Role of a PA14 domain in determining substrate specificity of a glycosidase hydrolase family 3 β-glucosidase from Kluyveromyces marxianus

Erina YOSHIDA†‡, Masafumi HIDAKA†‡†, Shinya FUSHINOBU†, Takashi KOYANAGI*†, Hiromichi MINAMI*†, Hisanori TAMAKI§, Motomitsu KITAOKA‡, Takane KATAYAMA†‡*† and Hidehiko KUMAGAI*‡

†Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University, Nonoi-ch, Ishikawa 921-8836, Japan, ‡Department of Biotechnology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan, §National Food Research Institute, National Agriculture and Food Research Organization, Kannondai, Tsukuba, Ibaraki 305-8642, Japan, and ¶Department of Biochemical Science and Technology, Faculty of Agriculture, Kagoshima University, Korimoto, Kagoshima 890-0065, Japan


INTRODUCTION

β-Glucosidases (EC 3.2.1.21) occur in most living organisms and play fundamental roles in various biological processes. The enzymes catalyse the hydrolysis of O-glycosidic bonds to release Glc (glucose) units from the non-reducing end. The β-glucosidase gene from Kluyveromyces marxianus ATCC12424 (formerly known as K. fragilis) has been cloned and its amino acid sequence determined [1]. The enzyme was classified in the GH3 (glycoside hydrolase family 3) [2], one of the largest families in the CAZy database (http://www.cazy.org) [3]. Currently, GH3 comprises approx. 2400 members. Despite the large number of the sequences, structural information are available for only five enzymes, i.e. β-D-glucosidase from Hordeum vulgare (ExoI) [4], two β-N-acetylhexosaminidases from Vibrio cholerae (NagZ) and Bacillus subtilis (YbbD) (PDB code 3BMX) [5,6], exo-1,3-1,4-β-glucanase (ExoP) from Pseudoalteromonas sp. BB1 (PDB code 3F93) and β-glucosidase from Thermotoga neapolitana (Bgl3B) [7]. Among them, ExoI, NagZ and Bgl3B have been characterized both structurally and enzymatically. The domain constitutions of the three enzymes are quite different. NagZ, which consists of 330 residues, is a one domain enzyme of the (β/α)8-barrel fold. ExoI, which is a protein of 605 residues, consists of two domains: the N-terminal (β/α)8-barrel domain and the C-terminal (α/β)6-sandwich domain (see Figure 1F). Bgl3B (721 amino acid residues) has a slightly longer C-terminal stretch than ExoI, and this C-terminal extension adopts a FnIII (fibronectin type III)-like fold, which makes Bgl3B a three-domain enzyme (see Figure 1E).

β-Glucosidase from K. marxianus ATCC12424 (845 residues) shares 26% and 27% amino acid sequence identities with ExoI and Bgl3B respectively. Interestingly, it possesses a distinct domain architecture that is clearly different to ExoI and Bgl3B, i.e. its GH3-C domain [structurally corresponding to a (α/β)6-sandwich] is divided into two units by the insertion of a PA14 domain (PDB code 3F93) [8]. The PA14 domain, named after its location in the protective antigen of anthrax toxin [9], was originally identified by an iterative database search starting from the insertion sequences of GH3 enzymes [10]. The crystal structure of anthrax toxin revealed that the PA14 domain consists of two layered β-sheet [9]. The PA14 domain is found in a variety of proteins such as glycosidases, glycosyltransferases, proteases, amidases, toxins, adhesins and signalling molecules [10]. The Flol protein of yeast is involved in the cell-wall sugar-mediated flocculation, and Kobayashi et al. [11] localized the sugar-recognition site of the protein at its N-terminal PA14 domain.

Key words: chain-length specificity, crystal structure, β-glucosidase, glycoside hydrolase family 3, PA14 domain, subsite (+1).

Abbreviations used: ASU, asymmetric unit; AtCbgI, Agrobacterium tumefaciens β-glucosidase; Bgl3B, Thermotoga neapolitana DSM4359 β-glucosidase; CtBglB, Clostridium thermocellum β-glucosidase; ExoI, Hordeum vulgare β-D-glucan glucohydrolase; FnIII, fibronectin type III; GH, glycoside hydrolase; K. fragilis, Kluvyveromyces marxianus; MAD, multiple anomalous dispersion; 4-MU, 4-methylumbelliferyl-β-D-glucoside; pNP, p-nitrophenyl; PrXyl3A, Prevotella ruminicola β-xylosidase; rmsd, root-mean square deviation; SeMet, selenomethionine; TbBglB, Thermotoga neapolitana ZZ706-AC1044 β-glucosidase; VbGlll, Volvariella volvacea β-glucosidase II; WT, wild-type.

1 These authors contributed equally to this work.
2 To whom correspondence should be addressed (email takane@ishikawa-pu.ac.jp).

The atomic co-ordinates and structure factors (PDB codes 3ABZ and 3AC0) have been deposited in the Protein Data Bank.
Human fibrocystin-L, mutation of which causes polycystic kidney and hepatic disease, also has the PA14 domain, and several mutations have been mapped in the domain [12]. These observations suggest that the PA14 domain has a carbohydrate-binding role, and, in some cases, it plays a fundamental role in biological events.

Currently, the PA14 domain occurs in approx. 820 sequences, and approx. 230 are present as the insertion sequences in the GH3 enzymes. The PA14 domain also occurs in GH2, 10, 20 and 31; however, these GH enzymes have the PA14 domains separately from the catalytic core domains [10]. Consequently, the GH3 enzymes with the PA14 domain insertions should be attractive targets to elucidate the structure–function relationship of GH3 enzymes and PA14 domains, individually and in combination. As representatives of GH3 enzymes with the PA14 domain insertion, two β-glucosidases from Agrobacterium tumefaciens (ArCbgI) [13,14] and Volvariella volvacea (VvBglII) [15] and one β-xylosidase from Prevotella ruminicola (PrXyl3A) [16] have been characterized. However, our knowledge regarding the PA14 domain still remains incomplete, due to the lack of structure-based biochemical analysis.

We have succeeded in molecular cloning, purification and crystallization of β-glucosidase from K. marxianus NBRC1777 (KmBglI) which shares 98% identity with β-glucosidase from K. marxianus NBRC1777.
**EXPERIMENTAL**

**Strains and chemicals**

*E. coli* BL21 (DE3) cells and pET3a were purchased from Novagen (EMD Chemicals). pNP (p-nitrophenyl)-glycosides were purchased from Wako Pure Chemical Industries. Sophorose, gentiobiose and 4-MU-Glc (4-methylumbelliferyl β-glucoside) were purchased from Sigma–Aldrich. Laminari-and cello-oligosaccharides were from Seikagaku Corporation. Laminarietraose was further purified by size-exclusion chromatography (Bio-Gel P2; Bio-Rad).

**Enzyme preparation**

The cDNA cloning of *KmBglII* has been described in our previous paper [17]. To prepare the N-terminal His-tagged protein, the *bglII* gene was amplified by PCR using KOD-plus polymerase (Toyobo) with the primer pairs (BglI-f and BglI-r) listed in Supplementary Table S1 (at http://www.BiochemJ.org/bj/431/bj4310039add.htm). The forward primer contained an Ndel site and six repeating CAC codons for histidine, and the reverse primer contained a BamHI site. The amplified fragment was inserted into the corresponding sites of a pET3a vector. The resulting plasmid pET3a-*KmBglII* was used to transform *E. coli* BL21 (DE3) cells. Expression and purification of the enzyme was performed as described in the Supplementary Experimental section (at http://www.BiochemJ.org/bj/431/bj4310039add.htm). The purified protein was dialysed against 20 mM citrate/phosphate buffer (pH 6.0) and concentrated against 20 mM citrate/phosphate buffer (pH 6.0) and concentrated.

**Enzyme assay**

Enzyme assays were performed at the optimal pHs of the WT (wild-type) *KmBglII*. The substrates used were pNP-Glc (pNP-β-d-glucopyranoside), pNP-Xyl (pNP-β-d-xlyopyranoside), pNP-Fuc (pNP-β-d-fucopyranoside), pNP-Ara (pNP-α-L-arabinofuranoside), pNP-Gal (pNP-β-d-galactopyranoside), pNP-GlcA (pNP-β-d-glucuronide), pNP-GlcNAc (pNP-N-acetyl-β-d-glucosaminide) and pNP-GalNAc (pNP-N-acetyl-β-d-galactosaminide). The reaction mixture contained 50 mM citrate/phosphate buffer (pH 5.5), substrate and enzyme in a total volume of 300 μl. After incubation for an appropriate time at 30°C, the reaction was stopped by adding 300 μl of 1 M Na2CO3. The amounts of released pNP were determined by measuring the absorbance at 405 nm. When 4-MU-Glc was used as the substrate, fluorescence was detected by excitation and emission at wavelengths of 360 and 460 nm respectively. Activities toward β-linked gluco-oligosaccharides were examined in 50 mM citrate/phosphate buffer (pH 6.0). The reactions were performed at 30°C, and stopped by heat treatment at 95°C for 2 min. The amount of Glc released from disaccharides was measured using a glucose hexokinase assay kit (Sigma–Aldrich), in which the concentrations of the released Glc were estimated to be half of the total Glc concentrations. Activities on laminaritirole and cellobiose were determined by analysing the reaction products using high-performance anion-exchange chromatography with a CarboPac PA1 column, followed by pulsed amperometric detection (Dionex ICS3000). The elution was performed by a linear gradient of 0–0.5 M sodium acetate in 125 mM NaOH at a flow rate of 1 ml/min for 30 min.

The kinetic parameters were calculated by curve fitting the experimental data with the Michaelis–Menten equation, using Grafit 4 (Erithacus Software). The substrate concentrations were varied from 0.3–2-fold the respective *Km* values, in which transglucosylation was not observed. When trisaccharides were used as the substrates, *kcat/Km* values were determined at the low substrate concentrations (less than one-third of the *Km* value for the respective disaccharides). In the case of the hydrolysis of laminaritirole by the wild-type enzyme, Glc was the sole observable product because the formed laminariobiase was immediately hydrolysed; therefore the activity was estimated by dividing the amount of liberated Glc by three. In the other assays, activities were determined within the range that the concentrations of mono- and di-saccharide formed from trisaccharide were identical.

**TLC**

Modes of action of WT and Δ(503–512) enzymes on oligosaccharides were examined by TLC. The reaction mixtures containing 0.1 mM laminari- or 1 mM cello-oligosaccharides in a total volume of 100 μl were incubated at 30°C for 30 min. The amounts of WT and Δ(503–512) enzymes added to the assay mixtures were standardized according to their enzyme activities on the respective disaccharides. The reaction products were concentrated by vacuum centrifugation and spotted on to a TLC plate (Silica gel 60; Merck). The plate was developed by 80% acetonitrile, and the carbohydrates were visualized by heating the plate after briefly soaking in orcinol/H2SO4 reagent.

**Crystallography**

Enzyme preparation, crystallization and MAD (multiple anomalous dispersion) data collection of SeMet (selenomethionine)-labelled protein have been described previously [17]. The SOLVE/RESOLVE programs were used for site detection of selenium, phase calculation and initial model building of the MAD data set [18]. The high-resolution data set [2.15 Å (1 Å = 0.1 nm)] of SeMet-*KmBglII* was collected using synchrotron radiation (BL17A; Photon Factory, Tsukuba, Japan). Non-labelled, non-tagged *KmBglII* was prepared by the same procedure as for SeMet-*KmBglII*. The non-labelled *KmBglII* was crystallized at 20°C using a reservoir solution consisting of 40 mM potassium dihydrogen phosphate (pH 5.1), 16% (w/v) PEG [poly(ethylene glycol)] 8000, 20% (v/v) glycerol and 10 mM Glc. The crystal was flash-cooled in a nitrogen stream at 100 K, and the X-ray diffraction data set was collected using synchrotron radiation (BL6A). The data sets were processed and scaled using HKL2000 [19]. The structure of non-labelled *KmBglII* was solved by starting...
from the refined SeMet-labelled structure. Visual inspection of the models was carried out using Coot [20]. Water molecules were added using the built-in find-water function of Coot and individually checked for significant signal and consistent contact with H-bond donor/acceptor. Cycles of refinement without non-crystallographic symmetry restraints were run using Refmac5 [21]. The Figures were prepared using PyMOL (DeLano Scientific; http://www.pymol.org), and the structural alignment was performed by LSQMAN [23]. The buried surface area was calculated by PISA [24].

RESULTS

Characterization of a recombinant KmBglII

A recombinant His-tagged KmBglII was used to determine the enzymatic properties. Although six histidine residues were inserted after the N-terminal methionine residue, the numbering of the residues has been given according to the non-tagged protein (1–845) throughout the present study. The purified protein migrated as a single band with an apparent molecular mass of 95000 Da on SDS/PAGE, which closely matches the calculated molecular mass (94524 Da). Size-exclusion chromatography indicated that the native molecular mass was 390 kDa, indicating a tetramer formation (Supplementary Figure S1 at http://www.BiochemJ.org/bj/431/bj4310039add.htm). The enzyme was stable up to 45°C for 30 min and in a pH range of 4.5–9.0 (Supplementary Figure S1). The optimum temperature was 55°C, and the optimal pH was 5.5 for pNP-Glc and 6.0 for laminaribiose (Supplementary Figure S1). Specific activities on pNP-Glc were identical between His-tagged and non-tagged enzymes (results not shown).

The glycon specificity of KmBglII was examined using pNP-monosaccharides (Table 1). The enzyme showed the highest activity towards pNP-Glc ($k_{cat} = 150$ s$^{-1}$; $k_{cat}/K_m = 1500$ mM$^{-1}$.s$^{-1}$) and a considerable activity towards pNP-Xyl ($k_{cat} = 17$ s$^{-1}$; $k_{cat}/K_m = 76$ mM$^{-1}$.s$^{-1}$), whereas it showed faint activity on pNP-Fuc, pNP-Ara and pNP-Gal ($k_{cat}/K_m = 2.0$, 1.5 and 0.3, 0.2, respectively). No activity was detected for pNP-GalNAc, pNP-GlcNAc and pNP-GlcA. The substrate specificity was also examined using $\beta$-glucosyl oligosaccharides (Table 1). KmBglII hydrolysed laminaribiose (47 s$^{-1}$) and cellobiose (54 s$^{-1}$) at a similar catalytic centre activity, but the $K_m$ value for cellobiose (17 mM) was 18-fold higher than that for laminaribiose (0.96 mM). The enzyme hydrolysed sophorose with a catalytic efficiency (3.5 mM$^{-1}$.s$^{-1}$) similar to that for cellobiose (3.2 mM$^{-1}$.s$^{-1}$), whereas it hydrolysed gentiobiose at the lowest efficiency (0.33 mM$^{-1}$.s$^{-1}$) among the disaccharides. The $k_{cat}/K_m$ values for laminaritriose (0.27 mM$^{-1}$.s$^{-1}$) and cellobiose (3.2 mM$^{-1}$.s$^{-1}$) were 180- and 37-fold lower than those for laminaribiose and cellobiose respectively.

Crystallography

The crystal of SeMet-labelled protein was obtained without Glc [17], whereas the crystal of the non-labelled protein grew only in the presence of Glc. Consequently, the crystal structures of the apo-form of SeMet-labelled protein and the Glc-complexed form of the non-labelled protein were determined at 2.15 and 2.55 Å resolutions respectively. The statistics for data collection and refinement are given in Tables 2 and 3 respectively. Both crystals belong to the space group C2 and contain four subunits per ASU (asymmetric unit). There is no remarkable difference between the SeMet-labelled and non-labelled structures. Figure 1(A) shows the Glc complex structure

Table 1 Substrate specificity of KmBglII

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_m$ (mM$^{-1}$.s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNP-sugar*</td>
<td>9.9 (± 1.4) × 10$^{-2}$</td>
<td>1.5 (± 0.1) × 10$^2$</td>
<td>1.5 (± 0.2) × 10$^3$</td>
</tr>
<tr>
<td>pNP-Glc</td>
<td>2.2 (± 0.1) × 10$^{-1}$</td>
<td>1.7 (± 0.0) × 10$^1$</td>
<td>7.6 (± 0.3) × 10$^1$</td>
</tr>
<tr>
<td>pNP-Fuc</td>
<td>3.1 (± 0.3)</td>
<td>6.4 (± 0.2)</td>
<td>2.0 (± 0.4)</td>
</tr>
<tr>
<td>pNP-Ara</td>
<td>3.5 (± 0.8)</td>
<td>5.2 (± 0.5)</td>
<td>1.5 (± 0.2)</td>
</tr>
<tr>
<td>pNP-Gal</td>
<td>9.8 (± 2.2)</td>
<td>3.1 (± 0.3)</td>
<td>3.2 (± 0.5) × 10$^{-1}$</td>
</tr>
<tr>
<td>$\beta$-linked disaccharides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sophorose</td>
<td>2.0 (± 0.3)</td>
<td>7.0 (± 0.4)</td>
<td>3.5 (± 0.3)</td>
</tr>
<tr>
<td>Laminaribiose</td>
<td>9.6 (± 0.1) × 10$^{-1}$</td>
<td>4.7 (± 0.0) × 10$^1$</td>
<td>4.9 (± 0.0) × 10$^1$</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>1.7 (± 0.2) × 10$^1$</td>
<td>5.4 (± 0.3) × 10$^1$</td>
<td>3.2 (± 0.3)</td>
</tr>
<tr>
<td>Gentibiose</td>
<td>2.9 (± 0.0) × 10$^1$</td>
<td>9.4 (± 0.1)</td>
<td>3.3 (± 0.0) × 10$^{-1}$</td>
</tr>
<tr>
<td>$\beta$-linked oligosaccharides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laminaritriose</td>
<td>2.7 (± 0.2) × 10$^{-1}$</td>
<td>8.6 (± 0.4) × 10$^{-2}$</td>
<td></td>
</tr>
<tr>
<td>Cellobiose</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Neither pNP-GlcNAc, pNP-GalNAc, nor pNP-GlcA were hydrolysed by the WT enzyme.

Table 2 Data collection statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>SeMet-labelled</th>
<th>Non-labelled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>High-resolution</td>
<td>Glc complex</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97831</td>
<td>1.00000</td>
</tr>
<tr>
<td>Beamline</td>
<td>BL17A</td>
<td>BL6A</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>245.7</td>
<td>245.8</td>
</tr>
<tr>
<td>b (Å)</td>
<td>148.6</td>
<td>148.4</td>
</tr>
<tr>
<td>c (Å)</td>
<td>119.8</td>
<td>119.6</td>
</tr>
<tr>
<td>$\beta$ (°)</td>
<td>112.9</td>
<td>112.8</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00–2.15</td>
<td>50.00–2.55</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>816392</td>
<td>498570</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>213699</td>
<td>129736</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8 (3.8)</td>
<td>3.8 (3.8)</td>
</tr>
<tr>
<td>Mean I/σ</td>
<td>20.1 (3.1)</td>
<td>16.6 (2.9)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>7.2 (29.8)</td>
<td>8.9 (44.0)</td>
</tr>
</tbody>
</table>
in the ASU. The four molecules form a tetramer related by a non-crystallographic 222 symmetry. The buried molecular surface areas are approx. 1040 Å² (AC and BD interface) and 1060 Å² (AB and CD interface).

### Overall structure of KmBglII

Figure 1(D) shows the monomeric structure of the Glc-complexed KmBglII. All four chains in the ASU have identical domain architectures consisting of an N-terminal (βα)-8-fold-like domain (residues 1–295; blue), an (αβ)6-sandwich domain (residues 307–381 and 560–658; green), a PA14 domain (residues 392–559; yellow) and a C-terminal domain (residues 700–845; orange). The rmsd (root-mean-square deviation) for the Cα atoms between each four chains was within 0.23 Å, except for the PA14 domains which shared a common β-sheet. The buried molecular surface areas are approx. 1560 Å² and 1100 Å² respectively. The PA14 domains of anthrax protective antigen (PDB code 1ACC) (ββ6-sandwich structure. The patterns of the secondary structures are designated as those of BglII superimpose well with those of Bgl3B (Z score = 46.2, rmsd = 2.1 Å for 652 residues) [7] (Figures 1E and 2) and Exol (Z score = 35.5, rmsd = 2.3 Å for 491 residues) [4] (Figures 1F and 2) (the secondary structures are designated as those of Exol). The N-terminal domain of Exol adopts a canonical (βα)6-barrel fold, whereas the N-terminal domain of KmBglII adopts a (βα)6-barrel-like fold, in which strand-a and strand-c are connected by a short loop (residues 22–28) and consequently two α-helices corresponding to helix-A and -B of Exol are deleted (Figures 1G and 1H). This large deletion is also observed in Bgl3B (Figure 1H). The topology of the N-terminal domain of KmBglII is thus interpreted as a ββ6-barrel-like fold, in which the second β-strand is antiparallel. The second domain has a six-stranded β-strand (β strand i-n) sandwiched by α-helices to form an (αβ)6-sandwich structure. The patterns of the secondary structure of this domain are similar to those of Bgl3B [7] and Exol [4] with some exceptions (discussed later) (Figures 1D–1F and 2). Interestingly, the chain of the second domain of KmBglII is separated into two units by the insertion of a PA14 domain (residues 392–559; yellow) between β-strand-k and helix-

### Glucose-binding site

In the Glc-complex structure, each subunit in the ASU holds one Glc molecule. All Glc molecules are present in the β-anomeric state and adopt a C4C chair conformation (Figure 1C). The Glc molecule was bound in a cleft located between the N-terminal and (αβ)6-sandwich domains, and the cleft is covered by the loops (including Phe445 and Phe506, see below) of the PA14 domain. The β-anomeric hydroxy group forms a hydrogen bond with Glu990 (2.6 Å), and the anomeric carbon is located at a distance of 2.9 Å from Asp235. In the SeMet-labelled protein, a glycerol molecule (cryoprotectant) occupies
Figure 2  Comparison of the amino acid sequences of KmBglII, Bgl3B and ExoI

The secondary structures and their designations are shown: arrows and coils represent the β-strands and α-helices respectively. The residues involved in substrate recognition are indicated by filled triangles. The structure-based sequences alignment was carried out by MATRAS [36] and the secondary structure was assigned with DSSP [37].

© The Authors Journal compilation © 2010 Biochemical Society
each of the four Glc-binding sites (Supplementary Figure S2 at http://www.BiochemJ.org/bj/431/bj4310039add.htm).

Characterization of mutant forms of KmBglI
To better understand the catalytic mechanism of KmBglI, mutant enzymes [D225A, F445A, F508A, E590A and Δ(503–512)] were constructed and characterized (see the Supplementary Experimental section). Asp225 and Glu590 were assumed to be the catalytic nucleophile and acid/base residues respectively. Phe445, Phe508 and the region 503–512 were assumed to be involved in the subsite (+1) formation. The kinetic parameters of D225A and E590A mutants for 4-MU-Glc are summarized in Supplementary Table S2 (at http://www.BiochemJ.org/bj/431/bj4310039add.htm). The Km and kcat values of F445A and F508A for pNP-Glc were slightly decreased as compared with those of the WT enzyme (Table 4). The F445A substitution did not significantly alter the kinetic parameters for laminari- and cello-biose, whereas the F508A replacement led to an approx. 6-fold increase of the Km values and 4-fold decrease of the kcat values for the disaccharides. The kinetic parameters of the deletion mutant Δ(503–512) for pNP-Glc, laminaribiose and cellobiose were not significantly different to those of the F508A mutant. The modes of action of WT and Δ(503–512) enzymes on laminari- and cello-oligosaccharides was examined by TLC (Figures 3A and 3B). The WT enzyme hydrolysed the disaccharides significantly faster than the tri- and tetra-saccharides, whereas the Δ(503–512) mutant was capable of releasing glucose from the oligosaccharides (degree of polymerization ≥3) at a comparable level with when the disaccharides were used as substrates. The kcat/Km values of Δ(503–512) mutant for laminari- and cello-trioses were calculated to be 0.17 and 0.02 mM⁻¹·s⁻¹, being approx. 63 % and 23 % as compared with the values of the WT enzyme respectively.

DISCUSSION

Tetramer formation of KmBglI
KmBglI forms a tetramer with tight subunit–subunit interactions in the crystal to reflect the tetrameric form in solution. Most of the reported GH3 enzymes, including ExoI [4], exist as monomers, but Clostridium thermocellum β-glucosidase (CtBglI) [27], VvBglII [15] and PrXyl3A [16] have been shown to form a dimer, tetramer and nonamer respectively. The crystal structure of KmBglI suggests that the tetramer formation is not essential for catalysis because the subunit interfaces are far from the Glc-binding site.

Structure and specificity of subsite (−1)
GH3 includes enzymes acting on β-glucoside, β-xyliside, β-N-acetylglucosaminide and α-L-arabinofuranoside. Interestingly, the glycon specificity is diverse across the members; some are quite specific to one glycosidic linkage, whereas others can act on two or more linkages. KmBglI hydrolyses pNP-Glc with a high catalytic efficiency and hydrolyses pNP-Xyl, pNP-Fuc, pNP-Ara and pNP-Gal with lower efficiencies (Table 1). TnBglI, the closest homologue of Bgl3B (96 % identity), hydrolyses pNP-Glc most efficiently and pNP-Xyl, pNP-Fuc and pNP-Gal at the levels of 2–43 % of pNP-Glc [28]. In contrast, ExoI is reported to be specific to β-glucoside [29].

The KmBglI–Glc complex structure was compared with the Glc complex structure of Bgl3B (Figure 4A) [7] and the S-laminaribioside (Glcβ1–3Sglcβ1–3Oglc–thiopNP) complex structure of ExoI (Figure 4B) [30]. The Glc molecule of KmBglI

![Table 4](image)

**Table 4.** Kinetic parameters of mutant KmBglI proteins on various substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Km (mM)</th>
<th>kcat (s⁻¹)</th>
<th>kcat/Km (mM⁻¹·s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNP-Glc</td>
<td>0.02</td>
<td>0.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>0.02</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>LaminariBiose</td>
<td>0.01</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Xyl3A</td>
<td>0.01</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Fuc</td>
<td>0.01</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Gal</td>
<td>0.01</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>pNP-Ara</td>
<td>0.01</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

© The Authors Journal compilation © 2010 Biochemical Society
As for Clostridium thermocellum laminari-oligosaccharides (G1, glucose; G2, laminaribiose; G3, laminaritriose; G4, laminartetraose) and Km uses this tryptophan residue for the formation of subsite (χ suggests that Phe445 is important, but not essential, for the catalytic activity. The ratios (fold) are estimated by comparing their kcat/Km values. As for Clostridium thermocellum β-glucosidase (CtBglB) [27], Azospirillum irakense β-glucosidase (AiCelA) [31] and ExoI [30], the specific activities (units/mg) were compared.

overlapped with the Glc in Bgl3B and the non-reducing end Glc of S-laminaribioside in ExoI; thus it must occupy the subsite (−1). The catalytic residues of Bgl3B (Asp242 and Glu491) and ExoI (Asp285 and Glu590) are well overlapped on Asp235 and Glu520 of KmBglI. The results of the mutational study revealed the importance of these residues (Supplementary Table S2). Pozzo et al. [7] reported that the rotamer angle of the tryptophan residue that is next to the nucleophile significantly differs between the structures of Bgl3B (χ = −92°) and ExoI (χ = −169°). ExoI uses this tryptophan residue for the formation of subsite (−1) [30], whereas Bgl3B uses it for the formation of subsite (−1). In KmBglI, the rotamer angle of the tryptophan residue (Trp236) is quite similar (χ = −90°) to that of Bgl3B, and the structure of subsite (−1) of KmBglI is almost identical with that of Bgl3B. The strict substrate specificity of ExoI and the relatively relaxed substrate specificities of Bgl3B and KmBglI may arise from their structural differences at subsite (−1).

Structure and specificity of subsite (+1)

KmBglI hydrolysed sophorose, laminaribiose, cellulbiose and gentiobiose with varied catalytic efficiencies (0.33–49 mM−1·s−1), whereas ExoI can hydrolyse them at a comparable level [30]. In the S-laminaribioside complex structure of ExoI, the reducing end Glc is clamped by Trp286 and Trp434 at subsite (+1) (Figures 4B and 4E). When a laminaribiose molecule was modelled into the structure of KmBglI by reference to S-laminaribioside in ExoI, the side chains of Phe445 and Phe508 from the PA14 domain sandwich the Glc molecule (Figures 4B and 4C). Phe445 of KmBglI is located at the structurally neighbouring position of Trp286 of ExoI, and its side chain forms hydrophobic interactions with C5 and C6 of the Glc. This residue is not strongly conserved among the PA14 domains inserted into GH3 enzymes (Figure 5), and the F445A substitution had a moderate effect on the activity (Table 4). These results suggest that Phe445 is important, but not essential, for the catalytic activity of KmBglI. On the other hand, the side chain of Phe508 in the loop linking β-strands 7 and 8 (7-8loop) (Figure 5) is located at an almost identical position with that of Trp434 of ExoI, although the positions of their Ca atoms are 8.3 Å apart (Figure 4B). The F508A substitution had a pronounced effect on the activities for disaccharides. The catalytic efficiencies toward laminaribiose and cellobiose decreased to 5 and 3% of the WT enzyme respectively. These results, and its conservation among the PA14 domain sequences inserted into GH3 (Figure 5), indicate that Phe438 plays a significant role in the catalytic activity. The PA14 domain-mediated subsite (+1) formation of KmBglI is thus totally different to that of ExoI, for which Trp286 from the (β/α)8-fold domain and Trp438 from the (α/β)8-sandwich domain are responsible. This structural difference may explain the difference in the catalytic efficiencies between the enzymes towards β-glucosides. In Bgl3B, the edge strand in the (α/β)8-sandwich domain (β-strand-k of ExoI and KmBglI) is missing and the corresponding region is replaced by an α-helix and a long unstructured loop (residues 409–434) (Figure 2). This loop, although its tip is not visible in the structure, is presumably involved in subsite (+1) formation (Figure 4D) [7]. These findings suggest that the three enzymes have independently evolved with respect to subsite (+1) formation.

PA14 domain

Several GH3 enzymes effectively hydrolyse oligosaccharides. The ratios of the kcat/Km values of ExoI [30] and the relative activities of CtBglB [27] and Azospirillum irakense β-glucosidase (AiCelA) [31] for cellobiose/celletroiose/celletroiose are reported to be 0.1:1:1.5, 3.2:1:0.9 and 2.6:1:1 respectively (Figure 3C). In contrast, KmBglI, AtCBgl and VvBGglI, all of which possess a PA14 domain insertion, are quite specific towards small substrates. The kcat/Km ratios of KmBglI for laminaribiose/laminaritriose and cellulbiose/celletroiose are 180:1 and 37:1 respectively (Figure 3C), and the release glucose from pNP-cellobiose [14,15]. The difference of the chain-length specificities between ExoI and KmBglI can be explained by the structural difference at the entrances of their catalytic pockets. The active site of ExoI is covered by a loop between strand-j and helix-J2 (30 residues including Trp434) from the (α/β)8-sandwich domain, which results in the formation of

Figure 3 Modes of action of WT KmBglI and the Δ(503–512) mutant on oligosaccharides

The amounts of WT and Δ(503–512) enzymes added were standardized according to enzyme activities on the respective disaccharides; i.e. 0.3 μg of WT and 10 μg of Δ(503–512) were used for the hydrolysis of laminari-oligosaccharides, and 4.5 μg of WT and 90 μg of Δ(503–512) were used for the hydrolysis of cello-oligosaccharides. TLC analysis of the reaction products containing laminari-oligosaccharides (G1, glucose; G2, laminaribiose; G3, laminaritriose; G4, laminartetraose) and cello-oligosaccharides (G1, glucose; G2, cellulbiose; G3, celletroiose; G4, celletroiose) revealed the preference for disaccharides relative to trisaccharides of several GH3 enzymes. With regard to WT and Δ(503–512) KmBglI, the ratios (fold) are estimated by comparing their kcat/Km values.
The crystal structure of GH3 \( \beta \)-glucosidase with an inserted PA14 domain

### Figure 4 Structural comparison among KmBglI, Bgl3B and ExoI at the active sites

(A and B) Wall-eyed stereoviews showing the superpositioning of KmBglI and Bgl3B (A), and KmBglI and ExoI (B). KmBglI and bound Glc are shown as in Figure 1. Bgl3B and bound Glc molecule, and ExoI and bound S-laminaribiose molecule (ball-and-stick) are coloured in grey. The broken lines indicate the C\( \alpha \)-atom distances between Phe508 of KmBglI and Trp434 of ExoI and between Trp226 of KmBglI and Met316 of ExoI. (C–E) Active-site pocket formations of KmBglI (C), Bgl3B (D) and ExoI (E). The molecular surface of the N-terminal domain (blue) and the (\( \alpha/\beta \))_{6}-sandwich domain (green) are shown. The surfaces are rendered transparent to show the bound molecules. (C) PA14 domain (yellow) and the side chains of Trp226, Phe445 and Phe508 of KmBglI are shown as a ribbon diagram and stick models respectively. The broken line indicates the disordered loop between residues 540–542. The bound Glc molecules in the KmBglI structure (magenta) and the modelled laminaribiose molecule (cyan) by reference to S-laminaribioside of ExoI are shown as stick models. (D) A long loop (residues 409–434, red) and the side chains of Trp243 of Bgl3B are shown as a ribbon diagram and stick models respectively. The broken line indicates the disordered structure between residues 418 and 422 (the amino acid sequence is DSWGT). (E) The side chains of Trp286 and Trp434, and the bound S-laminaribioside molecule are shown as stick models.

a deep pocket (Figure 4E). In contrast, the corresponding loop of KmBglII takes a different conformation and yields a shallow pocket. Instead, the PA14 domain covers the active site and consequently deepens it. The PA14 domain approaches the active site via two loops; one is a loop linking \( \beta \)-strand 3 and 4 (residues 444–459 including Phe445) and the other is the 7-8loop (residues 498–514 including Phe508). The latter may sterically hinder the binding of long-chain oligosaccharides (Figure 4C and see below). \( \text{Pr} \)Xyl3A, which also possesses a PA14 domain insertion, is able to act on xylo-oligosaccharides, as well as xyllobiose [16]. Sequence alignment of the PA14 domains from four enzymes showed that \( \text{Pr} \)Xyl3A lacks the region corresponding to the 7-8loop of KmBglII (Figure 5). Accordingly, we constructed the loop deletion mutant, \( \Delta (503–512) \), and examined its substrate specificity (Figure 3 and Table 4). The deletion caused a drastic decrease of the \( k_{\text{cat}}/K_{m} \) values for the disaccharides to 1% of the WT enzyme, probably due to the lack of Phe508. However, remarkably, the mutant retained 63% and 23% activities towards laminari- and cello-trioses respectively, as compared with the WT enzyme. Consequently, the ratios of the \( k_{\text{cat}}/K_{m} \) values of the mutant for laminaribiose/laminaritriose and cellobiose/cellotriose became 3:1 and 2:1 respectively, indicating that the mutant no longer showed the chain-length specificity (Figure 3C).

In ExoI, a helix-like loop that connects the (\( \beta/\alpha \))_{8}-barrel and (\( \alpha/\beta \))_{6}-sandwich domains could act as a hinge that allows the domains to move relative to each other during successive catalytic
events [32–34]. In *Km* BglI, the N-terminal and (α/β)6-sandwich domains are also connected by a linker, but the domains are unlikely to move because the tetrameric assembly would limit such movement. Alternatively, the subsite (+1) could become flexible by the significantly higher B-factor of the PA14 domain to facilitate the capture of the substrates (Figure 1B and Table 3). Formation of the catalytic pocket via a domain–domain interaction has been reported in several GHs. For example, an active site of *E. coli* β-galactosidase (LacZ) is formed by the interaction (known as α-complementation) between the (βα)8-barrel domain and the N-terminal domain that adopts a similar structure to the PA14 domain ([Z score = 8.3, rmsd = 3.2 Å for 121 residues]) [35]. Interestingly, both of the PA14 domain of *Km* BglI and the N-terminal domain of LacZ participate in their active site formation via loops extending from the same edge of the domains (results not shown).

The PA14 domain frequently occurs in GHs, bacterial toxins, yeast adhesins and mammalian signalling molecules [10], and the carbohydrate-binding function has long been claimed without any structural evidence. In the present study, using the *Km* BglI protein, we revealed the structural basis of the interaction between the PA14 domain and a carbohydrate. Unfortunately, since the sequences of PA14 domains are not strongly conserved, especially at the loop regions, the role(s) of the domain cannot be generalized at present. But, particularly for GH3 members, the domain could form the subsite(s) and affect the substrate specificity. The overall structure of *Km* BglI is similar to those of Bgl3B and ExoI, indicating that these enzymes have evolved from a common ancestor. However, their active-site structures are quite different. *Km* BglI appears to have evolved to become more active on small substrates by accepting the insertion of the PA14 domain into the (α/β)6-sandwich.

**AUTHOR CONTRIBUTION**

Erina Yoshida performed the biochemical experiments, interpreted the results and participated in the writing of the paper. Masafumi Hidaka performed the structural analysis, interpreted the results and participated in the writing of the paper. Shinya Fushinobu supervised the structural studies. Takashi Koyanagi, Hiromichi Minami and Hisanori Tamaki provided their expertise in genetic experiments. Motomitsu Kitaoka provided his expertise in enzymology. Takane Katayama designed the research, interpreted the results and edited the paper. Hidehiko Kumagai designed the research and obtained the funding for the work.

**ACKNOWLEDGEMENTS**

We thank the staff of the Photon Factory for the X-ray data collection and Dr N. Nagano for helpful comments on our manuscript.

**FUNDING**

This work was supported, in part, by a Grant-in-Aid from the New Energy and Industrial Technology Development Organization (NEDO), Japan.

**REFERENCES**

The crystal structure of GH3 \( \beta \)-glucosidase with an inserted PA14 domain

Received 8 March 2010/20 July 2010; accepted 27 July 2010
Published as BJ Immediate Publication 27 July 2010, doi:10.1042/BJ20100351


22 Reference deleted


© The Authors Journal compilation © 2010 Biochemical Society
SUPPLEMENTARY ONLINE DATA

Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 $\beta$-glucosidase from Kluyveromyces marxianus

Erina YOSHIDA$^*$, Masafumi HIDAKA†‡, Shinya FUSHINOBU†, Takashi KOYANAGI*, Hiromichi MINAMI*, Hisanori TAMAKI*, Motomitsu KITAOKA‡, Takane KATAYAMA*‡ and Hidehiko KUMAGAI*

*Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University, Nonoichi, Ishikawa 921-8836, Japan, †Department of Biotechnology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan, ‡National Food Research Institute, National Agriculture and Food Research Organization, Kanon-dai, Tsukuba, Ibaraki 305-8642, Japan, and §Department of Biochemical Science and Technology, Faculty of Agriculture, Kagoshima University, Korimoto, Kagoshima 890-0065, Japan

EXPERIMENTAL

Expression and purification of His-tagged KmBglI

The transformants were grown in LB (Luria–Bertani) medium containing 50 mg/l ampicillin at 18°C to an attenuation at 600 nm of 0.5. Isopropyl $\beta$-D-thiogalactopyranoside was subsequently added to a final concentration of 0.5 mM and the cultures were incubated for a further 20 h. Cells were harvested, suspended in 50 mM sodium phosphate buffer (pH 8.0) containing 300 mM NaCl and 10 mM imidazole, and disrupted by sonication. The protein was purified using Ni-NTA (Ni2+-nitrilotriacetate) affinity chromatography according to the manufacturer’s instruction (Qiagen). The active fractions were dialysed against 50 mM Tris/HCl buffer (pH 8.0) and further purified using MonoQ5/50GL anion-exchange chromatography (GE Healthcare), in which the protein was eluted by a liner gradient of 0–0.5 M NaCl. During the purification steps, 2 mM 2-mercaptoethanol was added to the buffers.

Size-exclusion chromatography

The native molecular mass of the enzyme was estimated using a Superdex 200 10/300GL column (GE Healthcare). The protein was eluted with a buffer consisting of 50 mM sodium phosphate (pH 6.0) and 150 mM NaCl with a flow rate of 0.25 ml/min. Thyroglobulin (669 kDa), ferritin (443 kDa), aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (43 kDa) were used as standards.

Temperature and pH profile

The temperature and pH stabilities were determined by measuring the residual pNP-Glc-hydrolysing activities after incubating the enzyme (1 mg/ml) in 50 mM potassium phosphate buffer (pH 6.0) for 30 min at various temperatures and after dialysing the enzyme overnight against various 10 mM buffers at 4°C respectively. The buffers used were citrate/phosphate (pH 3.0–7.0), sodium phosphate (pH 6.5–8.0) and Tris/HCl (pH 7.5–9.0). The optimal temperature was examined by incubating the reaction mixture containing 50 mM citrate/phosphate (pH 5.5) and 0.5 mM pNP-Glc at various temperatures for a short period of time (<1 min) to prevent damage to the enzyme. The optimal pH values were determined for pNP-Glc (0.5 mM) and laminaribiose (1 mM) using the buffers described above (50 mM).

Figure S1 Physicochemical properties of a recombinant KmBglI

(A) Determination of the native molecular mass by size-exclusion chromatography. (B) Thermostability. The value at 40°C was taken as 100%. (C) pH stability. The value at pH 9.0 was taken as 100%. (D) Optimal temperature. The value at 55°C was taken as 100%. (E and F) Optimal pH values. The values at pH 5.5 and 6.0 were taken as 100% for the hydrolysis of pNP-Glc and laminaribiose respectively. See the Supplementary Experimental section for detailed assay conditions.

Mutagenesis

Mutant enzymes, D225A, F445A, F508A and E590A, were constructed using a the QuikChange® site-directed mutagenesis...
Table S1  Oligonucleotides used in the present study

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of pET3a-KmBglI</td>
<td>(BglI-f) 5'-CCCCCATATGCACCCACCCACCCACCTCTAAATTGATGGTAAGCAT-3'</td>
</tr>
<tr>
<td></td>
<td>(BglI-r) 5'-CCCCGGATCTACACAGCACCCTTCCAAATACAATTCGCT-3'</td>
</tr>
<tr>
<td>Site-directed mutagenesis*</td>
<td></td>
</tr>
<tr>
<td>D225A</td>
<td>5'-GTTAATGTCCGGCTGTTGCGAC-3'</td>
</tr>
<tr>
<td>F445A</td>
<td>5'-GTTCATGCGCAGGATTACACT-3'</td>
</tr>
<tr>
<td>F508A</td>
<td>5'-GTAGTTCTGGCGGGGACGTTGAC-3'</td>
</tr>
<tr>
<td>E590A</td>
<td>5'-ATGGGAAACCACCGGTATGACAG-3'</td>
</tr>
<tr>
<td>Construction of Δ(503–512)</td>
<td>(forward) 5'-AAAGAAAAACCCACAAAGTTGAC-3'</td>
</tr>
<tr>
<td></td>
<td>(reverse) 5'-TTCTTGGTGTGCTTGGCTCAA-3'</td>
</tr>
</tbody>
</table>

*These primers and their complementary primers were used in pairs.

Figure S2  The liganded glycerol molecule in the high-resolution structure is shown as a ball-and-stick model with the \(|F_o|-|F_c|\) electron density map (2.5σ).

The broken lines indicate hydrogen bonds. Residues are coloured according to the domain as in Figure 1(A) of the main text.

Received 8 March 2010/20 July 2010; accepted 27 July 2010
Published as BJ Immediate Publication 27 July 2010, doi:10.1042/BJ20100351

Table S2  Activity of D225A and E590A mutants of KmBglI on 4-MU-Glc

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>(K_m) (mM)</th>
<th>(k_{cat}) (s(^{-1}))</th>
<th>(k_{cat}/K_m) (mM(^{-1})·s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>5.8 (± 1.4) × 10(^{-2})</td>
<td>6.1 (± 0.5) × 10(^{1})</td>
<td>1.0 (± 0.2) × 10(^{1})</td>
</tr>
<tr>
<td>D225A</td>
<td>2.0 (± 0.3) × 10(^{-2})</td>
<td>3.4 (± 0.2) × 10(^{-4})</td>
<td>1.7 (± 0.2) × 10(^{-2})</td>
</tr>
<tr>
<td>E590A</td>
<td>3.5 (± 0.3) × 10(^{-2})</td>
<td>1.1 (± 0.0) × 10(^{-1})</td>
<td>3.1 (± 0.2)</td>
</tr>
</tbody>
</table>

The mutagenic primers used are listed in Supplementary Table S1. The deletion mutant \(Δ(503–512)\) was constructed by inverse PCR using pET3a-KmBglI as a template and the primer pair (Supplementary Table S1). The entire fragments used for later manipulations were sequenced to ensure that no base change other than those planned had occurred. Expression and purification of the mutant enzymes were performed in a similar manner as described for the wild-type enzyme.