A novel \(\beta\)-xylosidase structure from Geobacillus thermoglucosidasius: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds

Giannina Espina, Kirstin Eley, Guillaume Pompidor, Thomas R. Schneider, Susan J. Crennell and Michael J. Danson

A novel β-xylosidase structure from \textit{Geobacillus thermoglucosidasius}: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds

\textit{Geobacillus thermoglucosidasius} is a thermophilic bacterium that is able to ferment both C6 and C5 sugars to produce ethanol. During growth on hemicellulose biomass, an intracellular β-xylosidase catalyses the hydrolysis of xyloligosaccharides to the monosaccharide xylose, which can then enter the pathways of central metabolism. The gene encoding a \textit{G. thermoglucosidasius} β-xylosidase belonging to CAZy glycoside hydrolase family GH52 has been cloned and expressed in \textit{Escherichia coli}. The recombinant enzyme has been characterized and a high-resolution (1.7 Å) crystal structure has been determined, resulting in the first reported structure of a GH52 family member. A lower resolution (2.6 Å) structure of the enzyme–substrate complex shows the positioning of the xylobiose substrate to be consistent with the proposed retaining mechanism of the family; additionally, the deep cleft of the active-site pocket, plus the proximity of the neighbouring subunit, afford an explanation for the lack of catalytic activity towards the polymer xylan. Whilst the fold of the \textit{G. thermoglucosidasius} β-xylosidase is completely different from xylosidases in other CAZy families, the enzyme surprisingly shares structural similarities with other glycoside hydrolases, despite having no more than 13% sequence identity.

1. Introduction

β-Xylosidases (xylan 1,4-β-xylosidases; EC 3.2.1.37) catalyse the hydrolysis of (1,4)-β-D-xylo-oligosaccharides (\textit{i.e.} xylobiose, xylotriose and xylotetrose) from their nonreducing termini into individual xylose units. β-Xylosidases are gaining increasing attention owing to their potential to reduce costs and environmental impact in several industrial and biotechnological processes. Their applications include the paper and pulp industry (\textit{e.g.} bio-bleaching recycled paper and processing wood pulp; Suurnäkki \textit{et al.}, 1997), the food industry (\textit{e.g.} the production of xylitol, improving bread dough in baking, and winemaking), oligosaccharide and thiglycoside synthesis (Jordan \& Wagschal, 2010) and second-generation biofuel production from renewable resources such as lignocellulose.

The efficient conversion of lignocellulosic biomass into liquid fuels is a major industrial goal, both to reduce dependency on fossil fuels and to reduce CO$_2$ emissions. To this end, the thermophilic ethanologen \textit{Geobacillus thermoglucosidasius} has been metabolically engineered to divert carbon flux from a mixed-acid fermentation to one in which ethanol is the major product (Cripps \textit{et al.}, 2009). However, the xylan component of lignocellulosic biomass remains
a microbially recalcitrant polymer. Xylan is a heteropoly-
saccharide containing different subunit groups in the
backbone and side chains (Sunna & Antranikian, 1997), with
the most common xylan substituents being arabinofuranose,
gluconic acid, methylglucoronic acid and acetyl groups.
Therefore, a range of carbohydrates are necessary for the
efficient utilization of typical biomass. However, once broken
down into xylo-oligosaccharides through physical and enzymic
treatment, intracellular β-xidoses play a key role in
converting these to the monosaccharide xylose, which can then be metabolized by organisms such as G. thermo-
glucosidasius.

To date, β-xidoses have been classified according to their
amino-acid sequence similarities in the Carbohydrate-
Active enZymes database (CAZy; Cantarel et al., 2009) within
ten different glycoside hydrolase families: GH1, GH3, GH30,
GH39, GH43, GH51, GH52, GH54, GH116 and GH120. High-
resolution structures are available for representatives of all
these families except for families GH52 and GH116. Czjzek et
al. (2004) described a preliminary crystallographic analysis of
a thermostable family 52 β-xidosase from G. stearo-
thermophilus, although to date no structure has been reported.

This paper reports the recombinant production, character-
ization and determination of both a high-resolution crystal
structure and a lower resolution structure of the substrate
complex of a G. thermoglucosidasiusβ-xidosase that belongs
to the CAZy glycoside hydrolase family GH52. This is the first
reported structure of a member of family GH52, and the
positioning of the substrate xylolose in the active site is
consistent with the retaining mechanism proposed for this
family. Moreover, the two-domain protein was found to have
high structural similarity to other glycosyl hydrolases with
which it shares less than 13% sequence identity. The evolu-
tionary relationships of these enzymes are discussed.

2. Materials and methods

2.1. Cloning of the β-xidosase (GH52) gene

G. thermoglucosidasius strain TM242 (ldh--; pdh<sup>op</sup>; pfl--; Cripps et al., 2009) was grown aerobically at 60°C for 16 h in
Tryptone Soya broth (Oxoid, Basingstoke, England) and its
chromosomal DNA was isolated using standard techniques
(Sambrook & Russell, 2001). The β-xidosase gene (RTMO
01743) was PCR-amplified using Phusion HF DNA poly-
merase (Thermo Fisher Scientific, Loughborough, England)
and the oligonucleotide primers 5'<sup>′</sup>-CGGCTAGCATGCCA-
AAAAACATGTATTATTAC-3' (forward) and 5'<sup>′</sup>-CGGCT-
CGGTTACCTTCCCCCTCCAC-3' (reverse). Following
amplification, the PCR product was further purified, digested with
NheI and XhoI restriction enzymes (Promega, Southampton,
England) and then ligated into the expression vector pET-
28a(+). N-terminal His tag. The sequence was confirmed to be correct.

2.2. Recombinant protein expression

Escherichia coli BL21(DE3) cells (Novagen) were trans-
formed with the pET-28a(+)-β-xidosase plasmid. Transfor-
mants were grown in TB medium [1.2% (w/v) tryptone,
2.4% (w/v) yeast extract, 72 mM K₂HPO₄, 17 mM KH₂PO₄,
0.4% (v/v) glycerol] containing kanamycin (50 µg ml⁻¹) at
37°C with shaking at 220 rev min⁻¹ until an OD₆₀₀ of 1 was
reached, at which point protein expression was induced for 5 h
by the addition of 0.3 mM IPTG (Melford, Ipswich, England).
The cells were harvested by centrifugation at 6000g for 15 min
at 4°C and resuspended in 10 ml 50 mM Tris–HCl buffer pH
7.4. 500 mM NaCl containing one EDTA-free protease-
inhibitor tablet (Roche, Welwyn Garden City, England).
The cells were lysed by five 15 s bursts of sonication using a
150 W Ultrasonic Disintegrator (MSE Scientific Instruments,
Crawley, England). A soluble cell extract was obtained by
centrifugation at 14 000g for 20 min at 4°C and was filtered
through a 0.22 µm filter (Millipore Ltd) prior to purification.

2.2.1. Selenomethionine-labelled protein. E. coli BL21 cells
harbouring the pET-28a(+)-β-xidosase plasmid were grown
overnight in 5 ml LB medium containing kanamycin
(50 µg ml⁻¹) at 37°C with shaking at 220 rev min⁻¹. The cells
were collected by centrifugation, washed twice with sterile
distilled water and then inoculated into 500 ml sterile
Selenomethionine Medium Complete (Molecular Dimensions
Ltd, Newmarket, England) in 2 l baffled flasks. The cells were
grown at 37°C with shaking at 220 rev min⁻¹ until an OD₆₀₀ of
1 was reached. The medium was then supplemented with
lysine, phenylalanine and threonine at 100 mg l⁻¹ and isoleuc-
ine, leucine and valine at 50 mg l⁻¹ to block methionine
biosynthesis in E. coli (Doublié, 2007). 2 ml Selenomethionine
Solution (250; 10 mg ml⁻¹; Molecular Dimensions) were
then added 15 min prior to induction with 0.3 mM IPTG for
5 h. The cells were harvested, resuspended and lysed as before,
adding 1 mM tris-(2-carboxyethyl)phosphine (Calbiochem, Millipore)
to the resuspension buffer.

2.3. Protein purification

β-xidosase was purified by nickel-affinity chromato-
graphy and gel filtration at 25°C on an AKTAsplorer FPLC
system (GE Healthcare, Chalfont St Giles, England). The
soluble cell lysate in 50 mM Tris–HCl buffer pH 7.4 containing
500 mM NaCl was loaded onto a pre-equilibrated HiTrap
5 ml HP column (GE Healthcare) at a flow rate of 1 ml min⁻¹.
The column was washed with the loading buffer followed by
the same buffer containing 60 mM imidazole; finally, the
β-xidosase was eluted with Tris–HCl buffer pH 8.0, 300 mM
NaCl, 200 mM imidazole.

Fractons containing β-xidosase activity were pooled and
concentrated using an Amicon Ultra 10K centrifugal filter
device (Millipore) and then further purified by gel filtration
on a Superdex 200 10/300 GL column (GE Healthcare) at a
flow rate of 0.2 ml min⁻¹ in 50 mM Tris–HCl pH 8 containing
150 mM NaCl. Fractons were analysed by SDS–PAGE and
those containing pure protein were pooled and concentrated.
to 20 mg ml\(^{-1}\) using an Amicon Ultra 10K centrifugal filter device.

Protein concentrations for the pure enzyme were determined from \(A_{280}\) nm readings using an absorption coefficient of 144 785 M\(^{-1}\) cm\(^{-1}\) as determined by ProtParam (Gasteiger et al., 2005).

2.4. Enzyme assays

\(\beta\)-Xylosidase activity was routinely assayed spectrophotometrically using the chromogenic substrate \(p\)-nitrophenyl-\(\beta\)-d-xylopyranoside (Glycosynth, Warrington, England). All assays were conducted in a volume of 1 ml at 65°C using preheated McIlvaine’s buffer (0.16 M Na\(_2\)HPO\(_4\), 0.018 M citric acid pH 7.0) and the appearance of \(p\)-nitrophenol was monitored at 410 nm; the molar absorption coefficient determined under these conditions was 8500 M\(^{-1}\) cm\(^{-1}\). Catalytic activity with \(p\)-nitrophenyl-\(d\)-arabinopyranoside (Glycosynth Ltd), \(p\)-nitrophenyl-\(d\)-arabinofuranoside (Carbosynth, Compton, Berkshire, England) and \(p\)-nitrophenyl-\(d\)-glucopyranoside (Sigma–Aldrich, Gillingham, England) was determined under the same conditions. All assays were initiated by the addition of enzyme, and one unit of \(\beta\)-xylosidase activity is defined as the release of 1 \(\mu\)mol of \(p\)-nitrophenol per minute.

Catalytic activity with the natural substrates xylobiose and xylotriose was assayed by determination of the xylose produced using NAD\(^+\)-xylose dehydrogenase as a coupling enzyme from the Megazyme \(d\)-Xylose Assay Kit (Megazyme, Bray, Ireland). Spectrophotometric assays (1 ml) were carried out at 45°C in Megazyme buffer, and the production of NADH was followed continuously at 340 nm (the molar absorption coefficient of NADH is 6220 M\(^{-1}\) cm\(^{-1}\)). One unit of \(\beta\)-xylosidase activity is defined as the production of 1 \(\mu\)mol of NADH per minute.

Assay data were analysed using the Enzyme Kinetics module in SigmaPlot 12 (Systat Software, Hounslow, England).

2.5. Mass spectrometry

Selenomethionine (SeMet) incorporation into the protein was verified by mass spectrometry. Native and SeMet-derivatized protein at 5 mg ml\(^{-1}\) concentration were analysed using a Nano LC-CHIP-MS system consisting of a 1200 Nano HPLC-Chip microfluidic device (Agilent Technologies, Santa Clara, California, USA) coupled to an Agilent 6520 Accurate-Mass Quadrupole-Time-of-Flight Liquid Chromatography-Mass Spectrometry (Q-TOF LCMS).

A 1 \(\mu\)l sample was loaded from the autosampler onto the Agilent Chip enrichment column [5 \(\mu\)m ZORBAX 300SB-C3 (300 Å), 40 nl] at a loading flow of 4 \(\mu\)l min\(^{-1}\), followed by passage to the analytical column [5 \(\mu\)m ZORBAX 300SB-C3 (300 Å), 75 \(\mu\)m \(\times\) 43 mm] at 300 nl min\(^{-1}\). The composition of mobile phase \(A\) was 0.1% (v/v) formic acid, and the liquid phase \(B\) consisted of 90% (v/v) acetonitrile and 0.1% (v/v) formic acid. Elution was obtained with increasing concentration of buffer \(B\): 3% for the first 8 min, 50% for 2 min and finally 100% for 2 min. The percentage of buffer \(B\) was then decreased to 3% and maintained at this level to allow column re-equilibration.

Samples were analyzed by ESI in positive-ion mode. Mass-spectrometric data were acquired in the \(m/z\) range 100–3000 with an acquisition rate of 1.35 spectra per second, averaging 10 000 transients. The source parameters were adjusted as follows: drying gas temperature 300°C, drying gas flow rate 5 \(l\) min\(^{-1}\), nebulizer pressure 310 kPa and fragmentor voltage 150 V. Data acquisition and processing were performed using the Agilent MassHunter Workstation acquisition software (B.02.01 Build 2116).

2.6. Structural studies

2.6.1. Protein crystalization. Crystallization was performed in 24-well plates by the hanging-drop vapour-diffusion method at 18°C. Native crystals, as well as crystals of protein derivatized with selenomethionine, were obtained from \(\beta\)-xylanase (10 mg ml\(^{-1}\) in 50 m\(M\) Tris–HCl buffer pH 8, 150 m\(M\) NaCl) mixed in equal volumes with well solution consisting of 0.1 M MES buffer pH 6.0, 4 M ammonium sulfate, 25% PEG 3350. Large crystals of approximately 0.1 \(\times\) 0.05 \(\times\) 0.05 mm in size appeared after three weeks at 18°C. Glycerol [10% (v/v) diluted in well solution] was used as cryoprotectant when the crystals were flash-cooled prior to data collection.

Co-crystallization of \(\beta\)-xylanase with its product xylitol (100 m\(M\)) was performed using the same conditions, although sodium formate (1.6 \(M\) in well solution containing 100 m\(M\) xylitol) was used as a cryoprotectant.

2.6.2. X-ray data collection and processing. Diffraction data were collected on EMBL beamline P13 at PETRA III (DESY, Hamburg, Germany). The beamline was equipped with a PILATUS 6M-F detector (DECTRIS, Baden, Switzerland) and an MD2 micro diffractometer (MAATEL, Moirans, France). Both SeMet-derivatized and native data sets were collected at \(\lambda = 0.9765\) Å (12.700 keV), slightly above the Se K edge (12.658 keV), in shutterless data-collection mode. For the SeMet-derivative crystals a wedge of 540° was collected from a single crystal with an exposure time of 250 ms per 0.2° oscillation (25% beam transmission) at a crystal-to-detector distance of 343 mm.

The native data set was measured over a total wedge of 140° with angular steps of 0.2°, an exposure time of 100 ms with a beam transmission of 100% and a crystal-to-detector distance of 319 mm. The data were processed using XDS (Kabsch, 2010) and were scaled with SCALA (Evans, 2006). Diffraction data for the native enzyme co-crystallized with xylose were collected in-house at 100 K using a Rigaku MicroMax-007 HF X-ray generator with a Saturn 94+ charge-coupled device detector at a wavelength corresponding to the Cu Ka edge (1.5418 Å). The data were processed using the HKL-2000 package (Otwinowski & Minor, 1997). Data-collection and processing statistics for both data sets are presented in Table 1.

2.6.3. Structure solution and refinement. The structure was solved via SAD phasing using the SeMet anomalous diffraction data with beta-test versions of SHELXC/D/E (Sheldrick, 2008) via a beta-test version of HKL2MAP (Pape &
Table 1
Data-collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data-collection statistics</th>
<th>Glycerol</th>
<th>Formate + xylose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C222₁</td>
<td>C222₁</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td>a = 73.56, b = 105.10, c = 195.60, α = β = γ = 90</td>
<td>a = 73.56, b = 105.35, c = 194.84, α = β = γ = 90</td>
</tr>
<tr>
<td>Total No. of reflections</td>
<td>429091</td>
<td>500871</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>83431</td>
<td>22164</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>97.3 (78.0)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.1 (5.3)</td>
<td>5.0 (2.1)</td>
</tr>
<tr>
<td>(I/</td>
<td>σ(I)</td>
<td>)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.045 (0.530)</td>
<td>0.138 (0.481)</td>
</tr>
<tr>
<td>Overall average B factor (Å²) from Wilson plot (Å²)</td>
<td>19.6</td>
<td>14.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure refinement</th>
<th>Glycerol</th>
<th>Formate + xylose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>97.78–1.70</td>
<td>7.31–2.634</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>83366</td>
<td>22136</td>
</tr>
<tr>
<td>Protein</td>
<td>5862</td>
<td>5674</td>
</tr>
<tr>
<td>Ligands/ions</td>
<td>167/21</td>
<td>26/1</td>
</tr>
<tr>
<td>Water</td>
<td>637</td>
<td>306</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.261</td>
<td>0.805</td>
</tr>
<tr>
<td>B factors (Å²)</td>
<td>23.55</td>
<td>13.75</td>
</tr>
<tr>
<td>Protein</td>
<td>47.36/69.10</td>
<td>20.69/10.63</td>
</tr>
<tr>
<td>Ligands/ions</td>
<td>35.59</td>
<td>12.5</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot analysis, residues in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured regions (%)</td>
<td>97.5</td>
<td>96.01</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MolProbity score</td>
<td>1.12 [99th percentile; 1.58 [99th percentile; N = 9248, N = 6042, 1.70 ± 0.25 Å] 2.63 ± 0.25 Å ]</td>
<td>1.70 ± 0.25 Å</td>
</tr>
<tr>
<td>Clashscore</td>
<td>2.64 [99th percentile; 2.63 [100th percentile; N = 819, N = 226, 1.70 ± 0.25 Å] 2.63 ± 0.25 Å ]</td>
<td></td>
</tr>
</tbody>
</table>

For solution of the selenium substructure, 20 sites were requested and the data were truncated to 3.0 Å resolution; for density modification a solvent content of 46% was specified, but default values were used for all other parameters. The electron-density map resulting from SHELXE after refinement of the selenium sites and ten macro-cycles consisting of 20 cycles of density modification followed by auto-tracing was automatically interpreted by ARPPWIN (Langer et al., 2008), building 697 residues.

Manual model building was performed with Coot (Emsley et al., 2010), refinement was performed using REFMAC5 (Murshudov et al., 2011) and PHENIX (Adams et al., 2010), and final model evaluation was performed using MolProbity (Chen et al., 2010).

The product-bound structure was solved by molecular replacement using the unbound structure as a model, and was rebuilt and refined as above. Both structures have been deposited in the PDB as entries 4c1o and 4c1p, respectively. Surface interface areas were determined using the PISA server (Protein Interfaces, Surfaces and Assemblies) at the European Bioinformatics Institute (Krissinel & Henrick, 2007). PDB searches and structural comparisons were performed using the DALI server (Holm & Rosenström, 2010) and PDBxView (Krissinel & Henrick, 2004).

3. Results and discussion

3.1. Characterization of the recombinant β-xylosidase

The recombinant β-xylosidase from G. thermoglucosidasius TM242 was expressed in a soluble, catalytically active form in E. coli and was purified to homogeneity by nickel-affinity and gel-filtration chromatography. Mass-spectrometric analysis gave a molecular mass of 81 980, suggesting that the N-terminal methionine had been removed (the expected molecular mass minus methionine is 81 975). Size-exclusion chromatography with standard proteins of known molecular masses gave an molecular mass of 168 000, indicating that the enzyme is a dimer.

Using p-nitrophenyl-sugar derivatives as substrates, it is clear that the enzyme has a strong preference for β-D-xylpyranoside substrates, although there was also significant activity with the α-L-arabinopyranoside derivative, albeit with a very high K_m (Table 2). The enzyme has excellent activity with the natural substrates xylobiose and xylotriose; however, these assays had to be carried out at 45°C owing to the mesophilic nature of the coupling enzyme, xylose dehydrogenase. From assays of the temperature dependence of enzyme activity using pNP-β-D-xylpyranoside, the activity of the enzyme with xylobiose (24.5 μmol pNP min⁻¹ mg⁻¹) would equate to approximately 100 μmol pNP min⁻¹ mg⁻¹ at 65°C and a k_cat of 135 min⁻¹. Similar adjustments to the values with xylotriose give 68 μmol pNP min⁻¹ mg⁻¹ at 65°C and a k_cat of 92 min⁻¹. No activity was found with xylan, thus confirming the identity of the enzyme as a β-xylosidase. The temperature optimum for the enzyme was determined to be 65°C, at which temperature the half-life was 50 min as measured by its time-dependent irreversible thermal inactivation. These data are those expected for an enzyme from G. thermoglucosidasius, which grows optimally at 60°C.

The selenomethionine-derivatized enzyme was also catalytically active (its specific activity using pNP-β-D-xylpyranoside as the substrate was 38 U mg⁻¹) and the determined molecular mass of 82 864 by mass spectrometry indicated that all of the methionine residues had been replaced by selenomethionine (expected molecular mass of 82 868).

3.2. β-xylosidase structure

The structure determined by SAD has two domains: an N-terminal domain folding into a β/α/β barrel. Analysis using the EBI PISA server (Krissinel & Henrick, 2007) indicated that the β-xylosidase exists as a stable dimer within the crystal, the symmetry-related molecule being rotated so the N-terminal domain of one molecule interacts with the C-terminal domain of the other, burying 
18 370 Å² of surface, with ten hydrogen bonds and four salt bridges across the interface (Fig. 1).

As the cryoprotectant, glycerol, can take up a conformation mimicking half of a sugar ring, the position of the active site was suggested by a glycerol molecule bound tightly in a cleft in the C11 domain and on the interface between the two monomers (the average B factor for this glycerol was less than half of the mean B factor for the 17 glycerols bound in the structure; 22.4 Å² compared with 48.3 Å²). A second glycerol molecule was bound outside the cleft but close to the first tightly bound glycerol (average B factor of 32.2 Å²) with hydrogen bonds being formed across the dimer interface, indicating a possible second sugar-binding site. Consistent with this is our finding that glycerol is a competitive inhibitor of the C12-xylosidase activity (data not shown).

Since the glycerol cryoprotectant is required at a concentration that would compete with substrate or product binding to the active site, formate was used as the cryoprotectant in the xylose co-crystal data collection. This structure confirmed the cleft as the predominant xylose-binding site, and sufficient electron density was seen to fit two xylose units having been combined into a single molecule of xylobiose (Fig. 2). It should be noted that xylose was in considerable excess over its Kₘ and therefore it is entirely possible that the enzyme has catalysed the reverse, glycosynthetic, reaction.

Xylose did not bind anywhere else in the structure. The overlapping positioning of glycerol (cryoprotectant) and the

| Table 2 |
| Catalytic activity and kinetic parameters of the recombinant β-xylosidase. |
| Errors on all values are <10%. Assays with the artificial substrates were carried out at 65°C and those with the natural substrates at 45°C. |
| (a) Artificial substrates. |
|                                | Vₘₐₓ  (µmol pNP min⁻¹ mg⁻¹) | kₗₜₙ⁻¹⁻¹ (s⁻¹) | Kₘ (mM) | kₗₜₙ/Kₘ (mM⁻¹ s⁻¹) |
|--------------------------------|----------------------------|
| pNP-β-d-xylopyranoside         | 40.5                       | 55.4          | 0.34    | 163               |
| pNP-β-d-glucopyranoside        | 1.3                        | 1.8           | 12.8    | 0.1               |
| pNP-α-l-arabinoxyranoside      | 7.6                        | 10.4          | 27.4    | 0.4               |
| pNP-α-l-arabinofuranoside      | 0.3                        | 0.4           | 0.8     | 0.5               |
| (b) Natural substrates.       |                            |              |         |                   |
|                                | Vₘₐₓ (µmol NADH min⁻¹ mg⁻¹) | kₗₜₙ⁻¹⁻¹ (s⁻¹) | Kₘ (mM) | kₗₜₙ/Kₘ (mM⁻¹ s⁻¹) |
|--------------------------------|----------------------------|
| xylobiose                      | 24.5                       | 33.5          | 2.4     | 13.9              |
| xylotriose                     | 16.9                       | 23.1          | 3.5     | 6.6               |

18 370 Å² of surface, with ten hydrogen bonds and four salt bridges across the interface (Fig. 1).

As the cryoprotectant, glycerol, can take up a conformation mimicking half of a sugar ring, the position of the active site was suggested by a glycerol molecule bound tightly in a cleft in the α/αC domain and on the interface between the two monomers (the average B factor for this glycerol was less than half of the mean B factor for the 17 glycerols bound in the structure; 22.4 Å² compared with 48.3 Å²). A second glycerol molecule was bound outside the cleft but close to the first tightly bound glycerol (average B factor of 32.2 Å²) with hydrogen bonds being formed across the dimer interface, indicating a possible second sugar-binding site. Consistent with this is our finding that glycerol is a competitive inhibitor of the β-xylosidase activity (data not shown).

Since the glycerol cryoprotectant is required at a concentration that would compete with substrate or product binding to the active site, formate was used as the cryoprotectant in the xylose co-crystal data collection. This structure confirmed the cleft as the predominant xylose-binding site, and sufficient electron density was seen to fit two xylose units having been combined into a single molecule of xylobiose (Fig. 2). It should be noted that xylose was in considerable excess over its Kₘ and therefore it is entirely possible that the enzyme has catalysed the reverse, glycosynthetic, reaction.

Xylose did not bind anywhere else in the structure. The overlapping positioning of glycerol (cryoprotectant) and the
xylobiose in the active site confirms that glycerol is mimicking the natural substrate (Fig. 2).

There are two distinct xylose-binding subsites, one of which is buried and accommodates the −1 nonreducing end of xylose, whilst the second is more open, with its main interaction with the +1 reducing end of xylose being a nondirectional stacking interaction with Tyr360. The −1 nonreducing end of xylose is held in an enclosed pocket by several hydrogen bonds (Table 3) and is distorted into a higher energy 4H3 half-chair conformation, in marked contrast to the low-energy chair conformation taken up by the +1 xylose.

A single hydrogen bond is made across the active-site cleft between O2A (of the +1 xylose) and a main-chain atom from the other molecule in the β-xylosidase dimer, showing a restricted access to the active site that prevents larger xylan polymers from entry (Fig. 3).

### 3.3. The catalytic site

Glycosides fall into two mechanistic classes: one hydrolysing the glycosidic bond with net inversion of the anomeric configuration and one doing so with net retention (McCarter & Withers, 1994). Both classes of enzyme utilize a pair of acidic residues, which are on average 9.0–9.5 Å apart in inverting enzymes but only 4.8–5.3 Å apart in those with a retaining mechanism. Previous studies have shown that β-xylosidases belonging to the family 52 glycoside hydrolases possess a retaining mechanism of catalysis in which hydrolysis of xylo-oligosaccharides to xylose occurs with net retention of configuration (Bravman et al., 2001). This is achieved via a two-step double-displacement mechanism involving a covalent glycosyl-enzyme intermediate, with each step passing through an oxocarbenium ion-like transition state (Koshland, 1953).

Within the active-site pocket of the *G. thermoglucosidasius* β-xylosidase, the two carboxylic residues that are positioned across the glycosidic O between the xylose units, Glu357 and Asp517, are 6.5 Å apart (Fig. 2), indicative of the enzyme

---

**Table 3**

<table>
<thead>
<tr>
<th>Xylobiose</th>
<th>GH52 BXP</th>
<th>Distance (Å)</th>
<th>GH39 XYP/ANX</th>
<th>Distance (Å)</th>
<th>GH120 BXP</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XylA stacking</td>
<td>Tyr360</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O2A</td>
<td>Pro100 O†</td>
<td>2.46</td>
<td>Glu160 O²</td>
<td>2.84</td>
<td>Glu405 O²</td>
<td>3.17</td>
</tr>
<tr>
<td>O3A</td>
<td></td>
<td></td>
<td>Glu160 O¹</td>
<td>2.75</td>
<td>Glu405 O²</td>
<td>2.59</td>
</tr>
<tr>
<td>O4A</td>
<td>Asp517 O²</td>
<td>3.13</td>
<td>Glu160 O¹</td>
<td>2.75</td>
<td>Asp382 O²</td>
<td>3.10</td>
</tr>
<tr>
<td>(O1 in ANX)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1B</td>
<td>Glu357 O²</td>
<td>2.99</td>
<td>Glu160 O¹</td>
<td>3.22†</td>
<td>Asp382 O²</td>
<td>2.76</td>
</tr>
<tr>
<td>O2B</td>
<td>Glu357 O²</td>
<td>3.11</td>
<td>Glu160 O²</td>
<td>2.90</td>
<td>Arg450 N⁹</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>Glu357 O¹</td>
<td>3.14</td>
<td>Glu160 O²</td>
<td>2.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His418 N²</td>
<td>2.99</td>
<td>Asn159 N²</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn159 O†</td>
<td>2.88</td>
<td>Glu324 O²</td>
<td>2.62</td>
<td>Trp383 N⁴</td>
<td>2.93</td>
<td></td>
</tr>
<tr>
<td>O3B</td>
<td>Asp367 O²</td>
<td>2.89</td>
<td>His54 N²</td>
<td>2.38</td>
<td>Glu289 O⁴</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Water</td>
<td>2.95</td>
<td></td>
</tr>
<tr>
<td>O4B</td>
<td>Thr515 O⁹</td>
<td>3.00</td>
<td>Glu324 O¹</td>
<td>2.53</td>
<td>Glu353 O²</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td>Arg715 N¹</td>
<td>3.00</td>
<td>Trp316 N⁴</td>
<td>3.18</td>
<td>Glu353 O⁴</td>
<td>3.09</td>
</tr>
<tr>
<td>Trp654 C³</td>
<td>3.10</td>
<td></td>
<td></td>
<td>Lys358 N⁷</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>O5B</td>
<td>Glu357 O²</td>
<td>3.12</td>
<td>Tyr230 O⁹</td>
<td>3.25</td>
<td>His352 O</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lys358 N⁷</td>
<td>3.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>His352 O</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3**

A surface representation, with 20% transparency, of the β-xylosidase active-site cleft with the xylobiose substrate shown in stick form. The two subunits of the dimeric enzyme are shown in green and blue.
displaying a retaining mechanism of catalysis. Furthermore, in a sequence alignment with the XynB2 β-xylosidase from *G. stearothermophilus* T-6 (GenBank code ABI49956.1), these two residues align perfectly with Glu335 and Asp495 of this enzyme. Through site-directed mutagenesis of the *G. stearothermophilus* enzyme, followed by detailed kinetic analyses including the pH-dependence of the catalytic constants, Bravman et al. (2003) have demonstrated that Glu335 functions as a nucleophile in the glycosyl-cleavage reaction, with Asp495 assisting as an acid. Asp495 then acts as a base to catalyse the hydrolysis of the glycosyl-glutamyl intermediate. These, and the other amino acids shown to interact with xylobiose, are conserved in all GH52 sequences in CAZy, and a sequence alignment of representative genera is given in Supplementary Fig. 1. The one exception is XylA from *G. stearothermophilus* 236, but this sequence is 60 amino acids shorter than any other members of the family, making alignments difficult.

3.4. Comparison with structures of enzymes with the same xylosidase function

β-Xylosidases, as identified by EC number, are found in nine glycoside hydrolase families in CAZy (Supplementary Table S1), the majority of which are of known fold, although most of the structures deposited are of proteins with other glycosidase activities (Supplementary Table S1). The available β-xylosidase structures are for members of families 39, 43 and 120, of which families 39 and 120 have mechanisms that retain product configuration, as does *G. thermoglucosidasius* β-xylosidase. The folds of these enzymes are completely different from that of the GH52 family described here, being a (β/α)8 barrel and β-sandwich in GH39 and a β-helix and a β-sandwich in family GH120 (Supplementary Table S1). However, comparison of their active sites suggests common features for protein active sites with β-xylosidase activity (Fig. 4).

Comparison of *G. thermoglucosidasius* β-xylosidase with the GH120 structure (PDB entry 3vsu; Huang et al., 2012), which also has xylobiose bound, reveals that both have a deep pocket binding the −1 xylose with many hydrogen bonds, while the +1 xylose is bound much less strongly, with few hydrogen bonds and stacking interactions with an aromatic residue. In both cases the +1 site is in close proximity to another monomer of the oligomer, although interactions are only by a water-mediated hydrogen-bond network in GH120. The closer proximity of the other monomer in *G. thermoglucosidasius* β-xylosidase (Fig. 4) causes the active-site cleft to be significantly narrower than in the GH120 structure. In both cases, the −1 sugar takes up a skew conformation, but the +1 xylose is in the relaxed chair conformation in *G. thermoglucosidasius* β-xylosidase, while that in GH120 is in the higher energy boat form.

The GH39 (PDB entry 2bfg; Czjzek et al., 2005) active site shares more similarities with the *G. thermoglucosidasius*

---

1 Supporting information has been deposited in the IUCr electronic archive (Reference: WD5231).
β-xylosidase: both have a deep cleft to accommodate the substrate, restricting access for larger substrates, although this is within one monomer in GH39. If aligned by bound substrate, the nucleophiles (Glu278 and Glu357, respectively) also align, while the general bases do not. The buried $\beta$-xylose is distorted into a similar conformation in both active sites, caused by a mixture of strong hydrogen bonds and hydrophobic interactions and a stacking interaction in GH39. The conformation of the $\alpha$-xylose cannot be determined as pNP occupies this site in 2bfg, but it is clear that both form stacking interactions with Tyr.

As in the *G. thermoglucosidasius* β-xylosidase (Fig. 3), solvent-accessibility analysis of the substrate bound in all three structures reveals that the $\beta$-xylose is completely buried (data not shown), while the $\alpha$-xylose has variable levels of exposure to solvent. This lack of exposure of the $\beta$-xylose to bulk solvent would render the enzyme incapable of cleaving bulk xylan, thus explaining the lack of xylanase activity.

3.5. Comparison with similar structures

Solving the *G. thermoglucosidasius* β-xylosidase structure by molecular replacement was not successful as no structures shared more than 20% sequence identity. However, once the structure had been solved, a search of the PDB using either the Protein Structure Comparison Service PDBeFold (Krissinel & Henrick, 2004) or the DALI server (Holm & Rosenström, 2010) showed that the *G. thermoglucosidasius* β-xylosidase shares striking structural similarities with other proteins, despite sharing no more than 13% sequence identity (Fig. 5).

As all of the PDBeFold results were also identified by DALI, only the DALI results are given in Supplementary Table S2. The structures identified as similar are usually glycoside hydrolases, from families GH15, GH63, GH65, GH78, GH92, GH94 and GH95. Of these, just GH15 and GH65 have a clan assigned, both belonging to the clan GH-L, which denotes an $(\alpha/\alpha)_6$ fold; this clan assignment would match the C-terminal domain of the *G. thermoglucosidasius* β-xylosidase. The classic position for an active site in an $(\alpha/\alpha)_6$ fold, in the groove at the centre of the barrel above the N-terminal ends of the inner helices, is also conserved in the β-xylosidase. In those structures containing both similar domains this location is conserved (Fig. 5). However, the dimeric form of the *G. thermoglucosidasius* β-xylosidase is not strictly conserved in the majority of these other families. Restriction of access to the C-terminal active site by the N-terminus of another monomer may be an
adaptation to restrict access to smaller xylo-oligosaccharides rather than bulk polymer xylan.

This identification suggests that an ancestral protein contained this combination of structural domains and has subsequently evolved to hydrolyse a wide range of glycosides to the extent that sequence identity is no longer discernible.

4. Related literature

The following references are cited in the Supporting Information: Canteral et al. (2009) and Gouet et al. (1999).

We thank Dr Rob Meijers (EMBL, Hamburg) and Dr Jean van den Elsen (University of Bath) for their expert advice, and Mr Mervyn Lewis (University of Bath) for assistance with the mass-spectrometric analysis. The research leading to the crystallization results received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 227764 (P-CUBE; http://www.p-cube.eu/). We also indebted to the National Commission for Scientific and Technological Research (CONICYT) of Chile for a studentship to GE.

References


