Insights into the structure and function of fungal -mannosidases from glycoside hydrolase family 2 based on multiple crystal structures of the Trichoderma harzianum enzyme

http://www.producao.usp.br/handle/BDPI/50708

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo
Insights into the structure and function of fungal β-mannosidases from glycoside hydrolase family 2 based on multiple crystal structures of the Trichoderma harzianum enzyme

Alessandro S. Nascimento1†, Joao Renato C. Muniz1†, Ricardo Aparício2†, Alexander M. Golubev3 and Igor Polikarpov1

1 Instituto de Física de São Carlos, Universidade de São Paulo, Brazil
2 Instituto de Química, Universidade Estadual de Campinas, Brazil
3 Petersburg Nuclear Physics Institute, Molecular and Radiation Biophysics Division, Gatchina, Leningrad District, Russia

Keywords
crystal structure; galactomannan binding; glycosyl hydrolase family 2; Trichoderma harzianum; β-mannosidase

Correspondence
I. Polikarpov, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador Sãocarlense, 400, Centro, São Carlos 13566-590, SP, Brazil
Fax: +55(16)33739881
Tel: +55(16)33738088
E-mail: ipolikarpov@ifsc.usp.br
A. M. Golubev, Petersburg Nuclear Physics Institute, Molecular and Radiation Biophysics Division, Orlova Roscha, Gatchina 188300, Leningrad District, Russia
Fax: +7(813-71) 36025
Tel: +7(813-71) 46025
E-mail: golubev@pnpi.spb.ru

†These authors contributed equally to this work.

Hemicellulose is an important part of the plant cell wall biomass, and is relevant to cellulosic ethanol technologies. β-Mannosidases are enzymes capable of cleaving nonreducing residues of β-D-mannose from β-D-mannosides and hemicellulose mannose-containing polysaccharides, such as mannans and galactomannans. β-Mannosidases are distributed between glycoside hydrolase (GH) families 1, 2, and 5, and only a handful of the enzymes have been structurally characterized to date. The only published X-ray structure of a GH family 2 mannosidase is that of the bacterial Bacteroides thetaiotaomicron enzyme. No structures of eukaryotic mannosidases of this family are currently available. To fill this gap, we set out to solve the structure of Trichoderma harzianum GH family 2 β-mannosidase and to refine it to 1.9-Å resolution. Structural comparisons of the T. harzianum GH2 β-mannosidase highlight similarities in its structural architecture with other members of GH family 2, reveal the molecular mechanism of β-mannoside binding and recognition, and shed light on its putative galactomannan-binding site.

Database
Coordinates and observed structure factor amplitudes have been deposited with the Protein Data Bank (4CVU and 4UOJ). The T. harzianum β-mannosidase 2A nucleotide sequence has GenBank accession number BankIt1712036 GeneMark.hmm KJ624918.

doi:10.1111/febs.12894

Abbreviations
BtMan2A, Bacteroides thetaiotaomicron glycosyl hydrolase family 2 β-mannosidase; CBM, carbohydrate-binding module; CmMan5A, Cellvibrio mixtus glycosyl hydrolase family 5 β-mannosidase; GH, glycosyl hydrolase; GM, β-galactomannan; Man3, mannotriose; Man4, mannotetraose; PDB, Protein Data Bank; ThMan2A, Trichoderma harzianum glycosyl hydrolase family 2 β-mannosidase.
**Introduction**

The biotechnological approach to increase the production of biofuels continues to be an important scientific and technological challenge. Since the last decade, the concept of transforming plant lignocellulose – the most abundant biological material on earth – into renewable biofuels has become increasingly attractive, leading to the development of several lines of investigation on how to efficiently break down lignocellulose into simple sugars and other molecules [1]. Sugarcane bagasse, one of the most promising plant biomasses for second-generation biofuel production, is composed of cellulose (54–55%), hemicellulose (24–25%), and lignin (27%) [2]. The second most abundant polysaccharide component of plant biomass, hemicellulose, is mainly composed of xyloglucans, xylans, galactomannans, glucomannans, and β-(1→3), 1→4)-glucans [3]. In order to complete the enzymatic degradation of galactomannans, at least three different types of glycoside hydrolase (GH) are required: β-mannanases, α-galactosidases, and β-mannosidases. β-Mannanases and β-mannosidases act synergistically: the former cleaves endomannosidic linkages randomly, and the latter hydrolyzes nonreducing β-D-mannose [4,5].

The β-mannosidases are distributed among three GH families, according to the Carbohydrate-Active Enzymes Database classification [6], and, so far, have been poorly characterized from a structural point of view. One plant (Oryza sativa) β-mannosidase in GH family 1 has been structurally characterized, an X-ray structure of one bacterial β-mannosidase in GH family 2 has been determined, and one GH family 5 proteobacterial β-mannosidase has had its structure solved.

*Trichoderma harzianum* is an ascomycete filamentous fungus that shows both cellulolytic and mycoparasitic activities. Enzymatic studies of several *T. harzianum* strains have demonstrated that these fungi are able to produce cellulolytic complexes with high β-glucosidase activity [7]. Moreover, *T. harzianum* secretes an effective and well-balanced enzymatic system that is able to completely hydrolyze cellulose substrates into monomeric glucose [7]. The hydrolytic efficiency of *T. harzianum* represents a considerable advantage over that of *Trichoderma reesei*, a current workhorse microorganism that is widely used for enzyme preparation production aimed at plant biomass saccharification, owing to the higher levels of hemicellulase activities and, consequently, more efficient biomass hydrolysis [8].

A recent analysis of *T. harzianum* transcriptome libraries generated under conditions of growth on delignified sugarcane bagasse, crystalline cellulose and lactose showed enhanced β-mannosidase expression [9]. *T. harzianum* GH family 2 β-mannosidase (ThMan2A) was functionally characterized [4,10], although, at that time, on the basis of morphological characteristics, the microorganism was classified as a *T. reesei* isolate.

ThMan2A has an apparent molecular mass of 105 kDa and shows specific activity against p-nitrophenyl-β-D-mannopyranoside of 3.2 U·mg⁻¹ at optimum pH (pH 3.5) [10]. Mannobiose (Man₂), mannotriose (Man₃) and mannotetraose (Man₄) are all substrates of this β-mannosidase, but the rate of β-mannooligosaccharide hydrolysis was found to decrease with increasing substrate length. The ThMan2A hydrolytic efficiency for Man₂ is approximately 10 times and 15 times higher than those of Man₃ and Man₄, respectively. Notably, ThMan2A was found to bind insoluble galactomannan via a noncatalytic binding site with an affinity that was highest at very acidic pH (pH 2.0–2.8) and dropped very rapidly as the pH approached 5. β-Galactomannan (GM) binding was not observed at neutral pH [10]. Furthermore, β-mannooligosaccharides (Man₂ in particular) were found to interfere with galactomannan sorption by the enzyme. It was speculated that an additional noncatalytic domain might mediate interactions with the galactomannan [10].

ThMan2 also has transglycosylation activity, which is higher in the presence of organic solvents [4]. With the use of small-angle X-ray scattering enhanced by X-ray crystallography, similarities between the molecular shape of ThMan2 and that of *Escherichia coli* β-galactosidase, another member of GH family 2, has been demonstrated [11].

Here, we determined the X-ray structure of ThMan2A, which is the first structurally characterized eukaryotic GH family 2 β-mannosidase. We solved and refined the structure in two different crystal forms to a highest resolution of 1.9 Å. We also compared the ThMan2A structure with the only publicly available structure of a bacterial GH family 2 β-mannosidase as well as with several other members of GH family 2.

**Results**

**3D structure of *T. harzianum* β-mannosidase**

One of the open challenges in protein crystallography is to push the boundaries of the technique, and, in addition to determining the atomic content of a single crystal, to provide a more dynamic view of macromolecular structure. This may achieved by using a combi-
nation of techniques [12], by an ensemble refinement of crystallographic data [13,14], or by solving multiple crystal structures of the same target, especially in different crystal forms (e.g. [15] and [16]).

Here, we applied the third approach in attempt to gain insights not only into the structure of ThMan2A, but also into its conformational mobility and flexibility. We solved the structures of the enzyme in both tetragonal and primitive orthorhombic space groups, and compared them with each other. The crystal structure of ThMan2A in tetragonal form was solved by the use cadmium single-wavelength anomalous dispersion data to 1.9-Å resolution (see experimental statistics in Table 1). The P4₁2₁2₁ crystal form contains a monomer in the asymmetric unit, whereas the orthorhombic P₂₁₂₁₂ form of the same enzyme, solved and refined to 2.5-Å resolution, contains two copies of the molecule in the asymmetric unit.

ThMan2A consists of five distinct domains (Fig. 1A), with the catalytic domain positioned in the central portion of the structure. Such modular organization is similar to that seen for other GH family 2 enzymes, such as E. coli β-galactosidase [17], Arthrobacter sp. β-galactosidase [18], and Bacteroides thetaiotaomicron GH family 2 β-mannosidase (BtMan2A) [19]. Domain 1 (colored pink in Fig. 1A), consisting of residues 26–221, is a β-sheet domain that contains a six-stranded antiparallel β-sheet, a four-stranded antiparallel β-sheet, and an α-helical region, and is reminiscent of carbohydrate-binding modules (CBMs). Domain 2 (residues 222–346; yellow) and domain 4 (residues 738–849; orange) are structurally very similar, and show an immunoglobulin/ CBM-like β-sandwich fold. The catalytic domain, domain 3 (blue), has a classic (β/α)₈ TIM-barrel fold, and consists of residues 347–737; domain 5 (green) consists of residues 850–942. Domain 5 is structurally related to the exo-β-D-glucosaminidase C-terminal domain and the bacterial β-mannosidase C-terminal domain, according to the DALI comparison with a Z-score threshold of 7.0 as computed by the server. A similar search conducted for domain 2 revealed no structurally similar domains with a Z-score threshold of 7.0, although β-galactosidase auxiliary domains and an immunoglobulin-like domain [Protein Data Bank (PDB) ID: 1EXU] were listed at a Z-score of 5. Finally, a similar analysis performed with domain D1 showed a number of structurally related protein domains encountered in β-glucosaminidases, β-galactosidases, and glucuronidases, and also revealed somewhat unexpected structural similarities with subunit A of cellulose synthase (PDB ID: 4P02) and the ligand domain of the ephrin receptor (PDB ID: 3NBU).

Structural superposition of the refined structures determined in two different crystal forms revealed high overall rigidity of the protein structure (rmsd values lower than 0.2Å for all of the compared structures when the Cα atoms were compared; Fig. 1). The most mobile portions of the protein are the longer loops connecting the two β-strand layers in domains D1, D2, and D4 (Fig. 1). The observed mobility pattern based on the comparison of the multiple X-ray structures of ThMan2A is well correlated with the crystallographic B-factors.

ThMan2A is a native enzyme, purified from a supernatant of T. arzianum culture, and, consistent with this fact, contains a large number of N-glycosylation sites;

Table 1. Statistics of X-ray data collection and structure refinement

<table>
<thead>
<tr>
<th>Data</th>
<th>Tetragonal crystal form</th>
<th>Orthorhombic crystal form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.545</td>
<td>1.545</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁2₁2₁</td>
<td>P₂₁₂₁₂</td>
</tr>
<tr>
<td>Resolution of data (Å)</td>
<td>48.30–1.90</td>
<td>45.51–2.50</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.15 (1.60)</td>
<td>0.24 (1.25)</td>
</tr>
<tr>
<td>Mn(II) half-set correlation</td>
<td>0.99 (0.65)</td>
<td>0.95 (0.75)</td>
</tr>
<tr>
<td>CC(1/2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>7.9 (1.7)</td>
<td>8.3 (2.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.1 (96.9)</td>
<td>99.6 (96.8)</td>
</tr>
<tr>
<td>Multiplicity (%)</td>
<td>7.4 (6.2)</td>
<td>7.8 (7.8)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R cryst/Rfree</td>
<td>23.4/26.3</td>
<td>17.9/20.1</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>7241</td>
<td>14 493</td>
</tr>
<tr>
<td>rmsd (bonds; Å)</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>rmsd (angles; °)</td>
<td>1.51</td>
<td>1.73</td>
</tr>
<tr>
<td>Mean B-value (Å²)</td>
<td>47.20</td>
<td>50.50</td>
</tr>
<tr>
<td>Macromolecules</td>
<td>46.60</td>
<td>51.95</td>
</tr>
<tr>
<td>Saccharide chains</td>
<td>53.70</td>
<td>52.79</td>
</tr>
<tr>
<td>Solvent</td>
<td>49.80</td>
<td>45.84</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored region (%)</td>
<td>97</td>
<td>97.3</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>MolProbity clashscore/scorea</td>
<td>2.54/1.27</td>
<td>2.47/1.37</td>
</tr>
<tr>
<td>PDB code</td>
<td>4CVU</td>
<td>4UOJ</td>
</tr>
</tbody>
</table>

*For both structures, the MolProbity Clashscores correspond to the 100th percentile, which is the best among structures of comparable resolution. The MolProbity score combines the clashscore, rotamer and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.
we identified 10 of these in the tetragonal crystal form, and 11 in the orthorhombic crystal form. During model building and refinement, it became evident from the electron density that Asn255 located near the interface of the catalytic domain and domain 2 is heavily glycosylated, with a large and complex saccharide chain protruding from the surface of the protein. This glycosylation has an important role in crystal lattice formation, as the saccharide chain forms a considerable number of direct hydrogen bonds and water-mediated hydrogen bonds with a crystallographically related monomer in both tetragonal and orthorhombic crystal forms (Fig. 2A; detailed in Fig. 2B). This explains both the partial reproducibility of the crystallization conditions and significant anisotropy in the diffraction datasets [11].

ThMan2A molecules determined in both tetragonal and primitive orthorhombic crystal forms are heavily glycosylated. There are differences in the size of sugar decorations between individual monomers of ThMan2A in the asymmetric unit cell of the orthorhombic space group, presumably selected by the crystallization process (Fig. 1B–D): glycoside decoration is longer at Asn86 of molecule B of the orthorhombic crystal form. This crystal form also has an additional glycosylation at Asn235 as compared with the tetragonal crystal form, and a longer sugar decoration at Asn667 (Fig. 1D). The differences in glycosylation patterns between two heavier glycosylated forms of ThMan2A seem to be the main reason for formation of the lower-symmetry orthorhombic crystals with two non-crystallographic symmetry-related monomers in the asymmetric unit cell, as opposed to the tetragonal crystals formed by only one less glycosylated form of the enzyme (Fig. 2A). Nevertheless, this had a very limited effect on the overall conformation of the main chain of ThMan2A. On the basis of these observations, we used the highest-resolution structure of ThMan2A, determined to 1.9-A resolution, for further structural analysis and comparison with other members of GH family 2.

Structural comparison with bacterial GH family 2 β-mannosidases

To date, BtMan2A is the only structurally characterized bacterial GH family 2 β-mannosidase [19–21]. The enzyme has a complex five-domain fold with a
central catalytic \( \alpha/\beta \)-barrel domain surrounded by four smaller \( \beta \)-domains, which seems to be the typical tertiary structure architecture for GH family 2 enzymes. Although the \( \beta \)-mannosidases from \( B. \) thetaiotaomicron and \( T. \) harzianum share only 26% amino acid sequence identity, their domain organization is well conserved. The quaternary structures of the enzymes are, however, different: ThMan2A is monomeric in solution [11], whereas BtMan2A forms dimers with the dimerization interface, mediated by interactions between auxiliary domains 5 of both monomers [19]. Superposition of ThMan2A with BtMan2A based on secondary structure matching [22] reveals an rmsd of 2.1 \( \AA \), computed for 733 aligned residues. Comparison of the superposed structures reveals minor rearrangements of the auxiliary \( \beta \)-domains with respect to the central catalytic domain (Fig. 3A). The central domain of ThMan2A is somewhat expanded, as judged by the movement of the helices of the \( \alpha/\beta \)-barrel (Fig. 3A).

The active site residues of the structures show considerable conservation, consistent with the same
enzymatic activity and catalytic mechanism of two enzymes. Both enzymes belong to the clan of GH-A glycosidases, in which the catalytic acid–base and nucleophile residues involved in catalysis with retention of the anomic configuration of substrates belong to, respectively, β-strands 4 and 7 of the catalytic TIM-barrel domain [23]. The position and orientation of the two catalytic residues, the acid–base Glu462 and the nucleophile Glu555, in BtMan2A are essentially identical to those of the corresponding residues, Glu489 and Glu594, respectively, in ThMan2A (Fig. 3B) [19,20]. Furthermore, for the residues surrounding the catalytic center, only two side chain substitutions are observed. Trp533 in BtMan2A is replaced by Tyr573 in ThMan2A, and Cys424 in BtMan2A is replaced by Asp451 in ThMan2A (Fig. 3B). Another marked difference between BtMan2A and ThMan2A is an insertion in the loop (residues 624–638 in ThMan2A) covering the active site, which restricts access to the active site and is unique to the latter structure (Fig. 3A).

Curiously, the distribution of glycosylation sites is not even over ThMan2A domains. Most of these sites are found in the N-terminal domain (D1) and between the central catalytic domain (D3) and domain D2. No glycosylation site is observed in domains D4 and D5. The overall effect of this asymmetry in the distribution of glycosylation sites can be observed in the structural alignment of bacterial and fungal structures. The rmsd obtained for the alignment of the central catalytic domains of BtMan2A and ThMan2A (Cα of residues 317–738) is 1.53 Å (334 residues), as compared with an rmsd of 2.1 Å when all structurally aligned Cα atoms of these two structures are used (733 Cα atoms). The largest differences are observed for domain D2, and stem from the rigid body movement of this domain with respect to the central TIM-barrel domain (Fig. 3A). An interface between domains D2 and D3 is highly glycosylated in ThMan2A (Fig. 2A), and this might be the reason for the observed displacement of domain D2 around the central catalytic domain (D3). It is also remarkable that the N-glycosylation sites of ThMan2A are located only on the domains conserved across GH family 2 (D1, D2, and the catalytic domain).

**Structural comparison with GH family 2 members**

Structural comparisons of ThMan2A with other members of GH family 2 reveal interesting patterns in the macromolecular architecture of this GH family. ThMan2A is structurally related to β-d-glucosaminidase (rmsd of 3.0 Å, as computed by DALI [24] for PDB structure 2VZO [25]), β-glucuronidase (rmsd of 3.7 Å for PDB structure 3HN3; not published), and β-galactosidase (rmsd of 4.3 Å for PDB structure 1F4A [26]).

Despite the low sequence identity (only 21%, covering 43% of the ThMan2A sequence), ThMan2A and Amycolatopsis orientalis β-glucosaminidase [25] (PDB ID: 2VZO) share the same domain organization, with four small domains folded as small β-sandwiches surrounding the central α/β-barrel that contains the active site. A secondary structure-based superposition reveals an rmsd of only 2.7 Å for 683 residues superimposed with secondary structure matching [22].
Human β-glucuronidase (PDB ID: 3HN3) also preserves a central α/β-barrel surrounded by CBM-like smaller domains. However, in contrast to what is observed for β-mannosidases and β-glucosaminidases, β-glucuronidases have only two N-terminal domains, and the two C-terminal domains are missing. The rmsd computed for the superposition of ThMan2A and human β-glucuronidase is 3.2 Å for 475 Ca atoms. Finally, a comparison of β-galactosidases and ThMan2A shows that β-galactosidases, in contrast to β-glucuronidases, have an extended C-terminus forming a two-layer β-sandwich with seven and eight strands per layer. The rmsd computed for the secondary structure-based superposition of ThMan2A and E. coli β-galactosidase [26] (PDB ID: 1F4A) is 3.2 Å for Ca atoms of 461 superimposed residues (Fig. 4). These structure-based comparisons of ThMan2A with different members of GH family 2 reveal that the proteins of the family adopt a similar fold and a similar domain organization that are strongly conserved for the two N-terminal β-domains and for the catalytic core TIM-barrel domain, whereas increased diversity is observed for C-terminal domains.

**Putative GM-binding site**

Kinetic studies of ThMan2A have revealed the existence of a noncatalytic binding site for GM [10]. The binding of GM is inhibited by Man₂, Man₃, and Man₄ but is not affected by mannose, the β-mannosidase competitive inhibitor. Additionally, the inhibition constant decreases with increasing β-oligosaccharide length [10]. Together, these results seem to indicate that ThMan2A has a binding site for GM that partially overlaps or at least connects with the catalytic site.

We inspected the ThMan2A crystal structure, looking for putative binding sites for GM over the van der Waals surface, by using a grid-based approach. Grid points distributed over the surface with favorable van der Waals interactions were selected as putative binding grid points, and the two largest grid point clusters were selected for visual inspection.

The first pocket with favorable van der Waals interactions found in the ThMan2A structure is located at the interface between domains D1, D2 and the catalytic domain (D3). Although domains D1 and D2 are highly glycosylated, the glycosylation sites are solvent-exposed, leaving an attractive cavity buried by the catalytic domain (Fig. 5A). This pocket is lined by the side chains of His478, His477, His470, Gln213, Asp110, Arg344, and Asn471, which shows that this is actually a polar pocket capable of favorable interactions (Fig. 5B). Three structural features make this pocket unique. First, the pocket is branched and is of sufficient volume to recognize and accommodate branched substrates such as GM (Fig. 5B). Second, histidines (His478 and His477) give to this pocket a pH-dependent binding potential for polar binders. Finally, the rear end of the pocket connects to the exterior of the enzyme, which would be an advantage in binding extended GM polysaccharides.

A second putative binding site is observed at the interface between domain D4 and the catalytic domain. Although the pocket volume is much smaller than the volume of the first pocket, it is formed by the side chains of Tyr656, Asp681, Tyr886, Ser685, Asn847, Lys684, and Asn688, and, notably, a histidine (His845) (Fig. 5C). Interestingly, this pocket is located between the first β-sheet of domain D4 and the β-barrel of the catalytic domain, and is perfectly positioned for cooperative communication among the sites.
In order to obtain a more detailed picture of the putative interactions between ThMan2A and GM, we used the atomic coordinates of the former and the coordinates and partial charges of the latter, as available in the databank PUBCHEM [27] (CID 439336), in docking simulations to map the possible low-energy conformations of GM in the putative binding sites previously identified. As the visual inspection suggested, the second binding site was found to be too small to favorably accommodate a galactomannan molecule, and, after 100 independent simulation attempts, no favorable conformation (with negative interaction energy) was found. In contrast, simulation of docking of GM into the first putative site resulted in interaction energies in the range of $-60$ to $-70$ kcal-mol$^{-1}$ (Table 2; Fig. 5D), and at least two important polar interactions can be readily visualized from the docked poses. Asp434 interacts with one of the mannose hydroxyl groups, whereas Arg344 interacts with a hydroxyl group of the mannose unit on the opposite site. Arg344 and Ser342 also contribute their main chain atoms to the network of polar interactions (not shown in Fig. 5D). His477 and His478 do not directly interact with GM, according to the conformation proposed by molecular docking, but their location seems to be crucial for maintaining the correct geometry of this putative site. His478 is located between the side chains of Asp437, Asp434, and Asp430. The latter acidic side chains are positioned close to each other, and the presence of a protonated histidine sandwiched in between appears to be important to prevent electrostatic repulsion and structural destabilization of the site. Interestingly, His478, Asp434 and Asp437 are conserved in BtMan2A (His451, Asp407, and Asp410), which has a similar binding cavity. At the same time, ThMan2A Asp430 is replaced by Asn403 in BtMan2A,

![Figure 5](image_url)

**Fig. 5.** (A) Putative attractive pockets in the ThMan2A crystal structure. (B) Zoom into the first and the larger putative pocket at the interface between domains D1, D2, and D3. (C) Details of the second putative binding pocket at the interface between domains D3 and D4. (D) Interactions of the top 10 poses of galactomannan (yellow sticks) with ThMan2A residues of the first binding pocket. Interactions with Arg344 (side chain) and Arg325 (main chain) are shown as blue lines. Interactions between Asp430, His478 and Asp434 are shown as dotted yellow lines.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Electrostatic</th>
<th>VDW</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-43.632</td>
<td>26.864</td>
<td>70.217</td>
</tr>
<tr>
<td>2</td>
<td>-41.513</td>
<td>27.664</td>
<td>69.176</td>
</tr>
<tr>
<td>3</td>
<td>-41.781</td>
<td>27.266</td>
<td>69.047</td>
</tr>
<tr>
<td>4</td>
<td>-41.823</td>
<td>26.95</td>
<td>68.772</td>
</tr>
<tr>
<td>5</td>
<td>-41.347</td>
<td>27.408</td>
<td>68.754</td>
</tr>
<tr>
<td>6</td>
<td>-41.386</td>
<td>27.382</td>
<td>68.768</td>
</tr>
<tr>
<td>7</td>
<td>-41.502</td>
<td>26.634</td>
<td>68.136</td>
</tr>
<tr>
<td>8</td>
<td>-39.723</td>
<td>27.382</td>
<td>67.105</td>
</tr>
<tr>
<td>9</td>
<td>-39.122</td>
<td>27.495</td>
<td>66.617</td>
</tr>
<tr>
<td>10</td>
<td>-34.78</td>
<td>29.889</td>
<td>64.669</td>
</tr>
</tbody>
</table>

**Table 2.** Interaction energies of GM in the first putative binding site obtained from docking simulation.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Electrostatic</th>
<th>VDW</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-43.632</td>
<td>26.864</td>
<td>70.217</td>
</tr>
<tr>
<td>2</td>
<td>-41.513</td>
<td>27.664</td>
<td>69.176</td>
</tr>
<tr>
<td>3</td>
<td>-41.781</td>
<td>27.266</td>
<td>69.047</td>
</tr>
<tr>
<td>4</td>
<td>-41.823</td>
<td>26.95</td>
<td>68.772</td>
</tr>
<tr>
<td>5</td>
<td>-41.347</td>
<td>27.408</td>
<td>68.754</td>
</tr>
<tr>
<td>6</td>
<td>-41.386</td>
<td>27.382</td>
<td>68.768</td>
</tr>
<tr>
<td>7</td>
<td>-41.502</td>
<td>26.634</td>
<td>68.136</td>
</tr>
<tr>
<td>8</td>
<td>-39.723</td>
<td>27.382</td>
<td>67.105</td>
</tr>
<tr>
<td>9</td>
<td>-39.122</td>
<td>27.495</td>
<td>66.617</td>
</tr>
<tr>
<td>10</td>
<td>-34.78</td>
<td>29.889</td>
<td>64.669</td>
</tr>
</tbody>
</table>

VDW: Van der Waals interactions.

Energies are given in kcal-mol$^{-1}$. 

© 2014 FEBS
whereas ThMan2A Arg344 is replaced by His326 in BtMan2A.

Discussion

Our structural characterization of ThMan2A reveals that this β-mannosidase has the canonical fold of GH family 2 members, consisting of a centrally positioned catalytic TIM (α/β)8-barrel and four auxiliary, predominantly β-fold, domains.

Although previous studies have shown that new and unexpected structural information can be obtained by the collection of multiple X-ray diffraction datasets from the same macromolecular target (e.g. [16]), this was not case for multiple structures of ThMan2A. X-ray structures distributed among two different crystal forms revealed the same overall structural conformation of the polypeptide chains and, apart from the loop regions, very low mobility of the enzyme. The differences between the two crystal forms mostly reflect different glycosylation patterns of the glycoforms of the enzyme selected during the crystallization process. Small deviations in the atomic positions of the polypeptide chains of the enzyme might reflect overall rigidity of the structure and a lack of conformational changes in the absence of substrates.

The overall structure and the molecular envelope of ThMan2A are very similar to those of the other members of GH family 2, including BtMan2A [19], A. orientalis α-glucosaminidase [25], and E. coli β-galactosidase [26], and also resemble those of human β-galactosidase, although two C-terminal domains are missing in the latter structure. This is consistent with the previous small-angle X-ray scattering studies on ThMan2A, which showed similarity of its molecular envelope with that of E. coli β-galactosidase [26] (PDB ID: 1BGL [17]). At the same time, there is considerable divergence in the oligomeric forms adopted by the enzymes in solution. E. coli β-galactosidase is a tetramer, BtMan2A forms dimers, whereas ThMan2A is active as a monomer. This fact indicates that, in general, different quaternary structure arrangements do not interfere with the enzymatic activity and interactions of the GH family 2 members with the soluble substrates.

Most of the active site residues are strongly conserved in ThMan2A, and resemble, in their positions and orientations, the active site of BtMan2A. However, there is an insertion of the lid loop (residues 624–638) in ThMan2A, which covers the active site and restricts access of the ligands to the catalytic center of the enzyme. The lid loop covering the active site of ThMan2A is likely to have a functional role for selection of smaller substrates. This hypothesis agrees with the available experimental enzymatic data for both ThMan2A [10] and BtMan2A [19], which demonstrate that the former enzyme has preferences for smaller substrates (mannobiose), whereas the latter has high affinity for longer substrates.

ThMan2A is known to bind insoluble galactomannan via a noncatalytic binding site in a pH-dependent manner [10]. Our computational analysis of the ThMan2A crystallographic structure revealed two putative binding pockets at the interface between the central catalytic domain (D3) and the auxiliary domains (D1, D2, and D4). The larger of the pockets, formed by the interface between domains D1, D2, and D3 (pocket 3; Fig. 5A,B), is branched and has a sufficient volume to accommodate several residues of the mannose backbone and galactose side groups of galactomannan, thus providing a means for attachment of ThMan2A to the insoluble GM. The enzyme immobilization on the insoluble substrate might be beneficial for rapid hydrolysis of the small, soluble substrates (such as mannobiose, but also Man3 and Man4) released from GM by a concerted action of β-mannanases and α-galactosidases. Remarkably, both identified putative GM-binding sites contain histidines, which would mediate interactions between GM and the enzyme. The histidines become protonated at acidic pH, but not at alkaline pH, and this might strongly interfere with the interactions between ThMan2A and GM. This structural evidence correlates very nicely with the published experimental results showing significant increase in ThMan2A binding to and sorption on the galactomannan at acidic pH [10]. This effect is greatest at very acidic pH (down to pH 2.0) and ceases at neutral pH, when histidines become nonprotonated. This would explain the ability of GM to bind to ThMan2A at lower pH but not at pH above 6.5, as observed experimentally [10]. Docking of GM to this site confirmed that GM can undergo favorable interactions with ThMan2A through polar contacts involving a network of hydrogen bonds, with His478 playing a pivotal role in the formation of the binding site. Is ThMan2A unique in its ability to bind to GM? This question awaits additional experimental evidence, but the general conservation of the residues in the vicinity of the putative binding identified in this work may indicate that this binding capability could be more general than previously realized. The pocket is maintained in the structure of bacterial mannosidase (BtMan2A). Of the residues participating in the hydrogen bond network, His478, Asp434 and Asp437 are conserved in BtMan2A, whereas Asn325 is replaced by an aspartate, Arg344 is replaced by a histidine and
Asp430 is replaced by an asparagine. Thus, it is tempting to speculate that a putative interaction with GM mediated by the GM-binding cavity described for ThMan2A could also exist for bacterial mannosidases. Interestingly, His478 is also conserved across other structurally related members of GH family 2 – β-D-glucosaminidase (PDB ID: 2VZ0), β-glucuronidase (PDB ID: 3HN3), and β-galactosidase (PDB ID: 1F4A) – that also preserve the putative binding pocket between domains 4 and 5. In contrast, the putative GM-binding pocket is not observed in β-mannosidases from GH family 5, which adopt a much simpler single domain fold.

Indeed, the complexity of the molecular multidomain architecture shown by GH family 2 β-mannosidases is quite striking when compared with the only available structure of a GH family 5 β-mannosidase, the enzyme from Cellvibrio mixtus (CmMan5A) [28]. Dias et al. showed that CmMan5A is folded in a unique (α/β)8-barrel, which is structurally very similar to the catalytic central TIM-barrel domain of ThMan2A. The rmsd calculated for 218 superimposable Cα atoms of two structures is 2.4 Å [21]. Despite high structural similarity in their catalytic core structure, ThMan2A and CmMan5A have significant differences in their catalytic properties and substrate specificities. CmMan5A shows similar enzymatic activities against crystalline and amorphous mannans, preferentially hydrolyzes Man3 and Man4, but not mannobiose, and has markedly different catalytic efficiency against aryl-mannosides. For example, CmMan5A has an efficiency (kcat/Km) of 1.3 × 103 min⁻¹ M⁻¹ for p-nitrophenyl-β-D-mannopyranoside [28], whereas BtMan2A has a kcat/Km of 4.0 × 105 min⁻¹ M⁻¹ [19] and ThMan2A has a kcat/Km of 2.4 × 105 min⁻¹ M⁻¹ for the same substrate [10]. Furthermore, ThMan2A does not hydrolyze insoluble GM, and shows maximum activity against mannobiose [10]. The four orders of magnitude higher efficiency of GH family 2 enzymes than of CmMan5A and the substantially different substrate specificity probably reflect the diversity of β-mannosidase functions in biological settings and diverse biological rationales for their expression. It has been argued that CmMan5A is most likely attached to the inner or outer membrane of C. mixtus and has activity against both insoluble and soluble polysaccharides, as well as manno-oligosaccharides [21]. The products of its enzymatic activity, such as mannose and Man2, are released at the surface of the bacterium, and thus become available for preferential uptake by the host rather than competing microorganisms [21]. On the other hand, ThMan2A is an extracellular enzyme exported by the filamentous fungus T. harzianum to the milieu of the plant cell wall, and has the capacity to recognize and undergo adsorption to galactomannan and very efficiently hydrolyze Man2, making mannose available to its fungal producer as well as to other competing microorganisms. Immobilization of ThMan2A on GM would place the enzyme close to the source of its substrates, and thus increase the overall efficiency of GM degradation by a set of enzymes produced by T. harzianum. In this situation, the existence of additional noncatalytic domains in the ThMan2A structure (domains D1, D2, D4, and D5) is beneficial to guarantee recognition of the insoluble galactomannan and efficient adsorption on this polysaccharide, and might reflect divergent strategies for plant polysaccharide degradation adopted by the different microorganisms.

**Experimental procedures**

**Purification and crystallization of ThMan2A**

Expression and purification of native ThMan2A were carried out as described previously [29]. The purified protein was dialyzed and concentrated to 10 mg mL⁻¹ prior to crystallization. Initial crystallization conditions were as in the macromolecular crystallization reagent kits I and II (Hampton Research) and further optimized [30]. Diffraction-quality crystals were grown from mother solution containing 26% poly(ethylene glycol) 400, 0.13 M CdCl₂, and 0.1 M sodium acetate (pH 4.7). Removal of cadmium chloride was obstructive for crystal growth. As discovered during X-ray data collection and processing, the same crystallization condition rendered two crystal forms: tetragonal P43212 and orthorhombic P2₁2₁2₁. The former crystal form contains one molecule of ThMan2A in the assymetric unit cell, whereas the latter crystal form contains two molecules of the enzyme. Diffraction patterns were highly anisotropic, and maximum resolution varied strongly and consistently within the particular dataset, indicating partial lack of order for particular directions within the crystals.

**High-throughput sequencing**

Details of the molecular identification and high-throughput DNA sequencing of T. harzianum will be published elsewhere. In brief, extraction of genomic DNA from T. harzianum was performed with a protocol adapted from Chakraborty et al. [31]. DNA integrity was verified by electrophoresis on 0.8% agarose gel, and quantified with a NanoDrop spectrophotometer (Thermo Scientific). Ten micrograms of genomic DNA was used for library construction. The sample was fragmented by sonication with the CovarisTM S2 System, giving fragments of different sizes in the range from 100 bp to 200 bp. DNA ends were
repaired and phosphorylated with T4 DNA polymerase and T4 polynucleotide kinase, respectively. Adaptors P1 and P2 were ligated to the blunt-ended fragments, and DNA fragments containing both adaptors were purified from agarose gel and amplified by PCR. Finally, the DNA library was quantified, used for emulsion PCR, and submitted for sequencing on SOLiD 4 equipment. The T. harzianum genome was assembled from shotgun reads with the SOLiD System v2.0 de novo assembly pipeline, which uses the Velvet assembler [32] to assemble contigs with an error-corrected set of reads and implements a scaffolding method that fills gaps and merges fragmented contigs into longer scaffolds by using read pairing information. Coding sequences in the assembly were predicted with GeneMark-es [33]. β-Mannosidase from GH family 2 was annotated, and its deduced amino acid sequence was used in the subsequent X-ray structure determination and refinement.

X-ray structure determination

Prior to data collection at cryogenic temperature (100 K), the ThMan2A crystal was soaked in a cryoprotectant solution consisting of mother liquor containing 15% (v/v) ethylene glycol and flash cooled. The X-ray data collection was carried out on beamline MX1 at the Synchrotron Light National Laboratory (LNLS, Campinas, Brazil) [34] and on beamline ID14-2 at ESRF (Grenoble, France). All datasets were integrated by the use of xds [35] and scaled with Aimless [36]. The space group of the tetragonal crystals, which diffracted up to 1.9Å resolution, was determined to be P4_12_2_1, with cell dimensions a = b = 166.46 Å and c = 121.45 Å. The primitive orthogonal crystals (P2_12_1_2) with cell dimensions a = 165.16 Å, b = 165.63 Å and c = 123.56 Å diffracted to a maximum resolution of 2.5 Å. Details of dataset collections and structure refinement are summarized in Table 1.

The crystal structure for the highest-resolution dataset (tetragonal form) was solved with single-wavelength anomalous dispersion methods. Cadmium positions were determined with Shelxd [37,38], and phases were subsequently calculated with AUTOSOL from the PHENIX suite [39]. Solvent flattening and phase improvement were performed with the same PHENIX suite. Five per cent of the data were set aside for cross-validation analysis, and the behavior of Rfree was used to monitor and guide the refinement protocols. ARP/WARP [40] and autoBuild [41] cycles were performed to automatically build the sequence into the electron density. Structure refinement was performed with REFMAC [42] and 

BUSTER [43]. Manual inspection, correction and model building were performed with COOT [44]. The final and refined structure was used as the search model for molecular replacement as implemented in Phaser [45] to solve the ThMan2A structures in primitive orthogonal crystal form. Nine different ThMan2A structures were refined in the tetragonal space group, and two different structures in the primitive orthorhombic space group. Superposition of the structures revealed that all of them were very similar, and therefore only the highest-resolution structures were used for further comparison. Visualization and representation of the structures were carried out with PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) and UCSF CHIMERA [46].

Identification of the putative binding sites

For identification of putative binding sites on the ThMan2A structure, we computed the van der Waals potential in a grid by using a 60 × 60 × 60 Å computation box centered on the protein center of mass. A 1.5 Å grid size was used for computation. Atomic parameters from Amber force field FF99SB [47] were used for protein atoms, and an sp3 carbon atom was used as the probe. The points in the grid where a Lennard–Jones potential resulted in attractive interactions (less than −5.0 kcal·mol⁻¹), were selected as putative binding grid points. The two largest grid point clusters were selected for visual inspection. A mask was then created, covering the probe atom on grid points with favorable interactions, with NCSMask as implemented in CCP4 [48]. All calculations were performed with MCPocket, an in-house-developed code.

The putative binding pockets found in the protein structure and mapped by Mrpocket were used for galactomannan docking. The centers of mass of the grid atoms were used to define a docking box of 30 × 30 × 30 Å. Docking was achieved by global optimization of binding energies as computed by the Amber FF99SB force field and the general Amber force field for protein and ligand, respectively. In docking simulations, different conformations of the ligand were sampled with a genetic algorithm as implemented in OpenBabel [49], and then overlaid on grid atoms by use of the algorithm as implemented in MOLSHACS [50]. Finally, the initial ligand conformation based on volume superposition was energy optimized with a global optimization procedure [51] as implemented in NLOPT [52]. Libela (Muniz and Nascimento, in preparation) was used for the docking simulations, and contains an automated implementation of the concepts described here.

Acknowledgements

This work was supported by: the Brazilian funding agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via grants #2008/56255-9 and 2009/52840-7; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via grants #490022/2009-0, 301981/2011-6, 400045/2012-5, and 485950/2013-8; and CAPES and Universidade de São Paulo via grants ‘Centro de Instrumentação para estudos avançados de materiais nanoestruturados e
biosistemas’ and ‘Núcleo de Apoio à Pesquisa em Bioenergia e Sustentabilidade (NAPBS’). We are grateful to the LNLS and ESRF staff and to J. C. dos Santos for help with the X-ray data collection and structure refinement, respectively.

**Author contributions**

AMG, IP and RA planned experiments; RA, AMG, ASN and JRCM performed experiments; ASN, JRCM, RA, AMG and IP analyzed data; AMG contributed reagents or other essential material; ASN, JRCM and IP wrote the paper.

**References**

evidence for a boat-like transition state in beta-

21 Offen WA, Zeche DL, Withers SG, Gilbert HJ &
Davies GJ (2009) Structure of the Michaelis complex
of beta-mannosidase, Man2A, provides insight into
the conformational itinerary of mannoside hydrolysis.
Chem Commun (Camb) 18, 2484–2486.

matching (SSM), a new tool for fast protein structure
alignment in three dimensions. Acta Crystallogr D Biol
Crystallogr 60, 2256–2268.

23 Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon
JP & Davies G (1995) Proc Natl Acad Sci USA 92,
7090–7094.

24 Holm L & Rosenström P (2010) Dali server:
conservation mapping in 3D. Nucleic Acids Res 38,
W545–W549.

25 van Bueren AL, Ghinet MG, Gregg K, Fleury A,
Brzezinski R & Boraston AB (2009) The structural
basis of substrate recognition in an exo-beta-D-
glucosaminidase involved in chitosan hydrolysis. J Mol
Biol 385, 131–139.

26 Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber
resolution refinement of beta-galactosidase in a new
crystal form reveals multiple metal-binding sites and
provides a structural basis for alpha-complementation.
Protein Sci 9, 1685–1699.

27 Wang Y, Xiao J, Suzeck TO, Zhang J, Wang J &
system for analyzing bioactivities of small molecules.

28 Dias FM, Vincent F, Pell G, Prates JA, Centeno MS,
Tailford LE, Ferreira LM, Fontes CM, Davies GJ &
Gilbert HJ (2004) Insights into the molecular
determinants of substrate specificity in glycoside
hydrolase family 5 revealed by the crystal structure and
kinetics of Cellvibrio mixtus mannosidase 5A. J Biol

29 Colussi F, Serpa V, Delabona PS, Manzine LR,
Voltatodio ML, Alves R, Mello BL, Pereira N, Farinas
CS, Golubev AM et al. (2011) Purification, and
Biochemical and biophysical characterization of
cellobiohydrolase I from Trichoderma harzianum

30 Aparicio R, Eneiskaya EV, Kulminskaya AA, Savel’ev
AN, Golubev AM, Neustroev KN, Kobarg J &
Polikarpov I (2000) Crystallization and preliminary
X-ray study of beta-mannosidase from Trichoderma

31 Chakraborty BN, Chakraborty U, Saha A, Dey PL &
Sunar K (2010) Molecular characterization of
Trichoderma viride and Trichoderma harzianum isolated
from soils of North Bengal based on rDNA markers
and analysis of their PCR-RAPD profiles. Global J

de novo short read assembly using de Bruijn graphs.
Genome Res 18, 821–829.

33 Ter-Hovhannisyan V, Lomsadze A, Chernoff Y &
genomes using an ab initio algorithm with unsupervised

34 Guimaraes BG, Sanfelici L, Neuschwander RT,
Rodrigues F, Grizolli WC, Raulik MA, Piton JR,
MX2 macromolecular crystallography beamline: a
wiggler X-ray source at the LNLS. J Synchrotron
Radiat 16, 69–75.

Crystallogr 66, 125–132.

36 Evans PR (2011) An introduction to data reduction:
space-group determination, scaling and intensity statistics.

37 Sheldrick G (2010) Experimental phasing with
SHELXC/D: combining chain tracing with density
479–485.

38 Pape T & Schneider T (2004) HKL2MAP: a graphical
user interface for macromolecular phasing with SHELX

39 Adams P, Afonine P, Bunkoczi G, Chen V, Echols N,
Headd J, Hung L, Jain S, Kapral G, KuwSTEVE R et al.
(2011) The Phenix software for automated
determination of macromolecular structures. Methods
55, 94–106.

Automated macromolecular model building for X-ray
crystallography using ARP/wARP version 7. Nat

41 Terwilliger T, Grosse-Kunstleve R, Afonine P,
Moriarty N, Zwart P, Hung L, Read R & Adams P
(2008) Iterative model building, structure refinement
and density modification with the PHENIX AutoBuild

Refinement of macromolecular structures by the

43 Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea S &
structures with maximum likelihood in BUSTER-TNT.

tools for molecular graphics. Acta Crystallogr D Biol
Crystallogr 60, 2126–2132.

45 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn
MD, Storoni LC & Read RJ (2007) Phaser
crystallographic software. J Appl Crystallogr 40,
658–674.


