Structural and functional analysis of three β-glucosidases from bacterium \textit{Clostridium cellulovorans}, fungus \textit{Trichoderma reesei} and termite \textit{Neotermes koshunensis}

Wen-Yih Jeng\textsuperscript{a,c}, Nai-Chen Wang\textsuperscript{a,c}, Man-Hua Lin\textsuperscript{a,c}, Cheng-Tse Lin\textsuperscript{a}, Yen-Chywan Liaw\textsuperscript{b}, Wei-Jung Chang\textsuperscript{a}, Chia-I Liu\textsuperscript{a,c,d}, Po-Huang Liang\textsuperscript{a,d}, Andrew H.-J. Wang\textsuperscript{a,c,d,*}

\textsuperscript{a} Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan
\textsuperscript{b} Institute of Molecular Biology, Academia Sinica, Taipei 115, Taiwan
\textsuperscript{c} Core Facility for Protein Production and X-ray Structural Analysis, Academia Sinica, Taipei 115, Taiwan
\textsuperscript{d} Institute of Biochemical Sciences, National Taiwan University, Taipei 106, Taiwan

\textbf{A R T I C L E  I N F O}

\textbf{ABSTRACT}

β-Glucosidases (EC 3.2.1.21) cleave β-glucosidic linkages in disaccharide or glucose-substituted molecules and play important roles in fundamental biological processes. β-Glucosidases have been widely used in agricultural, biotechnological, industrial and medical applications. In this study, a high yield expression (70–250 mg/l) in \textit{Escherichia coli} of the three functional β-glucosidase genes was obtained from the bacterium \textit{Clostridium cellulovorans} (CcBglA), the fungus \textit{Trichoderma reesei} (TrBgl2), and the termite \textit{Neotermes koshunensis} (NkBgl) with the crystal structures of CcBglA, TrBgl2 and NkBgl, determined at 1.9 Å, 1.63 Å and 1.34 Å resolution, respectively. The overall structures of these enzymes are similar to those belonging to the β-retaining glycosyl hydrolase family 1, which have a classical (β/α)\textsubscript{8}-TIM barrel fold. Each contains a slot-like active site cleft and a more variable outer opening, related to its function in processing different lengths of β-1,4-linked glucose derivatives. The two essential glutamate residues for hydrolysis are spatially conserved in the active site. In both TrBgl2 and NkBgl structures, a Tris molecule was found to bind at the active site, explaining the slight inhibition of hydrolysis activity observed in Tris buffer. Manganese ions at 10 mM exerted an approximate 2-fold enzyme activity enhancement of all three β-glucosidases, with CcBglA catalyzing the most efficiently in hydrolysis reaction and tolerating Tris as well as some metal inhibition. In summary, our results for the structural and functional properties of these three β-glucosidases from various biological sources open important avenues of exploration for further practical applications.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

D-Glucose is widely utilized in the food industry and is also employed as a precursor in the production of ethanol and other chemicals. Sugar cane, starch and cellulose are the primary sources for glucose production. As such, it requires the exploitation of large tracts of agricultural land to satisfy the global demand for these products. In order to avoid this growing infringement on the amount of land available for food production, the use of agricultural and forestry wastes have been investigated as a preferable alternative to producing cellulosic ethanol (Service, 2007). The β-1,4 linked glucose polymer, cellulose, can be enzymatically hydrolyzed into glucose by three types of cellulase: endoglucanase (EC 3.2.1.8), which randomly digests the crystalline cellulose, exoglucanase or cellobiohydrolase (EC 3.2.1.91), which cleaves cellulobiose off from the non-reducing ends of the cellulose polymer chain, and β-glucosidase (EC 3.2.1.21), which breaks the β-1,4 glucosidic bonds of cellulose to produce glucose monomers (Sticklen, 2008).

Enzymatic hydrolysis of cellulose is a major limitation in lignocellulose conversion to glucose (Lynd et al., 2002; Zhang and Lynd, 2004). During enzymatic hydrolysis, insufficient β-glucosidase activity causes an accumulation of cellulose with the consequence...
that both endoglucanase and exoglucanase activities are severely inhibited (Bommarius et al., 2008). In the overall degradation route of cellulose, glucose accumulation reduces the efficiency of enzymatic saccharification. The conversion processes from lignocellulose to bioethanol, such as simultaneous saccharification and fermentation (SSF) or simultaneous saccharification and co-fermentation (SSCF) present possible strategies for reducing the product inhibition of glucose (Lynd et al., 2002).

β-Glucosidases, a heterogeneous group of exo-type glycosyl hydrolases, cleave β-glucosidic linkages in disaccharide or glucose-substituted molecules. β-Glucosidases occur in all kingdoms of life and play fundamental biological roles in processes such as degradation of cellulose and other carbohydrates for nutrient uptake (Beguin, 1990; Bhatia et al., 2002) and developmental regulation or chemical defense against pathogen attack (Brzobohaty et al., 1993; Leah et al., 1995; Zagrebelny et al., 2008). According to the classification of Carbohydrate-Active enzymes (CAZY) (http://www.cazy.org) (Cantarel et al., 2009; Henrissat, 1991), β-glucosidases have been classified into glycosyl hydrolase families 1 and 3. In a structural view, glycosyl hydrolase family 1 enzymes fold into an (β/α)8 barrel structure and act as a β-retaining mechanism using Glu as the catalytic nucleophile (Wang et al., 1995); whereas glycosyl hydrolase family 3 β-glucosidases, with only a few structures solved, exploit an Asp residue in nucleophile attacks.

In this study, we report the crystallization and X-ray analysis of three β-glucosidases: CcBglA from Clostridium cellulovorans (anaerobic bacterium), TrBgl2 from Trichoderma reesei (Ascomycota fungus), and NkBgl from Neotermes koshunensis (termite), all highly expressed in Escherichia coli using the pET system (Novagen). A Tris molecule in the active site of both TrBgl2 and NkBgl was located, and this may be responsible for the inhibition of the hydrolase activity. Interestingly, Mn2+ in the 2–10 mM concentration, significantly enhanced the hydrolase activity of all three β-glucosidases, CcBglA, TrBgl2 and NkBgl at a rate of 150% to 230%. According to our results, CcBglA exhibits the highest potential for use in the process of lignocellulosics ethanol production, especially for the processes of SSF and SSCF due to its high enzyme activity at a moderate reaction temperature near neutral pH.

2. Materials and methods

2.1. Enzymes, chemicals and bacterial strains

E. coli strains used in this study were DH5α, BL-21 (DE3) (Novagen), and XL1-Blue (Stratagene) cells. The enzymes for DNA manipulation were purchased from New England Biolabs. Novozyme 188, salicin, glucose, 3,5-dinitrosalicylic acid, p-nitrophenyl-β-D-glucopyranoside (pNPG), resorufin-β-D-glucopyranoside and other major chemicals were purchased from Sigma–Aldrich. The β-glucosidase of Novozyme 188 was further fractionated by size exclusion chromatography using a Sephacryl S-200 column equilibrated with 100 mM NaCl in 50 mM Hepes buffer pH 8.0 (Boisset et al., 2000). The active fractions were pooled and concentrated using 30 kDa cut-off size membranes of Amicon-Ultra-15 (Millipore).

2.2. Gene cloning

Genes from C. cellulovorans (CcBglA; NCBI accession number AF268940), T. reesei (TrBgl2; AB003110) and N. koshunensis (NkBgl; AB073638) were synthesized and optimized to the favored codon usage for E. coli by the Geneart company (Regensburg, Germany). The CcBglA gene fragment was amplified by polymerase chain reaction (PCR) with forward 5′ GGAATTCCATATGGATGTTGCCTCCGGACAC 3′ (Ndel site underlined) and reverse 5′ CCAGCTCAGGTTGGTCGCTGCTCGATCAC 3′ (Xhol site underlined) primers.

The TrBgl2 gene fragment was amplified by PCR with forward 5′ GAAGATTCCATATGGATGTTGCCTCCGGACAC 3′ (Ndel site underlined) and reverse 5′ CCGAGGCTCCGGGCAATTACGTC 3′ (EcoRI site underlined) primers. The NkBgl gene fragment was amplified by PCR with forward 5′ GAAGATTCCATATGGATGTTGCCTCCGGACAC 3′ (Ndel site underlined) and reverse 5′ CGCGCTCGAGGTCTCTGAATCTCTCTGGATC 3′ (Xhol site underlined) primers. Each PCR product of the three genes was subcloned into the expression vector of pET-21a (Novagen) to generate pET-21-CcBglA, pET-21-TrBgl2 and pET-21-NkBgl. The N-terminal 20 residues of authentic NkBgl were predicted to be an N-terminal signal peptide and were obviated in the construction of pET-21-NkBgl. For easier purification, a His6 extension was added on the N-terminus of TrBgl2 and the C-termini of CcBglA and NkBgl. The inactive E193D mutant of NkBgl was generated by using the PCR-based QuickChange site-directed mutagenesis kit (Stratagene) using pET-21-NkBgl as the template.

2.3. Protein expression and purification

All DNA constructs were confirmed by nucleotide sequencing. The correct constructs were transformed into E. coli BL-21 (DE3) competent cells for protein expression. An overnight culture of a single transformant (100 ml) was used to inoculate six liters of fresh Luria–Bertani (LB) medium containing 100 μg/ml ampicillin, and was subsequently incubated at 30 °C with shaking until the medium reached an A600 nm of 0.8–1.0. The temperature was then reduced to 13 °C (for CcBglA and TrBgl2) or 16 °C (for NkBgl), and the bacteria were allowed to cool for one hour before adding the inducer, isopropyl β-thiogalactopyranoside (IPTG), to a final concentration of 0.2 mM (for CcBglA), 0.5 mM (for TrBgl2) or 0.4 mM (for NkBgl). After further growth for 16 h, the bacterial cells were harvested by centrifugation at 7000g and immediately resuspended in a lysis buffer containing 20 mM Tris–HCl, 400 mM NaCl, 10 mM imidazole, pH 7.5. The cells were then destroyed by using Cell Disruption System (Constant Systems Ltd.) and cell debris was removed by centrifugation at 17,000g. The cell-free extract was loaded onto a Ni2+-nitrilotriacetic acid column, which had been equilibrated with the lysis buffer. The column was washed using the same buffer, and the His6-tagged proteins were subsequently eluted by a linear gradient of 10–300 mM imidazole. After checking by SDS–PAGE, the purified His6-tagged CcBglA, TrBgl2 and NkBgl proteins were concentrated and changed to enzyme storage buffer (50 mM Tris–HCl, 100 mM NaCl, pH 8.0) using 30 kDa cut-off size membranes with Jumbosep (Pall) and Amicon-Ultra-15 (Millipore). As the Tris molecule was found to occupy each active site of CcBglA, TrBgl2 and NkBgl in our crystal structures and acted as an inhibitor, the Tris–HCl was subsequently replaced by Hepes buffer for enzyme purification and storage. The expression and purification of NkBgl E193D mutant was performed using the same protocol established for native NkBgl. About 70 mg/l of CcBglA, 240 mg/l of TrBgl2 and 80 mg/l of NkBgl were obtained, with a purity of greater than 98% as determined by SDS–PAGE.

2.4. Optimum pH and temperature determination of enzyme activity

The CcBglA, TrBgl2 and NkBgl were prepared and stored by using Hepes buffer in the production procedure to avoid Tris inhibition in enzyme assay. Enzyme activity of β-glucosidase toward salicin was measured according to the method of Miller (Miller, 1959) using glucose as the standard. A total 200 μl reaction mixture containing 0.5% salicin and enzyme (10 μl enzyme stored and diluted in 50 mM Hepes and 100 mM NaCl, pH 8.0) in 100 mM phosphate-citrate buffer under different pH conditions (3.0–7.0) was incubated at temperatures from 30 to 70 °C for 30 min. The 3,5-dinitrosalicylic acid reagent (200 μl) was added
and the solution was incubated at 95 °C for 5 min. Finally, 600 μl distilled water was added and the absorbance was measured at 540 nm by Lambda Bio 40 spectrometer (Perkin–Elmer). One unit of enzyme activity corresponds to one μmol of glucose per minute in the reaction. The commercial Novozyme 188 is a crude protein mixture produced by Aspergillus niger. In addition to β-glucosidase, it contains another major protein which appears to be α-amylase with a molecular weight of about 65 kDa. The size of the enzyme was determined by zymogram and protein identification using mass spectrometry. Protein concentrations were determined as described by Bradford (Bradford, 1976) using the Bio-Rad Protein Assay Kit with BSA as the standard. All experiments were performed in triplicate using a single prepaped protein.

2.5. Effect of chemical reagents and metal ions on β-glucosidase activity

The effect of chemicals and metal ions on each β-glucosidase activity was assayed in 100 mM phosphate–citrate buffer using 0.5% salicin as the substrate at its optimal pH and temperature for 30 min. Various chemicals, such as Tris, EDTA, dithiothreitol (DTT), and 2-mercaptoethanol (2-ME), and metal ions such as CaCl2, CoCl2, CuSO4, MgCl2, MnCl2, NaN3, and ZnSO4 were added to the reaction mixture to determine their effect at different concentrations. To avoid changing the pH values of the reaction mixture, the stock solution of Tris was adjusted to the optimal assay pH, i.e. pH 5.5 for Nkβgl and pH 6.0 for both CcβglA and Trβgl2. The relative activities were defined using the activity ratio of each β-glucosidase activity without chemicals added. All experiments were performed in triplicate using a single prepaped protein.

2.6. Enzyme kinetic determination

Two substrates, pNPG and resorufin-β-d-glucopyranoside, were used for kinetic analysis for CcβglA, Trβgl2 and Nkβgl. Both reactions were performed in 100 mM phosphate–citrate buffer in pH 5.5 or 6.0, depending on the optimal condition of each β-glucosidase. The former substrate was assayed at 40 or 45 °C using a temperature controller system (Jasco EHC-716), while the latter substrate was investigated at 25 °C. Reactions were carried out using 0.01–15 mM pNPG in a total volume of 2 ml. The reaction was initiated by adding 20 μl of CcβglA, Trβgl2 or Nkβgl to give a final concentration of 10–100 nM of enzyme. The released p-nitrophenol was monitored continuously at 410 nm by Jasco V-630 Bio spectrophotometer. Reactions were carried out using a final concentration of 0.1–150 μM resorufin-β-d-glucopyranoside in a total volume of 1 ml. The reaction was initiated by the addition of 10 μl of CcβglA, Trβgl2 or Nkβgl to give a final concentration of 0.1–4 nM of enzyme. The resorufin release was monitored continuously at excitation and emission wavelengths of 535 nm and 595 nm, respectively by Fluorolog-3 spectrofluorometer (Horiba). Km and kcat values were calculated by fitting the data to the Michaelis–Menten equation using GraphPad Prism 5.0.

2.7. Crystallization

All proteins were crystallized by the sitting-drop vapor diffusion method at 25 °C, with each sitting-drop prepared by mixing 2 μl each of protein solution and the reservoir solution. The Combiclover Junior sitting plate (Emerald BioSystems) and 250 μl each well of reservoir solution were used in the crystallization. The concentration of both CcβglA and Trβgl2 is 10 mg/ml in 50 mM Tris–HCl buffer (pH 8.0) containing 100 mM NaCl and 5 mM dithiothreitol, whereas Nkβgl is 25 mg/ml. The crystals of CcβglA were obtained with a reservoir solution containing 21–23% (w/v) PEG 3350 (J.T. Baker) and 0.3–0.45 M Li2SO4 (J.T. Baker) in 0.1 M Hepes (J.T. Baker) buffer at pH 7.5. The crystals of Trβgl2 were obtained with a reservoir solution containing 20–22% (w/v) PEG 3350 and 0.23 M MgCl2 (Fluka) and 12.5 mM sodium formate (Fluka) in 0.1 M Tris buffer at pH 8.0. The crystals of Nkβgl were obtained with a reservoir solution containing 18–21% (w/v) PEG 3350 and 0.1–0.25 M MgCl2 in 0.1 M Bis-Tris (Fluka) buffer at pH 6.5. Crystals of Nkβgl-E193D mutant were obtained at similar conditions to that of native Nkβgl though the Tris–HCl buffer was replaced by Hepes buffer.

2.8. Diffraction data collection and processing

All crystals were flash-cooled to 100 K in a stream of cold nitrogen prior to data collection. Prior to flash cooling, the crystals were soaked briefly in a cryoprotectant solution. The cryoprotectant solution for CcβglA crystals contains 38% (w/v) PEG 3350 and 0.45 M Li2SO4 in 0.1 M Hepes buffer at pH 7.5, while for Trβgl2 and Nkβgl crystals it contains 8% (w/v) glycerol, 23% (w/v) PEG 3350, 0.12 M NaCl and 0.25 M MgCl2. The data of Nkβgl-E193D pNPG complex was collected from the inactivate Nkβgl-E193D crystal after soaking for one minute with the cryoprotectant solution of Nkβgl containing saturated pNPG. X-ray diffraction data were collected at SPXF beamlines BL13B1 and BL13C1 at the National Synchrotron Radiation Research Center (NSRC) in Taiwan and beamline BLSA at the Photon Factory (PF) in Japan. The diffraction images were processed using the program package of HKL2000 (Otwinowski and Minor, 1997). The crystal data and data collection statistics of CcβglA, Trβgl2, Nkβgl and Nkβgl-E193D-pNPG complex are summarized in Table 1.

2.9. Structure determination and refinement

The structures of CcβglA, Trβgl2 and Nkβgl were solved by molecular replacement with CNS (Brünger et al., 1998) using the structures of the PDB entries of 1O0D (Thermotoga maritima β-glucosidase A), 2E3Z (Phanerochaete chrysosporium β-glucosidases 1A) and 1CBG (Trifolium repens cyanogenic β-glucosidase) as the search models, respectively. The structural models were built and refined using Xtalview (McRee, 1999), Coot (Emsley and Cowtan, 2004), CNS, and Refmac (Winn et al., 2003). The structure of Nkβgl-E193D-pNPG was solved by molecular replacement using CNS with the refined structure of Nkβgl as a search model. A subset of 5% randomly selected reflections was excluded from computational refinement to calculate the Rfree factor throughout the refinement (Brünger, 1993). All final refinements were carried out using Refmac with TLS group tensor and anisotropic B factor without NCS restraints. The CcβglA crystals contained four molecules in an asymmetric unit and the models were refined to a final Rwork of 14.6% and Rfree of 19.8%. The Trβgl2 crystals contained four molecules in an asymmetric unit and the models were refined to a final Rwork of 16.4% and Rfree of 21.7%. Both Nkβgl and Nkβgl-E193D-pNPG crystals contained one molecule in an asymmetric unit. The Nkβgl model was refined to a final Rwork of 11.5% and Rfree of 14.4%, while a Rwork of 12.2% and Rfree of 15.3% was obtained for Nkβgl-E193D-pNPG. The stereochemistry and structure of the final models were analyzed by RAMPAGE (Lovell et al., 2003) and SFCHECK (Vagin et al., 1999) of the CCP4 program suite (Bailey, 1994). The refinement statistics are summarized in Table 1. The molecular figures were produced by using PyMOL (DeLano Scientific, http://www.pymol.org).

2.10. PDB accession codes

The atomic coordinates and structure factors for β-glucosidases from C. cellulosovorans, T. reesei, and N. koshunensis have been deposited in the Protein Data Bank (PDB accession code: 3AHX, 3AHY, 3AHZ). The Nkβgl-E193D in complex with pNPG was deposited as 3AIO.
3. Results

3.1. Enzyme activities

The effect of pH and temperature on the activity of β-glucosidases toward salicin substrate was characterized. Both CcBglA and TrBgl2 were optimally active at pH 6, while NkBgl was optimal at pH 5.5 (Fig. 1). The optimum pH for Novozyme 188, a commercial β-glucosidase enzyme, is pH 4, at which condition our enzymes show less activity. However, at their optimum pH, TrBgl2, NkBgl, and CcBglA show at least 2-, 5-, and 10-fold increases in activity in comparison to the commercial enzyme Novozyme 188 in equal protein quantity by weight. The temperature for maximum activity was found to be 40°C for TrBgl2 and 45°C for CcBglA and NkBgl at 0.5% salicin in 100 mM phosphate–citrate buffer at pH 3.0–7.0 and 30–70°C for 30 min. One unit of β-glucosidase activity is defined as 1 μmol of glucose produced from salicin per minute. The functional protein fraction separated by Sephacryl S-200 size exclusion chromatography of Novozyme 188 was used in the assay. The CcBglA, TrBgl2 and NkBgl were prepared and stored by using Heps buffer to avoid the inhibited effect of Tris in the assay. All measurements were analyzed with triplicates with error bars indicated. Error bars represent the standard deviation of three replicates from a single prepped protein.

![Fig. 1.](image-url) Effect of pH and temperature on the activity of β-glucosidases. The optimal conditions of (A) pH and (B) temperature affected profiles of β-glucosidases were determined by salicin as substrate. The optimal temperature of CcBglA (45°C), TrBgl2 (40°C), NkBgl (40°C) and Novozyme 188 (60°C) were used to determine the pH effect in (A). The optimal pHe of CcBglA (pH 6.0), TrBgl2 (pH 6.0), NkBgl (pH 5.5) and Novozyme 188 (pH 4.0) were used to determine the temperature effect in (B). The enzyme activities were measured in 0.5% salicin in 100 mM phosphate–citrate buffer at pH 3.0–7.0 and 30–70°C for 30 min. One unit of β-glucosidase activity is defined as 1 μmol of glucose produced from salicin per minute. The functional protein fraction separated by Sephacryl S-200 size exclusion chromatography of Novozyme 188 was used in the assay. The CcBglA, TrBgl2 and NkBgl were prepared and stored by using Heps buffer to avoid the inhibited effect of Tris in the assay. All measurements were analyzed with triplicates with error bars indicated. Error bars represent the standard deviation of three replicates from a single prepped protein.
their optimum pH. At temperatures higher than the optimum, CcBglA activity declines drastically and shows similar activity to Novozyme 188 at 60 °C.

To determine the affinity of β-glucosidases for β-1,4 linked glucose derivatives, the kinetics properties with pNPG and resorufin-β-D-glucopyranoside were investigated. All three of these β-glucosidases show higher affinity toward the latter substrate, especially Novozyme 188 at 60°C/Cc

3.2. Effect of chemical reagents and metal ions

To investigate the effect of metal ions and reducing agents on enzyme activity, β-glucosidases were assayed in the presence of 2, 5, or 10 mM of metal ions or chemicals as shown in Fig. 2. Among the metal ions, Mn$^{2+}$ significantly enhanced the β-glucosidase activity by 150–230% in all three tested concentrations. As to the Cu$^{2+}$ ion, it notably reduced up to 90% of CcBglA activity at 10 mM Cu$^{2+}$, although the same concentration was tolerable for CcBglA where 60% activity remained. A higher concentration of Zn$^{2+}$ (10 mM) severely inhibited more than 90% of CcBglA and TrBgl2 activities; however, no obvious inhibition was observed in NkBgl. The chelating agent EDTA exerts 11–41% inhibition on enzyme activity, suggesting the presence of some trace elements in the original purified enzymes or reaction buffer in addition to those that we studied, i.e., divalent cations other than Ca$^{2+}$, Mg$^{2+}$, etc., which contributed to enzyme activation. In the presence of reducing agent, dithiothreitol had no effect on β-glucosidase activity; however, 2-mercaptoethanol resulted in 30–50% inhibition of β-glucosidase activity. With regard to the metals and chemicals used in this study, TrBgl2 is more affected by their presence than CcBglA and NkBgl.

In both the TrBgl2 and NkBgl crystal structures, a Tris molecule was observed in the active site. In order to gain an insight into the effect of Tris on β-glucosidases, different concentrations of Tris molecules were applied. At all concentrations tested, Tris inhibits the three β-glucosidases as shown in Fig. 2. Severe inhibition was detected both in TrBgl2 and NkBgl when the Tris concentration was higher than 20 mM.

3.3. Overall structures of CcBglA, TrBgl2 and NkBgl

The primary sequences of CcBglA, TrBgl2 and NkBgl, belonging to the glycoside hydrolase family 1, are composed of 445, 465 and 498 residues, respectively (Supplementary Fig. S1). The N-terminal signal peptide of 20 residues of NkBgl is obviated in the recombinant protein. The overall structures of CcBglA, TrBgl2 and NkBgl are similar to the PDB-deposited structures of glycosyl hydrolase family 1 enzymes, which have a classical (χ/β)α-TIM barrel fold (Fig. 3A–C). Amino acid sequence alignment of nine glycosyl hydrolase family 1 enzymes and the secondary structure of CcBglA, TrBgl2 and NkBgl are shown in Supplementary Fig. S1.

The crystal structure of CcBglA was determined at 1.9 Å. There are four protein molecules and two PEG fragments in an asymmetric unit. The structures of four chains of CcBglA with root-mean-square deviations (RMSDs) ranged from 0.23 to 0.30 Å over 441–444 Cx atoms. The two PEG fragments were located on the packing surface of chains A–C and B–D.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>p-Nitrophenyl-β-D-glucopyranoside</th>
<th>Resorufin-β-D-glucopyranoside</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Parameters</td>
<td></td>
</tr>
<tr>
<td>CcBglA</td>
<td>50.67 ± 1.00</td>
<td>0.15 ± 0.01</td>
</tr>
<tr>
<td>TrBgl2</td>
<td>6.91 ± 0.16</td>
<td>0.86 ± 0.07</td>
</tr>
<tr>
<td>NkBgl</td>
<td>1.39 ± 0.03</td>
<td>0.29 ± 0.02</td>
</tr>
</tbody>
</table>

Values are shown with standard deviations calculated from three replicates using a single prepped protein.
The crystal structure of TrBgl2 was determined at 1.63 Å and contained four protein and Tris molecules in the asymmetric unit. The structures of four chains of TrBgl2 with RMSDs ranged from 0.23 to 0.59 Å over 464–465 C\textsubscript{\alpha} atoms.

The crystal structures of NkBgl were determined at 1.34 Å and contained one protein, Tris, and glycerol molecule in the asymmetric unit. The crystal structure of NkBgl-E193D-\textit{pNPG} was determined at 1.40 Å and contained one protein, \textit{pNPG}, and glycerol molecule in the asymmetric unit. The RMSD was 0.15 Å over 472 C\textsubscript{\alpha} atoms of structures of NkBgl and NkBgl-E193D-\textit{pNPG}.

The accessible surface areas of the monomer to solvent of CcBglA, TrBgl2 and NkBgl are about 17,650, 17,000 and 17,300 Å\textsuperscript{2}, respectively. Although four protein molecules of CcBglA and TrBgl2 were present in the asymmetric unit, they did not reveal any specific interactions that could result in the formation of stable quaternary structures by the Protein Interfaces, Surfaces, and Assemblies (PISA) server (Krissinel and Henrick, 2007). Interestingly, one protein molecule of NkBgl was present in the asymmetric unit, but two monomers interacted to form a dimer through crystallographic packing. However, NkBgl exists as a monomeric form in solution as determined by gel filtration chromatography (data not shown). The results suggest that CcBglA, TrBgl2 and NkBgl likely function as a monomer.

The active sites of CcBglA, TrBgl2 and NkBgl formed a 15–20 Å deep slot-like cleft and were located on connecting loops at the C-terminal end of the \textit{\beta}-sheets of the TIM barrel. The bottom of each active site is surrounded by negatively charged residues (Supplementary Fig. S2A–C). The catalytic acid/base of CcBglA, TrBgl2 and NkBgl are Glu166, Glu165 and Glu193, respectively, located on the TXNEP motif at the end of \textit{\beta}-strand 4 (where X is a hydrophobic amino acid residue). The catalytic nucleophile of CcBglA is Glu352, located on the I/VTENG motif at the end of \textit{\beta}-strand 7;

---

**Fig. 3.** Comparison of the structures of \textit{\beta}-glucosidases from bacteria, fungus and termite. Ribbon representation of the structures of (A) CcBglA, (B) TrBgl2, (C) NkBgl and (D) stereoview of the three superimposed \textit{\beta}-glucosidases. The overall structures of \textit{\beta}-glucosidases adopt the typical TIM barrel fold. The highly superimposed helices are shown in red, whereas strands in yellow, loops in gray. Variant regions of CcBglA, TrBgl2 and NkBgl different from the three superimposed structures are shown in green, magenta and cyan, respectively. The two catalytic glutamates of NkBgl and the Tris molecule located in the active site of NkBgl are shown as stick and ball models. Molecular surface representation of (E) substrate-free CcBglA, (F) TrBgl2 and (G) NkBgl in complex with Tris molecules. The view and color painting in (E–G) are the same as in (D). The active sites of (E–G) are indicated by the yellow circles. For displayed glutamates and Tris molecules, oxygen atoms are in red, nitrogen atoms are in blue, and carbon atoms are in green (for glutamates) or orange (for Tris).
whereas for \( \text{TrBgl2} \) it is Glu367, and for \( \text{NkBgl} \) it is Glu402 (Supplementary Fig. S1) (Wang et al., 1995). Two cis-peptide bonds were found at Ala181-Pro182 and Trp399-Ser400 of \( \text{CcBglA} \), at Ala180-Pro181 and Trp417-Ser418 of \( \text{TrBgl2} \), at Ala208-Pro209 and Trp444-Ser445 of \( \text{NkBgl} \). The presence of such bonds is typical for the glycosyl hydrolase family 1 (Seshadri et al., 2009). In addition, one more cis–peptide bond was found between Glu350 and Pro351 of \( \text{NkBgl} \).

3.4. Comparison of structures of \( \text{CcBglA}, \text{TrBgl2} \) and \( \text{NkBgl} \)

The one chain crystal structure of \( \text{NkBgl} \) was compared with the four chain structures of \( \text{CcBglA} \) and \( \text{TrBgl2} \) with RMSDs ranging from 1.27 to 1.32 Å over 441–444 Cα atoms, and 1.15–1.21 Å over 465–466 Cα atoms, respectively. A comparison of the four chain crystal structures of \( \text{CcBglA} \) and \( \text{TrBgl2} \) produced RMSDs ranging from 1.17 to 1.34 Å over 441–444 Cα atoms. The ribbon represen-

Fig. 4. The active sites of \( \beta \)-glucosidases from bacteria, fungus and termite. Stereoview of the surrounding residues of the active site of (A) substrate-free \( \text{CcBglA} \), (B) \( \text{TrBgl2} \) and (C) \( \text{NkBgl} \) in complex with Tris molecules. \( |F_o| - |F_c| \) difference Fourier maps for the Tris molecules contoured at 3.0σ are shown in cyan. 2|F_o| – |F_c| maps for the side chains of selective residues are shown in gray, whereas water molecules contoured at 1.0σ are in yellow-orange. The side chains of residues around the active site of the \( \beta \)-glucosidases are shown as stick models. The Tris molecules are shown as ball-and-stick models. The glutamates labeled in magenta are the catalytic nucleophile and acid–base residues of these \( \beta \)-glucosidases. Significant variance of the side chain of Trp325 of \( \text{CcBglA} \) was observed. The hydrogen-bond interactions of the residues and Tris molecules are labeled in magenta dotted lines. The carbon atoms of proteins and Tris molecules are shown in green and yellow, respectively. The oxygen atoms are shown in red, nitrogen in blue and sulfur in gold.
tation of superimposed structures of CcBglA, TrBgl2 and NkBgl is shown in Fig. 3D. The Cα traces of superimposed structures of CcBglA, TrBgl2 and NkBgl are shown in Supplementary Fig. S3. The structures of CcBglA, TrBgl2 and NkBgl share many similarities. The core structures of the TIM barrel forming at the bottom of its active site are especially highly conserved (Fig. 3E–G and Supplementary Fig. S2D–F) where the glycone (subsite – 1) and aglycone (subsite + 1) binding sites are located (Supplementary Fig. S1). However, with the exception of the right lobe, numerous variances do exist in the structures of CcBglA, TrBgl2 and NkBgl in the outer opening regions around their active sites (Fig. 3E–G).

3.5. The active site of TrBgl2 and NkBgl is occupied by a Tris molecule

In the structures of TrBgl2 and NkBgl a Tris molecule was found to bind at each active site. The electron density of the difference Fourier map, calculated including all refined atoms except the Tris, clearly defined all atoms of the Tris molecules (Fig. 4B–C). Although no Tris molecule was observed in the active site of CcBglA, some small, irregular and unidentified extra electron density was observed. The inhibition of Tris on the hydrolase activity of CcBglA, TrBgl2 and NkBgl were in concert with that observed in the crystal structures. Among residues around the active sites of CcBglA, TrBgl2 and NkBgl, only the side chain of Trp325 of CcBglA shows a significant variance (Fig. 4A). The Trp339 of TrBgl2 and Trp374 of NkBgl have similar chi-1 (~58°) and chi-2 (~91°) angles; however, the Trp325 of CcBglA has different chi-1 (~56°) and chi-2 (~22°) angles. The location and orientation of Tris molecules and side chains of the three key glutamates (Glu165, Glu367 and Glu424) are highly superimposed in each chain of TrBgl2 (Supplementary Fig. S4A). Although the side chains of three key glutamates (Glu165/193, Glu367/402 and Glu424/451) of TrBgl2 and NkBgl were located at similar positions, the Tris molecules of TrBgl2 and NkBgl were bound in different orientations (Supplementary Fig. S4B). A total of three residues of TrBgl2 formed five direct hydrogen bonds with Tris, and five residues of NkBgl formed seven hydrogen bonds (Supplementary Fig. S5).

3.6. Analysis of the crystal structure of NkBgl-E193D in complex with the pNPG substrate

The structure of NkBgl-E193D in complex with the pNPG substrate (Fig. 5B and Supplementary Fig. S4C) was determined at 1.40 Å resolution from crystals that were soaked with the cryoprotectant solution containing saturated pNPG. The pNPG is a chromogenic substrate for β-glucosidases and has the Michaelis–Menten constant (k_m) of 0.29 mM for NkBgl in 100 mM phosphate-citrate buffer with pH 5.5 and at 40 °C (Table 2). The electron density of the Fourier difference map, calculated by including all refined atoms except pNPG, clearly defined all atoms of the pNPG molecule (Supplementary Fig. S4C). Comparing the structures of Tris bound NkBgl with pNPG bound NkBgl-E193D, significant difference in these structures were observed at the side chains of the residues of Glu/Asp193, Asn253, Trp374, Trp444, Asn449 and Glu451. The residues of Asn253 and Trp374 were located within the aglycone binding pocket and the residues of Trp444, Asn449 and Glu451 were located within the aglycone binding pocket. In addition, the catalytic acid/base of NkBgl was mutated from Glu193 to the smaller Asp193 which creates room in the substrate binding pocket and a subsequent rotation of the Asn253 side chain toward the substrate binding pocket was observed (Fig. 5). Furthermore, it seems that the β-carboxyl group of Asp193 possesses the ability to attack the O1 atom of pNPG through a water molecule (Supplementary Figs. S4C and S6).

Fig. 5. Comparison of the active site in NkBgl complexes. Stereoview of the superimposed surrounding residues of the active site of (A) NkBgl-Tris and (B) NkBgl-E193D-pNPG complexes. The side chains of residues around the active site of NkBgl are shown as line models. Carbon atoms of residues are shown in green for NkBgl-Tris and orange for NkBgl-E193D-pNPG. The Tris and pNPG molecules are shown as ball-and stick models (carbon atoms in gray). Oxygen atoms are shown in red and nitrogen in blue. Cyan dotted lines indicate the hydrogen-bond interactions between residues and Tris/pNPG molecules.
The residues Gln45, His148, Asn255, Glu402, Trp444, Glu451 and Trp452 of NkBgl-E193D formed direct hydrogen bonds with pNPG (Fig. 5B and Supplementary Fig. S6). A total of seven residues of NkBgl formed nine direct hydrogen bonds with pNPG. Residues Asp193, Asn253, Asn255, Asn335, Tyr337, Trp374 and Glu458 of NkBgl-E193D formed water-mediated hydrogen bonds with pNPG and the residues Trp149, Asn253, Tyr337, Trp374, Trp444 and Phe460 of NkBgl-E193D interacted with pNPG through hydrophobic contacts (Supplementary Fig. S6). As mentioned above, the residues Gln45, His148, Trp149, Asn335, Trp452, Glu458 and Phe460 were located within the glycone binding pocket and the residues of Asn253, Asn255, Tyr337 and Trp374 were located within the aglycone binding pocket (Fig. 6A and Supplementary Fig. S1). Fig. 6B represents a retaining β-glucosidase mechanism in accordance with the structure of the NkBgl-E193D-pNPG complex. Furthermore, the Glu193 residue of NkBgl, the catalytic acid/base, was mutated into Asp193 to reduce enzyme activity in order to investigate the enzyme structure in complex with pNPG. Notably, the binding mode of the hydroxyl groups with Tris was similar to that of the glucoside geometry in the structure of the NkBgl-E193D-pNPG complex.

4. Discussion

C. cellulovarans, one of the most efficient cellulolytic organisms, produces an extracellular non-cellulosomal β-glucosidase A to degrade small oligosaccharides cooperatively with cellulosomes. The C-terminal His6-tagged recombinant CcBglA has been expressed in E. coli BL-21 (DE3) by Kosugi et al. (2006). The recombinant CcBglA in our study shows a higher affinity for the substrates...
pNPG ($K_m = 0.15 \text{ mM}$) and resorufin-β-β-glucopyranoside ($K_m = 0.0087 \text{ mM}$) than the cellobiose ($K_m = 2.9 \text{ mM}$) previously studied by Kosugi et al. (2006).

Some Trichoderma species exhibit strong cellulose-degrading activities. The TrBgl2 gene has been previously isolated from assexual fungus *Trichoderma reesei* (teleomorph *Hypocrea jecorina*), and the recombinant TrBgl2 has been produced by *Aspergillus oryzae* (Takashima et al., 1999). Using salicin as the substrate instead of the pNPG, the recombinant TrBgl2 demonstrates the same pH and temperature optima during catalysis with the fungus-expressed recombinant TrBgl2. As for the kinetic parameters, TrBgl2 expressed in *E. coli* shows approximately 2.5-fold higher affinity to pNPG than the one expressed in *A. oryzae*. Interestingly, there is another β-glucosidase II previously purified from *T. reesei* by Chen et al. (1992), which optimally catalyzes the degradation of pNPG in more acidic and thermophilic conditions (pH 4, 60 °C) and shows 100-fold higher catalytic efficiency. However, there is no structural information about this extra long β-glucosidase II, which has an amino acid sequence almost twice the length of TrBgl2.

With regard to β-glucosidasises in insects, since β-glycans are key dietary sources for some insects, the substrate specificity of insect β-glucosidases seems to be broad. The recombinant NkBgl we studied catalyzed the degradation of pNPG more efficiently than β-glucosidase from *Spodoptera frugiperda* larval midgut (Marana et al., 2001). The NkBgl has not only a higher affinity toward the substrate but a 2.5-times higher turnover number. Compared with the native NkBgl purified from *N. koshunensis* salivary glands, our recombinant NkBgl exhibits similar pH and temperature optima using salicin as the substrate, and shows a higher affinity toward pNPG ($K_m = 0.29 \text{ mM}$) and resorufin-β-β-glucopyranoside ($K_m = 0.00116 \text{ mM}$) than the cellobiose ($K_m = 2.5 \text{ mM}$) previously studied by Tokuda et al. (2002). To our knowledge, NkBgl is the first insect β-glucosidase structure to have been determined.

Comparing the primary sequences and tertiary structures of CcBglA, TrBgl2 and NkBgl with the structures deposited in PDB by Blast ([http://blast.ncbi.nlm.nih.gov/]) and PDBFold (also known as SSM) (Krissinel and Henrick, 2004), it is no surprise to find that the structure most similar to CcBglA is β-glucosidase A from a bacterial source – *Thermotoga maritima*, which shows 49% identity, 73% similarity, and an RMSD of 0.97 Å over 430 Ca atoms (Zechel et al., 2003). The TrBgl2 demonstrates most similarity with a fungal β-glucosidase structure from *Phanerochaete chrysosporium*, which has 54% identity, 73% similarity and an RMSD of 0.90 Å over 451 Ca atoms for (Nijikken et al., 2007). The insect NkBgl shows 45% identity, 66% similarity, and 0.90 Å over 446 Ca atoms for aphid *Brevicoryne brassicae* myrosinase (Husebye et al., 2005). *Aphid B. brassicae* myrosinase (EC 3.2.1.147), hydrolyzing glucosinolates into isothiocyanates, showed considerable similarity in gene sequence to glycosyl hydrolase family 1 members (Husebye et al., 2005). According to the protein sequence alignment and the secondary structure analysis, the linker regions between α4—β5 and β6—β7 of bacterial β-glucosidases are shorter than those of the fungal and insect proteins (Supplementary Fig. S1). Compared to bacterial and fungal β-glucosidases, both NkBgl and myrosinase in insects show shorter linker regions between αF—αG and longer ones for β7—α7 (Supplementary Fig. S1). All the above regions are located on the surfaces of β-glucosidases. These distal regions seem not to affect the active site directly; however, we cannot obviate their indirect effects on enzyme properties among proteins from different species.

The activity of the three β-glucosidases was inhibited by most cations we examined; however, Mn$^{2+}$ significantly enhanced their enzyme activity by at least 50%, which implies that this cation may be involved in enzyme function. In previous studies, strong activation by Mn$^{2+}$ was observed in two ascomycetes strains; 65% toward pNPG in *Hanseniaspora vineae* (Vasserot et al., 1989) and 77% in *Aspergillus oryzae* (Riou et al., 1998). Nevertheless, our study demonstrates the highest stimulation, a 226% increase in relative activity, in TrBgl2-catalyzed degradation of salicin in the presence of 10 mM MnCl$_2$. These results suggest that Mn$^{2+}$ ions might assist the attacking reaction of the catalytic residues (Dismukes, 1996; Riou et al., 1998). However, as the Mn$^{2+}$-coordinated β-glucosidase structure is yet to have been solved, it is hard to clarify the real role of Mn$^{2+}$ in enzyme function.

In the active sites of the three β-glucosidases, Cys168 of TrBgl2 is supposed to be more reactive with metal ions than CcBglA Val169 and NkBgl Thr196, whose residues are located at a similar position as shown in Fig. 4. We speculated that the thiol-group of Cys168 of TrBgl2 may play an important role in the inhibition of heavy metal ions. Uzun et al. (2008) have reported the following metal ion affinity order of a cysteine functionalized poly(4-hydroxyethyl methacrylate) monolith: Cd$^{2+}$ > Cu$^{2+}$ > Hg$^{2+}$ > Pb$^{2+}$ > Zn$^{2+}$ > -Co$^{2+}$. This is consistent with our findings that CuSO$_4$ and ZnSO$_4$ caused a drastic inhibition of TrBgl2, while CoCl$_2$ did not (Fig. 2B). Additionally, 10 mM ZnSO$_4$ caused severe inhibition of CcBglA but not in NkBgl (Fig. 2). It is supposed that Zn$^{2+}$ ions at high concentration may coordinate to two His ligands (His121 and His180) in the active site of CcBglA (Fig. 4) (Tainer et al., 1992). This metal attaching, Cu$^{2+}$ or Zn$^{2+}$ with Cys in TrBgl2 as well as Zn$^{2+}$ with His in CcBglA, may hinder the substrate binding and inhibit catalytic reaction in the active site.

In previous reports, Tris served as an inhibitor to reduce the hydrolase activity of β-glucosidases (Ait et al., 1982; Ferreira and Terra, 1983; Park et al., 2001). Here, the crystal structures of TrBgl2 and NkBgl in complex with the Tris, and NkBgl-E193D in complex with the pNPG, show the Tris molecule interacting directly with the residues on the glycone-binding site (Fig. 6A and Supplementary Figs. S5 and 6). Furthermore, some chemical reagents containing the poly-hydroxyl group, such as Tris, glycerol and tartaric acid, were also observed to bind to the active site of glycosyl hydrolase family 1 β-glucosidases (Nam et al., 2008; Sanz-Aparicio et al., 1998; Seshadri et al., 2009). Although there is no obvious interaction between Tris molecules and the Trp339 of TrBgl2 or Trp374 of NkBgl as shown in the crystal structures, it is supposed that the Tris tolerance of CcBglA is relevant to the different orientation of the side chain of Trp325 in CcBglA (Figs. 2 and 4).

In the conversion processes of lignocellulose to bioethanol, SSF and SSCF offer promising strategies for reducing the product inhibition of glucose. In both SSF and SSCF processes, the hydrolysis of cellulose by different kinds of cellulases proceeds simultaneously with the microbial fermentation to produce ethanol. However, the optimum conditions for cellulase catalysis and microbial fermentation are usually not in concert. Reaction conditions at a near neutral pH and moderate temperature are favorable for microbial growth and fermentation in SSF and SSCF processes. In our study, the bacterial CcBglA exhibits the highest activity and shows the most tolerance to various metal ions. Synergism between CcBglA and Celluclasta 1.5L (with endoglucanase and exoglucanase) was also observed on the saccharification of Avicel (microcrystalline cellulose) or acid-pretreated rice straw (data not shown). Furthermore, its optimal reaction parameters demonstrate that CcBglA should be a good candidate for utilization in the SSF and SSCF processes (Lynd et al., 2002). In order to further the industrial application of lignocelluliosics ethanol production, we have employed a fermentor to produce active enzymes and increase production yield. Further research into enzyme yield improvement is ongoing.

**Acknowledgments**

Portions of this research were carried out at the National Synchrotron Radiation Research Center (NSRRC), a national user facil-
ity supported by the National Science Council, Taiwan. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine. We thank NSRRC of Taiwan, SPring-8 and the Photon Factory of Japan for beam time allocations. This work was supported by Academia Sinica and the National Science Council of Taiwan (Grant Numbers NSC 96-3112-B-001-014, NSC 96-3114-P-004-001, NSC 97-3112-B-001-017, and NSC 97-3114-P-001-001 to A.H.J.W.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2010.07.008.

References


