博士論文

A study on β-glucosidases derived from wood-feeding insect and symbiotic protist of termite

（材食性昆虫およびシロアリ腸内原生生物に由来するβ-グルコシダーゼに関する研究）

李 宜海

博士論文

A study on β-glucosidases derived from wood-feeding insect and symbiotic protist of termite

（材食性昆虫およびシロアリ腸内原生生物に由来するβ-グルコシダーゼに関する研究）

応用生命工学専攻
平成 24 年度博士課程入学
氏　名　　李　宜海
指導教員名　　有岡　学

A study on β-glucosidases derived from wood-feeding insect and symbiotic protist of termite

（材食性昆虫およびシロアリ腸内原生生物に由来するβ-グルコシダーゼに関する研究）

by

LI Yihai

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Ph. D.

Department of Biotechnology
Graduate School of Agricultural and Life Sciences
The University of Tokyo

Abbreviations

AOX	alcohol oxidase promoter
AEC	anion exchange chromatography
AS	ammonium sulfate precipitate
Avi	Avicel
B. R.	Britton-Robinson buffer
BG	β-glucosidase
^{o}C	degree Celsius
CAI	codon adaptation index
CBB	coomassie brilliant blue
CBH	cellobiohydrolase
cDNA	complementary cDNA
CMC	carboxymethyl cellulose
c-myc	c-myc epitope (protein tag for Western blot analysis)
CS	culture supernatant
Ctrl	control
Da	Dalton
DEAE	diethylaminoethyl
DMSO	dimethyl sulfoxide
DP	degree of polymer
DTT	dithiothreitol
EAEA repeat	polypeptide recognized by ste13
Elu	elution (fraction)
EDTA	ethylenediaminetetraacetic acid
EG	endoglucanase
Endo H	endoglycosidase H
FT	flow-through
Gal	galactose
GAP	3-phosphate dehydrogenase promoter
GH	glycosyl hydrolase
glaA	glucoamylase A promoter
Gen	gentiobiose
Glc	glucose
GlcNAc	*N*-acetylglucosamine
His_6	hexahistidine tag

IMAC	immobilized metal affinity chromatography
imi	imidzaole
k_{cat}	catalytic constant
Kex2	Kex2 endoprotease
k_{cat}/K_m	catalytic efficiency
Ki	inhibition constant
K_m	Michaelis constant
KR	dibasic amino acid pair recognized by Kex2
Lac	lactose
Lam	larminarinbiose
Lar	Larminarin
Mal	Maltose
Mix	mixture of culture supernatant and buffer
Mut^s	methanol utilization slow
Ni-NTA	Nickel-nitrilotriacetic acid
OD	optical density
ORF	open reading frame
PB	sodium phosphate buffer
PCR	polymerase chain reaction
*p*NP	*p*-nitrophenol
*p*NPLac	*p*-nitrophenyl-β-D-lactopyranosidase
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sop	sophorose
Spec. at.	Specific activity
Ste13	Ste13 dipeptidyl aminopeptidase
Suc	sucrose
TLC	thin-layer chromatography
Tris	hydroxymethyl aminomethane
UV	ultra violet
W	wash fraction
Xln	xylan

General introduction and objectives

0.1 Lignocellulosic biomass for bioethanol production

Lignocellulosic biomass is the most abundant biomass on earth and is estimated to account for approximately 50% of the biomass in the world, with an annual yield of 10 billion tons (Claassen et al., 1999). Lignocellulosic biomass, including agriculture and forestry residues, waste papers, etc., is mainly comprised of three polymers: cellulose (35–50%), hemicellulose (25–30%), and lignin (15–30%) (Wyman, 1994). Cellulose, as the major fraction of lignocellulose, is a linear polymer of glucose linked by β-1,4-glycosidic bonds with chain length ranging from 2,000 to 20,000 glucose residues (Delmer and Amor, 1995). The chains are arranged in parallel and form highly ordered, crystalline domains interspersed by more disordered amorphous regions (Béguin and Aubert, 1994). In nature, cellulose is synthesized in a form of microfibril, a composite of many chains ranging from ~36 to more than 200 (Delmer and Amor, 1995). In the secondary cell walls of plants, cellulose microfibrils usually form sheets and embedded in a matrix of hemicellulose and lignin (Delmer and Amor, 1995). The interchain hydrogen bonding between adjacent chains in a cellulose sheet is rather strong and makes crystalline cellulose highly resistant to chemical and biological hydrolysis, and such resistance of plant cell walls towards microbial and enzymatic deconstruction is referred to as ‘biomass recalcitrance’ (Himmel et al., 2007). A scheme of the structure of cellulose in the plant cell walls is shown in Fig. 0-1A.

Ethanol could be used as a primary fuel in a neat form or as a gasoline blend (Lynd et al., 1991). According to the definition of the International Energy Angency (IEA), bioethanol refers to ethanol produced from sugar (including plants or cereal crops used as gasoline substitute; Mabee and Saddler, 2008). Bioethanol has been employed as a fuel or gasoline blend in USA, Brazil, China, and other countries (Sanchez et al., 2008).

According to the statistics of the U.S. Energy Information Administration (EIA), the volumetric share of ethanol was up to 9.8% of total U.S. motor gasoline supply in 2014, which accounts for 14.3 billion gallons of ethanol mainly from corn (http://www.eia.gov/todayinenergy/detail.cfm?id=21212). In Europe, ethanol is also deemed as the most cost-effective and readily available means of substantial decarbonization in transport sectors by the European Commission (Reboredo et al., 2016). Bioethanol could be produced from either energy crops such as sugar and starch crops, or lignocellulosic biomass (Wyman, 1999). However, the production of bioethanol is restricted by the supply of energy crops, and competes with demands for food and feed (Wyman, 1999).

Bioethanol produced from lignocellulosic feedstock, which is referred to as the second-generation bioethanol (Mabee and Saddler, 2008), is considered to be a promising renewable energy with advantages of reducing more greenhouse gas emissions than food-based ethanol (Farrell et al., 2006), low-cost (Hill et al., 2006), not occupying extra farmland, and decreasing the reliance on fossil fuel (Peplow et al., 2014). The development of cellulosic ethanol has been drawn intensive research attention worldwide. Facilities for the production of the second generation biofuel have been in construction in the U.S. (Peplow et al., 2014; Service, 2014).

Cellulose can be hydrolyzed into soluble sugar by acid treatment or enzymes. However, acid process has drawbacks such as product degradation, corrosion of equipment, and impurities in the syrups, etc. (Ryu and Mandels, 1980). Although there is a variety of process for hydrolysis of cellulosic biomass, current process adopted in the production of bioethanol from lignocellulosic biomass is mainly based on enzymatic hydrolysis of cellulose (Wyman, 1999). Generally, processing of bioethanol (and

chemicals) from lignocellulosic biomass includes pretreatment, enzymatic hydrolysis, and fermentation (or catalytic upgrading) steps (Payne et al., 2015; Fig. 0-1B). In the production of bioethanol, the conversion of biomass into simple sugars is the key bottleneck and the development of new biotechnological approaches to promote conversion efficiency is needed (Lynd et al., 2008). The barriers are partially from the high cost of cellulases. Despite the cost reduction of about 20-folds had been achieved in recent years, the cost of cellulases is still five- to ten-folds of that of amylase which converts corn starch to glucose for fermentation (Schubert, 2006). According to the U.S. National Renewable Energy Laboratory (NREL), cellulases comprise up to 20% of the total ethanol production costs (Lambertz et al., 2014). Therefore, exploring new enzyme paradigms or improve the performance of cellulases are crucial to promote production efficiency to a cost-competitive level (Himmel et al., 2007).

0.2 Hydrolysis of cellulose by cellulases

More than a half century ago, Reese et al. firstly proposed a multienzyme system of hydrolysis of cellulose (Reese et al., 1950). The repertoire of cellulose digestion discovered in nature hitherto is comprised of two primary paradigms (Payne et al., 2015): (1) the 'free' enzyme paradigm, in which cellulose digestion is achieved by sequential actions of synergistic cellulases, is represented by the enzyme suite from the filamentous fungus *Trichoderma reesei* (anamorph of the ascomycete *Hypocrea jecorina*), and (2) the cellulosomal paradigm, in which cellulose is degraded by complementary enzymes recruited to a common protein scaffold, is represented by the macromolecular assemblies called cellulosomes in the anaerobic rumen bacterium, *Clostridium thermocellum*. Aside for the two well-characterized paradigms, a novel

mechanism which employs an intermediate strategy of secreting many free cellulases containing multiple catalytic domains has been pushed forward (Brunecky et al., 2013).

Currently, commercial enzyme products for biomass degradation are derived from fungi, as they secrete cellulases into the growth medium with high catalytic efficiency and at high yields (Merino and Cherry, 2007). *T. reesei* is an outstanding cellulolytic microorganism capable of secreting a complete set of extracellular cellulases which synergistically break down crystalline cellulose into soluble sugar (Ryu and Mandels, 1980). In the well-studied cellulolytic paradigm of *T. reesei*, endoglucanase (EG; endo-1,4-β-D-glucan glucanohydrolase; EC 3.2.1.4) firstly hydrolyzes the cellulose chains in its amorphous regions to generate free ends, then cellobiohydrolase (CBH; exo-1,4-β-D-glucan cellobiohydrolase; EC 3.2.1.91) attaches to the cellulose chain ends and releases cellobiose from both the reducing and non-reducing ends of cellulose chains, and lastly, β-glucosidase (BG, β-D-glucopyranoside glucohydrolases; β-D-glucoside glucohydrolase; EC 3.2.1.21) completes the final step by cleaving cellobiose into glucose (Henrissat et al., 1985; Béguin and Aubert, 1994; Fig. 0-2). The cleavage of cellulose chains by EGs and CBHs is assumed to be non-processive and processive, respectively (Divne et al., 1994; Teeri, 1997).

Recently, lytic polysaccharide monooxygenases (LPMOs; 'PMO' in Fig. 0-2) derived from bacteria and fungi were discovered to be capable of breaking down chitin (a crystalline analog of cellulose) and the crystalline region of cellulose (Vaaje-Kolstad et al., 2010; Horn et al, 2012). The action of LPMOs can boost the yield of cellulose degradation (Cannella et al., 2012) by generating entry points for glycoside hydrolases such as CBHs and EGs (Fushinobu, 2014). A simplified scheme of the current understanding on the enzymatic degradation of cellulose by the 'free' cellulases

paradigm is shown in Fig. 0-2.

0.3 Substrate specificity and functions of BGs

BG is an enzyme that hydrolyzes terminal, non-reducing β-D-glucosyl residue with the release of β-D-glucose (ExplorEnz; http://www.enzyme-database.org/query.php?ec=3.2.1.21). The activities of BGs are not limited to the β-1,4 glucosidic linkage, but also apply to β-1,2, β-1,3, and β-1,6 linkages (Han and Srinivasan, 1969; Sano et al., 1975). In some cases BGs were even characterized to have dozens or hundreds folds higher activities on laminaribiose (β-1,3 glucosidic linkage) than on cellobiose (Zverlov et al., 1997; Decker et al., 2001). It was proposed that fungus-derived BGs whose preferred substrate is β-1,3-glucosidic linkage over β-1,4-glucosidic linkage might be glucan 1,3-β-glucosidases (EC 3.2.1.58) rather than BGs, and play a role in fungal cell wall metabolism such as reutilization of the cell wall components (Igarashi et al., 2003; Tsukada et al., 2006). Furthermore, many of BGs are reported to be associated with the activities of β-D-galactosidase (Dahlqvist, 1961), α-L-arabinosidase, β-D-xylosidase, and β-D-fucosidase (Chinchetru et al., 1989). In some cases, glycosidases characterized as BGs were also found to possess much higher activities on β-D-fucoside (Peralta et al., 1990) and/or β-D-galactoside (Walker and Axelrod, 1978; Souza et al., 2010) than on β-D-glucoside. Recently, BGs displaying *N*-acetyl-β-D-glucosaminidase activity were also reported (Koffi et al., 2012; Ferrara et al., 2014), which broadens the substrate spectrum of BGs.

BG is ubiquitously distributed in many taxa, including bacteria (Han and Srinivasan, 1969), fungi (Crook and Stone, 1957), yeasts (Fleming and Duerksen, 1967), insects (Terra and Ferreira, 1996), plants (Heyworth and Walker, 1962), and animals

(McMahon et al., 1997), and plays pivotal roles in many crucial biological pathways owing to its hydrolytic activity and transferase activity (Bhatia et al., 2002). Aside for the hydrolase role of BG in the hydrolysis of glucosides, BGs were frequently reported to have transglycosylation activities (Crook and Stone, 1953; Umezurike, 1975), and this property in fungal BGs might associate with the induction of cellulases. For example, it was reported that a cell-bound β-glucosidase of *T. reesei* transglycosylated cellobiose to form sophorose, an inducer of cellulase expression (Vaheri et al., 1979; Kubicek, 1987), and the intracellular BG of *Penicillium purpurogenum* converts cellobiose to gentiobiose to induce the synthesis of cellulases (Kurasawa et al., 1992). Secondly, BGs are found to be implicated in hormone metabolism (Hösel and Barz, 1975; Falk and Rask, 1995) and cyanogenesis of plants (Zhou et al., 2002) in relation to the synthesis of irritating compound in the defensive secretions of insects and arthropods (Sillam-Dussès et al., 2012), and take roles in the production process of egg-recognition pheromone in several species of termites and wood-feeding cockroaches (Matsuura et al., 2009; Shimada and Maekawa, 2014).

0.4 Classification, structure, and catalytic mechanisms of BGs

Based on the homology of amino acid sequences, BGs are classified into the glycoside hydrolase (GH) families 1, 3, 5, 9, 30, and 116 in the Carbohydrate-Active enZYmes database (CAZy; http://www.cazy.org/) (Henrissat, 1991; Lombard et al., 2014). BGs constitute one of the major groups among GHs, and are mostly affiliated with GH1 and GH3 (Bhatia et al., 2002). With the exception of one putative endogenous GH3 BG identified from the salivary gland EST library of *Hodotermopsis sjostedti* (Yuki et al., 2008), all endogenous BGs of lower termites belong to GH1, whilst BGs

expressed by the symbiotic protists in the lower termites are primarily affiliated with GH3 (Ni and Tokuda, 2013).

As the folding of proteins is better conserved than their sequences, some of the families are also grouped in ’clans’ (Henrissat and Bairoch, 1996). The GH1 BGs fall in Clan A (Table 0-1), which is consisted of enzymes with a common $(\beta/\alpha)_8$-barrel architecture and the two catalytic glutamates located at the C-terminal end of β-strands 4 and 7 (Jenkins et al., 1995; Henrissat and Bairoch, 1996). On the other hand, all of the available three-dimensional structures of GH3 BGs to date have multidomain architectures, with the active center at the domain interface and catalytic residues derived from different domains (Agirre et al., 2016), and half of the structurally-resolved GH3 BGs are oligomers (detailed explanation is made in Table 3-2 in Chapter 3).

The overall topologies of the active sites of glycoside hydrolases fall into only three general classes: pocket or crater, cleft or groove, and tunnel (Davies and Henrissat, 1995). The active site of BGs presents a topology of pocket or crater, which is optimal for the recognition of a saccharide non-reducing end but with low efficiency for fibrous substrates lacking free chain ends such as native cellulose (Davies and Henrissat, 1995). The shapes of pockets of BGs are like a funnel (Hrmova et al., 1998) or coin slot (Varghese et al., 1999). For GH1 BGs, the active residues are usually located at the bottom of the pocket (Hrmova et al., 1998; Matsuzawa et al., 2016). Some BGs are capable of hydrolyzing soluble cellodextrin (Shewale, 1982; Schmid and Wandrey, 1987) or even insoluble cellooligosaccharides with the average degree of polymerization (DP) up to 20 (Sakamoto et al., 1985b). Furthermore, two BGs, isoenzyme βII (GH1) of germinated barley (*Hordeum vulgare* L.) and *Hj*Cel3A (GH3)

of *H. jecorina*, prefer to hydrolyze longer cellooligosaccharides than cellobiose (Hrmova et al., 1998; Karkehabadi et al., 2014). These properties might be because their pockets are longer and narrower at the exit such that they can accommodate substrates (Varghese et al., 1999). AaBGL1, the GH3 BG of *Aspergillus aculeatus* capable of hydrolyzing insoluble cellooligosaccharides with high DP up to 20, was found to have a long cleft extending from sugar-binding subsite +1 (subsite denotes the position of a specific sugar unit relative to the bond actually undergoing hydrolysis, with the reducing end as +*n*; Davies et al., 1997), which appears to be suitable for binding long cellooligosaccharides (Suzuki et al., 2013).

Glycoside hydrolases catalyze the hydrolysis reaction via two major mechanisms: either a retention or an inversion of the configuration of the anomeric carbon (Henrissat and Davies, 1997; White and Rose, 1997; Fig. 0-3). Aside for BGs in GH9, all BGs belong to the retaining type, in which the reaction is mediated by two carboxyl groups in the catalytic center, with one acts as a nucleophile and the other as an acid/base catalyst, respectively. The retaining type BGs perform catalysis via a double-displacement mechanism including two steps of glycosylation and deglycosylation (Koshland, 1953; White et al., 1996). In the glycosylation step, the aglycone accepts a proton from the catalytic acid/base and departs, whilst the anomeric carbon of the glycone is attacked by the catalytic nucleophile and form a covalent enzyme-glycone intermediate. In the deglycosylation step, the catalytic acid/base extracts a proton from a water molecule, then the free hydroxide radical attacks the anomeric carbon of the glucose moiety of the enzyme-glycone intermediate, and displaces the catalytic nucleophile to release the enzyme (Ketudat Cairns and Esen, 2010). A comparison of the catalytic mechanisms, catalytic residues, and structural

features of BGs affiliated with different GHs are shown in Table 0-1.

0.5 Transglycosylation and product/substrate inhibition of BGs

Under defined conditions, BG can synthesize glycosyl-bond between different molecules via reverse hydrolysis and transglycosylation (Bhatia et al., 2002). In transglycosylation reaction, an enzyme-glycosyl intermediate is initially formed by the glycosylation action of BG to a donor glycoside (e.g., a disaccharide or aryl-linked glucoside), then a nucleophile (such as monosaccharide, disaccharide, and alcohols, etc.) other than water (in the case of hydrolysis) displaces the enzyme of the intermediate and yields a new elongated product (Bhatia et al., 2002).

High concentrations of glucose can inhibit BGs by either blocking the active site for the substrate or preventing the reaction products from leaving (Krogh et al., 2010). The manner of product inhibition could be competitive (Bissett and Sternberg, 1978) or non-competitive (Gong et al., 1977). In contrast, when using *p*-nitrophenyl β-D-glucopyranoside (*p*NPG) as the substrate, certain BGs are found being stimulated by low concentrations of glucose (Maguire 1977; Zanoelo et al., 2004; Uchiyama et al., 2013), but gradually inhibited by elevated concentrations of glucose (Pérez-Pons et al., 1995; Fang et al., 2010; Nascimento et al., 2010; Uchima et al., 2011; Pei et al., 2012; Zhao et al., 2012; Uchiyama et al., 2015). These characteristics are attributable to their high transglycosylation activity, in which the BGs preferentially use glucose rather than water as an acceptor for the glycosyl moiety in the catalytic reaction (Uchiyama et al., 2013). BGs with transglycosylation activities can generate glucodisaccharides such as sophorose, laminaribiose, cellobiose, and gentiobiose (Uchiyama et al., 2013), or trisaccharides (Saloheimo et al., 2002; Bohlin et al., 2013). The transglycosylation

phenomenon was found to be exacerbated when the concentration of cellobiose (Smaali et al., 2007) or glucose increased (Bohlin et al., 2013), and the transglycosylation products were generated in a time-dependent manner (Yang et al., 2013; Guo et al., 2015; Zhang et al., 2016).

Substrate inhibition of BGs by cellobiose is often observed (Han and Srinivasan, 1969; Sternberg, 1976; Shewale and Sadana, 1981), which might also be due to the competition between the water molecular and the acceptor sugar for the glucosyl-enzyme intermediate to conduct hydrolysis or transglycosylation reaction (Kawai et al., 2004; Bohlin et al., 2013). Some BGs are found being competitively inhibited by cellobiose in the hydrolysis of *p*NPG (Han and Srinivasan, 1969; Ait et al., 1982; Harhangi et al., 2002). High concentration of cellobiose might slow down the hydrolysis rate of BGs, not because of the delays the catalytic cycle but due to transglycosylation (Bohlin et al., 2013).

0.6 Role of BGs in industrial cellulose degradation

Industrial enzymatic cellulose degradation is based on the repertoire of cellulases of *T. reesei* (Merino and Cherry, 2007). To achieve effective hydrolysis of cellulose, concerted actions of EGs, CBHs, and BGs are required. Cellobiose is the hydrolysis product of CBHs and partially generated from actions of EGs. Both CBHs and EGs are inhibited by cellobiose, and BGs alleviate this product inhibition of EGs and CBHs by hydrolysis of cellobiose (Halliwell and Griffin, 1973; Wood and McCrae, 1975). On the other hand, BGs in commercial cellulase preparations such as Novozym 188 and CTec2 were found to be easily absorbed to lignin (residue of feedstock pretreatment), which led to the loss of the majority of BG activity (Haven and Jørgensen, 2013).

Trichoderma, particularly *T. reesei* and its mutants such as RUT-C30 (Tangnu et al., 1981) and PC-3-7 (Kawamori et al., 1986b), are excellent producers of cellulases (Coughlan, 1985), especially strains . Cellulases of *T. reesei* had set up the current industry standard (Merino and Cherry, 2007). However, *T. reesei* are characterized by their low secretion ability of BGs (Sternberg, 1976), presumably because low levels of BG are sufficient for growth of *T. reesei* on cellulose, but are insufficient for large scale *in vitro* cellulose hydrolysis for bioethanol production (Lynd et al., 2002). The low yield of BG was also found in another commercialized cellulase producer strain *T. viride* (Gong, 1977). Solutions to promote the overall BG activity include the addition of extra BGs from other sources (Sternberg et al., 1977), optimizing the culture condition for higher BG yield (Tangnu et al., 1981), co-cultivation of fungal strains producing cellulase and BG (Duff et al., 1985), mutation of cellulase producing strains (Kawamori et al., 1986a), and over-expression of either endogenous (Barnett et al., 1991) or exogenous (Nakazawa et al., 2012) BGs in the recombinant *T. reesei* strains.

Currently, the common industrial technique for relieving product inhibition of CBH and EG is to add BG into the cellulases cocktails. For example, Celluclast 1.5L (Sørensen et al., 2011) and Cytolase CL (Ju et al., 2014) are supplemented with the BG preparation Novozym 188. In addition, the BG activity of the novel commercial cellulase preparation Cellic CTec2 is further promoted to a level of over 10-fold of that of the Novozym 188 (Cannella and Jørgensen, 2014). The activities of commercial BG preparations such as Novozym 188 (Dekker, 1986; Sørensen et al., 2011; Ng et al., 2011) and Cellic CTec (Sørensen et al., 2011; Kawai et al., 2012) towards cellobiose are within 40 U/mg level. Novozym 188 is the most common commercial BG preparation and composed of two isoforms of BGs from *Aspergillus niger* (Himmel et al., 1993). A

structurally-characterized BG, *An*Bgl1 was purified from Novozym 188, and exhibited a specific activity towards cellobiose up to 98.7 U/mg (Lima et al., 2013).

0.7 Exploring BGs for bioethanol production

Present pre-treatment of lignocellulose dependents on acid and heat processes (Rubin, 2008). In terms of cellulose conversion, the cellulases produced by *Trichoderma* are stable in stirred tank reactors at pH 4.8, 50°C for 48 h or longer (Ryu and Mandels, 1980), and the hydrolysis conditions adopted in the current lignocellulosic ethanol production process are typically at pH 5 and 50°C for 24-120 h (Merino and Cherry, 2007). Therefore, it is clear that desirable BG should be stable and active at acidic ambient pH and temperature higher than 50°C. In addition, the industrial process operated at high-solid concentrations is particularly useful in reducing the cost of heating by lowering the volume of processing (Jørgensen et al., 2007). Henceforth, BGs resistant to both glucose and cellobiose inhibition are desirable when operating at the condition of high concentrations of feedstock.

Production of BG from *Aspergillus* had been commercialized for over half a century (Crook and Stone, 1957). *Aspergilli* are believed to be the most productive organisms in industrial BG production (Sternberg, 1976), and some of the BGs possess high activity along with hyperstability at 50°C. For instance, the BG preparation from *Aspergillus phonenicis* had a specific activity towards cellobiose up to 160 U/mg, and 85% of its activity still remained after 4 days at 50°C (Sternberg, 1976). Novozym 188 had a temperature stability up to 60°C (at pH 5 for 0.5 h) and a half-life at 50°C for over 48 h (Dekker, 1986). However, as far as the glucose tolerance is concerned, except for some exceptions (Riou et al., 1998), BGs from *Aspergillus* genera are generally either

sensitive to glucose (Bissett and Sternberg, 1978; Dekker, 1986; Zhou et al., 2016), or highly tolerant to glucose but have poor efficiencies on hydrolyzing cellobiose (Yan and Lin, 1997; Günata and Vallier, 1999; Decker et al., 2001). Furthermore, *Aspergillus*-derived BGs with high activities on cellobiose are usually inhibited by cellobiose at low concentration. For instance, Sternberg et al. reported that some *Aspergillus*-derived BGs were inhibited at the cellobiose concentration over 10 mM (Sternberg et al., 1977), and Dekker found that Novozym 188 was inhibited by cellobiose at the concentration of around 10 mM (Dekker, 1986). Conversely, some BGs are found to have advantages on the activity and cellobiose resistance but are weak in glucose inhibition, as shown in the case of β-Glu II, an extracellular β-glucosidase from *A. niger* CCRC 31494, which possesses an activity of 464 U/mg on cellobiose, and is tolerant to cellobiose over 50 mM, but has a K_i value of merely 5.7 mM on glucose (Yan et al., 1998). To sum up, current commercialized BG products from *Aspergilli* do not fully satisfy the demand of industrial bioethanol production, which has been prompting people to screen and develop high efficiency enzymes through protein engineering.

In recent years, due to consistent efforts of researchers, more and more BGs having great potential for industrial application had been isolated, characterized, and/or heterologously expressed. For instance, BGL, a recombinant β-glucosidase from the anaerobic bacteria *Thermoanaerobacterium aotearoense* P8G3#4 heterologously expressed in *Escherichia coli*, demonstrated a specific activity towards cellobiose up to 740.4 U/mg (Yang et al., 2015). RfBGluc-1, which was derived from the lower termite *Reticulitermes flavipes* and expressed by a baculovirus-insect expression system, exhibited a V_{max} value up to 638 U/mg against cellobiose (Scharf et al., 2010). BGL I, a

BG from the ascomycete fungus *Periconia* sp. BCC2871 and expressed in *Pichia pastoris*, was reported to have a V_{max} value of 627 U/mg (Harnpicharnchai et al., 2009). A list of the key kinetic properties of BGs with high activities documented is summarized in Table 0-2, using the criterion of specific activity or V_{max} over 100 U/mg towards cellobiose. It is noteworthy, however, that the definition of activity unit toward cellobiose in these reports is not uniform: some are defined as the amount of enzyme capable of producing 1 μmol of glucose per min under assay conditions (Yan et al., 1998; Wu et al., 2012; Ni et al., 2007), whilst others are defined as the amount of enzyme that produces 2 μmol of glucose per minute (Nakazawa et al., 2012; Treebupachatsakul et al., 2016; Sakamoto et al., 1985a). In addition, there are also many reports in which the definition of activity units was not clearly specified. Therefore, comparing activities among BGs is relatively difficult.

0.8 Objective of this study

Presently there are at least 68 phytophagous or xylophagous insect species affiliated with eight different taxonomic orders described to have the cellulolytic activities in the gut or head-derived fluids (Oppert et al., 2010). Insects thriving on lignocellulose digest cellulose either by their own cellulases or by the combination of endogenous cellulases and cellulolytic enzymes from the cellulolytic microbes such as protozoa, bacteria, or fungi (Martin, 1983). For instance, the intestinal protozoa of the lower termites are essential for the survival of their host in cellulose digestion (Cleveland, 1923), and referred to as ‘symbiont’, whilst xylophagous higher termites are capable of degrading cellulose by their own cellulases (Tokuda et al., 2012) except for *Macrotermes natalensis* which establish symbiotic relationship with the fungus *Termitomyces* sp

(Martin and Martin, 1978). The cellulose digestion paradigms in insects are generally similar to that in fungi (Martin, 1983), and BG activities are commonly present in insects (Terra et al., 1996). Therefore, those insects and their symbionts could be treasuries harboring powerful BGs. Some termite-origin BGs characterized exhibited high activities, as shown in the cases of RfBGluc-1 mentioned above and other termite-derived BGs listed in Table 0-2.

BG activity was detected in guts of the cockroach long time ago (Banks, 1963). Recently, PaBG1b (GenBank accession number: LC125463), an endogenous BG derived from the wood feeding cockroach *Panesthia angustipennis spadica* (Blaberidae: Panesthiinae) collected from Tsukuba city, Japan, was purified by Arakawa et al. from the midgut enzyme extract of the nymphs (Arakawa et al., 2016). PaBG1b is affiliated with GH1 and the native enzyme displayed extremely high specific activity up to 708 U/mg towards cellobiose and a V_{max} of 1,020 U/mg, which might be the highest among BGs reported (Arakawa et al., 2016; Table 0-2). PaBG1b has high catalytic efficiency with a k_{cat}/K_m value of 184 mM/s, which is at least among the top 3 of BGs listed in Table 0-2, and displays an optimum temperature at 60°C. In terms of the thermostability, the native enzyme exhibited a thermo-tolerance of 30 min at 50°C (Arakawa et al., 2016). All these characteristics endow PaBG1b with remarkable potential for industrial application. In addition, although BGs from omnivorous cockroach *Periplaneta americana* (Koffi et al., 2012) and wood-feeding cockroach *Panesthia cribrata* Saussure (Scrivener and Slaytor, 1994) were isolated, to the best of my knowledge thus far there has been no case of heterologous expression of BGs from cockroaches being documented. Therefore, overexpression and characterization of PaBG1b, elucidating the molecular basis that builds up its high activity, improving its robustness through

mutagenesis, and employing it as an additive of commercial cellulase cocktails are of great value.

On the other hand, termite gut symbionts are also novel reservoirs of cellulases, within which some high performance BGs had been screened out as listed in Table 0-2. RsBG is a GH3-like BG from the hindgut symbiotic protist of the subterranean lower termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae), whose sequence was isolated from the cDNA library of the symbiotic protist community of *R. speratus* through environmental cDNA analysis (Todaka et al., 2007). Another objective of this study is the heterologous expression of RsBG displaying unique sequence features as a member of GH3 and its biochemical characterization.

Table 0-1. Comparison of BGs in various GH families (Source: CAZy database)

GH family	GH1	GH3	GH5	GH9	GH30	GH116
Hydrolysis mechanism	Retaining	Retaining	Retaining	Inverting	Retaining	Retaining
Clan	GH-A	-	GH-A	-	GH-A	-
Catalytic nucleophile[a] (or catalytic base, for inverting mechanism)	E	D	E	D	E	E
Catalytic acid/base[a] (or catalytic proton donor for inverting mechanism)	E	E	E	E	E (inferred)	D
3D Structure status	$(\beta/\alpha)_8$	-[b]	$(\beta/\alpha)_8$	$(\alpha/\alpha)_6$	$(\beta/\alpha)_8$	-

-: not specified.
[a]: catalytic residues are determined experimentally, unless otherwise noted.
[b]: structurally characterized GH3 BGs thus far generally have a N-terminal $(\beta/\alpha)_8$-barrel domain (or $\beta\beta(\beta/\alpha)_6$-barrel domain), and an $(\alpha/\beta)_6$-sandwich domain, with the N-terminal domain harboring the nucleophile catalyst and the α/β-sandwich domain housing the catalytic acid/base residue (c.f. Table 3-2 in Chapter 3).

Table 0-2. Comparison of kinetic properties of selected BGs with high activity towards cellobiose (specific activity or V_{max} >100 U/mg)

Enzyme Name	Origin	Expression Host	Spec. at. (U/mg)	V_{max} (U//mg)	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	Reference
AaBG1	*Aspergillus aculeatus*	(Native enzyme)	180	N/A	N/A	N/A	Nakazawa *et al*., 2012
BGL1	*Aspergillus aculeatus*	*Saccharomyces cerevisiae*	194	N/A	N/A	N/A	Treebupachatsakul *et al*., 2016
β-Glucosidase 1	*Aspergillus aculeatus* No. F-50	(Native enzyme)	132	N/A	N/A	N/A	Sakamoto *et al*., 1985a
β-Glucosidase 2	*Aspergillus aculeatus* No. F-50	(Native enzyme)	157	N/A	N/A	N/A	Sakamoto *et al*., 1985a
(Unnamed)	*Aspergillus japonicus*	(Native enzyme)	124[b]	N/A	1.0	350[b]	Korotkova *et al*., 2009
β-Glu II	*Aspergillus niger* CCRC 31494	(Native enzyme)	N/A	232[a]	15.4	N/A	Yan *et al*., 1998
PBG	*Aspergillus niger* SK34.002	(Native enzyme)	209.9[b]	1221[b]	8.9	N/A	Zhou *et al*., 2016
HGT-BG	*Aspergillus oryzae*	(Native enzyme)	N/A	353[b]	7	N/A	Riou *et al*., 1998
(Unnamed)	*Aspergillus phonenicis*	(Native enzyme)	160[b]	N/A	0.8	N/A	Sternberg *et al*., 1977
β-Glucosidase I	*Aspergillus tubingensis C*BS 643.92	(Native enzyme)	N/A	151.5[b]	1	N/A	Decker *et al*., 2001
(Unnamed)	*Aspergillus wentii*	(Native enzyme)	113[b]	N/A	N/A	N/A	Sternberg *et al*, 1977
(Unnamed)	*Clostridium thermocellum*	*Escherichia coli*	322[b]	N/A	N/A	N/A	Kadam and Demain, 1989
(Unnamed)	*Coprinopsis cinerea* ATCC 56838	(Native enzyme)	175.3[b]	N/A	N/A	N/A	Zhang *et al*., 2016
SRF2g14	Cow rumen metagenome	*Escherichia coli*	186.0[a]	N/A	8.0	23.8[a]	Del Pozo *et al*., 2012
SRF2g18			426.4[a]	N/A	25.5	1.7[a]	
LAB20g4			153.0[a]	N/A	7.8	14.4[a]	
LAB25g2			768.6[a]	N/A	4.9	10.4[a]	
Glu1B	*Coptotermes formosanus*	*Escherichia coli*	N/A	462.6[b]	2.3	N/A	Zhang *et al*., 2012
(Unnamed)	*Fomitopsis pinicola* KMJ812	(Native enzyme)	117[b]	N/A	N/A	N/A	Joo *et al*., 2009
Ks5A7	Kusaya gravy metagenome	*Escherichia coli*	170[b]	155[b]	0.4	386[b]	Uchiyama *et al*., 2015
MbmgBG1	*Macrotermes barneyi*	*Escherichia coli*	103[a]	N/A	N/A	N/A	Wu *et al*., 2012
β-Glucosidase A	*Macrotermes muelleri*	(Native enzyme)	199[a]	N/A	1	N/A	Rouland *et al*., 1992

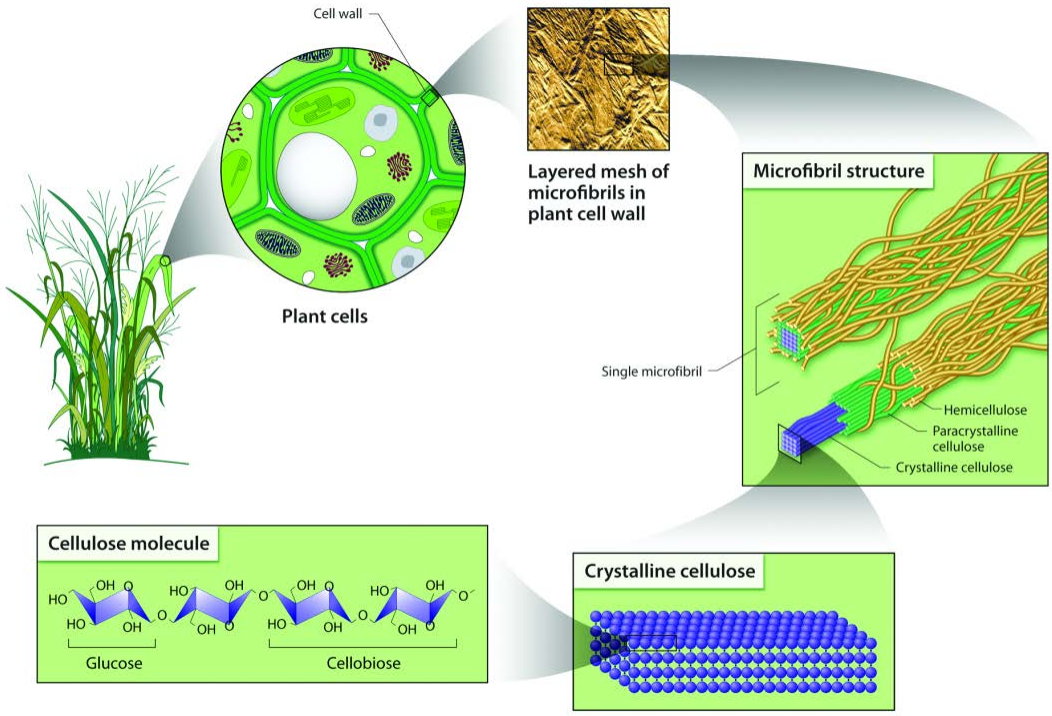
(Continued)

Table 0-2. Comparison of kinetic properties of selected BGs with high activity towards cellobiose (specific activity or V_{max} >100 U/mg) **(Continued from the previous page)**

Enzyme Name	Origin	Expression Host	Spec. at. (U/mg)	V_{max} (U//mg)	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	Reference
reBglM1	*Marinomonas* MWYL1	*Escherichia coli*	N/A	508[b]	1.1	395.8[b]	Zhao *et al*., 2012
BglU	*Micrococcus antarcticus*	*Escherichia coli*	198.3[a]	N/A	N/A	N/A	Fan *et al*., 2011
NfBGL1	*Neosartorya fischeri* P1	*Pichia pastoris* GS115	101.9[a]	N/A	N/A	N/A	Yang *et al*., 2014
NkBG	*Neotermes koshunensis*	*Escherichia coli*	78.4[a]	110[a]	3.8	N/A	Ni *et al*., 2007
(Unnamed)	*Paecilomyces thermophila* J18	(Native enzyme)	49.1	272.1	0.7	7	Yang *et al*., 2008
PtBglu1	*Paecilomyces thermophila* J18	*Pichia pastoris* GS115	65.2	306.3	1.0	5.1	Yang *et a*l., 2013
PaBG1b	*Panesthia angustipennis spadica*	(Native enzyme)	708	1020	5.3	184	Arakawa *et al*., 2016
Bgl6	*Penicillium funiculosum* NCL1	*Escherichia coli*	N/A	166[b]	0.3	N/A	Ramani *et al*., 2015
BGL	*Penicillium purpurogenum* KJS506	(Native enzyme)	432[b]	N/A	N/A	N/A	Jeya *et al*., 2010
BGL I	*Periconia* sp. BCC2871	*Pichia pastoris*	N/A	627[b]	0.5	N/A	Harnpicharnchai *et al*., 2009
*Rf*BGluc-1	*Reticulitermes flavipes*	Baculovirus-insect system	N/A	638[b]	1.4	N/A	Scharf *et al*., 2010
(Unnamed)	*Rhizomucor miehei* NRRL 5282	(Native enzyme)	N/A	115.5[b]	0.1	N/A	Krisch *et al*., 2012
β-Glucosidase B	*Termitomyces* sp.	(Native enzyme)	103.5[a]	N/A	2.8	N/A	Rouland *et al*., 1992
BGL	*Thermoanaerobacterium aotearoense* P8G3#4	*Escherichia coli*	370.2[a]	370.3[a]	25.5	157.2[a]	Yang *et al*., 2015
(Unnamed)	*Thermoanaerobacterium thermosaccharolyticum* DSM 571	*Escherichia coli*	N/A	120[b]	7.9	13.3[b]	Pei *et al*., 2012

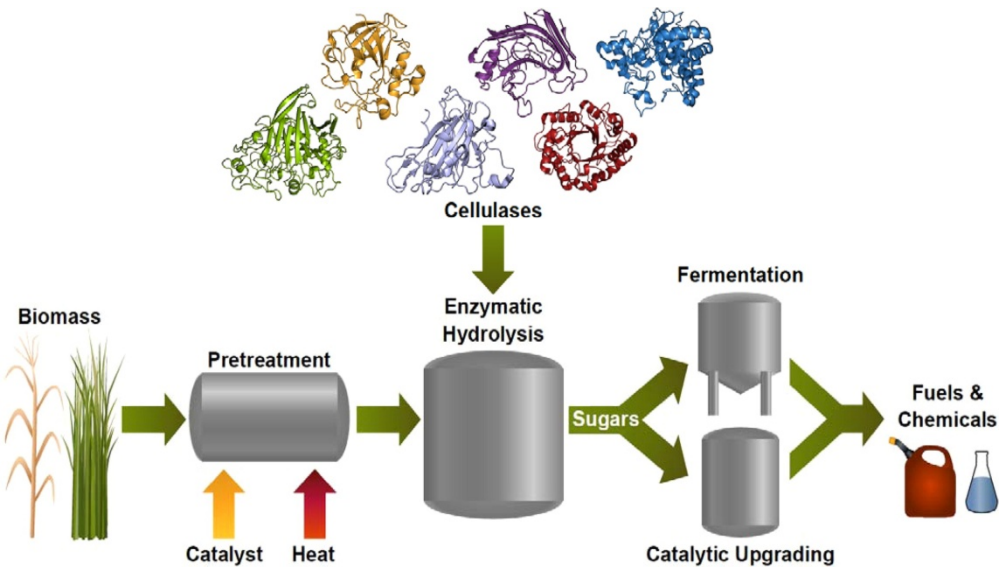
Spec. at.: specific activity. One unit of activity was defined as the amount of enzyme that produced 2 μmol of glucose per minute.
[a]: for comparison, data in the original references were re-calculated to show the enzyme unit as defined above.
[b]: the definition of unit of activity was not clearly specified in the source references.

A



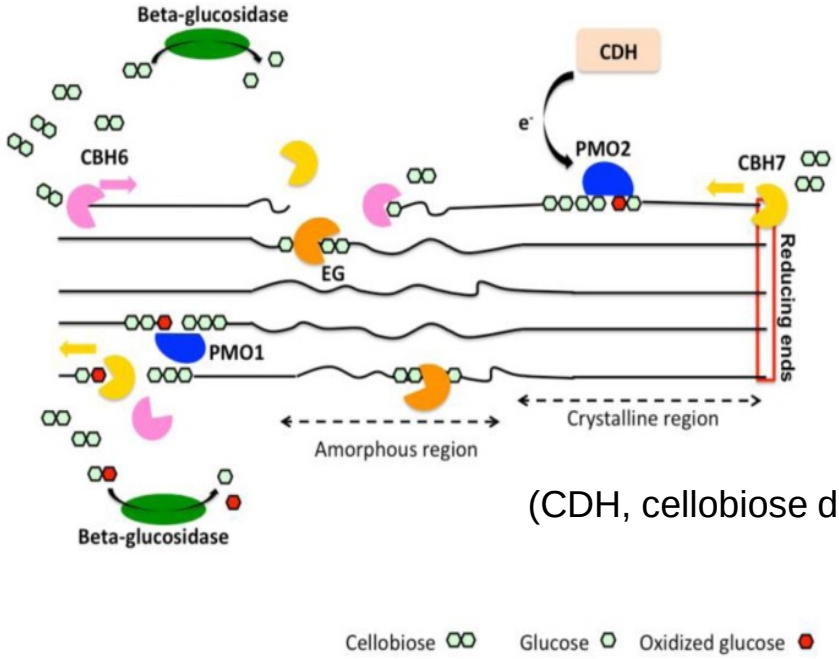
Source: http://www.intechopen.com/source/html/44414/media/image1.png

B



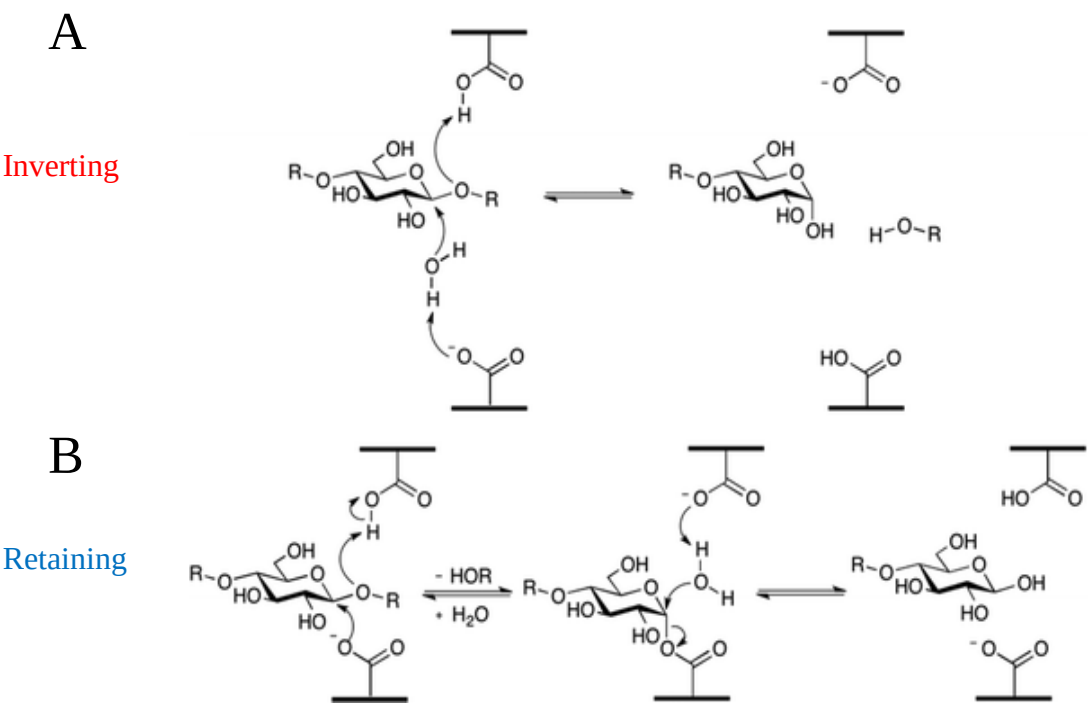
Source: Payne *et al*. (2015)

Fig. 0-1. Structure of cellulose in the plant cell walls (A) and the schemes of biofuel production (B)



Source: Dimarogona *et al*. (2012)

Fig. 0-2. A simplified scheme of the current view on the enzymatic degradation of cellulose by the 'free' cellulases paradigm



Source: Payne *et al*. (2015)

Fig. 0-3. A scheme of catalytic mechanisms of inverting (A) and retaining (B) carbohydrate-active enzymes (GHs)

Chapter 1

Heterologous expression of PaBG1b in *P. pastoris*

1.1 Introduction

1.1.1 BGs in cockroaches

Cockroaches have close phylogenetic relationship with termites. The wood-feeding cockroaches were evolved from the same ancestor with termites (Lo et al., 2000), and *Cryptocercus* (Cryptocercidae) is the last common ancestor of wood-feeding cockroaches and termites (Lo et al., 2000). It is notable that *Cryptocercus punctulatus* is unique in that it is possibly the most primitive extant cockroach and has both endogenous and symbiotic cellulases in the gut (Scrivener et al., 1997), whereas wood feeding cockroaches *Salganea* (Blaberidae: Panesthiinae) (Maekawa et al., 2008) and *Panesthia* (Blaberidae: Panesthiinae) (Scrivener et al., 1989) mainly depend on their endogenous cellulases. As far as BGs are concerned, based on the gene expression and molecular phylogenetic analysis, Shimada and Maekawa suggested that before the divergence of termites and cockroaches, they shared at least two types of BG gene homologues, one acting as a digesting enzyme and another being involved in pheromonal communication (Shimada and Maekawa, 2014). Therefore, cockroaches, especially xylophagous cockroaches, might likewise possess powerful BGs as termites do.

Before this study, Scrivener and Slaytor purified two BGs (GD1 and GD2) from *Panesthia cribrata* Saussure (Scrivener and Slaytor, 1994), both of which had pretty low activities (V_{max} values of GD1 and GD2 against cellobiose were only 1.0 U/mg). Shimada and Maekawa isolated partial cDNA sequences of four BGs from three xylophagous cockroaches, i.e. CpBGI from *Cryptocercus punctulatus*, SeBGI and II from *Salganea esakii*, and PaBGI from *Panesthia angustipennis spadica* (Shimada and Maekawa, 2014), but neither the native enzymes were purified nor the proteins were

heterogeneously expressed. For cockroaches other than the wood-feeding cockroaches, Koffi et al. purified and biochemically characterized a BG from the omnivorous cockroach *Periplaneta americana* (Koffi et al., 2012) with low V_{max} of 39.12 U/mg towards cellobiose, and suggested that the role of this BG was not for the digestion of cellulosic material but for hydrolysis of glycoside toxins ingested. Tamaki et al. cloned three BG cDNAs from the transcriptome of the midgut of *P. americana* (Tamaki et al., 2014), one of which having the complete sequence (GenBank accession number KJ576835). They also identified two BG activities from the electrophoretic and zymolographical data, with one of which being suggested to be the same BG reported by Koffi et al. In addition, Cornette et al. purified the native enzyme of Lma-p72 of the Madeira cockroach *Leucophaea maderae* (Blaberidae: Oxyhaloinae), and the complete cDNA of Lma-p72 was cloned (Cornette et al., 2003). Lma-p72 is a GH1 glycosidase (EC 3.2.1) which exhibits a β-galactosidase-like activity and is fully inhibited by the general BG inhibitor D-gluconolactone, and its physiological function was suggested to be cleaving a pheromone-sugar conjugate to release the pheromonal compounds on to the epicuticular surface in *L. maderae* males (Cornette et al., 2003).

P. angustipennis is a gregarious wood-feeding cockroach living in and feeding on rotting logs during its whole development period (Nalepa et al., 2008). *P. angustipennis spadica* is one of its subspecies and was found in Kyushu, Shikoku, Honshu, and Taiwan Island (Maekawa et al., 1999). By the study of the midgut enzyme extract of the nymph *P. a. spadica* (Fig. 1-1A), PaBG1b (Uniprot ID: BAU51446) with an apparent molecular weight of 56.7 kDa was isolated and purified, and the complete cDNA sequence of PaBG1b with 1,551 bp in length (GenBank accession number: LC125463) was cloned from the cDNA library (Arakawa et al., 2016). This cDNA sequence shows

99% identity with the partial cDNA sequence encoding a putative BG fragment PaBGI (GenBank accession number: AB915872; Uniprot ID: BAO85050) obtained by Shimada and Maekawa (Shimada and Maekawa, 2014), and the deduced amino acid sequence of PaBG1b (502 amino acids) differs in only one amino acid (R221) from the partial sequence of PaBGI (303 amino acids).

1.1.2 Host for expression of PaBG1b

E. coli and yeast are often employed as the host organisms for over-expression of BGs. Although *E. coli* has the advantages of rapid growth and high transformation efficiency, many proteins of eukaryotic origins expressed in *E. coli* have been found to be insoluble, which might be attributable to improper folding and/or lack of necessary post-translational modifications (Romanos et al., 1992). Besides, the codon usage of the expression host affects the translation efficiency and is considered to be a potential factor for successful expression (Romanos et al., 1992). Currently, there are certain online analysis tools available such as the GenScript (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) for rare codon analysis and selecting desired expression organism and/or codon optimization, which could contribute to successful expression.

As far as the yeast expression system is concerned, although *Saccharomyces cerevisiae* is capable of performing post-translational modifications, there are still some drawbacks such as the lack of potent and tightly-regulated promoters, and hyperglycosylation of expression product which might alter the enzymatic or immunogenic properties (Ballou, 1970; Romanos et al., 1992). Instead, the methylotrophic yeast *Pichia pastoris* is another excellent eukaryotic system for

heterologous expression, which could grow to very high densities up to 60 g/L dry weight of cells (Cregg et al., 1987), and produce large amounts of protein up to 12 g/L (Clare et al., 1991a). *P. pastoris* appears to have the advantage in secreting proteins with high molecular weight compared to *S. cerevisiae* in which such proteins mostly retained in the periplasm (Çelik and Çalık, 2012). The structures of *N*-linked oligosaccharides of the secreted expression products by *P. pastoris* are generally $Man_{8-14}GlcNAc_2$ (Grinna and Tschopp, 1989), composed of a core of $Man_3GlcNAc_2$ (893 Da) and variable amounts of hexoses (Tull et al., 2001), which is much shorter than that of hyper-mannosylated products (>50) expressed in *S. cerevisiae* (Romanos et al., 1992). The glycoproteins expressed in *S. cerevisiae* have large numbers of α-1,3-linked mannose units in the outer chain, which might be immunogenic (Ballou, 1970), and this does not occur in the products expressed in *P. pastoris* (Trimble et al., 1991) due to the absence of α-1, 3-linked mannosyl transferase (Çelik and Çalık, 2012).

P. pastoris has two alcohol oxidases, AOX1 and AOX2, which catalyze the first step of the methanol-utilization pathway (Cregg et al., 1989). As the yield of alcohol oxidases can be induced to the level of about 33% of the proteins of cell free extracts (Couderc and Baratti, 1980) and AOX1 is the primary alcohol oxidase (Cregg et al., 1989), the powerful, tightly regulated and methanol-inducible *AOX1* promoter is particularly suitable for driving the expression of target genes (Romanos et al., 1992). Furthermore, multi-copy integrant *P. pastoris* strains of foreign protein genes can be obtained by using vectors containing tandem copies of the expression cassettes (Clare et al., 1991a) or by screening integrants harboring multiple copies of foreign genes (Clare et al., 1991b), and employing such multi-copy integrants frequently achieved remarkably high yields of products (Romanos et al., 1992). To improve the

post-translational processing of the target protein by limiting the growth rate of *P. pastoris*, the *AOX1*-defective mutant (i.e. Mut^{s} strain, which means 'methanol utilization slow') was developed (Cregg et al., 1987).

Lastly, *P. pastoris* secretes endogenous proteins at relatively low levels, which greatly facilitates the purification of expression products (Çelik and Çalık, 2012).

1.1.3 Plasmids for expression of PaBG1b

For extracellular expression in *P. pastoris*, two expression plasmids, pBGP3 and pPICZα-A, have been used most often in our laboratory. The expression vector pBGP3 was derived from pBGP1, an expression vector customized for screening mutant gene library of *P. pastoris* (Lee et al., 2005; Fig. S1-1 A and B, in page 45). pBGP1 was designed to be an episomal plasmid containing both *E. coli* and *P. pastoris* replication origins (colE1 and PARS1, respectively), which allows it to be a shuttle vector. The prepro α-factor sequence of *S. cerevisiae* is included upstream of the multicloning site (MCS), which allows the expression product to be secreted extracellularly into the media, which facilitates subsequent product collection and purification. The α-factor precursor of *S. cerevisiae* contains a canonical signal sequence of around 20 hydrophobic amino acid residues, followed by a pro-region of about 60 hydrophilic amino acids, and four tandem copies of mature α-factor sequence separated by spacer peptides of six or eight residues (variations of Lys-Arg-Glu-Ala-Asp-Ala-Glu-Ala, i.e., KREADAEA) (Kurjan and Herskowitz, 1982; Julius et al., 1983). The pro-region of α-factor precursor was suggested to be involved in the translocation of the precursor from the ER to the Golgi apparatus (Julius et al., 1984). In *S. cerevisiae*, endopeptidase Kex2 (Julius et al., 1984) and dipeptidyl aminopeptidase Ste13 (Julius et al., 1983) are

membrane-bound proteases involved in the processing of the α-factor precursor. To release the mature α-factor sequences, Kex2 cleaves C-terminus to the lysine-arginine (KR) dibasic amino acid pair of the spacer peptide (Julius et al., 1984; Achstetter and Wolf, 1985), whereas Ste13 excises the three EA repeats preceding each of the four mature pheromone sequences (Julius et al., 1983). The proteolytic processing by Kex2 presumably occurs in the late Golgi, prior to or simultaneously with the vacuolar protein sorting (Graham and Emr, 1991). *P. pastoris* has the homologues of both Kex2 and Ste13 (De Schutter et al., 2009), and the coding sequence of polypeptide KREAEA is employed in the expression vector as the proteolytical cleavage sites for the post-translational processing of the expression product. As for the promoter, *AOX1* and *GAP* genes are to date the most prominently employed for achieving heterologous expression in *P. pastoris* (Ahmad, 2014). pBGP1 employs the *S. cerevisiae GAP* promoter to drive the gene expression in a constitutive manner. *GAP* promoter, which is constitutively active on glucose medium and to a lesser extent on glycerol and methanol media, could be employed for circumventing the drawbacks of *AOX1* promoter, as the feeding of methanol might be unsuitable, especially in the food processing (Ergün et al., 2015). The ampicillin-resistance gene and the Zeocin-resistance gene (*zeoR*) are used as selection markers in *E. coli* and *P. pastoris*, respectively (Lee et al., 2005). pBGP3 was developed from pBGP1 in our laboratory (Uchima and Arioka, 2012). pBGP3 contains a sequence encoding c-*myc* epitope (c-myc) and six consecutive histidines (i.e., hexahistidine, or His_6) at the upstream of the MCS, for the purpose of expressing a chimeric fusion protein with the c-myc and His_6 tags at the N-terminus, allowing the detection of the recombinant protein by Western blot analysis, and the purification via Ni-NTA column chromatography. pBGP3 has been successfully applied to heterologous

expression of termite-derived cellulases, NkBG, mgNtBG1, sgNtBG1, RsEG, and NtEG in our laboratory (Uchima and Arioka, 2012; Uchima et al. 2012; Akemi Uchima et al., 2013).

pPICZα-A is a commercial, integration-type vector for high-level, methanol-inducible expression and secretion of recombinant proteins in *P. pastoris*. The expression of recombinant proteins is under the control of *AOX1* promoter, and the secretion of the expressed product is directed by *S. cerevisiae* prepro α-factor sequence. Zeocin resistance gene is employed for selection in both *E. coli* and *P. pastoris*. The c-myc and His_6 tags are placed at the C-terminus of MCS for detection and purification of the recombinant fusion protein.

1.2 Results

1.2.1 Bioinformatic analysis of PaBG1b

The open reading frame of *pabg1b* (1,509 bp) encodes a polypeptide of 502 amino acids. The deduced amino acid sequence of PaBG1b was analyzed by BLAST server (Altschul et al., 1997; https://blast.ncbi.nlm.nih.gov/Blast.cgi), and the results showed that the hypothetical BG of omnivorous cockroach *P. americana* (GenBank accession number: KJ576835), cDNA sequence of which was isolated by Tamaki et al. (Tamaki et al., 2014), shares the highest amino acid sequence identity (63 %), followed by NkBG (GenBank accession number: AB073638) of the lower termite *Neotermes koshunensis* with 62% of identity, and SgNtBG4 (GenBank accession number: AB508957) of the higher termite *Nasutitermes takasagoensis* (Tokuda et al., 2009) with 58% of identity. NkBG is the first endogenous BG of termite isolated from the salivary glands of *N. koshunensis* (Shiraki) and was characterized in 2002 (Tokuda et al., 2002). NkBG was

heterologously expressed in *E. coli* (Ni et al., 2007) and *A. oryzae* (Uchima et al., 2011), and was structurally resolved in 2011 (Jeng et al., 2011; PDB ID: 3AHZ). The partial amino acid sequence of the putative BG PaBGI shares 99% identity with PaBG1b. Phylogenetic tree inferred from the amino acid sequences of the homologues mentioned above is shown in Fig. 1-1B. It is noteworthy that two of the termite origin BGs, *Rf*Gluc-1 and Glu1B, both of which were reported to have high activities (Table 0-2), also have the highest homologies (E-value 0.0) with PaBG1b and share 57% and 56% identities with PaBG1b, respectively.

The first 20 amino acid residues of the deduced amino acid sequence of PaBG1b were predicted to be the secretion signal by SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/), which suggested that PaBG1b might be a secretion protein. Two potential *N*-glycosylation sites (N265 and N416) and one *O*-glycosylation site (S315) were predicted by NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetOGlyc 4.0 (http://www.cbs.dtu.dk/services/NetOGlyc/) servers, respectively. Amino acid sequence alignment of homologues of PaBG1b (described below) indicated that two glutamic acids, E196 and E406, might serve as the catalytic acid/base and catalytic nucleophile, respectively. The presence of a lysine-arginine (KR) dipeptidyl sequence in the N-terminal region of putative mature polypeptide of PaBG1b suggests the possibility that the translation product of PaBG1b in yeasts might be excised by Kex2 at this site and lose 9 amino acids from the N-terminus. The nucleotide and deduced amino acid sequences of PaBG1b are shown in Fig. 1-2.

Multiple alignment of PaBG1b against its homologues in Fig. 1-1B are shown in Fig. 1-3. PaBG1b has a GH1 signature described by the PROSITE motif PS00572, which is

depicted as the consensus sequence [LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN] with 'E' being the active site residue (http://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS00572). It is interesting that insect-origin BGs used for the alignment in Fig. 1-3, together with other insect-derived BGs mentioned in Table 0-2, i.e. *Rf*Gluc-1 and Glu1B, do not contain any KR dibasic site in their whole amino acid sequences of the mature region. In the budding yeast, dibasic amino acid pairs such as KR, RR, and RK are the potential processing sites of protein precursors such as prepro-α-factor and killer toxin precursor (Julius et al., 1984). However, it is unknown whether this mechanism also exists in insects. A summary of information of PaBG1b inferred from the amino acid sequence is shown in Table 1-1.

The cDNA sequence of PaBG1b was uploaded to the online tool GenScript for rare codon analysis (http://www.genscript.com/quick_order_menu.html). The results showed that PaBG1b sequence has CAI (Codon Adaptation Index) values of 0.73 and 0.63 when using yeast and *E. coli* as the expression host, respectively. Since a CAI of >0.8 is rated as good for expression in the desired expression organism, it might be more favorable to choose yeast as the expression host of PaBG1b.

1.2.2 Expression of N- or C-terminally-tagged PaBG1b in *P. pastoris*

In this study, *P. pastoris* was chosen for the heterologous expression of PaBG1b. For this purpose, the episomal expression plasmid pBGP3 and the integration-type expression plasmid pPICZα-A harboring the inducible *AOX1* promoter were used.

The cDNA fragment encoding PaBG1b without the putative signal peptide was amplified by PCR from the cDNA library provided by our collaborators (Arakawa,

Kamino, Tokuda, and Watanabe; personal communication), and served for the construction of the plasmid pBGP3-PaBG1b (Fig. 1-4 A). The calculated size of expression product fused with N-terminal c-myc and His_6 tags is 58.5 kDa. As a KR dibasic site is present at the N-terminus of the mature region of PaBG1b, the N-terminal tags of the expression product might suffer Kex2 excision and a product with the calculated size of 54.3 kDa might be generated (Fig. 1-4 B). Transformation of pBGP3-PaBG1b to *P. pastoris* was conducted by electroporation and three transformants screened by colony PCR from the selective plates containing Zeocin were further used for expression test. The culture supernatant was sampled daily for the analyses including Western blot, CBB staining, and BG activity assay. In this study, *p*NPG was used as the substrate for regular BG activity assay, unless otherwise noted. According to the Western blot analysis and BG activity assay results, the expression of N-terminally c-myc- and His_6-tagged PaBG1b was detected on the 3rd day of the main culture and the yield significantly increased on the 4th day. Although PaBG1b could not be observed on the CBB-stained SDS-PAGE gel, a clear immunoblot band with the approximate size of 61.8 kDa was detected, suggesting the successful expression of PaBG1b (Fig. 1-5 A). Enzyme assay of the culture supernatants indicated an average of 1.5-fold higher activity was detected in the pBGP3-PaBG1b transformants compared to the vector-transformed control strains (Fig. 1-5 B). However, the immunoblot bands in the Western blot faded out on the 5th and 6th days, and the activity decreased likewise. On the 6th day, the activity of pBGP3-PaBG1b transformants fell to a negligible level, which implied that PaBG1b might be proteolytically degraded by extracellular proteases of *P. pastoris* (Fig. 1-5 C).

To increase the yield of PaBG1b, the inducible expression plasmid pPICZα was

employed. The construction of pPICZα-A-PaBG1b (Fig. 1-6 A), linearlization of the plasmid, transformation, and screening of *P. pastoris* transformants were performed as described in Materials and Methods. The c-myc and His_6 tags were placed at the C-terminus of PaBG1b. The proposed scheme for trimming of expressed polypeptide is shown in Fig. 1-6 B. Three pPICZα-A-PaBG1b *P. pastoris* transformants were chosen for expression test. After 4 days of methanol induction, the culture was stopped and the culture supernatant was harvested. Judged by the immunoblot bands on Western blot membrane, the CBB-stained bands on the SDS-PAGE gel, and the BG activity assay, two of pPICZα-A-PaBG1b transformants were found to express PaBG1b (Fig. 1-7A), and the transformant No. 3 was subsequently chosen for further analysis, as it displayed the highest yield of protein and BG activity.

1.2.3 Time-course analysis of expression of PaBG1b

To determine the optimum induction time, a time-course analysis of the expression of PaBG1b was done using pPICZα-A-PaBG1b transformant and the control strain transformed with the linearized empty vector. The expression of PaBG1b gradually increased after the methanol feeding, which was prolonged for 7 days. Results of Western blot analysis, CBB-stained SDS-PAGE gel, and the BG activity assay showed that the expression of PaBG1b reached the maximum on the 6th day of methanol feeding (Fig. 1-7 B).

Discussion

P. pastoris is an excellent host for heterologous expression of eukaryotic genes. In this study, PaBG1b was expressed successfully in *P. pastoris*, driven by constitutive and

inducible promoters, with the protein tags fused at either N- or C-terminus, and in episomal- and integration-type vectors. When the expression of PaBG1b using two different vectors, pBGP3 and pPICZα-A, was compared, the former was quicker and easier in verifying the expression of a specific gene, but the yield was relatively low and cost was higher since the persistent presence of expensive Zeocin was needed to maintain the expression plasmid in *P. pastoris,* whilst the latter needs more labor, but results in higher yield with less cost and probability of losing the expression unit as it is integrated to the genome of *P. pastoris*. Although PaBG1b has a KR dibasic amino acid pair on the N-terminus of its amino acid sequence which might be recognized by Kex2 resulting in the loss of its N-terminal tags in the Golgi apparatus, N-terminally c-myc-tagged recombinant PaBG1b was still detected in the culture supernatant of pBGP3-PaBG1b transformant on the 4th day. This implies that at least a portion of the recombinant PaBG1b was free from Kex2 cleavage at the second KR site of the precursor of the expression product (Fig. 1-4B). However, the expression product seemed to be vulnerable to degradation, because the immunoreactive band decreased on the 5th and 6th days of culture, and the bands with lower molecular weight were detected on the 6th day (Fig. 1-5 C). It was reported that the yield of expression product was reduced by extracellular proteases of *P. pastoris* (Clare et al., 1991b). In the case of pPICZα-A-PaBG1b transformant, the cell wash step prior to the methanol feeding might have removed the extracellular proteins, especially endogenous proteases of *P. pastoris* secreted in cell growth stage, such that the secreted PaBG1b was basically intact during the whole 7 days of induction.

As to the placement of N- and C- terminal tags, in the previous case of expression of a termite-derived BG, G1mgNtBG, it was reported that N-terminal tags were better than

C-terminal tags (Uchima and Arioka, 2012). However, in the induced expression of pPICZα-A-PaBG1b transformant, there was almost no sign of the proteolytic degradation of the C-terminal tags throughout 7 days of methanol feeding, as observed by Western blot as well as CBB-staining of sister SDS-PAGE gel. These facts might reflect the effect of removing endogenous protease(s) through cell was step before the methanol induction, or just imply that difference in the stability of tags having different placement to a given protein is case by case.

Overall the production of PaBG1b by pBGP3-PaBG1b transformant was pretty low, and the expressed product could not be observed on the CBB-stained SDS-PAGE gel. The yield of PaBG1b was improved by using pPICZα-A-PaBG1b transformant, but the total protein of the culture supernatant was less than 0.2 mg/ml. Optimization of the amount of daily-fed methanol could be an option to increase the yield. According to the user manual of Invitrogen, the methanol dosage for induction is recommended to be 0.5% for Mut^S stains and can be increased up to 3% without negative effects (Invitrogen, 2010). In this study, I tested both decreasing and increasing the methanol concentration (data not shown). However, no significant improvement of production was observed when the methanol concentration was decreased by half, and the expression of PaBG1b was abolished when the methanol concentration was doubled from 1% (v/v) to 2%. To maximize the output, optimizing both the G+C content and the codon usage of the coding sequence to those of *P. pastoris* could be effective approaches (Sinclair and Choy, 2002). Screening of transformants containing multiple copies of *pabg1b* genes or employing vectors harboring multiple copies of the expression cassette, and optimizing the culture and induction conditions might also be of help for increasing the expression level.

The cell growth of pPICZα-A-PaBG1b transformant in the YPG medium was at the normal rate, and it took 24 h to reach an OD_{600} of around 14-17 in both the pre-culture and main culture. However, the production speed of pPICZα-A-PaBG1b transformant was rather slow, which took 6 days to reach the maximum yield (Fig. 1-7B). The long induction time might be attributable to the low methanol utilization rate of the host strain. The expression host, *P. pastoris* KM71H used in this study, is phenotypically a Mut^S stain. In the Mut^S strain, *AOX1* gene encoding the primary alcohol oxidase was disrupted and the growth of Mut^S strain on methanol was much slower than with *AOX1* (Koutz et al., 1989). According to the Invitrogen user manual, the doubling time of log phase of Mut^+ strain in the methanol medium is 4-6 h, whereas that of Mut^S strain is approximately 18 h (Invitrogen, 2010). Therefore, it is reasonable that expression by Mut^S strains would have a longer induction time.

Another possible explanation for the long induction time is the poor oxygen supply in the flask culture. The expression of heterologous proteins in *P. pastoris* was found highly sensitive to aeration, and the productivity of shake-flask induction is usually less efficient than that of the controlled fermenter (Clare et al., 1991a). The user manual of Invitrogen recommends an agitation rate of 250-300 rpm for a baffled flask containing the induction culture with a volume of 1/5 to 1/10 of its size (Invitrogen, 2010). However, the shaking incubator used in this study has a limitation of 200 rpm and was usually operated at 150 rpm for a safety concern. As a comparison, two GH3 BGs heterologously expressed in *P. pastoris* KM71H strain using pPICZα (Kawai et al., 2003; Hong et al., 2007) and shaking flasks took a long induction time of 7 to 8 days. In contrast, additional two GH3 BGs expressed in X33 strain (Mut^+ phenotype) using pPICZα (Liu et al., 2012; Karkehabadi et al., 2014) took 3-4 days for methanol

induction. Thus employing the Mut^+ strain might be another viable option to shorten the induction time.

Table 1-1. Summary of PaBG1b amino acid sequence analysis.

GenBank accession number	LC125463
UniProt ID	BAU51446
Length of open reading frame (bp)	1509
Length (aa)	502
Length of mature polypeptide (aa)	482
Predicted signal sequence (aa)	20
Average molecular weight of mature region (kDa)	55.5
Isoelectric point	4.70
Catalytic acid/base	E196
Catalytic nucleophile	E406
Putative *N*-glycosylation sites	2 (N265 and N416)
Putative *O*-glycosylation site	1 (S315)

A

Source: Arakawa *et al*. (2016)

B

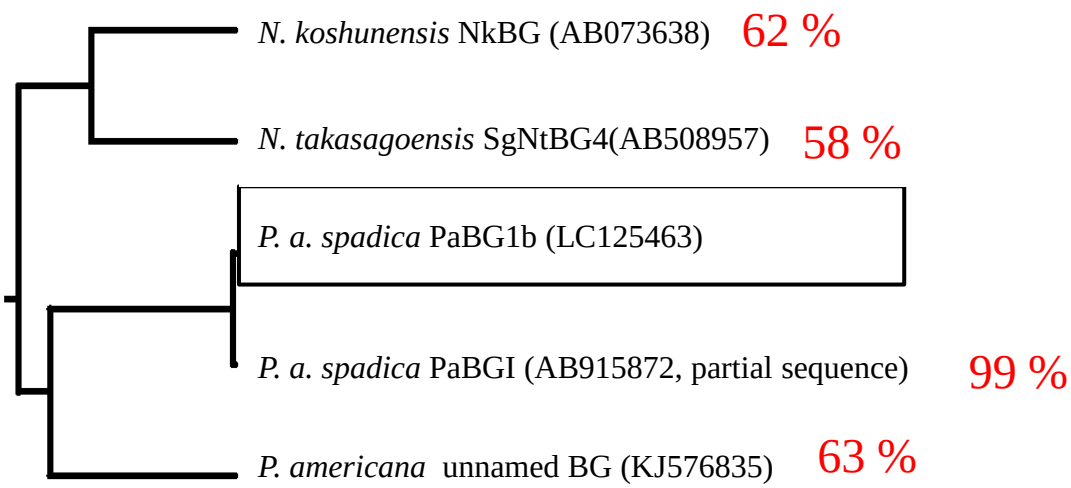


Fig. 1-1. Adult and nymph of Wood feeding cockroach *P. a. spadica* (A) and phylogenetic tree inferred from amino acid sequences of selected BGs with the highest homologies (E-value 0.0) towards PaBG1b by BLAST search (B)

In (B), the Genebank accession numbers of NkBG of lower termite *N. koshunensis*, SgNtBG4 of higher termite *N. takasagoensis*, PaBG1b of wood-feeding cockroach *P. a. spadica*, as well as the partial, deduced amino acid sequence of PaBGI of *P. a. spadica*, and the unnamed, hypothetical BG of omnivorous cockroach *P. americana* are given in parentheses. The phylogenetic tree mapping was done by ClustalW server (http://www.genome.jp/tools/clustalw/) based on the UPGMA method. Amino acid sequence identities towards PaBG1b are shown by numbers in red.

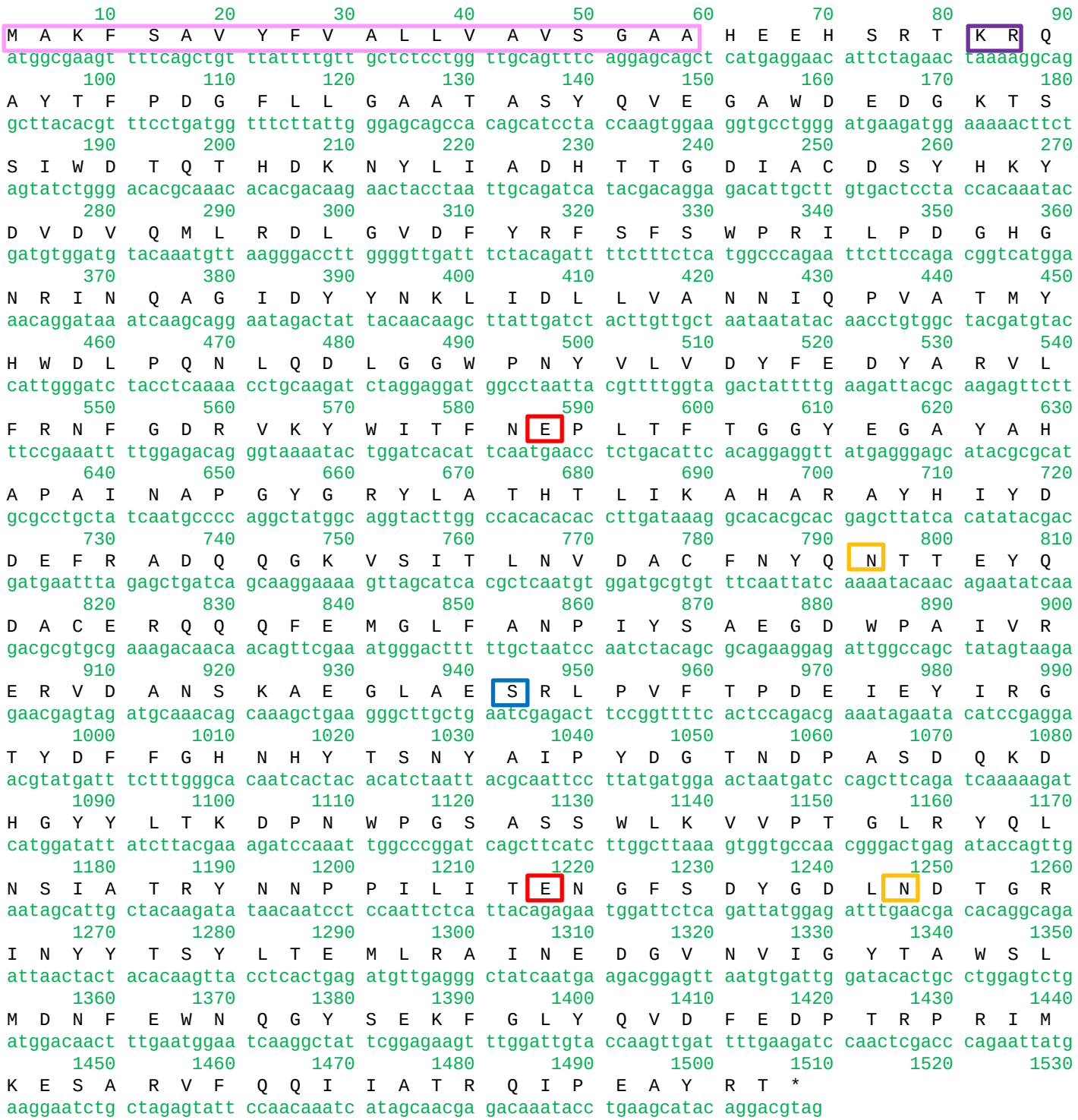


Fig. 1-2. Nucleotide and deduced amino acid sequences of PaBGb1

Nucleotide sequence and numbering thereof are shown in green at the top and bottom rows of each line. Amino acid sequence is shown in the middle row in black capital letters. The putative signal sequence and the KR dibasic amino acid pair at the N-terminus of PaBG1b are boxed in pink and purple, respectively. Putative catalytic acid/base (E196) and nucleophile (E406) residues are boxed in red. Two potential *N*-glycosylation sites (N265 and N416) and one *O*-glycosylation site (S315) are boxed in yellow and blue, respectively.

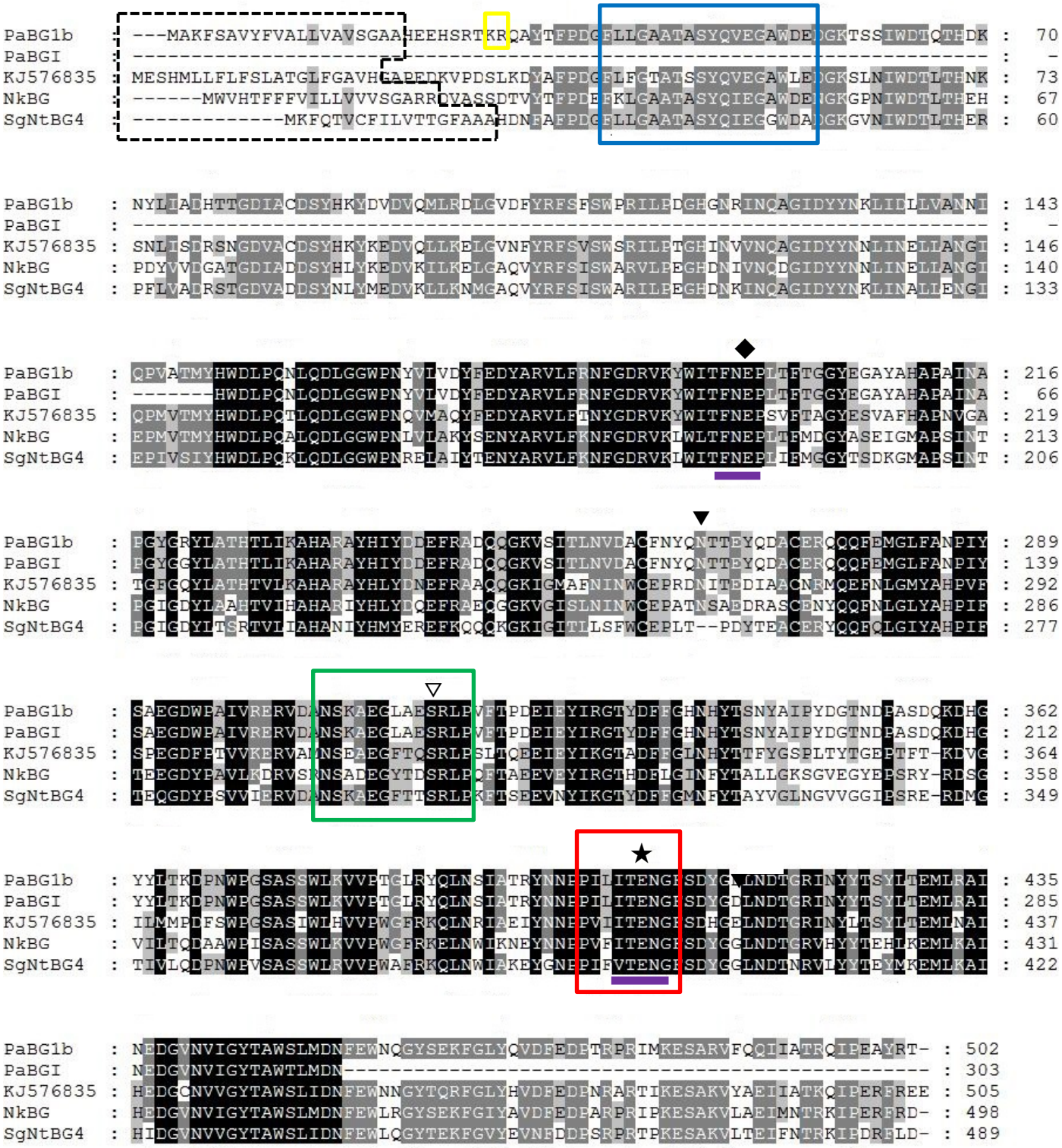


Fig. 1-3. Multiple alignment of PaBG1b and selected representative GH1 BGs

Amino acid sequences of three insect-origin BGs in Fig. 1-1B are aligned with PaBG1b. The unnamed BG of *P. americana* is indicated by its GenBank accession number (KJ576835). Catalytic acid/base and nucleophile residues (E196 and E406 in PaBG1b, respectively) are marked by a diamond (◆) and an asterisk (★). Two potential *N*-glycosylation sites (N265 and N416) and one *O*-glycosylation site (S315) are indicated by solid (▼) and open (▽) arrowheads, respectively. The putative signal sequence is boxed by dash lines. The GH1 signature described by the PROSITE motif PS00572 is boxed in red. The conserved GH1 N-terminal signature (Tsukada *et al*., 2006) is boxed in blue, whereas an internal conserved sequence among NtBGs (Tokuda *et al*., 2009) is boxed in green. Two conserve motifs (Verdoucq *et al*., 2004), i.e., 'TXNEP' and 'I/VTENG', are denoted by purple bars. The KR dibasic amino acid pairs (Kex2 recognition site) of PaBG1b is boxed in yellow.

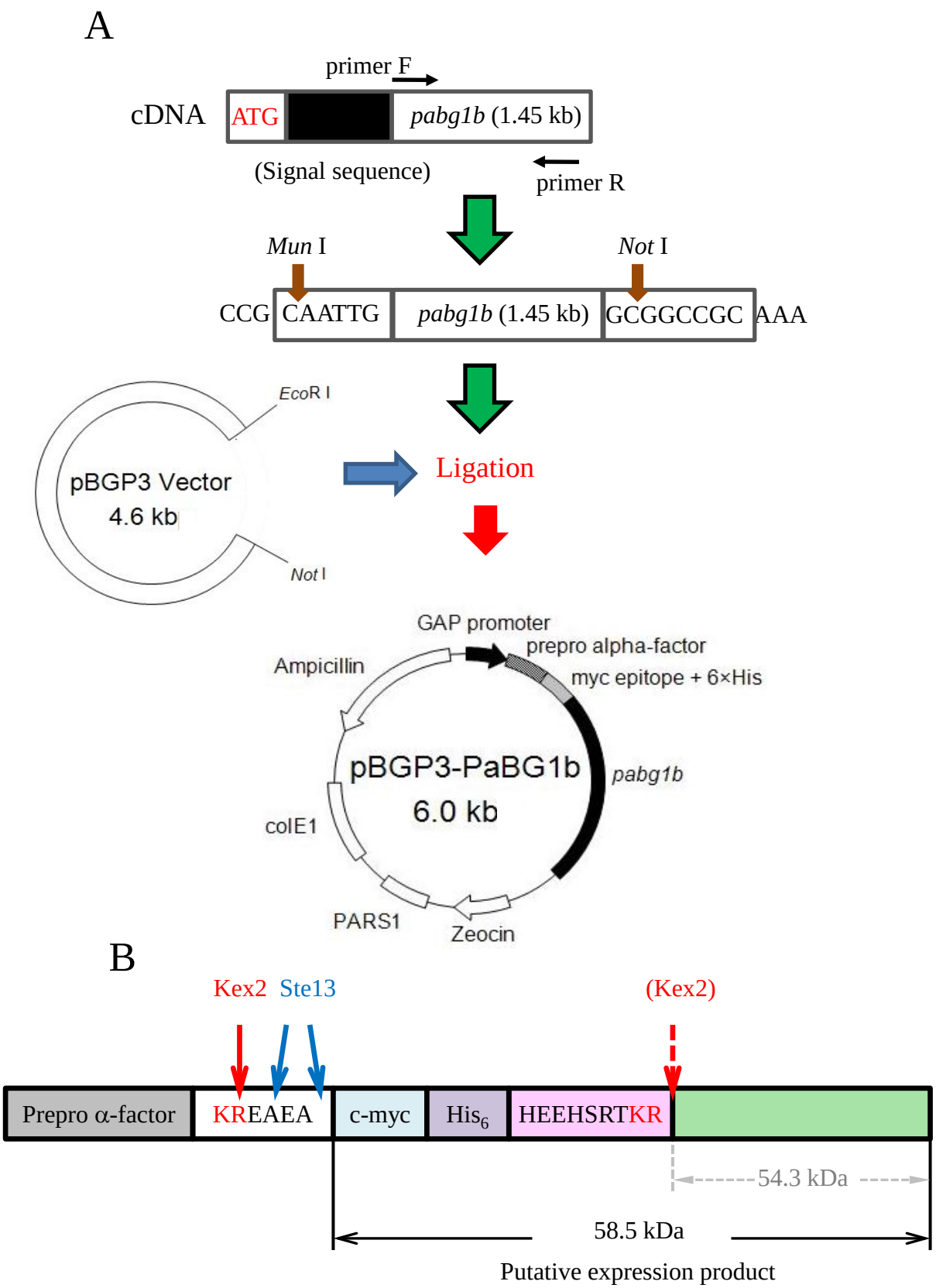


Fig. 1-4. Construction of pBGP3-PaBG1b (A) and proposed scheme for the preocessing of the precursor of the expression product (B)

Protein tags are placed at the N-terminus of PaBG1b. PaBG1b might suffer cleavage at the second Kex2 recognition site (KR dibasic amino acid pair) present in the putative mature region of PaBG1b, which is located 33 amino acid residues (4.2 kDa in size) downstream of the Ste13 recognition site.

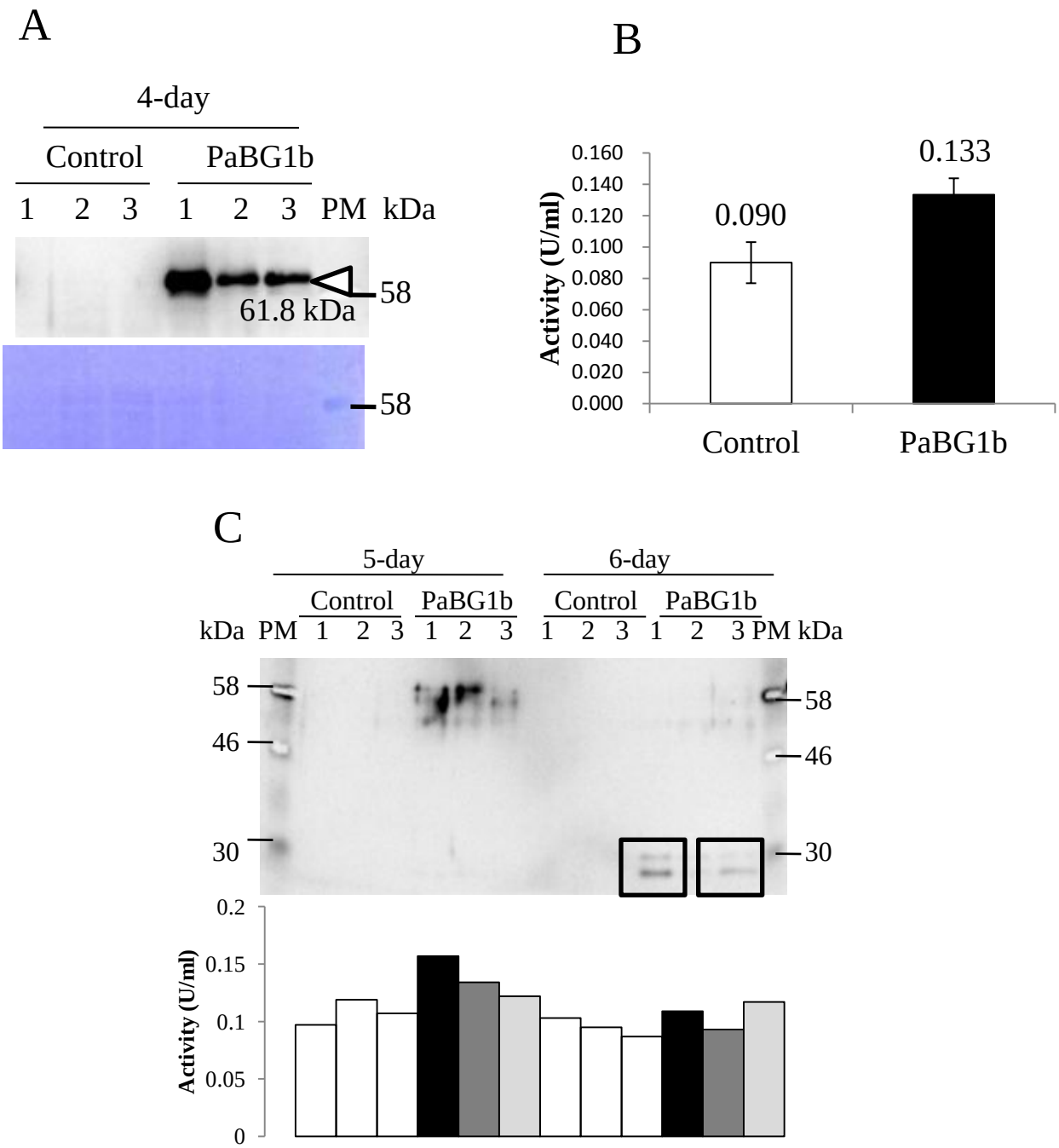


Fig. 1-5. Expression of N-terminally c-myc- and His_6-tagged PaBG1b by pBGP3-PaBG1b

(A) Western blot analysis (top) and CBB-stained SDS-PAGE gel (bottom) of the culture supernatants of *P. pastoris* transformants after 4-day culture. PM: protein marker; Control 1, 2, 3: negative control strains transformed with the empty vector; PaBG1b 1, 2, 3: transformants harboring PaBG1b expression cassette. Arrowhead indicates the band of PaBG1b. (B), BG activity of the transformants after 4-day culture. *p*NPG was used as the substrate for BG activity assay in this study unless otherwise noted. Data are means ± SD of three independent experiments. (C), Western blot analysis of the 5-day and 6-day culture supernatants. The lane numbers are the same as those in (A). Squares indicate the PaBG1b fragments with c-myc tag, presumably due to proteolytic degradation.

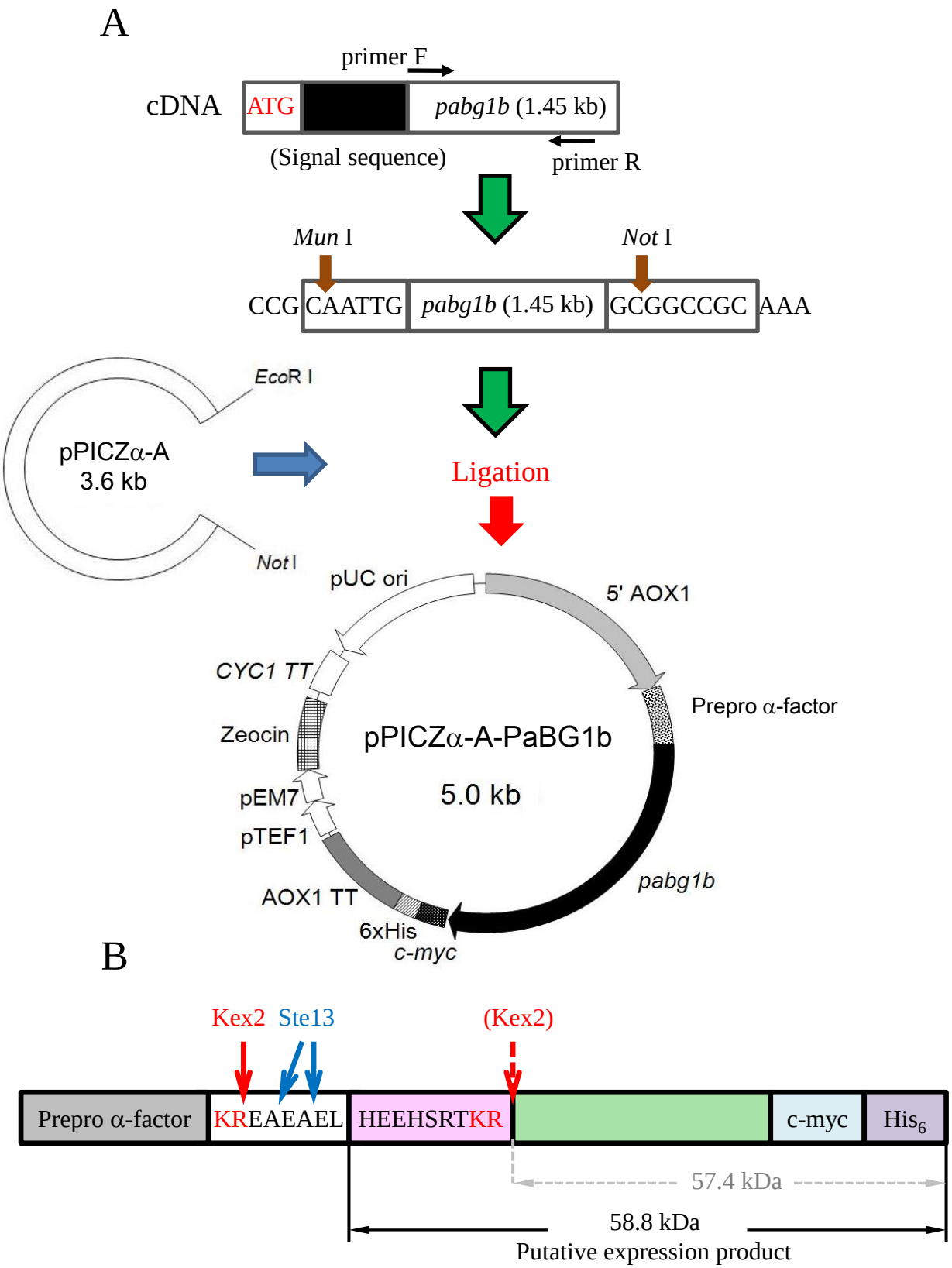


Fig. 1-6. Construction of pPICZα-A-PaBG1b (A) and proposed scheme for trimming of the precursor of the expressed product (B)

Protein tags are placed at the C-terminus of PaBG1b. PaBG1b might suffer cleavage of Kex2 due to a KR dibasic amino acid pair 11 residues downstream of the Ste 13 recognition site (two amino acids derived from the *Mun* I site and 9 amino acids of the N-terminal of the mature region of PaBG1b).

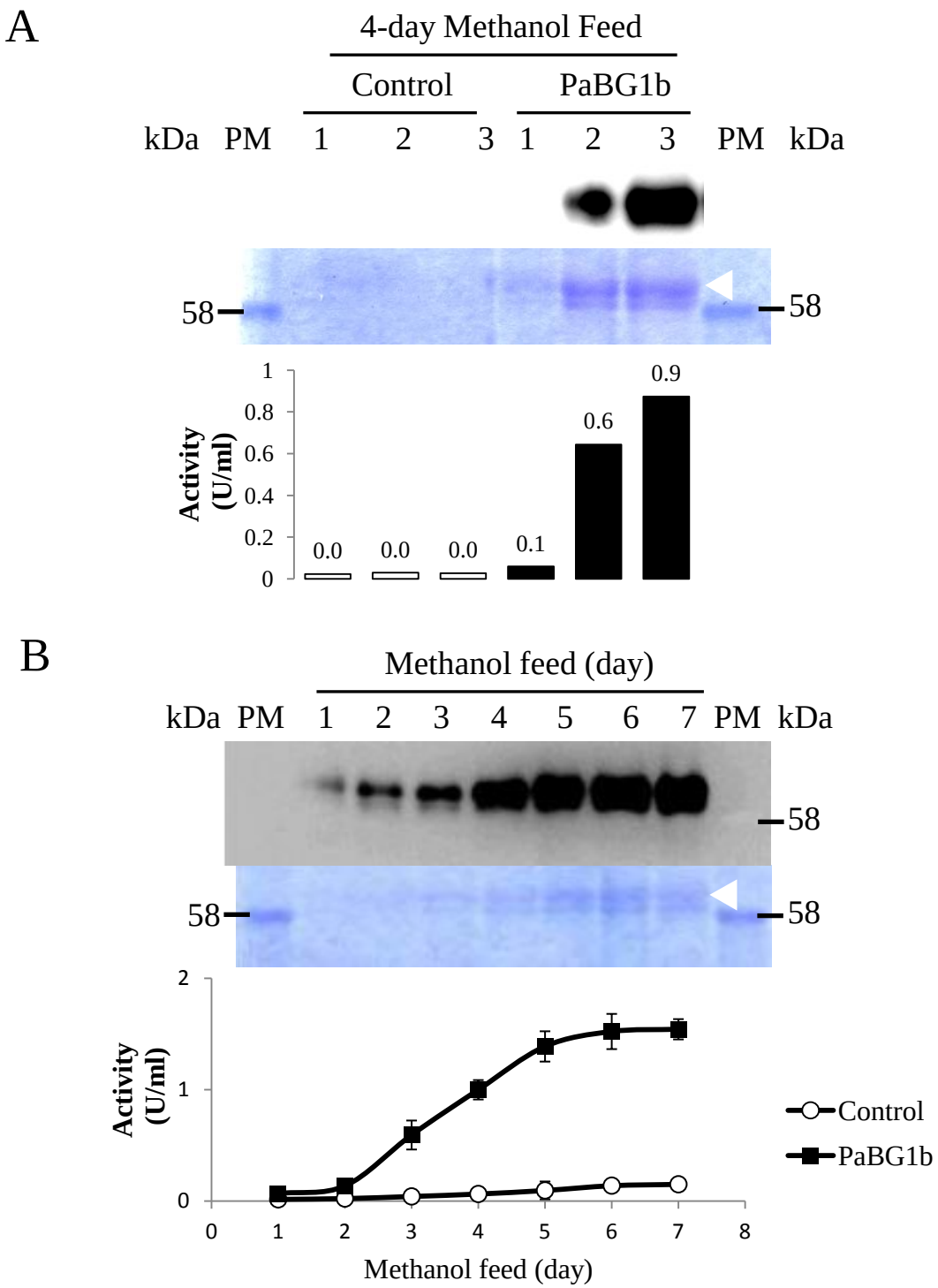
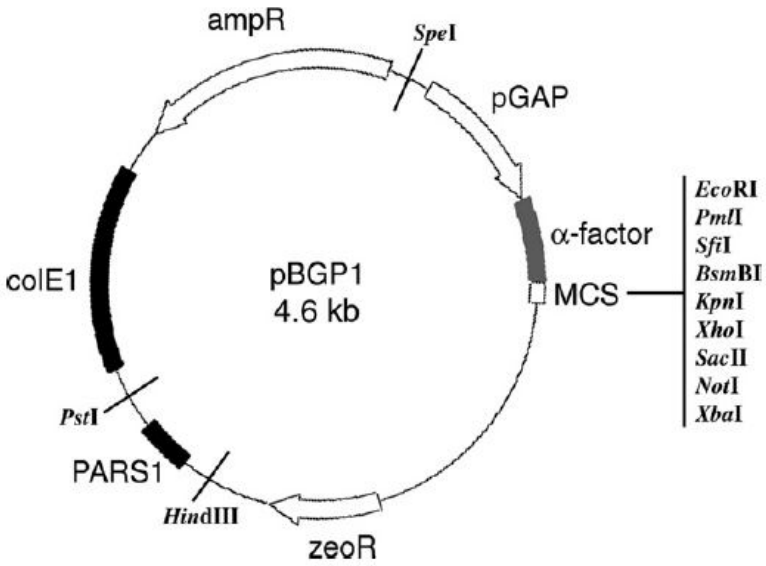


Fig. 1-7. Expression of C-terminally c-myc- and His_6-tagged PaBG1b by pPICZα-A-PaBG1b

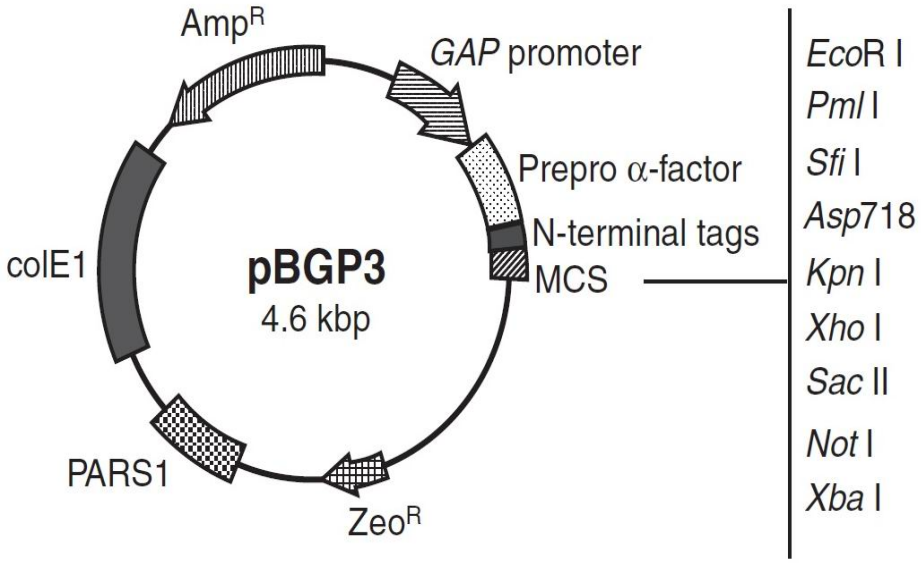
(A) The expression of PaBG1b in the pPICZα-A-PaBG1b transformants on the 4th day of methanol feeding. (B) Time-course analysis of the culture supernatant of the pPICZα-A-PaBG1b transformant during the methanol induction period. Top and middle panels in (A) and (B) show the representative results of Western blot and SDS-PAGE analysis. PM: protein marker. Open arrowheads indicate the position of PaBG1b on SDS-PAGE gel, with an apparent molecular weight of 61.8 kDa. In the bottom panel, BG activities of the pPICZα-A-PaBG1b transformant (closed squares) and the negative control strain transformed with the linearized empty vector (open circles) are shown. Data are means ± S.D. of three independent experiments.

A



Source: Lee, *et al*., 2005

B



Source: Uchima and Arioka, 2012

Fig. S1-1. Plasmid maps of expression vectors pBGP1 (A) and pBGP3 (B).
pGAP, *GAP* (glyceraldehydes-3-phosphate dehydrogenase) promoter; α-factor in (A) and prepro α-factor in (B), prepro α-factor sequence of *S. cerevisiae*; MCS, multiple cloning site; zeoR, Zeocin-resistance gene; PARS1, *P. pastoris* autonomous replication sequence; colE1, bacterial replication origin; Amp^R, ampicillin-resistance gene.

Chapter 2

Purification and biochemical characterization of PaBG1b

2.1. Introduction

2.1.1 Protein purification through Ni-NTA system

Expression of tagged recombinant protein combined with immobilized metal affinity chromatography (IMAC) is a powerful strategy to achieve high specificity, single step purification of target protein. The principle of IMAC is based on the specific coordinate covalent bond of amino acids, particularly histidine, to metals (mostly nickel cation). Imidazole is a diazole which is usually employed for binding to the nickel ions and disrupting the binding of dispersed histidine residues in non-tagged background proteins or competes with the polyhistidine-tagged proteins. Proteins bound to nickel column can also be eluted by other methods such as lowering pH within the range of 2.5–7.5 as suggested by the company manual of Amersham Biosciences (Amersham Biosciences, 2004). According to the company manual of QIAGEN, the histidine residues in the His_6 tag have a pK_a of approximately 6.0 and become protonated when pH is decreased to 4.5–5.3 (QIAGEN, 2003). Under these conditions the His_6-tagged protein dissociates from the Ni-NTA resin. Reagents such as EDTA or EGTA can also be employed for elution of His_6 tagged proteins, but these reagents can chelate the nickel ions and remove them from the NTA groups, which leads to the His_6-tagged protein being eluted as a protein–metal complex (QIAGEN, 2003). Overall, optimal purification conditions such as pH, type of buffers, and imidazole concentrations vary for each protein and must be optimized empirically.

2.1.2 Evaluating BGs in view of industrial cellulose conversion

As the hydrolysis product of *p*NPG can be directly quantified in the colorimetric assays, and *p*NPG is more economical and readily available than other artificial

substrates (Scharf et al., 2010), *p*NPG is popularly used as an artificial substrate in the activity assay of BG. However, the preferences to cellobiose and/or aryl-linked glucosidic substrates vary among BGs. The hydrolysis activities of a specific BG toward cellobiose (previously referred to as cellobiase) and *p*NPG can be presented in three types: the activity on cellobiose is (1) significantly higher (Scharf et al., 2010; Wu et al., 2012; Uchiyama et al., 2015) or (2) comparable (Peralta et al., 1990; Koffi et al., 2012; Meleiro et al., 2014), or (3) markedly lower (Zverlov et al., 1997; Venturi et al., 2002; Joo et al., 2009; Kalyani et al., 2012; Uchiyama et al., 2013; Kaur and Chadha; 2015) than that on *p*NPG. Some BGs even possess activity on *p*NPG but are incapable of hydrolyzing cellobiose (Cicek et al., 1999; Marques et al., 2003; Lee et al., 2015; Nakajima et al., 2016; Watanabe et al., 2016). The differences in the affinities among BGs for a particular substrate are reflections of the physiological functions and locations of those enzymes (Woodward and Wiseman, 1982). The high activity on aryl-β-glucosides of BGs might imply that the hosts of those BGs are capable of degrading flavones and anthocyanins that constitute pigments in flowers (Woodward and Wiseman, 1982), complex polysaccharides from plant residues (Uchiyama et al., 2013), or glucosidic flavor precursors in fruits (Günata and Vallier, 1999). In view of the production of bioethanol from cellulose, it is important to evaluate the kinetic parameters of BGs towards cellobiose. Besides, BGs capable of dealing with high DP cellooligosaccharides are of value, as they can improve the saccharification of cellulose via degrading cellodextrins generated by the actions of EG and CBH (Shewale, 1982).

2.2. Results

2.2.1 Purification of PaBG1b

2.2.1.1 Ammonium sulfate precipitation

The preliminary ammonium sulfate precipitation experiment was performed as shown in Fig. 2-1A. The results demonstrated that PaBG1b achieved the highest recovery of 72.1% of the total activity at 75% saturation of ammonium sulfate (Fig. 2-1B). Therefore, the ammonium sulfate precipitation in the succeeding operation was conducted in a stepwise manner: contaminant proteins in the culture supernatant of the pPICZα-A-PaBG1b were removed first by ammonium sulfate precipitation at 45% saturation, and then PaBG1b was precipitated at 75% saturation.

2.2.1.2 Purification of PaBG1b by Ni-NTA column chromatography

2.2.1.2.1 Purification of PaBG1b under standard condition

To optimize the purification condition, small scale Ni-NTA purification was conducted under the standard condition described in the company manual of Invitrogen (Invitrogen, 2010). Unfortunately, although PaBG1b was successfully bound and eluted from the column, no BG activity was detected in the elution fractions containing 250 mM imidazole (Fig. 2-2A). To examine if high imidazole concentration inhibited the activity of PaBG1b, lower concentration of imidazole was tested (Fig. 2-2B). The results showed that PaBG1b could be eluted by 125 mM imidazole. However, the PaBG1b-containing eluate only showed negligible activity, even after dilution by 5-folds with 50 mM sodium phosphate buffer (pH 8.0) prior to enzyme assay.

To remove imidazole, the PaBG1b-containing eluates were subjected to dialysis against the enzyme assay buffer (50 mM sodium acetate, pH 5.5) overnight at 4°C. The pH of the resultant product was confirmed to be pH 5.5. However, no activity was detected in the dialysis product (data not shown). Next, removal of imidazole was

conducted by buffer exchange as follows: the elution fractions (ca. 3 ml) were concentrated by using disposable ultrafiltration device to less than 1 ml, then 20 ml of the enzyme assay buffer were added, and the mixture was then concentrated again to less than 1 ml. This procedure was repeated three times. The pH of the resultant solution was confirmed to be 5.5. However, there was no activity detected (data not shown).

2.2.1.2.2 Investigation the effect of imidazole on PaBG1b

The loss of activity after Ni-NTA purification might be because of either inhibition by imidazole, or deviation of assay pH from the optimum pH for PaBG1b, since the elution buffer (50 mM sodium phosphate buffer, 300 mM NaCl containing 125 mM or 250 mM imidazole, pH 8.0) has strong buffer capacity. To examine these possibilities, the pH profile of PaBG1b should be determined at first. By the activity assay of the culture supernatant of PaBG1b-expressing transformant, conducted in a series of 50 mM Britton-Robinson buffers (H_3BO_3: CH_3COOH: H_3PO_4=1:1:1; pH was adjusted by NaOH; Britton and Robinson, 1931) at pH ranging from 3.0 to 9.0, the optimum pH of the culture supernatant was determined to be pH 5.0. The relative activity was 92% at pH 6.0, and almost zero at pH 8.0 (Fig. 2-3A). In the pH stability experiment, the culture supernatant was mixed with the same volume of a series of 50 mM Britton-Robison buffers with their pH ranging from 3.0 to 9.0, left on ice for 1 h, and the activities of the resultant mixture were assayed. The activity of mixture was stable at pH 6.0, and over 75% of the activity remained from pH 3.0 to pH 8.0, which implied that shifting the ambient pH to 8.0 would not irreversibly inactive PaBG1b (Fig. 2-3B).

To clarify whether imidazole inhibits PaBG1b, an experiment was designed as follows (Fig. 2-3C). Firstly, the binding buffers (50 mM sodium phosphate, 300 mM

NaCl) containing 0 or 500 mM imidazole, pH of which was pre-adjusted to 6.0 or 8.0, were prepared. Next, the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant was mixed with the same volume of these buffers to make the enzyme solution containing 0 or 250 mM imidazole at pH 6.0 or 8.0. The mixtures were left on ice for 1 h, and the activity assay was performed under the routine condition. The results demonstrated that imidazole inhibited PaBG1b. As described above, at pH 6, PaBG1b retained 92% of its maximum activity. However, when pH of the imidazole solution was pre-adjusted to 6.0 and mixed with the culture supernatant (Fig. 2-3C, center bar), the activity of the mixture was only the half of the control reaction without imidazole addition (Fig. 2-3C, left bar). Furthermore, when pH of the imidazole solution was set at 8.0, which corresponded to the standard elution condition of Ni-NTA purification, the activity was even lower (9.1% of the relative activity; Fig. 2-3C, right bar, boxed by a dotted rectangle in red).

In the design of the BG activity assay, to assure that the reaction is conducted at pH 5.5, the substrate (*p*NPG) is dissolved in the enzyme assay buffer (50 mM sodium acetate buffer, pH 5.5), and mixed with one-tenth volume of the enzyme solution. It was assumed that the 10-fold volume of enzyme assay buffer could adjust the pH of enzyme-substrate mixture to an acidic pH. However, if the buffering action of imidazole solution is too strong to offset that of the enzyme assay buffer, the pH of the enzyme reaction mixture should deviate from pH 5.5 to an unfavorable (neutral or alkaline) pH for PaBG1b, and henceforth PaBG1b appears to be inhibited by imidazole. To clarify this possibility, the elution buffer of Ni-NTA purification (the binding buffer containing 250 mM imidazole, pH 8.0) was diluted with the enzyme assay buffer and the pH of the resultant mixture was examined. The results showed that the buffering action of enzyme

assay buffer was rather poor. To shift the pH to 6.0, the elution buffer had to be diluted by 40- to 50-fold with the enzyme assay buffer (Fig. 2-3D). This result indicates that the low activity of culture supernatant mixed with the imidazole solution set at pH 8.0 was at least partially due to the shift in pH to the alkaline condition (Fig. 2-3C, right bar), since in that experiment imidazole solution was only diluted by 10-fold with the enzyme assay buffer.

Based on this result, I re-organized the experiment to confirm the inhibitory effect of imidazole (Fig. 2-4). In this experiment, the culture supernatant was first mixed with the same volume of '2x imi' buffer (50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0) to generate a mixture containing 250 mM imidazole (mixture I in Fig. 2-4A). The mixture was left on ice for 1 h and diluted by 5-fold with the enzyme assay buffer to generate the mixture II. Mixture II was subsequently mixed with the 10-fold volume of the substrate solution to start the enzyme reaction (Fig. 2-4 A). In this protocol, '2x imi' buffer was finally diluted by 100-fold with the enzyme assay buffer, which is equivalent to 50-fold dilution of the elution buffer. The resultant pH of the reaction mixture (mixture III in Fig. 2-4A) was 5.65. The relative activity was compared with the controls where '2x imi' buffer was replaced by the solutions without imidazole, i.e. distilled water or the binding buffer of Ni-NTA purification. The results demonstrated that the presence of imidazole severely, and probably irreversibly, abolished over 80% of the activity of PaBG1b, and this effect was not due to the shift of pH (Fig. 2-4B).

It had been shown that the Ni-NTA purification is still applicable when the operating pH is lowered down to pH 7.0 (Del Pozo et al., 2012). To clarify whether imidazole was capable of irreversibly inhibiting PaBG1b under the alkaline condition, I conducted the

purification at pH 6.0, a pH where PaBG1b remained active. The results demonstrated that although PaBG1b still bound to and was eluted from the column at pH 6.0, nearly equal amount of PaBG1b was eluted by 125 and 250 mM imidazole (Fig. 2-5A). This is presumably because of two reasons: firstly, the histidine residues in the His_6 tag have a pK_a of approximately 6.0 (QIAGEN, 2003) and were therefore partially protonated at pH 6.0, which led to a weak binding with the column. Secondly, imidazole has a pK_a of 7.1 (Li and Byers, 1989) and at pH 6.0 it was mostly protonated and thus the affinity with Ni-NTA agarose weakened. Then I performed the purification of PaBG1b at pH 6.0, eluting with 250 mM imidazole. The results showed that PaBG1b was purified in an active form, and 14.8% of the total activity was recovered (Fig. 2-5B). This might indicate that imidazole does not irreversibly inhibit the activity of PaBG1b at pH 6.0. However, it was also found that when imidazole was not removed by dialysis, purified PaBG1b gradually lost its activity and precipitated after repeated freezing and thawing. The activity of precipitated PaBG1b could not be recovered through dialysis or buffer exchange (data not shown).

2.2.1.2.3 Elution of PaBG1b by EDTA and pH

According to the QIAGEN manual, among the three elution methods of Ni-NTA chromatography, i.e. imidazole, EDTA, and pH, imidazole is the mildest and thus recommended to be used in the purification under native conditions, because the protein might be damaged by reduction in pH, and removal of metal ions in the eluate might have adverse effect on the purified protein (QIAGEN, 2003). As imidazole might have some unfavorable effect on PaBG1b, EDTA and pH elution methods were also tested. In the EDTA elution experiment, the buffers containing 1 mM and 10 mM EDTA were

used as the wash and elution buffers, respectively. Although PaBG1b was successfully purified, PaBG1b in the elution fractions showed very low activity (Fig. 2-6A). The activity was recovered after dialysis against the enzyme assay buffer (Fig. 2-6 B and C), but the recovery of total activity was around 21%.

Eluting PaBG1b by lowering pH was also tested. The culture supernatant was mixed with the same volume of binding buffer, loaded onto the Ni-NTA column, and washed by the same buffer whose pH was 6.0, and then eluted by 50 mM sodium acetate buffer, pH 5.0. The results showed that PaBG1b could be purified through pH elution (Fig. 2-7). However, the recovery of total activity was only 16.2%.

2.2.1.3 Purification of PaBG1b by anion exchange chromatography

To compare the purification efficiency and specific activities of purified products, the anion exchange chromatography was employed for purification of PaBG1b from the culture supernatant. The culture supernatant was subjected to ammonium sulfate precipitation and the resultant crude enzyme solution was loaded onto a HiTrap DEAE FF column. The elution was conducted by a linear gradient of NaCl. The results showed that PaBG1b could be purified through anion exchange chromatography (Fig. 2-8), but with less purity than PaBG1b purified by Ni-NTA affinity chromatography as shown in Fig. 2-5 and Fig. 2-7.

A summary of different purification methods is listed in Fig. 2-9. The results showed that the purification efficiency of Ni-NTA system and the specific activity of the purified products thereof were not better than those of the anion exchange chromatography. However, judged by the CBB-stained SDS-PAGE gel, the anion exchange chromatography product was not pure as that by Ni-NTA system, which

implies that the Ni-NTA purification techniques applied might lead to the partial loss of activity of PaBG1b.

2.2.1.4 Introducing of Tris in Ni-NTA purification

Scharf et al. reported that recombinant *Rf*BGluc-1, a GH1 BG from *R. flavipes*, displayed nearly 1.5-fold greater activity in the sodium acetate buffer at pH 7 than in the potassium phosphate buffer (Scharf et al., 2010). To investigate whether the phosphate buffer employed for Ni-NTA purification in this study affected the activity of PaBG1b, Tris-HCl buffer was used to substitute the sodium phosphate buffer. In the Ni-NTA purification, 50 mM Tris-HCl buffer (pH 8.0) containing 300 mM NaCl was used as the binding buffer and the same buffers containing 5 mM and 100 mM imidazole were employed as wash and elution buffers, respectively. The results showed that the resultant product had significant activity (Fig. 2-10). Therefore, Tris-HCl buffer was chosen for subsequent Ni-NTA purification of PaBG1b. Besides, it was found that if imidazole was not removed, the purified product gradually lost its activity after repeated freezing and thawing, which was similar to the phenomenon mentioned in 2.2.1.2.2 (page 52). If imidazole was removed by dialysis, this phenomenon could be avoided.

2.2.1.5 Determination of the final strategy for purification of PaBG1b

Finally, a large scale purification of PaBG1b was conducted. A total volume of 300 ml of 6-day methanol induction culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant was subjected to the ammonium sulfate precipitation, and the resultant precipitate was re-suspended in 50 mM Tris-HCl buffer, pH 8.0. Ammonium sulfate was removed by dialysis against the same buffer at 4°C. Then Ni-NTA

purification was conducted using the column equilibrated by 50 mM Tris-HCl buffer, 300 mM sodium chloride, pH 8.0. Wash and elution were done using the same buffer containing 5 and 100 mM of imidazole, respectively. The result is shown in Fig. 2-11 and summarized in Table 2-1. This result demonstrates that by simply replacing sodium phosphate buffer with Tris-HCl, purification by Ni-NTA system achieved a comparable efficiency (recovery of 42.9 %) to that by anion exchange chromatography. Judged by the specific activity, the impairment of PaBG1b activity during Ni-NTA purification had been circumvented. The sample was subsequently subjected to the anion exchange chromatography (Fig. 2-12 A and B). As the Ni-NTA purification product already had high purity, only limited increase in the specific activity was obtained by the anion exchange chromatography (Table 2-1). In the end of this study, 3 mg of PaBg1b was obtained by this process and the purity of the resultant product was check by CBB staining (Fig 2-12C).

The reason why the Ni-NTA purification products eluted with imidazole in two types of buffers (i.e. sodium phosphate buffer vs Tris-HCl buffer) demonstrated different results remained obscure, and this phenomenon prompted me to investigate the effect of imidazole and Tris towards PaBG1b in the following study (2.2.2.8 and 2.2.2.9).

2.2.1.6 Post-translational modification analysis

PaBG1b has two potential *N*-glycosylation sites (Fig. 1-2). To analyze the post-translational modification of PaBG1b, the purified protein was treated with glycopeptidase F (GPF) or endoglycosidase H (Endo H) (Fig. 2-13A). GPF removes the entire *N*-glycan from the protein, whereas EndoH leaves the innermost

N-acetylglucosamine (GlcNAc) residue (Lee et al., 2009) as shown in the scheme of Fig. 2-13B. Compared to the mobility of the untreated control sample, both deglycosylated products showed reduced sizes. The untreated control exhibited a smear band with the size of approximately 61.8 kDa, whereas the GPF-treated product showed a dominant band with the size of 58 kDa and two bands with smaller sizes, and the Endo H-treated product displayed two distinct bands with molecular sizes of 58 kDa (upper band) and 56.3 (lower band). As mentioned in Chapter 1 (page 29), the mature region of recombinant PaBG1b might be cleaved carboxyl to the KR dibasic site by Kex2 during the post-translational processing, and two types of processing products (different in 1.4 kDa) with distinct N-terminal sequences could be generated when the Kex2 digestion is incomplete (Fig. 2-13C). Assuming that the upper (58 kDa) band of GPF- or Endo H-treated products were derived from the full length recombinant PaBG1b (and the lower (56.3 kDa) band was derived from PaBG1b excised by Kex2 at its N-terminus), the increment in size by *N*-glycosylation was 3.8 kDa (61.8 minus 58), which means that each *N*-glycan contributed 1.9 kDa in size on average. As a comparison, Tull et al. heterologously expressed the alkalophilic *Bacillus* α-amylase (ABA) in *P. pastoris* and found that seven sites were glycosylated and totally contributed to an increase of approximately 11.4 kDa (Tull et al., 2001), which implied an average increment of 1.63 kDa per site. In the case of G1sgNtBG1 of the higher termite *N. takasagoensis*, a GH1 BG which had only one glycosylation site and heterologously expressed in *P. pastoris*, the glycan was supposed to contribute to an increase of 2 kDa (Akemi Uchima et al., 2013). Therefore, the size increase of PaBG1b by *N*-glycosylation was in the reasonable level.

To further clarify the identity of the expression product, N-terminal amino-acid

sequencing was performed. As purified PaBG1b was presented in a smear band (at least two closely-overlapping bands of 61.8 kDa and a smaller protein), they were sent for sequencing as a whole. The E(K/L)H and QAY sequences were identified (Supplemental File 1 in page 170-172), the former of which might be derived from the correct processing at the first KR dibasic amino acid pair, whilst the latter might be generated by Kex2 cleavage at the second KR dibasic amino acid pair (Fig. 2-13C). These results indicate that the incomplete cleavage at the second Kex2 recognition site actually occurred.

2.2.1.7 Native PAGE

To determined whether the recombinant PaBG1b is a monomer, 0.01 mg of purified product was resolved by Native PAGE, in which condition the SDS and 2-mercaptoethanol (2-ME) were removed from the regents used in SDS-PAGE. A sample was boiled as an indicator denatured PaBG1b. However, the intact PaBG1b displayed a long smear on the gel, whereas the boiled PaBG1b presented a distinct band with a size of approximately 110 kDa (Fig 2-13 D).

2.2.2. Biochemical characterization of PaBG1b

2.2.2.1 Optimum temperature and thermostability

To investigate the optimum temperature of PaBG1b, the purified protein was subjected to enzyme assay at different temperatures, and the relative activities were calculated. The results show that PaBG1b displayed the highest activity when assayed at 45°C (Fig. 2-14A), which is 15°C lower than the native enzyme (Arakawa et al., 2016). However, it is notable that the assay of the native enzyme was conducted using

cellobiose as the substrate, and reacting for 5 min (Arakawa et al., 2016), which is much shorter than this study. At 60°C, there was only 38% of the relative activity left. In the evaluation of thermostability, PaBG1b retained over 87% of activity after incubation at 50°C for 30 min prior to the enzyme assay (Fig. 2-14B), which is similar to the thermostability of the native enzyme, as the thermal tolerance of the native enzyme was described as '(the native enzyme) showed thermal stability up to 50°C for 30 min of pre-incubation' (Arakawa et al., 2016). At temperatures higher than 50°C PaBG1b rapidly lost its activity. These results indicate that PaBG1b is a mesophilic enzyme.

2.2.2.2 Optimum pH and pH stability

Generally, the optimum pH of the microbial (especially fungal) BGs is within the range of pH 5.0 to 6.0 (Woodward and Wiseman, 1982; Bhatia et al., 2002; Eyzaguirre et al., 2005). PaBG1b exhibited the highest relative activity at pH 5.5 (Fig. 2-14C), which was within the range of the optimum pH of the majority of BGs, and compatible with the cellulose hydrolysis processed by *T. reesei*-origin cellulases (described in page 12). The pH stability experiment indicated that PaBG1b was stable (retaining over 80% of its activity) at the pH range of 5.5-7.0 (Fig. 2-14D). The optimum pH of the native PaBG1b was reported to be pH 5.0, and it retained over 65% of activity between pH 4.0 and 6.5 (Arakawa et al., 2016). Thus the recombinant PaBG1b has slight difference with the native enzyme in terms of the optimum pH and pH stability.

2.2.2.3 Substrate specificity

A variety of aryl-glycosides as well as saccharides were employed in the substrate specificity analysis of PaBG1b, and the results are shown in Table 2-2. For

aryl-glycosides, the most preferred substrate for PaBG1b was *p*-nitrophenyl-β-D-fucopyranoside (*p*NPFuc), followed by *p*NPG, *p*-nitrophenyl-β-D-galactopyranoside (*p*NPGal), and *p*-nitrophenyl-α-L-arabinofuranoside (*p*NPA*f*). PaBG1b could not hydrolyze *p*-nitrophenyl-β-D-*N*-acetylglucosamine (*p*NPGlcNAc). For saccharides, the most preferred substrate was laminaribiose (β-1,3 linkage), followed by cellobiose, cellotriose, and cellohexaose. The relative activity towards sophorose (β-1,2 linkage; 35.9% of activity compared to laminaribiose) was 8-fold higher than that of gentiobiose (β-1,6-linkage; 4.2%). PaBG1b showed weak activity towards lactose (3.7%) and salicin (1.9%). Overall, these results indicate that PaBG1b has a broad substrate specificity. PaBG1b failed to show activity towards sucrose and maltose, which showed that PaBG1b does not hydrolyze either α-1,2 or α-1,4 linked glycosides. Activities towards polysaccharides, i.e. Avicel, carboxymethyl cellulose (CMC), and laminarin were not detected.

2.2.2.4 Transglycosylation analysis by thin layer chromatography

To investigate the transglycosylation activity of PaBG1b, sugar substrates were reacted with PaBG1b and subjected to thin layer chromatography (TLC) analysis (Fig. 2-15). Cellobiose (C2), cellotriose (C3), cellohexaose (C6), gentiobiose, sophorose, laminaribiose, and lactose were all hydrolyzed to monosaccharides. No transglycosylation products were generated under the defined reaction condition (37°C for 1 h). Although there are a few BGs without transglycosylation activity (Shewale and Sadana, 1981; Saha and Bothast, 1996; Guo et al., 2016), GH1 BGs were frequently reported to have transglycosylation activity (Harhangi et al., 2002; Park et al., 2005;

Nguyen et al., 2010; Park et al., 2010; Uchiyama et al., 2013; Ramani et al., 2015). As mentioned in page 10, the transglycosylation reaction usually occurs at high concentration of glucose or cellobiose in a time-dependent manner. For instance, when 10 U/ml of purified PtBglu1, a GH1 BG from *Paecilomyces thermophila* and expressed in *P. pastoris*, was incubated with 10% (w/v) cellobiose and cellotriose at 55°C for 6 h, the transglycosylation products were detected during the reaction time between 15 min to 1 h (Yang et al., 2013). As a comparison, in this study, the reaction mixtures containing 0.2 U/ml of PaBG1b and 1% of substrates (i.e., cellobiose, cellotriose, cellohexaose, gentiobiose, sophorose, sucrose, laminaribiose, lactose, maltose), were incubated for 1 h at 37°C. Henceforth, further time-course analysis with higher concentration of substrate is need to verify whether PaBG1b has the transglycosylation activity.

2.2.2.5 Effect of cations and reagents

The effect of cations and reagents on the activity of PaBG1b was examined and is shown in Table 2-3. Overall, metal ions (5 mM) can more or less affect the activity of PaBG1b. K^+, Mg^{2+}, Mn^{2+}, Ca^{2+}, and Cu^{2+} displayed little effect on the activity of PaBG1b (over 80% activity remained), while the presence of Zn^{2+}, Fe^{2+}, and Fe^{3+} lead to the reduction of activity by 30% to 40%. The enzyme activity was augmented by 17% in the presence of Al^{3+}, which is in contrast to Bgl-gs1, a GH1 BG, which was inactivated by 10 mM Al^{3+} (Wang et al., 2012). It seems that the effect of Al^{3+} on a given BG is dose-dependent, as demonstrated for GH1 BG, NpaBGS, which was stimulated in the presence of 1 mM $AlCl_3$ but inhibited at 10 mM (Chen et al., 2012). Besides, $NiCl_2$ decreased more than 30% of the activity of PaBG1b. Some GH1 BGs

were characterized to be inhibited in the presence of $NiCl_2$ (Harnpicharnchai et al., 2009; Uchiyama et al., 2015) or Ni^{2+} (Fan et al., 2011; Wang et al., 2012; Yang et al., 2015; Zhao et al., 2015). A family-unknown BG of *Penicillium italicum* isolated and purified from rotten citrus peel was also observed to be sensitive to Ni^{2+} (Park et al., 2012). On the other hand, there is also an exceptional case where a GH1 BG was slightly stimulated by Ni^{2+} (Ramani et al., 2015). The inhibition or stimulation of GH1 BGs by Ni^{2+} might be in relation to the concentration of Ni^{2+} and the property of the protein, as demonstrated in the case of GH1 BG from *Humicola insolens*, which was stimulated in the presence of 1 mM Ni^{2+}, but inhibited by 10 mM Ni^{2+} (Meleiro et al., 2014). On the other hand, Ni_2SO_4 has negligible effect on PaBG1b, which is consistent with BGL2, a GH3 BG from *Neurospora crassa* and heterologously expressed in *P. pastoris* (Pei et al., 2016), which implies that the inhibition by Ni^{2+} might also be in relation to the anion species in solution.

As for the effect of chemical reagents, dithiothreitol, glycerol, dimethyl sulfoxide, and EDTA had negligible effects on PaBG1b, while SDS abolished the enzyme activity, and imidazole exhibited a distinct inhibitory effect on PaBG1b. In addition, the enzyme activity was increased by 15% in the presence of Triton X-100, which is similar to two alkaline BGs, As-Esc6 and AS-Esc10, isolated by functional metagenomics from soil (Biver et al., 2014).

The fact that PaBG1b was not inhibited by the chelating agent EDTA indicates that divalent cations are not required for enzyme action (Riou et al., 1998). Generally EDTA does not significantly inhibit the activity of BGs, but two BGs purified from the higher termite *M. muelleri* and its symbiotic fungus *Termitomyces* sp. were reported to be sensitive to 1 mM EDTA (Rouland et al., 1992), suggesting that these BGs need metal

cations for their activity.

2.2.2.6 Kinetics parameters (K_m, V_{max}, and k_{cat})

The kinetic parameters of PaBG1b towards both *p*NPG and cellobiose were examined. The results are shown in Figs. 2-16 and 2-17, and summarized in Table 2-4. PaBG1b has relatively poor affinity (K_m=28.0±1.7 mM) and lower V_{max} (59.9±0.8 U/mg) towards the artificial substrate *p*NPG. On the other hand, PaBG1b exhibited high activity (V_{max}=436.7±6.3 U/mg) and catalytic efficiency (k_{cat}/K_m=109.8 $mM^{-1}\cdot s^{-1}$) towards cellobiose, which is comparable to those of BGs listed in Table 0-2, but lower than those of the native enzyme (V_{max}=1,020 U/mg, k_{cat}/K_m=184 $mM^{-1}\cdot s^{-1}$). As to the K_m towards cellobiose, the value of PaBG1b (4.1±0.3 mM) was similar to that of the native enzyme (5.3 mM). A comparison of the catalytic efficiency of PaBG1b and native enzyme towards cellobiose is shown in Table 2-5. In addition, PaBG1b was not inhibited by cellobiose up to the highest concentration tested (100 mM), which is 16-fold higher than its K_m value. This characteristic makes PaBG1b a member of the best cellobiose-tolerant BGs documented (resistant to 100 mM cellobiose; Bohlin et al., 2013; Uchiyama et al., 2015), and thus endows it with an advantage in terms of industrial application (Discussed in page 68 and Table 2-7).

2.2.2.7 Glucose inhibition analysis

The K_i of PaBG1b towards glucose determined by the Dixon plot was 200.3±1.1 mM (Fig. 2-18). Except for some BGs being extremely tolerant to glucose (Riou et al., 1998; Fang et al., 2010; Pei et al., 2012; Lu et al., 2013; Yang et al., 2015), BGs generally exhibit a K_i values lower than 100 mM (Eyzaguirre et al., 2005; Sørensen et

al., 2013). Therefore, PaBG1b has moderate glucose tolerance.

2.2.2.8 Inhibition constant of imidazole

Imidazole and its derivatives were found to be potent inhibitors of BGs (Patchett et al., 1987; Li and Byers, 1989; Field et al., 1991; Li et al., 1998, cited in Fig. S2-1; Panday et al., 2000), as they were transition state analogues of the enzyme glycosyl intermediate. The imidazole ring was supposed to be capable of forming hydrogen bonds with the catalytic residues of the BGs (Heightman and Vasella, 1999; cited in Fig. S2-2). This suggestion was supported by the X-ray crystallography studies of the complexes of two GH1 BGs with the inhibitors, i.e. *Tm*GH1 from *Thermotoga maritima* with glucoimidazoles and *Ss*GH1 from *Sulfolobus solfataricus* with phenethyl-substitued glucoimidazoles (Gloster et al., 2006; cited in Fig. S2-3). Among the insect-derived BGs, *Rf*BGluc-1 of the lower termite *R. flavipe*s was found to be almost completely inhibited by 10 mM cellobioimidazole (Scharf et al., 2010). PaBG1b was also inhibited by 10 mM imidazole at pH 5.5 (Table 2-3). To investigate the reason why PaBG1b was inactivated by imidazole in the initial Ni-NTA purification trial, the inhibition constant of imidazole at the optimum pH of PaBG1b was evaluated (Fig. 2-19). Dixon plot analysis demonstrated that imidazole competitively inhibits PaBG1b with the inhibition constant (K_i) of 4.3±0.3 mM at pH 5.5, which indicates that imidazole is a modest competitive inhibitor of PaBG1b at its optimal pH.

2.2.2.9 Inhibition constant of Tris

Previously, it had been reported that some BGs, such as the BG from *C. thermocellum* (Ait et al., 1982) and the commercial BG preparation Novozym 188

(Dekker, 1986), were sensitive to Tris. Tris was found to be a strong inhibitor capable of binding to certain BGs by forming multiple hydrogen bonds with the residues in their catalytic center (Jeng et al., 2011, cited in Figs. S2-4 A and S2-4 B; Trofimov et al., 2013, cited in Fig. S2-4 C), and the inhibition on Asβ-Gly (a GH1 BG) was competitive (Trofimov et al., 2013). Furthermore, Tris inhibition was not limited to BGs but also seen in other glycosidases, such as α-glucosidase (Jørgensen and Jørgensen, 1967) and β-galactosidase (Karasová-Lipovová et al., 2003). The inhibition was suggested to be due to the formation of Tris-enzyme complex causing a steric hindrance for the substrate, or the change in the conformation or the charge distribution of the enzyme upon binding of Tris to the enzyme (Jørgensen and Jørgensen, 1967). In this study, to investigate the reason why the use of Tris buffer instead of sodium phosphate buffer resulted in circumvention of the inactivation of PaBG1b by imidazole at alkaline pH, the inhibition constant of Tris at the optimum pH of PaBG1b was evaluated. Dixon plot showed that Tris was a competitive inhibitor of PaBG1b (Fig. 2-20). The K_i of Tris was 5.9±0.2 mM at pH 5.5, which implied that Tris is also a modest competitive inhibitor of PaBG1b, but the potency is slightly weaker than that of imidazole.

Discussion

In the purification of PaBG1b by Ni-NTA column, the inhibition by imidazole unexpectedly occurred, which was solved by examining various purification conditions, especially by replacing sodium phosphate buffer with Tris-HCl buffer. Although both imidazole and Tris are modest competitive inhibitors for PaBG1b with comparable K_i at pH 5.5, the inhibition by Tris at either alkaline or acidic condition can be easily recovered by dilution or dialysis to decrease Tris concentration and restoring the

solution pH to acidic pH. In contrast, the inhibition by imidazole at alkaline pH could not be recovered, although at acidic pH the inhibition was reversible. These results suggest that the inhibition mechanisms of imidazole on PaBG1b under acidic and alkaline pH were different, and under alkaline condition, the conformation of PaBG1b might change forming a specific binding with imidazole. Flannelly et al. studied the influence of pH on the conformational and substrate binding dynamics of two GH1 BGs at pH 5-7.5 through molecular dynamics simulations (Flannelly et al., 2015), and concluded that the pH-dependent changes of ionization states of the non-catalytic residues outside the active site caused the disruption of the active site conformations by interfering with the formation of favorable hydrogen bonding between catalytic residues and the substrate. Likewise, the conformation of PaBG1b might change at pH 8.0, to which imidazole bound so tightly that it was hardly removed through regular methods such as dialysis or buffer exchange. However, Tris did not have such an effect and simply behaves as a competitive inhibitor under both alkaline and acidic pH. When Tris was employed for Ni-NTA purification, it might have preoccupied the binding site(s) of PaBG1b, and when imidazole was introduced in the subsequent wash and elution stages, it failed to compete with Tris and bind to PaBG1b. Thus the purified PaBG1b showed activity when pH was shifted to acidic pH and imidazole was removed.

The expression product of PaBG1b contained two types of polypeptides, probably due to the incomplete processing by Kex2 at the N-terminal KR site in the putative mature region. Although very short in sequence (9 amino acids in the native sequence of PaBG1b), this region might function as a pro-sequence. Pro-peptides have versatile functions such as ensuring correct folding (Steiner and Clark, 1968) and directing carboxylation of protein (Jorgensen et al., 1987). Proteases are often synthesized as

inactive precursors (zymogen), and subsequently activated by limited proteolysis (Neurath and Walsh, 1976). The role of pro-peptides of proteases might not only be suppression of the enzyme activity before reaching their destinations, but also containing the information for the delivery to the vacuole (Valls et al., 1987). Previously, it was also found that PcCel45A, a GH45 EG of *Phanerochaete chrysosporium*, had a 7-amino-acid presequence as a Kex2 protease site (KR) lies before the native N-terminal sequence, and the putative pro-sequence of PcCel5A was excised in the recombinant protein heterologously expressed in *P. pastoris* (Igarashi et al., 2008). However, to our knowledge thus far there has been no insect-origin BG precursors containing pro-peptides, and it is not known whether insects have such a post-translational processing mechanism. Therefore the KR dibasic pair in the mature region of PaBG1b might not be a recognition site for Kex2-like protease, but rather a coincidental occurrence in the amino acid sequence of PaBG1b. Owing to the failure in the N-terminal sequencing of the native PaBG1b (Arakawa et al., 2016), the N-terminal residue of native PaBG1b is unknown.

The optimum temperature (45°C) and thermostability (after 0.5 h at 50°C, over 80% activity remained) of the recombinant PaBG1b marginally match the temperature demand of the industrial cellulose processing (page 12). The optimum temperature of the majority of fungal BGs is at 55°C or above (Eyzaguirre et al., 2005). Regarding that insects are poikilothermic in general and the host organism of PaBG1b lives in the temperate zone, it is quite natural that PaBG1b has not been developed to be highly thermostable as the enzymes from fungi like *Trichoderma* spp. do (Arakawa et al., 2016). To overcome the drawback, amino acid substitution through site-directed mutagenesis or DNA shuffling could be a viable approach, and to date there were some

successful cases already presented (Lopez-Camacho et al., 1996; Ni et al., 2005; Lee et al., 2012).

Substrate specificity analysis revealed that PaBG1b has a broad substrate spectrum. The most favored aryl glycoside and glucoside substrates of PaBG1b were aryl β-fucosyl linkage and laminaribiose (3-β-D-glucosyl-D-glucose), respectively, rather than aryl β-glucosyl linkage and cellobiose (β-1,4-linked glucose). This preference is the same as that of NkBG (GH1) of the lower termite *N. koshunensis* (Uchima et al., 2011) and Ks5A7 (GH1) cloned from the metagenome of Kusaya gravy (Uchiyama et al., 2015); this is in reversed order of that of *Rf*BGluc-1 (GH1) from the lower termite *R. flavipes* (Scharf et al., 2010). Laminaribiose is a component of the cell walls of microbial insect pathogens, thus the relatively high activities of insect-derived BGs towards laminaribiose might be related to their function of immunity (Scharf et al., 2010). On the other hand, two BGs purified from the higher termite *M. muelleri* and its symbiotic fungus *Termitomyces* sp. are incapable of hydrolyzing laminaribiose and gentiobiose, although they display relatively high activity on cellobiose and moderate activity on *p*NPG (Rouland et al., 1992), which might imply that these BGs are merely play a role of cellulose digestion. PaBG1b can split off glucosyl units from the non-reducing end of cellooligosaccharides up to DP=6, which endows it with an advantage in terms of saccharification of short chain cellodextrins.

The results of kinetic analysis showed that PaBG1b demonstrated high activity and catalytic efficiency towards cellobiose (Table 2-5), which make it among the best BGs in terms of activity and catalytic efficiency, in comparison to other BGs listed in Table 0-2. Previously, mutagenesis experiments on plant-origin BGs demonstrated that replacing the residues forming the aglycone-binding sites of GH1 BGs led to significant

changes in k_{cat} and catalytic efficiency (Verdoucq et al., 2003). To gain insight on the high activity and catalytic efficiency of PaBG1b, an investigation the amino acid sequences of BGs highly active towards cellobiose (specific activity or V_{max} over 200 U/mg) in Table 0-2 was performed. Amino acid sequence alignment of PaBG1b, NkBG, and four GH1 BGs whose GenBank accession numbers are available, i.e. Glu1B, RfBGluc-1, PtBglu1, and BGL of *T. aotearoense* P8G3#4, were performed (Fig. 2-21). NkBG was reported to have 9 and 4 residues in the glycone- and the aglycone-binding pockets, respectively, which formed interactions with *p*NPG (Jeng et al., 2011). The alignment result shows that despite 11 sites are conserved in the BGs aligned, the site equivalent to N255 of NkBG is highly variable, i.e. D258 in PaBG1b, H252 in Glu1B, H252 in RfBGluc-1, D241 in PtBglu1, and T222 in BGL of *T. aotearoense* P8G3#4 (Fig. 2-21, indicated with a red arrow). Regarding that N255 is within the aglycone-binding pocket and capable of forming direct and indirect hydrogen bonds with *p*NPG (Jeng et al., 2011) and cellobiose (Jeng et al., 2012), it was suggested that D258 of PaBG1b might play an important role in catalytic hydrolysis and contribute to the high performance of PaBG1b compared to other BGs in Table 0-2.

With regard to cellobiose inhibition, *Aspergillus*-origin BGs with high activity but with sensitivity to cellobiose had been reviewed in the Introduction (page 13). Some fungal BGs are resistant to cellobiose up to 50-100 mM, but their hydrolysis activities are considerably reduced (Bohlin et al., 2013) or relatively weak (Shewale and Sadana, 1981). Judged by the Michaelis-Menten plot in Fig. 2-17A, PaBG1b was not apparently inhibited by 100 mM of cellobiose, which make it one of the most cellobiose-tolerant BGs.

As far as glucose tolerance is concerned, a summary of BGs documented with K_i

values over 100 mM is shown in Table 2-6. PaBG1b has modest glucose tolerance with a K_i of 200.3 mM. However, it is of note that the majority of glucose-tolerant BGs have poor activities towards cellobiose. For example, Bgl6 from metagenomic library of Turpan Depression and its mutant M3, which were extremely resistant to glucose (K_i over 3M), displayed low specific activities (10.0 and 41 U/mg) on cellobiose (Cao et al., 2015). Thus it can be emphasized that PaBG1b is endowed with the balanced characteristics of strong catalytic activity towards cellobiose and significant glucose tolerance. Interestingly, N223 in Td2F2 was identified to be the key residue for glucose tolerance, substrate specificity, and transglycosylation activity (Matsuzawa et al., 2016), and the equivalent residue (N256) is conserved in PaBG1b, although they share only 36% amino acid sequences identity with the E-value of 9e-85. As the glucose tolerance of two GH1 BGs, Bgl1B from a marine microbe (Liu et al., 2011) and Cel1A of *T. reesei* (Guo et al., 2016) were benefited from the mutagenesis study. Likewise, L198W and/or G203L mutation of PaBG1b might also have chances to improve its glucose tolerance, or even thermostability (Fig. S2-5).

The application of BGs with drawbacks in either glucose or cellobiose resistance might require additional consideration in the processing. For an overall comparison, a summary of BGs whose activities, glucose inhibition constant (K_i), and cellobiose inhibition concentration are available is shown in Table 2-7. It is apparent that PaBG1b is prominent in that it displays a high activity together with significant tolerance to glucose and cellobiose, which will potentially facilitate industrial application than other BGs with shortcomings.

As to the differences in the catalytic properties of recombinant PaBG1b and native enzyme (Table 2-5), in addition to the significant difference in the optimal temperature

and the marginal differences in the optimum pH and K_m towards cellobiose, the specific activity and V_{max} of the recombinant PaBG1b were approximately half of those of the native enzyme. Presently a number of BGs were successfully expressed in *P. pastoris*, most of which were fungal origin (Chen et al., 2011; Yang et al., 2013; Yang et al., 2014; Zhao et al., 2015; Pei et al., 2016), and some were derived from termites (Uchima and Arioka, 2012; Uchima et al., 2012; Akemi Uchima et al., 2013) or plants (Suthangkornkul et al., 2016). However, it is uncommon that the biochemical properties of the native enzymes and heterologously-expressed products were compared, possibly because the native enzymes were not available or hardly obtained. In the cases of fungal-origin BGs heterogeneously expressed in *P. pastoris*, the recombinant BGs basically shared the same characteristics with those of the native enzymes in terms of optimum temperature and pH (Harnpicharnchai et al., 2009), thermostability (Karkehabadi et al., 2014) as well as kinetic characteristics (Kawai et al., 2003; Liu et al., 2012, Hong et al., 2007). To explain the decrease in activity of the recombinant PaBG1b, six hypotheses are pushed forward as described in the following paragraph.

Firstly, the data are not based on the direct comparison between the native and recombinant enzymes. Although the assay conditions in this study are the same as those defined by the authors working on the native enzyme (Arakawa et al., 2016), the results could be slightly varied owing to the difference in the lab equipment and persons. This might be similar to the case of SBgl3His, a *Solanum torvum* GH3 BG heterologously expressed in *P. pastoris* with a His_6 tag fused at its C-terminus (Suthangkornkul et al, 2016). The turnover rate of SBgl3His was determined to be approximately half of that of the native enzyme, although the data of native enzyme were cited from a literature (Arthan et al., 2006), whereas the other biochemical and kinetic properties were

generally consistent with the native enzyme (Suthangkornkul et al, 2016). Secondly, because of the existence of a KR site in the N-terminus of mature PaBG1b, the recombinant PaBG1b expressed in *P. pastoris* were present in a mixture of two polypeptides with different N-terminal sequences, although whether or not the absence of small polypeptide (9 amino acid residues) from the N-terminus resulted in the difference in the activity is unclear. It is practically difficult to separate these two polypeptides. Tsukada et al. characterized two intracellular GH1 BGs of the basidiomycete *P. chrysosporium*, BGL 1A and BGL1 B, which shared high sequence identity, and found that BGL 1B having a C-terminal extension of 63-amino acids hydrolyzed cellobiose more effectively than BGL1 A. In addition, the substrate recognition patterns were quite different from each other (Tsukada et al., 2006). Thirdly, the recombinant PaBG1b has two C-terminal fusion tags which might have affected the catalytic function, although it may not always be the case that the tags added at either N- or C-terminus has any unfavorable effects. For other cellulases, Zhang et al. has ever found that the C-terminally-tagged recombinant EG derived from *C. formosanus* and expressed in *E. coli* was less active and stable than its native form (Zhang et al., 2009). Fourthly, the difference in the post-translational modification such as *N*-glycosylation was suggested to have effects on the catalytic efficiency of BG (Suthangkornkul et al, 2016) and the hyper-glycosylation of recombinant PaBG1b might affect the catalytic activity. Lastly, since the recombinant PaBG1b was produced in the heterologous host, the proportion of improperly-folded, inactive PaBG1b might be higher than that of the preparation purified from its native host, which led to the lower specific activity. However the last two hypotheses do not have exemplar cases.

To overexpress PaBG1b with the full-length mature region, expression in *A. oryzae*

and *E. coli* could be options. Changing the expression host might also have opportunities to improve the activity of the resultant product, as demonstrated in the case of NkBG expressed in *E. coli* (Ni et al., 2007), which exhibited a higher activity on cellobiose than the native enzyme (Tokuda et al., 2002).

Table 2-1. Summary of purification of PaBG1b

Purification step	Total protein (mg)	Total activity (U)[a]	Specific activity (U/mg)	Recovery (%)	Purification fold
Culture supernatant	26.4	360.7	13.7	100.0	1.0
Ammonium sulfate precipitation	16.2	295.8	18.3	82.0	1.3
Ni-NTA chromatography	3.0	127.4	43.0	35.3	3.2
HiTrap DEAE FF chromatography	1.6	72.7	45.6	20.2	3.3

[a]: one enzyme unit (U) was defined as the amount of enzyme required to release 1 μmol of *p*-nitrophenol from the substrate per min.

Table 2-2. Substrate specificity of PaBG1b

Substrate	Linkage of glycosyl groups	Relative activity (%)[a]
Aryl glycosides		
p-nitrophenyl β-D-fucopyranoside	(βFuc)	100±0.9
p-nitrophenyl β-D-glucopyranoside	(βGlc)	61.7±1.3
p-nitrophenyl β-L-arabinofuranoside	(βAra*f*)	1.0 ±0.2
p-nitrophenyl β-D-galactopyranoside	(βGal)	6.9±0.3
p-nitrophenyl β-D-*N*-acetylglucosamine	(βGlcNAc)	ND
Saccharides		
Cellobiose	(β-l,4) Glc	68.3±9.9
Cellotriose	(β-l,4) Glc	59.7±1.6
Cellohexaose	(β-l,4) Glc	35.9±0.7
Laminaribiose	(β-1,3) Glc	100±3.1
Sophorose	(β-1,2) Glc	34.9±1.1
Gentiobiose	(β-1,6) Glc	4.2±0.8
Lactose	(β-1,4) Gal	3.7±0.4
Salicin	(βGlc)	1.9±0.2
Surcose	(α-1,2)Glc	ND
Maltose	(α-1,4)Glc	ND
Avicel	(β-l,4) Glc	ND
CMC	(β-l,4) Glc	ND
Laminarin	(β-1,3; β-1,6)Glc	ND

ND: not detected.

[a]: the activities toward aryl glycosides (10 mM) and saccharides (1%) were determined by measuring the release of *p*-nitrophenol at A_{410} and glucose at A_{505}, respectively. Data are means ± S.D. of three independent experiments. The most preferentially-hydrolyzed aryl glycoside and saccharide substrates (i.e. *p*-nitrophenyl β-D-fucopyranoside and laminaribiose, respectively) were taken as 100%.

Table 2-3. Effect of cations and reagents

Item	Relative activity (%)[a]
Control[b]	100±2.46
Cations (5 mM)	
KCl	88.6±2.1
$MgCl_2$	96.3±2.4
$MnCl_2$	92.1±1.1
$CaCl_2$	82.1±2.1
$NiCl_2$	68.1±3.6
$NiSO_4$	97.7±6.5
$ZnSO_4$	72.7±4.0
$CuSO_4$	105.8±3.3
$FeSO_4$	58.6±5.3
$Fe(NO_3)_3$	64.3±1.8
$AlCl_3$	117.3±8.1
Reagents (10 mM)	
SDS	1.4±1.3
Imidazole	74.6±3.6
Dithiothreitol	92.7±4.9
Glycerol	93.3±2.7
Dimethyl sulfoxide	97.2±7.1
EDTA	102.6±8.0
Tween 80	106.1±3.2
Triton X-100	115.6±6.8

[a]; data are means ± S.D. of three independent experiments.
[b]; control assays were done under optimal conditions without additives.

Table 2-4. Summary of kinetic parameters of PaBG1b

Substrate	Spec. act. (U/mg)[a]	V_{max} (U/mg)[a]	K_m (mM)	k_{cat} (s^{-1})[b]	k_{cat}/K_m (mM^{-1} s^{-1})
*p*NPG	45.5 ± 0.5	59.9 ± 0.8	28.0 ± 1.7	61.7±0.8	2.2
Cellobiose	338.5 ± 2.8	436.7 ± 6.3	4.1 ± 0.3	450.2±6.5	109.8

Spec. act.: specific activity.
[a]: one unit (U) of BG activity was defined as the amount of enzyme required to release l μmol of *p*-nitrophenol (*p*NP) from *p*NPG, or 2 μmol of glucose from cellobiose per minute.
[b]: the turnover number (*k*cat) was calculated based on the apparent molecular weight of the recombinant PaBG1b (61,800 Da) with V_{max}.
For spec act., V_{max} and K_m, data are mean ±SD of three independent experiments.

Table 2-5. Comparison of the catalytic efficiency of the native and the recombinant PaBG1b toward cellobiose[a].

Enzyme	Opt. temp (° C)	Opt. pH	K_m (mM)	Spec. act. (U/mg)	V_{max} (U/mg)	k_{cat}/K_m (mM^{-1} ·S^{-1})	Reference
Native	60	5.0	5.3	708	1020	184	Arakawa *et al*., 2016.
Recombinant	45[b]	5.5	4.1	338.5	436.7	109.8	This study

[a]: enzyme unit (U) was defined as the amount of enzyme required to release 2 μmol of glucose from cellobiose per minute. Enzyme reaction was conducted the same as described by Arakawa et al., i.e., mixing of 25 μl of enzyme with 100 μl of 1% cellobiose (approximately 30 mM) in 100 mM sodium acetate solution, pH 5.5, incubating at 37° C for 5 min, then stop the reaction by boiling for 5 min (Arakawa *et al*., 2016).
Spec. act.: specific activity.
Opt. temp: optimum temperature.
Opt. pH: optimum pH.
[b]: the reaction condition for optimum temperature analysis was using *p*NPG as the substrate and reacting for 30 min, whilst that of the native enzyme was using cellobiose and reacting at 37° C for 5 min.

Table 2-6. Comparison of kinetic properties on cellobiose of selected BGs with high glucose tolerance (K_i >100 mM).

Name	Origin	Expression host	Spec. at. (U/mg)	V_{max} (U/mg)	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	K_i (mM)	Reference
BGII	*A. niger* CBS 55464	(Native enzyme)	(very low)	N/A	N/A	N/A	953	Günata and Vallier, 1999
(Unnamed)	*A. niger* CCRC 31494	(Native enzyme)	2.6[a]	N/A	N/A	N/A	543	Yan and Lin, 1997
HGT-BG	*A. oryzae*	(Native enzyme)	N/A	353[b]	7	N/A	1360	Riou *et al*., 1998
β-glucosidase III	*A. tubingensis* CBS 643.92	(Native enzyme)	N/A	Low[c]	N/A	N/A	470	Decker *et al*., 2001
β-glucosidase IV	*A. tubingensis* CBS 643.92	(Native enzyme)	N/A	2.9[a]	N/A	N/A	600	Decker *et al*., 2001
(Unnamed)	*Candida peltata* NRRL Y-6888	(Native enzyme)	N/A	75[b]	66	N/A	1400	Saha and Bothast, 1996
Td2F2	Compost microbial	(Native enzyme)	0.6[b] (or 4.6[b])	N/A	4.4	1.61[b]	Stimulated at 1 M	Uchiyama *et al*., 2013 (Matsuzawa *et al*., 2016)
(Unnamed)	*Debaryomyces vanrijiae*	(Native enzyme)	N/A	42.15[a]	57.9	N/A	439	Belancic *et al*., 2003
reBglM1	*Marinomonas* MWYL1	*E.coli*	N/A	508[b]	1.1	395.8	>400	Zhao *et al*., 2012
Bgl6	Metagenomic library of Turpan Depression	*E. coli*	10.9[a]	N/A	38.5	0.3[a]	3.5[d]	Cao *et al*., 2015
M3	Mutant of Bgl6	*E. coli*	~41[a]	N/A	49.2	0.9[a]	3.0[d]	Cao *et al*., 2015
NkBG	*N. koshunensis*	*A. oryzae*	Yes[e]	N/A	N/A	N/A	>600	Uchima *et al*., 2011
Ks5A7	Kusaya gravy	*E. coli*	170[b]	155[b]	0.4	386[b]	Stimulated at 0.1-0.4 M	Uchiyama *et al*., 2015
PaBG1b	*P. a. spadica*	*P. pastoris*	338.5	436.7	4.1	109.8	200.3	This study
(Unnamed)	*Scytalidium thermophilum* CBS 619.91	(Native enzyme)	1.9[a]	2.1[a]	1.61	N/A	>200	Zanoelo *et al*., 2004
Bgl1269	Soil samples	*E. coli*	N/A	N/A	N/A	N/A	4.3	Li *et al*., 2012
BGL	*T. aotearoense* P8G3#4	*E. coli*	370.2[a]	370.3[a]	25.5	157.2[a]	800	Yang *et al*., 2015
(Unnamed)	*T. thermosaccharolyticum* DSM 571	*E. coli*	N/A	120[b]	7.9	13.3[b]	600	Pei *et al*., 2012
Tt-BGL	*Thermotoga thermarum* DSM 5069T	*E. coli*	N/A	19[b]	35.5	N/A	1500[f]	Zhao *et al*., 2013

(Continued)

Table 2-6. Comparison of kinetic properties on cellobiose of selected BGs with high glucose tolerance (K_i >100 mM). (Continued from the previous page)

Name	Origin	Expression host	Spec. at. (U/mg)	V_{max} (U/mg)	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	K_i (mM)	Reference
167/172 Mutant	*T. reesei*	*E. coli*	N/A	N/A	N/A	N/A	650	Guo et al., 2016
Bgl1A	Uncultured bacterium	*E. coli*	N/A	15.5[b]	20.4	N/A	1000	Fang *et al*., 2010
Unbgl1A	Uncultured bacterium	*E. coli*	N/A	N/A	N/A	N/A	1500	Lu *et al*., 2013

Spec. at.: specific activity. One unit of activity was defined as the amount of enzyme that produced 2 μmol of glucose per min.
[a]: for comparison, data in the original references were re-calculated to show the enzyme unit as defined above.
[b]: the definition of enzyme unit was not clearly specified in the original references.
[c]: the physico-chemical and kinetic properties of β-glucosidase III were similar to those of IV according to the reference.
[d]: data are presented as half maximal inhibitory concentration (IC50).
[e]: the activity towards cellobiose was reported as relative activity by the authors.
[f]: the K_i value towards glucose was defined as amount of glucose required to inhibit 50% of the BG activity.

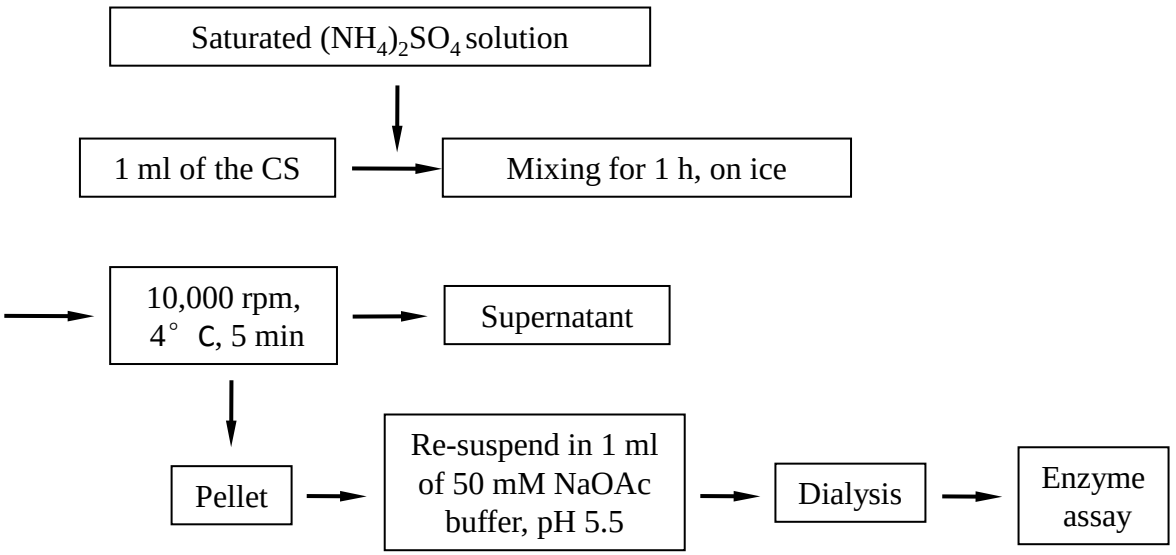
(Continued)

Table 2-7. Comparison of kinetic properties of selected BGs on cellobiose, and tolerance to glucose and cellobiose

Name	Origin	Expression host	Spec. at. (U/mg)	V_{max} (U/mg)	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	K_i (to glucose, mM)	Inhibited by Cellobiose (mM)	Reference
Novozym 188	*A. niger*	(Native enzyme)	N/A	16.9[a]	5.6	N/A	3	>10	Dekker, 1986
β-Glu II	*A. niger* CCRC 31494	(Native enzyme)	N/A	232[a]	15.4	N/A	5.7	>50	Yan *et al.*, 1998
HGT-BG	*A. oryzae*	(Native enzyme)	N/A	353[b]	7	N/A	1360	~438[c]	Riou *et al.*, 1998
Unnamed	*A. phonenicis*	(Native enzyme)	160[b]	N/A	0.8	N/A	N/A	10	Sternberg, 1977
Unnamed	*A. wentii*	(Native enzyme)	113[b]	N/A	N/A	N/A	N/A	10	Sternberg, 1977
Ks5A7	Kusaya gravy metagenome	*E. coli*	170[b]	155[b]	0.4	386[b]	Stimulated at 0.1-0.4 M	100	Uchiyama *et al.*, 2015
Bgl6	Metagenomic library of Turpan Depression	*E. coli*	10.9[a]	N/A	38.5	0.3[a]	3.5[d]	440	Cao *et al.*, 2015
M3	Mutant of Bgl6	*E. coli*	~41[a]	N/A	49.2	0.9[a]	3.0[d]	>440	
PaBG1b	*P. a. spadica*	*P. pastoris*	338.5	436.7	4.1	109.8	200.3	100	This study
BGL I	*Periconia* sp. BCC2871	*P. pastoris*	N/A	627[b]	0.5	N/A	20	<10[f]	Harnpicharnchai *et al.*, 2009
Bgl1269	Soil samples	*E. coli*	N/A	N/A	N/A	N/A	4.3	N/A	Li *et al.*, 2012
BG-3	*S. rolfsii*	(Native enzyme)	87.5[a]	87.5[a]	5.8	N/A	0.6[e]	>58 .4[g]	Shewale and Sadana, 1981

Spec. at.: specific activity.
One unit of activity was defined as the amount of enzyme that produced 2 μmol of glucose per minute
[a] : for comparison, data in the original references were re-calculated to show the enzyme unit as defined above
[b]: the definition of enzyme unit was not clearly specified in the original references.
[c]: the initial hydrolysis rate was decrease half in the presence of 15% (wt/vol) cellobiose (438 mM).
[d]: data are presented as half maximal inhibitory concentration (IC50).
[e]:the substrate employed for inhibition constant determination experiment is cellobiose.
[f]: when 10 mM cellobiose was presented as a inhibitor, 79.8% of the relative activity on *p*NPG was remained.
[g]: deduced by the description of no inhibition at cellobiose concentration up to 10 fold of K_m in the source reference.

A



B

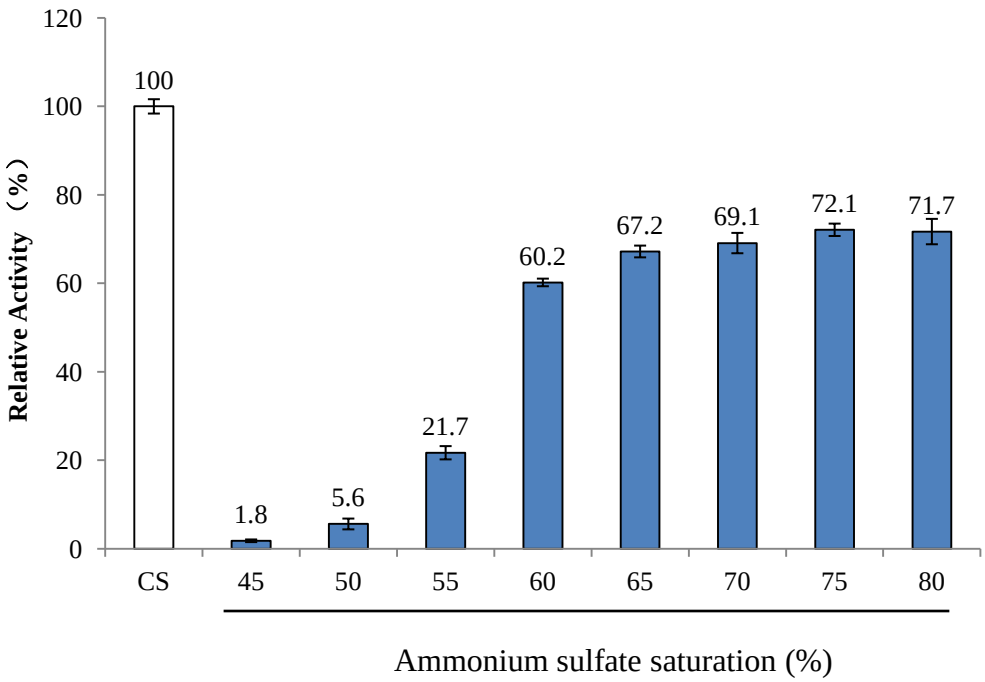
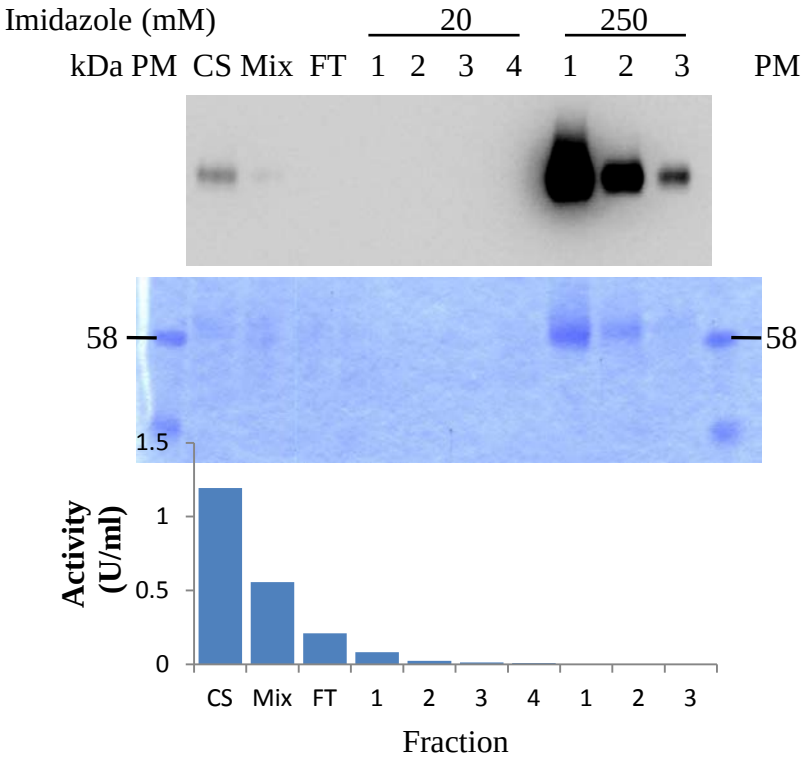


Fig. 2-1. Determination of the condition for ammonium sulfate precipitation
(A), scheme of ammonium sulfate precipitation experiment, (B) relative recovery of PaBG1b activity. In (B), the activity of the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant was taken as 100%, and the relative activities of the pellets which were dissolved in 50 mM sodium acetate buffer (pH 5.5) in (A) are shown. Data are means ± S.D. of three independent experiments.

A



B

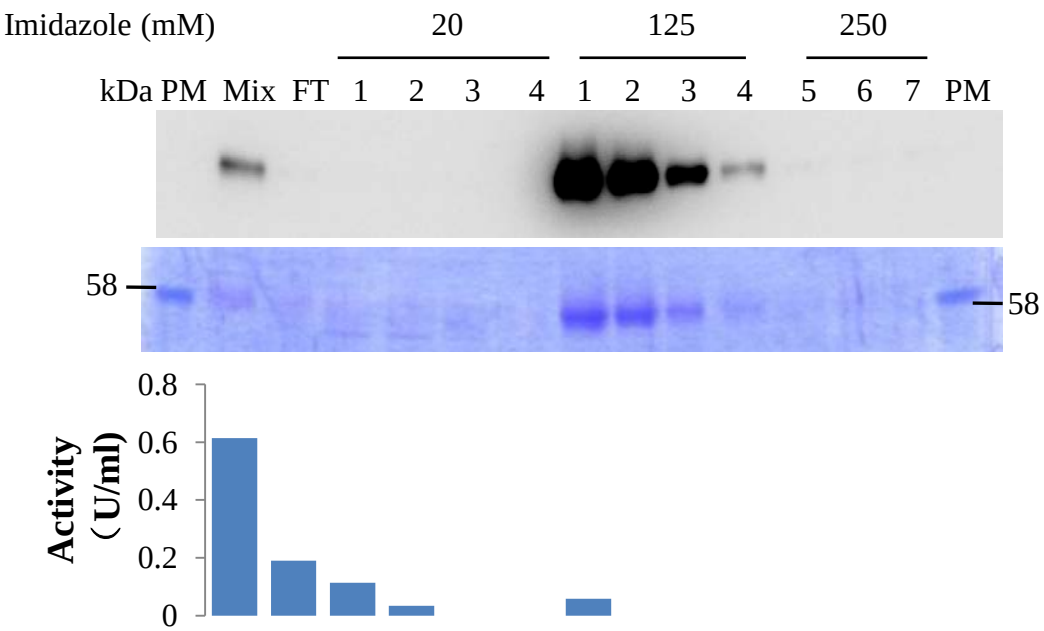


Fig. 2-2. Purification of PaBG1b by Ni-NTA affinity chromatography at pH 8.0
Elution by 250 mM imidazole (A) and stepwise elution by 125 mM and 250 mM imidazole (B). Representative Western blot analysis, CBB staining, and BG activity assay results are shown from the top to the bottom panels, respectively. PM: protein marker; CS: culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; Mix: mixture of the culture supernatant with the same volume of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8.0; FT, flow-through fraction. The volume of wash and elution fractions which contained 20 mM and 125/250 mM imidazole, respectively, was 1 ml for each.

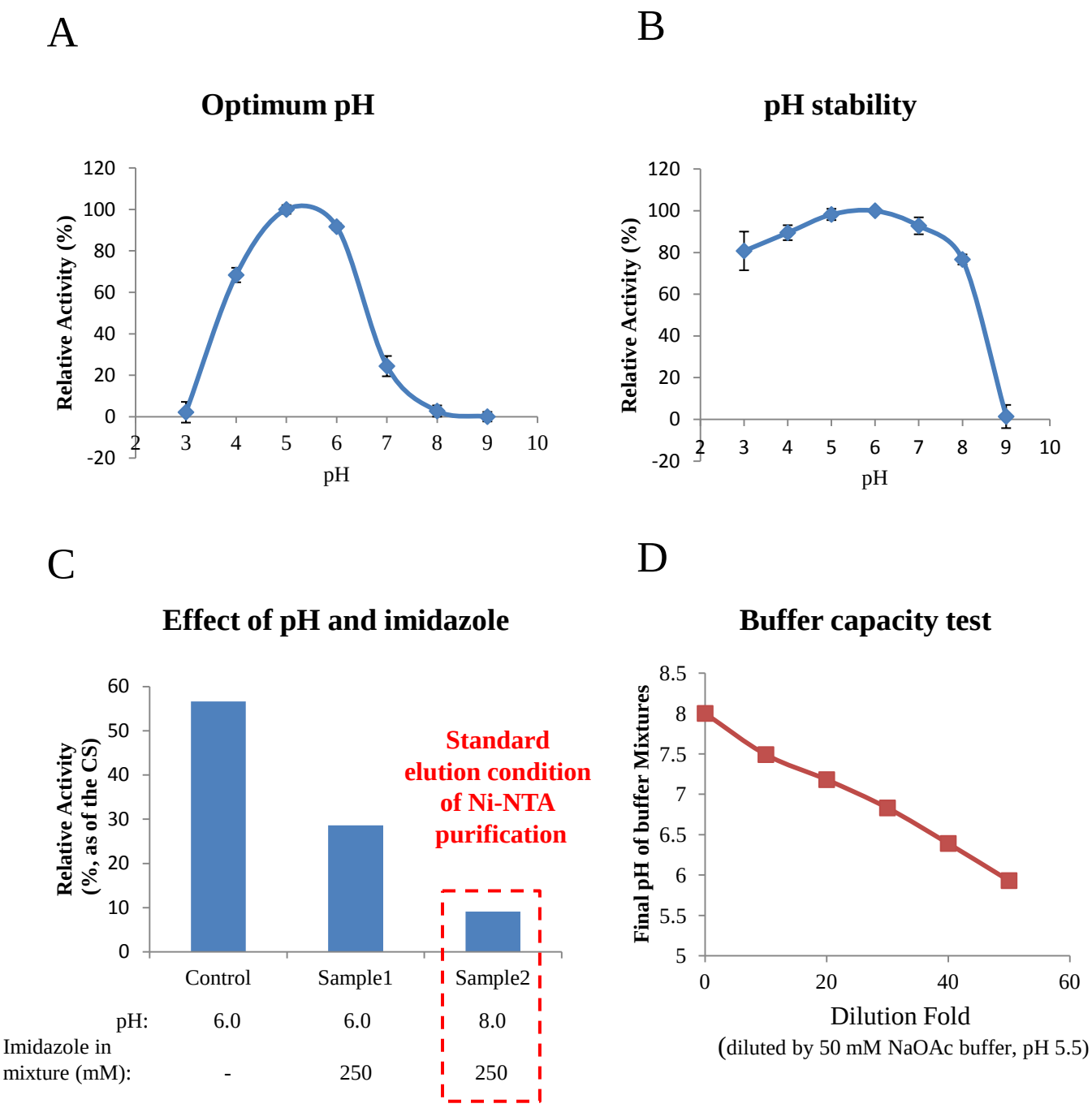


Fig. 2-3. Optimum pH (A), pH stability (B) of the culture supernatant of the pPICZα-A-PABG1b *P. pastoris* transformant, the effect of pH and imidazole on the activity of the culture supernatant (C), and the buffer capacity test (D).

(A), Optimum pH was determined by using a series of 50 mM of Britton-Robinson buffers (50 mM H_3BO_3 + CH_3COOH + H_3PO_4) ranging from pH 3 to 9 to prepare the substrate solutions. (B), the culture supernatant was firstly mixed with the same volume of Britton-Robinson buffers, left on ice for 1 h, and then the enzyme assay was performed. Data are means ± S.D. of three independent experiments. (C), the culture supernatant was mixed with the same volume of 50 mM sodium phosphate buffer, containing 300 mM NaCl and 0 or 500 mM imidazole, pH 6.0 or 8.0, to reach a final imidazole concentration of 0 or 250 mM, then left on ice for 1 h, and the enzyme assay was performed. (D), pH of the solution after the elution buffer of Ni-NTA purification (50 mM phosphate buffer containing 300 mM NaCl and 250 mM imidazole, pH 8.0) was diluted with the enzyme assay buffer (50 mM sodium acetate, pH 5.5). The dilution-fold is indicated in the X-axis.

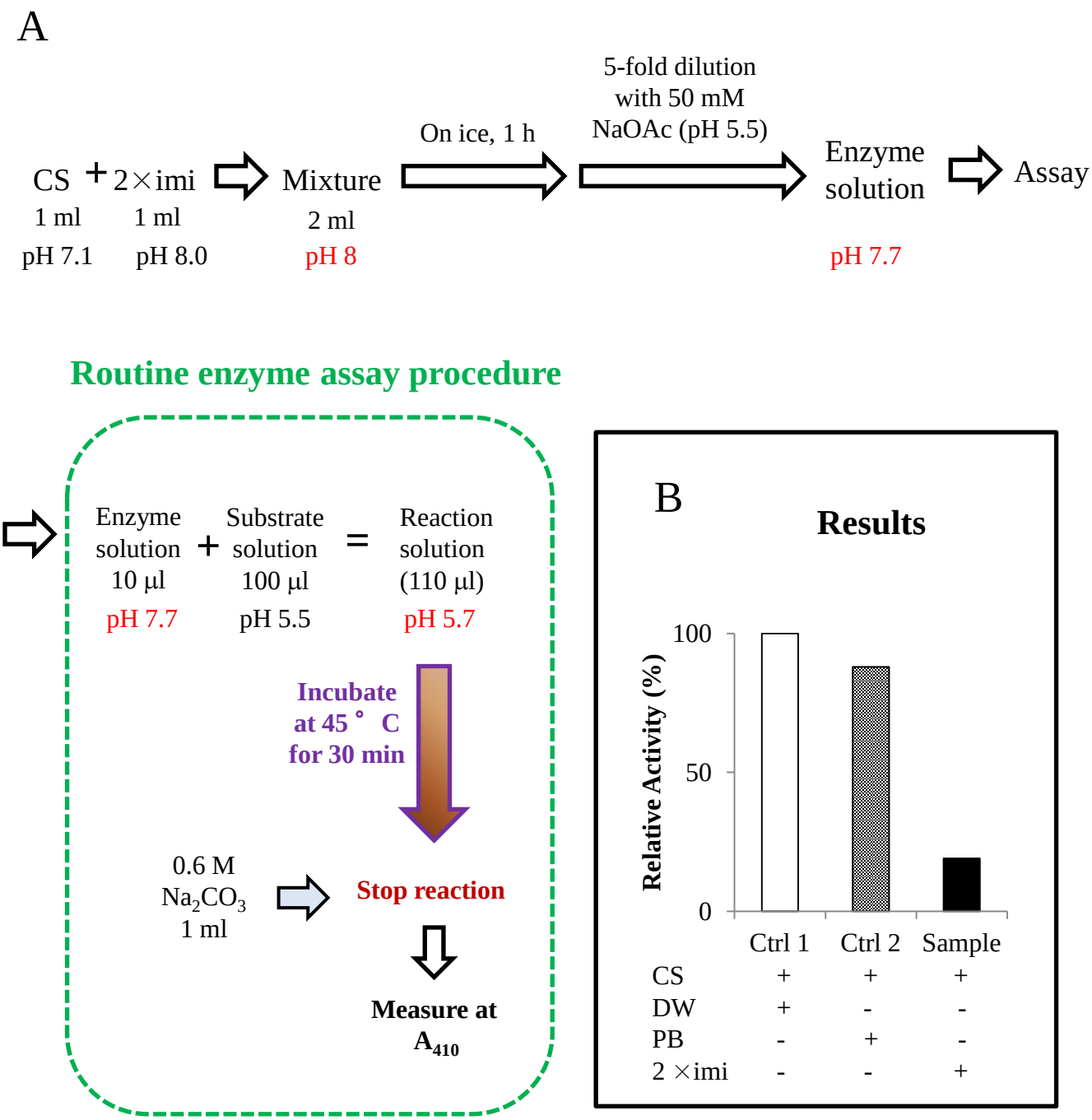


Fig. 2-4. Imidazole inactivates PaBG1b

(A), schematic representation of the design of the experiment. CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; 2×imi, 50 mM phosphate buffer containing 600 mM NaCl and 500 mM imidazole, pH 8.0; NaOAc: sodium acetate buffer; Substrate solution: 10 mM *p*NPG in 50 mM sodium acetate buffer (pH 5.5). (B) BG activity of the Mixture in (A). In the control reactions (Ctrl 1 and 2), the Mixture were prepared by using distilled water (DW) or phosphate buffer (PB; 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8.0) instead of 2×imi in (A).

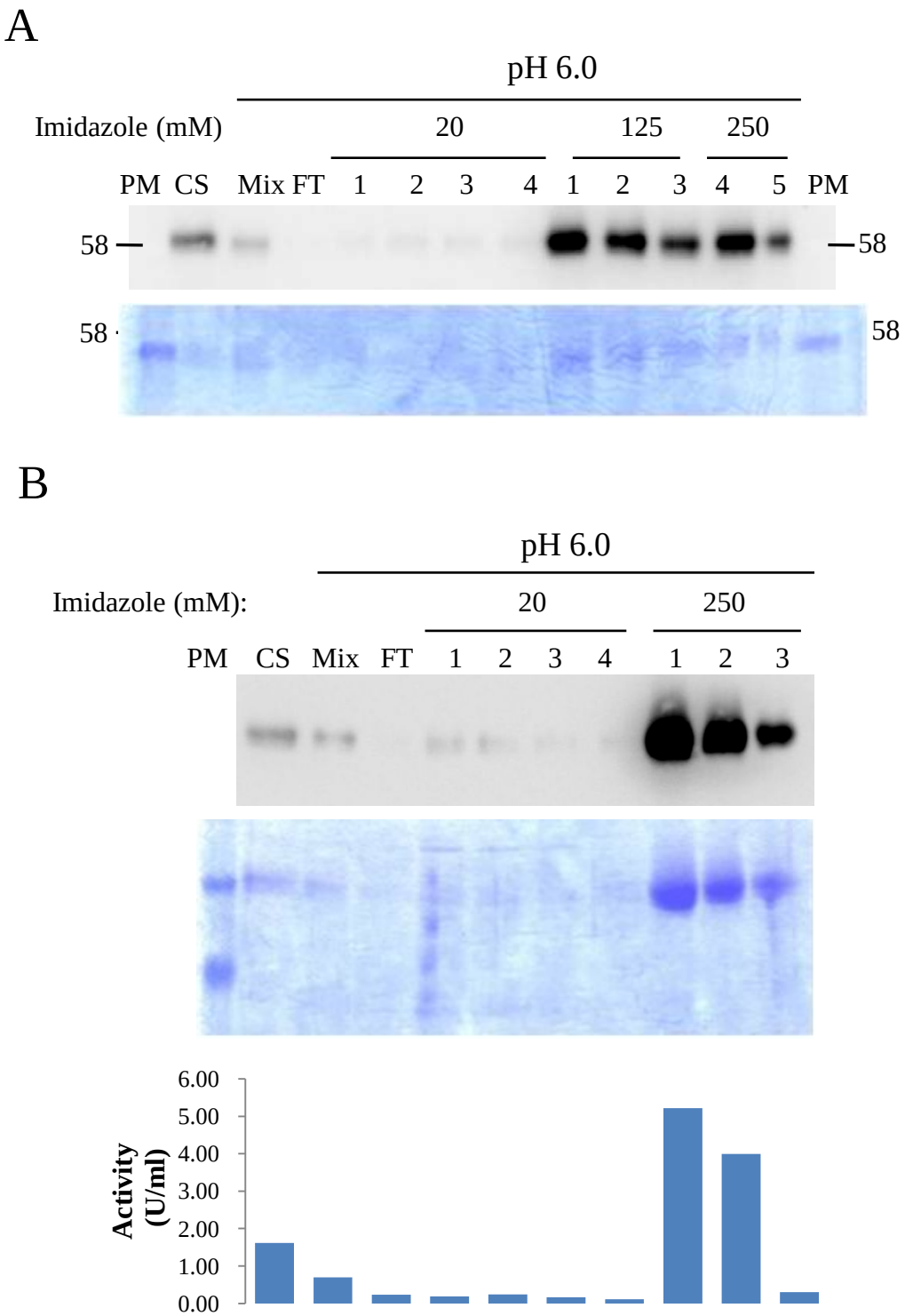


Fig. 2-5. Purification of PaBG1b by Ni-NTA affinity chromatography at pH 6.0, eluting with 125/250 mM imidazole (A) and 250 mM imidazole (B).
Western blot analysis, CBB staining, and BG activity assay results are shown from the top to the bottom panels, respectively. PM, protein marker; CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; Mix, culture supernatant mixed with the same volume of 50 mM phosphate buffer, 300 mM NaCl, pH 6.0; FT, flow-through fraction. The column was washed and eluted with buffers containing 20 mM and 125-250 mM imidazole, respectively. Approximately 14.8% of the total activity was recovered.

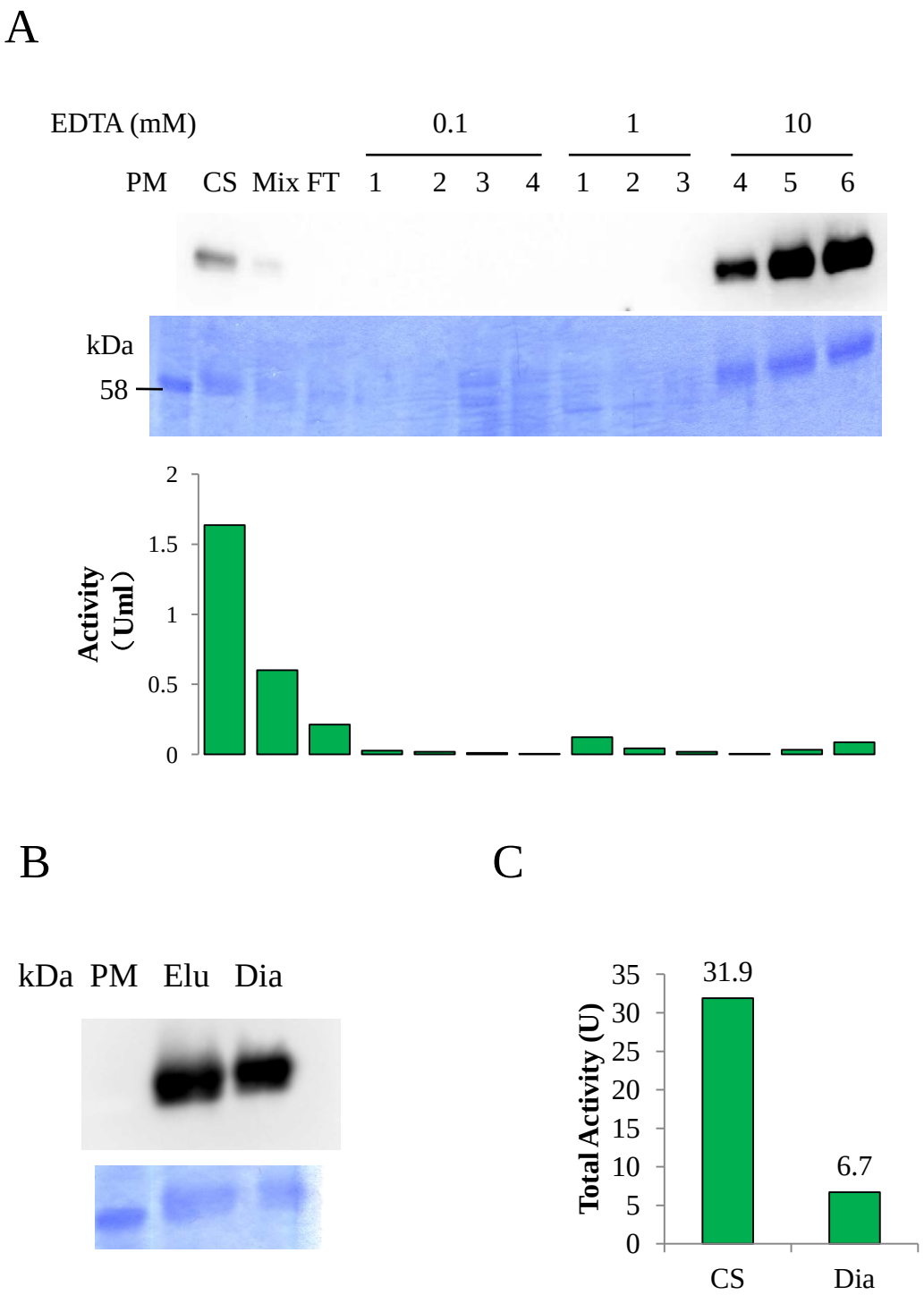


Fig. 2-6. Purification of PaBG1b by EDTA elution

(A) from the top to the bottom, Western blot analysis, CBB-staining, and BG activity assay results are shown. CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; Mix, culture supernatant mixed with the same volume of 50 mM phosphate buffer, 300 mM NaCl, pH 8.0; FT, flow through fraction. EDTA concentrations are shown at the top. (B) Elu, mixture of elution fractions 4 to 6 in (A); Dia, dialysis product of Elu against 50 mM sodium acetate buffer, pH 5.5. (C) the total activity of culture supernatant and Dia. Recovery of total activity was 21.0%.

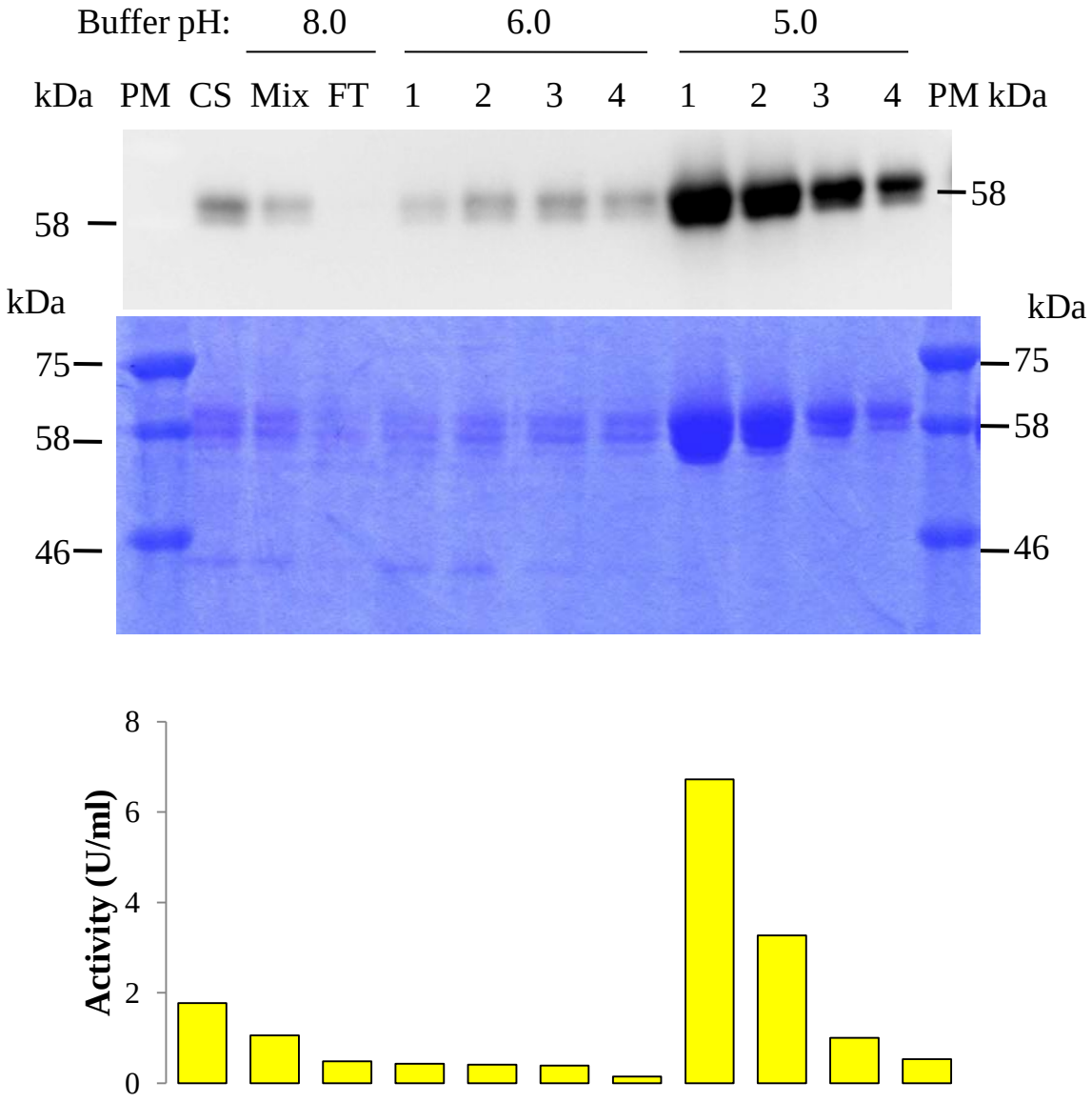


Fig. 2-7. Purification of PaBG1b by pH elution
From the top to the bottom, Western blot analysis, CBB-stained SDS-PAGE gel, and BG activity assay results are shown. CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; Mix, culture supernatant mixed with the same volume of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8.0; FT, flow through fraction. The column was washed by 50 mM phosphate buffer, 300 mM NaCl, pH 6.0, and eluted by 50 mM sodium acetate buffer, pH 5.0, as indicated at the top. The recovery of total activity was 16.2%.

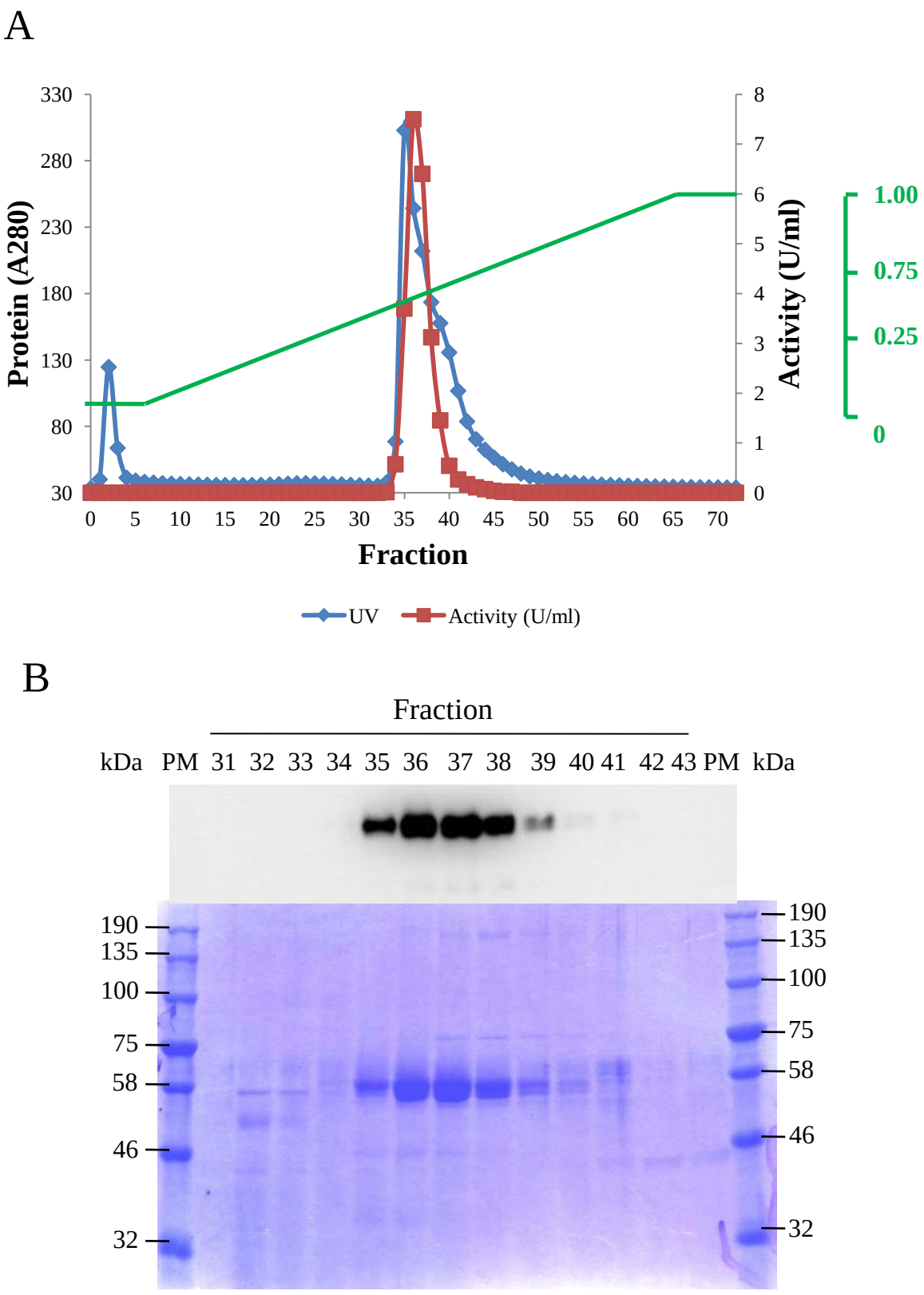


Fig. 2-8. Purification of PaBG1b by anion exchange chromatography
Chromatography profile (A) and Western blot analysis/CBB-stained SDS-PAGE gel (B) are shown. Crude enzyme solution from ammonium sulfate precipitation of the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant was purified using the HiTrap DEAE FF column. Starting buffer: 50 mM HEPES buffer, pH 7.0; elution buffer: 1 M NaCl in the same buffer.

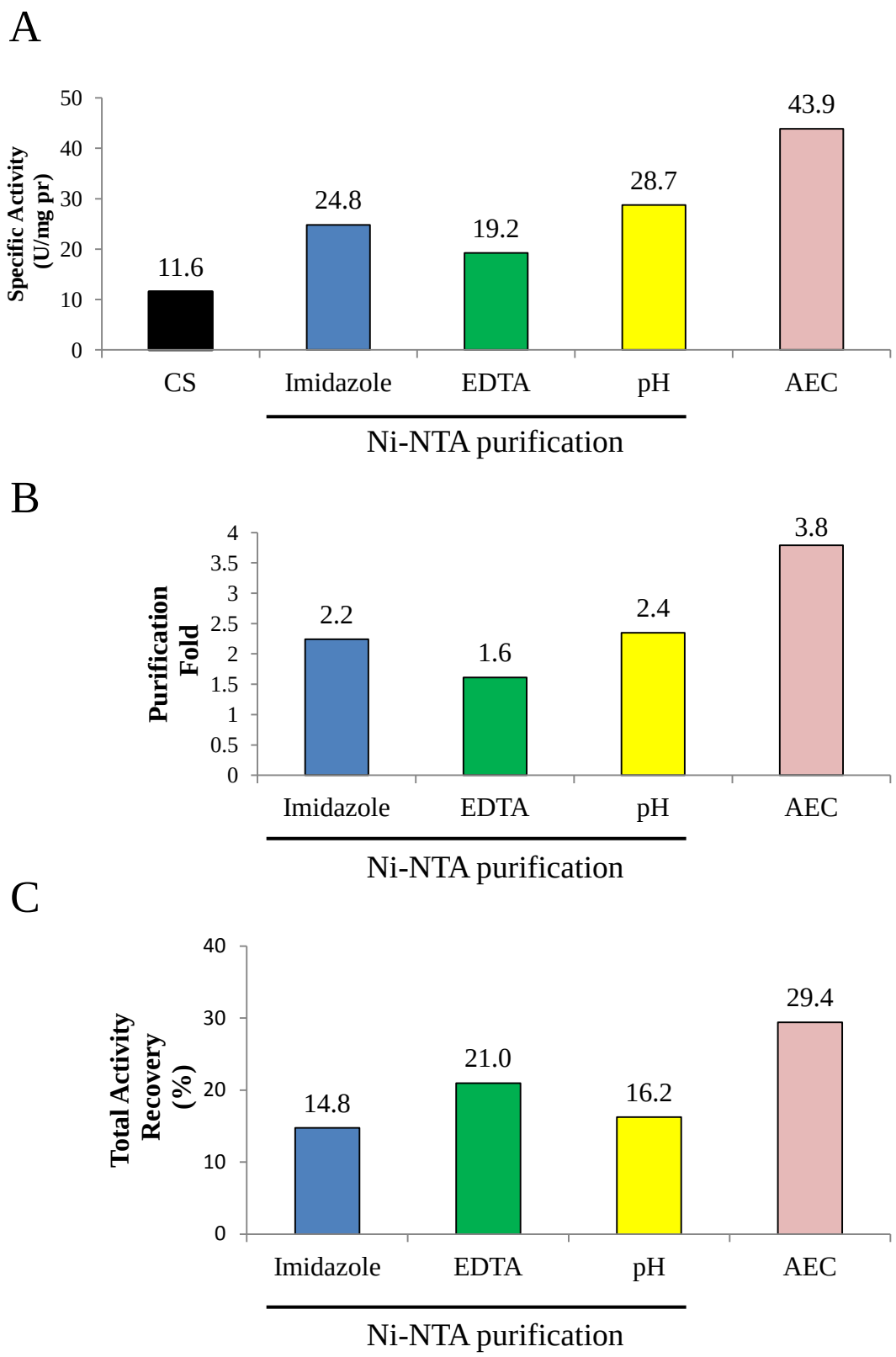


Fig. 2-9. Comparison of specific activities of purified PaBG1b (A), purification fold (B), and total recovery (C) of different purification methods

CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant; Imidazole and EDTA, elution substance for Ni-NTA purification; pH, Ni-NTA purification by pH elution; AEC, anion exchange chromatography.

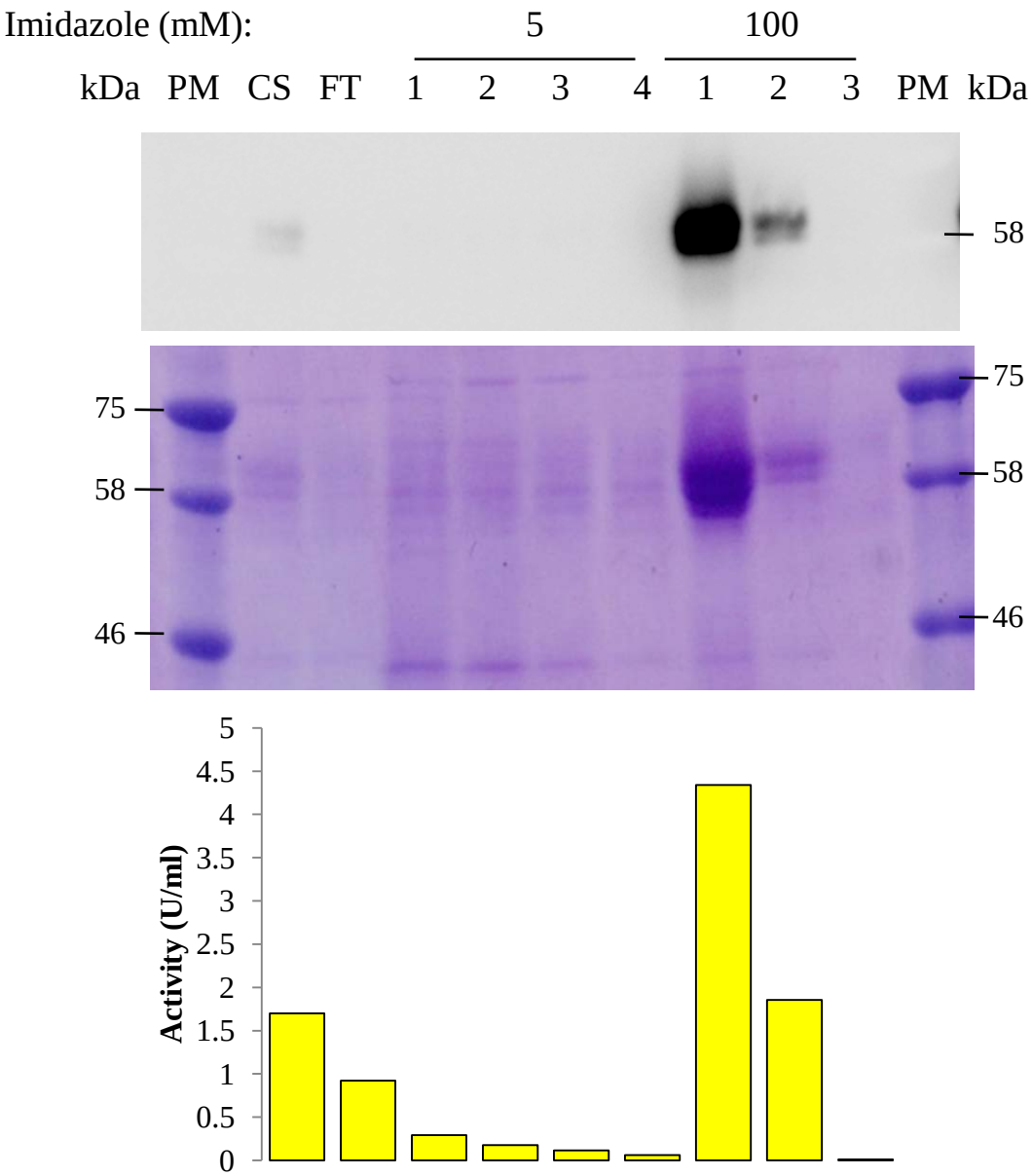


Fig. 2-10. Purification of PaBG1b using Tris-HCl buffer
From the top to the bottom, Western blot analysis, CBB staining, and enzyme assay results are shown. CS, culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant. The Ni-NTA column was pre-equilibrated with 50 mM Tris-HCl buffer, 300 mM NaCl, pH 8.0, prior to sample loading.

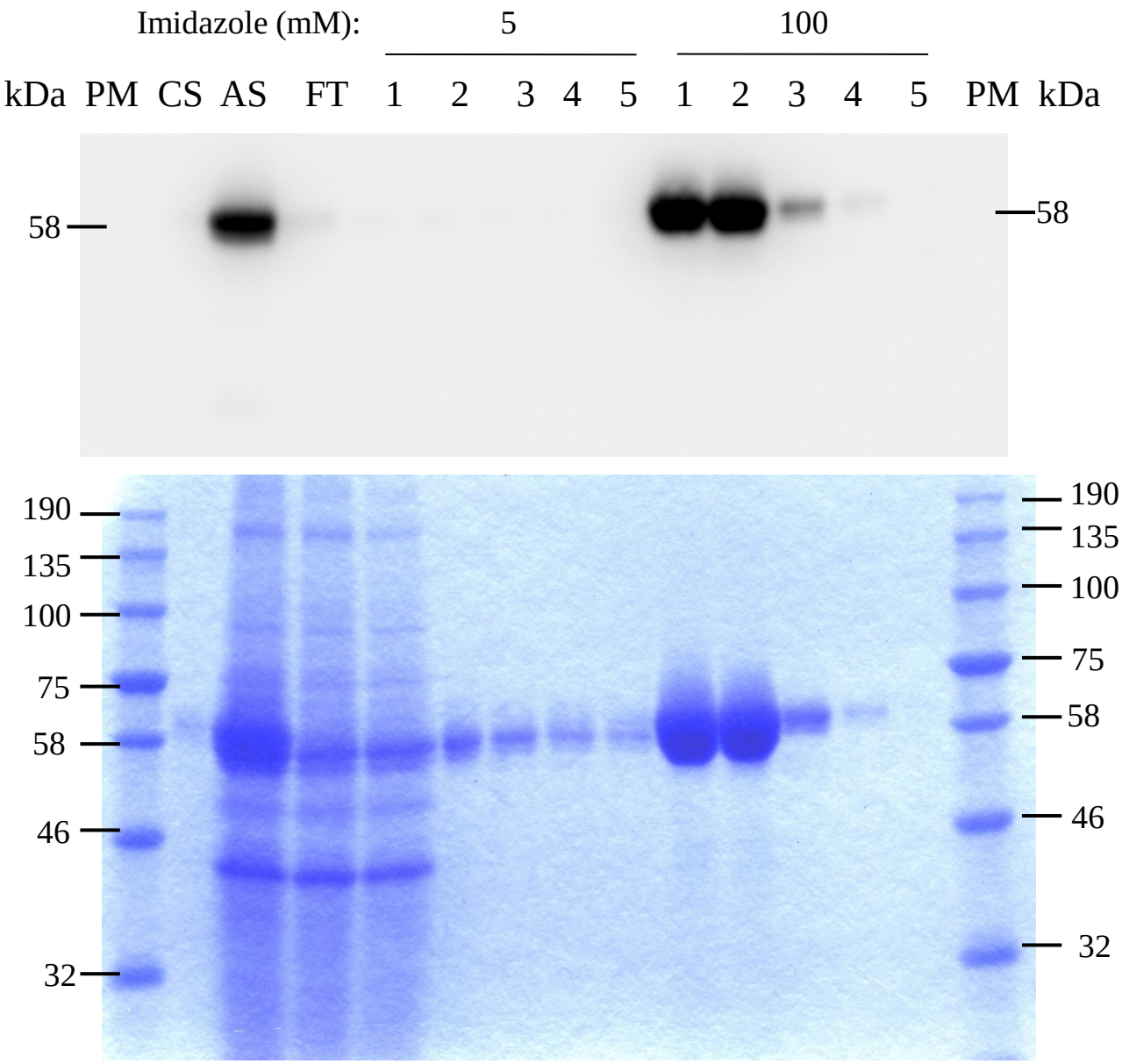


Fig. 2-11. Purification of PaBG1b from crude enzyme solution
The results of Western blot analysis (upper) and CBB staining (lower) of samples from Ni-NTA affinity chromatography are shown. PM, protein marker; AS, crude enzyme solution prepared by re-suspending the ammonium sulfate precipitate of the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant in 50 mM Tris-HCl buffer, 300 mM NaCl, pH 8.0; FT, flow through fraction. The imidazole concentrations are indicated at the top, with 5 mM in the wash buffer and 100 mM in the elution buffer. Purified PaBG1b was instantly dialyzed for 4 h at 4° C to remove imidazole.

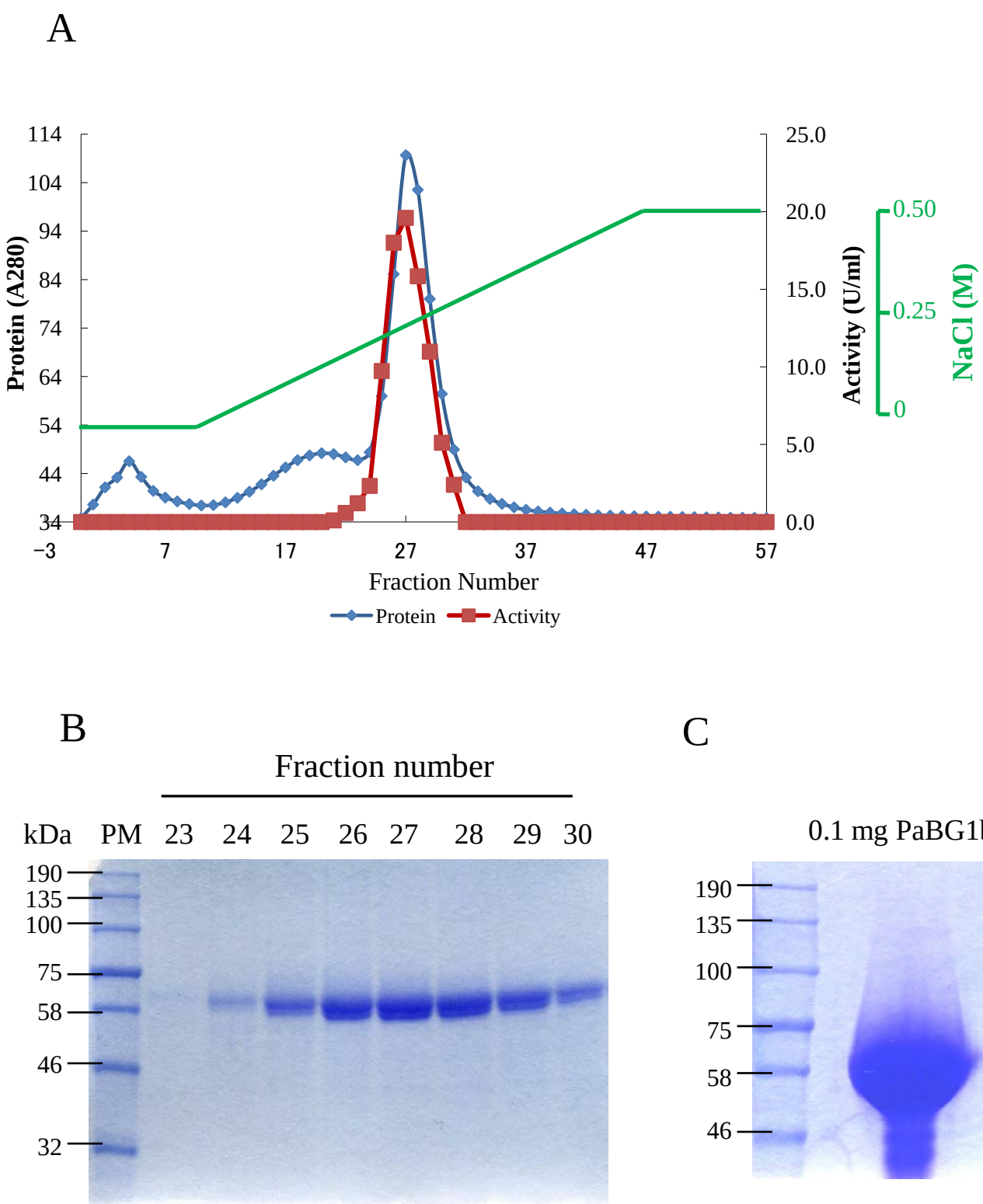
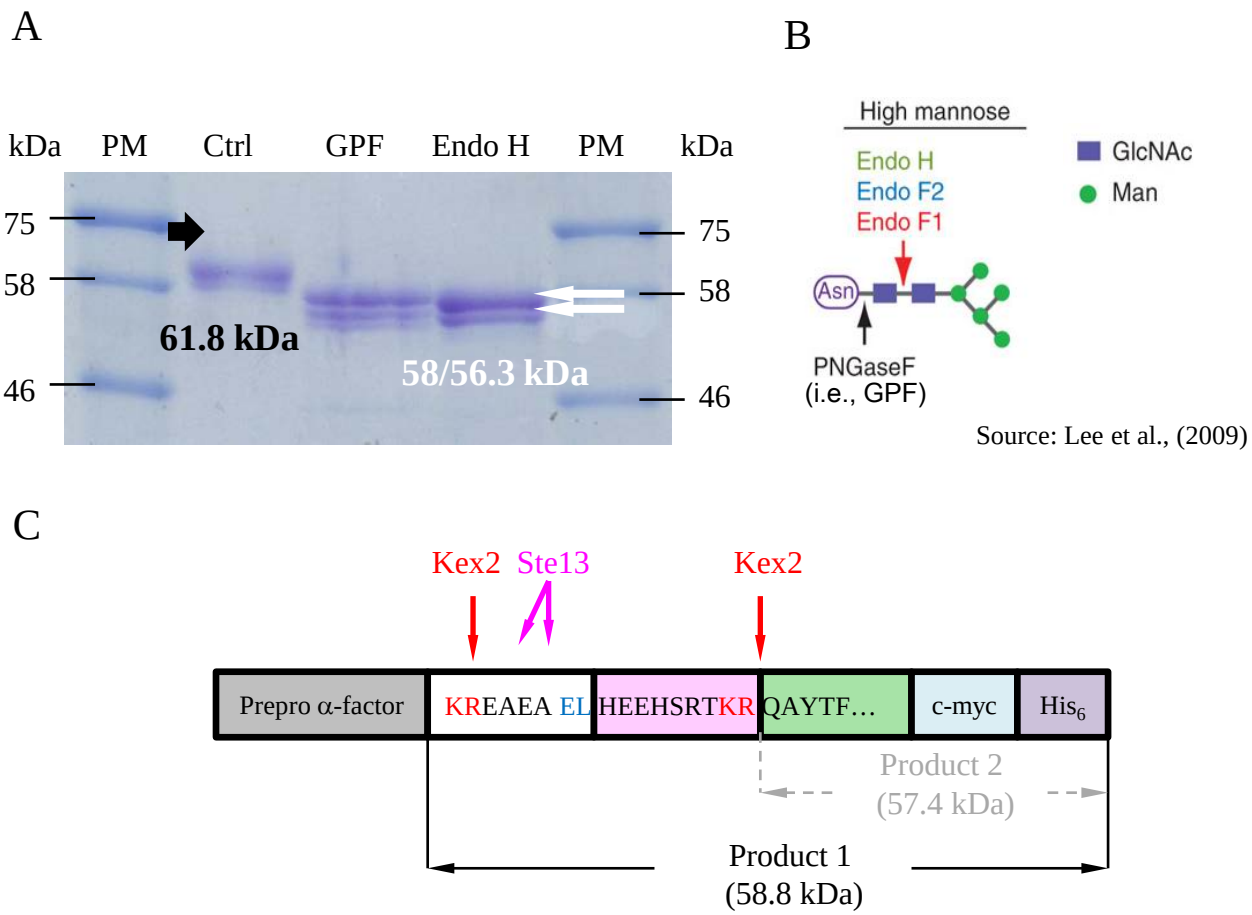


Fig. 2-12. Purification of PaBG1b by anion exchange chromatography
(A) The profile of HiTrap DEAE FF column chromatography. (B) CBB-stained gel of the fractions containing PaBG1b. (C) 0.1 mg of purified PaBG1b was resolved on SDS-PAGE gen for purity check.



Source: Lee et al., (2009)

Fig. 2-13. Deglycosylation of PaBG1b

(A), Two mg of purified PaBG1b was treated with glycopeptidase F (GPF) or endoglycosidase H (Endo H). The sample was boiled to denature the protein prior to GPF and Endo H treatment, as described in the manuals of two enzymes. Ctrl, untreated control. The solid arrow indicates the band of intact PaBG1b with the size of approximately 61.8 kDa. The bands of endo H-deglycosylated PaBG1b are indicated by two open arrows with the apparent sizes around 58 kDa and 55.8 kDa, respectively. (B), Scheme for deglycosylation by different glycosidases. (C), Putative processing of PaBG1b precursor by proteases and two resultant products. Dipeptide 'EL' in blue is a vector-derived spacer dipeptide. (D) Native PAGE of PaBG1b.

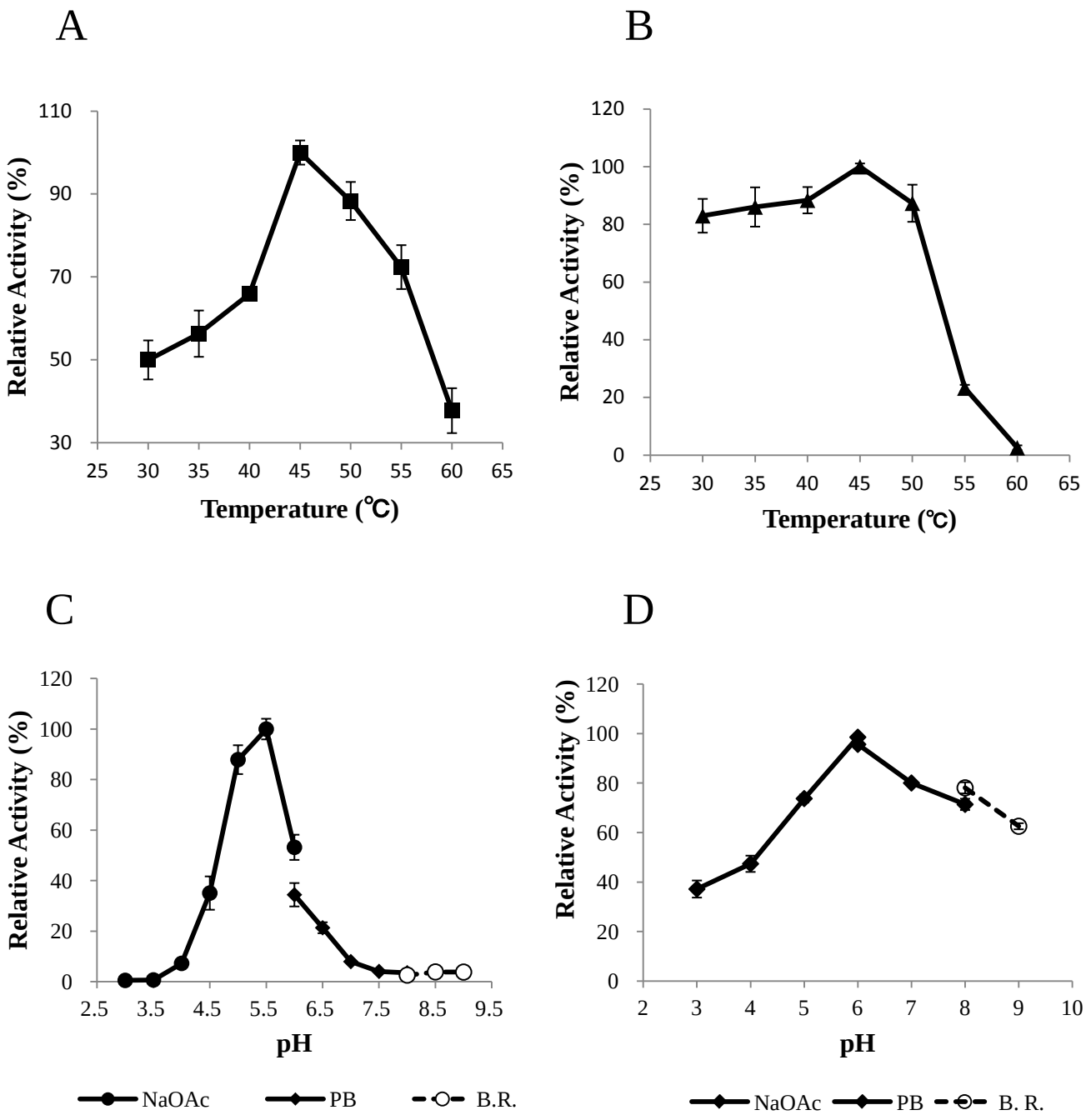


Fig. 2-14. Optimum temperature (A), thermostability (B), optimum pH (C), and pH stability (D) of PaBG1b

(A), purified PaBG1b was subjected to routine enzyme assay at temperatures ranging from 30° C to 60° C. (B), PaBG1b was pre-incubated at different temperatures for 30 min, and the residual activity was measured. (C), the enzyme reaction was conducted at different pH over the range of pH 3.0 to 9.0 to determine the optimum pH of PaBG1b. (D) PaBG1b was mixed with different buffer sets at pH ranging from 3.0 to 9.0, incubated for 30 min at 30° C, and the residual activity was measured. The buffers employed were: 50 mM sodium acetate (pH 3.0 to 6.0, closed circles), 50 mM sodium phosphate (pH 6.0 to 8.0, closed diamonds), 50 mM Britton-Robinson (H_3BO_3 + CH_3COOH + H_3PO_4, pH 8.0 to 9.0, open circles). Data are means ± S.D. of three independent experiments.

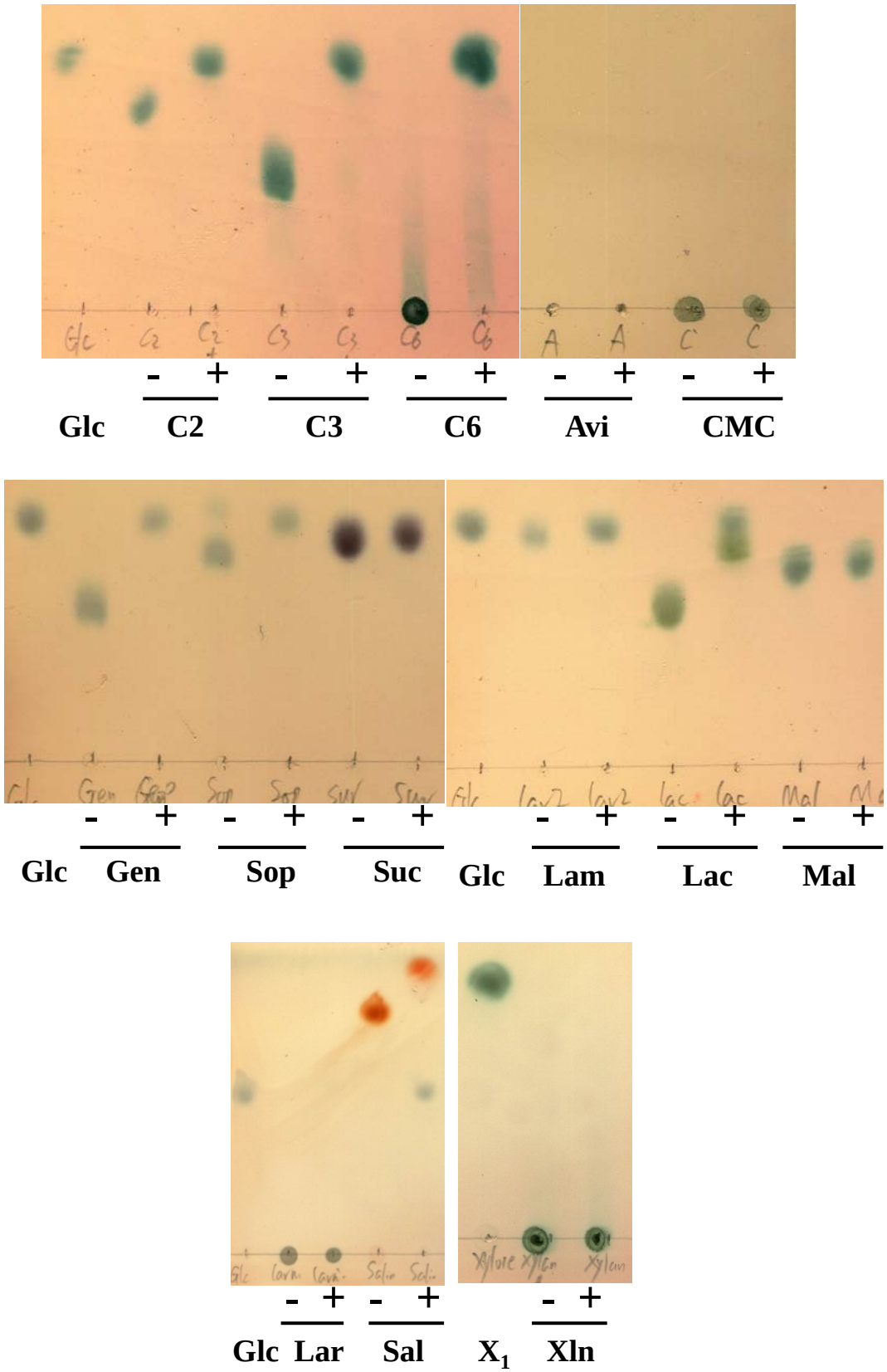
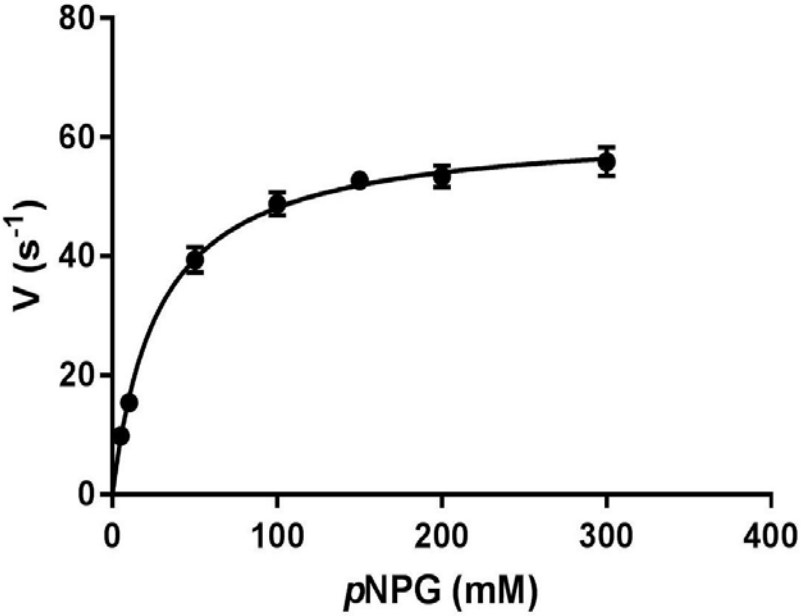


Fig. 2-15. Thin layer chromatography (TLC) analysis of hydrolyzed products by PaBG1b
Glc, glucose; C2, C3, and C6: cellobiose, cellotriose, and cellohexaose; Avi, Avicel; CMC, carboxymethyl cellulose; Gen, gentiobiose; Sop, sophorose; Suc, sucrose; Lam, laminaribiose; Lac, lactose; Mal, maltose; Lar, laminarin; Sal, salicin; X_1, xylose; Xln, xylan. Saccharides (1% (w/v) in 50 mM sodium acetate buffer, pH 5.5) were mixed with (+) or without (-) PaBG1b and incubated for 1 h at 37° C. The enzyme-substrate ratio was 1: 10, and the enzyme concentration in the reaction mixture is 0.2 U/ml (towards cellobiose).

A



B

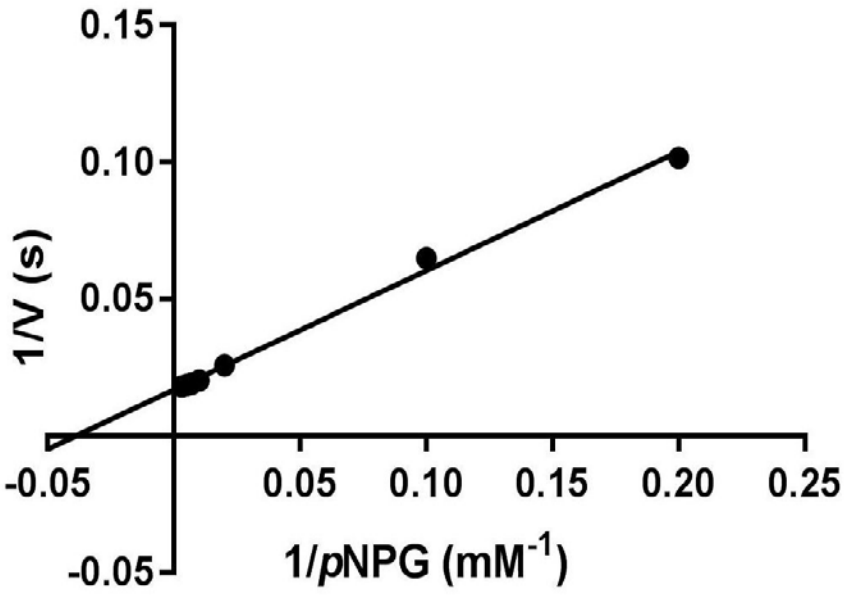
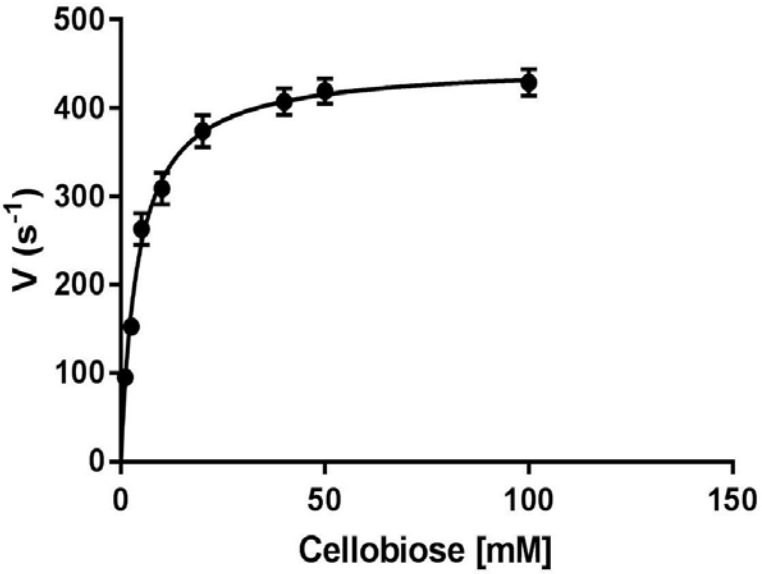


Fig. 2-16. Determination of kinetic parameters of PaBG1b towards *p*NPG
Michaelis-Menten (A) and Hanes-Woolf (B) plots of PaBG1b using *p*NPG as a substrate are shown. Reactions were conducted at 45° C, pH 5.5, and initiated by adding 10 μl of enzyme solution (diluted to the final concentration of approximately 0.008 mg/ml) into 100 μl of substrate solution containing different concentrations of *p*NPG ranging from 5 mM to 300 mM. After incubation for 10 min, the reaction was stopped by adding 1 ml of sodium carbonate, and the reaction velocity was determined by measuring the released *p*-nitrophenol at A_{410}. V_{max}, K_m and k_{cat} were calculated by a nonlinear regression of the Michaelis-Menten equation using GraphPad PRISM 7.

A



B

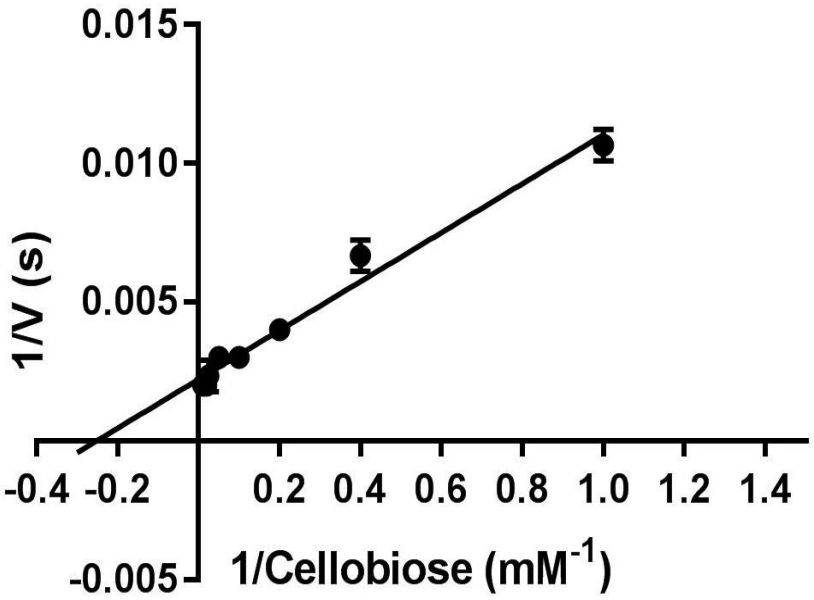


Fig. 2-17. Determination of kinetic parameters of PaBG1b towards cellobiose
Michaelis-Menten (A) and Hanes-Woolf (B) plots of PaBG1b using cellobiose as a substrate are shown. Reactions were conducted at 37° C, pH 5.5, and initiated by adding 25 μl of enzyme solution (diluted to the final concentration of approximately 0.008 mg/ml) into 100 μl of substrate solution containing different concentrations of cellobiose ranging from 2.5 mM to 150 mM. After incubation for 5 min, the reaction was stopped by boiling for 5 min, and the reaction velocity was determined by measuring the release of glucose using glucose oxidase kit. One unit (U) of BG activity was defined as the amount of enzyme that degrades 1 μmol of cellobiose per min. V_{max}, K_m and k_{cat} were calculated by a nonlinear regression of the Michaelis-Menten equation using GraphPad PRISM 7.

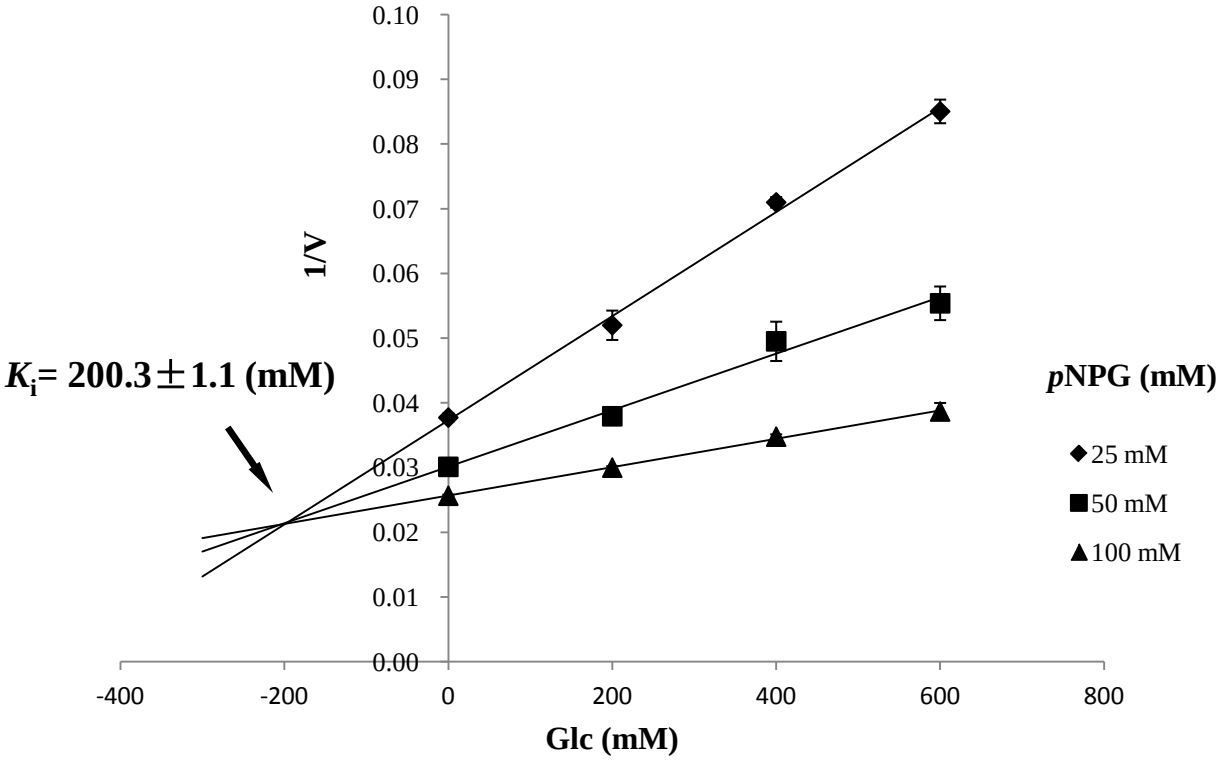


Fig. 2-18. Inhibition constant (K_i) of glucose towards PaBG1b
K_i value, expressed as a mean ± S.D. of three independent experiments with similar results, was obtained by Dixon plot. Data shown are from one representative experiment conducted in triplicate.

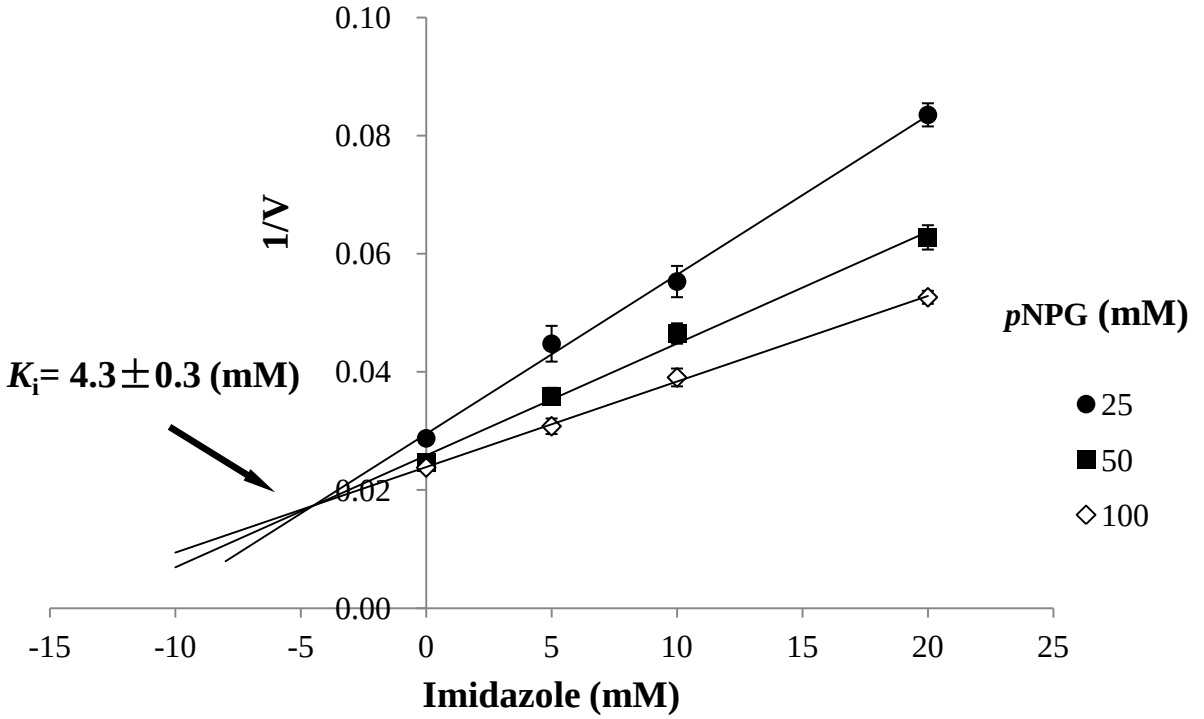


Fig. 2-19. Inhibition constant (K_i) of imidazole towards PaBG1b at pH 5.5
*K*i value, expressed as a mean ± S.D. of three independent experiments with similar results, was obtained by Dixon plot. Data shown are from one representative experiment conducted in triplicate.

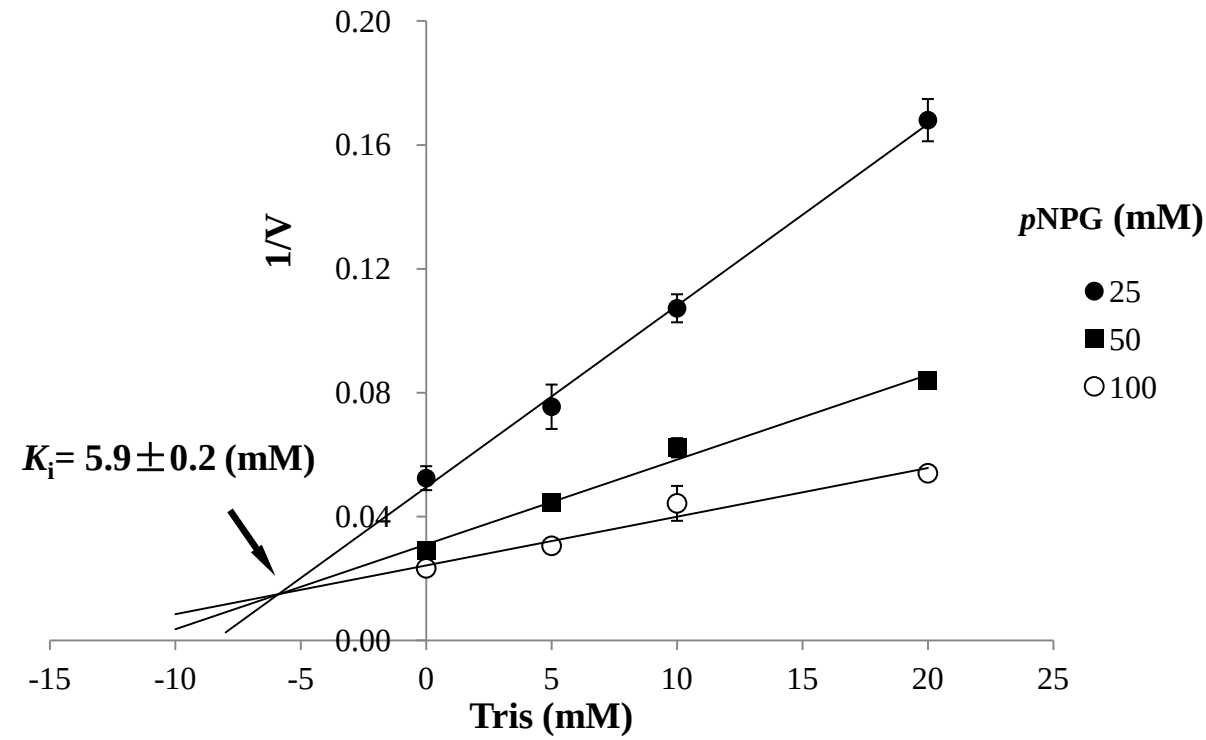


Fig. 2-20. Inhibition constant (K_i) of Tris towards PaBG1b at pH 5.5
*K*i value, expressed as a mean ± S.D. of three independent experiments with similar results, was obtained by Dixon plot. Data shown are from one representative experiment conducted in triplicate.

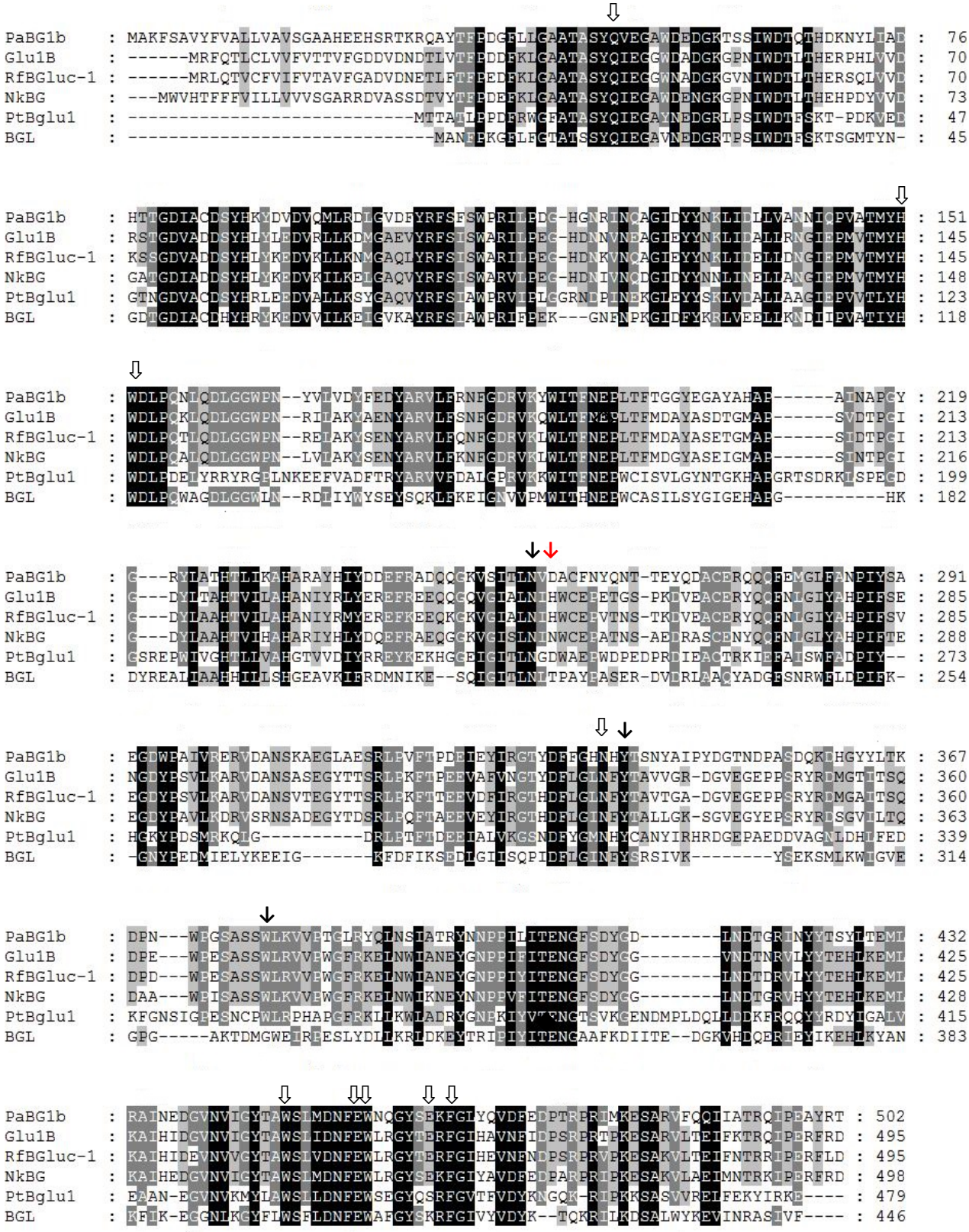
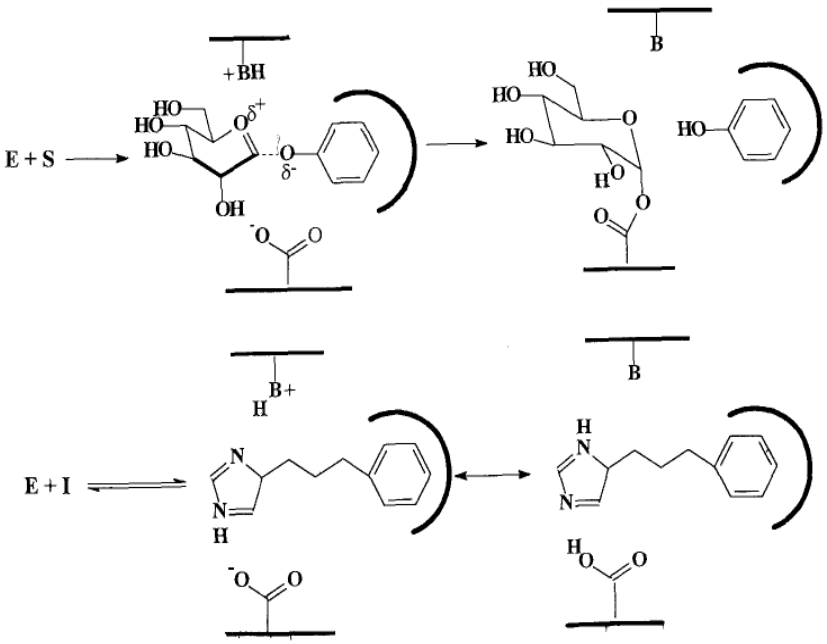
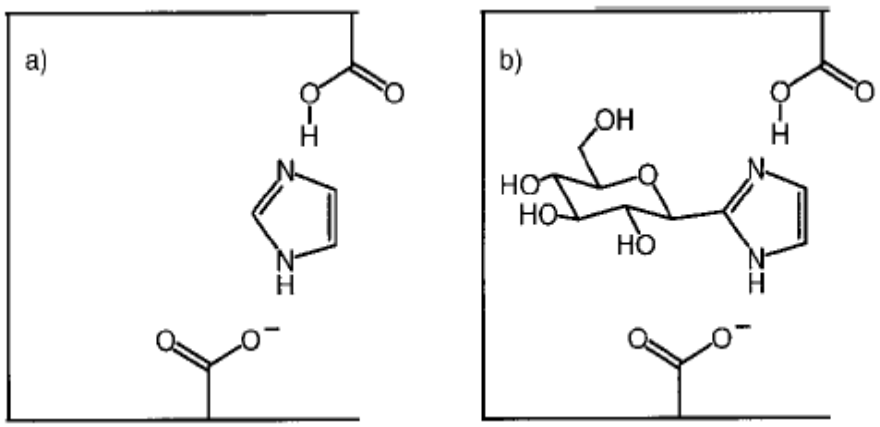


Fig. 2-21. Alignment of PaBG1b, NkBG and GH1 BGs in Table 0-2 whose specific activity or V_{max} over 200 U/mg. PaBG1b is aligned with NkBG and Glu1B (GenBank accession number: GQ911585), PtBglu1 (HM036350), *Rf*BGluc-1 (HM152540), and BGL of *T. aotearoense* P8G3#4 (KP772230), is shown. Residues equivalent to those of which formed interactions with pNPG and located in the glycone binding pocket (QG45, H148, W149, N335, W444, E451, W452, E458 and F460) and the aglycone binding pocket (N253, N255, Y337 and W374) of NkBG (Jeng *et al*., 2011) are marked by empty (⇩) and filled (↓) arrows, respectively. Residues equivalent to N255 of NkBG, which are highly variable among BGs aligned, are boxed with a solid line.



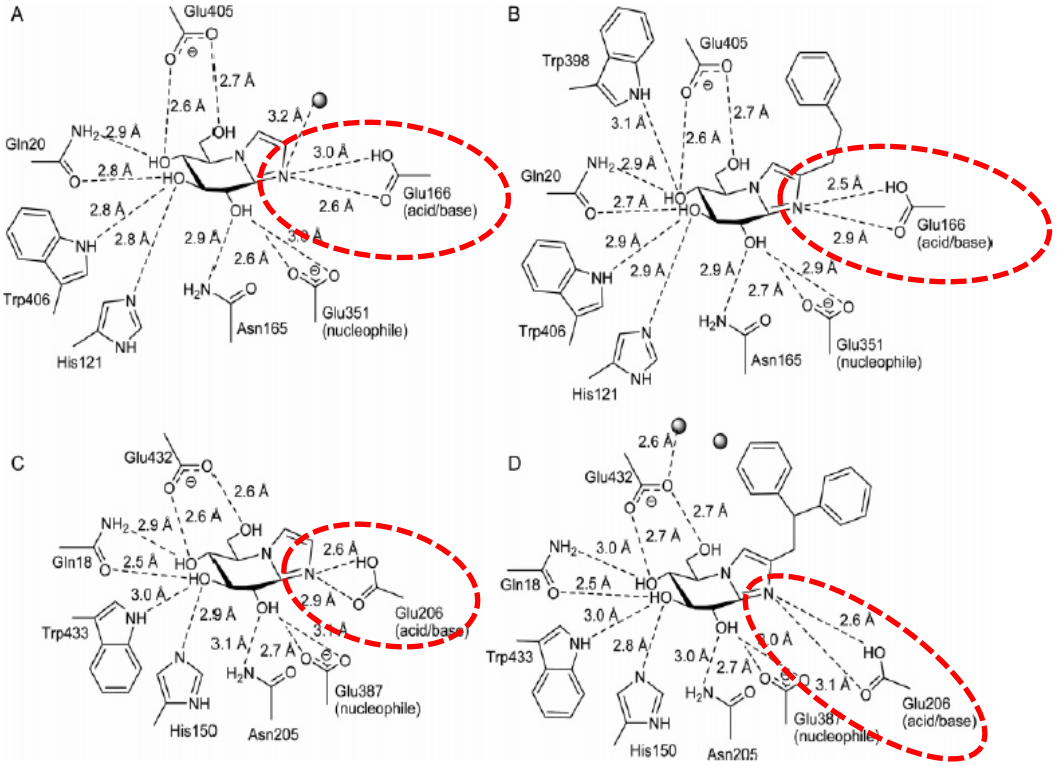
Source: Li *et al*., (1998)

Fig. S2-1. Schemes of the proposed binding models for an artificial substrate (upper) and 4-(3'-phenylpropyl) imidazole (lower) with almond β-glucosidase
The aryl ring was supposed to bind with the aglycone binding subsite of the enzyme.



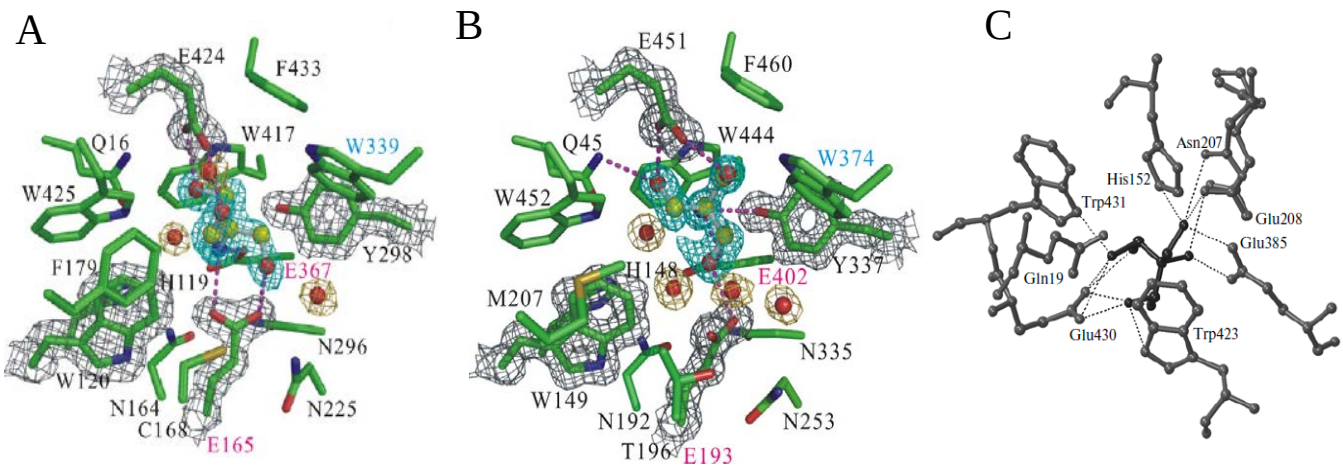
Source: Heightman and Vasella (1999)

Fig. S2-2. Mechanism of inhibition of glycosidases by imidazole (a) and β-D-glucopyranosylimidazole (b).



Source: Gloster *et al*. (2006)

Fig. S2-3. Interactions of *Tm*GH1 with glucoimidazole (A) and phenethyl-substitued glucoimidazole (B), and *Ss*GH1 with glucoimidazole (C) and phenethyl-substitued glucoimidazole (D) by X-ray crystallography.
The dotted circles in red indicate the nitrogen atoms of the imidazole moiety which form hydrogen bonds with the catalytic acid/base residues Glu166 of *Tm*GH1 and Glu206 of *Ss*GH1.



Source: Jeng *et al*. (2011) (for A and B); Trofimov *et al*. (2013) (for C)

Fig. S2-4. Binding of Tris in the active sites of three GH1 BGs.
Three-dimensional ball-and-stick models of Tris in complex with *Tr*BGl2 from *T. reesei* (A), *Nk*BGl from *N. koshunensis* (B), and Asβ-Gly from *Acidilobus saccharovorans* (C). In (A) and (B), the carbon atoms of Tris molecules are shown in yellow, whereas the hydrogen bond interactions of the residues and Tris molecules are labeled by dotted lines in magenta. In (C), the hydrogen bonds are indicated by dashed lines.

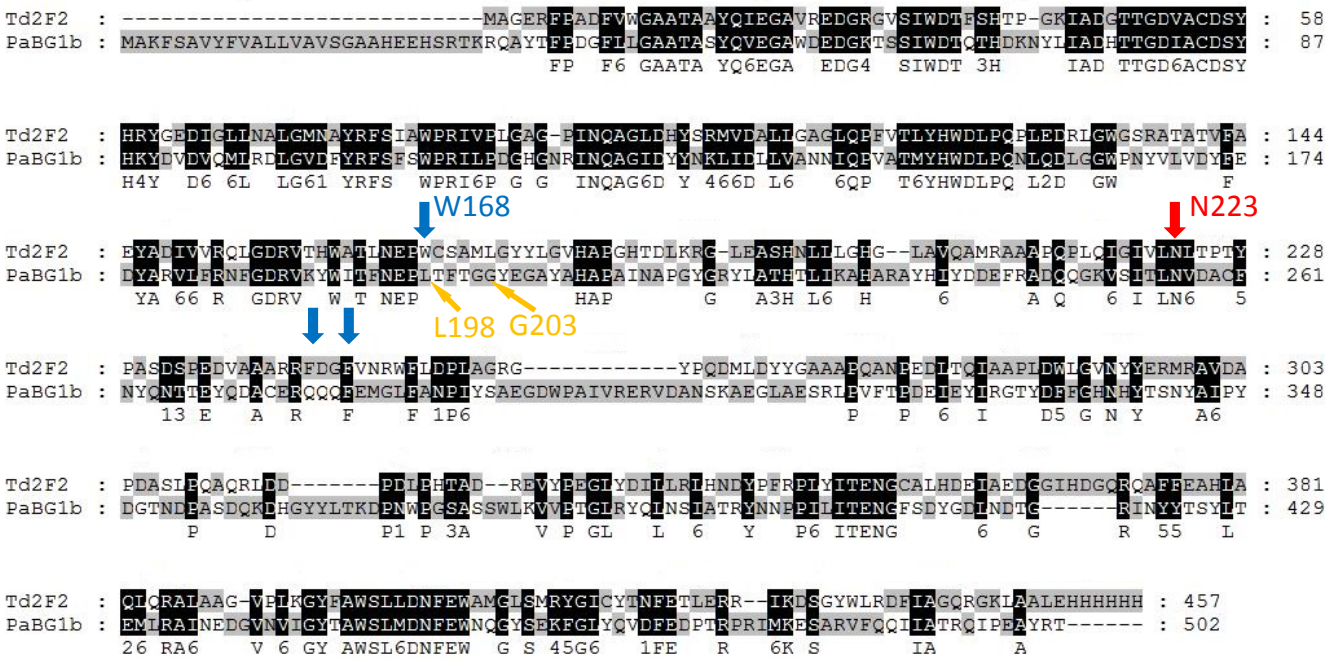


Fig. S2-5. L198W and/or G203L mutation of PaBG1b might improve its glucose tolerance and thermostability.

N223 of Td2F2, which is highly conserved in other GH1 BGs and strongly related to glucose tolerance/activation and substrate specificity, is indicated by a red arrow, and the aromatic side chains of W168, F243, and F246 of Td2F2 which forming a partially hydrophobic pocket located near subsite +2 and critical for the glucose tolerance and substrate specificity of Td2F2 by providing an additional binding site for glucose, are indicated by blue arrows (Matsuzawa *et al*., 2016). On the other W168 and L173 of HiBG which play a role in glucose tolerance (de Giuseppe PO *et al*., 2014) are conserved in Td2F2, and their equivalents in Td2F2 are conserved (W168 and L173), but not occured in PaBG1b (L198 and G203, indicated by yellow arrows). Therefore, single or double mutation of L198W and G203L might possibly improve the performance of PaBG1b in glucose tolerance. In addition, the L167W mutation of TrBgl2 led to an increase of optimum temperature (10 ° C), and glucose tolerant (3 fold), as compared with the wild-type (Lee *et al*., 2012). The counterpart of L167 of TrBgl2 is W168 of HiBG and Td2F2, and L198 of PaBG1b. Therefore, L198W mutation of PaBG1b might be an viable approach to promote its thermostability.

Chapter 3

Heterologous expression in *P. pastoris*, purification, and characterization of RsBG

3.1 Introduction

3.1.1 Termites and its gut symbionts are novel reservoirs of cellulases

Termites (Isoptera) are eusocial cockroaches (Inward et al., 2007) and mostly populate in the tropics and subtropics (Breznak and Brune, 1994). Presently, there are over 2,600 described termite species in 281 genera (Kambhampati and Eggleton, 2000), and around 500-1,000 species are still left to be described (Eggleton, 2011). Termites are phylogenetically divided into two groups: lower and higher termites. The lower termites harbor a diverse array of bacteria and cellulose-digesting flagellate protozoa in their alimentary tract, whereas higher termites typically lack protozoa (Breznak and Brune, 1994). Traditionally, lower termites are classified into six extant families including Mastotermitidae, Kalotermitidae, Hodotermitidae, Rhinotermitidae, Serritermitidae (Breznak and Brune, 1994), and Termopsidae (Eggleton, 2001). Higher termites, which are comprised of the majority of all termite species, are affiliated with the family Termitidae, and have a more sophisticate external and internal anatomy and social organization than those of the lower termites (Breznak and Brune, 1994).

It is estimated that termites consume approximately 3-7 billion tons of lignocellulose annually, which accounts for an important process in the tropical and subtropical ecosystems and is unique for a single order of invertebrates (Tokuda et al., 2014). Furthermore, ruminants generally feed on herbaceous plants such as soft grasses but are incapable of digesting woody plants (Watanabe and Tokuda, 2010), whereas termites use wood as the primary nutrition source and are capable of digesting 74-99% of cellulose they ingest (Prins and Kreulen, 1991), which makes them to be the most efficient decomposers of wood on earth (Watanabe and Tokuda, 2010). Termite guts are believed to be the world's smallest bioreactors in which the lignocellulose is digested

through both aerobic and anaerobic processes (Brune, 1998). As the strategies of termites employed for breakdown of lignified plant cell walls resemble the industrial processes, and digestion of wood requires overcoming the barrier imposed by the recalcitrant lignin matrix, termites are deemed to be a high efficiency model for the industrial conversion of lignocellulose and a promising reservoir of cellulolytic enzymes (Brune, 2014).

Although termites are obviously capable of utilizing cellulose, it took a long time to understand their digestion systems. Almost one century ago, Cleveland studied the viability of some species of lower termites (mostly *R. flavipes* Kollar) on wood and provided the first experimental evidence of the symbiosis nature between the termites and intestinal protozoa (Cleveland, 1923). Half a century ago, in 1964, Yokoe firstly hypothesized that *Leucotermes* (=*Reticulitermes*) *speratus* itself had endogenous cellulases (Watanabe et al., 1997). In 1975, Yamaoka and Nagatani firstly demonstrated that the EG activity was present in the salivary glands of *R. speratus* and its hindgut protozoa possessed the ability to hydrolyze crystalline cellulose (Watanabe et al., 1997), and set forth a proposal of dual source of cellulases in *R. speratus*, although the original paper was written in Japanese (Watanabe and Tokuda, 2010). After two decades, Watanabe et al. further purified two EGs from the salivary glands of *R. speratus* (Watanabe et al., 1997), and obtained the cDNA of the first termite-origin EG, *rseg*, by using the antiserum raised against one of the EGs (Watanabe et al., 1998). Later, Ohtoko et al. isolated 15 cellulase genes from the symbiotic protists in the hindgut of *R. speratus* (Ohtoko et al., 2000). Nakashima et al. investigated the distribution of EG components in the digestive system of the lower termite *C. formosanus* Shiraki and concluded that *C. formosanus* had two independent cellulose-digesting systems: one in

the midgut, where cellulose is digested by endogenous cellulases, and another in the hindgut, in which cellulose degradation is possibly carried out by symbiotic flagellates-derived cellulases (Nakashima et al., 2002b). They subsequently isolated the cDNA and functionally expressed a CBH of the symbiotic flagellate in the hindgut of *C. formosanus* (Nakashima et al., 2002a). Owing to these milestone works, the hypothesis that the cellulose degradation system of the lower termites is a combination of endogenous cellulases and cellulolytic symbionts, which has been proposed for a long time, was finally evidenced. A scheme of the dual cellulose-digesting system of the lower termites is shown in Fig. 3-1A.

As for the higher termites, generally many species of higher termites do not feed on wood, and xylophagous higher termites are well adapted to a wood diet without the aid of symbiotic protists (Tokuda et al., 2012). Almost four decades ago, Martin and Martin demonstrated that the fungus-growing termite *Macrotermes natalensis* secreted CBH and BG from its midgut epithelium and salivary gland, whilst acquired EG from the fungus grown in their nests (Martin and Martin, 1978). O'Brien further showed that higher termite *Nasutitermes exitiosus* had endogenous cellulases and was not dependent on its gut flora for the digestion of cellulose (O'Brien et al., 1979). Their works are the early experimental evidences that suggest that higher termites had their own cellulolytic enzymes. Additionally, metagenomic analysis (Warnecke et al., 2007) and zymogram analysis (Tokuda and Watanabe, 2007) revealed that the symbiotic bacteria of higher termite also harbored considerable cellulolytic activities and could be a novel source of cellulases.

Presently, an increasing number of cellulase genes (EGs and CBHs) from termites were isolated (Watanabe et al., 2002), heterologously expressed, and characterized

(Tokuda et al., 1999; Zhang et al., 2009; Zhou et al., 2010; Todaka et al., 2011). Heterologous expression of EGs (Inoue et al., 2005; Sasaguri et al., 2008; Otagiri et al., 2013) and CBHs (Sethi et al., 2013) derived from the termite symbiotic protists were also achieved. Two EGs (RsEG and NtEG) of termites (Hirayama et al., 2010) and one EG (RsSymEG) from the symbiotic protist of *R. speratus* (Todaka et al. 2010) were expressed in *A. oryzae* in our lab, and their activities were found to be superior to extant EGs in that they displayed higher activity than those like commercial EGs from *Trichoderma* spp. As for the termite-origin BGs, NkBG (Tokuda et al., 2002) and sgNtBGs (Tokuda et al., 2009) are the first endogenous BGs of the lower and higher termites characterized, respectively. Some other termite-origin BGs displaying high activities are summarized in Table 0-2. As for the BGs derived from termite gut microbiota, until now, there are at least two characterized BGs. One is BglB, a bacterial enzyme of the gut microbiota of the lower termite *Reticulitermes santonensis* (Mattéotti et al., 2011), which demonstrated both BG and β-xylosidase activities. Another is Bgl-gs1, whose gene was isolated from the metagenomic library of the gut of the higher termite *Globitermes sulphureus*, and the recombinant enzyme was highly thermostable showing a half-life of approximately 1 h at 90°C (Wang et al., 2012). In addition, β-Glucosidase B, a native enzyme purified from the symbiotic fungus *Termitomyces* sp. of the higher termite *M. muelleri*, displayed a specific activity towards cellobiose of 207 U/mg (Rouland et al., 1992, as cited in Table 0-2).

3.1.2 BGs from *R. speratus* and its symbiotic protists

The lower termite *R. speratus* is the most common subterranean termite in Japan and known as "*Yamato-shiroari*" which literally means "Japanese white-ant" (Inoue et

al., 1997; Watanabe et al., 1997). Furthermore, *R. speratus* is one of the most extensively studied termites in which both the cellulolytic and xylanolytic systems were analyzed, together with the molecular phylogeny of its symbiotic protists (Ohtoko et al., 2000). In terms of BGs from *R. speratus* and the symbiont protists thereof, Inoue et al. analyzed the distribution of enzymes of cellulose metabolism and found that only 23.9% of the BG activity was in the salivary grand, and at least 70% of the remaining BG activity of *R. speratus* located in the hindgut (Inoue et al., 1997). In the same year, Watanabe et al. isolated one major BG from the whole body extracts of *R. speratus*, which was suggested to be endogenous, and three minor BGs in the hindgut extracts, which were suggested to be protozoan origin (Watanabe et al., 1997). From these results it could be inferred that the BGs from the symbiotic protists might have high specific activities and/or catalytic efficiencies, as they contribute to the major BG activity with little amount of protein. However, the cellulolytic flagellates in the hindgut of termites live in an anaerobic environment and often harbor intracellular bacterial symbionts, which makes it extremely difficult to establish axenic cultures and impedes the study of cellulases they produce (Yamin, 1978). Todaka et al. performed the transcriptome analysis in relation to the lignocellulose digestion in the symbiotic protist community of *R. speratus*, and found that the BG cDNAs are solely from GH3 (Todaka et al., 2007). Among them, the full-length cDNA of a GH3-like BG, RsBG, was obtained and deposited into the GenBank database with the accession number BJ978994. A previous effort in our laboratory to express RsSym3BG1 (called RsBG in this study) in *A. oryzae* was a modest success (Sasaguri et al., 2008), since the BG activity of the culture supernatant of the RsBG transformant was only marginally higher than that of the control strain, and the target protein was not purified to homogeneity. In this Chapter, *P.*

pastoris was employed as the expression host of RsBG.

Besides, Shimada and Maekawa cloned two types of BG homologues (RsBGI and RsBGII) of *R. speratus* and suggested that RsBGI was involved in cellulose digestion, whereas RsBGII might be a main component involved in the social communication, for example, the egg-recognition pheromone (Shimada and Maekawa, 2014). It is unclear, however, which of the two genes correspond to the BG purified by Watanabe et al. (Watanabe et al., 1997). In conclusion, to date neither endogenous nor symbiotic protist-derived BG of *R. speratus* has been characterized. Therefore, overexpression and characterization of RsBG is particularly meaningful for acquiring a powerful BG and revealing the complete gene set in relation to cellulose digestion in the exemplary lower termite *R. speratus*.

3.1.3 Structural features of GH3 BGs

According to Table 0-1, BGs in both GH1 and GH3 employ the same retaining mechanism for hydrolysis reaction, and use glutamic acid (E) as the catalytic acid/base residue. However, GH1 and GH3 BGs are experimentally characterized to use glutamic acid (E) and the aspartic acid (D) as their catalytic nucleophile, respectively. Although GH3 is one of the major families that BGs are affiliated with, to date only a few GH3 BGs have been structurally characterized. Up to May 2016, there are totally 14 crystallized GH3 BGs deposited in the CAZy database (http://www.cazy.org/GH3_structure.html; summarized in Table 3-1), but two BGs (BglB (larminaribiase) of *Thermotoga neapolitana* Z2706-MC24 and tomatinase from *Septoria lycopersici*) are not available in PDB bank. The remaining 12 BGs can be divided into three groups by their origins: (1) three from bacteria, i.e. ExoP (PDB ID:

3F93) of *Pseudoalteromonas* sp. BB1 (Nakatani et al., 2012), *Sv*DesR (4I3G) of *Streptomyces venezuelae* ATCC 15439 (Zmudka et al., 2013), and *Tn*Bgl3B (2X41) of *T. neapolitana* DSM 4359 (Pozzo et al., 2010); (2) eight from eukaryota, i.e. AaBGL1 (4IIB) of *A. aculeatus* F-50 (Suzuki et al., 2013), *Af*βG (5FJI) of *A. fumigatus* Af293, *Ao*βG (5FJJ) of *A. oryzae* RIB40 (Agirre et al., 2016), *An*Bgl1 of *A. niger* (Lima et al., 2013), ExoI (1EX1) of *H. vulgare* subsp. *vulgare* (which is an exo-β-1,3-1,4-glucanase (EC 3.2.1.-) (Varghese et al., 1999)), *Km*Bgl1 (3ABZ) of *Kluyveromyces marxianus* NBRC1777 (Yoshida et al., 2010), *Re*Cel3A of *Rasamsonia emersonii* IMI 392299 (Gudmundsson et al., to be published), and *Hj*Cel3A (3ZYZ) of *T. reesei* QM9414 (Karkehabadi et al., 2014); and (3) one unclassified BG, JBM19063 (3U4A) screened from the compost microbe (McAndrew et al., 2013). All of the structurally-resolved GH3 BGs employ multidomain architectures, with the three-domain organization being the most common, as summarized in Table 3-2. With three exceptions (ExoI, *Sv*DesR, and *Km*BglI), all of the architectures of GH3 BGs mentioned above contain (1) N-terminal domain with $(\beta/\alpha)_8$-barrel (i.e. canonical TIM barrel) or $\beta\beta(\beta/\alpha)_6$-barrel structure, (2) an α/β-sandwich domain placed in the middle (usually $(\alpha/\beta)_6$-sandwich), and (3) C-terminal fibronectin type III-like (FnIII-like) domain. The architecture of ExoI is more sparing since it lacks the FnIII-like domain, whilst those of *Sv*DesR and *Km*BglI are more complicated in that they possess an extra PA14 domain.

Base on the analysis of the three-dimensional structure of ExoI, Varghese et al. found that the active sites of ExoI are located at the domain interface of the N-terminal $(\beta/\alpha)_8$-barrel and the $(\alpha/\beta)_6$-sandwich domains, with the N-terminal domain harboring the nucleophile and the $(\alpha/\beta)_6$-sandwich domain housing the catalytic acid/base residues, respectively, and suggested that other plant homologues of ExoI have similar structures

(Varghese et al., 1999). The structures of recently-resolved GH3 BG are in agreement with this suggestion (Agirre et al., 2016), and this fact is also apparent as shown in Table 3-1 (for the catalytic residue placement) and Table 3-2 (for the domain arrangement). Harvey et al. investigated the three-dimensional structures of approximately 100 GH3 members (mostly microbial BGs) via comparative modeling analysis, and concluded that their domain orientations are generally the same as ExoI, although there are still some exceptions such as two bacterial GH3 BGs from *Butyrivibrio* and *Ruminococcus* that show reverse orientation of domains (Harvey et al., 2000).

The catalytic nucleophile residue of GH3 BG is rather conserved and readily identified by the amino acid sequence alignment. In 1980, Bause and Legler labeled the catalytic nucleophile residue (D) of β-glucosidase A3 from *Aspergillus wentii* in the active site by radioactive inhibitor conduritol B-epoxide (Bause and Legler, 1980). In 1992, by the amino acid sequence alignment of *A. wentii* A3 and its homologues, the consensus sequence 'GFVMSDW' encompassing the active site was firstly proposed (Castle et al., 1992), and in the same year, the PROSITE motif of PS00775 which is depicted as the consensus pattern of [LIVM](2)-[KR]-x-[EQKRD]-x(4)-G-[LIVMFTC]-[LIVT]--[LIVMF]-[ST]-D-x(2)-[SGADNIT] (http://prosite.expasy.org/PS00775) at the active site region of GH3 enzymes was established, with 'D' being the active site residue (http://prosite.expasy.org/PDOC00621). The amino acid sequence of 'S/TDW' was later identified to be the conserved motif of GH3 BGs (Janbon et al., 1995), and the PROSITE PS00775 harboring the SDW motif was suggested to be used as a signature region for the active site of GH 3 enzymes (Iwashita et al., 1999). Base on the

comparative modeling of the three-dimensional structures of ExoI and other GH3 members, Harvey et al. concluded that the aspartic acid (D) serving as the catalytic nucleophile in the GFVISDW sequence of ExoI is highly conserved, and the 'D' is absolutely conserved in all GH3 enzymes (Harvey et al., 2000).

As far as the catalytic acid/base residue is concerned, contrary to the easily-identified catalytic nucleophile residue, the acid/base residue is not highly conserved among GH3 BGs, especially in the distantly-related members of the family (Harvey et al., 2000). Recently, a novel bioinformatic approach had been developed for the assignment of the phylogenetically-variable acid/base catalyst of GH3 members via three-dimensional structure homology modeling and detailed kinetic analysis of site-directed mutants (Thongpoo et al., 2013).

As for the tertiary structures of GH3 BG, five of structurally-resolved BGs are dimers, whilst KmBglI is a tetramer, as shown in Table 3-1.

3.2 Results

3.2.1 Sequence analysis of RsBG

The full-length cDNA of *rsbg* encodes a polypeptide of 573 amino acids containing an N-terminal signal sequence of 13 amino acids predicted by SignalP 4.1 Server (Fig. 3-2A). BLAST P search suggested that mature RsBG is composed of two putative domains: the BglX domain (periplasmic BG and related glycosidases) and the Glyco_hydro_3_C domain (GH3 C-terminal domain), respectively (Fig. 3-2A). BglX is a BG derived from *E. coli* and falls into GH3 (Yang et al., 1996). The domain architecture of RsBG is different from the three-domain architecture of structurally-resolved GH3 BGs, but similar to that of ExoI from barley. On the other

hand, although RsBG is originated from a protist, it shared the highest homology (E-values ranging from $2e^{-126}$ to $1e^{-121}$) with bacterial BGs or hypothetical proteins (Fig. 3-2B), and the amino acid sequence identity of these homologues with RsBG is relatively low (38-42%). By the amino acid sequence alignment of RsBG with structurally-characterized GH3 BGs, E275 and E471 (or E483) were tentatively assigned to be the putative catalytic nucleophile and acid/base residues, respectively (see explanation hereinafter). Five potential *N*-glycosylation sites (N47, N78, N171, N449, and N478) were identified by NetNGlyc 1.0 server with a putative *N*-glycosylation site (N443) was missed by the algorithm (Fig. 3-3), whilst NetOGlyc server 4.0 figured out one putative *O*-glycosylation site (T426). A summary of features of RsBG inferred from the amino acid sequence thereof is shown in Table 3-3.

Amino acid sequence alignment of RsBG with five structurally-characterized GH3 BGs is shown in Fig. 3-4. The position of the catalytic nucleophile of RsBG was determined by the consensus sequence of 'GFVMSDW' of GH3 BGs aligned. However, the putative nucleophile residue of RsBG is glutamic acid (E275), rather than the conserved aspartic acid (D). A conserved sequence 'KHFV' is the consensus in all proteins aligned including RsBG in Fig. 3-4, in which the 'KH' dipeptide is a carbohydrate-binding site (Harvey et al., 2000). The putative catalytic acid/base residue of RsBG (E471 or E483) was tentatively assigned by the alignment (Fig. 3-4).

The alignment of the partial amino acid sequence of RsBG surrounding the putative nucleophile residue with all structurally-characterized GH3 BGs is shown in Fig. 3-5. The sequence of *An*Bgl1 was manually cited from the supplemental information 1 and 2 of the source reference (Lima et al., 2013), and deposited in the Supplemental File 2 of this study (page 173), as the authors did not reveal neither the PBD number, GenBank

accession number, nor UniProt ID of *An*Bgl. Although the catalytic residues of three BGs, *Af*βG, *Ao*βG (Agirre et al., 2016), and *Re*Cel3A, (Gudmundsson et al., to be published), were not specified, the alignment shown in Fig. 3-5 indicates that 'D' in the consensus sequence 'GFVMSDW' is highly conserved among all structurally-characterized GH3 BGs.

To examine whether there is any GH3 BG that employs glutamic acid as the catalytic nucleophile, an investigation of all eukaryotic proteins/fragment entries (up to May, 2016) annotated to be GH3 family BGs and deposited in UniProt database was conducted, by aligning them with the structurally-identified eukaryotic GH3 BGs to predict their hypothetical catalytic nucleophiles. Totally 109 proteins were analyzed and the result is summarized in Table 3-4. Aside for 5 small fragments (Q8X1K2, Q9ZU04, Q69FJ8, P29091, and Q69FF5) that are too short to identify the sequence similar to 'GFVISDW', all of the other 104 proteins/fragments had an aspartic acid at the position equivalent to the catalytic nucleophile. Thus, to date there is no eukaryotic GH3 BG employing glutamic acid as its catalytic nucleophile, and RsBG is a GH3-like BG which is rather peculiar in the putative nucleophile residue.

Aside for the difference in the catalytic nucleophile residue, RsBG has another exceptional feature in terms of conserved residues of GH3 BGs. The conserved arginine (R), i.e. R156 of *Aa*BGL1 (Suzuki et al., 2013), R156 of Cel3A (Karkehabadi et al., 2014), R130 of *Tn*Bgl3B (Pozzo et al., 2010), R183 of ExoI (Varghese et al., 1999), and R142 of JMB19063 (McAndrew et al., 2013), is reported to be capable of forming hydrogen bonds with the substrate, and the counterpart of this residue in *Km*BglI (R113) was reported to be involved in the substrate recognition (Yoshida et al., 2010). However, the corresponding residue in RsBG is glycine (G163), not arginine (Fig. 3-4, indicated

by a red arrow).

On the other hand, although BGs in GH1 employ glutamic acid as the nucleophile, they seem to be harboring their function within one domain (the TIM barrel), and the placement of catalytic residues of GH1 BG is vice versa to that of GH3 BG (i.e. the catalytic acid/base residue lies N-terminal to the nucleophile) as shown in Fig. 1-3.

The cDNA sequence of RsBG was uploaded to the online tool GenScript for rare codon analysis and the results showed that RsBG sequence has a CAI value of 0.62 for yeast and 0.64 for *E. coli*, both of which are less than 0.8, a value indicating a desired expression organism (Chapter 1, in page 30). Considering that these two expression systems got the similar scores and RsBG has 5 putative *N*-glycosylation sites, *P. pastoris* was chosen as the host for the expression of RsBG.

3.2.2 Expression of RsBG by pBGP3-RsBG

The *rsbg* fragment was amplified by PCR, which was then ligated with pBGP3 to generate the expression plasmid pBGP3-RsBG (Fig. 3-6A). Transformation to *P. pastoris* was done and the transformants were screened from YPDS plates containing 100 μg/ml Zeocin. The pBGP3-RsBG transformants were verified by colony PCR, transferred to 10 ml of YPD medium containing 100 μg/ml Zeocin in 50 ml of flask, and incubated at 30°C for 1 day. The pre-culture was then transferred to 500 ml flask containing 50 ml of YPD medium containing 100 μg/ml of Zeocin. After 3 days of culture, the culture supernatant was taken for BG activity assay and Western blot analysis. The Western blot result revealed that the culture supernatant of pBGP3-RsBG transformants (RsBG 1, 2, and 3 in Fig. 3-6B; boxed) exhibited immunoblot bands with apparent size of 66.4 kDa, which was not found in the samples from the culture

supernatant of the control strains transformed with the empty vector pBGP3 (i.e. Control 1, 2, and 3). Two N-terminal tags (c-myc and His_6) were fused with RsBG and the calculated molecular weight of the expression product is 62.7 kDa. Therefore this band should be the target protein. As RsBG has 5 potential *N*-glycosylation sites, the increase in molecular weight might be attributable to the post-translational modification. The average BG activity of the culture supernatant of three pBGP3-RsBG transformants was 52.2% higher than that of the controls. From these results it was concluded that RsBG was successfully expressed in *P. pastoris*.

To optimize the production of RsBG, a time-course experiment was carried out to study the expression level of RsBG in different culture duration. The BG assay results indicated that the culture supernatant of the 5th day of the main culture displayed the highest activity (Fig. 3-7A). In general, the BG activity of pBGP3-RsBG *P. pastoris* transformant was slightly higher than that of the control strain harboring the empty vector and a significant difference appeared in the first day. However, the analysis of activity of individual culture flask also showed that the BG activity of pBGP3-RsBG *P. pastoris* transformant was not always higher than that of the control strain.

Subsequently, pBGP3-RsBG transformant was cultured for 5 days in the main culture and 10 ml of the culture supernatant was subjected to filtration through 0.45 μm filter, concentrated by 8-folds by ultrafiltration, and RsBG was purified by Ni-NTA chromatography. Judged by the Western blot analysis result, RsBG was detected in the fractions eluted by 250 mM imidazole. However, no BG activity was detected in the eluate (Fig. 3-7B). As *p*NPG is an artificial substrate and might not be hydrolyzed by RsBG, cellobiose was used for the enzyme assay. The reaction mixture was kept overnight at 37°C, and then analyzed by thin layer chromatography (TLC). The results

showed that there was no sign of hydrolysis of cellobiose by the elution fraction 1 of Ni-NTA purification (Fig. 3-8).

3.2.3 Expression of RsBG by pPICZα-A-RsBG

As the poor activity of RsBG might be attributable to low expression level of pBGP3-RsBG *P. pastoris* transformants, the construction of plasmid pPICZα-A-RsBG and transformation of *P. pastoris* were conducted (Fig. 3-9A). The resultant expression plasmid was linearized by *Bgl* II and transformation of *P. pastoris* was performed. Then the transformants were screened and used for the culture and methanol induction.

The Western blot analysis and CBB-stained SDS-PAGE gel shown in Fig. 3-9B i, ii, and iii revealed that the expression of RsBG was induced from the first day of methanol induction and kept increasing until the 5th day, and the expression level was significantly higher than that of RsBG expressed by pBGP3-RsBG, as the RsBG band on the CBB-stained SDS-PAGE gel could be figured out (red arrows). The BG assay of 5th day culture supernatant showed that the average activity of the culture supernatant of the pPICZα-A-RsBG transformants (RsBG 1, 2, and 3 in Fig. 3-9B iv) was higher than that of the vector-transformed control strains (control 1, 2, and 3), and a significant difference appeared from the 2nd day of methanol induction. The clone number 2 was chosen for further expression of RsBG. A total volume of 84 ml of the culture supernatant of 5-day methanol induction was employed for Ni-NTA purification and RsBG was successfully purified (Fig. 3-10, top and middle panels). However, again the purified RsBG did not show activity (Fig. 3-10, bottom panel). In case that imidazole inhibited the activity of RsBG, dialysis of the fractions containing purified RsBG was performed, but the resultant product did not show activity (data not shown). Then I

tested the activity of purified RsBG towards different natural substrates of BG including cellobiose, cellotriose, cellohexaose, laminaribiose, lactose, sucrose, maltose, laminarin, CMC, Avicel, and salicin, by conducting the reaction at 30°C for 1 day and checking the reaction mixtures by TLC analysis (Fig. 3-11, and the results are summarized in Table 3-5). The results showed that the culture supernatants of both the control strain harboring the empty vector and RsBG had poor activities towards cello-oligosaccharides. Various disaccharides as well as polysaccharides were also tested, but none of these were hydrolyzed by purified RsBG, or the results were variable (for laminaribiose and sucrose). Aryl-glycosides such as *p*NPG, *p*NPFuc, and *p*NPGal were also tested, but no activity was detected. RsBG purified by Ni-NTA chromatography and dialyzed against sodium acetate buffer (pH 5.5) was employed for the enzyme assay, but likewise no activity was detected (data not shown).

3.2.4 Purification of RsBG by anion exchange chromatography

In case that RsBG was inactivated during Ni-NTA purification as was observed in the purification of PaBG1b, anion exchange chromatography was performed for purification of RsBG. A total volume of 50 ml of 5-day methanol induction culture supernatant of pPICZα-A-RsBG *P. pastoris* transformant was used for ammonium sulfate precipitation at a saturation of 70%. The resultant pellet was re-dissolved into 50 mM Tris-HCl buffer (pH 7.0), and then dialyzed to remove ammonium sulfate. The resultant product was then subjected to anion exchange chromatography using HiTrap DEAE FF column (Fig. 3-12). The results showed that RsBG was purified through anion exchange chromatography, but the protein peak was separated from the activity peak, which indicated that the purified RsBG still did not show activity.

3.2.5 Expression of untagged RsBG by pPICZα-RsBGnt

Generally, the protein tags are considered to have no effect on the function of recombinant proteins. However, there are also exceptions, as the case of EG mentioned in Chapter 2 (page 71). In case that the addition of the tags inhibited the activity of RsBG, RsBGnt (nt means 'no tag') was expressed using pPICZα-A-RsBGnt (Fig. 3-13 A i and ii). Four *P. pastoris* transformants were picked up for culture and methanol induction. The methanol induction was prolonged for 3 days and the culture supernatants were harvested for analysis. RsBGnt was successfully expressed, as judged by the distinct bands with apparent size of approximately 64.1 kDa in the CBB-stained gel (Fig. 3-13B), and slightly (43% on average) higher activity of the culture supernatants of pPICZα-RsBGnt transformants compared to those of controls (Fig. 3-13C). However, the RsBGnt-containing supernatant failed to show activity towards cellobiose, as shown in the TLC analysis (Fig. 3-13D). These results show that it is unlikely that the activity of RsBG was affected by the fusion tags.

3.2.6 Mutational analysis of RsBG

To examine the possibilities that (1) *rsbg* is a pseudogene (discussed hereinafter) encoding a non-functional protein, or (2) the catalytic nucleophile of RsBG is actually aspartic acid (D275), but for unknown reason, e.g. occurrence of pyrimidine-to-purine point mutation (GAT/C→GAA) during the cloning of *rsbg*, it was changed to glutamic acid (E275), resulting in the loss of its activity, site-directed mutagenesis was conducted using pBGP3-RsBG as a template to substitute E275 with aspartic acid (Fig. 3-14A). PCR was performed to generate the DNA fragment carrying E275D mutation (Fig.

3-14B). The resultant plasmid pBGP3-RsBG E275D was transformed to *P. pastoris* and 8 transformants were screened and used for expression test. The main culture was continued for 3 days and the culture supernatants were analyzed (Fig. 3-14C). As judged by the results of Western blot analysis and CBB-stained SDS-PAGE gel, it seemed that one of the transformants (12#) produced the mutated protein RsBG E275D, but substantial degradation seemed to have occurred, since the band was much smaller than the expected size (62.7 kDa). Enzyme assay showed no difference between the culture supernatant of the pBGP3-RsBG E275D transformants and the control strains harboring the empty vector (data not shown). In the next experiment, the culture was sampled daily until the 5th day and both the culture supernatant and the cells were kept for analysis. Only in the samples of the 5th day, two faint bands (in the transformants 5# and 13#) showed up on the Western blot membrane with sizes lower than the target protein (Fig. 3-14D). Meanwhile, an intense band showed up on the CBB-stained gel, suggesting that the culture supernatant of 13# might contain secreted but degraded target protein. Enzyme assay of the culture supernatant was carried out but no significant difference was detected (data not shown). To test if the target protein was expressed but was not secreted, the cell samples from the 2nd day to 5th day were harvested and analyzed by Western blot analysis and CBB staining. The result showed that the immunoblot bands on the Western blot membrane were very faint and might be generated by non-specific binding to intracellular proteins (Fig. 3-15). Therefore it was unlikely that the expressed target protein was enclosed inside the cells of *P. pastoris*.

Discussion

Although the heterologous expression in *P. pastoris* and purification of RsBG was

successfully achieved, purified RsBG did not show activity towards regular substrates of BG. As mentioned in 3.2.1, the amino acid sequence of RsBG is rather unique, as its domain arrangement is similar to a GH3 member, but the residue corresponding to the nucleophile is glutamic acid (E275), rather than the highly-conserved aspartic acid. BGs in GH1 employ two glutamic acids for catalysis, and they harbor two residues within a $(\beta/\alpha)_8$ TIM barrel domain. It has been reported that the substitution of glutamic acid to aspartic acid resulted in the inactivation of the enzyme. As to the catalytic acid/base residue (which lies N-terminal to the nucleophile residue), when E193 of NkBG, the catalytic acid/base, was mutated to aspartic acid, NkBG became inactive (Jeng et al., 2011). The same mutation of catalytic acid/base residue was applied to the study of other GH1 BGs such as E191D mutant of maize-derived ZmGlu1 (Verdoucq et al., 2003), E186D mutant of maize-derived Zm-p60.1 (Zouhar et al., 2001), and E189D mutant of sorghum-derived SbDhr-1 (Verdoucq et al., 2004); in all these cases, the resultant products were inactivated. As to the catalytic nucleophile, in the case of Abg, a GH1 BG of *Agrobacterium* (UniProt ID: P12614), when E358 was mutated to aspartic acid, the E358D mutant basically lost its activity (Withers et al., 1992). Therefore, it seems that the catalytic residues for GH1 are conserved and cannot be substituted by aspartic acid, although glutamic acid has only one more methylene group than aspartic acid. In contrast, to my knowledge there has been no study reporting the effect of D/E substitution in GH3 BGs. The only example is JMB19063 where the catalytic nucleophile aspartic acid was mutated to asparagine, which inactivated the enzyme for making a crystal of enzyme-substrate complex.

Pseudogenes are functionless relatives of genes that have lost their gene expression in the cell or their ability to code for a protein (Vanin, 1985). Cho et al. found that the

bacteria in the gut of *R. speratus* have unexpectedly low level of EG activity and concluded that they have adapted to symbiosis in the terminal digestive system and eliminated the EG function in the proteins they synthesized (Cho et al., 2010). To examine the possibility that whether RsBG is a protein without the catalytic function, or an unexpected occurrence of pyrimidine-to-purine point mutation (GAT/C→GAA) during the cloning of *rsbg* led to the substitution of aspartic acid to glutamic acid, E275D mutation was introduced. Unfortunately, it seemed that the expression product was proteolytically degraded (Fig. 3-14 C and D) or failed to be expressed in *P. pastoris* (Fig. 3-15). This problem might be solved by codon optimization or changing the expression system to others such as *E. coli* or the baculovirus-insect expression system.

As mentioned in the introduction of this study, BGs take a variety of roles in physiological process. Considering that some residues of RsBG (G163 and N171) are rather peculiar as compared to the highly conserved residues among GH3 BGs, and the culture supernatants of the *P. pastoris* transformants harboring *rsbg* displayed consistently higher BG activity than those of the control strains, which is similar to the result obtained in *A. oryzae*-employed expression (Sasaguri et al., 2008), RsBG might play a role in enhancing the BG activity in the culture supernatant. Therefore, further study, including the expression of E275D RsBG mutant and X-ray crystallography of enzyme-substrate complex, will be helpful for unveiling the true biological function of RsBG.

Table 3-1. Structurally characterized GH3 family BGs (up to Mar, 2016)[a]

Protein	Organism	Nucleophile	Acid/base	PDB ID	Source
Bacteria					
ExoP	*Pseudoalteromonas* sp. BB1	D320[b]	E520[b]	3F93	Nakatani *et al*., 2012
*Sv*DesR	*Streptomyces venezuelae* ATCC 15439	D273	E578	4I3G	Zmudka *et al*., 2013
*Tn*Bgl3B	*Thermotoga neapolitana* DSM 4359	D242	E458	2X41	Pozzo *et al*., 2010
Eukaryota					
AaBGL1	*A. aculeatus* F-50	D280	E509	4IIB	Suzuki *et al*., 2013
*Af*βG	*A. fumigatus* Af293	Not specified		5FJI	Agirre *et al*., 2016
*An*Bgl1[c]	*A. niger*	D280	E509	N/A	Lima *et al*., 2013
*Ao*βG	*A. oryzae* RIB40	Not specified		5FJJ	Agirre *et al*., 2016
ExoI[d]	*H. vulgare* subsp. *vulgare*	D310[e]	E516[e]	1EX1	Varghese *et al*., 1999
*Km*BglI	*Kluyveromyces marxianus* NBRC1777	D225	E590	3ABZ	Yoshida *et al*., 2010
*Re*Cel3A	*Rasamsonia emersonii* IMI 392299	N/A	N/A	4D0J	Gudmundsson *et al*., (to be published)
*Hj*Cel3A	*T. reesei* QM9414	D267[f]	E472[f]	3ZYZ	Karkehabadi *et al*., 2014
Unclassified					
JBM19063	compost microbe	D261	E488	3U4A	McAndrew *et al*., 2013

[a]: two GH3 BGs, i.e. BglB (larminaribiase) of *T. neapolitana* Z2706-MC24 (GenBank accession number: CAB01407) and tomatinase from *Septoria lycopersici* (AAB08445), were crystallized, but their structures are not available in the PDB bank.
[b]: the catalytic residues are D293 and E493 according to the reference, as the authors did not include the number of the signal peptide.
[c]: the authors did not describe the PBD number, GenBank accession number, or UniProt ID of *An*Bgl1 in their paper, but provided the amino acid sequence of *An*Bgl1 in the supplemental data, which is fully identical with the deduced amino acid sequence of a BG precursor of *A. niger* (GenBank accession number: ABH01182) except for a 74-residue insert in the middle of the precursor.
[d]: ExoI is an exo-β-1,3-1,4-glucanase (EC 3.2.1.-).
[e]: the catalytic residues are D285 and E491 according to the authors for the same reason as in a.
[f]: the catalytic residues are D236 and E441 according to the authors for the same reason as in a.
N/A: not available. For *An*Bgl1, the structure is not available in PBD database. For the catalytic residues of *Re*Cel3A, the relevant information has not been published.

Table 3-2. Multidomain architectures of structurally characterized GH3 BGs[a]

Protein	Domain Numbers	N-terminal	2nd (and 3rd) Domain	C-terminal	Source
Bacteria					
ExoP	3	$(\beta/\alpha)_8$-barrel	$(\alpha/\beta)_6$-sandwich	β-sheets	Nakatani et al., 2012
*Sv*DesR[b]	3	Catalytic core	PA14	Ig-like fold	Zmudka et al., 2013
*Tn*Bgl3B	3	ββ$(\alpha/\beta)_6$-barrel	$(\alpha/\beta)_5$-sandwich	Fn-III like	Pozzo et al., 2010
Eularyota					
AaBGL1 (dimer)	3	ββ$(\alpha/\beta)_6$-barrel	$(\alpha/\beta)_6$-sandwich	FnIII-like	Kentaro et al., 2013
*Af*βG (dimer)	3	ββ$(\alpha/\beta)_6$-barrel	$(\alpha/\beta)_6$-sandwich	FnIII-like	Agirre *et al*., 2016
*An*Bgl1	3	ββ$(\beta/\alpha)_6$-barrel	$(\alpha/\beta)_6$-sandwich	FnIII-like	Lima *et al*., 2013
*Ao*βG (dimer)	3	ββ$(\alpha/\beta)_6$-barrel	$(\alpha/\beta)_6$-sandwich	FnIII-like	Agirre *et al*., 2016
ExoI[c]	2	$(\beta/\alpha)_8$-barrel	-	$(\alpha/\beta)_6$-sandwich	Varghese et al., 1999
*Km*BglI (tetrameric)[d]	4	ββ$(\beta/\alpha)_6$-barrel	PA14 + $(\alpha/\beta)_6$-sandwich	FnIII-like	Yoshida et al., 2010
*Hj*Cel3A (dimer)	3	ββ$(\beta/\alpha)_6$-barrel	$(\alpha/\beta)_5$-sandwich	FnIII-like	Karkehabadi et al., 2014
Unclassified					
JMB19063[e] (dimer)	3	$(\beta/\alpha)_8$-barrel	$(\alpha/\beta)_6$-sandwich	Ig-like region	McAndrew et al., 2013

[a]: the list is based on Table 3-1 with the exclusion of *Re*Cel3A which is to be published.
[b]: the catalytic core of *Sv*DesR is similar to the [$(\beta/\alpha)_8$-barrel + $(\alpha/\beta)_6$-sandwich] structure of the ExoI and interrupted by the insertion of a PA14 domain, which in most cases are thought to be involved in carbohydrate binding. The Ig-like fold domain of *Sv*DesR is similar to the FnIII-like domain.
[c]: ExoI is an exo-β-1,3-1,4-glucanase (EC 3.2.1.-).
[d]: the $(\alpha/\beta)_6$-sandwich domain is separated into two units by the insertion of a PA14 domain.
[e]: the Ig-like region of JMB19063 is similar to the FnIII-like domain.

Table 3-3. Summary of the features of RsBG inferred from the amino acid sequence

Item	Features
GenBank accession number	BJ978994
Length (aa)	573
Length of mature polypeptide (aa)	560
Predicted signal sequence (aa)	13
Average molecular weight of mature region of native RsBG (kDa)	59.8 kDa
Isoelectric point	4.68
Catalytic nucleophile	E275 (putative)
Catalytic acid/base	E471 or E483 (putative)
Putative *N*-glycosylation sites	5 (N47, N78, N171, N449, N478)
Putative *O*-glycosylation site	1 (T426)

Table 3-4. Consensus sequence 'GFVMSDW' found in eukaryotic GH3 BGs (1)

No.	Organism	Uniprot ID	Consensus Sequence[a]	Note
1	*Aspergillus aculeatus* F-50	P48825	GFVMSDW	*Aa*BGL1
2	*Aspergillus avenaceus*	Q58IJ4	GFVMSDW	
3	*Aspergillus awamori* BTMFW032	C7F6Y0	GFVMSDW	BglX fragment
4	*Aspergillus fumigatus* Af293	Q4WR62	GYVMTDW	
5	*A. fumigatus* Af293	Q4WJJ3	GFVMSDW	*Af*βG
6	*Aspergillus kawachii* IFO4308	P87076	GFVMSDW	
7	*Aspergillus nidulans* FGSC A4	Q5BCC6	GIVLTDW	
8	*A. nidulans* FGSC A4	Q5BFG8	GLVMSDW	
9	*A. nidulans* FGSC A4	C8VJG1	GHVLSDW	
10	*A. nidulans* FGSC A4	Q5B9F2	GHVLSDW	
11	*Aspergillus niger* AS3.350	Q30BH9	GFVMSDW	
12	*A. niger* AS3.796	A8WE01	GFVMSDW	
13	*A. niger* B1	Q9P8F4	GFVMSDW	
14	*A. niger* BCRC 31494	C9E8N1	GFVMSDW	
15	*A. niger* CBS 513.88	A2RAL4	GFVMSDW	
16	*A. niger* FTA008	B7U4V8	GFVMSDW	fragment
17	*A. niger* NIBGE	C0M0K9	GFVMSDW	
18	*A. niger* NIBGE 06	B0YIR9	GFVMSDW	
19	*A. niger* NRRL 3135	C7C4Z9	GFVMSDW	
20	*A. niger* zju-2	B6V747	GFVMSDW	
21	*Aspergillus oryzae* RIB40	Q2UUD6	GFVMSDW	*Ao*βG
22	*A. oryzae* RIB40	Q2UN12	GRVVTDW	
23	*A. oryzae* RIB40	Q2UIR4	GFVVSDW	
24	*A. oryzae* RIB40	Q2U325	GYVQSDW	
25	*A. oryzae* RIB40	Q2TYS6	GFVMSDW	
26	*Aspergillus terreus* NIH2624	Q0CEF3	GYVMSDW	
27	*A. terreus* SUK-1	D0VKF5	GFVMSDW	
28	*Aspergillus wentii*	P29090	GFVMSDW	fragment
29	*Bipolaris maydis* C4	O13385	GFVMSDW	
30	*B. maydis* C5	O13391	GFVMSDW	
31	*Botryotinia fuckeliana* SAS56	Q9UVJ6	GYVVSDW	
32	*Chaetomium thermophilum* CT2	A6YRT4	GFVMSDW	
33	*Coccidioides posadasii*	Q8X1K2	-	fragment
34	*C. posadasii* C735	O14424	GFVMTDW	
35	*Debaryomyces hansenii* 4025 DBVPG	O42827	GFVITDW	fragment
36	*Dictyostelium discoideum* AX3	Q23892	GVAVTDW	
37	*Glycine max*	Q9ZU04	-	fragment
38	*Gossypium hirsutum*	Q4F885	GFVISDW	
39	*Hordeum vulgare* subsp. *vulgare*	Q9XEI3	GFVISDW	ExoI, (EC 3.2.1.-)
40	*H. vulgare* subsp. *vulgare*	Q42835	GFVISDW	ExoII, (EC 3.2.1.-)

(Continued)

Table 3-4. Consensus sequence 'GFVMSDW' found in eukaryotic GH3 BGs (2)

41	*Kluyveromyces marxianus* ATCC 12424	P07337	GMLMSDW	
42	*K. marxianus* NBRC1777	D1GCC6	GMLMSDW	*Km*BglI
43	*Kuraishia capsulata* 35M5N	Q12653	GFVVSDW	
44	*Lilium longiflorum*	Q75Z80	GFVISDW	exo-1,3-β-glucanase
45	*L. longiflorum*	Q6UY81	GFVISDW	exo-β-glucanase
46	*Magnaporthe grisea* Y34	Q5EMW3	GFVMSDW	
47	*M. grisea* 70-15	G4N7Z0	GFIMLDW	
48	*M. grisea* 70-15	G4NI45	GYVMSDW	
49	*Neurospora crassa* OR74A	Q7SGM9	GFVMSDW	
50	*N. crassa* OR74A	Q9C2C4	GFVMSDW	
51	*Nicotiana tabacum*	O82151	GFVISDW	
52	*Oxybasis rubra*	Q7XJH8	GYIVSDC	
53	*Parastagonospora nodorum* S-83-2	Q68KX9	GFIMSDW	
54	*P. nodorum* SN26-1	Q68KX7	GFIMSDW	
55	*P. nodorum* Sn37-1	Q68KX7	GFIMSDW	
56	*P. nodorum* SN48-1	Q68KX6	GFIMSDW	
57	*Penicillium brasilianum* IBT 20888	A5A4M8	GFVMSDW	
58	*Penicillium decumbens* 114-2 / JU-A10	B3GK87	GFVMSDW	
59	*P. decumbens* CICC 40361	D3JUX2	GFVMSDW	
60	*Penicillium occitanis*	A7LKA2	GFVMTDW	fragment
61	*Periconia* sp. BCC 2871	A9UIG0	GFTMSDW	
62	*Phaeosphaeria avenaria* f. sp. *avenaria* 1920WRS	Q68HW4	GFIMSDW	
63	*P. avenaria* f. sp. *avenaria* 1921WRS	Q68HW4	GFIMSDW	
64	*P. avenaria* f. sp. *avenaria* 5413	Q68HW4	GFIMSDW	
65	*P. avenaria* f. sp. *avenaria* ATCC 12277	Q68HW4	GFIMSDW	
66	*P. avenaria* f. sp. *avenaria* SAA001NY-85	Q68HW1	GFIMSDW	
67	*P. avenaria* f. sp. *avenaria* SAT002NY-84	Q68HW4	GFIMSDW	
68	*P. avenaria* f. sp. *tritici* ATCC 26370	Q68KY2	GFIMSDW	
69	*P. avenaria* f. sp. *tritici* S-81-W10	Q68KY1	GFIMSDW	
70	*P. avenaria* f. sp. *tritici* SA38-1	Q68HW6	GFIMSDW	
71	*P. avenaria* f. sp. *tritici* SA39-2	Q68HW5	GFIMSDW	
72	*P. avenaria* f. sp. *tritici* SAT24-1	Q68HW5	GFIMSDW	
73	*Phaeosphaeria avenaria* WAC1293	Q9P879	GFIMSDW	
74	*Phaeosphaeria* sp. S-93-48	Q68KY0	GFIMSDW	
75	*Phanerochaete chrysosporium* K-3	Q8TGC6	GYVMSDW	
76	*P. chrysosporium* OGC101	O74203	GYVMSDW	
77	*Physarum polycephalum*	Q3V6T3	GYIMSDW	
78	*Pichia etchellsii*	B5B423	GMIMSDW	
79	*Piromyces* sp. E2	Q875K3	GFVMSDW	
80	*Pleospora* sp. P56	Q69FJ8	-	fragment
81	*Rasamsonia emersonii* IMI 392299	Q8TGI8	GFVMTDW	*Re*Cel3A

(Continued)

Table 3-4. Consensus sequence 'GFVMSDW' found in eukaryotic GH3 BGs (3)

82	*Rhizomucor miehei* NRRL 5282	B0JE65	G**LI**MS**D**W	
83	*Saccharomycopsis fibuligera*	P22507	GFV**V**S**D**W	
84	*S. fibuligera* ATCC 36309	P22506	GFV**V**S**D**W	
85	*S. fibuligera* BCRC 20455	B5TWK3	GFV**V**S**D**W	
86	*Schizophyllum commune*	P29091	-	fragment
87	*Schizosaccharomyces pombe* 972h-	Q9P6J6	G**TII**S**D**W	
88	*Secale cereale*	Q6PQF3	GFV**I**S**D**W	
89	*Septoria lycopersici*	Q99324	G**Y**V**V**S**D**W	tomatinase
90	*Stemphylium xanthosomatis* EGS17-137	Q69FF5	-	fragment
91	*Talaromyces purpureogenus* KJS506 (KACC 93053P)	C9E9M9	GFVM**T**D**D**W	
92	*Thermoascus aurantiacus* IFO 9748	Q4U4W7	GFVMS**D**W	
93	*T. aurantiacus* IFO 9748	Q0ZUL0	GFVM**T**D**D**W	
94	*T. aurantiacus* var. *levisporus*	A9QUC3	GFVM**T**D**D**W	
95	*T. aurantiacus* var. *levisporus*	A9QUC4	GFVM**T**D**D**W	
96	*T. aurantiacus* var. *levisporus*	A9UFC6	GFVMS**D**W	
97	*Trichoderma reesei* QM6A	Q7Z9M5	GFVMS**D**W	Cel3B
98	*T. reesei* QM6A	Q7Z9M4	GLIMS**D**W	Cel3C
99	*T. reesei* QM6A	Q7Z9M1	PLIVS**D**W	Cel3D, fragment
100	*T. reesei* QM6A	Q7Z9M0	GFVML**D**W	Cel3E
101	*T. reesei* QM9414	Q12715	GYVMT**D**W	*Hj*Cel3A
102	*Trichoderma* sp. SSL	B5TYI5	G**Y**VM**T**D**D**W	
103	*T. viride* AS 3.3711	C6GGC9	G**Y**VM**T**D**D**W	
104	*T. viride* AS 3.3711	Q6UJY0	G**Y**VM**T**D**D**W	
105	*Tropaeolum majus*	O82074	GFV**I**S**D**W	
106	*Uromyces viciae-fabae*	Q70KQ7	G**VMVT**D**D**W	
107	*Volvariella volvacea* V14	Q9C3Z9	G**LI**MS**D**W	
108	*Wickerhamomyces anomalus* VAR. ACETAETHERIUS	P06835	GFVM**T**D**D**W	
109	*Zea mays*	Q9LLB8	GFV**IT**D**D**W	(EC 3.2.1.-)

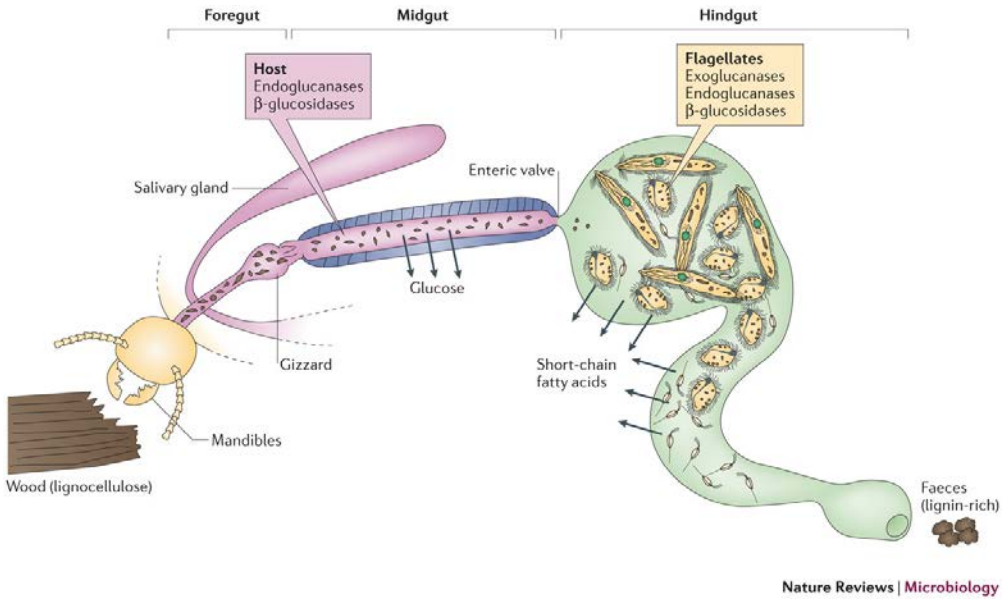
[a]: The highly conserved aspartic acid (D) is highlighted in red, and the residues mismatched from the consensus sequence 'GFVMSDW' are indicated in blue.

Table 3-5. Substrate specificity of the culture supernatant of pPICZα-A-RsBG and purified RsBG

Substrates	V	Rs	E	D
Cello-oligosaccharides				
Cellobiose	✖	✖	✖	-
Cellotriose	✖	✖	✖	✖
Cellohexaose (insoluble)	✔	✔	✖	✖
Disaccharides				
Laminaribiose		Unclear		
Lactose	✖	✖	✖	-
Sucrose		Unclear		
Maltose	✖	✖	✖	-
Polysaccharides				
Laminarin	✖	✖	✖	✖
CMC	✖	✖	✖	✖
Avicel (insoluble)	✖	✖	✖	✖
Salicin	✖	✖	✖	✖
Aryl-glycosides				
p-Nitrophenyl-β-D-glucopyranoside	✔	✔	✖	✖
p-Nitrophenyl-β-D-fucopyranoside	✖	✖	✖	✖
p-Nitrophenyl-β-D-galactopyranoside	✖	✖	✖	✖

V, Rs, E, and D are the same as those in Fig. 3-11. '✔' and '✖', capability and incapability of hydrolyzing the substrate, respectively. '-', not tested.

A



Source: Brune (2014)

B

Source: http://www.riken.jp/en/research/rikenresearch/highlights/5414/

Fig. 3-1. The scheme of dual-cellulolytic system of lower termites (A) and photos of *R. speratus* and intestinal symbionts thereof (B)

(A), Lignocellulose digestion in the lower termites involves activities of both the host and its symbiotic protists. (B), from the left to the right, *R. speratus*, the gut of the termite, and the protist community.

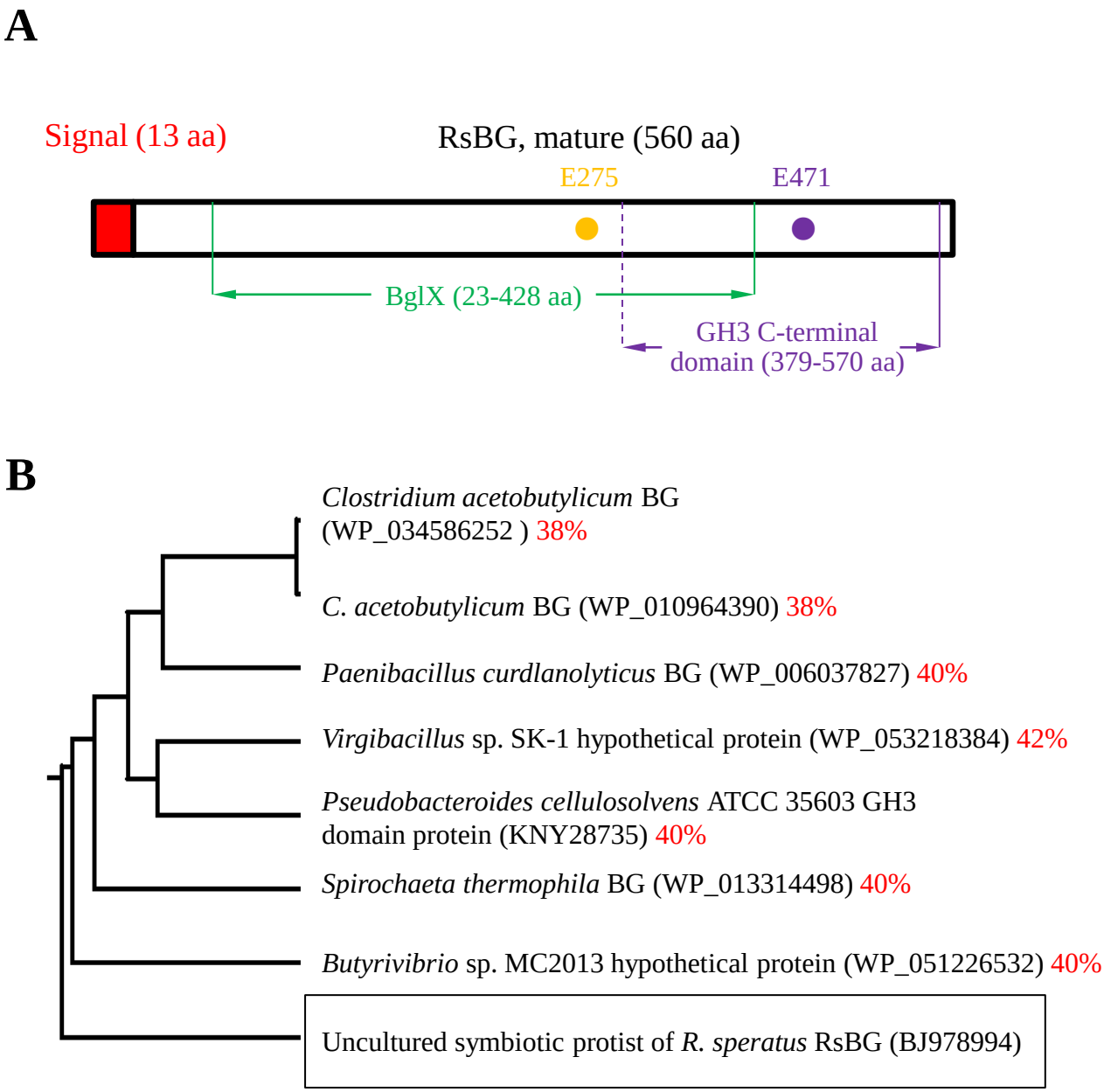


Fig. 3-2. Domain architecture of RsBG (A) and phylogenetic tree inferred from the amino acid sequences of selected BGs with the highest homology (E-value from 2e-126 to 1e -121) towards RsBG by Blast search

In (A), the domain architecture of RsBG predicted by BLAST P server. BglX domain refers to the periplasmic BG and related glycosidases (carbohydrate transport and metabolism), whereas GH 3 C-terminal domain is involved in the catalysis and might be involved in binding with β-glucan. Two putative catalytic residues are assigned by amino acid sequence alignment with structurally characterized GH3 BGs (cf. Figs. 3-4 and 3-5). In (B), the origins and the description of the homologues are indicated. The GenBank accession numbers are given in parentheses. The phylogenetic tree mapping was drawn by using ClustalW server (http://www.genome.jp/tools/clustalw/) based on the UPGMA method. The identity of each BG towards RsBG is shown by the numbers in red.

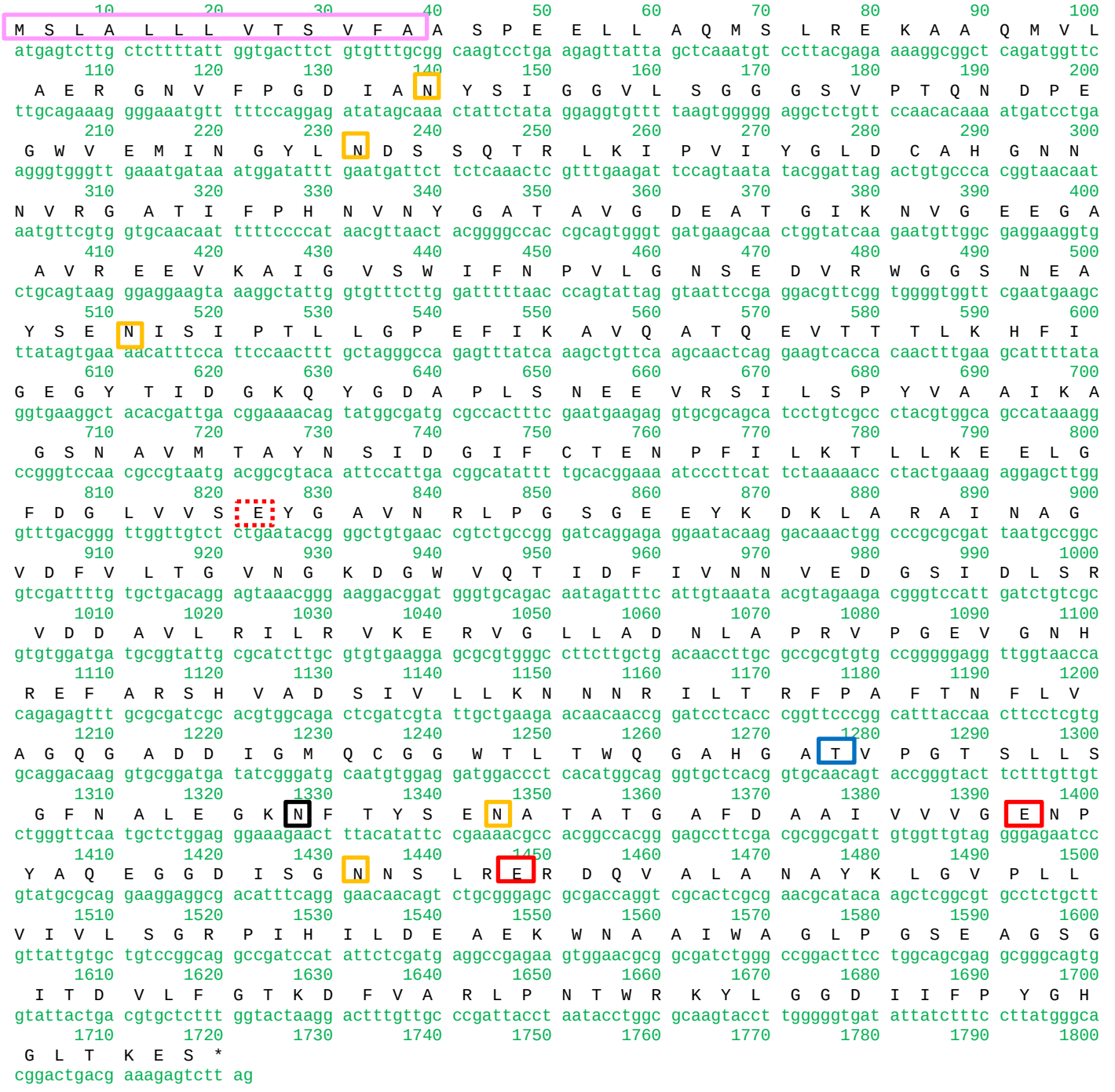
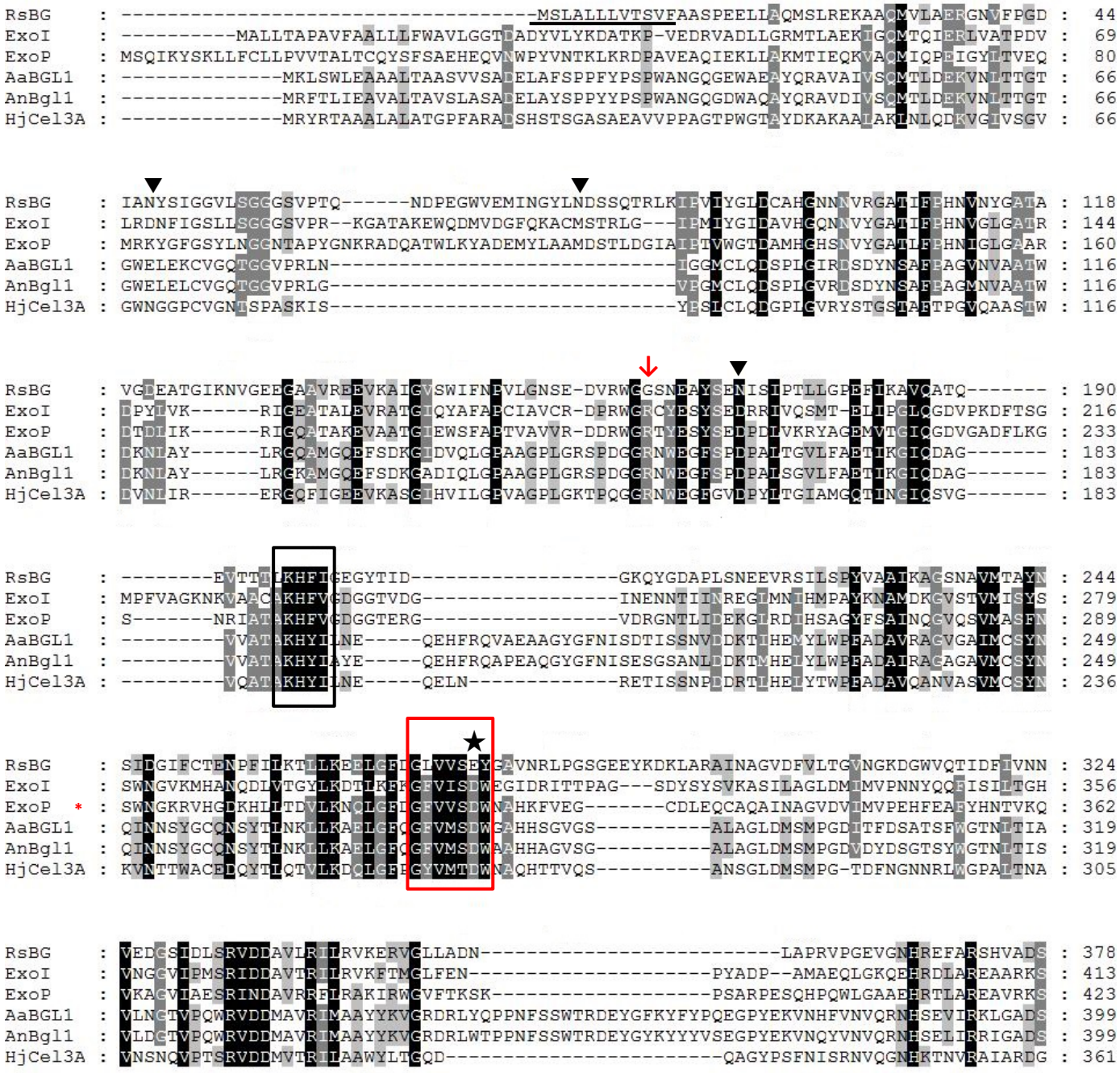


Fig. 3-3. Nucleotide and deduced amino acid sequences of RsBG (573 amino acids)
Nucleotide sequence and numbering thereof are shown in green at the top and bottom rows of each line, respectively. The amino acid sequence is shown in the middle row of each line in black and capital letters. The putative signal sequence is boxed in pink. Putative catalytic nucleophile (E275) and acid/base (E471 or E483, see Fig 3-4) residues are boxed in red, five potential *N*-glycosylated sites (N47, N78, N171, N449, and N478) and one *O*-glycosylated site (T426) are boxed in yellow and blue, respectively. A putative *N*-glycosylation site (N443; boxed in black) was identified, but omitted from the potential *N*-glycosylation site in the prediction by NetNGlyc 1.0 server algorithm.



(Continued to the next page)

Fig. 3-4. Multiple alignment of RsBG and selected structurally-characterized GH3 BGs and ExoI (1)

The putative signal sequence of RsBG is underlined. Putative catalytic nucleophile of RsBG (E275) within the consensus sequence of 'GFVMSDW' of GH3 BGs (boxed in red) is marked by the solid star (★). Potential *N*-glycosylation sites of RsBG (N47, N78, and N171) are indicated by solid arrowheads (▼). A red arrow (↓)indicates the conserved arginine (R) which was reported to form hydrogen bonds with the substrate or be involved in the substrate recognition among BGs aligned, whereas its counterpart in RsBG is glycine (G163). The putative conserved carbohydrate-recognition motif 'KHFV' of GH3 BG is boxed in black.

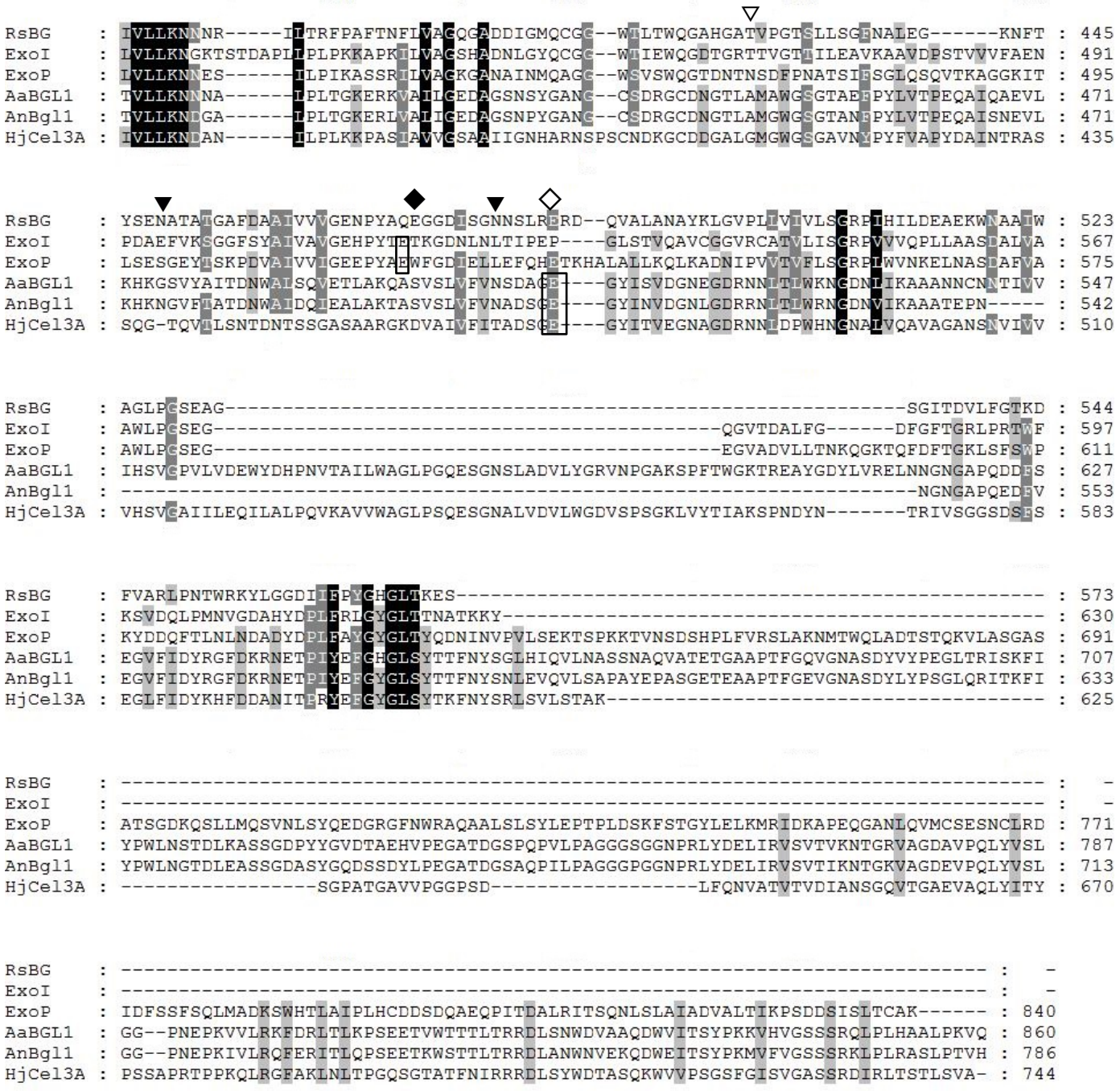


Fig. 3-4. Multiple alignment of RsBG and selected structurally-characterized GH3 BGs and ExoI (2)

Putative catalytic acid/base residue of RsBG (E471 or E483) tentatively assigned by the alignment is marked by the diamond (◆ and ◇). Potential *N*-glycosylation sites (N449 and N478) and *O*-glycosylation site (T426) of RsBG are indicated by solid (▼) and open (▽) arrowheads, respectively. Catalytic acid/base residues of aligned BGs, i.e. Exo I (E516), Exo P (E520), *Aa*BGL1 (E509), *An*Bgl1 (E509, see supplemental file 2 at page 168) and *Hj*Cel3A (E472), are boxed.

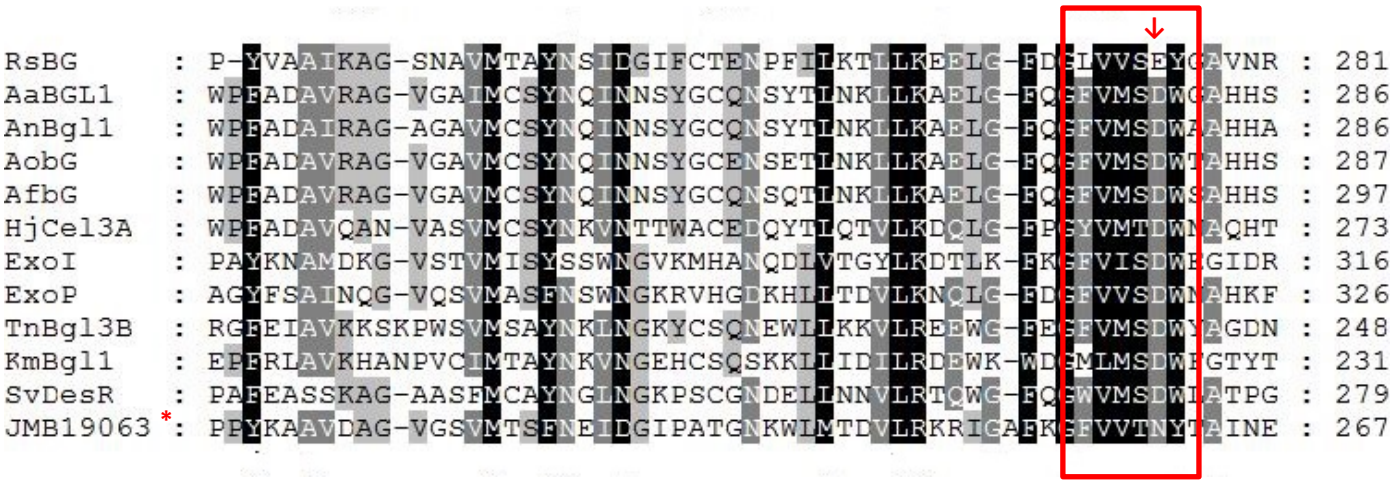


Fig. 3-5. Alignment of partial sequences of RsBG and all structurally-characterized GH3 BGs and ExoI

The highly conserved sequence 'GFVMSDW' is highlighted by the box in red, and the putative catalytic nucleophile residue of RsBG (E275) is indicated by the red arrow. The sequence of *An*Bgl1 was cited from the supplemental 1 and 2 of the paper by Lima et al. as the authors did not describe the PBD number, GenBank accession number, or UniProt ID of *An*Bgl1 (Lima *et al*., 2013). The sequence of *An*Bgl1 is fully identical to the deduced amino acid sequence of a BG precursor of *A. niger* (GenBank accession number ABH01182) with the exception of a 74 residue-length insert in the middle of the precursor. *: JMB19063 is a variant of which putative catalytic nucleophile D261 was substituted by asparagine and catalytically inactivated.

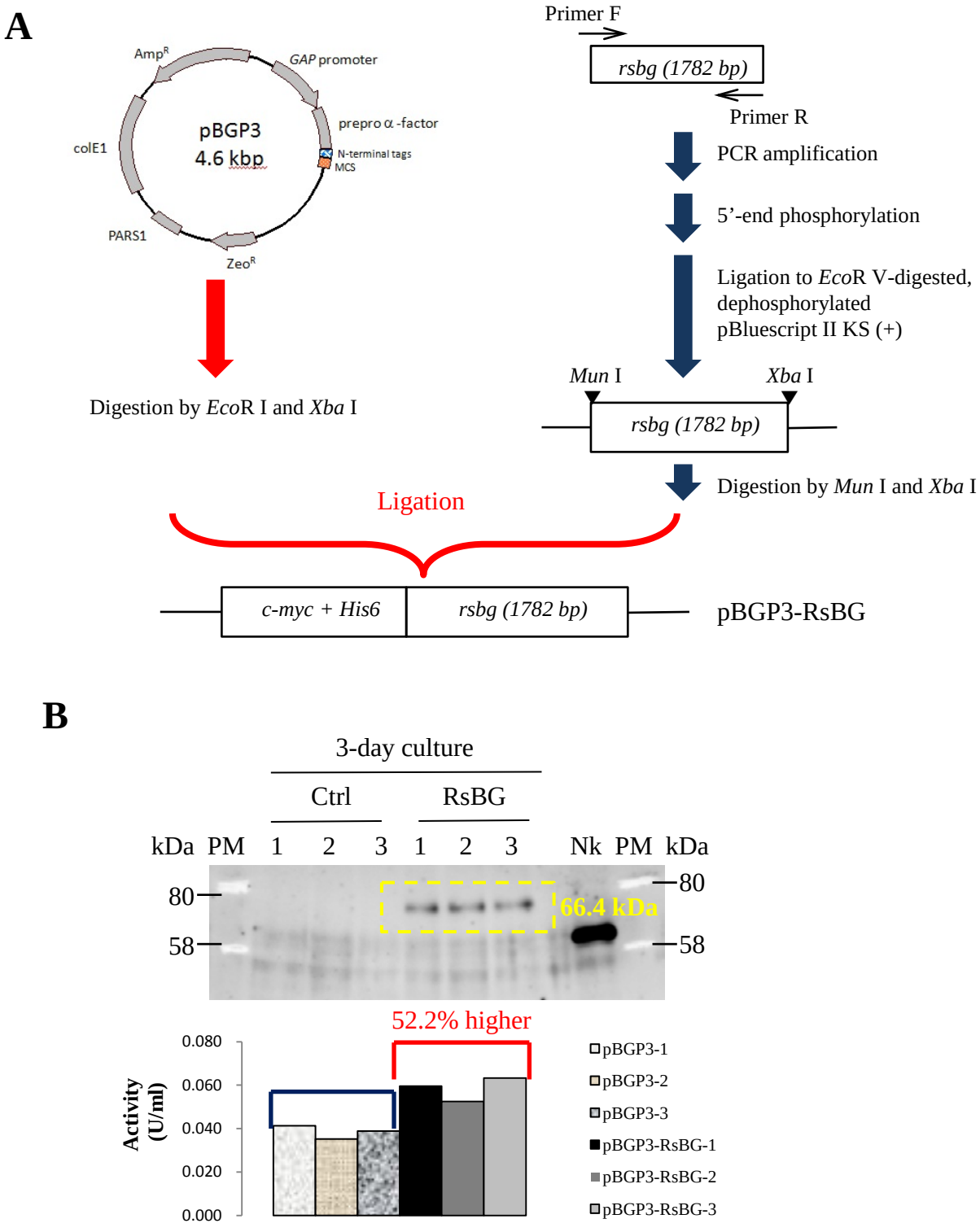


Fig. 3-6. Construction of pBGP3-RsBG (A) and expression of RsBG in *P. pastoris* (B)
In (B), Western blot analysis (top) and BG activity assay result (bottom) of the supernatant of the 3-day culture are shown. PM: Protein marker; Ctrl 1, 2, 3: control strains transformed with the empty vector; RsBG 1, 2, 3: pBGP3-RsBG transformants; Nk: NkBG positive control. The bands with the size of 66.4 kDa were detected in the Western blot. In the activity assay, RsBG transformants displayed slightly higher activity compared to the controls.

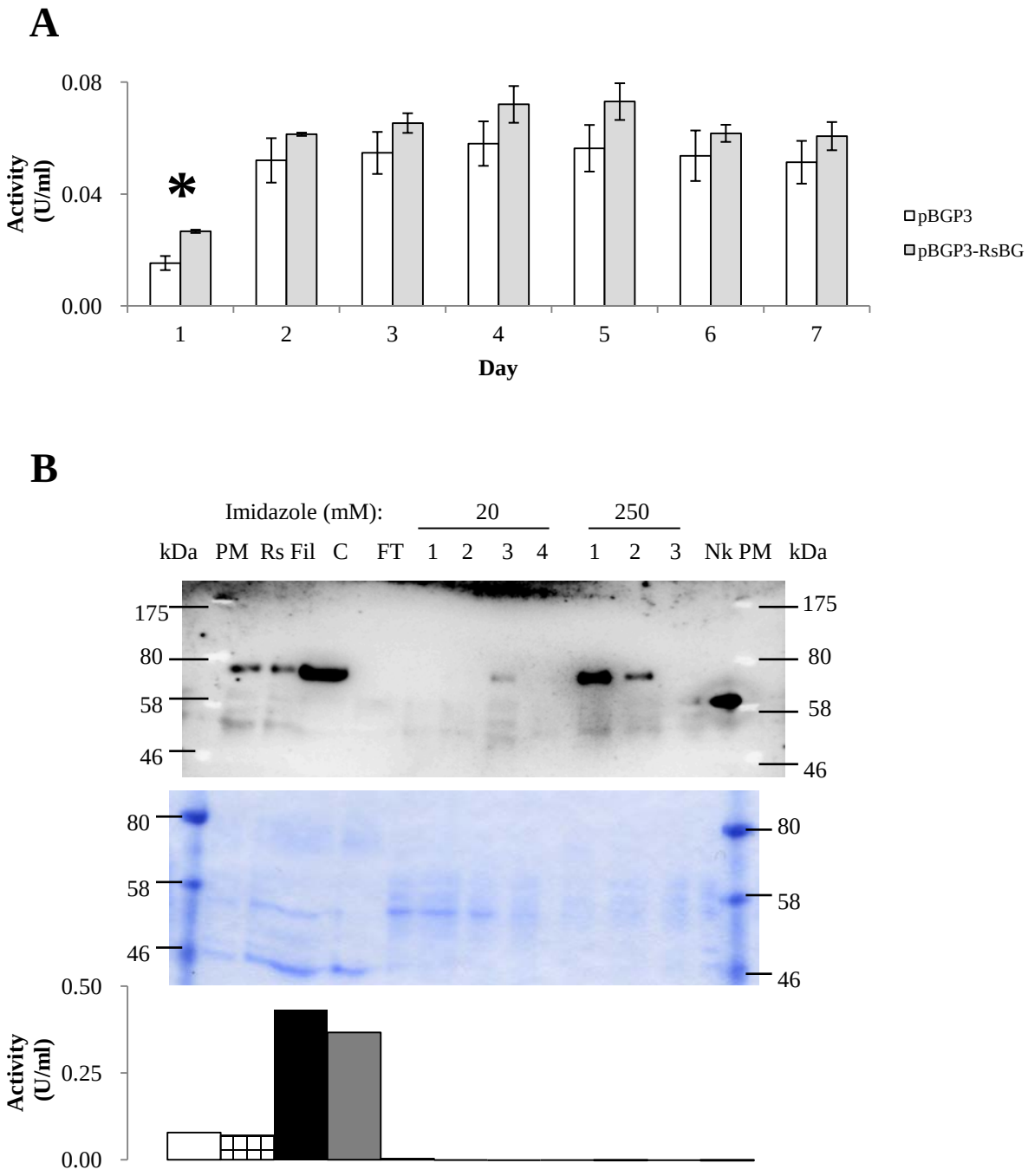


Fig. 3-7. Time-course analysis of BG activity of the culture supernatant of pBGP3 and pBGP3-RsBG *P. pastoris* transformant (A) and purification of RsBG through Ni-NTA column chromatography (B)

(A), the result of BG assay. Data are means ± S.D. of three independent experiments. *, $p<0.05$ vs control by the Student's *t* test. In (B), from the top to the bottom, Western blot, CBB-stained SDS-PAGE gel, and BG assay results are shown. PM: protein marker; Rs: pBGP3-RsBG *P. pastoris* transformant; Fil: filtrate of culture supernatant by 0.45 μm filter; C: 8-fold-concentrated sample of the culture supernatant; Nk: NkBG positive control.

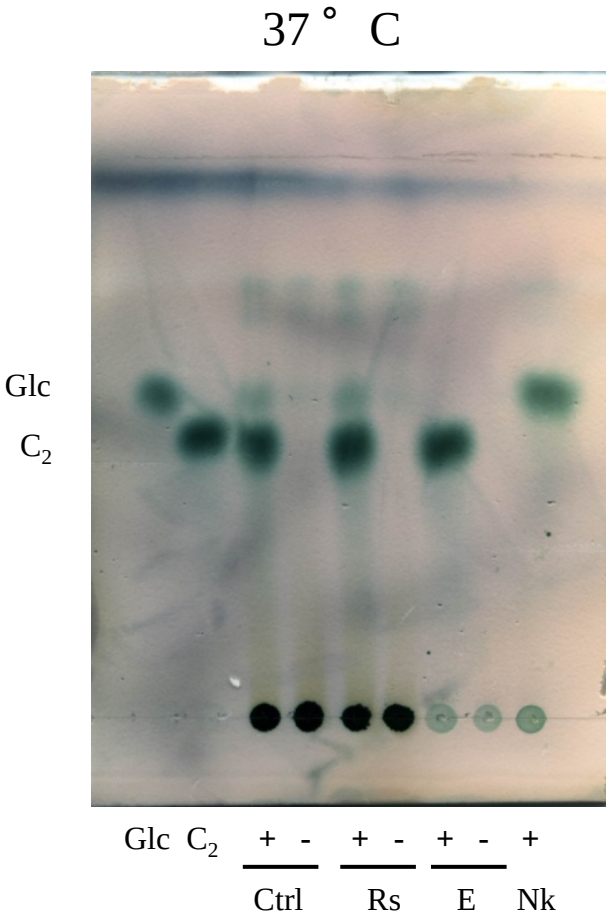


Fig. 3-8. TLC analysis of the products of enzyme reaction conducted at 37° C using cellobiose as a substrate

G: Glucose; C_2: Cellobiose. Ctrl: culture supernatant of the control strain harboring the empty vector pBGP3. Rs: culture supernatant of pBGP3-RsBG; E: purified RsBG through Ni-NTA purification system (fractions eluted at 250 mM imidazole; Fig. 3-7B); Nk: NkBG positive control; +, the reaction was done in the presence of cellobiose; -, reaction was done in the absence of cellobiose.

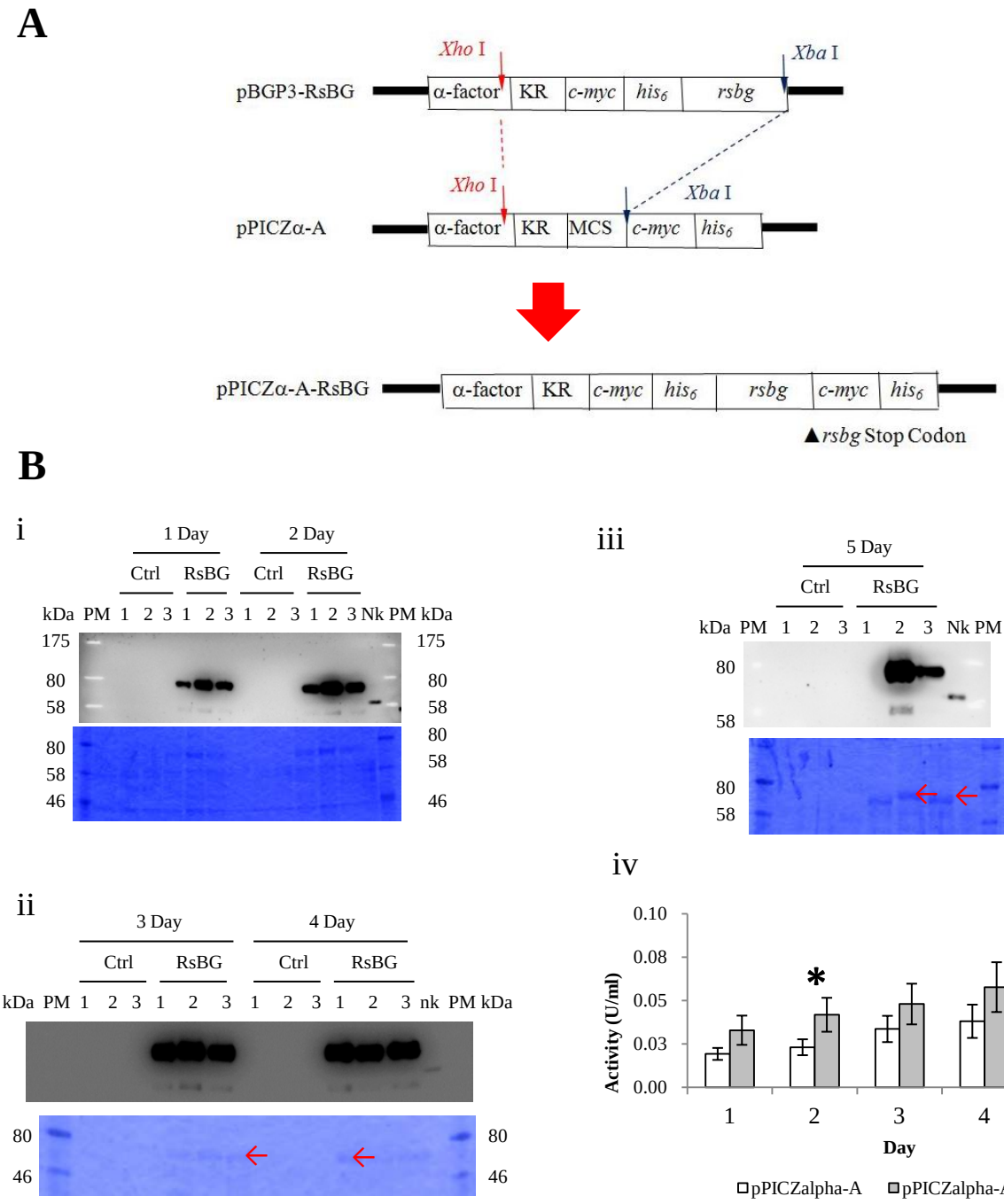
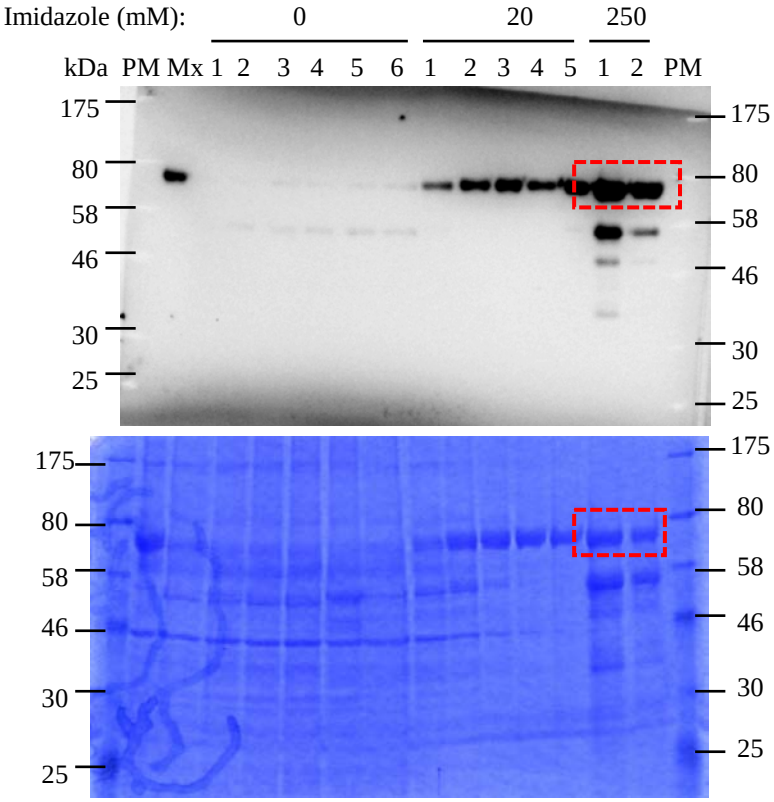


Fig. 3-9. Construction of pPICZα-A-RsBG (A) and expression of RsBG (B)
In (B), from i to ii: Western blot results of culture supernatants from 1- to 5-day methanol induction; Ctrl 1, 2, and 3: control strain transformed with pPICZα-A vector; RsBG 1, 2, and 3: pPICZα-A-RsBG transformants. On the 5th day of methanol induction, it seemed that c-myc tag of pPICZα-A-RsBG transformant no. 1 was lost, whereas RsBG in no. 2 was partially degraded as revealed by the smaller band on the Western blot membrane. Red arrows indicate the bands of RsBG on SDS-PAGE gels; iv: average BG activities of culture supernatants of three control strains (open bars) and three pPICZα-A-RsBG transformants. Data are means ± S.D. of three independent experiments. *, $p<0.05$ vs control by the Student's *t* test.



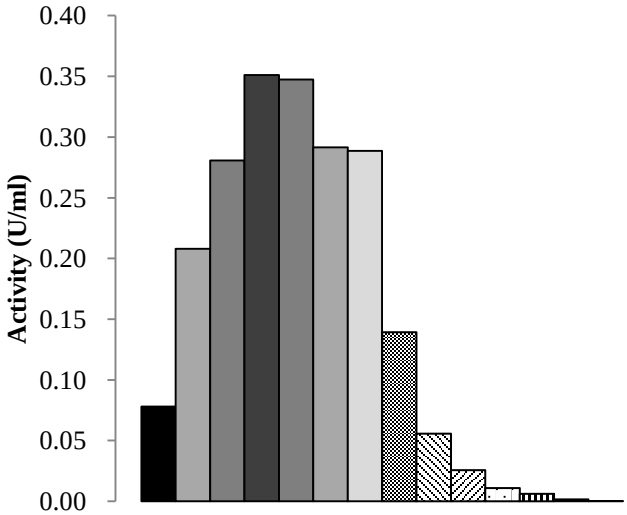


Fig. 3-10. Analysis of Ni^{2+}-NTA affinity chromatography purification fractions by Western blot analysis (A), CBB staining (B), and BG activity assay (C)

About 84 ml of culture supernatant of the pPICZα-A-RsBG transformant were concentrated into 8 ml by ultrafiltration and mixed with the same volume of 50 mM sodium phosphate buffer, 300 mM NaCl, pH 8.0, then loaded onto the Ni^{2+}-NTA column. Mx: mixture of the concentrated culture supernatant and buffer. The boxes in red indicate partially-purified myc-His_6-RsBG in the elution fractions. Elution fractions containing RsBG were named as solution E hereafter. The volume of each fraction is 1 ml, except for Mx (8 ml).

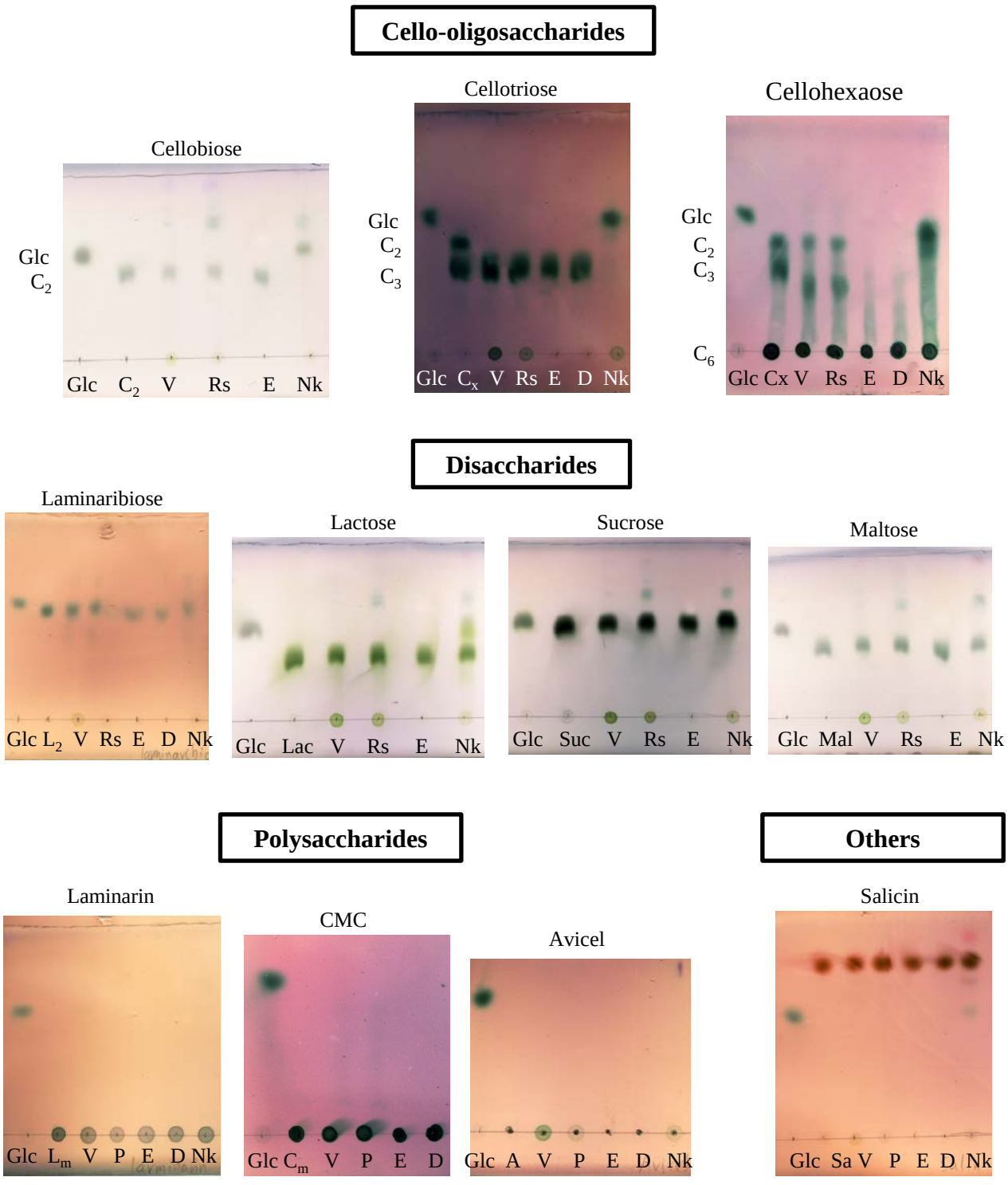


Fig. 3-11. TLC analysis of reaction products using various saccharides as substrates

Glc: glucose; V, Rs, and Nk: culture supernatants of the control strain, pPICZα-A-RsBG, and the positive control NkBG; E and D indicate the elution fractions of Ni-NTA purification and dialyzed products, respectively; Cx: cello-oligosaccharide mixture (C2+C3 or C2+C3+C6); C_2, C_3, C_6: cellobiose, cellotriose, and cellohexaose, respectively. L_2: laminaribiose; Lac: lactose; Suc: sucrose; Mal: maltose; L_m: laminarin; C_m: CMC; A: Avicel; Sa: salicin. Enzymatic reactions were performed at 30° C for 1 day, with the 1:100 ratio of enzyme versus substrate. Substrate concentrations were: 10 mM disaccharides, 1% polysaccharides, and 1% salicin in 50 mM NaOAc buffer, pH 5.5.

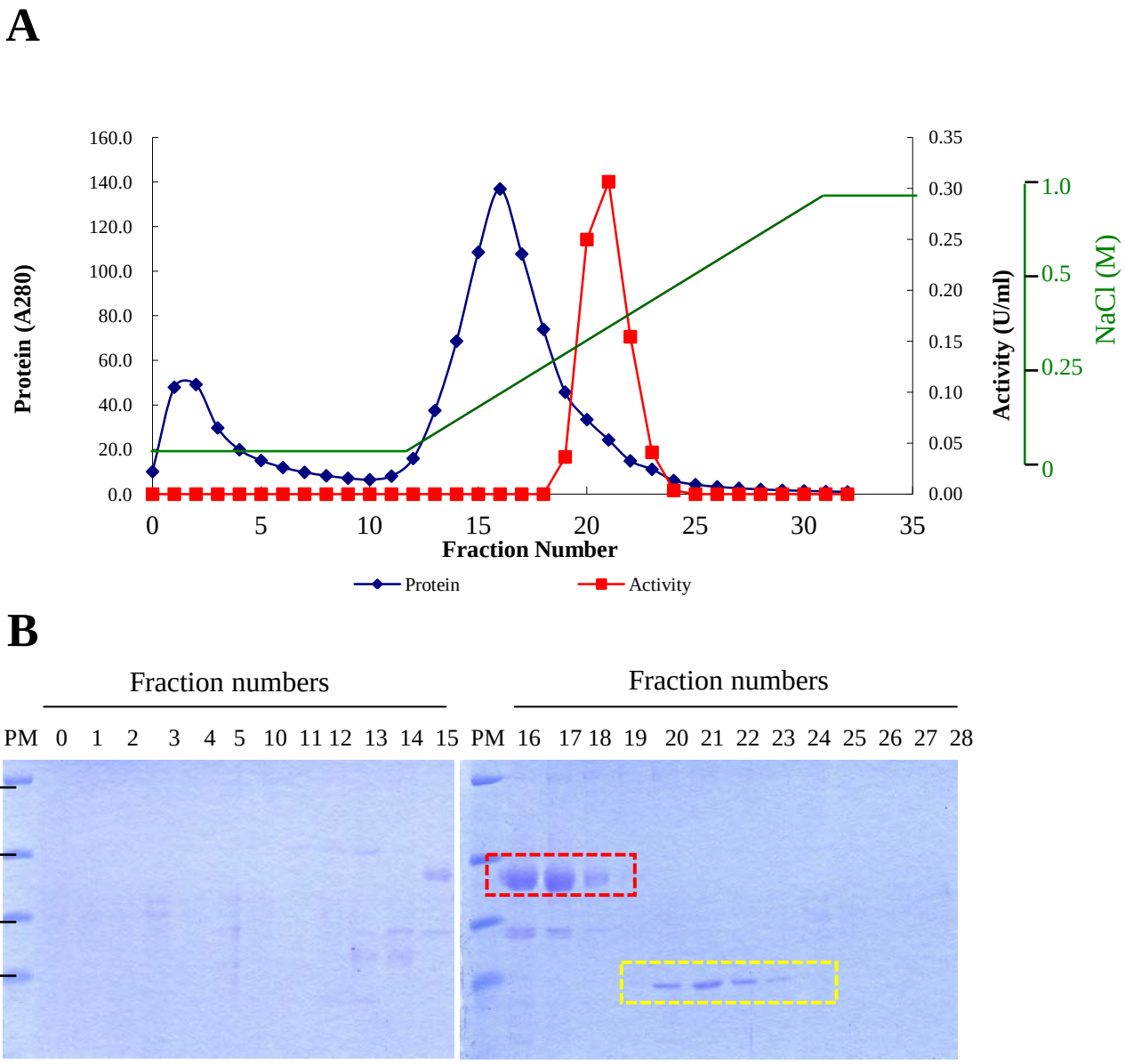


Fig. 3-12. Purification of RsBG by anion exchange chromatography

(A), two ml of ammonium sulfate precipitation products from 50 ml culture supernatant of the pPICZα-A-RsBG transformant was loaded onto the HiTrap DEAE FF column equilibrated by the starting buffer (50 mM Tris-HCl, pH 7.0). Proteins were eluted by 0-1 M linear gradient of NaCl at a flow rate of 1 ml/min. The protein and BG activity profiles are shown. The protein peaks did not match the activity peak. (B), fractions in (A) were checked by CBB staining. RsBGnt is boxed in red. The result demonstrated that the BG activity was mainly derived from the contaminants in the fractions 19 to 22. The proteins with small size of ~46 kDa which displayed activities (boxed by yellow dash lines) might be EXG1 of *P. pastoris*. EXG1 is an extracellular exo-β-(1,3)-glucanase from *P. pastoris* (Huang *et al*., 2011) and with a size of approximately 47 kDa and capable of hydrolyzing *p*NPG and cellobiose (Xu *et al*., 2006).

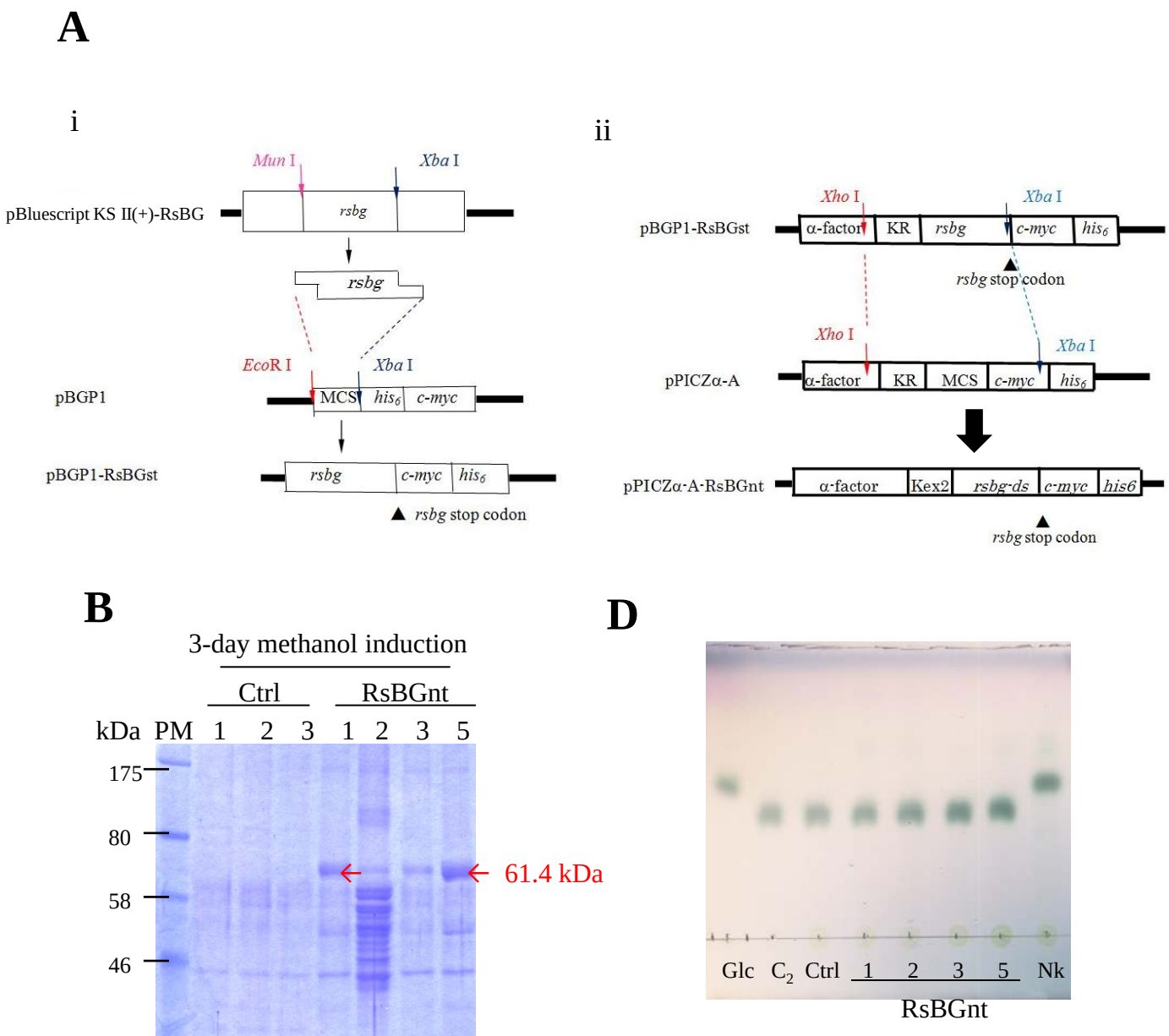


Fig. 3-13. Construction of pPICZα-A-RsBGnt (A) and analysis of RsBGnt (no tags) by CBB staining (B), BG assay (C), and TLC analysis using cellobiose as a substrate (D)

(A), construction of plasmids pBGP1-RsBGst and pPICZα-A-RsBGnt is shown. (B), PM: protein marker; Ctrl 1, 2, 3, and RsBGnt 1, 2, 3, 5: 3-day methanol induction culture supernatants of the control strain and the pPICZα-A-RsBGnt transformant strains no. 1, 2, 3, and 5. Red arrows indicate RsBGnt with the apparent size of 64.1 kDa; In (D), enzymatic reactions were carried out at the same conditions as those in Fig. 3-11.

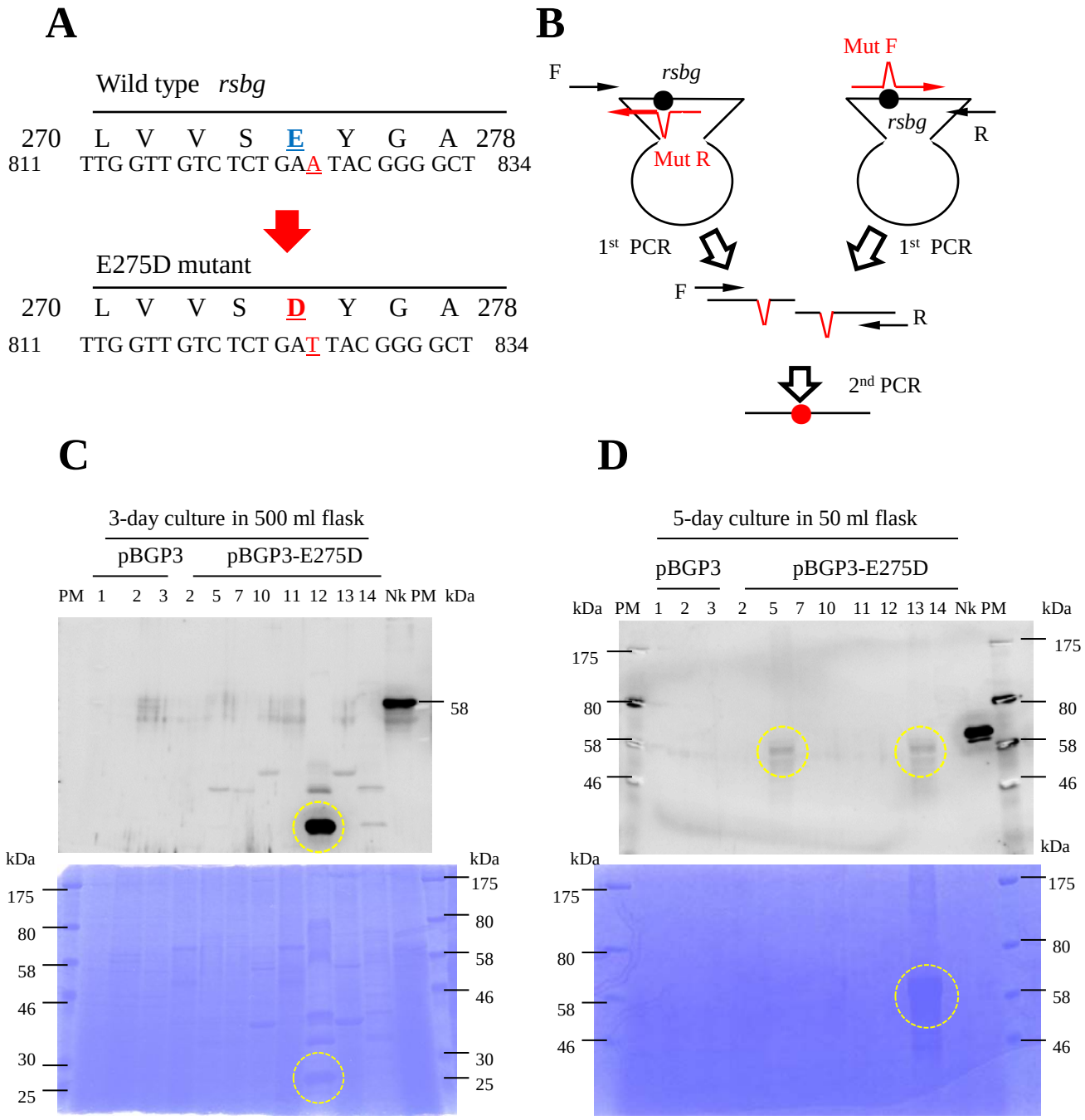


Fig. 3-14. Expression of E275D mutant of RsBG

(A and B) Site-directed mutagenesis was performed to generate E275D mutant of RsBG. Mutation (●) was introduced by two rounds of PCR. The construction of the plasmid pBGP3-RsBG E275D and transformation of *P. pastoris* were subsequently preformed. (C and D) Cultures of the resultant transformants were carried out in 500 ml flasks (C) and 50 ml flasks (D) containing 50 ml and 20 ml media, respectively, and the culture supernatants of 8 pBGP3-RsBG E275D transformants together with 3 control strains harboring the empty vector (pBGP3) were subjected to SDS-PAGE followed by Western blot and CBB staining analyses. Circles in yellow indicate the putative secreted but degraded RsBG E275D. PM: protein marker.

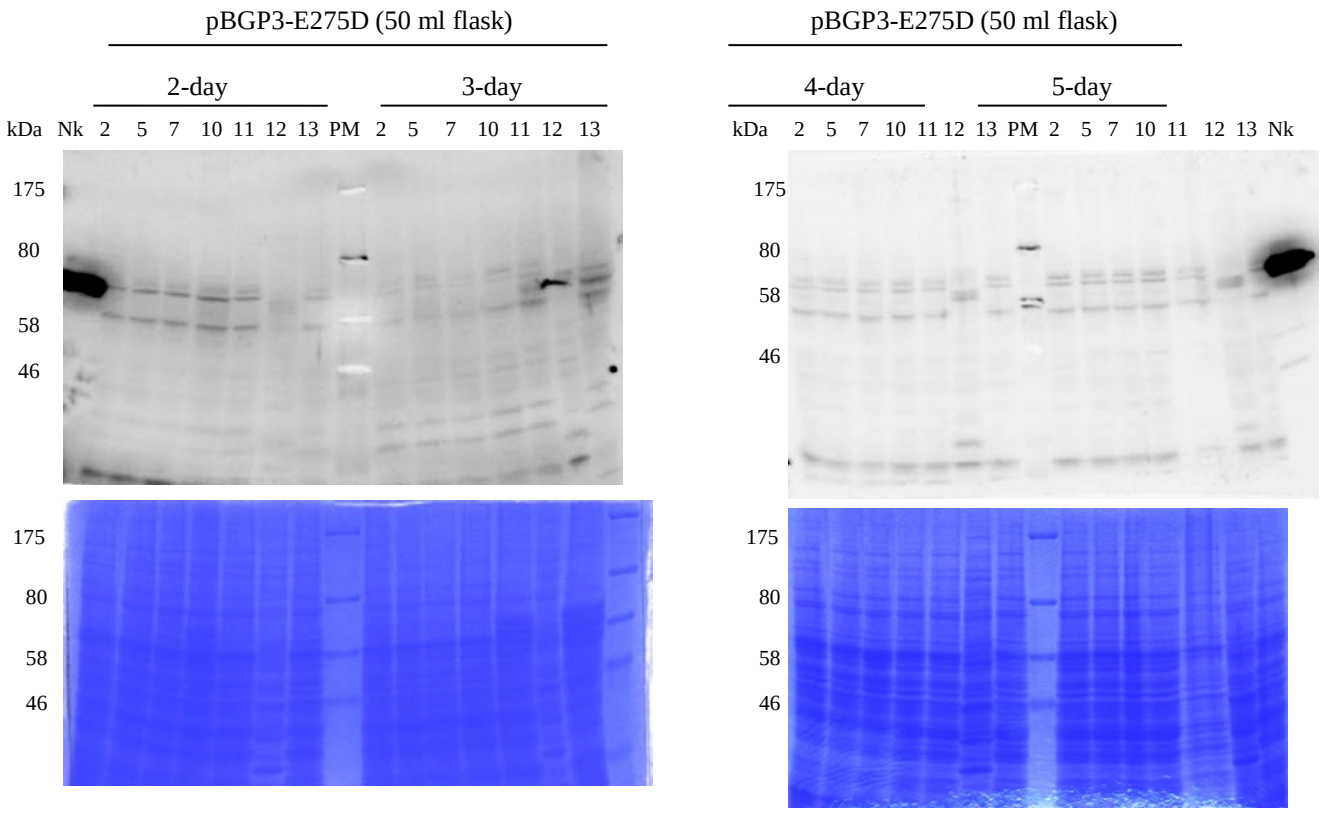


Fig. 3-15. Western blot and CBB staining analyses of the cellular fractions of pBGP3-RsBG E275D transformant

Culture of transformants were carried out in 50 ml flasks containing 20 ml of medium for 5 days. The cellular fractions from 2nd day to 5th day were collected and subjected to SDS-PAGE, followed by Western blot and CBB staining analyses. Nk: NkBG positive control.

Final conclusion

In this study, PaBG1b, a potentially valuable BG of the wood-feeding cockroach *Panesthia angustipennis spadica*, which was isolated and purified by Arakawa et al. and displayed extremely high activity towards cellobiose, was successfully expressed in *P. pastoris*. Economical bioethanol production from cellulose demands cost-competitive enzyme preparations. Preliminary characterization of PaBG1b revealed high activity (V_{max} of 436.7±6.3U/mg) and catalytic efficiency (109.8 $mM^{-1}\cdot s^{-1}$) towards cellobiose, which makes it one of the best BGs in terms of hydrolyzing cellobiose. PaBG1b displayed modest glucose tolerance and high resistance to cellobiose, which endows it with advantages in the industrial purposes. However, the activity of the recombinant PaBG1b was lower than that of the native enzyme. Although the reason for this is unclear, if the presence of smaller-sized product due to the partial Kex2 processing affected the activity, then by changing the expression host it might be expected to obtain the full-length, fully-active product. On the other hand, in terms of the optimum temperature, that for the native enzyme (60°C) was higher by 15°C than that of the recombinant PaBG1b (45°C). This difference might be attributable to the difference in the reaction time and the substrate used in the BG assays, as mentioned in 2.2.2.1 of Chapter 2 (in page 57). As for the thermostability, the performances of the recombinant PaBG1b were approximately the same level as those of the native enzyme, and protein engineering technology such as site-directed mutagenesis might be employed for improving the thermostability of PaBG1b. The glucose resistance of PaBg1b might also benefit from targeted mutagenesis, as discussed in Chapter 2 (page 69). Furthermore, as PaBG1b is a powerful BG, it could be in turn a good template to understand how the subtle differences in the amino acid sequences lead to significant change in the catalytic efficiency, and how we can improve the performance of BGs. Therefore, further studies

of PaBG1b, including the structural analysis, would be of great value.

In this study, another putative BG from the symbiotic protist, RsBG, was expressed in *P. pastoris*. Unfortunately, attempts to show BG activity in the purified RsBG preparation failed, and the real role of RsBG as well as the reason of the weak but consistently higher activity of the culture supernatant of the transformants harboring RsBG remain obscure. The incapability of RsBG to hydrolyze substrates might imply that a small spatial change in the active site residue severely affects the catalytic activity.

In conclusion, it is highly hoped and expected that the studies of PaBG1b and RsBG contribute to deeper understanding of BGs in general and benefit the commercial bioethanol production from cellulose.

Materials and Methods

1. Strains

E. coli strain DH5α (*supE4* Δ*lacU169*(Φ*80lacZ* Δ*M15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was used for DNA manipulation. *P. pastoris* strain KM71H (*his4 aox1*::*ARG4*, from Invitrogen, Carlsbad, USA) was used as a host for heterologous expression of recombinant proteins. *P. pastoris* transformant strains were grown on YPD (1% Bacto yeast extract, 2% Bacto peptone, 2% dextrose), YPDS (YPD medium containing 1 M sorbitol), and YPG (1% Bacto yeast extract, 2% Bacto peptone, and 1% glycerol) medium containing 100 μg/ml zeocin (Invitrogen, catalog No. R25001). YP (1% Bacto yeast extract, 2% Bacto peptone) medium was employed for methanol indiction of *P. pastoris* strains transformation with pPICZα-A constructs. The concentration of zeocin in YPD, YPDS, and YPG mediums was 100 μg/ml, unless otherwise indicated.

2. Construction of expression plasmids

2.1. Construction of pBGP3-PaBG1b and pPICZα-A-PaBG1b for expression of PaBG1b in *P. pastoris*

The cDNA fragment encoding PaBG1b without the putative signal peptide was amplified from the library of *P. a. spadica* by PCR using PrimeStar (Takara Bio, Shiga, Japan) and the following primer pairs: PaBG1b-Mun I F (5'-CCGCAATTGCATGAGGAACATTCTAGAACTAAAAGG-3') and PaBG1b-Not I R (5'-AAAGCGGCCGCCTACGTCCTGTATGCTTCAGGTATT-3'), with *Mun* I and *Not* I sites are underlined. The amplified fragment was inserted into the *Eco*R I and *Not* I sites of the *P. pastoris* expression vector pBGP3 to generate the expression plasmid pBGP3-PaBG1b. The *pabg1b* fragment on the resultant plasmid was verified by DNA

sequencing. The *P. pastoris* strain KM71H was employed as the expression host. Transformation of the pBGP3-PaBG1b to *P. pastoris* by electroporation was performed.

The construction of pPICZα-A-PaBG1b was similar to that of the pBGP3-PaBG1b, except for the stop codon of the reverse primer was removed. The cDNA fragment encoding PaBG1b without the putative signal peptide was amplified from the library of *P. a. spadica* by PCR using PrimeStar and the following primer pairs: PaBG1b-Mun I F (5'-CCGCAATTGCATGAGGAACATTCTAGAACTAAAAGG-3'; *Mun* I site was underlined) and PaBG1b-Not I R (5'-AAAGCGGCCGCCGTCCTGTATGCTTCAGGTATT-3'; *Not* I site was underlined and the stop codon was removed). The amplified fragment was inserted into the *Eco*R I and *Not* I sites of the *P. pastoris* expression vector pPICZα-A (Invitrogen) to generate the expression plasmid pPICZα-A-PaBG1b. After verification of the *pabg1b* fragment on the resultant plasmid by DNA sequencing, the plasmid DNA was linearized with *Bpu*1102I (New England Biolabs) for integration into the chromosomal DNA of *P. pastoris* strain KM71H (Invitrogen, Carlsbad, USA). Transformation of the pPICZα-A-PaBG1b by electroporation and screening of transformants were performed by Colony PCR.

2.2. Construction of pBGP3-RsBG, pPICZα-A-RsBG, pBGP1-RsBGst and pPICZα-A-RsBGnt for expression of RsBG in *P. pastoris*

The *rsbg* sequence (without the signal sequence) was firstly amplified by PCR. Then 5'-end phosphorylation of the PCR product was performed, and subsequently digested by *Eco*R V to generate blunt ends. The resultant fragment was subcloned into the *Eco*R V-digested, dephosphorylated pBluescript II KS (+). Then the *rsbg* insert was

cut from the vector by *Mun* I and *Xba* I, and used for ligation with the *Eco*R I and *Xba* I double-digested pBGP3 vector. The resultant plasmid, pBGP3-RsBG, was confirmed by DNA sequencing and transformed to *P. pastoris* KM71H strain by electroporation.

To construct the pPICZα-A-RsBG, the DNA fragment carrying the C-terminal part of prepro α-factor and *rsbg* was transferred from the pBGP3-RsBG to pPICZα-A by simply cutting and re-ligation of the *Xho* I-*Xba* I fragment to generate the pPICZα-A-RsBG. The pPICZα-A-RsBG was linearized by digestion with *Bgl* II and transformed to *P. pastoris* KM71H strain by electroporation.

To construct the pBGP1-RsBGst, the *rsbg* sequence (without the signal sequence) was amplified by PCR from pBluescript II KS (+)-RsBG. Then the *rsbg* insert was cut from the vector by *Mun* I and *Xba* I, and used for ligation with the pBGP1 vector double digested by *Eco*R I and *Xba* I to generate pBGP1-RsBGst.

To construct the pPICZα-A-RsBGnt, the *rsbg* sequence with the stop codon was excised from pBGP1-RsBG by digesting with *Xho* I *and Xba* I. Then the DNA fragment was employed for ligation with the vector, pPICZα-A double-digested with the same enzyme set, and in the end generating the pPICZα-A-RsBGnt. The rsulting plasmid was then transformed to *P. pastoris* KM71H strain by electroporation.

2.3 Construction of pBGP3-RsBG E273D for expression RsBG E273D in *P. pastoris*

Single site mutation was conducted with $pBGP_3$-RsBG as the template for mutation of the E275 to the aspartic acid by two rounds of PCR. In the first round of PCR, two fragments, each size is approximately half of the *rsbg*, were amplified with a mutation A to T at position 825 of RsBG conding sequence without signal sequence. Then the two fragments were employed in the second round of PCR to generat a full

length RsBG E273D coding sequence. The amplified sequence was used for ligation with pBGP3 vector to generate pBGP3-RsBG E273D. The resulting product was then transformed to *P. pastoris* KM71H strain by electroporation.

3. Transformation of expression plasmid in *P.pastoris* by electroporation

To prepare the competent cells, a clone of *P. pastoris* host strain was picked from the YPD plate, transferred to a 50 ml Erlenmeyer flask containing 10 ml YPD medium, and incubated at a rotary shaker at 30°C for 24 hr. The culture was subsequently transferred to a 500 ml Erlenmeyer flask containing 200 ml of YPD medium and incubated for approximately 12 h until A_{600} of the culture reached to 0.8-1.5. Cells in forty milliliter of the culture was collected in a 50 ml Falcon tube and centrifuged at 1,700 × *g* for 5 min at 4°C, and the pellet was re-suspended by 40 ml of Milli Q water. Then perform centrifugation again to wash the cells. The operation was repeated for twice, then washed the cells for twice with the same procedure except for using 1 M sorbitol instead of Milli Q water. In the end, the competent cells were collected in 1 ml of 1 M sorbitol.

Transformation of *P. pastoris* KM71 strain was done by electroporation following the standard protocol (Lin-Cereghino *et al.*, 2005). Briefly, 10 μl of transforming plasmid containing 2.5 mg of total DNA was mixed with 50 μl of competent cells and cooled down on ice for 5 min. Then the cell suspension was put into the iced-cuvette (Bio-Rad, catalog No. 165-2086) and pulsed at 2.1 kV. Immediately, 1 ml of 1 M sorbitol was added and transferred to a sterilized 1.5 ml tube, following the incubation at 30°C for 1 h with vigorous shaking. After incubation, the tube was centrifuged at 1,500 *g* for 5 min at room temperature and the pellet was spread into an YPDS + zeocin

agar plate for the selection of transformants. After 3 or 4 days, colonies were selected for culture in YPD + zeocin liquid medium by Colony PCR.

4. Growth of transformant strains in the liquid media

4.1. Expression in *P. pastoris* transformantion with pBGP3 constructs.

A clone of the *P. pastoris* transformant strain was picked up and transferred into a 50 ml Erlenmeyer flask containing 10 ml YPD + zeocin medium, and pre-incubated at 30°C in a rotary shaker at 150 rpm for 1 day. Then 100 μl of the cell suspension was transferred to a 500 ml Erlenmeyer flask containing 50 ml YPD + zeocin medium and incubated at 30°C, 150 rpm for 5 days of the main culture. The culture supernatant was harvested by centrifugation at 6,000 g, 4°C for 5 min, for subsequent analysis. To make a glycerol stock of *P. pastoris* transformant strain, 500 μl of the pre-cultured cell suspension was mixed with an equal volume of 50% sterile glycerol and stored at -80°C.

4.2. Expression of *P. pastoris* transformantion with pPICZα-A constructs.

The *P. pastoris* transformant was grown in a 50 ml Erlenmeyer flask containing 10 ml of YPG medium with zeocin on a rotary shaker at 30°C, 150 rpm for 24 h, then the culture was transferred to a 500 ml Erlenmeyer flask containing 200 ml of YPG medium for cell proliferation. After incubating for 24 h, the OD_{600} was reached 14-16, and the cells were harvested by centrifugation at 1,700 g, 4°C for 5 min, washed by 40 ml of Mill Q water for twice, and transferred to the same flask containing 40 ml of YP medium for methanol induction. Methanol feed was performed at the rate of 500 μl per day, and aliquots of the culture supernatant samples were collected daily for the analysis.

The methanol feed was performed for around 4 to 6 days.

5. Time-course analysis of expression

To determine the optimum expression time, time-course analysis was performed, by incubating three *P. pastoris* transformants harboring pBGP3-PaBG1b or pBGP3-RsBG in YPD + zeocin medium and conducting the main culture for 5 to 6 days. For pPICZα-A constructs, the methanol feed was performed for 6 to 7 days. Three transformants harboring the empty vectors were employed as the control strains, and the culture supernatant of the transformants were sampled every day for analysis.

6. Ammonium sulfate precipitation analysis

Ammonium sulfate solution was added to 1 ml of the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant to reach a series of saturation, from 45% to 90%, constantly stirred for 3 h at 4°C. The mixture was subjected to centrifugation at 12,000 *g*, 4°C for 5 min to remove the contaminants. Subsequently, the pellet was re-suspended in 1 ml of 50 mM sodium acetate buffer, pH 5.5 and dialysis towards 1 L of the same buffer at 4°C for overnight to desalt. Then perform enzyme assay.

After a appropriate saturation of ammonium sulfate was determined, for processing large bulk volumes of the culture supernatant, ammonium sulfate was added to a lower saturation and constantly stirred for 3 h at 4°C for precipitating the contaminant proteins, then increased to the desired saturation and constant stirred for 3 h at 4°C to precipitate the target protein. After centrifugation at 12,000 *g*, 4°C for 5 min, the crude target protein in the pellet was re-suspended in an appropriate volume of 50 mM sodium phosphate buffer, pH 8.0 (or 50 mM Tris-HCl, pH 8.0) and dialysis towards 1 L of the

same buffer at 4°C for overnight for desalting.

7. Standard Ni-NTA purification (for both PaBG1b and RsBG)

The Ni-NTA column (Invitrogen) packed with 1 ml of Nickel-chelating resin and equilibrated with the binding buffer (50 mM sodium phosphate buffer, 300 mM sodium chloride, pH 8.0). A certain volume (for example, 20 ml) of the culture supernatant was mixed with the same volume of binding buffer, and loaded onto the column. The column was then washed and eluted manually in a stepwise manner using the same binding buffer containing 20 mM and 250 mM imidazole. The elution fractions, with each volume was 1 ml, were diluted for at least 5 folds against 50 mM sodium acetate buffer, pH 5.5 to restore the pH to circumneutral or acidic pH for enzyme assay.

8. Buffer exchanging to remove imidazole

To remove imidazole, The Ni-NTA purified products was concentrated to less than 1 ml by a Vivaspin20 ultrafiltration cell (10,000 MWCO, Sartorius Stedim, Germany) and centrifugation at 12,000g, 4°C, then added 20 ml of 50 mM sodium acetate buffer (pH 5.5) to the cell and repeat the operation for three times to remove imidazole.

9. Ni-NTA purification of PaBG1b by EDTA eluting

The operation was basically the same as the standard Ni-NTA purification, except that substituting imidazole in wash and elution buffer with 1 mM and 10 mM of EDTA, respectively. The eluted fractions was diluted for at least 5 folds against 50 mM sodium acetate buffer (pH 5.5) to restore the pH to circumneutral or acidic, and then performed enzyme assay, or dialysis to remove EDTA towards 1 L of the same beffer at 4°C, for

overnight, and perform enzyme assay.

10. Ni-NTA purification of PaBG1b by pH eluting

The operation was basically the same as the standard Ni-NTA purification, except employing 50 mM of sodium phosphate buffer, 300 mM sodium chloride (pH 7.0) as the wash buffer, and 50 mM of sodium acetate buffer (pH 5.0) as the elution buffer. The eluates can be directly used for enzyme assay.

11. Anion exchange chromatography for purification of PaBG1b

Fifty milliliter of the culture suernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant was subjected to ammonium sulfate precipitation treatment and the resultant precipitate was re-suspended in 50 mM HEPES buffer, pH 7.0, and dialyzed against 1 L of the same buffer at 4°C, for overnight. The resultant crude enzyme solution was loaded onto a HiTrap DEAE FF anion exchange column (1 ml; GE Healthcare, catalog No. 17-5154-01) pre-equilibrated with the strating buffer (i.e., 50 mM HEPES buffer, pH 7.0). The column was washed with 10 ml of the starting buffer and eluted with 60 ml of linear gradient to 1.0 M sodium chloride in the same buffer at a flow rate of 1 ml/min, with the collection of 1.0 ml fractions. The chromatography procedure was controlled by AKTAprime plus (GE Healthcare), and the fractionated samples were subsequently subjected to protein concentration determination and enzyme activity assay. Active fractions were collected for further analysis.

12. Introducing Tris in Ni-NTA purification

The operation is basically the same as the standard Ni-NTA purification, except that

substituting sodium phosphate buffers with 50 mM Tris-HCl buffers, 300 mM sodium chloride (pH 8.0), and washing and eluting with 5 mM and 100 mM of imidazole, respectively. Piror to enzyme assay, the eluates were diluted for at least 5 folds against 50 mM sodium acetate buffer (pH 5.5) to restore the pH.

13. Large scale purification of PaBG1b

To purify PaBG1b from 300 ml of the culture supernatant of the pPICZα-A-PaBG1b *P. pastoris* transformant, the saturated ammonium sulfate solution was added to the culture supernatant to reach a saturation of 45%, which was constantly stirred for 3 h at 4°C. The mixture was centrifuged at 12,000 *g*, 4°C for 5 min to remove contaminants. Then the saturation of ammonium sulfate was increased to 65% and the resultant mixture was stirred for 3 h at 4°C to precipitate PaBG1b. After centrifuging at 12,000 *g*, 4°C for 5 min, the crude PaBG1b protein in the pellet was re-suspended in an appropriate volume of the binding buffer (50 mM Tris-HCl, 300 mM sodium chloride, pH 8.0) and loaded onto a Ni-NTA column (Invitrogen) packed with 1 ml of Nickel-chelating resin and pre-equilibrated with the binding buffer. The column was washed and eluted manually in a stepwise manner using the same buffer containing 5 mM and 100 mM imidazole, respectively. The elutes were immediately subjected to dialysis against 1 L of the starting buffer (50 mM Tris-HCl, pH 7.0) at 4°C for 4 h to remove imidazole. The resultant solution was concentrated to less than 2 ml by a Vivaspin20 ultrafiltration cell, and applied to a HiTrap DEAE FF anion exchange column (1 ml; GE Healthcare) pre-equilibrated with the starting buffer. The column was washed with 10 ml of the starting buffer and eluted with 0-0.5 M linear gradient of sodium chloride in 60 ml of the same buffer at a flow rate of 1 ml/min, with the

collection of 1 ml fractions. The chromatography procedure was controlled by AKTAprime plus (GE Healthcare) and the fractionated samples were subsequently subjected to protein concentration determination and enzyme activity assay. Active fractions were collected for further analysis.

14. Anion exchange chromatography of RsBG

The culture supernatant of the pPICZα-A-RsBG *P. pastoris* transformant was subjected to ammonium sulfate precipitation at the saturation of 80%, and the resultant precipitate was dialyzed towards 1L of 50 mM Tris-HCl buffer (pH 7.0), at 4°C for overnight. The crude RsBG solution was concentrated into less than 2 ml by a Vivaspin20 ultrafiltration cell, and loaded onto a column of HiTrap DEAE FF anion exchange column (1 ml; GE Healthcare) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.0, then eluted with a linear gradient of 0 to 1 M sodium chloride in 40 ml of the same buffer at a flow rate of 1 ml/min, with the collection of 1 ml fractions. The chromatography procedure was controlled by AKTAprime plus (GE Healthcare) and the fractionated samples were subsequently subjected to protein concentration determination and enzyme activity assay. Active fractions were collected for further analysis.

15. Standard activity assay

Standard β-glucosidase assay was initiated by adding 10 μl of the enzyme solution appropriately diluted in 50 mM sodium acetate buffer (pH 5.5) to 100 μl of the substrate solution containing 10 mM *p*-nitrophenyl β-D-glucopyranoside (*p*NPG, Sigma-Aldrich) in the same buffer. After incubating at 45°C for 30 min, 1 ml of 0.6 M sodium carbonate

was added to the mixture to stop the reaction. The activity was determined by measuring the amount of *p*-nitrophenol released by A_{410} using a spectrophotometer (JASCO, V-630 BIO). A standard curve was plotted by using the solution containing serial dilutions of *p*-nitrophenol (Wako, Japan) in 50 mM sodium acetate buffer, pH 5.5. One enzyme unit (U) was defined as the amount of enzyme required to release l μmol of *p*-nitrophenol from the substrate per min. This protocol was applied to routine BG assays, unless otherwise indicated, and the reading of the spectrophotometer should not exceed 1.0.

16. Optimum temperature and thermostability

The optimum temperature was determined by measuring the relative activity at various temperatures ranging from 30°C to 60°C. The thermostability was determined by incubating the enzyme solution at different temperatures from 30°C to 60°C for 30 min, then evaluating the residual activity under the standard enzyme assay conditions.

17. Optimum pH and pH stability

The optimum pH was determined by conducting the assay over the range of pH 3.0 to 9.0. Buffers used were: 50 mM sodium acetate buffer (pH 3.0 to 6.0); 50 mM sodium phosphate buffer (pH 6.0 to 8.0), and 50 mM Britton-Robinson buffer (H_3BO_3: CH_3COOH: H_3PO_4=1:1:1, pH was adjusted by NaOH), pH 8.0 to 9.0. The pH stability was determined by mixing 10 μl of the enzyme solution with the same volume of different buffer sets ranging from pH 3.0 to 9.0, incubated at 30°C for 30 min, and then evaluating the residual activity under the standard enzyme assay condition.

18. Substrate specificity analysis

Analysis of substrate specificity was performed by employing several *p*-nitrophenyl derivatives, and native substrates. For the *p*-nitrophenyl derivatives including *p*-nitrophenyl β-D-fucopyranoside (Tokyo Chemical Industry, Tokyo, Japan), *p*-nitrophenyl-β-D-galactopyranoside, (Nacalai Tesque, Kyoto, Japan), *p*-nitrophenyl α-L-arabinofuranoside and *p*-nitrophenyl β-D-*N*-acetylglucosamine (Sigma-Aldrich), the activity assays followed the routine method except for substitution of *p*NPG with the same concentration of *p*-nitrophenyl derivatives (10 mM). For polysaccharides such as larminarin (Nacalai Tesque, Kyoto, Japan), carboxymethyl cellulose (CMC; Seikagaku, Tokyo, Japan) and Avicel, as well as oligosaccharides including D-(+)-cellobiose (Sigma-Aldrich), cellotriose, cellohexaose, and larminaribiose (Seikagaku, Tokyo, Japan), gentiobiose (Tokyo Chemical Industry, Tokyo, Japan) and β-lactose (Nacalai Tesque, Kyoto, Japan), α-sophorose (Serva Electrophoresis GmbH), sucrose and maltose (Kokusan Chem, Japan), the enzyme assays were conducted by adding 25 μl of enzyme solution to 100 μl of 1% (w/v) polysaccharides in 50 mM sodium acetate buffer (pH 5.5) and incubating at 37°C for 5 min, followed by boiling for 5 min to stop the reaction. The activity was determined by measuring the release of glucose using glucose oxidase-mutarotase reagent (Glucose CII Test Wako, Wako Pure Chemical Co., Tokyo, Japan). The glucose standard solution I in the kit was diluted to five concentrations (0.25, 0.5, 1.0, 1.25, and 1.5 mg/ml) for plotting the standard curve. One enzyme unit (U) was defined as the amount of enzyme required to release 2 μmol of glucose from the substrate per minute. Salicin (Nacalai Tesque, Kyoto, Japan) is also assayed by the same Kit.

19. Effect of cations and reagents on PaBG1b

To evaluate the effect of cations and reagents, 10 μl of enzyme solution was pre-incubated with 50 μl of 5 mM cations or 10 mM reagents preared in 50 mM sodium acetate buffer (pH 5.5) at 30°C for 30 min. Then the activities were determined by routine assay method.

20. Kinetic analysis of PaBG1b

To determine the kinetic constants towards *p*NPG, 10 μl of appropriately diluted enzyme solution (approximately 8 μg/ml) was added into 100 μl of substrate sets containing different concentrations of *p*NPG (from 5 mM to 300 mM), and the mixture was incubated at 45°C for 10 min, the reaction was then stopped by adding 1 ml of sodium carbonate, and the reaction velocity was determined by measuring the released *p*-nitrophenol (*p*NP) at A_{410}.

To determine the kinetic constants towards cellobiose, 25 μl of appropriately diluted enzyme solution (approximately 8 μg/ml) was added into 100 μl of substrate sets containing different concentrations of cellobiose (from 2.5 mM to 150 mM), and the mixture was incubated at 37°C for 5 min, the reaction was then stopped by boiling for 5 min, and the reaction velocity was determined by measuring the released of glucose by using the glucose oxidase-mutarotase reagent (Glucose CII Test Wako; Wako Pure Chemical Co., Tokyo, Japan, catalog No. 439-90901) under the instructions of Wako. Briefly, take 20 μl of the reaction mixture and mixed with 3 ml of Chromogen Reagent, incubate at 37°C for 5 min, then measure at A_{505} within 1 h.

The maximum velocities (V_{max}), Michaelis–Menten (K_m), and the turnover number (k_{cat}) constants were calculated by a nonlinear regression of the Michaelis-Menten

equation using GraphPad PRISM 7 (GraphPad Software, La Jolla, CA) and the catalytic efficiency was presented as k_{cat}/K_m.

For the evaluation of the inhibition constant (K_i) of glucose (Kokusan Chem, Japan), imidazole (Tokyo Chemical Industry, Tokyo, Japan) and Tris (Invitrogen, Carlsbad, CA, USA), 10 μl of enzyme solution was added to 100 μl of solution sets containing various concentrations of *p*NPG (25, 50, and 100 mM) and inhibitors (0, 200, 400, and 600 mM for glucose; 0, 5, 10, and 20 mM for imidazole and Tris) in 50 mM sodium acetate buffer, pH 5.5, and incubated at 45°C for 10 min. After adding 1 ml sodium carbonate to stop the reaction, the reaction velocity was determined by measuring the release of *p*-nitrophenol at A_{410}, and the K_i of corresponding inhibitor was calculated from the Dixon plot.

21. Thin layer chromatography (TLC) analysis for transglycosylation activity of PaBG1b

The TLC analysis was conducted by adding 5 μl of the enzyme solution to 50 μl of 1 % (w/v) substrates in 50 mM sodium acetate buffer, pH 5.5, and incubating at 37°C for 1 h. Two μl of the resultant mixture were developed on a silica gel 60 TLC plate (Merck) with a solvent system of 1-butanol: ethanol: distilled water (2:2:1), dried at room temperature and the oligosaccharides released were visualized by staining with a reagent comprised of 2.5% (w/v) anisaldehyde, 1% (v/v) acetic acid, and 3.4% (v/v) concentrated sulfuric acid in ethanol and baked at 100°C for 5 min in an oven.

22. Post-translational modification

Protein deglycosylation analysis was carried out by digesting the purified PaBG1b

with glycopeptidase F (Takara, Japan) or endoglycosidase H (New England Biolabs, Ipswich, USA) according to the manufacturer's instructions. Two micrograms of purified PaBG1b were denatured by boiling for 10 min, as described by manuals of manufactures, and digested in the denatured condition. The resultant products were resolved by SDS-PAGE and visualized by CBB staining.

23. SDS-PAGE and Native PAGE

The protein samples was examined by SDS-PAGE in gels containing 10% polyacrylamide and according to the method introduced by Laemmli (1970). For the Native PAGE, the reagents employed were the same with those of SDS-PAGE except for excluding of SDS and 2-mercaptoethanol (2-ME). After the electrophoresis, the gels were subjected to CBB staining and/or Western blot analysis.

In the Western blot analysis, the proteins were transferred from the SDS-PAGE gel to a PVDF membrane, and incubated with anti-c-Myc mouse monoclonal antibody (1:1,000 dilution, Clontech) and a peroxidase-labeled anti-mouse-immunoglobulin G antibody (1:500 dilution, Vector), respectively. Then the PVDF membrane was rinsed by ELC detection reagents (Pierce) and the blots were detected by a luminescent image analyzer LAS-4000miniEPUV (Fujifilm, Japan).

24. Protein assay

Protein concentration was analyzed according to the Bradford method (1976), by measuring the absorbance at 595 nm (A_{595}) using the Bio-Rad protein assay solution (Bio-Rad) 5 fold diluted in distilled water. Bovine serum albumin (Takara, Japan) was used as the standard with concentration of 0.2, 0.4, 0.6, and 0.8 mg/ml. Specific activity

was expressed as enzyme units of per milligram of protein (U/mg).

25. Microsequencing of purified PaBG1b

Approximately 10 μg of purified PaBG1b was developed on a SDS-PAGE gel and transferred to a polyvinylidene difluoride membrane with cathode buffer without glycine (pH 9.4) and CAPS buffer (10 mM 3-[cyclohexylamino]-1-propanesulfonic acid, 10% methanol, pH 11.0) and stained for proteins using 0.1% Coomassie Blue R-250 in a 50% (v/v) methanol solution according to Matsudaira (Matsudaira, 1987). Then the membrane was destained with a 50% methanol solution containing 10% acetic acid, dried by air at room temperature and subjected to sequence analysis using Procice 491HT (Applied Biosystems, Foster City, USA).

26. Bioinformatic analysis

The deduced amino acid sequence of PaBG1b and RsBG were uploaded to the BLAST server (Altschul et al., 1997; https://blast.ncbi.nlm.nih.gov/Blast.cgi) for homology analysis. The signal sequence of PaBG1b and RsBG were predicted by SignalP 4.1 server (Petersen et al., 2011; http://www.cbs.dtu.dk/services/SignalP/), whilst the amino acid sequence alignment of homologues was performed by ClustalW (http://www.genome.jp/tools/clustalw/) and visualized by GeneDoc (http://genedoc.software.informer.com/2.7/). Potential *N*-and *O*-glycosilation sites were predicted by NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetOGlyc 4.0 (http://www.cbs.dtu.dk/services/NetOGlyc/) servers, respectively.

Rare codon analysis was performed by GenScript (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis). The Amino acid

sequences of PaBG1b and RsBG were analyzed by ScanProsite (Gattiker et al., 2002; http://prosite.expasy.org/scanprosite/) for signature motifs.

27. Polymerase Chain Reaction (PCR) of PaBG1b

27.1 Amplification of *pabg1b* from cDNA template

① Primers for construction of pBGP3-PaBG1b:

PaBG Mun I F: 5'-CCGCAATTGCATGAGGAACATTCTAGAACTAAAAGG-3'

PaBG Not I R: 5'-AAAGCGGCCGCCTACGTCCTGTATGCTTCAGGTATT-3'

R Primer for construction of pBGP3-PaBG1b (F primer is Mun I F):

PaBG1b Not I R2 5'-AAAGCGGCCGCCGTCCTGTATGCTTCAGGTATT-3'

② Reagent for each PCR tube

5X PrimeStar Buffer: 10 μl

2.5 mM dNTPs : 4 μl

10 μm Mun I F : 15 μl

10 μm Not I R (or R2): 15 μl

cDNA template : 0.5 μl

Prime Star : 0.5 μl

D. W. : 4.0 μl

Total 50 μl

③ Procedure:

98°C ——98°C——57°C——72°C————16°C

2 min 10s 5s 1 min/kb* ∞

20 cycles

*: set at 1 min 30 s

27.2 Colony PCR for *E.coli* transformants

① Reagent for each PCR tube

10X Ex-Taq Buffer: 2 μl

2.5 mM dNTPs : 2 μl

10 μm Mun I F : 2 μl

10 μm Not I R (or R2): 2 μl

r-TAq : 0.1 μl

D. W. : 11.9 μl

Total 20 μl

② Procedure：

94°C	94°C	55°C	72°C	72°C	4°C
5 min	30s	30s	1 min/kb*	7 min	∞

30 cycles

*: set at 1 min 30 s

27.3 Colony PCR for *P. pastoris* transformants

① Reagent for each PCR tube

2X KOD FX NEO Buffer: 10 μl

2 mM dNTPs : 4 μl

10 μm Mun I F : 0.4 μl

10 μm Not I R (or R2): 0.4 μl

KOD FX NEO : 0.4 μl

TE : 2 μl

D. W. : 2.8 μl

Total 20 μl

② Procedure：

94°C——98°C——65°C—— 68°C———4°C

2 min 10s 30s 30 s/kb** ∞

40 cycles

**: set at 48 s

28 Polymerase Chain Reaction (PCR) of RsBG

28.1 Amplification of *rsbg* from cDNA template

① Primers for construction of pBGP3-RsBG

G3RsBG Mun I F: 5'-CAATTGGCAAGTCCTGAAGAGTTATTAGC-3'

G3RsBG Xba I R: 5'-TCTAGACTAAGACTCTTTCGTCAGTCC-3'

② Reagent for each PCR tube

10X PCR Buffer:		5 μl
2.5 mM dNTPs	:	4 μl
10 μm G3 RsBG Mun I F	:	2.5 μl
10 μm G3 RsBG Xba I R:		2.5 μl
cDNA template	:	1 μl
Pyrobest	:	0.5
D. W.	:	4.5 μl
Total		20 μl

③ Procedure：

98°C——98°C ——57°C ———— 72°C———16°C

10s 10s 30s 1 min/kb* ∞

30 cycles

*: set at 1 min 42 s

28.2 Colony PCR for *E.coli* transformants

① Reagent for each PCR tube

10X Ex-Taq Buffer: 2 μl

2.5 mM dNTPs : 2 μl

10 μm G3 RsBG Mun I F : 2 μl

10 μm G3 RsBG Xba I R 2 μl

r-TAq : 0.1 μl

D. W. : 11.9 μl

Total 20 μl

② Procedure：

94°C——94°C——55°C—— 72°C————72°C——4°C

5 min 30s 30s 1 min/kb* 5 min ∞

30 cycles

*: set at 1 min 42 s

28.3 Colony PCR for *P. pastoris* transformants

① Reagent for each PCR tube

2X KOD FX NEO Buffer: 10 μl

2 mM dNTPs : 4 μl

10 μm G3 RsBG Mun I F: 0.4 μl

10 μm G3 RsBG Xba I R: 0.4 μl

KOD FX NEO : 0.4 μl

TE : 2 μl

D. W. : 2.8 μl

Total 20 μl

② Procedure：

94°C——98°C——65°C——68°C————4°C

2 min 10s 30s 30 s/kb** ∞

40 cycles **: set at 51 s

29. Polymerase Chain Reaction (PCR) for site mutation of RsBG

29.1 Amplificaition of two site mutation fragments

① Mutation primers:

G3RsBGD Mt F GTTGTCTCTGATTACGGGGC

G3RsBGD Mt R GCCCCGTAATCAGAGACAAC

② Reagent for fragment 1:

5X PCR Buffer:		10 μl
2.5 mM dNTPs	:	4 μl
10 μm G3 RsBG Mun I F	:	15 μl
10 μm G3 G3RsBGD Mt R:		15 μl
pBGP3-RsBG	:	1 μl
PrimeStar	:	0.5
D. W.	:	4.5 μl
Total		50 μl

③ Procedure：

98°C——98°C——52°C——72°C————16°C

2 min 10 s 5 s 1 min/kb* ∞

20 cycles *: set at 52 s

④ Reagent for fragment 2:

5X PCR Buffer: 10 μl

2.5 mM dNTPs : 4 μl

10 μm G3RsBGD Mt F : 15 μl

10 μm G3 G3RsBGD Xba I: 15 μl

pBGP3-RsBG : 1 μl

PrimeStar : 0.5 μl

D. W. : 4.5 μl

Total 50 μl

⑤ Procedure：

98°C——98°C——52°C——72°C————16°C

2 min 10 s 5 s 1 min/kb* ∞

20 cycles *: set at 52 s

29.2 Amplificaition of RsBG E273D fragment

① PCR reagent:

5X PCR Buffer: 10 μl

2.5 mM dNTPs : 4 μl

20 μm G3RsBGD Mt F : 7.5 μl

20 μm G3 G3RsBGD Xba I: 7.5 μl

Fragment 1 : 10 μl

Fragment 2 : 10 μl

PrimeStar : 0.5 μl

D. W. : 0.5 μl

Total 50 μl

② Procedure：

98°C ——98°C——52°C——72°C————16°C

2 min 10 s 5 s 1 min/kb* ∞

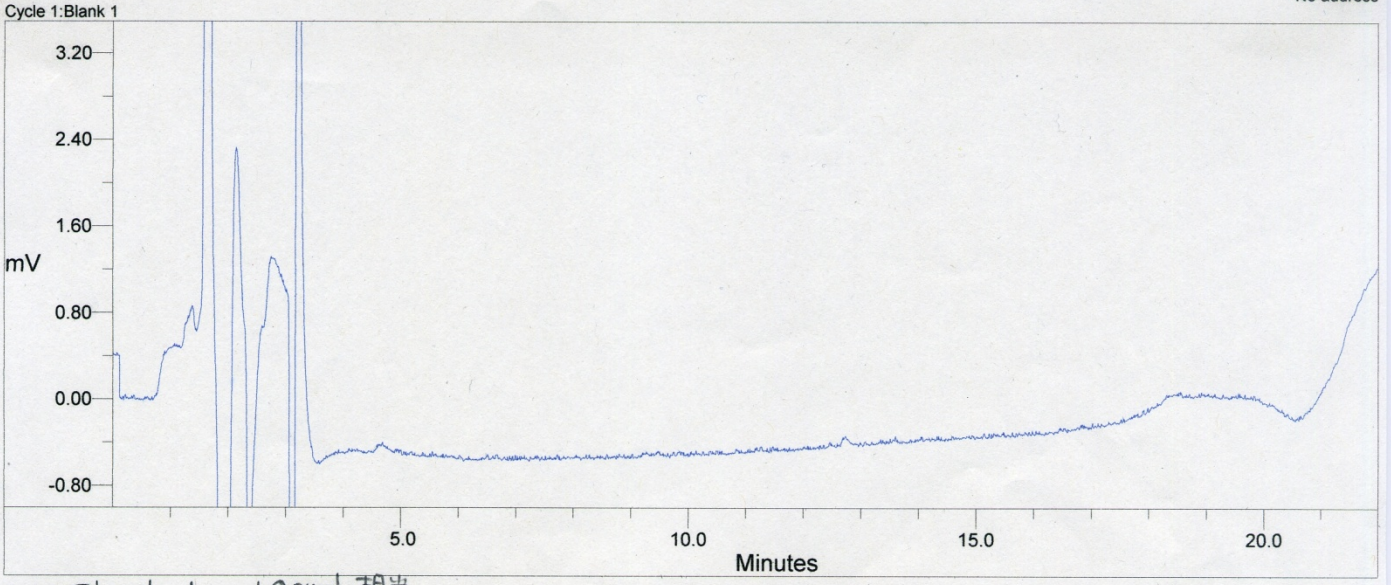
20 cycles

*: set at 1 min 42 s

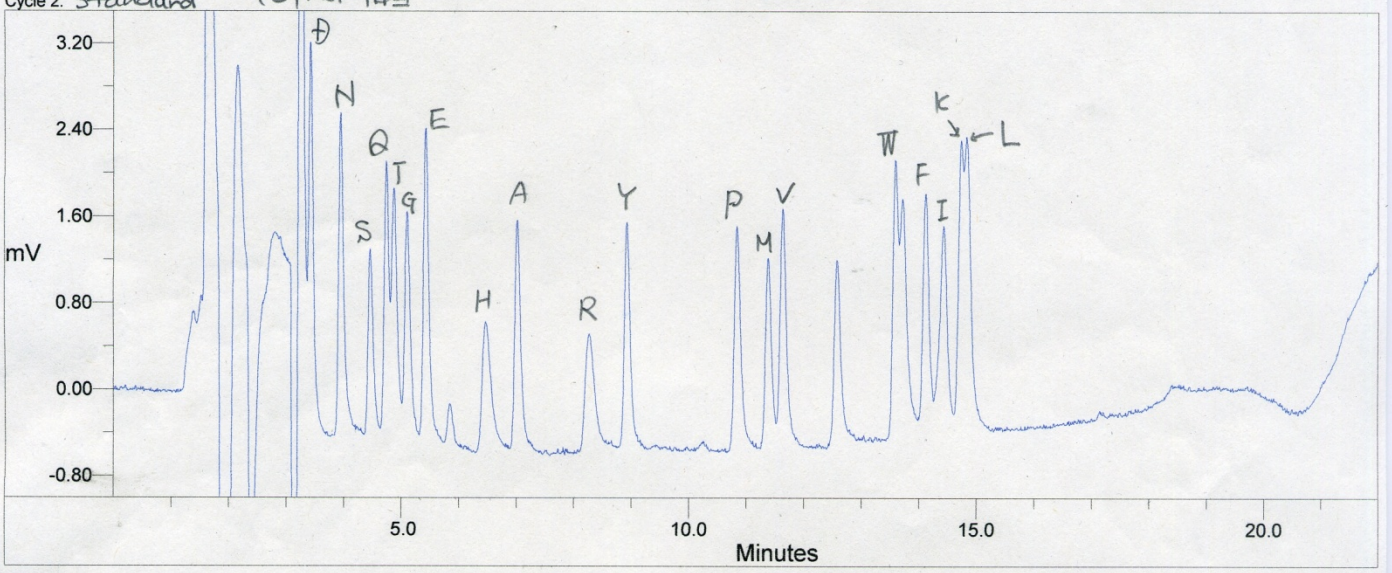
Supplemental File 1

No lab name
No address

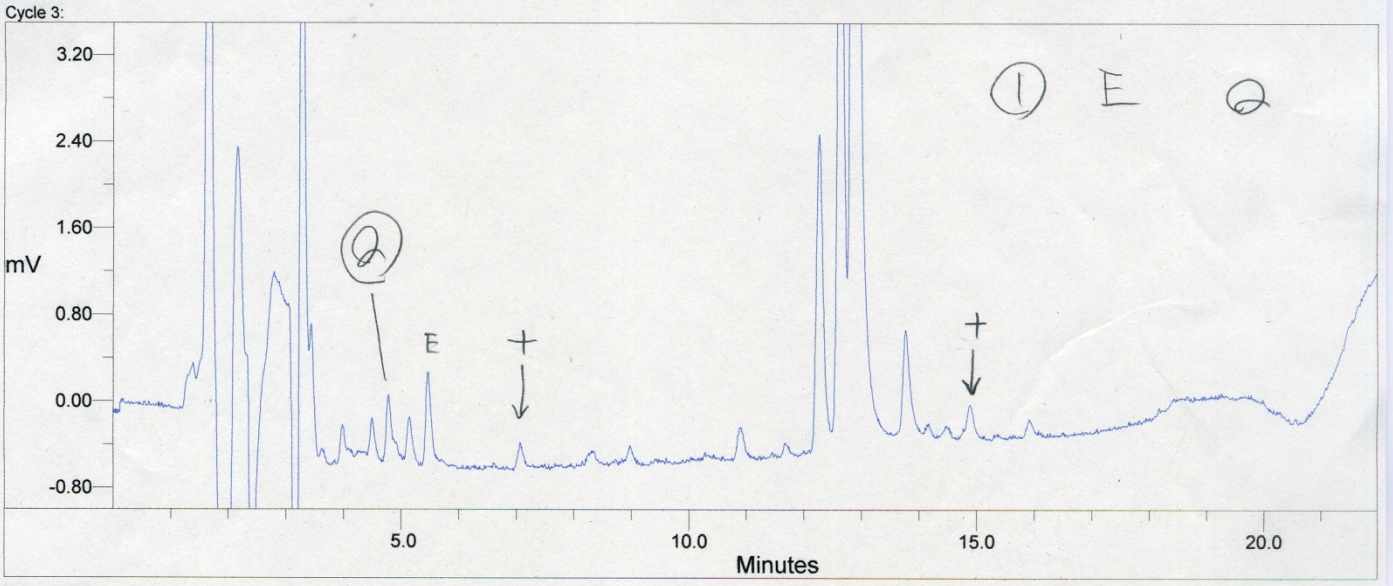
Cycle 1:Blank 1



Cycle 2: Standard 10pmol 相当

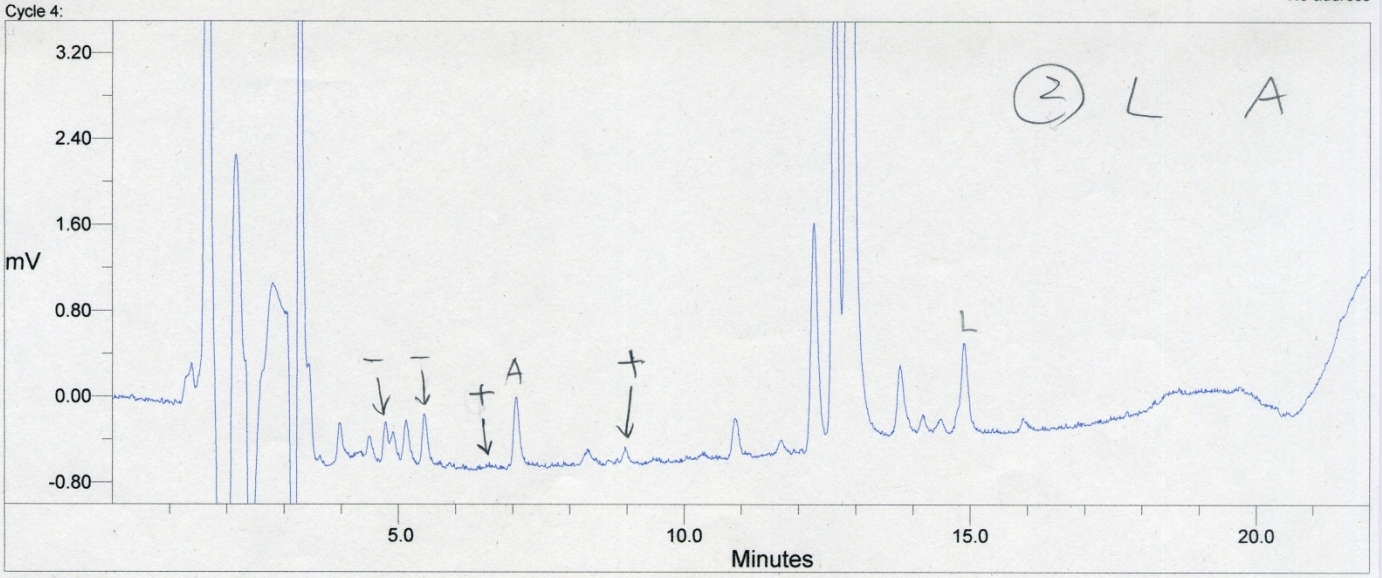


Cycle 3:

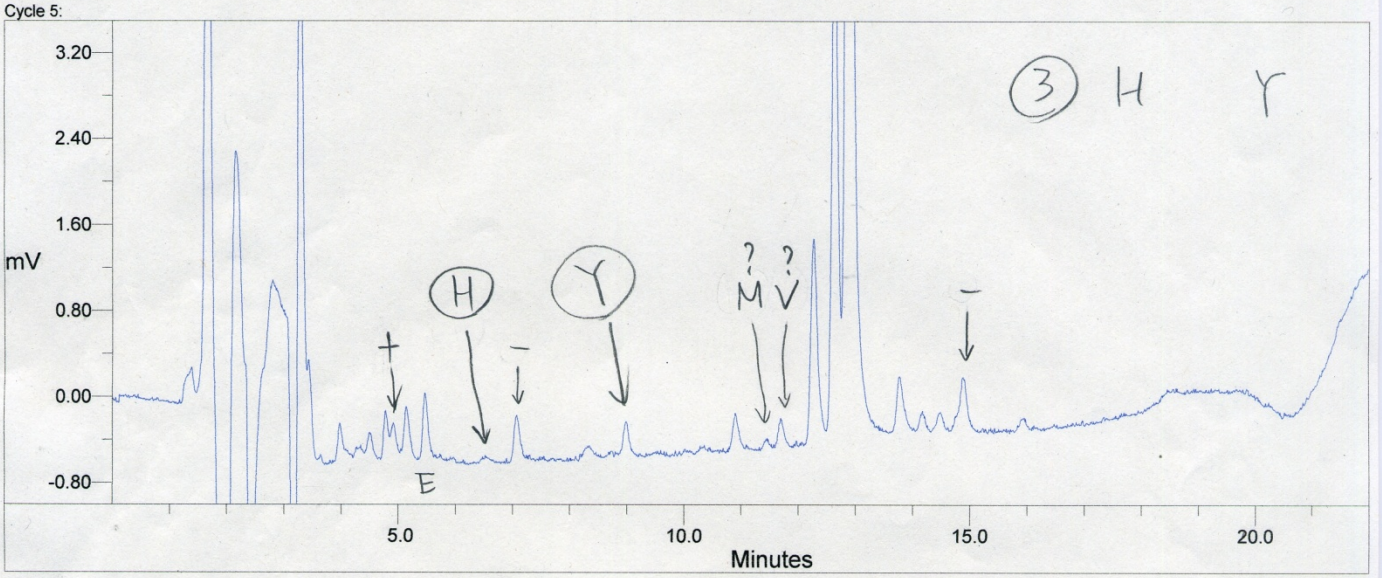


No lab name
No address

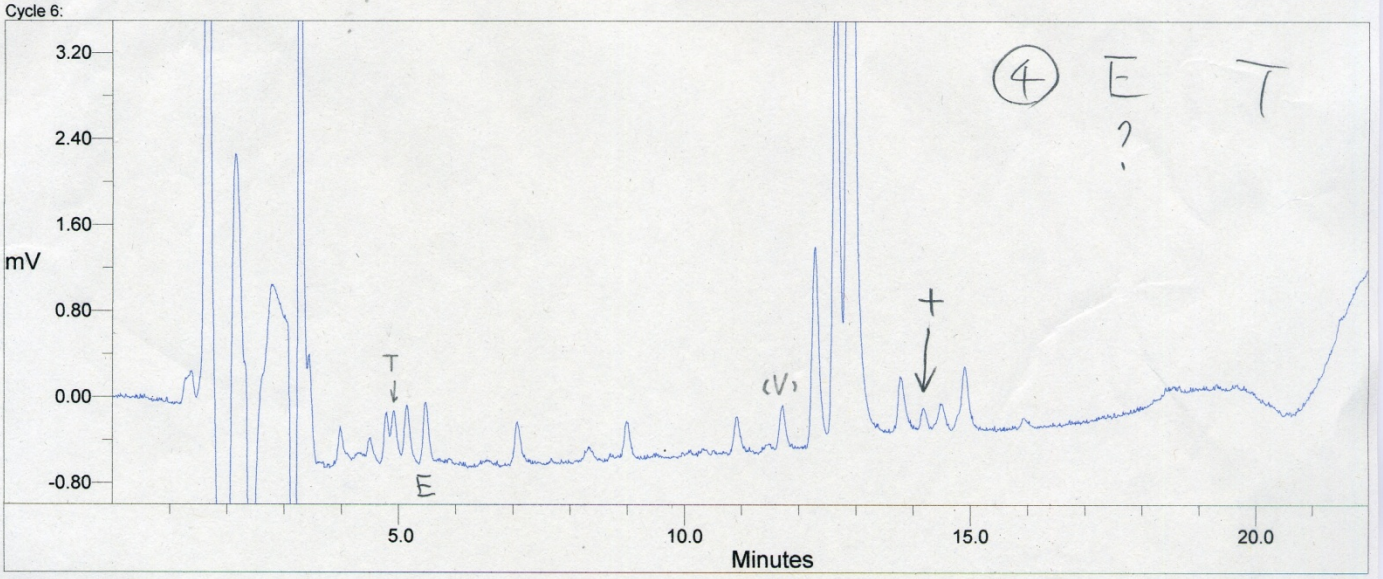
Cycle 4:



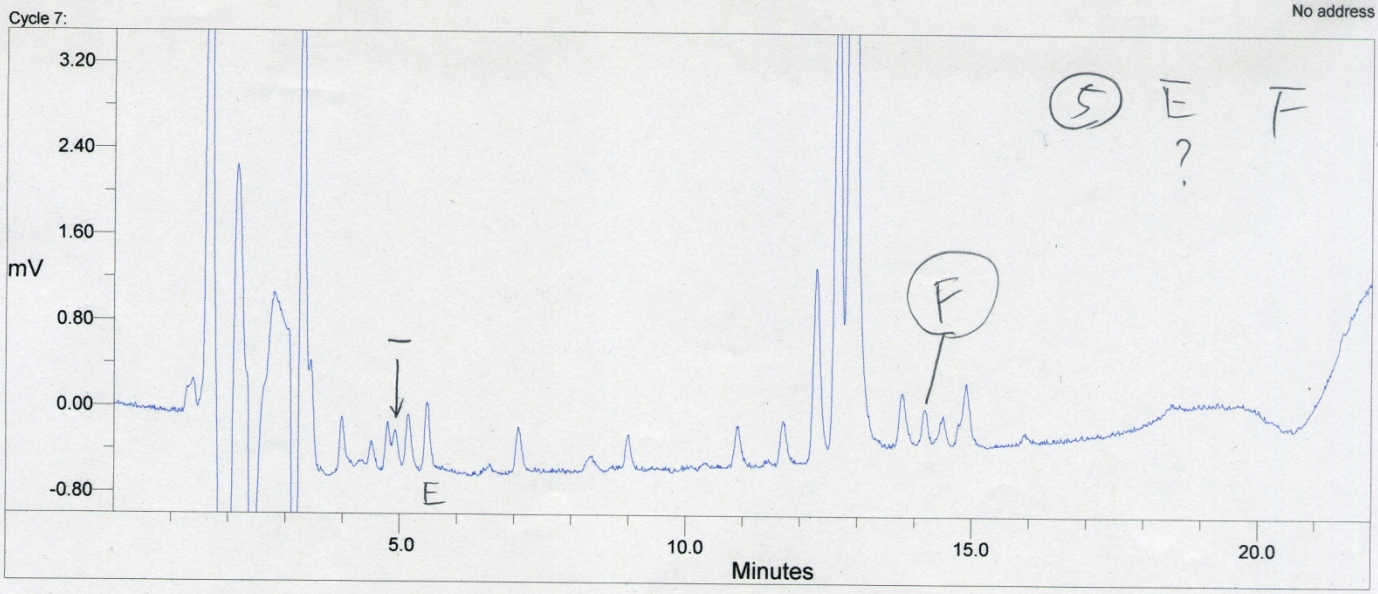
Cycle 5:



Cycle 6:



No lab name
No address



Supplemental File 2

```
        10         20         30         40         50         60
MRFTLIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AQAYQRAVDI VSQMTLDEKV
        70         80         90        100        110        120
NLTTGTGWEL ELCVGQTGGV PRLGVPGMCL QDSPLGVRDS DYNSAFPAGM NVAATWDKNL
       130        140        150        160        170        180
AYLRGKAMGQ EFSDKGADIQ LGPAAGPLGR SPDGGRNWEG FSPDPALSGV LFAETIKGIQ
       190        200        210        220        230        240
DAGVVATAKH YIAYEQEHFR QAPEAQGYGF NISESGSANL DDKTMHELYL WPFADAIRAG
       250        260        270        280        290        300
AGAVMCSYNQ INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
       310        320        330        340        350        360
GDVDYDSGTS YWGTNLTISV LDGTVPQWRV DDMAVRIMAA YYKVGRDRLW TPPNFSSWTR
       370        380        390        400        410        420
DEYGYKYYYV SEGPYEKVNQ YVNVQRNHSE LIRRIGADST VLLKNDGALP LTGKERLVAL
       430        440        450        460        470        480
IGEDAGSNPY GANGCSDRGC DNGTLAMGWG SGTANFPYLV TPEQAISNEV LKHKNGVFTA
       490        500        510        520        530        540
TDNWAIDQIE ALAKTASVSL VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAATE
       550        560        570        580        590        600
PNNGNGAPQE DFVEGVFIDY RGFDKRNETP IYEFGYGLSY TTFNYSNLEV QVLSAPAYEP
       610        620        630        640        650        660
ASGETEAAPT FGEVGNASDY LYPSGLQRIT KFIYPWLNGT DLEASSGDAS YGQDSSDYLP
       670        680        690        700        710        720
EGATDGSAQP ILPAGGGPGG NPRLYDELIR VSVTIKNTGK VAGDEVPQLY VSLGGPNEPK
       730        740        750        760        770        780
IVLRQFERIT LQPSEETKWS TTLTRRDLAN WNVEKQDWEI TSYPKMVFVG SSSRKLPLRA
       790        800        810        820        830        840
SLPTVH
```

Source: Lima *et al*. (2013)

Amino acid sequence of *An*Bgl1.

The amino acid sequence of *An*Bgl1, manually cited from the supplemental file 1 and 2 of the source reference, was found identical to the deduced amino acid sequence of a BG precursor of *A. niger* (GenBank accession number ABH01182), with the exception of a 74 amino acid-long insert in the middle of the BG precursor. Two catalytic residues, D280 and E509, are hightligted in red.

References

Achstetter T, Wolf DH. Hormone processing and membrane-bound proteinases in yeast. EMBO J 1985;4(1):173-177.

Agirre J, Ariza A, Offen WA, Turkenburg JP, Roberts SM, McNicholas S, Harris PV, McBrayer B, Dohnalek J, Cowtan KD, Davies GJ, Wilson KS. Three-dimensional structures of two heavily N-glycosylated *Aspergillus* sp. family GH3 β-D-glucosidases. Acta Crystallographica Section D: Structural Biology 2016;72(2):254-265.

Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 2014;98(12):5301-5317.

Ait N, Creuzet N, Cattaneo J. Properties of β-glucosidase purified from *Clostridium thermocellum*. Microbiology 1982;128(3):569-577.

Akemi Uchima C, Tokuda G, Watanabe H, Kitamoto K, Arioka M. A novel glucose-tolerant β-glucosidase from the salivary gland of the termite *Nasutitermes takasagoensis*. J Gen Appl Microbiol 2013;59(2):141-145.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389-3402.

Arakawa G, Kamino K, Tokuda G, Watanabe H. Purification, characterization, and

cDNA cloning of a prominent β-glucosidase from the gut of the xylophagous cockroach Panesthia angustipennis spadica. J. Appl. Glycosci 2016;63(3):51-59

Arthan D, Kittakoop P, Esen A, Svasti J. Furostanol glycoside 26-O-β-glucosidase from the leaves of *Solanum torvum*. Phytochemistry 2006;67(1):27-33.

Ballou CE. A study of the immunochemistry of three yeast mannans. J Biol Chem 1970;245(5):1197-1203.

Banks WM. Carbohydrate digestion in the cockroach. Science 1963;141(3586):1191-1192.

Barnett CC, Berka RM, Fowler T. Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. Nat Biotechnol 1991;9(6):562-567.

Bause E, Legler G. Isolation and structure of a tryptic glycopeptide from the active site of β-glucosidase A 3 from *Aspergillus wentii*. Biochim Biophys Acta. 1980;626(2):459-465.

Béguin P, Aubert J. The biological degradation of cellulose. FEMS Microbiol Rev 1994;13(1):25-58.

Belancic A, Gunata Z, Vallier M, Agosin E. β-glucosidase from the grape native yeast

Debaryomyces vanrijiae: purification, characterization, and its effect on monoterpene content of a muscat grape juice. J Agric Food Chem 2003;51(5):1453-1459.

Bhatia Y, Mishra S, Bisaria V. Microbial β-glucosidases: cloning, properties, and applications. Crit Rev Biotechnol 2002;22(4):375-407.

Bissett F, Sternberg D. Immobilization of *Aspergillus* beta-glucosidase on chitosan. Appl Environ Microbiol 1978;35(4):750-755.

Biver S, Stroobants A, Portetelle D, Vandenbol M. Two promising alkaline β-glucosidases isolated by functional metagenomics from agricultural soil, including one showing high tolerance towards harsh detergents, oxidants and glucose. J Ind Microbiol Biotechnol 2014;41(3):479-488.

Bohlin C, Praestgaard E, Baumann MJ, Borch K, Praestgaard J, Monrad RN, Westh P. A comparative study of hydrolysis and transglycosylation activities of fungal β-glucosidases. Appl Microbiol Biotechnol 2013;97(1):159-169.

Breznak JA, Brune A. Role of microorganisms in the digestion of lignocellulose by termites. Annu Rev Entomol 1994;39(1):453-487.

Britton HTS, Robinson RA. CXCVIII.--Universal buffer solutions and the dissociation constant of veronal. Journal of the Chemical Society (Resumed) 1931:1456-1462.

Brune A. Symbiotic digestion of lignocellulose in termite guts. Nat Rev Microbiol 2014;12(3):168-180.

Brune A. Termite guts: the world's smallest bioreactors. Trends Biotechnol 1998;16(1):16-21.

Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MW, Lunin VV, Himmel ME, Bomble YJ. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 2013;342(6165):1513-1516.

Cannella D, Hsieh CW, Felby C, Jørgensen H. Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. Biotechnol Biofuels. 2012;5(1):26.

Cannella D, Jørgensen H. Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production? Biotechnol Bioeng 2014;111(1):59-68.

Cao LC, Wang ZJ, Ren GH, Kong W, Li L, Xie W, Liu YH. Engineering a novel glucose-tolerant β-glucosidase as supplementation to enhance the hydrolysis of sugarcane bagasse at high glucose concentration. Biotechnology for biofuels 2015;8(1):1.

Castle LA, Smith KD, Morris RO. Cloning and sequencing of an *Agrobacterium tumefaciens* β-glucosidase gene involved in modifying a *vir*-inducing plant signal molecule. J Bacteriol 1992;174(5):1478-1486.

Çelik E, Çalık P. Production of recombinant proteins by yeast cells. Biotechnol Adv 2012;30(5):1108-1118.

Chen HL, Chen YC, Lu MY, Chang JJ, Wang HT, Ke HM, Wang TY, Ruan SK, Wang TY, Hung KY, Cho HY, Lin WT, Shih MC, Li WH. A highly efficient β-glucosidase from the buffalo rumen fungus *Neocallimastix patriciarum* W5. Biotechnol Biofuels 2012;5(1):24-6834-5-24.

Chen P, Fu X, Ng TB, Ye X. Expression of a secretory β-glucosidase from *Trichoderma reesei* in *Pichia pastoris* and its characterization. Biotechnol Lett 2011;33(12):2475-2479.

Chinchetru MA, Cabezas JA, Calvo P. Purification and characterization of a broad specificity β-glucosidase from sheep liver. Int J Biochem 1989;21(5):469-476.

Cho M, Kim Y, Shin K, Kim Y, Kim Y, Kim T. Symbiotic adaptation of bacteria in the gut of *Reticulitermes speratus*: low endo-β-1, 4-glucanase activity. Biochem Biophys Res Commun 2010;395(3):432-435.

Chuenchor W, Pengthaisong S, Robinson RC, Yuvaniyama J, Oonanant W, Bevan DR,

Esen A, Chen CJ, Opassiri R, Svasti J, Cairns JR. Structural insights into rice BGlu1 β-glucosidase oligosaccharide hydrolysis and transglycosylation. J Mol Biol 2008;377(4):1200-1215.

Cicek M, Esen A. Expression of soluble and catalytically active plant (monocot) β-glucosidases in *E. coli*. Biotechnol Bioeng 1999;63(4):392-400.

Claassen P, Van Lier J, Contreras AL, Van Niel E, Sijtsma L, Stams A, De Vries SS Weusthuis RA. Utilisation of biomass for the supply of energy carriers. Appl Microbiol Biotechnol 1999;52(6):741-755.

Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA. High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Biotechnology (N Y) 1991a;9(5):455-460.

Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA. Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. Gene 1991b;105(2):205-212.

Cleveland LR. Symbiosis between termites and their intestinal protozoa. Proc Natl Acad Sci U S A 1923;9(12):424-428.

Cornette R, Farine J, Abed-Viellard D, Quennedey B, Brossut R. Molecular

characterization of a male-specific glycosyl hydrolase, Lma-p72, secreted on to the abdominal surface of the Madeira cockroach *Leucophaea maderae* (Blaberidae, Oxyhaloinae). Biochem J 2003;372:535-541.

Couderc R, Baratti J. Oxidation of methanol by the yeast, *Pichia pastoris*. Purification and properties of the alcohol oxidase. Agric Biol Chem 1980;44(10):2279-2289.

Coughlan MP. The properties of fungal and bacterial cellulases with comment on their production and application. Biotechnol Genet Eng Rev 1985;3(1):39-110.

Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol Cell Biol 1989;9(3):1316-1323.

Cregg JM, Tschopp JF, Stillman C, Siegel R, Akong M, Craig WS, Buckholz RG, Madden KR, Kellaris PA, Davis GR. High-Level Expression and Efficient Assembly of Hepatitis B Surface Antigen in the Methylotrophic Yeast, *Pichia Pastoris*. Nat Biotechnol 1987;5(5):479-485.

Crook EM, Stone BA. Formation of oligosaccharides during the enzymic hydrolysis of β-glucosides. Biochem J 1953;55(320th Meeting):xxv.

Crook EM, Stone BA. The enzymic hydrolysis of β-glucosides. Biochem J 1957;65(1):1-12.

Dahlqvist A. Pig intestinal β-glucosidase activities I. Relation to β-galactosidase (lactase). Biochim Biophys Acta 1961;50(1):55-61.

Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure 1995;3(9):853-859.

Davies GJ, Wilson KS, Henrissat B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J 1997;321 (Pt 2)(Pt 2):557-559.

de Giuseppe PO, Souza Tde A, Souza FH, Zanphorlin LM, Machado CB, Ward RJ, Jorge JA, Furriel Rdos P, Murakami MT. Structural basis for glucose tolerance in GH1 β-glucosidases. Acta Crystallogr D Biol Crystallogr. 2014;70(Pt 6):1631-1639.

De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouzé P, Van de Peer Y, Callewaert N. Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat Biotechnol 2009;27(6):561-566.

Decker C, Visser J, Schreier P. β-Glucosidase multiplicity from *Aspergillus tubingensis* CBS 643.92: purification and characterization of four β-glucosidases and their differentiation with respect to substrate specificity, glucose inhibition and acid tolerance. Appl Microbiol Biotechnol 2001;55(2):157-163.

Dekker RF. Kinetic, inhibition, and stability properties of a commercial β-D-glucosidase

(cellobiase) preparation from *Aspergillus niger* and its suitability in the hydrolysis of lignocellulose. Biotechnol Bioeng 1986;28(9):1438-1442.

Del Pozo MV, Fernández-Arrojo L, Gil-Martínez J, Montesinos A, Chernikova TN, Nechitaylo TY, Waliszek A, Tortajada M, Rojas A, Huws SA, Golyshina OV, Newbold CJ, Polaina J, Ferrer M, Golyshin PN. Microbial β-glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail. Biotechnol Biofuels. 2012;5(1):73

Delmer DP, Amor Y. Cellulose biosynthesis. Plant Cell 1995;7(7):987-1000.

Dimarogona M, Topakas E, Christakopoulos P. Cellulose degradation by oxidative enzymes. Computational and structural biotechnology journal 2012;2(3):1-8.

Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT, Jones TA. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 1994;265(5171):524-528.

Duff SJ, Cooper DG, Fuller OM. Cellulase and beta-glucosidase production by mixed culture of *Trichoderma reesei* RUT C30 and *Aspergillus phoenicis*. Biotechnol Lett 1985;7(3):185-190.

Eggleton P. Termites and trees: a review of recent advances in termite phylogenetics. Insectes Soc 2001;48(3):187-193.

Eggleton P. An introduction to termites: biology, taxonomy and functional morphology. *In*: Biology of termites: a modern synthesis. *Edited by* Bignell D, Roisin Y, Lo N. Springer; 2011. p. 1-26.

Ergün BG, Çalık P. Lignocellulose degrading extremozymes produced by *Pichia pastoris*: current status and future prospects. Bioprocess Biosyst Eng 2015:1-36.

Eyzaguirre J, Hidalgo M, Leschot A. 23 β-Glucosidases from filamentous fungi: properties, structure, and applications. *In*: handbook of carbohydrate engineering. *Edited by* Kevin J. Yarema. Boca Raton, FL, USA. 2005;645-685

Falk A, Rask L. Expression of a zeatin-O-glucoside-degrading β-glucosidase in *Brassica napus*. Plant Physiol 1995;108(4):1369-1377.

Fan HX, Miao LL, Liu Y, Liu HC, Liu ZP. Gene cloning and characterization of a cold-adapted β-glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium *Micrococcus antarcticus*. Enzyme Microb Technol 2011;49(1):94-99.

Fang Z, Fang W, Liu J, Hong Y, Peng H, Zhang X, Sun B, Xiao Y. Cloning and characterization of a β-glucosidase from marine microbial metagenome with excellent glucose tolerance. J Microbiol Biotechnol 2010;20(9):1351-1358.

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM. Ethanol can contribute to energy and environmental goals. Science 2006;311(5760):506-508.

Ferrara MC, Cobucci-Ponzano B, Carpentieri A, Henrissat B, Rossi M, Amoresano A, Moracci M. The identification and molecular characterization of the first archaeal bifunctional exo-β-glucosidase/N-acetyl-β-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. Biochim Biophys Acta 2014;1840(1):367-377.

Field RA, Haines AH, Chrystal EJ, Luszniak MC. Histidines, histamines and imidazoles as glycosidase inhibitors. Biochem J. 1991;274 (Pt 3):885-889.

Flannelly DF, Aoki TG, Aristilde L. Short-time dynamics of pH-dependent conformation and substrate binding in the active site of beta-glucosidases: A computational study. J Struct Biol 2015;191(3):352-364.

Fleming LW, Duerksen JD. Purification and characterization of yeast β-glucosidases. J Bacteriol 1967;93(1):135-141.

Fushinobu S. Metalloproteins: A new face for biomass breakdown. Nat Chem Biol 2014;10(2):88-89.

Gattiker A, Gasteiger E, Bairoch AM. ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinformatics. 2002;1(2):107-8.

Gloster TM, Roberts S, Perugino G, Rossi M, Moracci M, Panday N, Terinek M, Vasella A, Davies GJ. Structural, kinetic, and thermodynamic analysis of glucoimidazole-derived glycosidase inhibitors. Biochemistry 2006;45(39):11879-11884

Gong C, Ladisch MR, Tsao GT. Cellobiase from *Trichoderma viride*: purification, properties, kinetics, and mechanism. Biotechnol Bioeng 1977;19(7):959-981.

Graham TR, Emr SD. Compartmental organization of Golgi-specific protein modification and vacuolar protein sorting events defined in a yeast sec18 (NSF) mutant. J Cell Biol 1991;114(2):207-218.

Grinna LS, Tschopp JF. Size distribution and general structural features of *N*-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. Yeast 1989;5(2):107-115.

Gudmundsson M, Larsson AM, Hansson H, Stals I, Larenas E, Kaper T, Sandgren M. Crystal structure of glycoside hydrolase family 3 Beta-Glucosidase Cel3A from the moderately thermophilic fungus *Rasamsonia emersonii* (to be published). (Message cited from http://www.rcsb.org/pdb/explore/explore.do?structureId=4D0J)

Günata Z, Vallier M. Production of a highly glucose-tolerant extracellular β-glucosidase by three *Aspergillus* strains. Biotechnol Lett 1999;21(3):219-223.

Guo B, Amano Y, Nozaki K. Improvements in glucose sensitivity and stability of *Trichoderma reesei* β-glucosidase using site-directed mutagenesis. PLoS One. 2016;11(1):e0147301.

Guo Y, Yan Q, Yang Y, Yang S, Liu Y, Jiang Z. Expression and characterization of a novel β-glucosidase, with transglycosylation and exo-β-1,3-glucanase activities, from *Rhizomucor miehei*. Food Chem 2015;175:431-438.

Halliwell G, Griffin M. The nature and mode of action of the cellulolytic component C_1 of *Trichoderma koningii* on native cellulose. Biochem J 1973;135(4):587-594.

Han YW, Srinivasan VR. Purification and characterization of β-glucosidase of *Alcaligenes faecalis*. J Bacteriol 1969;100(3):1355-1363.

Hanson SR, Culyba EK, Hsu TL, Wong CH, Kelly JW, Powers ET. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. Proc Natl Acad Sci U S A 2009;106(9):3131-3136.

Harhangi HR, Steenbakkers PJ, Akhmanova A, Jetten MS, van der Drift C, Op den Camp HJ. A highly expressed family 1 β-glucosidase with transglycosylation capacity from the anaerobic fungus *Piromyces* sp. E2. Biochim Biophys Acta. 2002;1574(3):293-303.

Harnpicharnchai P, Champreda V, Sornlake W, Eurwilaichitr L. A thermotolerant

β-glucosidase isolated from an endophytic fungi, *Periconia* sp., with a possible use for biomass conversion to sugars. Protein Expr Purif 2009;67(2):61-69.

Harvey AJ, Hrmova M, De Gori R, Varghese JN, Fincher GB. Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. Proteins 2000;41(2):257-269.

Haven MØ, Jørgensen H. Adsorption of β-glucosidases in two commercial preparations onto pretreated biomass and lignin. Biotechnol Biofuels 2013;6(1):165.

Heightman TD, Vasella AT. Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases. Angew Chem Int Edit 1999;38(6):750-770.

Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280 (Pt 2):309-316.

Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem J. 1996;316 (Pt 2):695-696.

Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol. 1997;7(5):637-644.

Henrissat B, Driguez H, Viet C, Schülein M. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. Nat Biotechnol 1985;3(8):722-726.

Heyworth R, Walker PG. Almond-emulsin β-D-glucosidase and β-D-galactosidase. Biochem J 1962;83:331-335.

Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A 2006;103(30):11206-11210.

Himmel ME, Adney WS, Fox JW, Mitchell DJ, Baker JO. Isolation and characterization of two forms of β-D-glucosidase from *Aspergillus niger*. Appl Biochem Biotechnol 1993;39(1):213-225.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 2007;315(5813):804-807.

Hirayama K, Watanabe H, Tokuda G, Kitamoto K, Arioka M. Purification and characterization of termite endogenous β-1, 4-endoglucanases produced in *Aspergillus oryzae*. Biosci Biotechnol Biochem 2010;74(8):1680-1686.

Hong J, Tamaki H, Kumagai H. Cloning and functional expression of thermostable β-glucosidase gene from *Thermoascus aurantiacus*. Appl Microbiol Biotechnol 2007;73(6):1331-1339.

Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. Novel enzymes for the degradation of cellulose. Biotechnol Biofuels 2012;5(1):1-13.

Hösel W, Barz W. β-Glucosidases from *Cicer arietinum* L.: purification and properties of isoflavone-7-O-glucoside-specific-β-glucosidases. Eur J Biochem. 1975;57(2):607-616.

Hrmova M, MacGregor EA, Biely P, Stewart RJ, Fincher GB. Substrate binding and catalytic mechanism of a barley β-D-glucosidase/(1,4)-β-D-glucan exohydrolase. J Biol Chem 1998;273(18):11134-11143.

Huang C, Damasceno LM, Anderson KA, Zhang S, Old LJ, Batt CA. A proteomic analysis of the *Pichia pastoris* secretome in methanol-induced cultures. Appl Microbiol Biotechnol 2011;90(1):235-247.

Igarashi K, Ishida T, Hori C, Samejima M. Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete *Phanerochaete chrysosporium*. Appl Environ Microbiol 2008;74(18):5628-5634.

Igarashi K, Tani T, Kawai R, Samejima M. Family 3 β-glucosidase from cellulose-degrading culture of the white-rot fungus *Phanerochaete chrysosporium* is a glucan 1, 3-β-glucosidase. J Biosci Bioeng 2003;95(6):572-576.

Inoue T, Moriya S, Ohkuma M, Kudo T. Molecular cloning and characterization of a

cellulase gene from a symbiotic protist of the lower termite, *Coptotermes formosanus*. Gene 2005;349:67-75.

Inoue T, Murashima K, Azuma J, Sugimoto A, Slaytor M. Cellulose and xylan utilisation in the lower termite *Reticulitermes speratus*. J Insect Physiol 1997;43(3):235-242.

Inward D, Beccaloni G, Eggleton P. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. Biol Lett 2007;3(3):331-335.

Iwashita K, Nagahara T, Kimura H, Takano M, Shimoi H, Ito K. The *bglA* gene of *Aspergillus kawachii* encodes both extracellular and cell wall-bound β-glucosidases. Appl Environ Microbiol 1999;65(12):5546-5553.

Janbon G, Magnet R, Arnaud A, Galzy P. Cloning and sequencing of the β-glucosidase-encoding gene from *Candida molischiana* strain 35M5N. Gene 1995;165(1):109-113.

Jeng WY, Wang NC, Lin CT, Chang WJ, Liu CI, Wang AH. High-resolution structures of *Neotermes koshunensis* β-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates. Acta Crystallogr D Biol Crystallogr 2012;68(Pt 7):829-838.

Jeng WY, Wang NC, Lin MH, Lin CT, Liaw YC, Chang WJ, Liu CI, Liang PH, Wang AH. Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. J Struct Biol 2011;173(1):46-56.

Jenkins J, Lo Leggio L, Harris G, Pickersgill R. β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. FEBS Lett 1995;362(3):281-285.

Jeya M, Joo AR, Lee KM, Tiwari MK, Lee KM, Kim SH, Lee JK. Characterization of β-glucosidase from a strain of *Penicillium purpurogenum* KJS506. Appl Microbiol Biotechnol 2010;86(5):1473-1484.

Joo AR, Jeya M, Lee KM, Sim WI, Kim JS, Kim IW, Kim YS, Oh DK, Gunasekaran P, Lee JK. Purification and characterization of a β-1, 4-glucosidase from a newly isolated strain of *Fomitopsis pinicola*. Appl Microbiol Biotechnol 2009;83(2):285-294.

Jørgensen BB, Jørgensen OB. Inhibition of barley malt α-glucosidase by tris (hydroxymethyl) aminomethane and erythritol. Biochim Biophys Acta. 1967;146(1):167-172.

Jørgensen H, Vibe-Pedersen J, Larsen J, Felby C. Liquefaction of lignocellulose at high-solids concentrations. Biotechnol Bioeng 2007;96(5):862-870.

Jorgensen MJ, Cantor AB, Furie BC, Brown CL, Shoemaker CB, Furie B. Recognition site directing vitamin K-dependent γ-carboxylation resides on the propeptide of factor IX. Cell 1987;48(2):185-191.

Ju X, Bowden M, Engelhard M, Zhang X. Investigating commercial cellulase performances toward specific biomass recalcitrance factors using reference substrates. Appl Microbiol Biotechnol 2014;98(10):4409-4420.

Julius D, Blair L, Brake A, Sprague G, Thorner J. Yeast α factor is processed from a larger precursor polypeptide: the essential role of a membrane-bound dipeptidyl aminopeptidase. Cell 1983;32(3):839-852.

Julius D, Brake A, Blair L, Kunisawa R, Thorner J. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-α-factor. Cell 1984;37(3):1075-1089.

Kadam SK, Demain AL. Addition of cloned β-glucosidase enhances the degradation of crystalline cellulose by the *Clostridium thermocellum* cellulase complex. Biochem Biophys Res Commun 1989;161(2):706-711.

Kalyani D, Lee KM, Tiwari MK, Ramachandran P, Kim H, Kim IW, Jeya M, Lee JK. Characterization of a recombinant aryl β-glucosidase from *Neosartorya fischeri* NRRL181. Appl Microbiol Biotechnol 2012;94(2):413-423.

Kambhampati S, Eggleton P. Taxonomy and phylogeny of termites. *In*: Termites: evolution, sociality, symbioses, ecology. *Edited by* Abe T, Bignell DE, Higashi M. Springer; 2000. p. 1-23.

Karasová-Lipovová P, Strnad H, Spiwok V, Malá Š, Králová B, Russell NJ. The cloning, purification and characterization of a cold-active β-galactosidase from the psychrotolerant Antarctic bacterium *Arthrobacter* sp. C2-2. Enzyme Microb Technol 2003;33(6):836-844.

Karkehabadi S, Helmich KE, Kaper T, Hansson H, Mikkelsen NE, Gudmundsson M, Piens K, Fujdala M, Banerjee G, Scott-Craig JS, Walton JD, Phillips GN Jr, Sandgren M. Biochemical characterization and crystal structures of a fungal family 3 β-glucosidase, Cel3A from *Hypocrea jecorina*. J Biol Chem 2014;289(45):31624-31637.

Kaur A, Chadha BS. *Penicillium janthinellum*: a source of efficient and high levels of β-glucosidase. Appl Biochem Biotechnol 2015;175(2):937-949.

Kawai R, Igarashi K, Kitaoka M, Ishii T, Samejima M. Kinetics of substrate transglycosylation by glycoside hydrolase family 3 glucan (1→3)-β-glucosidase from the white-rot fungus *Phanerochaete chrysosporium*. Carbohydr Res 2004;339(18):2851-2857.

Kawai R, Yoshida M, Tani T, Igarashi K, Ohira T, Nagasawa H, Samejima M.

Production and characterization of recombinant *Phanerochaete chrysosporium* β-glucosidase in the methylotrophic yeast *Pichia pastoris*. Biosci Biotechnol Biochem 2003;67(1):1-7.

Kawai T, Nakazawa H, Ida N, Okada H, Tani S, Sumitani J, Kawaguchi T, Ogasawara W, Morikawa Y, Kobayashi Y. Analysis of the saccharification capability of high-functional cellulase JN11 for various pretreated biomasses through a comparison with commercially available counterparts. J Ind Microbiol Biotechnol 2012;39(12):1741-1749.

Kawamori M, Ado Y, Takasawa S. Preparation and application of *Trichoderma reesei* mutants with enhanced β-glucosidase. Agric Biol Chem 1986a;50(10):2477-2482.

Kawamori M, Morikawa Y, Takasawa S. Induction and production of cellulases by L-sorbose in Trichoderma reesei. Appl Microbiol Biotechnol 1986b;24(6):449-453.

Ketudat Cairns JR, Esen A. β-Glucosidases. Cell Mol Life Sci 2010;67(20):3389-3405.

Koffi YG, Konan K, Kouadio E, Dabonné S, Kouamé L. Purification and biochemical characterization of β-glucosidase from cockroach *Periplaneta americana*. J Anim Plant Sci 2012;13:1747-1757.

Korotkova OG, Semenova MV, Morozova VV, Zorov IN, Sokolova LM, Bubnova TM, Okunev ON, Sinitsyn AP. Isolation and properties of fungal β-glucosidases.

Biochemistry (Mosc). 2009;74(5):569-577.

Koshland DE. Stereochemistry and the mechanism of enzymatic reactions. Biol Rev 1953;28(4):416-436.

Koutz P, Davis GR, Stillman C, Barringer K, Cregg J, Thill G. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast 1989;5(3):167-177.

Krisch J, Bencsik O, Papp T, Vágvölgyi C, Takó M. Characterization of a β-glucosidase with transgalactosylation capacity from the zygomycete *Rhizomucor miehei*. Bioresour Technol 2012;114:555-560.

Krogh KB, Harris PV, Olsen CL, Johansen KS, Hojer-Pedersen J, Borjesson J, Olsson L. Characterization and kinetic analysis of a thermostable GH3 β-glucosidase from *Penicillium brasilianum*. Appl Microbiol Biotechnol 2010;86(1):143-154.

Kubicek CP. Involvement of a conidial endoglucanase and a plasma-membrane-bound β-glucosidase in the induction of endoglucanase synthesis by cellulose in *Trichoderma reesei*. Microbiology 1987;133(6):1481-1487.

Kurasawa T, Yachi M, Suto M, Kamagata Y, Takao S, Tomita F. Induction of Cellulase by Gentiobiose and Its Sulfur-Containing Analog in *Penicillium purpurogenum*. Appl Environ Microbiol 1992;58(1):106-110.

Kurjan J, Herskowitz I. Structure of a yeast pheromone gene (MFα): a putative α-factor precursor contains four tandem copies of mature α-factor. Cell 1982;30(3):933-943.

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680-685.

Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U. Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. Biotechnol Biofuels 2014 Oct 18;7(1):135-014-0135-5. eCollection 2014.

Lee CC, Williams TG, Wong DW, Robertson GH. An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. Plasmid 2005;54(1):80-85.

Lee HL, Chang CK, Jeng WY, Wang AH, Liang PH. Mutations in the substrate entrance region of β-glucosidase from *Trichoderma reesei* improve enzyme activity and thermostability. Protein Eng Des Sel 2012;25(11):733-740.

Lee JE, Fusco ML, Saphire EO. An efficient platform for screening expression and crystallization of glycoproteins produced in human cells. Nat Protoc 2009;4(4):592-604.

Lee JM, Kim YR, Kim JK, Jeong GT, Ha JC, Kong IS. Characterization of salt-tolerant β-glucosidase with increased thermostability under high salinity conditions from *Bacillus* sp. SJ-10 isolated from jeotgal, a traditional Korean fermented seafood. Bioprocess Biosyst Eng 2015;38(7):1335-1346.

Li G, Jiang Y, Fan X, Liu Y. Molecular cloning and characterization of a novel β-glucosidase with high hydrolyzing ability for soybean isoflavone glycosides and glucose-tolerance from soil metagenomic library. Bioresour Technol 2012;123:15-22.

Li YK, Byers LD. Inhibition of β-glucosidase by imidazoles. Biochim Biophys Acta 1989;999(3):227-232.

Li YK, Hsu HS, Chang LF, Chen G. New imidazoles as probes of the active site topology and potent inhibitors of beta-glucosidase. J Biochem 1998;123(3):416-422.

Lima MA, Oliveira-Neto M, Kadowaki MA, Rosseto FR, Prates ET, Squina FM, Leme AF, Skaf MS, Polikarpov I. *Aspergillus niger* β-glucosidase has a cellulase-like tadpole molecular shape: insights into glycoside hydrolase family 3 (GH3) β-glucosidase structure and function. J Biol Chem 2013;288(46):32991-33005.

Lin-Cereghino J1, Wong WW, Xiong S, Giang W, Luong LT, Vu J, Johnson SD, Lin-Cereghino GP. Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. Biotechniques. 2005;38(1):44, 46, 48.

Liu D, Zhang R, Yang X, Zhang Z, Song S, Miao Y, Shen Q. Characterization of a thermostable β-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. Microb Cell Fact. 2012;11:25.

Liu J, Zhang X, Fang Z, Fang W, Peng H, Xiao Y. The 184th residue of β-glucosidase Bgl1B plays an important role in glucose tolerance. J Biosci Bioeng 2011;112(5):447-450.

Lo N, Tokuda G, Watanabe H, Rose H, Slaytor M, Maekawa K, Bandi C, Noda H. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. Curr Biol 2000;10(13):801-804.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42(Database issue):D490-495.

Lopez-Camacho C, Salgado J, Lequerica JL, Madarro A, Ballestar E, Franco L, Polaina J. Amino acid substitutions enhancing thermostability of *Bacillus polymyxa* β-glucosidase A. Biochem J 1996;314 (Pt 3):833-838.

Lu J, Du L, Wei Y, Hu Y, Huang R. Expression and characterization of a novel highly glucose-tolerant β-glucosidase from a soil metagenome. Acta Biochim Biophys Sin (Shanghai) 2013;45(8):664-673.

Lynd LR, Cushman JH, Nichols RJ, Wyman CE. Fuel ethanol from cellulosic biomass. Science 1991;251(4999):1318-1323.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. How biotech can transform biofuels. Nat Biotechnol 2008;26(2):169-172.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 2002;66(3):506-577.

Mabee W, Saddler J. From 1st-to 2nd-generation biofuel technologies. An overview of current industry and RD&D activities. *In*: IEA Bioenergy. *Edited by*: Taylor M, Sims R. OECD/IEA, Paris 2008. p. 14-16.

Maekawa K, Lo N, Kitade O, Miura T, Matsumoto T. Molecular phylogeny and geographic distribution of wood-feeding cockroaches in East Asian islands. Mol Phylogenet Evol 1999;13(2):360-376.

Maekawa K, Matsumoto T, Nalepa C. Social biology of the wood-feeding cockroach genus *Salganea* (Dictyoptera, Blaberidae, Panesthiinae): did ovoviviparity prevent the evolution of eusociality in the lineage? Insectes Soc 2008;55(2):107-114.

Maguire RJ. Kinetics of the hydrolysis of cellobiose and *p*-nitrophenyl-β-D-glucoside by cellobiase of *Trichoderma viride*. Can J Biochem 1977;55(1):19-26.

Marques AR, Coutinho PM, Videira P, Fialho AM, Sa-Correia I. *Sphingomonas paucimobilis* β-glucosidase Bgl1: a member of a new bacterial subfamily in glycoside

hydrolase family 1. Biochem J 2003;370(Pt 3):793-804

Martin M. M. Cellulose digestion in insects. Comp Biochem Physiol A 1983;75(3): 313–324

Martin MM, Martin JS. Cellulose Digestion in the Midgut of the Fungus-Growing Termite *Macrotermes natalensis*: The Role of Acquired Digestive Enzymes. Science 1978;199(4336):1453-1455.

Matsudaira P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem 1987;262(21):10035-10038.

Matsuura K, Yashiro T, Shimizu K, Tatsumi S, Tamura T. Cuckoo fungus mimics termite eggs by producing the cellulose-digesting enzyme β-glucosidase. Curr Biol 2009;19(1):30-36.

Matsuzawa T, Jo T, Uchiyama T, Manninen JA, Arakawa T, Miyazaki K,4, Fushinobu S, Yaoi K. Crystal structure and identification of a key amino acid for glucose tolerance, substrate specificity and transglycosylation activity of metagenomic β-glucosidase Td2F2. FEBS J. 2016;283(12):2340-2353.

Mattéotti C, Haubruge E, Thonart P, Francis F, De Pauw E, Portetelle D, Vandenbol M. Characterization of a new β-glucosidase/β-xylosidase from the gut microbiota of the termite (*Reticulitermes santonensis*). FEMS Microbiol Lett 2011;314(2):147-157.

McAndrew RP, Park JI, Heins RA, Reindl W, Friedland GD, D'haeseleer P, Northen T, Sale KL, Simmons BA, Adams PD. From soil to structure, a novel dimeric β-glucosidase belonging to glycoside hydrolase family 3 isolated from compost using metagenomic analysis. J Biol Chem 2013;288(21):14985-14992.

McMahon LG, Nakano H, Levy MD, Gregory JF III. Cytosolic pyridoxine-β-D-glucoside hydrolase from porcine jejunal mucosa: purification, properties, and comparison with broad specificity β-glucosidase. J Biol Chem 1997;272(51):32025-32033.

Meleiro LP, Zimbardi AL, Souza FH, Masui DC, Silva TM, Jorge JA, Furriel RP. A novel β-glucosidase from *Humicola insolens* with high potential for untreated waste paper conversion to sugars. Appl Biochem Biotechnol 2014;173(2):391-408.

Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol. 2007;108:95-120.

Nakajima M, Yoshida R, Miyanaga A, Abe K, Takahashi Y, Sugimoto N, Toyoizumi H, Nakai H, Kitaoka M, Taguchi H. Functional and Structural Analysis of a β-glucosidase Involved in β-1, 2-Glucan Metabolism in *Listeria innocua*. PloS One 2016;11(2):e0148870.

Nakashima K, Watanabe H, Azuma J. Cellulase genes from the parabasalian symbiont

Pseudotrichonympha grassii in the hindgut of the wood-feeding termite *Coptotermes formosanus*. Cell Mol Life Sci. 2002a;59(9):1554-1560.

Nakashima K, Watanabe H, Saitoh H, Tokuda G, Azuma J. Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. Insect Biochem Mol Biol 2002b;32(7):777-784.

Nakatani Y, Cutfield SM, Cowieson NP, Cutfield JF. Structure and activity of exo-1, 3/1, 4-β-glucanase from marine bacterium *Pseudoalteromonas sp*. BB1 showing a novel C-terminal domain. FEBS Journal 2012;279(3):464-478.

Nakazawa H, Kawai T, Ida N, Shida Y, Kobayashi Y, Okada H, Tani S, Sumitani J, Kawaguchi T, Morikawa Y, Ogasawara W. Construction of a recombinant *Trichoderma reesei* strain expressing *Aspergillus aculeatus* β-glucosidase 1 for efficient biomass conversion. Biotechnol Bioeng 2012;109(1):92-99.

Nalepa CA, Maekawa K, Shimada K, Saito Y, Arellano C, Matsumoto T. Altricial development in subsocial wood-feeding cockroaches. Zool Sci 2008;25(12):1190-1198.

Nascimento CV, Souza FH, Masui DC, Leone FA, Peralta RM, Jorge JA, Furriel RP. Purification and biochemical properties of a glucose-stimulated β-D-glucosidase produced by *Humicola grisea* var. *thermoidea* grown on sugarcane bagasse. The Journal of Microbiology 2010;48(1):53-62.

Neurath H, Walsh KA. Role of proteolytic enzymes in biological regulation (a review). Proc Natl Acad Sci U S A 1976;73(11):3825-3832.

Ng IS, Tsai SW, Ju YM, Yu SM, Ho TH. Dynamic synergistic effect on *Trichoderma reesei* cellulases by novel β-glucosidases from *Taiwanese fungi*. Bioresour Technol 2011;102(10):6073-6081.

Nguyen NP, Lee KM, Lee KM, Kim IW, Kim YS, Jeya M, Lee JK. One-step purification and characterization of a β-1,4-glucosidase from a newly isolated strain of *Stereum hirsutum*. Appl Microbiol Biotechnol 2010;87(6):2107-2116.

Ni J, Takehara M, Watanabe H. Heterologous overexpression of a mutant termite cellulase gene in *Escherichia coli* by DNA shuffling of four orthologous parental cDNAs. Biosci Biotechnol Biochem 2005;69(9):1711-1720.

Ni J, Tokuda G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol Adv 2013;31(6):838-850.

Ni J, Tokuda G, Takehara M, Watanabe H. Heterologous expression and enzymatic characterization of β-glucosidase from the drywood-eating termite, *Neotermes koshunensis*. Appl Entomol Zool 2007;42(3):457-463.

O'Brien G, Veivers P, McEwen SE, Slaytor M, O'Brien R. The origin and distribution of cellulase in the termites, *Nasutitermes exitiosus* and *Coptotermes lacteus*. Insect

Biochemistry 1979;9(6):619-625.

Ohtoko K, Ohkuma M, Moriya S, Inoue T, Usami R, Kudo T. Diverse genes of cellulase homologues of glycosyl hydrolase family 45 from the symbiotic protists in the hindgut of the termite *Reticulitermes speratus*. Extremophiles 2000;4(6):343-349.

Oppert C, Klingeman WE, Willis JD, Oppert B, Jurat-Fuentes JL. Prospecting for cellulolytic activity in insect digestive fluids. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2010;155(2):145-154.

Otagiri M, Lopez CM, Kitamoto K, Arioka M, Kudo T, Moriya S. Heterologous expression and characterization of a glycoside hydrolase family 45 endo-β-1, 4-glucanase from a symbiotic protist of the lower termite, *Reticulitermes speratus*. Appl Biochem Biotechnol 2013;169(6):1910-1918.

Panday N, Canac Y, Vasella A. Very strong inhibition of glucosidases by C (2)-substituted tetrahydroimidazopyridines. Helv Chim Acta 2000;83(1):58-79.

Park AR, Hong JH, Kim JJ, Yoon JJ. Biochemical characterization of an extracellular β-glucosidase from the fungus, *Penicillium italicum*, isolated from rotten citrus peel. Mycobiology 2012;40(3):173-180.

Park AR, Kim HJ, Lee JK, Oh DK. Hydrolysis and transglycosylation activity of a thermostable recombinant β-glycosidase from *Sulfolobus acidocaldarius*. Appl Biochem

Biotechnol 2010;160(8):2236-2247.

Park TH, Choi KW, Park CS, Lee SB, Kang HY, Shon KJ, Park JS, Cha J. Substrate specificity and transglycosylation catalyzed by a thermostable β-glucosidase from marine hyperthermophile *Thermotoga neapolitana*. Appl Microbiol Biotechnol 2005;69(4):411-422.

Patchett ML, Daniel RM, Morgan HW. Purification and properties of a stable β-glucosidase from an extremely thermophilic anaerobic bacterium. Biochem J 1987;243(3):779-787.

Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Ståhlberg J, Beckham GT. Fungal cellulases. Chem Rev 2015;115(3):1308-1448.

Pei J, Pang Q, Zhao L, Fan S, Shi H. *Thermoanaerobacterium thermosaccharolyticum* β-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose. Biotechnol Biofuels. 2012;5(1):31.

Pei X, Zhao J, Cai P, Sun W, Ren J, Wu Q, Zhang S, Tian C. Heterologous expression of a GH3 β-glucosidase from *Neurospora crassa* in *Pichia pastoris* with high purity and its application in the hydrolysis of soybean isoflavone glycosides. Protein Expr Purif 2016;119:75-84.

Peplow M. Cellulosic ethanol fights for life. Nature 2014;507(7491):152-153.

Peralta RM, Terenzi HF, Jorge JA. β-D-Glycosidase activities of *Humicola grisea*: biochemical and kinetic characterization of a multifunctional enzyme. Biochim Biophys Acta. 1990;1033(3):243-249.

Pérez-Pons JA, Rebordosa X, Querol E. Properties of a novel glucose-enhanced β-glucosidase purified from *Streptomyces* sp. (ATCC 11238). Biochim Biophys Acta. 1995;1251(2):145-153.

Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods 2011;8(10):785-786.

Pozzo T, Pasten JL, Karlsson EN, Logan DT. Structural and functional analyses of β-glucosidase 3B from *Thermotoga neapolitana*: a thermostable three-domain representative of glycoside hydrolase 3. J Mol Biol 2010;397(3):724-739.

Prins R, Kreulen D. Comparative aspects of plant cell wall digestion in insects. Anim Feed Sci Technol 1991;32(1):101-118.

Ramani G, Meera B, Rajendhran J, Gunasekaran P. Transglycosylating glycoside hydrolase family 1 β-glucosidase from *Penicillium funiculosum* NCL1: Heterologous expression in *Escherichia coli* and characterization. Biochem Eng J 2015;102:6-13.

Reboredo FH, Lidon F, Pessoa F, Ramalho JC. The fall of oil prices and the effects on

biofuels. Trends Biotechnol 2016;34(1):3-6.

Reese ET, Siu RG, Levinson HS. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. J Bacteriol 1950;59(4):485-497.

Riou C, Salmon JM, Vallier MJ, Gunata Z, Barre P. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β-glucosidase from *Aspergillus oryzae*. Appl Environ Microbiol 1998;64(10):3607-3614.

Romanos MA, Scorer CA, Clare JJ. Foreign gene expression in yeast: a review. Yeast 1992;8(6):423-488.

Rouland C, Matoub M, Mora P, Petek F. Properties of two β-glucosidases purified from the termite *Macrotermes muelleri* and from its symbiotic fungus *Termitomyces* sp. Carbohydr Res 1992;233:273-278.

Rubin EM. Genomics of cellulosic biofuels. Nature 2008;454(7206):841-845.

Ryu DD, Mandels M. Cellulases: biosynthesis and applications. Enzyme Microb Technol 1980;2(2):91-102.

Saha BC, Bothast RJ. Production, purification, and characterization of a highly glucose-tolerant novel β-glucosidase from *Candida peltata*. Appl Environ Microbiol

1996;62(9):3165-3170.

Sakamoto R, Arai M, Murao S. Enzymic properties of three β-glucosidases from *Aspergillus aculeatus* No. F-50. Agric Biol Chem 1985a;49(5):1283-1290.

Sakamoto R, Kanamoto J, Arai M, Murao S. Purification and physicochemical properties of three β-glucosidases from *Aspergillus aculeatus* No. F-50. Agric Biol Chem 1985b;49(5):1275-1281.

Saloheimo M, Kuja-Panula J, Ylosmaki E, Ward M, Penttila M. Enzymatic properties and intracellular localization of the novel *Trichoderma reesei* β-glucosidase BGLII (cel1A). Appl Environ Microbiol 2002;68(9):4546-4553.

Sanchez OJ, Cardona CA. Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresour Technol 2008;99(13):5270-5295.

Sano K, Amemura A, Harada T. Purification and properties of a β-1, 6-glucosidase from *Flavobacterium*. Biochim Biophys Acta. 1975;377(2):410-420.

Sasaguri S, Maruyama J, Moriya S, Kudo T, Kitamoto K, Arioka M. Codon optimization prevents premature polyadenylation of heterologously-expressed cellulases from termite-gut symbionts in *Aspergillus oryzae*. J Gen Appl Microbiol 2008;54(6):343-351.

Scharf ME, Kovaleva ES, Jadhao S, Campbell JH, Buchman GW, Boucias DG. Functional and translational analyses of a beta-glucosidase gene (glycosyl hydrolase family 1) isolated from the gut of the lower termite *Reticulitermes flavipe*s. Insect Biochem Mol Biol 2010;40(8):611-620.

Schmid G, Wandrey C. Purification and partial characterization of a cellodextrin glucohydrolase (β-glucosidase) from *Trichoderma reesei* strain QM 9414. Biotechnol Bioeng 1987;30(4):571-585.

Schubert C. Can biofuels finally take center stage? Nat Biotechnol 2006;24(7):777-784.

Scrivener A, Slaytor M. Properties of the endogenous cellulase from *Panesthia cribrata* Saussure and purification of major endo-β-1, 4-glucanase components. Insect Biochem Mol Biol 1994;24(3):223-231.

Scrivener A, Slaytor M, Rose H. Symbiont-independent digestion of cellulose and starch in *Panesthia cribrata* Saussure, an Australian wood-eating cockroach. J Insect Physiol 1989;35(12):935-941.

Scrivener A, Zhao L, Slaytor M. Biochemical and immunological relationships between endo-β-1, 4-glucanases from cockroaches. Comp Biochem Physiol B Biochem Mol Biol 1997;118(4):837-843.

Service RF. Renewable energy. cellulosic ethanol at last? Science 2014;345(6201):1111.

Sethi A, Kovaleva ES, Slack JM, Brown S, Buchman GW, Scharf ME. A GHF7 cellulase from the protist symbiont community of *Reticulitermes flavipes* enables more efficient lignocellulose processing by host enzymes. Arch Insect Biochem Physiol 2013;84(4):175-193.

Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proc Natl Acad Sci U S A 2008;105(24):8256-8261.

Shewale JG. β-Glucosidase: its role in cellulase synthesis and hydrolysis of cellulose. Int J Biochem 1982;14(6):435-443.

Shewale JG, Sadana J. Purification, characterization, and properties of β-glucosidase enzymes from *Sclerotium rolfsii*. Arch Biochem Biophys 1981;207(1):185-196.

Shimada K, Maekawa K. Gene expression and molecular phylogenetic analyses of β-glucosidase in the termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae). J Insect Physiol 2014;65:63-69.

Sillam-Dussès D, Krasulová J, Vrkoslav V, Pytelková J, Cvačka J, Kutalová K, Bourguignon T, Miura T, Šobotník J. Comparative study of the labial gland secretion in termites (Isoptera). PloS one 2012;7(10):e46431.

Sinclair G, Choy FY. Synonymous codon usage bias and the expression of human

glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. Protein Expr Purif 2002;26(1):96-105.

Smaali I, Maugard T, Limam F, Legoy M, Marzouki N. Efficient synthesis of gluco-oligosaccharides and alkyl-glucosides by transglycosylation activity of β-glucosidase from *Sclerotinia sclerotiorum*. World J Microb Biot 2007;23(1):145-149.

Sørensen A, Lübeck M, Lübeck PS, Ahring BK. Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials. Biomolecules 2013;3(3):612-631.

Sørensen A, Lübeck PS, Lübeck M, Teller PJ, Ahring BK. β-Glucosidases from a new *Aspergillus* species can substitute commercial β-glucosidases for saccharification of lignocellulosic biomass. Can J Microbiol 2011;57(8):638-650.

Souza FHM, Nascimento CV, Rosa JC, Masui DC, Leone FA, Jorge JA, Furriel, RPM. Purification and biochemical characterization of a mycelial glucose-and xylose-stimulated β-glucosidase from the thermophilic fungus *Humicola insolens*. Process Biochem 2010;45(2):272-278.

Steiner DF, Clark JL. The spontaneous reoxidation of reduced beef and rat proinsulins. Proc Natl Acad Sci U S A 1968;60(2):622-629.

Sternberg D. β-glucosidase of *Trichoderma*: its biosynthesis and role in saccharification of cellulose. Appl Environ Microbiol 1976;31(5):648-654.

Sternberg D, Vuayakumar P, Reese E. β-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. Can J Microbiol 1977;23(2):139-147.

Suthangkornkul R, Sriworanun P, Nakai H, Okuyama M, Svasti J, Kimura A, Senapin S, Arthan D. A *Solanum torvum* GH3 β-glucosidase expressed in *Pichia pastoris* catalyzes the hydrolysis of furostanol glycoside. Phytochemistry. 2016;127:4-11.

Suzuki K, Sumitani J, Nam YW, Nishimaki T, Tani S, Wakagi T, Kawaguchi T, Fushinobu S. Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from *Aspergillus aculeatus*. Biochem J 2013;452(2):211-221.

Tamaki FK, Pimentel AC, Dias AB, Cardoso C, Ribeiro AF, Ferreira C, Terra WR. Physiology of digestion and the molecular characterization of the major digestive enzymes from *Periplaneta americana*. J Insect Physiol 2014;70:22-35.

Tangnu SK, Blanch HW, Wilke CR. Enhanced production of cellulase, hemicellulase, and β-glucosidase by *Trichoderma reesei* (Rut C-30). Biotechnol Bioeng 1981;23(8):1837-1849.

Terra W, Ferreira C, Jordao B, Dillon R. Digestive enzymes. Biology of the insect midgut: Springer; 1996. p. 153-194.

Teeri TT. Crystalline cellulose degradation: new insight into the function of

cellobiohydrolases. Trends Biotechnol 1997;15(5):160-167.

Thongpoo P, McKee LS, Araújo AC, Kongsaeree PT, Brumer H. Identification of the acid/base catalyst of a glycoside hydrolase family 3 (GH3) β-glucosidase from *Aspergillus niger* ASKU28. Biochim Biophys Acta. 2013;1830(3):2739-2749.

Todaka N, Lopez CM, Inoue T, Saita K, Maruyama J, Arioka M, Kitamoto K, Kudo T, Moriya S. Heterologous expression and characterization of an endoglucanase from a symbiotic protist of the lower termite, *Reticulitermes speratus*. Appl Biochem Biotechnol 2010;160(4):1168-1178.

Todaka N, Moriya S, Saita K, Hondo T, Kiuchi I, Takasu H, Ohkuma M, Piero C, Hayashizaki Y, Kudo T. Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. FEMS Microbiol Ecol 2007;59(3):592-599.

Todaka N, Nakamura R, Moriya S, Ohkuma M, Kudo T, Takahashi H, Ishida N. Screening of optimal cellulases from symbiotic protists of termites through expression in the secretory pathway of *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 2011;75(11):2260-2263.

Tokuda G, Lo N, Watanabe H, Slaytor M, Matsumoto T, Noda H. Metazoan cellulase genes from termites: intron/exon structures and sites of expression. Biochim. Biophys. Acta 1999;1447(2):146-159.

Tokuda G, Miyagi M, Makiya H, Watanabe H, Arakawa G. Digestive β-glucosidases from the wood-feeding higher termite, *Nasutitermes takasagoensis:* intestinal distribution, molecular characterization, and alteration in sites of expression. Insect Biochem Mol Biol 2009;39(12):931-937.

Tokuda G, Saito H, Watanabe H. A digestive β-glucosidase from the salivary glands of the termite, *Neotermes koshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5′-and 3′-RACE amplifications with degenerate primers. Insect Biochem Mol Biol 2002;32(12):1681-1689.

Tokuda G, Tsuboi Y, Kihara K, Saitou S, Moriya S, Lo N, Kikuchi J. Metabolomic profiling of ^{13}C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. Proc Biol Sci 2014;281(1789):20140990.

Tokuda G, Watanabe H. Hidden cellulases in termites: revision of an old hypothesis. Biol Lett 2007;3(3):336-339.

Tokuda G, Watanabe H, Hojo M, Fujita A, Makiya H, Miyagi M, Arakawa G, Arioka M. Cellulolytic environment in the midgut of the wood-feeding higher termite *Nasutitermes takasagoensis*. J Insect Physiol 2012;58(1):147-154.

Treebupachatsakul T, Nakazawa H, Shinbo H, Fujikawa H, Nagaiwa A, Ochiai N, Kawaguchi T, Nikaido M, Totani K, Shioya K, Shida Y, Morikawa Y, Ogasawara W,

Okada H. Heterologously expressed *Aspergillus aculeatus* β-glucosidase in *Saccharomyces cerevisiae* is a cost-effective alternative to commercial supplementation of β-glucosidase in industrial ethanol production using *Trichoderma reesei* cellulases. J Biosci Bioeng 2016;121(1):27-35.

Trimble RB, Atkinson PH, Tschopp JF, Townsend RR, Maley F. Structure of oligosaccharides on *Saccharomyces* SUC2 invertase secreted by the methylotrophic yeast *Pichia pastoris*. J Biol Chem 1991;266(34):22807-22817.

Trofimov AA, Polyakov KM, Tikhonov AV, Bezsudnova EY, Dorovatovskii PV, Gumerov VM, Ravin NV, Skryabin KG, Popov VO. Structures of β-glycosidase from *Acidilobus saccharovorans* in complexes with tris and glycerol. Dokl Biochem Biophys 2013;449:99-101.

Tsukada T, Igarashi K, Yoshida M, Samejima M. Molecular cloning and characterization of two intracellular β-glucosidases belonging to glycoside hydrolase family 1 from the basidiomycete *Phanerochaete chrysosporium*. Appl Microbiol Biotechnol 2006;73(4):807-814.

Tull D, Gottschalk TE, Svendsen I, Kramhøft B, Phillipson BA, Bisgård-Frantzen H, Olsen O, Svensson B. Extensive *N*-glycosylation reduces the thermal stability of a recombinant alkalophilic *Bacillus* α-amylase produced in *Pichia pastoris*. Protein Expr Purif 2001;21(1):13-23.

Uchima CA, Arioka M. Expression and one-step purification of recombinant proteins using an alternative episomal vector for the expression of N-tagged heterologous proteins in *Pichia pastoris*. Biosci Biotechnol Biochem 2012;76(2):368-371.

Uchima CA, Tokuda G, Watanabe H, Kitamoto K, Arioka M. Heterologous expression and characterization of a glucose-stimulated β-glucosidase from the termite *Neotermes koshunensis* in *Aspergillus oryzae*. Appl Microbiol Biotechnol 2011;89(6):1761-1771.

Uchima CA, Tokuda G, Watanabe H, Kitamoto K, Arioka M. Heterologous expression in *Pichia pastoris* and characterization of an endogenous thermostable and high-glucose-tolerant β-glucosidase from the termite *Nasutitermes takasagoensis*. Appl Environ Microbiol 2012;78(12):4288-4293.

Uchiyama T, Miyazaki K, Yaoi K. Characterization of a novel β-glucosidase from a compost microbial metagenome with strong transglycosylation activity. J Biol Chem 2013;288(25):18325-18334

Uchiyama T, Yaoi K, Miyazaki K. Glucose-tolerant β-glucosidase retrieved from a Kusaya gravy metagenome. Frontiers in microbiology 2015;6:548.

Umezurike G. Kinetic analysis of the mechanism of action of β-glucosidase from *Botryodiplodia theobromae* Pat. Biochim Biophys Acta. 1975;397(1):164-178.

Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VG. An

oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science 2010;330(6001):219-222.

Vaheri M, Leisola M, Kauppinen V. Transglycosylation products of cellulase system of *Trichoderma reesei*. Biotechnol Lett 1979;1(1):41-46.

Valls LA, Hunter CP, Rothman JH, Stevens TH. Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide. Cell 1987;48(5):887-897.

Vanin EF. Processed pseudogenes: characteristics and evolution. Annu Rev Genet 1985;19(1):253-272.

Varghese JN, Hrmova M, Fincher GB. Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. Structure 1999;7(2):179-190.

Venturi LL, Polizeli Mde L, Terenzi HF, Furriel Rdos P, Jorge JA. Extracellular β-D-glucosidase from *Chaetomium thermophilum* var. *coprophilum*: production, purification and some biochemical properties. J Basic Microbiol 2002;42(1):55-66.

Verdoucq L, Czjzek M, Moriniere J, Bevan DR, Esen A. Mutational and structural analysis of aglycone specificity in maize and sorghum β-glucosidases. J Biol Chem 2003;278(27):25055-25062.

Verdoucq L, Morinière J, Bevan DR, Esen A, Vasella A, Henrissat B, Czjze M. Structural determinants of substrate specificity in family 1 β-glucosidases: novel insights from the crystal structure of sorghum dhurrinase-1, a plant β-glucosidase with strict specificity, in complex with its natural substrate. J Biol Chem 2004;279(30):31796-31803.

Walker DE, Axelrod B. Evidence for a single catalytic site on the "β-D-glucosidase-β-D-galactosidase" of almond emulsin. Arch Biochem Biophys 1978;187(1):102-107.

Wang Q, Qian C, Zhang XZ, Liu N, Yan X, Zhou Z. Characterization of a novel thermostable β-glucosidase from a metagenomic library of termite gut. Enzyme Microb Technol 2012;51(6-7):319-324.

Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 2007;450(7169):560-565.

Watanabe A, Suzuki M, Ujiie S, Gomi K. Purification and enzymatic characterization of

a novel β-1, 6-glucosidase from *Aspergillus oryzae*. J Biosci Bioeng. 2016;121(3):259-264.

Watanabe H, Nakamura M, Tokuda G, Yamaoka I, Scrivener AM, Noda H. Site of secretion and properties of endogenous endo-β-1, 4-glucanase components from *Reticulitermes speratus* (Kolbe), a Japanese subterranean termite. Insect Biochem Mol Biol 1997;27(4):305-313.

Watanabe H, Nakashima K, Saito H, Slaytor M. New endo-β-1,4-glucanases from the parabasalian symbionts, *Pseudotrichonympha grassii* and *Holomastigotoides mirabile* of *Coptotermes* termites. Cell Mol Life Sci. 2002;59(11):1983-1992.

Watanabe H, Noda H, Tokuda G, Lo N. A cellulase gene of termite origin. Nature 1998;394(6691):330-331.

Watanabe H, Tokuda G. Cellulolytic systems in insects. Annu Rev Entomol 2010;55:609-632.

White A, Rose DR. Mechanism of catalysis by retaining β-glycosyl hydrolases. Curr Opin Struct Biol 1997;7(5):645-651.

White A, Tull D, Johns K, Withers SG, Rose DR. Crystallographic observation of a covalent catalytic intermediate in a β-glycosidase. Nat Struct Biol. 1996;3(2):149-154.

Withers SG, Rupitz K, Trimbur D, Warren RA. Mechanistic consequences of mutation of the active site nucleophile Glu 358 in *Agrobacterium* β-glucosidase. Biochemistry 1992;31(41):9979-9985.

Wood TM, McCrae SI. The cellulase complex of *Trichoderma koningii In* Symposium on enzymatic hydrolysis of cellulose. *Edited by* Bailey M, Enari TM, and Linko M. SITRA. Helsinki, Finland. 1975:231-254

Woodward J, Wiseman A. Fungal and other β-D-glucosidases--their properties and applications. Enzyme Microb Technol 1982;4(2):73-79.

Wu Y, Chi S, Yun C, Shen Y, Tokuda G, Ni J. Molecular cloning and characterization of an endogenous digestive β-glucosidase from the midgut of the fungus-growing termite *Macrotermes barneyi*. Insect Mol Biol 2012;21(6):604-614

Wyman CE. Biomass ethanol: technical progress, opportunities, and commercial challenges. Annu Rev Energy Environ 1999;24(1):189-226.

Wyman CE. Ethanol from lignocellulosic biomass: technology, economics, and opportunities. Bioresour Technol 1994;50(1):3-15.

Xu Z, Shih M, Poulton JE. An extracellular exo-β-(1, 3)-glucanase from *Pichia pastoris*: purification, characterization, molecular cloning, and functional expression. Protein Expr Purif 2006;47(1):118-127.

Yamin MA. Axenic cultivation of the cellulolytic flagellate *Trichomitopsis termopsidis* (Cleveland) from the termite *Zootermopsis*. J Protozool 1978;25(4):535-538.

Yan TR, Lin CL. Purification and characterization of a glucose-tolerant β-glucosidase from *Aspergillus niger* CCRC 31494. Biosci Biotechnol Biochem 1997;61(6):965-970.

Yan TR, Lin YH, Lin CL. Purification and characterization of an extracellular β-glucosidase II with high hydrolysis and transglucosylation activities from *Aspergillus niger*. J Agric Food Chem 1998;46(2):431-437.

Yang F, Yang X, Li Z, Du C, Wang J, Li S. Overexpression and characterization of a glucose-tolerant β-glucosidase from *T. aotearoense* with high specific activity for cellobiose. Appl Microbiol Biotechnol. 2015;99(21):8903-8915.

Yang M, Luoh SM, Goddard A, Reilly D, Henzel W, Bass S. The bglX gene located at 47.8 min on the *Escherichia coli* chromosome encodes a periplasmic β-glucosidase. Microbiology. 1996;142 (Pt 7):1659-1665.

Yang S, Hua C, Yan Q, Li Y, Jiang Z. Biochemical properties of a novel glycoside hydrolase family 1 β-glucosidase (PtBglu1) from *Paecilomyces thermophila* expressed in *Pichia pastoris*. Carbohydr Polym 2013;92(1):784-791

Yang S, Jiang Z, Yan Q, Zhu H. Characterization of a thermostable extracellular

β-glucosidase with activities of exoglucanase and transglycosylation from *Paecilomyces thermophila*. J Agric Food Chem. 2008;56(2):602-608.

Yang X, Ma R, Shi P, Huang H, Bai Y, Wang Y, Yang P, Fan Y, Yao B. Molecular characterization of a highly-active thermophilic β-glucosidase from *Neosartorya fischeri* P1 and its application in the hydrolysis of soybean isoflavone glycosides. PloS one 2014;9(9):e106785.

Yoshida E, Hidaka M, Fushinobu S, Koyanagi T, Minami H, Tamaki H, Kitaoka M, Katayama T, Kumagai H. Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. Biochem J 2010;431:39-49.

Yuki M, Moriya S, Inoue T, Kudo T. Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its hindgut. Zool Sci 2008;25(4):401-406.

Zanoelo FF, Polizeli Mde L, Terenzi HF, Jorge JA. β-glucosidase activity from the thermophilic fungus *Scytalidium thermophilum* is stimulated by glucose and xylose. FEMS Microbiol Lett 2004;240(2):137-143.

Zhang D, Allen AB, Lax AR. Functional analyses of the digestive β-glucosidase of Formosan subterranean termites (*Coptotermes formosanus*). J Insect Physiol 2012;58(1):205-210.

Zhang D, Lax AR, Raina AK, Bland JM. Differential cellulolytic activity of native-form and C-terminal tagged-form cellulase derived from *Coptotermes formosanus* and expressed in *E. coli*. Insect Biochem Mol Biol 2009;39(8):516-522.

Zhang W, Kang L, Yang M, Zhou Y, Wang J, Liu Z, Yuan S. Purification, characterization, and function analysis of an extracellular β-glucosidase from elongating stipe cell walls in *Coprinopsis cinerea*. FEMS Microbiol Lett. 2016;363(9).

Zhao J, Guo C, Tian C, Ma Y. Heterologous expression and characterization of a GH3 β-glucosidase from thermophilic fungi *Myceliophthora thermophila* in *Pichia pastoris*. Appl Biochem Biotechnol 2015;177(2):511-527.

Zhao L, Xie J, Zhang X, Cao F, Pei J. Overexpression and characterization of a glucose-tolerant β-glucosidase from *Thermotoga thermarum* DSM 5069T with high catalytic efficiency of ginsenoside Rb1 to Rd. J Mol Catal B Enzym 2013;95:62-69.

Zhao W, Peng R, Xiong A, Fu X, Tian Y, Yao Q. Expression and characterization of a cold-active and xylose-stimulated β-glucosidase from *Marinomonas* MWYL1 in *Escherichia coli*. Mol Biol Rep 2012;39(3):2937-2943

Zhou J, Hartmann S, Shepherd BK, Poulton JE. Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. Plant Physiol 2002;129(3):1252-1264.

Zhou L, Li S, Zhang T, Mu W, Jiang B. Properties of a novel polydatin-β-D-glucosidase from *Aspergillus niger* SK34.002 and its application in enzymatic preparation of resveratrol. J Sci Food Agric. 2016;96(7):2588-2595.

Zhou X, Kovaleva ES, Wu-Scharf D, Campbell JH, Buchman GW, Boucias DG, Scharf ME. Production and characterization of a recombinant beta-1, 4-endoglucanase (glycohydrolase family 9) from the termite *Reticulitermes flavipes*. Arch Insect Biochem Physiol 2010;74(3):147-162.

Zmudka MW, Thoden JB, Holden HM. The structure of DesR from *Streptomyces venezuelae*, a β-glucosidase involved in macrolide activation. Protein Sci 2013;22(7):883-892.

Zouhar J, Vevodova J, Marek J, Damborsky J, Su XD, Brzobohaty B. Insights into the functional architecture of the catalytic center of a maize β-glucosidase Zm-p60.1. Plant Physiol 2001;127(3):973-985.

Zverlov VV, Volkov IY, Velikodvorskaya TV, Schwarz WH. *Thermotoga neapolitana bglB* gene, upstream of *lamA*, encodes a highly thermostable β-glucosidase that is a laminaribiase. Microbiology. 1997;143 (Pt 11):3537-3542.

Company manuals

Amersham Biosciences, HisTrap HP, 1 ml and 5 ml Instructions, 71-5027-68 Edition

AB 2004:1-6

Invitrogen, User manual of EasySelect™ *Pichia* Expression Kit for expression of recombinant proteins using pPICZ and pPICZα in Pichia pastoris. Invitrogen. 2010

Invitrogen, Ni-NTA Purification system for purification of polyhistidine-containing recombinant proteins, Version C, 2006:4

QIAGEN, A handbook for high-level expression and purification of 6xHis-tagged proteins, fifth edition. 2003:18-74.

論文の内容の要旨

応用生命工学専攻
平成 24 年度博士課程入学
氏　　名　李　宜海
指導教員名　有岡　学

論文題目

A study on β-glucosidases derived from wood-feeding insect and symbiotic protist of termite
（材食性昆虫およびシロアリ腸内原生生物に由来するβ-グルコシダーゼに関する研究）

Introduction

Wood-feeding insects and termites evolved from the same ancestor, and share a similar digestive system with high efficiency for decomposition of lignocellulose. These organisms, together with their symbiotic protists which live in their guts and contribute to entire degradation of cellulose ingested by their endosymbiosis host, could be new sources of cellulolytic enzymes for development of the second generation biofuel derived from lignocellulosic biomass.

Cellulose is a linear polymer consisting of β-1,4-linked D-glucopyranosyl unit and contributes to about 20–40% of dry weight of the plant primary cell walls. In the sequential degradation of cellulose by cellulases, endoglucanase (EG; endo-1,4-β-glucanase; EC 3.2.1.4) firstly cleaves the cellulose chains in its amorphous regions, then cellobiohydrolase (CBH; exo-1,4-β-glucanase; EC 3.2.1.91) attaches to cellulose chain ends and hydrolyzes cellulose chains typically into disaccharide units (cellobiose), and β-glucosidase (BG) completes the final step by cleaving cellobiose into glucose. The recently-discovered lytic polysaccharide monooxygenases (LPMOs) which cleave the crystalline regions of cellulose possibly have synergistic effects with cellulases. BG can alleviates the product inhibition of EG and CBH by cellobiose, whereas most BGs suffer product inhibition by glucose. Commercial cellulases are mostly from filamentous fungi *Trichoderma*, especially *T. reesei*. However, *Trichoderma* spp. are characterized to have low yield of BG and thus high titers of BG supplements are required in the commercial cellulase cocktails. Therefore, screening of new powerful BG is of great importance to improve the efficiency of commercial bioethanol production from cellulose.

BG is an enzyme that hydrolyzes the terminal, non-reducing β-D-glucosyl residues with release of β-D-glucose. The activities of BGs are not limited to the β-1,4 glucosidic linkages, but also apply to β-1,2, β-1,3, and β-1,6 linkages. Furthermore, many of BGs are reported to be associated with the activities of β-D-galactosidase, α-L-arabinosidase, β-D-xylosidase, and β-D-fucosidase. On the basis of the homology of amino acid sequences, BGs are classified into the glycoside hydrolase (GH) families 1, 3, 5, 9, 30, and 116 in the Carbohydrate-Active enZYmes database (CAZy). In terms of glycosyl-transferring types, BGs are classified into 2 groups, based on whether they invert or retain the configuration of the anomeric carbon (C1). Aside for BGs in GH9, all BGs belong to the retaining type, in which the reaction is

mediated by two carboxyl groups in the catalytic center, with one acting as a nucleophile and the other an acid/base catalysts, respectively, and perform catalysis via a double-displacement mechanism (i.e. glycosylation and deglycosylation). In the GH1 BG, both the catalytic nucleophile and acid/base residues are glutamic acids, while in the GH3, aspartic acid plays a role as a catalytic nucleophile. In this study, PaBG1b, a powerful GH1 BG from the gregarious wood-feeding cockroach *Panesthia angustipennis spadica*, and RsBG, a GH3 β-glucosidase-like protein from the hindgut symbiotic protist of the lower termite *Reticulitermes speratus*, were heterologously expressed in the methylotrophic yeast *Pichia pastoris*, purified, and biochemically characterized.

Chapter 1. Heterologous expression of PaBG1b in *P. pastoris*

The full-length cDNA of *pabg1b* encodes a putative GH1 BG of 502 amino acids with a KR dibasic amino acid pair located at 7 amino acids downstream from the amino terminus of the mature polypeptide. The expression of recombinant PaBG1b was initially conducted by using the episomal expression plasmid pBGP3, which harbors the autonomous replication sequence and the constitutive glyceraldehyde 3-phosphate dehydrogenase (*GAP*) promoter. PaBG1b was designed to be fused with c-myc-epitope and hexahistidine (His_6) tags at its N-terminus, which allows detection of the recombinant protein by Western blot analysis, and purification via Ni-NTA column chromatography. The expression of PaBG1b was evidenced by Western blot analysis after 4 days of culture. BG activity assay also confirmed that PaBG1b was successfully expressed in the culture supernatant of pBGP3-PaBG1b transformant. The immunoblot bands of PaBG1b was found to fade out from the fifth day of culture, suggesting that PaBG1b was susceptible to proteolytic degradation and/or lost its N-terminal tags by the Kex2 endoprotease cleavage at the KR site.

To increase the production level of PaBG1b, an integration-type plasmid, pPICZα, in which the expression of heterologous protein is achieved by the inducible alcohol oxidase 1 (*AOX1*) promoter, was employed. In this case, the tags were moved to the C-terminus. Time-course analysis showed that the expression of PaBG1b was induced on the first day of induction by methanol and the yield of PaBG1b in the culture supernatant reached the highest after 6 days of induction. Western blot analysis showed that the bands of PaBG1b remained clear and intense even the culture was extended to 7 days, which implied that the protein tags at the C-terminal were stable.

Chapter 2. Purification and biochemical characterization of PaBG1b

PaBG1b in the culture supernatant was purified though three successive steps including ammonium sulfate precipitation, followed by Ni-NTA and anion exchange chromatographies. Ni-NTA column combined with imidazole elution is a simple and quick approach for purification of target proteins fused with His_6 tag. However, imidazole was found to strongly inhibit the BG activity of PaBG1b such that the inhibition could hardly be reversed either through dialysis or buffer exchange. Fortunately, substitution of sodium phosphate buffer with Tris (2-Amino-2-hydroxymethyl-propane-1,3-diol) buffer and instantly removing imidazole through dialysis successfully recovered the activity of purified product, although Tris

itself was also an competitive inhibitor of PaBG1b as described below.

To clarify the molecular nature of PaBG1b, deglycosylation analysis was conducted by treating purified PaBG1b with glycopeptidase F (GPF) or endoglycosidase H (Endo H). SDS-PAGE analysis showed that PaBG1b was *N*-glycosylated and two closely-migrating bands were present. N-terminal amino acid sequence analysis suggested that incomplete cleavage at the KR dibasic amino acid pair in the mature region of PaBG1b by Kex2 generated two different N-termini.

PaBG1b demonstrated the maximum specific activity of 45.5 U/mg and V_{max} of 59.9 U/mg with *p*-nitrophenyl β-D-glucopyranoside (*p*NPG) as a substrate, whereas it displayed high specific activity of 338.5 U/mg and V_{max} of 436.7 U/mg towards cellobiose. PaBG1b was also found to have relatively high catalytic efficiency with k_{cat}/K_m of 109.8 $mM^{-1} \cdot s^{-1}$ for cellobiose. Thermostability and pH stability analyses demonstrated that PaBG1b is a mesophilic enzyme with an optimum temperature of 45°C and favors slightly acidic condition (optimum pH at pH5.5). PaBG1b was capable of degrading diverse saccharides and aryl-glycosides. PaBG1b displayed the ability of (but might be not limit to) degrading cello-oligosaccharides up to cellohexaose, and did not show transglycosylation activity under the condition tested. PaBG1b has a moderate tolerance to glucose (K_i=200.3 mM), and was not inhibited by cellobiose up to 100 mM. These properties endow PaBG1b with advantages in terms of conversion of cellulose to glucose. Although in general metal cations are not required for the catalytic activity of BGs, Al^{3+} and Cu^{2+} were found to stimulate the activity of PaBG1b, whereas Zn^{2+}, Ni^{2+}, Fe^{3+}, and Fe^{2+} significantly inhibited the activity. To elucidate the interactions of PaBG1b with inhibitors, imidazole and Tris, the inhibition constants (K_is) at pH 5.5 were analyzed, and determined as 4.3 mM and 5.9 mM, respectively. These results indicate that both imidazole and Tris are modest competitive inhibitors for PaBG1b with comparable affinity. Inactivation of PaBG1b under standard Ni-NTA affinity chromatography condition was suggested to be due to a specific binding of imidazole with PaBG1b under alkaline pH. Although PaBG1b displayed high activity towards cellobiose, the activity is approximately half of the native enzyme. One of the possible reasons was that the expression products were comprised of two kinds of polypeptides with different N-termini.

Chapter 3. Heterologous expression in *P. pastoris* and purification of RsBG

The full-length cDNA of *rsbg* encodes a polypeptide of 573 amino acids which is composed of two domains: the BglX domain (periplasmic BG and related glycosidases) and the GH3 C-terminal domain. However, the putative catalytic nucleophile residue of RsBG is glutamic acid (E275) rather than aspartic acid which is almost completely conserved in the canonical GH3 BG orthologues. Although BGs in GH1 employ two glutamic acids as the nucleophile and acid/base catalysts, the orientation is reversed to that of GH3 BGs in that the nucleophile is located C-terminal to the acid/base residue in GH1. Furthermore, unlike most GH1 BGs that consist of only one common N-terminal $(\beta/\alpha)_8$ TIM barrel domain, all of nine structurally-characterized GH3 β-glucosidases are multidomain proteins. Thus RsBG shares part of the features of both GH3 and GH1 BGs, but to be precise, it belongs to neither of them.

To elucidate the biochemical properties of RsBG, pBGP3-based expression system was employed.

Expression of RsBG was confirmed by Western blot analysis after 3 to 4 days of culture. BG assay demonstrated, however, that the activity of RsBG-transformant was only a little higher than the control strain harboring the empty vector. Furthermore, RsBG purified by Ni-NTA column chromatography showed no activity. To increase the production level of RsBG, pPICZα was employed to produce N-terminally-tagged RsBG. Although the culture supernatant of RsBG transformant exhibited slightly higher activity than that of the control strain, again the purified RsBG did not exhibit any activity towards either natural or artificial substrates of BG. Considering the possibility that Ni-NTA purification caused inactivation of RsBG, traditional purification techniques were employed. RsBG was purified through ammonium sulfate precipitation followed anion exchange chromatography, but the purified RsBG did not display any activity. RsBG without tags was also expressed, but the assay result showed that the BG activity of culture supernatant of the transformant was similar to that of the transformant expressing tagged RsBG. These results demonstrate that RsBG does not function as a BG. Finally, to examine if the presence of an extra methylene group in the side chain of putative catalytic nucleophile, i.e. glutamic acid (E275) of RsBG, compared to aspartic acid in the counterpart of other GH3 BGs, hindered the access of the substrates, a mutation study was conducted to substitute glutamic acid with aspartic acid (E275D mutation). However, the expression of RsBG E275D mutant failed and the product was not observed presumably due to degradation.

Conclusion

In this study, two proteins from different organisms were successfully expressed in *P. pastoris* by using various expression settings. Biochemical characterization of PaBG1b revealed features potentially suitable for commercial purposes. Both imidazole and Tris were shown to be competitive inhibitors with the similar level of affinity towards PaBG1b at pH 5.5. However, they exhibited different inhibition potency in the pH shifting experiments. As the inhibition by Tris was easily recovered through dilution and restoring the pH to acidic pH, the inhibition by imidazole, at least when applied at pH 8.0, was hardly recovered. It might be attributable to a specific binding of imidazole and PaBG1b at alkaline pH. Intensive studies of PaBG1b including the structural analysis would be informative to clarify the key residues important for the catalytic function, substrate specificity, thermostability, and glucose tolerance, etc.

RsBG was successfully expressed in *P. pastoris* and the culture supernatant of transformants displayed limited but consistently higher activity compared to that of the control stain. However, no activity was detected in the purified RsBG preparation. These results might imply that the precise role of RsBG is not a function as a BG, but might be a product of a pseudogene gene, or a protein capable of boosting the activities of other BGs. Further studies are surely needed to elucidate the unexpected and unusual function(s) of RsBG distinct from those of canonical GH3 BGs.

Acknowledgement

First and foremost I want to thank my first supervisor Professor Katsuhiko Kitamoto who accepted me as a graduate student at the Laboratory of Microbiology. And I would like to extend my grateful regards to Mrs. Kitamoto that was always very kind to receive me in their house. It is a pleasure to thank Dr. Manabu Arioka who is my second supervisor and guided my PhD course. Thanks for his patience, corrections, and guidance during all my studies in this lab. I thank Dr. Jun-ichi Maruyama for taking care of me when I was in Obeya. Although I only stay there for only one year, but it was the most joyful days I have in University of Tokyo.

Acknowledgments go to Drs. Watanabe, Tokuda, Arakawa, Moriya for providing the cDNAs of PaBG1b and RsBG. Also Dr. Mayumi Kuroiwa of the Technology Advancement Center, Faculty of Agriculture, The University of Tokyo, for amino acid sequencing. I also thank Professor Shin-ya Fushinobu and Dr. Kiyohiko Igarashi for their precious advices in composing of my journal paper.

I am grateful to Ms. Cristiane Akemi Uchima, who taught me experimental skills at the initial stage of my study. I also highly appreciate the help of my tutor, Ms. Hikaru Tsukagoshi, whose talent, industriousness and luck deeply impressed me and always to be my model in research work. Furthermore, I would like to thank Ms. Hsiang-Ting Huang, and Ms. Yun-han Hsu, whose constructive suggestions enlightened me and led to two breakthroughs in this study. I also highly appreciate the help and moral support of Ms. Jin Fengjie, Ms. Wuhanqimuge, Ms. Zhu Lin, Ms. Han Pei, Ms. Yun-han Hsu, Mr. Hidetoshi Nakamura, Mr. Bi Chun, and all members in the Laboratory of Microbiology. It is a great experience to get acquantence with them and experience hard days together. I will cherish our friendship and keep the good memories with them in my heart forever.

I want to extend my thankfulness to the China Scholarship Committee (CSC) and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan for the scholarship received during the three and half years I have been studying in Japan. With this financial support I could fully concentrate on my studies.

Finally, I would like to deliver my greatest and deepest gratitude to my family in China for their enormous support during my study in Japan. Although they could not understand what I was studing, they trust me, comfort me and encourage me me all the time. Their unconditional love always fulfilled in my heart and inspired me to bear any burden and meet any hardship.

李宜海

Dec6, 2016