ^{bg} , H119 ^{bg} , F179, N426
460	Y331 ^p , W444 ^{bg}	Y284, W385 ^{bg}	Y296, W398 ^{bg}	Y333, W437 ^{bg}	Y331, W453 ^{bg}	W457 ^{bg}	Y309, W417 ^{bg}	Y298, W417 ^{bg}

The numbering of the sector A positions was based on the 5f8g sequence. * identifies catalytic glutamic acids; ^p – indicates residues involved in the modulation of the pKa of the catalytic glutamic acids; ^{bg} – shows residues involved in the binding of the substrate glycone; ^{ab} – indicates residues that form the aglycone binding region. Data regarding the role of individual residues in substrate binding and catalysis were retrieved from the literature [5,27,28,29,9,30]. β -Glucosidase from *Spodoptera frugiperda* 5f8g; β -glucosidase A from *Thermus thermophilus* (1U66); β -glucosidase Zmgli from *Zea mays* (1E56); β -glucosidase SbDhr from *Sorghum bicolor* (1V03); myrosinase from *Sinapis alba* (1E65); Human cytosolic β -glucosidase (2ZOX); β -glucosidase from *Trichoderma reesei* (3AHY).

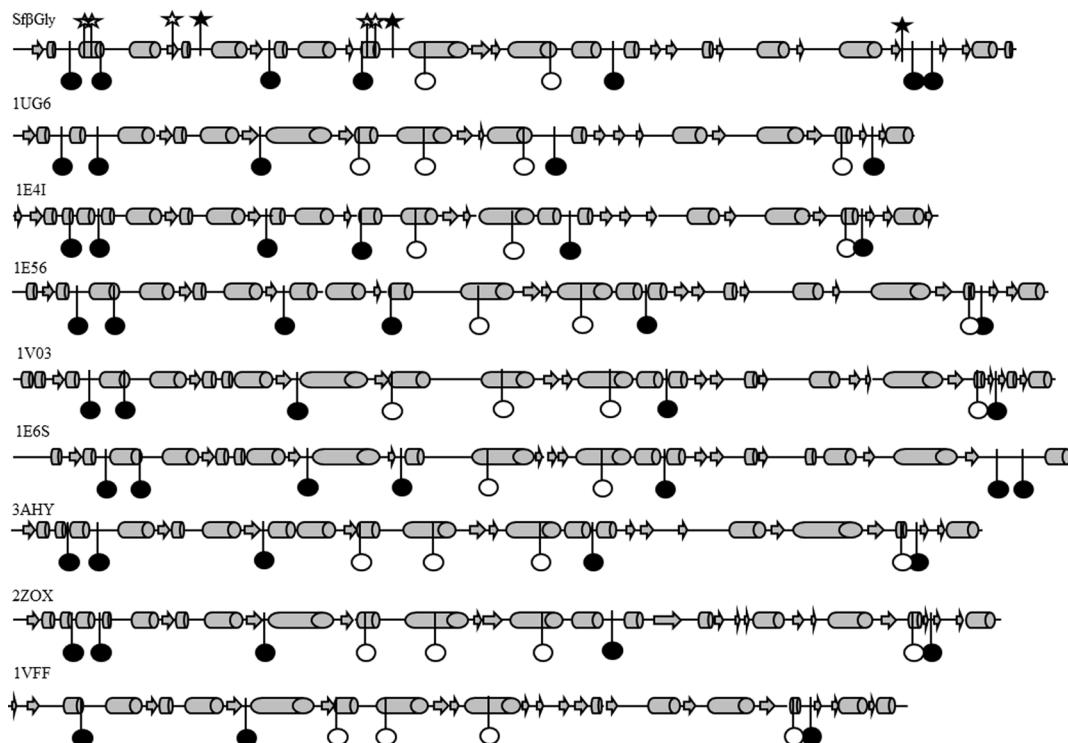


Figure 2. Distribution of sector S positions on the secondary structure of β -glucosidases. β -glucosidase from *Spodoptera frugiperda* Sf β Gly; β -glucosidase from *Thermus thermophilus* (1UG6); β -glucosidase A from *Paenibacillus polymyxa* (1E4I); β -glucosidase Zmglu from *Zea mays* (1E56); β -glucosidase SbDhr from *Sorghum bicolor* (1V03); myrosinase from *Sinapis alba* (1E6S); β -glucosidase from *Trichoderma reesei* (3AHY); Human cytosolic β -glucosidase (2ZOX); β -glucosidase from *Pyrococcus horikoshii* (1VFF). α -Helices are represented by cylinders, β -strands by arrows and loops by lines. Sector S positions are shown as circles, whereas non-sector S positions are shown as stars. The symbols (circles or stars) in black indicate positions placed at loops, whereas white symbols mark positions at helices or strands.

doi:10.1371/journal.pone.0096627.g002

SC positions related to the thermal stability of β -glucosidases

The coupled positions in which replacements significantly affect Sf β Gly T_m (Table 3) are mostly occupied by prolines (4 out 9 positions) or positively charged residues (K49 and H223). Conversely, only one proline residue was identified among the eight coupled positions not related to Sf β Gly thermal stability, and no charged residue is present among them. In addition, sequence comparison shows that positions 62, 188, 278 and 309 (Sf β Gly numbering) are predominantly occupied by proline residues among family 1 β -glucosidase sequences (frequencies higher than 50%), whereas positively charged residues are dominant at positions 49 and 223 (90% and 63%, respectively). Moreover, structural comparison shows that positions related to thermal stability are concentrated at the loop segments of the β -glucosidases (8 out 9 for Sf β Gly), whereas positions not related to thermal stability are mainly located in their α -helices and β -strands (5 out 8 for Sf β Gly). Naturally, this distribution mainly results from the properties of the proline residues, which favor loop segments [33]. Indeed, these data are in agreement with previous observations demonstrating that proline and charged residues and loop length and mobility are related to the thermal stability of proteins [34,35]. Thus, based on their similarities, coupled positions related to the thermal stability of β -glucosidases were labeled sector S (Figure 2).

Therefore, in contrast to sectors observed in the serine protease [13] and sector A of β -glucosidases (discussed above), which are composed of residues contacting each other, sector S of the β -glucosidases is formed by positions with no physical interactions

but is characterized by the prevalence of proline residues. Because of their placement in loops, sector S positions are jointly related to the thermal stability of β -glucosidases.

Supporting Information

Figure S1 SDS-PAGE of the purified wild-type and mutant Sf β Gly proteins. Mutant enzymes are identified by the residue and number of the covariant position followed by A, representing the replacement that they contain. Wt stands for wild type Sf β Gly. Details regarding the protein purification procedures are found in the Materials and Methods section. The gels (12% polyacrylamide) were stained using Coomassie blue R-250. Approximately 2 μ g of protein was loaded in each lane. MW, molecular weight markers.

(TIF)

Figure S2 The effect of substrate concentration on the activity of the wild-type and mutant Sf β Gly proteins. Mutant enzymes are identified by the residue and number of the covariant position in which A was introduced. The substrates are p -nitrophenyl β -glucoside (NP β Glc), p -nitrophenyl β -galactoside (NP β Gal) and p -nitrophenyl β -fucoside (NP β Fuc). Substrates were prepared in 100 mM sodium citrate – sodium phosphate buffer pH 6.0. Experiments were performed at 30°C. Dots are the experimental data. The continuous line represents the values calculated based on the best fitting of the experimental data into the Michaelis-Menten equation. Fitting was performed using EnzFitter software.

(TIF)

Figure S3 Differential scanning fluorimetry of the wild-type and mutant Sf β gly proteins. Samples (10 ng) were incubated with SYPRO stain, and the fluorescence was recorded at different temperatures. Mutant enzymes are identified by the residue and number of the covariant position in which A was introduced. Dots are the experimental data, whereas lines represent the calculated values produced based on the best fitting. Fitting process was performed using the software Enzfitter and the equation deduced on [22]. (TIF)

Figure S4 Structural comparison between the active site of β -glucosidases (A) and 6-phospho β -glucosidases (B) from family GH1. A glucose unit (gray and red; A) is represented in the glycone subsite of the β -glucosidases Sf β gly (green; Phyre2 model), *Neotermes koshunensis* (blue; PDB 3VIF) and wheat *Triticum aestivum* (yellow; PDB 3AIQ), whereas a 6-phospho glucose unit (gray and red, B) is presented for the 6-phospho β -glucosidases from *Streptococcus pneumoniae* (cyan, PDB 4IPN) and *Lactococcus lactis* (formerly known as *Streptococcus lactis*) (magenta, PDB 4PBG). Residues numbering was based on Sf β gly. Dotted lines

References

1. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl Acids Res* 42: D490–495.
2. Marana SR, Mendonça LMF, Andrade EHP, Terra WR, Ferreira C (2003) The role of residues R97 and Y331 in modulating the pH optimum in an insect beta-glucosidase of family 1. *Eur J Biochem* 270: 4866–4875.
3. Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In Gilbert HJ, Davies G, Henrissat B, Svensson B, (editors). *Recent Advances in Carbohydrate Bioengineering*. Cambridge: The Royal Society of Chemistry.
4. Marana SR, Terra WR, Ferreira C (2002) The role of amino acid residues Q39 and E451 in the determination of the specificity of an insect β -glycosidase. *Eur J Biochem* 269: 3705–3714.
5. Marana SR (2006) Molecular Basis of Substrate Specificity in Family 1 Glycoside Hydrolases. *IUBMB Life* 58 (2): 67–73.
6. Cairns JRK, Esen A (2010) β -glucosidases. *Cell Mol Life Sci* 67: 3389–3405.
7. Badieyan S, Bevan DR, Zhang C (2012) Probing the active site chemistry of β -glucosidases along the hydrolysis reaction pathway. *Biochemistry* 51: 8907–8918.
8. Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B, et al. (2000) The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. *Proc Natl Acad Sci USA* 97: 13555–13560.
9. Verdoucq L, Morinère J, Bevan DR, Esen A, Vasella A, et al. (2004) Structural determinants of substrate specificity in family 1 beta-glucosidases: novel insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with strict specificity, in complex with its natural substrate. *J Biol Chem* 279: 31796–31803.
10. Mendonça LMF, Marana SR (2008) The role in the substrate specificity and catalysis of residues forming the substrate aglycone-binding site of a β -glucosidase. *Biochim Biophys Acta* 1814: 1616–1623.
11. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, et al. (2005) Evolutionary information for specifying a protein fold. *Nature* 437: 512–518.
12. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, et al. (2008) Surface sites for engineering allosteric control in proteins. *Science* 322: 438–442.
13. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138: 774–786.
14. McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosai WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491: 138–144.
15. Marana SR, Jacobs-Lorena M, Terra WR, Ferreira C (2001) Amino acid residues involved in substrate binding and catalysis in an insect digestive β -glucosidase. *Biochim Biophys Acta* 1545: 41–52.
16. Marana SR, Andrade EHP, Ferreira C (2004) Investigation of the substrate specificity of a beta-glucosidase from *Spodoptera frugiperda* using site-directed mutagenesis and bioenergetics analysis. *Eur J Bioch* 271: 4169–4177.
17. Mendonça LMF, Marana SR (2011) Single mutations outside the active site affect the substrate specificity in a β -glucosidase. *Biochim Biophys Acta Proteins and Proteomics* 1814: 1616–1623.
18. Watanabe L, de Moura PR, Bleicher L, Nascimento AS, Zamorano LS, et al. (2010) Crystal structure and statistical coupling analysis of highly glycosylated peroxidase from royal palm tree (*Roystonea regia*). *J Struct Biol* 169: 226–242.
19. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286: 295–299.
20. Ramos CRR, Abreu PAE, Nascimento ALTO, Ho PL (2004) A high-copy T7 *Escherichia coli* expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Braz J Med Biol Res* 37: 1103–1109.
21. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.
22. Layton CJ, Hellinga HW (2010) Thermodynamics Analysis of Ligand-Induced Changes in Protein Thermal Unfolding Applied to High-Throughput Determination of Ligand Affinities with Extrinsic Fluorescent Dyes. *Biochemistry* 49: 10831–10841.
23. Kelley LA, Sternberg MJE (2009) A case study using the Phyre server. *Nat Protoc* 3: 363–371.
24. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723.
25. Fersht AR, Leatherbarrow RJ, Wells TNC (1986) Binding energy and catalysis: a lesson from protein engineering of the tyrosyl-tRNA synthetase. *TIBS* 11: 321–325.
26. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nature Struct Biol* 10: 59–69.
27. Sanz-Aparicio J, Hernoso JA, Martinez-Ripoll M, Lequerica JL, Polaina J (1998) Crystal structure of β -glucosidase A from *Bacillus polymyxa*: Insights into the catalytic activity in family 1 glycosyl hydrolases. *J Mol Biol* 275: 491–502.
28. Burmeister WP, Cottaz S, Driguez H, Iori R, Palmieri S, et al. (1997) The crystal structure of *Sinapis Alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5: 663–675.
29. Berrin J-G, Czjzek M, Kroon PA, McLauchlan WR, Puigserver A, et al. (2003) Substrate (aglycone) specificity of human cytosolic β -glucosidase. *Biochem J* 373: 41–48.
30. Jeng W-Y, Wang N-C, Lin M-H, Lin C-T, Liaw Y-C, et al. (2011) Structural and functional analysis of three β -glucosidases from bacterium Clostridium cellulovorans, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *J Struct Biol* 173: 46–56.
31. Wiesmann C, Hengstenberg W, Schulz GE (1997) Crystal structures and mechanism of 6-phospho-beta-galactosidase from *Lactococcus lactis*. *J Mol Biol* 269: 851–860.
32. Yu WL, Jiang YL, Pikis A, Cheng W, Bai XH, et al. (2013) Structural insights into the substrate specificity of a 6-phospho- β -glucosidase Bgl-A from *Streptococcus pneumoniae* TIGR4. *J Biol Chem* 288: 14949–14958.
33. Creighton TE (1992) *Proteins: Structures and Molecular Properties*. 2nd Edition. W H Freeman & Co.
34. Kumar S, Nussinov R (2001) How do thermophilic proteins deal with heat? *Cell Mol Life Sci* 58: 1216–1233.
35. Razvi A, Scholtz JM (2006) Lessons in stability from thermophilic proteins. *Protein Science* 15: 1569–1578.