Graduate School of Bioresources
Mie University

Ph. D. Thesis

Characterization of plant cell wall degrading enzymes from *Paenibacillus* sp.

(*Paenibacillus* 属の植物細胞壁多糖分解酵素の特性)

Daichi Ito

March, 2023

CONTENTS

Abbreviation

AA: Auxiliary Activity

BMC: Ball-milled cellulose

BSA: Bovine serum albumin

BWX: Birch wood xylan

CBM: Carbohydrate-binding module

CBH: Cellobiohydrolase

CD: Catalytic Domain

CMC: Carboxymethyl cellulose

DMP: 2,6-dimethoxyphenol

EC number: Enzyme Commission number

Fn III: Fibronectin type III module

GH: Glycoside hydrolase

HRP: Horseradish peroxidase

IPTG: Isopropyl-1-thio-β-D-galactopyranoside

KPB: Potassium phosphate buffer

PAGE: Polyacrylamide gel electrophoresis

PCR: Polymerase chain reaction

*Px*Xyl43A: Family 43 glycoside hydrolase from *Paenibacillus xylaniclasticu*s strain TW1

*Px*Xyl43A-UM: Family 91 carbohydrate-binding module from *Paenibacillus xylaniclasticu*s strain TW1

*Px*AA10A: Family 10 auxiliary activity enzyme from *Paenibacillus xylaniclasticu*s strain TW1

*Px*AA10A-CD: Catalytic domain of Family 10 auxiliary activity enzyme from *Paenibacillus xylaniclastic*us strain TW1

SDS: Sodium dodecyl sulfate

*p*NP: *p*-nitrophenol

*p*NPAf: *p*-nitrophenyl-α-L-arabinofuranoside

*p*NPG: *p*-nitropheny-β-D-glucopyranoside

*p*NPGal: *p*-nitropheny-β-D-galactopyranoside

*p*NPX: *p*-nitropheny-β-D-xylopyranoside

G1: glucose

G2: cellobiose

G3: cellotriose

G4: cellotetraose

G5: cellopentaose

G6: cellohexaose

X2: xylobiose

X3: xylotriose

X4: xylotetraose

X5: xylopentaose

X6: xylohexaose

Chapter 1
Introduction

1-1. General Introduction

Cellulose is the main structural polysaccharide in plant cell walls, and also the main constituent of the biomass. Cellulose is attracting attention as a renewable energy source, because it can be converted into bioethanol and other chemicals. Currently, corn and sugarcane are used as materials for bioethanol production, in addition to the inedible parts of rice straw and other agricultural products (Hamelinck *et al*. 2005).

To convert the cellulosic biomass into chemicals such as bioethanol, the biomass is first broken down into monosaccharides and oligosaccharides, and then these sugars are fermented to produce bioethanol and other alternative fuels (Tan *et al*. 2016). However, while the cellulosic biomass has the advantage that it is abundant, its complex structure makes it difficult to break down into the requisite sugars (Hamelinck *et al*. 2005).

To overcome this problem, microorganisms that degrade the cellulosic biomass in natural settings and microbially derived carbohydrate active enzymes are often used to break down the cellulosic biomass. *Trichoderma reesei* (also called *Hypocrea jecorina*) and *Clostridium thermocellum* (also called *Acetivibrio thermocellus*) are two enzyme-producing microorganisms often used for this purpose (Tindall *et al*. 2019).

T. reesei is a filamentous fungus and is known as a high cellulase producer (Peterson *et al*. 2012; Bischof *et al*. 2016). *C. thermocellum*, an anaerobic, thermophilic bacterium, exhibits a high capacity for biomass-degradation through the formation of extracellular enzyme complexes, cellulosomes, that contain cellulases and xylanases (Lamed *et al*. 1983). *Paenibacillus* spp. are also notable cellulolytic bacteria; some of these species

hydrolyze cellulose by producing cellulases while others form extracellular enzyme complexes (Ratanakhanockchai *et al*. 2012; Tachaapaikoon *et al*. 2012).

1-2. Structure of the Plant Biomass

The cellulosic biomass is the most abundant biomass and constitutes the cell wall of plant cells. Its structure is complex, consisting of cellulose intertwined with hemicellulose and lignin. Cellulose is a polymer of D-glucose polymerized by β-1,4 bonds. The cellulose chains consist of 5,000 to 14,000 polymerized glucose units. In addition, 36 cellulose chains are hydrogen bonded in regular rows to form the crystalline structure of microfibers. Therefore, cellulose is insoluble in water and requires high pressure and high temperature conditions to decompose its crystalline structure (Robak *et al*. 2018).

Hemicellulose has a branched structure that is shorter than cellulose chains but modified by many sugars. It is also composed of several types of sugar units. Among them, the main component is xylan. Xylan is a polymer of D-xylose polymerized by β-1,4 glycoside bonds. Xylan is the main chain, which is modified with L-arabinose, D-glucose, D-mannose and D-galactose, and the acidic sugars D-glucuronic acid and D-galacturonic acid as side chains (Robak *et al*. 2018).

Lignin is an irregularly polymerized structure of phenylpropanoids. Among these phenylpropanoids, ferulic acid and *p*-coumaric acid are mainly bound to L-arabinose and D-galactose, which are side chains of hemicellulose. Therefore, the cellulosic biomass has a complex structure because of the bonding between lignin and hemicellulose (Robak *et al*. 2018).

1-3. Glycoside Hydrolases

Glycoside hydrolases (GHs) cleave glycosyl bonds and are classified into 173 families according to their amino acid sequences. The GHs can be divided into two groups, anomer-inverting and anomer-retaining GHs. In the anomer-inverting enzymes, the anomeric configuration of the product is the inverse of that of the substrate, while in anomer-retaining enzymes, the anomeric configuration of the product is the same as that of the substrate. In the anomer-inverting enzymes, a general acid catalyst and a general base catalyst are located in the active center, and two carboxyl groups are generally responsible for catalysis. On the other hand, in the anomer-retaining enzymes, the general base catalyst activates a water molecule that nucleophilically attacks the anomeric carbon. This water molecule nucleophilically attacks from the opposite side of the glycosidic bond, thus inverting the anomeric form of the product (Okuyama, 2011).

The structure is also an important factor for classifying GHs. The GHs have diverse structures, with the most common being $(\beta/\alpha)_8$ barre (TIM barrel), $(\alpha/\alpha)_6$ barrel, β-jerry roll, β-helix and β-propeller structures (Rabinovich *et al*. 2002). GHs are classified into GH Clans based on their structures, and also into GH families based on their amino acid sequences. Some GH families are $(\beta/\alpha)_8$ barrel, including GH5, which has 56 subfamilies and is the largest GH family. GH43 has 39 subfamilies with a 5-bladed-β-propeller structure. These structures affect the substrate recognition of GHs.

1-4. Lignocellulolytic Enzymes

The main cellulases included in cellulose hydrolysis are of three types: endo-glucanases (EC 3. 2. 1. 4), cellobiohydrolases (EC 3. 2. 1. 91 and EC 3. 2. 1. 176) and β-glucosidases (EC 3. 2. 1. 21). Endo-glucanases can recognize the surface of cellulose

chains and cleave glycoside bonds randomly. The enzymes belonging to GH families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 124 and 148 have endo-glucanase activities.

Cellobiohydrolases (CBHs) can recognize the ends of cellulose chains and release products from there. The enzymes belonging to GH families 5, 6, 7 and 48 have CBH activity. The enzymes belonging to GH6 and GH7 are representative CBHs. One of the GH6 enzymes produced by *T. reesei* is specifically known as cellobiohydrolase II (CBH II). CBH II continuously releases cellobiose from the non-reducing end of the cellulose chain. Such an enzyme is called a "processive enzyme" and can capture the cellulose chain in a tunnel-like structure within the enzyme (Marana, 2012). This feature allows the degradation to proceed continuously, rather than allowing cellulases to leave the cellulose chain after each hydrolysis reaction. Similarly, One of the GH7 enzymes produced by *T. reesei* is specifically known as cellobiohydrolase I (CBH I). CBH I is a "processive enzyme" that continuously releases cellobiose from the reducing end of the cellulose chain (Marana, 2012).

The β-glucosidases belong to GH1, 2, 3, 5, 13, 16, 30, 39, 116 and 131 and cleave cellobiose and cellooligosaccharides from the end of the cellulose chain into glucose. β-Glucosidases can be inhibited by the glucose they themselves produce. β-Glucosidases with product tolerance are required when the concentration of the product is higher.

Enzymes that degrade xylan are broadly classified into xylanases (EC 3. 2. 1. 8) and β-xylosidases (EC 3. 2. 1. 37). The function of each is similar to that of cellulases. Xylanases belong to GH5, 6, 8, 10, 11, 26, 30, 43 and 98 and degrade the xylan chain, releasing xylose or xylooligosaccharides. β-Xylosidases belong to GH1, 2, 3, 5, 10, 30, 39, 43, 51, 52, 54 and 120 and degrade xylobiose or xylooligosaccharides into xylose or shorter xylooligosaccharides.

1-5. Carbohydrate-binding Modules

Carbohydrate-binding modules (CBMs) are non-catalytic proteins appended to glycolytic enzymes and related proteins, and they are currently classified into 95 families according to their amino acid sequences. In general, a short polypeptide, called a linker, links CBMs to the catalytic domains. The role of CBMs in glycolytic enzymes that degrade insoluble polysaccharides is to concentrate the catalytic domain on the substrate to be degraded by the enzyme or on a specific region of the substrate. In this way, CBM contributes to increasing both the effective enzyme concentration at the substrate surface and the rate of polysaccharide hydrolysis (Karita, 2016).

The binding of a CBM to a substrate is mostly accomplished by hydrophobic side chains of tyrosine and tryptophan, aromatic amino acid residues present on the surface of CBMs (Furtado *et al*. 2018; Jam *et al*. 2016). The cellulose that makes up the cellulosic biomass has horizontal hydroxy groups, and the hydrogen bonds between these hydroxy groups bind the celluloses together, forming microfibers. Therefore, the hydrophobic side chains of the aromatic amino acid residues of CBM and the hydrophobic surface of the cellulose microfibers form a hydrophobic interaction, allowing CBM to bind on cellulose

Three types of binding modes govern the binding of CBM to substrates: Types A, B, and C (Boraston *et al*. 2004; Gilbert *et al*. 2013).

The type A binding involves the hydrophobic side chains of aromatic amino acid residues on the surface of the CBM (Boraston *et al*. 2004). An example of a type A CBM is CBM1, which has a surface with three hydrophobic side chains of tyrosine and tryptophan parallel to the cellulose chain that recognizes and binds to the crystalline structure of cellulose. In type B, groove-like structures in the CBM, called clefts, bind to the substrate in a sandwiched manner (Prates *et al*. 2012). Thus, type B CBMs recognize single-stranded sugar chains and amorphous cellulose with loose crystalline structures;

CBM4, CBM6, CBM22, CBM29 and CBM36 are examples of type B binding modes (Boraston *et al*. 2004). CBM9, classified as type C, recognizes cellulose ends inside a pocket-like dent (Notenboom *et al*. 2001). Tryptophan with hydrophobic side chains is also present inside this pocket structure and binds to the cellulose chain.

The ABC classification reflects the function of the CBM structure and the binding mode for each CBM; however, it is not considered to represent the types of substrates. The cellulosic/xylanic substrates included in the plant cell wall are complex and diverse; for example, the xylanic substrates are branched and modified by other saccharides. To reflect the diversity of substrates in CBM-substrate binding, a classification based on the substrate structure rather than on the binding mode has also been suggested (Liu *et al*. 2022). This classification divides CBMs into Types I through IV. Type I CBMs recognize unbranched chains. Type II CBMs and Type III CBMs can bind both unbranched chains and branched chains. Type II CBMs prefer unbranched chains, while Type III CBMs have a higher affinity to branched chains. Type IV CBMs bind only side residues branched from the main chain.

There are also classifications focused on amino acid sequences and CBM structures. Each of these classifications would be needed in order to achieve a full understanding of the CBM–substrate interaction.

1-6. Lytic Polysaccharides Monooxygenases

Lytic polysaccharides monooxygenases (LPMO) are the one of enzymes for the degradation of plant biomass except for hydrolases, and they have been found relatively recently. An endo-glucanase with a weak activity which was formerly classified into GH family 61 was reclassified into the auxiliary activity (AA) family AA9 as a result of the amino acid sequence and the results of structure analysis (Harris *et al*. 2010; Levasseur

et al. 2013). LPMOs in the AA10 family were formerly classified into carbohydrate-binding module (CBM) family 33 with chitin-binding activity (Forsberg *et al*. 2011; Suzuki *et al*. 1998). These were structurally similar to each other and boosted cellulase activity by adding electron donors (Langsto *et al*. 2011). Since then, the families AA11, 13, 14, 15, 16 and 17 have also been identified (Hemsworth *et al*. 2014; Vu *et al*. 2014; Couturier *et al*. 2018; Sabbadin *et al*. 2018; Filiatrault-Chastel *et al*. 2019; Sabbadin *et al*. 2021). Their substrate specificities are diverse: Substrates of AA9, 15, and 16 are cellulose, AA10 and 11 are chitin, AA14 is xylan, AA13 is starch, and AA17 is pectin (Hemsworth *et al*. 2014; Vu *et al*. 2014; Couturier *et al*. 2018; Sabbadin *et al*. 2018; Filiatrault-Chastel *et al*. 2019; Sabbadin *et al*. 2021).

LPMOs are oxidoreductases and require a copper ion in the catalytic center (Quinlan *et al*. 2011). The copper ion is chelated into a "histidine brace" formed of conserved histidine residues (Hemsworth *et al*. 2013). The reaction for LPMO is similar to the "Fenton reaction," in which a hydroxyl radical is generated by iron ions and peroxide. The extra electrons from cellobiose dehydrogenase and ascorbic acid are essential to start an oxidizing reaction (Kracher *et al*. 2016; Vaaje-Kolstad s *et al*. 2010). The copper ion Cu (II) is reduced to Cu (I), and subsequently forms a Cu (II) superoxyl-complex by binding O_2 to Cu(I) (Kim *et al*. 2014). The Cu (II) superoxyl-complex gets an extra electron again, transforming the electron into a Cu (II) oxyl-intermediate that extracts a proton located at the C1 or C4 position of pyranose (Kim *et al*. 2014). The intermediate hydroxylates the C1 or C4 position and makes it unstable, and then cleaves a glycosidic bond (Kim *et al*. 2014). Polysaccharides cleaved by LPMOs were converted into oligo-saccharides having oxidized saccharides at the ends.

LPMOs are characterized into three types based on the oxidized position: C1-oxidizing LPMOs oxidize C1 carbon of pyranose and generate aldonic acid; C4-oxidizing LPMOs

oxidize C4 carbon of pyranose and generate 4-keto aldose; and C1/C4-oxidizing LPMOs randomly oxidize C1/C4 carbon of pyranose (Forsberg *et al*. 2014). The selectivity for the oxidized position is influenced by the position of the copper ion, which is caused by amino acid residues near the active center (Zhou *et al*. 2019).

1-7. *Paenibacillus xylaniclasticus* strain TW1

Paenibacillus xylaniclasticus strain TW1, a Gram-positive facultative anaerobic bacterium, was isolated from sludge in an anaerobic digester fed with pineapple waste in Thailand and could grow on oat spelt xylan (OSX) as the sole carbon source (Ratanakhanockchai *et al*. 2012). This bacterium produced many xylanolytic enzymes and extracellular multienzyme complexes when incubated with birch wood xylan (BWX) (0.5%) and cellulose-binding proteins when incubated with corn hull (0.5%) (Tachaapaikoon *et al*. 2012). Total genome sequencing of *P. xylaniclasticus* strain TW1 was performed and 27,693 open reading frames (ORFs) were predicted. Analysis of the function of these ORFs revealed that this bacterium had many genes for carbohydrate-active enzymes, including genes for at least 29 glycoside hydrolases (GH), 11 carbohydrate esterases, a polysaccharide lyase, a lytic polysaccharide monooxygenase and 24 carbohydrate-binding modules (Table 1-1). *P. xylaniclasticus* strain TW1 digested 50% of the fibril of milled-corn husk within 3 days without chemical pretreatment (unpublished), which is comparable to the digestion ability of *C. thermocellum*, one of the most famous cellulolytic thermo-anaerobic bacteria (Lamed *et al*. 1983) (Fig. 1-1, 1-2). So, *P. xylaniclasticus* strain TW1 would be one of the candidate bacteria for efficiently breaking down the lignocellulose biomass.

1-8. Purpose of This Study

P. xylaniclasticus strain TW1 produces many cellulolytic and xylanolytic enzymes and has a digestion ability comparable to that of *C. thermocellum*. The key advantages of TW1 are that it produces and secretes many enzymes that degrade the plant cell wall, utilizes these hydrolases efficiently, and alternatively forms extracellular enzyme complexes. As to the processes underlying the digestion ability of strain TW1, several clues have been uncovered, but the overall mechanism remains unclear.

It would be predicted that the key enzymes which completely degrade the plant cell wall are xylanolytic enzymes and lytic polysaccharide monooxygenases. The structure of the plant cell wall is very complex, and includes hemicelluloses, including xylan, intertwined with cellulose. To solubilize cellulose completely, the xylan must first be removed. Therefore, a xylanolytic enzyme, *Px*Xyl43A was a focus of this study.

*Px*Xyl43A has an unknown module, and this unknown module is likely to be a carbohydrate-binding module that is appended to many modular glycoside hydrolases. Thus, how the unknown module functions on *Px*Xyl43A was investigated.

Moreover, *P. xylaniclasticus* strain TW1 has an LPMO gene having similarity with known LPMOs. LPMOs have garnered much attention because they show a synergistic effect when reacting with cellulases. It would be reasonable to expect that *Px*AA10A from *P. xylaniclasticus* strain TW1 would also contribute to the digestion ability.

Therefore, in this study, *Px*Xyl43A and *Px*AA10A, which would provide the *P. xylaniclasticus* strain TW1 with its ability to digest cellulose and other components of the cellulosic biomass efficiently, were investigated and characterized.

Figures and Tables

Table 1-1. Putative cellulolytic and xylanolytic enzymes from *P. xylaniclasticus* TW1

Enzyme
GH11 xylanase CBM36
GH74 xyloglucanase CBM3-unknown
GH10 xylanase Fn3-CBM3-unknown
GH43 beta-xylosidase-uknown2
GH9 cellulase
GH5 cellulase CBM17_CBM28-Fn3-SLH
GH10 xylanase
GH9 cellulase Fn3-Fn3-CBM3-unknown
GH48 cellobiohydrolase Fn3-Fn3-Fn3-CBM3-unknown
GH6 cellulase Fn3-Fn3-Fn3-CBM3-unknown
GH5 cellulase Fn3-CBM3-unknown
GH5 cellulase Fn3-CBM3-unknown
GH43 alpha-arabinofruranosidase
GH11 xylanase
GH43 beta-xylosidase CBM6
E-Cellulase domain 2-GH9
GH5 cellulase CBM11-CBM3-unknown
GH5 cellulase Fn3-Fn3-CBM3-unknown
GH8 xylanase
GH43 beta-xylosidase CBM6
GH43 beta-xylosidase
GH11 xylanase CBM36
GH26 mannnanase unknown
GH26 mannanase unknown3
GH43 beta-xylosidase CBM6-CBM36
GH10 xylanase
GH43 xylosidase
GH10 xylanase CBM9-Fn3-Fn3-CBM3-unknown
GH10 xylanase CBM9-SLH

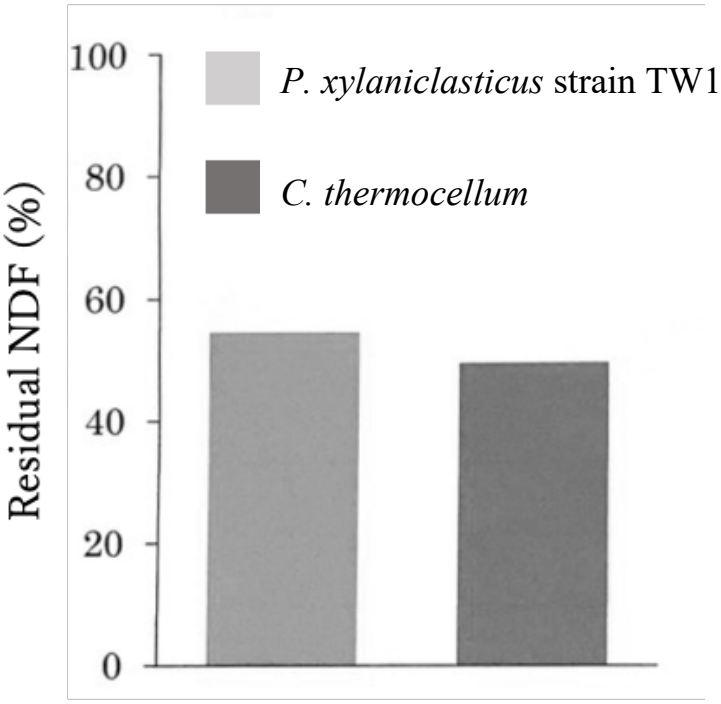


Fig. 1-1. Degradation of untreated corn husk (Ichikawa, 2015).

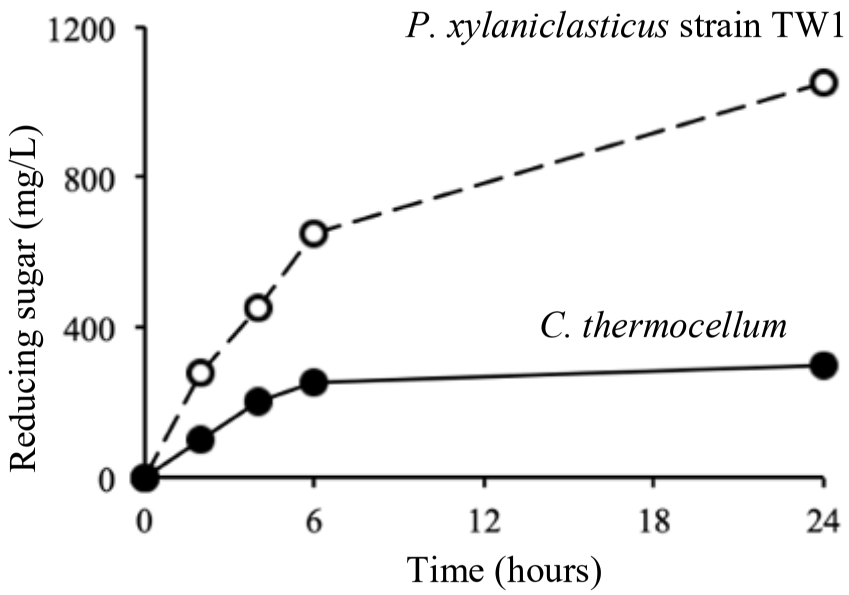


Fig. 1-2. Concentration of released reducing sugar.

White circle: *P. xylaniclasticus* strain TW1, Black circle: *C. thermocellum*.

Chapter2
Characterization of a GH family 43 β-xylosidase

Characterization of a GH family 43 β-xylosidase having a novel carbohydrate-binding module from *Paenibacillus xylaniclasticu*s strain TW1 (Ito *et al*. 2022)

2-1. Summary

P. xylaniclasticus strain TW1, a Gram-positive facultative anaerobic bacterium, has a lot of genes encoding carbohydrate-active enzymes containing many xylanolytic enzymes, for example, GH10 and GH11 xylanases and a GH74 xyloglucanase. Sequence analysis revealed that a GH43 family β-xylosidease from *P. xylaniclasticu*s strain TW1 (*Px*Xyl43A) belonged to GH43 subfamily 12, and had an unknown module (*Px*Xyl43A-UM) like carbohydrate-binding module. To characterize the function of *Px*Xyl43A, in this study we produced recombinant *Px*Xyl43A and *Px*Xyl43A-UM by means of heterologous expression in *E. coli* strain BL21 (DE3), investigated the ability of hydrolysis and binding to substrates. *Px*Xyl43A has activities against *p*-nitrophenyl-β-D-xylopyranoside and *p*-nitrophenyl-β-L-arabinoside, the specific activities are 250mU/mg and 306mU/mg, respectively. The optimal pH value is 7.0 and the optimal temperature is 54°C, it has a K_m of 1.2 mM and V_{max} of 0.012 μmol/min/μg for *p*-nitrophenyl-β-D-xylopyranoside at 54°C and pH 7.0. The *Px*Xyl43A-UM bound to birchwood xylan, oat spelt xylan and ball-milled cellulose, having a K_a of 2.0×10^5 $M^{-1}$and $[PC]_{max}$ of 10.1 μmol/g for 0.05% oat spelt xylan. It was suggested that the *Px*Xyl43A-UM was a carbohydrate-binding module by a K_a value measured. These results suggest that *Px*Xyl43A possesses a novel carbohydrate-binding module, named as CBM91, specific for xylan-containing polysaccharides.

2-2. Introduction

A broad variety of enzymes engage in lignocellulose biomass degradation, cellulases, hemicellulases and β-glucosidases. A role of β-glucosidases in lignocellulose biomass degradation is to cleave β-glycoside bonds between two linked glucose units and to produce fermentable glucose which is able to be used in the following prosses to convert to bio-products (Ketudat *et al*. 2010; Srivastava *et al*. 2019). β-Xylosidases, as well as β-glucosidases, can cleave β-glycoside bonds between two linked xylose units (Rohman *et al*. 2019). GHs are classified into different families in Carbohydrate-Active Enzymes database (CAZy) (Henrissat *et al*.1991; Lombard *et al*. 2014), according to the similarity to the amino acid sequences. β-Xylosidases belong to 11 GH families, GH 1, 5, 30, 39, 43, 51, 52, 54, 116, and 120 (Rohman *et al*. 2018).

The GH43 family is the second biggest GH family following GH3 family and has 37 subfamilies in CAZy (http://www.cazy.org) (Mewis *et al*. 2016). It contains enzymes with β-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55), xylanase (EC 3.2.1.8), α-1,2- L-arabinofuranosidase (EC 3.2.1.-), exo-α-1,5- L-arabinofuranosidase (EC 3.2.1.-), exo-α-1,5-L-arabinanase (EC 3.2.1.-), α-1,3-xylosidase (EC 3.2.1.-), exo-α-1,5-L-arabinanase (EC 3.2.1.-), endo-α-1,5-L-arabinanase (EC 3.2.1.99), exo-α-1,3-galactanase (EC 3.2.1.145) and α-D-galactofuranosidase (EC 3.2.1.146). Some GH43 members are bifunctional or trifunctional, showing β-xylosidase activity, α-L-arabinofuranosidase activity, or xylanase activity simultaneously (Viborg *et al*. 2013; Teeravivattanakit *et al*. 2016). The enzymes in this family also were grouped into type I to type IV based on their modular structure. Type I has only a catalytic module, and type II and type III is accompanied by a carbohydrate-binding module (CBM) family 6 and an X19 module, respectively. CBMs are non-catalytic modules appended to glycoside

hydrolases and facilitate the enzymes to bind to the substrates (Karita, 2016). An X19 module is found only in seven GH43 subfamilies (subfamily 9, 10, 11, 12, 13, 14 and 36), the function is uncleared, but it predicted that the module enhances thermostability for the whole enzymes (Li *et al*. 2020). There is the enzyme that possesses the X19 module at *N*-terminus (Mroueh *et al*. 2013), while, in most enzymes in GH43 family, these additional modules exist in *C*-terminus. Type IV has a more complicated modular structure, for example, containing two or three GH43 domains, and appending to CBM13, 35 and 42.

P. xylaniclasticus strain TW1 has a gene encoding GH43 β-xylosidase (*Px*Xyl43A). *Px*Xyl43A is connected to an unknown region, named *Px*Xyl43A-UM, by linker peptides sequence. In this study, to characterize *Px*Xyl43A, the gene encoding *Px*Xyl43A was cloned and heterologously expressed in *E. coli* strain BL21(DE3). The recombinant *Px*Xyl43A was characterized and the function of *Px*Xyl43A-UM was predicted as a member of a new CBM family, CBM91.

2-3. Material & methods

Strain, plasmid and media

E. coli strain aINV and *E. coli* strain iVEC3 were used as the cloning host, and *E. coli* strain JM109 and *E. coli* strain BL21(DE3) were used as protein expression hosts. pCR2.1 plasmid was used as a cloning vector and, pQE30 plasmid and pET16b plasmid were used for the expression of recombinant His-tagged proteins. Transformed *E. coli* was cultivated in Luria-Bertani broth (LB) medium supplemented with ampicillin (50 μg/ml). Plasmid pET16-*Px*Xyl43A and pQE30-*Px*Xyl43A-UM were constructed for the expression of recombinant proteins *Px*Xyl43A and *Px*Xyl43A-UM, respectively.

Structural analysis of *Px*Xyl43A and *Px*Xyl43A-UM

3D structure of *Px*Xyl43A and *Px*Xyl43A-UM and other proteins were built using Alpha Fold 2 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) (Mirdita *et al*. 2022). The docking tests for *Px*Xyl43A-UM and ligands were performed using CB-DOCK 2 (https://cadd.labshare.cn/cb-dock2/php/index.php) (Liu *et al*. 2022). 3D structures of others proteins were referred from Protein Data Bank (https://www.rcsb.org/). Resulted structures were processed in Pymol (https://pymol.org/2/).

Cloning of *Px*Xyl43A and *Px*Xyl43A-UM gene

A DNA fragment (1585 bp) including the region encoding *Px*Xyl43A (1560 bp) was obtained from genomic DNA of *P. xylaniclasticus* strain TW1 by PCR using primers #1 and #2 (Table 2-1). A DNA fragment of *Px*Xyl43A-UM (684 bp) was amplified from the genomic DNA of *P. xylaniclasticus* strain TW1 using primers #7 and #8. The linear

pET16b for iVEC cloning was amplified using primers #5 and #6. The DNA fragment of *Px*Xyl43A-UM was inserted into pCR2.1. After digestion with *Hin*d III and *Bam*H I, the DNA fragment was ligated into the corresponding site of pQE30. The DNA fragment including the region encoding *Px*Xyl43A was amplified again using primers #3 and #4 and inserted into pET16b by *in vivo* cloning using *E. coli* strain ME9806 (iVEC3) (Nozaki *et. al.* 2019). Each of the inserted DNA fragments was sequenced to confirm the absence of mutations.

Expression and purification of recombinant proteins of *Px*Xyl43A

Transformed *E. coli* cells having pET16b-*Px*Xyl43A or pQE30-*Px*Xyl43A-UM were grown overnight at 37 ˚C in LB medium supplemented with ampicillin (50 μg/ml). When the optical density of the cells at 600 nm reached 0.4, the lac operon trigger isopropyl-1-thio-β-D-galactopyranoside was added to a final concentration of 0.5 mM. After several hours of incubation, cells were collected by centrifugation at 12,000 rpm and disrupted with sonication in 50 mM sodium phosphate buffer (pH 7.0). The recombinant proteins were purified from the cell-free extracts with the Profinia system (Bio-Rad Laboratories, CA, USA) according to the manufacturer's instructions. The purified proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the protein concentration was determined using a Protein Assay kit (Bio-Rad Laboratories, CA, USA) with bovine serum albumin (BSA) as the standard.

Enzyme assay

The activity of recombinant *Px*Xyl43A was measured by incubating protein in 50 mM sodium phosphate buffer (pH 7.0) in the presence of *p*-nitrophenyl-β-D-xylopyranoside (*p*NPX) at 37℃ for 10 min unless otherwise stated. The amount of *p*-nitrophenol released

from *p*NPX was measured by the absorbance at 420 nm. All enzyme assays were carried out at least in triplicate. Except for *p*NPX, *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG), *p*-nitrophenyl-α-L-arabinofuranoside (*p*NPAf), *p*-nitrophenyl-β-D-galactopyranoside (*p*NPGal) were used as soluble substrate. OSX was used as insoluble xylan. The concentration of all substrates is 1.5% (w/v). One international unit (IU) corresponds to 1μmol of xylose equivalent released per min.

Optimal temperature, optimal pH and thermostability

On measuring optimal pH, Britton-Robinson's buffer solutions of different pHs, *i. e.* from 3 to 10, were mixed with recombinant *Px*Xyl43A solution and incubated for 5 min at 54℃ with 1 μM *p*NPX. On measuring optimal temperature, a reaction mixture of recombinant *Px*Xyl43A solution with 1 μM *p*NPX was incubated at a different temperature, from 15℃ to 75℃, on 50 mM potassium phosphate buffer for 5 min. On measuring thermostability, recombinant *Px*Xyl43A solution was mixed with the substrate after incubation at different temperatures, from 40℃ to 57℃, for 2 h. The amount of *p*-nitrophenol released from *p*NPX was measured by the absorbance at 420 nm. All enzyme assays were carried out at least in triplicate.

Kinetic analysis

To determine the reaction kinetics of *Px*Xyl43A, the experiment was performed on 50 mM potassium phosphate buffer at pH 7.0 for 5 min and a substrate concentrate range of 1mM to 100mM. The amount of *p*-nitrophenol released from *p*NPX was measured by the absorbance at 420 nm. All enzyme assays were carried out at least in triplicate.

Macroarray assay

Polysaccharides were applied as 1 µL aliquots to untreated mixed nitrocellulose sheets (MS-Millipore membrane filters) (Millipore, Billerica, MA) in 5-fold dilution series. The sheets were left to dry at room temperature for 30 min. before blocking for 1 h with 5% (w/v) skim milk protein in 5× phosphate buffer (PBS). The nitrocellulose sheets were incubated with appropriate proteins in 5× PBS buffer for 1h, extensively washed with 5× PBS, and then incubated with a 1000-fold dilution of anti-His horseradish peroxidase (HRP) conjugate (Qiagen) in 5×PBS for 1h. After washing with distilled water, the sheets were incubated in freshly prepared HRP substrate (0.05% 3,3'-diaminobenzidine tetrahydrochloride [DAB] and 0.005% H_2O_2) to detect *Px*Xyl43A-UM binding. The reaction was stopped by washing the nitrocellulose sheets with distilled water. Birchwood xylan (BWX), oat-spelt xylan (OSX), ball-milled cellulose (BMC), carboxymethyl cellulose (CMC) and lichenan was used as substrates.

Determination of affinity parameter of *Px*Xyl43A-UM

*Px*Xyl43A-UM solution of appropriate concentrations was added to 20 µl of 3% OSX and incubated for 5 min on ice with gentle mixing. The solutions were centrifuged, and the free protein concentrate in the supernatant was measured by Lowry's method (Lowry *et al*. 1951). The concentration of bound protein was determined from the difference between the total protein concentration and the protein concentration in the supernatant. Absorbance parameters were estimated using the following equation based on Langmuir adsorption isotherm, where K_a and $[PC]_{max}$ are the equilibrium absorption constant and the maximum amount of bound protein respectively:

$$\frac{1}{[PC]} = \frac{1}{K_a[PC]_{max}} \times \frac{1}{[P]} + \frac{1}{[PC]_{max}}$$ (Gilkes *et al.* 1992)

Analysis of hydrolysis product of *Px*Xyl43A

Xylooligosaccharides (xylose to xylotetraose, 10 μg of each), *p*NPX were incubated with 0.1U of the purified *Px*Xyl43A in 10μl of 50mM potassium phosphate buffer (pH7.0) at 37℃ for 1h. Separation of hydrolysis products in the reaction mixture was performed by thin-layer chromatography (TLC) on a DC-Fertigplatten SIL G-25 plate (Macherey-Nage) developed with a butanol-acetic acid-H_2O (2:1:1, v/v/v) solvent mixture. Xylooligosaccharide products on TLC were visualized by incubating 130℃ after soaking the developed TLC plate with methanol (95%)-sulfuric acid (5%) in the reagent.

Binding assay of *Px*Xyl43A-UM

OSX, BWX, BMC and lichenan were used as substrates. Recombinant *Px*Xyl43A solution and bovine serum albumin solution (0.4 mg/ml) diluted with 50 mM potassium phosphate buffer were mixed with 5mg of substrates and incubated for 10 min at room temperature. The mixtures were centrifuged, then the supernatant was retrieved as unbound proteins, and the precipitate was rinsed with 50 mM potassium phosphate buffer and used as bound proteins. BSA was used as a no-binding protein. Unbound proteins and bound proteins were analyzed by SDS-PAGE.

Tolerance for xylose

The tolerance of *Px*Xyl43A was measured by incubating protein with *p*NPX in the presence of xylose (10 mM to 200 mM) at 37℃ for 10 min. The amount of *p*-nitrophenol released from *p*NPX was measured by the absorbance at 420 nm. All enzyme assays were carried out at least in triplicate.

2-4. Result

Sequence analysis of *Px*Xyl43A

Sequence analysis showed that the open reading frame of *Px*Xyl43A consists of 1,560 nucleotides encoding a protein of 520 amino acids with a predicted molecular mass of 57.1 kDa (Fig. 2-1, 2-2). A search of this sequence using BLAST revealed that the query cover with other genes was about 54%, the first half of the sequence (1 bp - 881 bp) was consistent with GH family 43 (GH43_XYL_1 from *P. lautus*; 71.8%, *P.* sp. JDR-2; 69%), but there were no high similarity sequences matched the latter half of the sequence (882 bp – 1560 bp). The DNA sequence of *Px*Xyl43A has registered with the DNA data of Japan with the accession number LC648486.

The *N*-terminus domain of *Px*Xyl43A exhibited similarity to the amino acid sequence of glycoside hydrolase family 43 protein from *P. lautus* E7593-69 (AYB47526) (71.8%). The three active site residues (Asp19, Asp125 and Glu179) were conserved in *Px*Xyl43A, which position is consistent with those of β-xylosidase from *Geobacillus thermoleovorans* IT-08 (ABC75004) and glycoside hydrolase family 43 protein from *P. lautus* E7593-69 (AYB47526) (Fig. 2-3).

The *C*-terminus domain of *Px*Xyl43A showed similarity to the amino acid sequence of β-xylosidase *C*-terminal concanavalin A-like domain, and was conserved in other GH43 (Fig. 2-4). However, this domain was not essential for GH43 β-xylosidases because two GH43 enzymes, for example, β-1,4-xylosidase from *Actinobacillus pleuropneumoniae* CTC11383 (VTR44805) and glycoside hydrolase family 43 protein from *P. taichungensis* (WP_175380580), did not have *C*-terminus domain.

In phylogenetic analysis, elven GH43 enzymes whose three-dimensional structures are registered in the PDB were used: glycoside hydrolase family 43B protein from

Cladosporium acalyphae SL-16 (AQX83203), xylosidase/arabinofuranosidase from *Aspergillus niger* DSM 26641(ALN49267), β-xylosidase/α-L-arabinofuranosidase from *Phanerochaete chrysosporium* BKM-F-1767 (AFW16059), β-1,4-xylosidase and xylan β-1,4-xylosidase from *Bifidobacterium adolescentis* ATCC 15703 (BAF39209 and BAF39984), unnamed protein product from *Aspergillus oryzae* RIB40 (BAE55534), Pc16g08140 from *Penicillium rubens* Wisconsin 54-1255 (CAP93484), β-xylosidase from *G. thermoleovorans* IT-08 (ABC75004), XynB from *Paenibacillus* sp. JDR-2 (ABV90487), glycoside hydrolase family 43 protein from *P. lautus* E7593-69 (AYB47526), glycoside hydrolase 43 family protein from *P. crassostreae* LPB0068 (AOZ93067). The result showed that *Px*Xyl43A belongs to GH43 subfamily 12 (Fig. 2-5).

This protein did not have a signal peptide as a result of a signal peptide analysis using SignalP 5.0. Therefore, this protein is a member of GH43 subfamily 12, an intercellular protein, and composed of a GH43 β-xylosidase at the *N*-terminus and a UM domain at the *C*-terminus.

Structural analysis of *Px*Xyl43A and *Px*Xyl43A-UM

To analyze the 3D structure of *Px*Xyl43A and *Px*Xyl43A-UM, Alpha Fold 2 was used. And, to compare the structures of *Px*Xyl43A and *Px*Xyl43A-UM to their homologous proteins, 3D structures of other proteins were referred from Protein data bank (PDB).

The structure of *Px*Xyl43A was a 5-blade β-propeller and that of *Px*Xyl43A-UM was β-sandwich (Fig. 2-6). A 5-blade β-propeller structure is found in GH43 family and an active site is in the center of it. Amino acids residues in the active site of *Px*Xyl43A, D19, D125 and E179, also exist in the center of it (Fig. 2-7). Y371 and Y497, Y422 are aromatic amino acids that form hydrophobic interaction with substrates. These three

residues also are on a straight line (Fig. 2-8). Others aromatic amino acids were in the surface and loop structure.

Only when the structure analysis on Alpha Fold 2 was performed, the amino acid residues that involve binding to substrates was not identified. So, CB-Dock2 was used to analyze docking substrates and *Px*Xyl43A or *Px*Xyl43A-UM. Xylohexaose (PubChem CID 13872872) and cellohexaose (Pub Chem CID 74539963) were used as a substrate. In the result of CB-Dock2, xylohexaose and cellohexaose bound to the surface of *Px*Xyl43A and *Px*Xyl43A-UM (Fig. 2-9, 10). The binding values of CB-Dock2 were shown as "vina score", and the lower score indicates a strong affinity. When xylohexaose was used as the substrate, the lowest vina scores for *Px*Xyl43A and *Px*Xyl43A-UM were -8.3 and -8.1, respectively. When cellohexaose was used as the substrate, the lowest vina scores for *Px*Xyl43A and *Px*Xyl43A-UM were -7.6 and -7.1, respectively. In the docking for *Px*Xyl43A and xylohexaose, three aromatic acids, Y457, Y459 and F489, were involved in binding to xylohexaose, while in the docking for *Px*Xyl43A-UM and xylohexaose, F489 involved in binding. This site was not on the surface of *Px*Xyl43A-UM, but in a loop structure.

Cloning, expression and purification of *Px*Xyl43A

The amplified DNA encoding *Px*Xyl43A and *Px*Xyl43A-UM were 1585bp and 684 bp, respectively (Fig. 2-1). pET16b-*Px*Xyl43A and pQE30-*Px*Xyl43A-UM were constructed by inserting the amplified DNA into plasmid vector pET16b or pQE30. *Px*Xyl43A and *Px*Xyl43A-UM were expressed heterologously in *E. coli*, the single bands which consisted of calculated molecular weight, 57 kDa and 25 kDa, respectively, were detected by SDS-PAGE. The yields of purified recombinant proteins *Px*Xyl43A and *Px*Xyl43A-UM were 11 mg/ml, and 1.1 mg/ml, respectively.

Enzyme assay

The hydrolytic activity of purified recombinant *Px*Xyl43A toward four *p*-nitrophenyl glycosides, *p*NPX, *p*NPAf, *p*NPG and *p*NPGal as soluble substrates, was measured. *Px*Xyl43A has activities for two *p*-nitrophenyl glycosides, *p*NPX and *p*NPAf. The activity for insoluble substrates was measured using OSX, but no activity was detected (Table 2-2).

The kinetic parameters, K_m and V_{max}, were calculated using the data measured when *p*NPX was used as a substrate (Table 2-2). The value of a K_m was 1.2 mM and the value of a V_{max} was 0.012 μmol/min/μg. The K_m value for *p*NPX of *Px*Xyl43A was lower than other GH43 β-xylosidases, compared with that of the others ranging from 0.38 mM to 17 mM (Michlmayr *et al*. 2013; Jordan *et al*. 2007; Shallon *et al*. 2005; Wagschal *et al*. 2009), indicating that *Px*Xyl43A had higher affinity for *p*NPX than other GH43 β-xylosidases.

The specific activities for *p*NPX and *p*NPAf are 250 mU/mg and 306 mU/mg, respectively, much lower than others, for example, BXA43 from *Bifidobacterium animalis* subsp. *lactis* BB-12(45 U/mg) and AxB8 from *C. thermocellum* B8 (184±4 U/mg), but the values were similar order to XylC from *P. woosongensis* (659.90 mU/mg) and XylC from *B. adolescentis* (164 mU/mg) (Viborg *et al*. 2013; de Camargo *et al*. 2018; Kim *et al*. 2010; Lagaert *et al*. 2011) (Table 2-2).

Optimal pH, optimal temperature and thermostability

The optimal pH of *Px*Xyl43A was pH 7.0, retaining over 80% of its maximum activity at pH 5.0 to 8.0. The optimal temperature was 54.0℃, retaining over 50% of its maximum activity at 40℃ to 60℃. The activity of *Px*Xyl43A after incubation for 2 hours at various temperatures, pH7.0 was maintained to 52.0℃, but decreased drastically after that (Fig. 2-11, 2-12, 2-13).

Tolerance to xylose of *Px*Xyl43A

*Px*Xyl43A had about 74% of its maximum activity in 200 mM D-xylose solution, and 85% of that in 100 mM D-xylose solution (Fig. 2-14). Inhibition by D-xylose was observed in other GH43 β-xylosidases, for example, the activity of XynB2, GH43 subfamily 12, from *Lactobacillus brevis* ATCC 14869 reduced by 66% in 100mM D-xylose solution (Michlmayr *et al*. 2013). The other GH43 β-xylosidase also reduced their activity by more than 50% in 100mM (Rohman *et al*. 2019). Comparing *Px*Xyl43A with other GH43 β-xylosidases, *Px*Xyl43A relatively has a high tolerance for D-xylose.

Analysis of hydrolysis products

To study the mode of action of *Px*Xyl43A for xylooligosaccharides, the products released from those (from xylobiose, X2, to xylohexaose, X6) were detected by TLC. When using X2 as substrates, xylose was released by the β-xylosidase activity of *Px*Xyl43A (Fig. 2-15a). In the cases of X3 to X5, the final product of each substrate was xylose (Fig. 2-15b – 2-15d). No products were produced from X6 (Fig. 2-15e).

Binding to substrates of *Px*Xyl43A-UM

To investigate the binding specificity and affinity of *Px*Xyl43A-UM for insoluble substrates, a macroarray assay and binding tests were performed (Fig. 2-16).

Insoluble fraction and soluble fraction were separated by centrifugation, after incubating to bind *Px*Xyl43A-UM and substrates (Fig. 2-16b – 2-16e). When a band appears in the insoluble fraction by SDS-PAGE, *Px*Xyl43A-UM was considered to bind to the substrates. In cases of OSX and BWX, almost all *Px*Xyl43A-UM existed in bound fraction, and it did not bind specifically to BSA, a non-binding protein, which indicates that *Px*Xyl43A-UM was a xylan-binding protein (Fig. 2-16c and 2-16d). *Px*Xyl43A-UM

did not bind to cellulosic substrates and lichenan because there are no differences between insoluble fraction and soluble fraction of *Px*Xyl43A-UM and BSA when using their (Fig. 2-16d and 2-16e).

Macroarray assay revealed that, when applied at up to 10 μg/dots, *Px*Xyl43A-UM bound to BWX, OSX and CMC, but not BMC and lichenan. The detection limit was 0.2 μg, 1 μg, and 5 μg for birchwood xylan, oat-spelt xylan and CMC, respectively, suggesting that the binding affinity was higher for BWX, followed by OSX and CMC (Fig 2-17).

Binding affinity of *Px*Xyl43A-UM

To determine the affinity of *Px*Xyl43A-UM for insoluble substrates, the unbound protein concentration was measured by Lowry' method. The absorption constant K_a and the maximum amount of the bound $[PC]_{max}$ was calculated from a 1/[P] versus 1/[PC] (Fig. 2-18). The value of the binding constant, K_a, for *Px*Xyl43A-UM was 2.0×10^5 M^{-1}, $[PC]_{max}$ was 10.1 μmol/g for 0.05% oat spelt xylan.

Discussion

Sequence and phylogenetic analysis of *Px*Xyl43A

Sequence analysis showed that many GH43 β-xylosidases have homology to *Px*Xyl43A and the module structure also has similarity to it (Fig. 2-2). Of the GH43 β-xylosidases, as well as *Px*Xyl43A, Xyl from *G. thermolevoran*s IT-08 has a 5-blade β-propeller catalytic module at *N*-terminus and β-sandwich domain at *C*-terminus (Rohman *et al*. 2018), Xyl43A from *Thermobifida fusca* has a GH43 catalytic module at *N*-terminus and an undefined function Module-A at *C*-terminus (Yosida *et al*. 2010), and FSU2269 from *Fibrobacter succinogenes* S85 has a GH43 catalytic module at *N*-terminus and β-sandwich-shaped unassigned domain, XX domain, at *C*-terminus of side (Moraïs *et al*. 2012), XynB3 from *G*. *stearothermophilus* T-6 has an *N*-terminal five-bladed β-propeller catalytic module and an additional β-sandwich domain (Brüx *et al*. 2006), and *Bl*XynB from *Bacillus licheniformis* (ATCC 14580 strain) has a five-bladed β-propeller catalytic module and a β-sandwich accessory domain (Zanphorlin *et al*. 2019). These GH43 β-xylosidases have *C*-terminus domains, but the function of these domains has uncleared.

*Px*Xyl43A has no signal peptide, suggesting either that this protein is an intracellular enzyme or that this bacterium has another expression system independent of signal peptides. Thus, *Px*Xyl43A is expected to be an intercellular enzyme belonging to GH43 subfamily 12, defined as being composed of a GH43 β-xylosidase catalytic module at the *N*-terminus and an unknown-function module, named *Px*Xyl43A-UM, at the *C*-terminus.

The substrate-specificity of *Px*Xyl43A

*Px*Xyl43A showed the activity for *p*NPX and *p*NPAf, but not *p*NPGal and *p*NPG, which was consistent with the trend seen in many GH43 that was able to degrade *p*NPX

and *p*NPAf, not *p*NPG. *Px*Xyl43A had no activity for OSX, insoluble polymerized xylan, indicating that *Px*Xyl43A showed high specificity for soluble substrates and is not active against insoluble sugars (Table 2-2).

GH43 shows a wide variety of substrate specificities, some of which target only insoluble substrates, some of which target only soluble substrates, and others that are active against both. *Pc*Axy43A from *P. curdlanolyticus* B-6 exhibited the activity of both soluble and insoluble, and XylC from *P. woosongensis* also show the activity for *p*NPX, *p*NPAf, OSX and BWX (Teeravivattanakit *et al*. 2013). In other GH43, Xyl43A and Xyl43B from *Humicola insolens* Y1, Abf43A from *Paenibacillus* sp. strain E18, FSU2269 from *Fibrobacter succinogene*s S85 and AxB8 from *C. thermocellum* B8 were able to catalyze *p*NP-substrates and insoluble xylan (Yoshida *et al*. 2010; de Camargo *et al*. 2018; Yang *et al*. 2014; Shi *et al*. 2013).

RuXyn1 from rumen metagenome and HJ14GH43 from *Bacillus* sp. HJ14 exhibit the activity for *p*NPX and *p*NPAf (Wagschal *et al*. 2009; Zhou *et al*. 2012). *Hi*Abf43 from *H. insolens* Y1 and *Sa*Araf43A from *Streptomyces avermitili*s NBRC14893 showed the activity *p*NPAf, but not *p*NPX (Yang *et al*. 2015; Ichinose *et al*. 2008). In contrast, there are few examples of enzymes that are active in *p*NPX but not in *p*NPAf. There were also examples of diversity for *p*NP substrates, XynB3 from *G. stearothermophilus* T-6 was able to react with many *p*NP substrates, *p*NPX, *p*NPAf, *p*NPGal, *p*-nitrophenyl-α-L-arabinofuranoside, *p*-nitrophenyl-α-L-rhamnopyranoside, *p*-nitrophenyl-β-D-fucopyranoside, *p*-nitrophenyl-β-D-glucopyranoside and *p*-nitrophenyl-β-D-mannoyranoside. *Hi*Abf43 from *H. insolens* Y1, BXA43 from *Bifidobacterium animalis* subsp. *lactis* BB-12, XylC from *P. woosongensis* and Xyl43A from *T. fusca* did not catalyze *p*NPG, as well as *Px*Xyl43A (Viborg *et al*. 2013; Kim *et al*. 2010; Moraïs *et al*. 2012; Yang *et al*. 2015).

*Rm*Arase from *Rhizomucor miehei*, *Ct*GH43 from *C. thermocellum,* Abn2 from *Bacillus subtilis*, AbnB from *G. stearothermophilus* exhibited the activity only polymerized arabinan, such as debranched arabinan and sugar beet arabinan, no activity for *p*NP substrates (Chen *et al*. 2015; Das *et al*. 2012; Inácio *et al*. 2008; Alhassid *et al*. 2009).

As mentioned above, while some GH43 enzymes are active only against insoluble sugars and others are active against both insoluble and soluble sugars, *Px*Xyl43A was the type that was active only against soluble sugars. Since *Px*Xyl43A showed activities against *p*NPX and *p*NPAf, it was concluded that *Px*Xyl43A was a bifunctional enzyme with specificity for *p*NPX and *p*NPAf. No signal peptide was found in *Px*Xyl43A, suggesting that it is an intracellular enzyme. Therefore, it was concluded that this enzyme would be active against relatively short soluble sugars because it showed no activity against high-molecular polysaccharides.

Optimal pH, optimal temperature, thermostability and tolerance to xylose

The optimal pH for *p*NPX of other β-xylosidases is between pH 5.5-7.5 (Fig 2-11); RuXyn1 from rumen metagenome is pH 6.5 – 7.5, HJ14GH43 from *Bacillus* sp. HJ14 is pH 7.0, XynB3 from *G. stearothermophilus* T-6 is pH 6.5, deAX from compost starter is pH 5.5-7.5, XylC from *P. woosongensis* is 6.5, AxB8 from *C. thermocellum* B8 is 6.0, while the optimal pH of BXA43 from *Bifidobacterium animalis* subsp. *lactis* BB-12 is from 4.0 to 5.0, having the activity for *p*NPX at low pH (Zhou *et al*. 2012; Zhang *et al*. 2019; Shallom *et al*. 2005; Alhassid *et al*. 2009; Wagschal *et al*. 2009; Kim *et al*. 2010; Das *et al*. 2012; Viborg *et al*. 2013). The optimal pH of *Px*Xyl43A, pH 7.0, is appropriate, judging from the optimal pH of other β-xylosidase and the optimal pH of growth of *P. xylaniclasticus* strain TW1.

RuXyn1 from rumen metagenome retained 20% and 40% of maximum activity at pH 5.0 and 8.0, respectively (Zhou *et al*. 2012). AxB8 from *C. thermocellum* B8 retained 50% of maximum activity at pH 7.0 (Das *et al*. 2012). As in these results, some β-xylosidases decrease activity at more than pH 8.0 or less than 5.0. As well as most β-xylosidases, *Px*Xyl43A retained over 80% of its maximum activity at pH 5.0 to 8.0. This pH range includes the pH suitable for the growth of *P. xylaniclasticus* strain TW1.

The optimal temperature for *p*NPX of other β-xylosidases is about 50-65℃ (Fig 2-12); BXA43 from *Bifidobacterium animalis* subsp. *lactis* BB-12 and XylC from *P. woosongensis* is 6.5 is 50℃, deAX from compost starter is 55℃ and XynB3 from *G. stearothermophilus* T-6 is 65℃ (Viborg *et al*. 2013; Kim *et al*. 2010; Wagschal *et al*. 2009; Shallom *et al*. 2005). RuXyn1 from rumen metagenome and HJ14GH43 from *Bacillus* sp. HJ14 showed the optimal temperature at 40℃ and 25℃, respectively. However, Enzymatic reactions are more efficient at higher temperatures (Zhou *et al*. 2012; Zhang *et al*. 2019). In enzymatic reactions, *Px*Xyl43A has an advantage over enzymes that exhibit a lower optimum temperature.

The activity of SXA from *Selenomona ruminantium* decreased drastically after incubating at 50℃ for 1 hour (Rohman *et al*. 2019). The other GH43 decreased drastically after incubating until 50℃ for less than 1 hour; BXA43 from *Bifidobacterium animalis* subsp. *lactis* BB-12, 50℃ 10 mins, deAX from compost starter, 48℃, 20 mins (Mewis *et al*. 2016; Wagschal *et al*. 2009). The thermostability of *Px*Xyl43A is higher than other GH43 because this enzyme retained the activity after incubating for 2 hours at 52.0℃.

When degrading lignocellulosic biomass using enzymes, cellulases and hemicellulases, the final products are monosaccharides, D-glucose, D-xylose and L-arabinose and so on. They inhibit β-glucosidases and β-xylosidases by binding active sites of those (Table 2-3). β-Xylosidases have subsite +1 and -1, and one xylose unit of xylobiose bind to subsite

+1, and the other xylose unit binds to subsite -1. Each xylose unit interacts with adjacent amino acid residues in the subsites through hydrogen bonds or hydrophobic stacking interaction: for example, W145 in the subsite of β-xylosidase SXA from *S. ruminantium* GA192 and E198 in the subsite of β-xylosidase Xyl from *G. thermoleovorans* IT-08 (Fan *et al*. 2010; Rohman *et al*. 2019) (Fig 2-19). In the case of β-xylosidase from *S. ruminantium* GA192, mutation W145 to other amino acid residues decreased inhibition by D-xylose, and W145G was responsible for most of the lost affinity for the monosaccharides (Fan *et al*. 2010). Mutation E198 of β-xylosidase, from *G. thermoleovorans* IT-08, to other amino acid residues resulted in lowering affinity for D-xylose (Rohman *et al*. 2016). From these mutation experiments, inhibition by D-xylose would be caused by W145 and E198, respectively. *Px*Xyl43A did not have amino acid residues that interact with D-xylose in subsite +1 and -1, Trp residue and Glu residue corresponding with W145 in SXA from *S. ruminantium* GA192 and E198 in Xyl from *G. thermoleovorans* IT-08, these residues in *Px*Xyl43A were replaced into Gly residue, which is a low affinity for D-xylose (Fig. 2-3) (Rohman *et al*. 2016). Therefore, *Px*Xyl43A relatively has a high tolerance for D-xylose, and would be used in lignocellulosic biomass degradation at high concentrations of final products.

Analysis of hydrolysis products

As the result of the mode of action of *Px*Xyl43A for xylooligosaccharides by TLC for products, *Px*Xyl43A was able to recognize and cleave X2 to X5, but no activity was detected from X6. XynB3 from *G. stearothermophilus* T-6 is an exo-acting β-xylosidase, and the final product of that when using X6 as the substrate is xylose units (Shallom *et al*. 2005). *Px*Xyl43A, however, did not release xylose units and smaller xylooligosaccharides from X6, so it seems not to recognize larger xylooligosaccharides,

the degree of polymerization (DP) is more than 6. The results indicated that *Px*Xyl43A was an exoglycosidase because it removed xylose from the terminal ends of the xylooligosaccharides, up to 5 of DP.

Structural analysis of *Px*Xyl43A and *Px*Xyl43A-UM

To analyze the 3D structure of *Px*Xyl43A and *Px*Xyl43A-UM, Alpha Fold 2 and CB-Dock2 were used. The catalytic module of *Px*Xyl43A was a 5-blade β-propeller structure and *Px*Xyl43A-UM was a β-sandwich structure. GH43 family enzymes are typically a 5-blade β-propeller structure and an active site is in the center of it. The structures of D-xylosidase from *S. ruminantium* (PDB ID 3C2U) and glycoside hydrolase family 43 β-1,4-xylosidase from *G. thermoleovorans* IT-08 (PDB ID 5Z5D) corresponded to that of *Px*Xyl43A. Two structures were solved and registered in PDB.

A β-sandwich structure is typically found in CBMs, two β-sheets face each together. From the amino acids sequence, *Px*Xyl43A-UM has many aromatic amino acid residues which would involve in binding to substrates (Boraston *et al*. 2004). Some GH43 family enzymes do not have *C*-terminus modules, but the hydrolysis activities retain. *C*-terminus modules like a *Px*Xyl43A-UM, therefore, are not necessary for the activities of GH43 family enzymes. *Px*Xyl43A-UM is likely a new CBM because the amino acid sequence has not been registered as CBM, and the structure and conserved amino acid residues are unique to CBM.

The amino acid residues which are on the surface are involved in binding to substrates, the surface curving to hold substrates (Boraston *et al*. 2004; Abbott *et al*. 2014). In some cases, Loops which is formed by β-hairpin structure connecting β- sheets make a groove that binds to substrates. The former is called a “concave face site (CFS)”, and the latter is called a “variable loop site (VLS)” (Abbott *et al*. 2014).

A binding site as a “concave face site” of *Px*Xyl43A-UM would be Y371, Y497 and Y422 which are on a straight line on the surface. This site is optimal for oligosaccharides to bind to this protein because three aromatic amino acid residues exist along oligosaccharides. CBM 1 has a binding site that accommodates three aromatic amino acids along oligosaccharides (Kraulis *et al*. 1989). These amino acid residues are involved in binding to substrates on the surface. This is an example of type A CBM. CBM 29 also has three aromatic amino acids in a cleft formed by CFS, which can bind to an oligosaccharide (Charnock *et al*. 2002). In the analysis of the structure, *Px*Xyl43A-UM has CFS that possesses three aromatic amino acids. Therefore, this site is likely to be one of the binding sites. On the other hand, VLS also is a candidate binding site. CBM 6, for example, has an aromatic amino acid in VLS, this site is one of the binding sites (Jam *et al*. 2006). A binding site of *Px*Xyl43A-UM as VLS would be F498 in a loop structure.
The result of docking simulation using CB-Dock2, xylohexaose bound to some sites on the surface of *Px*Xyl43A-UM. The binding site which showed the lowest “vina score”, the value is -8.1, was involved in Y457, Y459 and F489. These three aromatic amino acids existed in other *C*-terminus modules except for *Px*Xyl43A-UM, and were highly conserved, particularly Y459 and F489.

Y371, Y497 and Y422 which are on a straight line in the CFS were regarded as important residues to bind to substrates, but in the recent docking test, those were not involved in binding to substrates. In CB-Dock2, if there is not enough space, a protein does not bind to a ligand because CB-Dock2 performs a docking test by searching cavities in the proteins. The CFS of *Px*Xyl43A-UM was covered by a loop, and the loop would have caused a three-dimensional barrier for docking. Loop structures in proteins are flexible, and actually CFS covered by the loop of *Px*Xyl43A-UM is likely to accommodate substrates if the loop opens to bind to substrates.

Function of *Px*Xyl43A-UM

FSUAXH1 from *F. succinogenes* S85 has a β-sandwich-shaped unassigned domain, XX domain, at the *C*-terminus of its β-xylosidase. To investigate the affinity of XX domain, XX domain which was truncated from a GH43 β-xylosidase was expressed, then affinity-gel assays were performed. The result showed that XX domain did not bind its substrates, arabinoxylan, and GH43 without XX domain also did not show binding affinity. Yet, GH43 with XX domain was able to bind to substrates. Both XX domain and a GH43 β-xylosidase were required for FSUAXH1 to bind to the substrate, and each of them alone resulted in a loss of binding activity. Therefore, FSUAXH1 binds to substrates by interdependent interaction between XX domain and a GH43 β-xylosidase (Yoshida *et al.* 2010).

An ancillary of an undefined function, named as module Module-A, at the *C*-terminus was derived from *T. fusca* Xyl43A (Moraïs *et al.* 2012). Module-A, when binding to insoluble polysaccharides and assessed by SDS-PAGE, was found in bound fraction and unbound fractions in almost equal proportions. Thus, Module-A plays a minor role in the binding interaction, while Xyl43A binds to substrates due mainly to the residues present in the catalytic module (Moraïs *et al.* 2012).

In the case of *Px*Xyl43A, *Px*Xyl43A-UM which truncated from the catalytic module was able to bind substrates, suggesting that *Px*Xyl43A-UM binds to substrates through the interaction between carbohydrates and CBM, which is observed in general CBM, and no interaction with GH43 is required for substrate binding. Thus, *Px*Xyl43A-UM was a carbohydrate-binding module because it showed binding activity independently.

The K_a value of CBM is broad variety, the range is approximately from 10^2 to 10^5 M^{-1} (van Bueren *et al.* 2005; Boraston *et al.* 2004) (Table 2-4). For example, The K_a values to ball-milled cellulose of *Cj*CBM17, family 17 CBM, and *Cj*CBM28, family 28 CBM, from

the endoglucanase Cel5A of an anaerobic cellulolytic bacterium, *Clostridium josui* was 7.2×10^5 M^{-1} and 3.5×10^5 M^{-1}, respectively (Araki *et al.* 2009). The K_a value to acid-swollen cellulose of *Cj*CBM3 was 4.1×10^6 M^{-1} at the first binding site (Ichikawa *et al.* 2014). The K_a values to crystalline cellulose of family 1 CBMs of cellobiohydrolase Cel7A and Cel6A from the filamentous fungus *T. reesei* (also called *Hypocrea jecorina*) also 1.9×10^5 M^{-1} and 2.7×10^5 M^{-1}, respectively (Guo *et al.* 2013).

As in the above example, CBM from bacteria and filamentous fungus showed a K_a value of about 10^5 M^{-1}. Moreover, xylan-specific CBMs belonging to CBM family 4, 6, 15, 22 and 36 all show similar K_a values. CBM4, CBM15 and CBM22 were able to bind OSX with the K_a value of 6.67×10^5 M^{-1}, 13.5×10^4 M^{-1} and 7.6×10^4 M^{-1}, respectively (Cicortas Gunnarsson *et al.* 2007; Szabó *et al.* 2001; Charnock *et al.* 2000). CBM6 and CBM36 were able to bind xylooligosaccharides with the K_a value in the rage of 10^2 M^{-1} to 10^5 M^{-1} and 13×10^3 M^{-1}, respectively (Sakka *et al.* 2003; Jamal-Talabani *et al.* 2004). *Px*Xyl43A-UM bound to OSX with the K_a value of 2.0×10^5 M^{-1}, which was higher than those of CBM15 and CBM22. And CBMs, generally, bind to the substrates that their catalytic modules target. Therefore, it was presumed that *Px*Xyl43A-UM is a xylan-specific binding module because the catalytic module was a β-xylosidase and it bound to insoluble xylan. These results suggest that *Px*Xyl43A-UM is a novel xylan-specific CBM. Therefore, it was proposed that this module was a novel CBM. This proposal was accepted and created CBM family 91 in the international CAZy database.

In ABC classification, CBM91 would be a type B CBM because CBM91 does not have a flat surface, but a cleft that binds to the substrates. *Px*Xyl43A-UM bound to oat spelt xylan, which has branched chains, and cellulose, indicating that CBM91 would be type II or type III CBM. This study, however, was not able to determine whether CBM91 would be type II or type III CBM.

Figures and Tables

Table 2-1. List of primers.

1	AACAGTATAGCAGTTAAATACGAAA
2	TTAGGACAATCCTCATTAAGCTCCG
3	CGGCCATATCGAAGGTCGTCATATGAACAGTATAGCAGTTAAATA
4	GCCCCAAGGGGTTATGCTAGTTAGGACAATCCTCATTAAGCTCCG
5	CATATGACGACCTTCGATATGGC
6	TAACTAGCATAACCCCTTGGGG
7	GGATCCATGAATGTCGAAATTCTCGCAGGA
8	AAGCTTTCAAGACAGCGGTGTATAGTCCAG

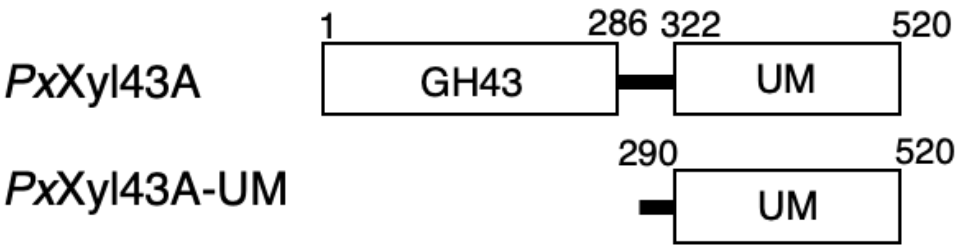


Fig. 2-1. Schematic diagram of *Px*Xyl43A and *Px*Xyl43A-UM. The numbers represent the number of amino acids. *Px*Xyl43A consists of a GH43 catalytic domain (GH43) and an unknown function module (UM), which are connected by a linker peptide between 287 and 321.

Fig. 2-2. Multiple protein aliment of *Px*Xyl43A, *Bo*GH43A (ALJ47681) and *Bo*GH43B (ALJ47682) from *B. ovatus* and Xyl from *G. thermoleovorans* (ABC75004). Each aliment was obtained by using NCBI and aligned by using Clustal Omega.

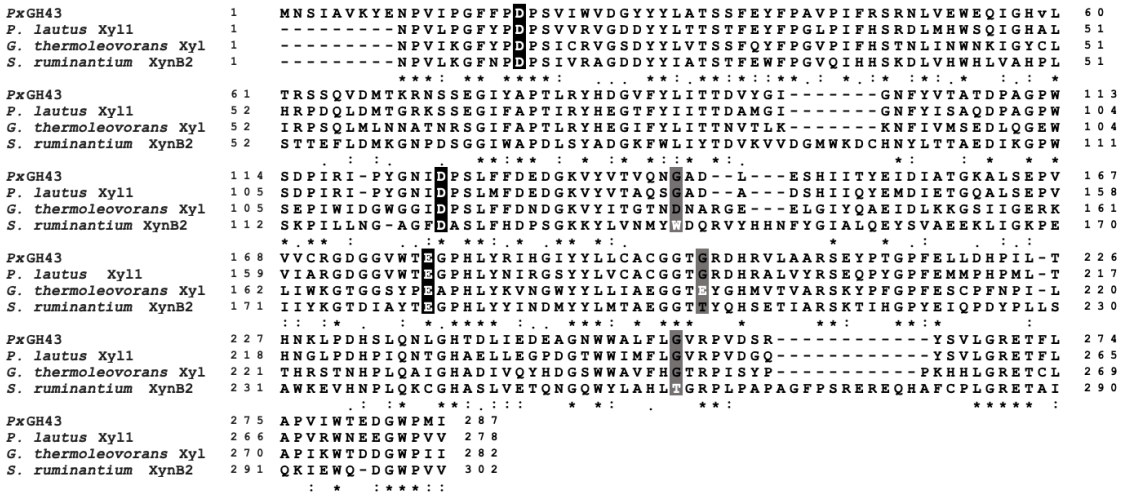


Fig. 2-3. Multiple protein aliment of the catalytic domains of *Px*Xyl43A (*Px*GH43), Xyl1 from *P. lautus* (AYB47526), Xyl from *G. thermoleovorans* (ABC75004) and Xynb2 from S. *ruminantium* (AAB97967). Three conserved catalytic residues are highlighted in solid black, and the inhibition residues are highlighted with outline characters in solid gray. Each aliment was obtained by using NCBI and aligned by using Clustal Omega.

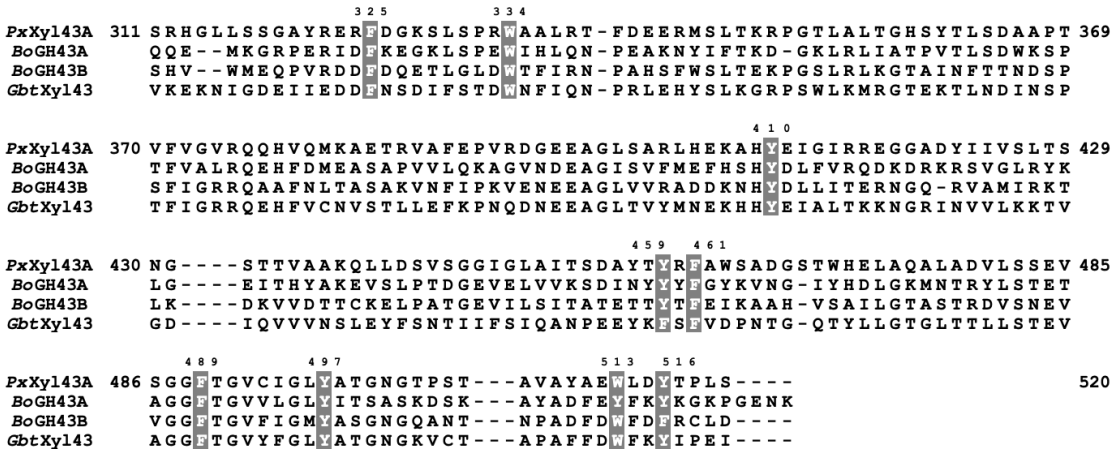


Fig. 2-4. Multiple protein aliment of *Px*Xyl43A, *Bo*GH43A (ALJ47681) and *Bo*GH43B (ALJ47682) from *B. ovatus* and Xyl from *G. thermoleovorans* (ABC75004). Each aliment was obtained by using NCBI and aligned by using Clustal Omega. The conserved catalytic residues are highlighted in solid gray. Each aliment was obtained by using NCBI and aligned by using Clustal Omega.

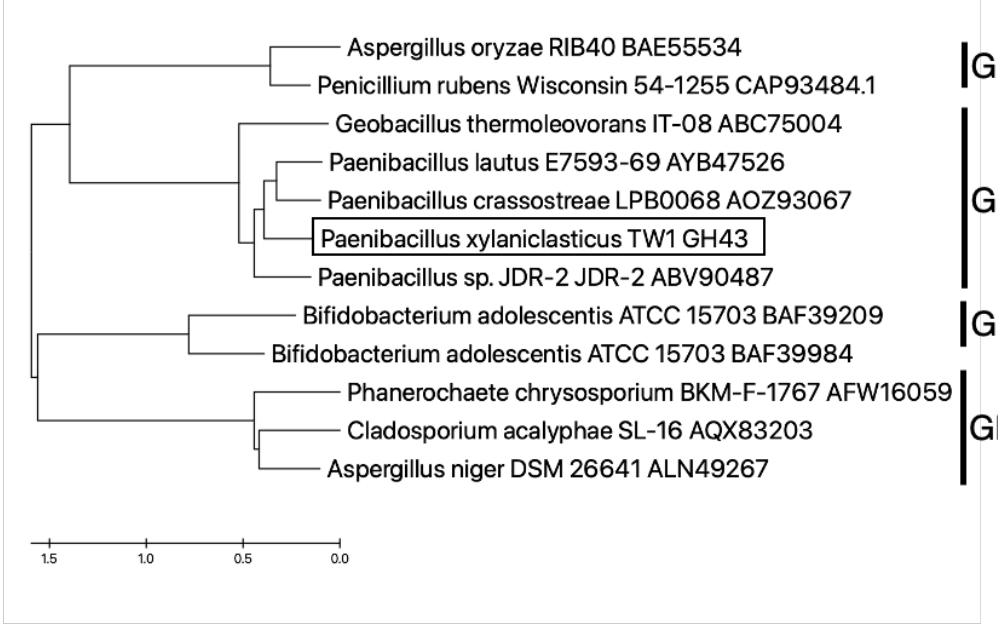


Fig. 2-5. Phylogenetic tree of GH43 subfamily 11 to 14. Each aliment was obtained by using NCBI. The phylogenetic tree was built using the neighbor-joining method in Mega X. The following enzymes were used in this analysis: glycoside hydrolase family 43B protein from *Cladosporium acalyphae* SL-16 (AQX83203), xylosidase/ arabinofuranosidase from *Aspergillus niger* DSM 26641 (ALN49267), β-xylosidase/α-L-arabinofuranosidase from *Phanerochaete chrysosporium* BKM-F-1767 (AFW16059), β-1,4-xylosidase and xylan β-1,4-xylosidase from *Bifidobacterium adolescentis* ATCC 15703 (BAF39209 and BAF39984), unnamed protein product from A*spergillus oryzae* RIB40 (BAE55534), Pc16g08140 from *Penicillium rubens* Wisconsin 54-1255 (CAP93484), β-xylosidase from *G. thermoleovorans* IT-08 (ABC75004), XynB from *Paenibacillus* sp. JDR-2 (ABV90487), glycoside hydrolase family 43 protein from *P. lautus* E7593-69 (AYB47526), glycoside hydrolase 43 family protein from *P. crassostreae* LPB0068 (AOZ93067).

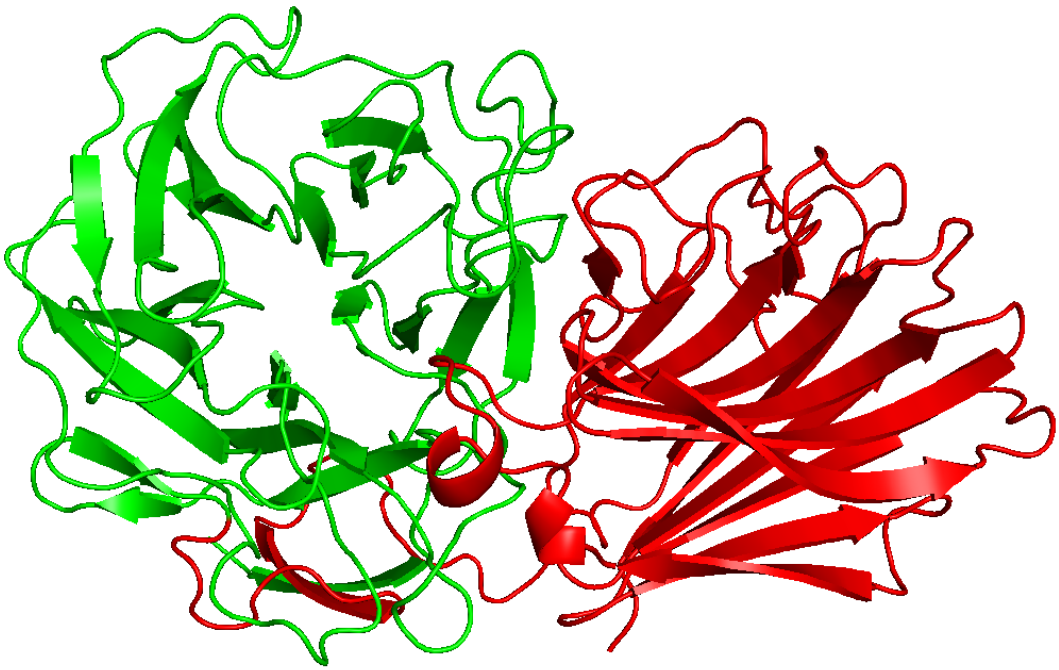
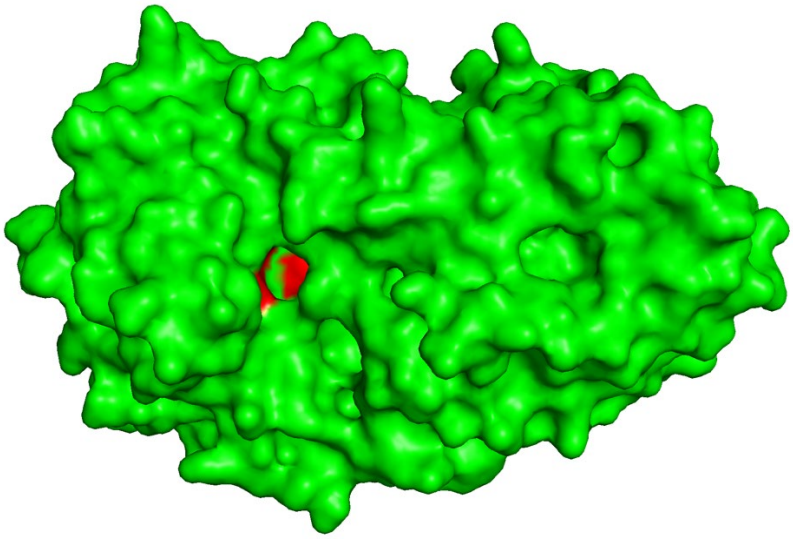


Fig. 2-6. Structure of *Px*Xyl43A. The prediction structure was performed by using Alpha Fold 2 and the structure was built by Pymol. *Px*Xyl43A is composed of the catalytic domain of *Px*Xyl43A (1-287) (green) and CBM91 (288-520) (red). The catalytic domain is 5-bladed-β-propeller, and CBM91 is β-sandwich.

(a)



(b)

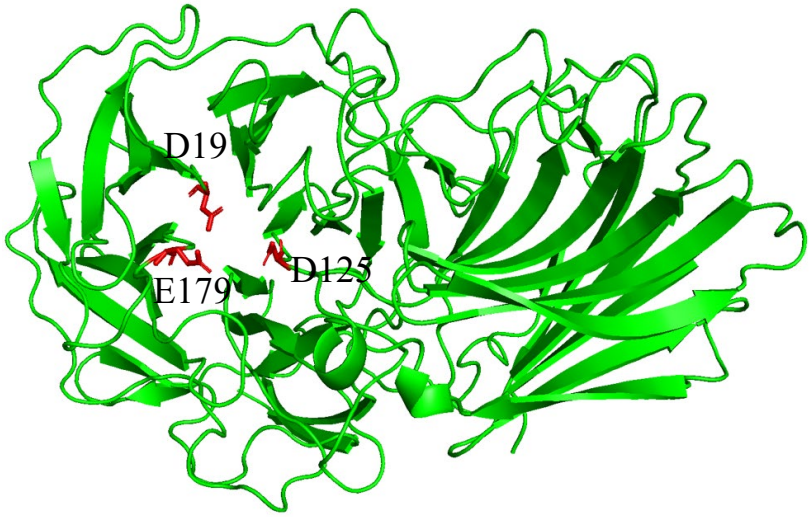


Fig. 2-7. The active site of *Px*Xyl43A. The predicted structure was performed by using Alpha Fold 2 and the structure was built by Pymol. Amino acids residues, D19, D125 and E179 in the catalytic center are colored in red.

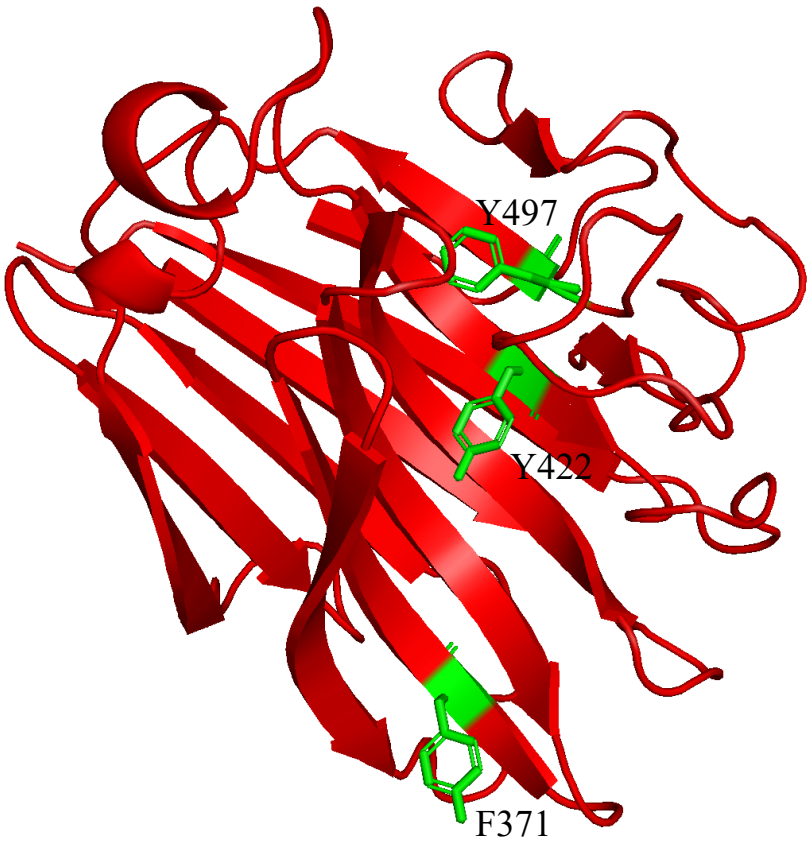


Fig. 2-8. The predicted substrate-binding site of CBM91 from *Px*Xyl43A. The prediction structure was performed by using Alpha Fold 2 and the structure was built by Pymol. Three aromatic amino acid residues, F371, Y422 and Y497 which are colored in green, are in range with each other on β-sheets.

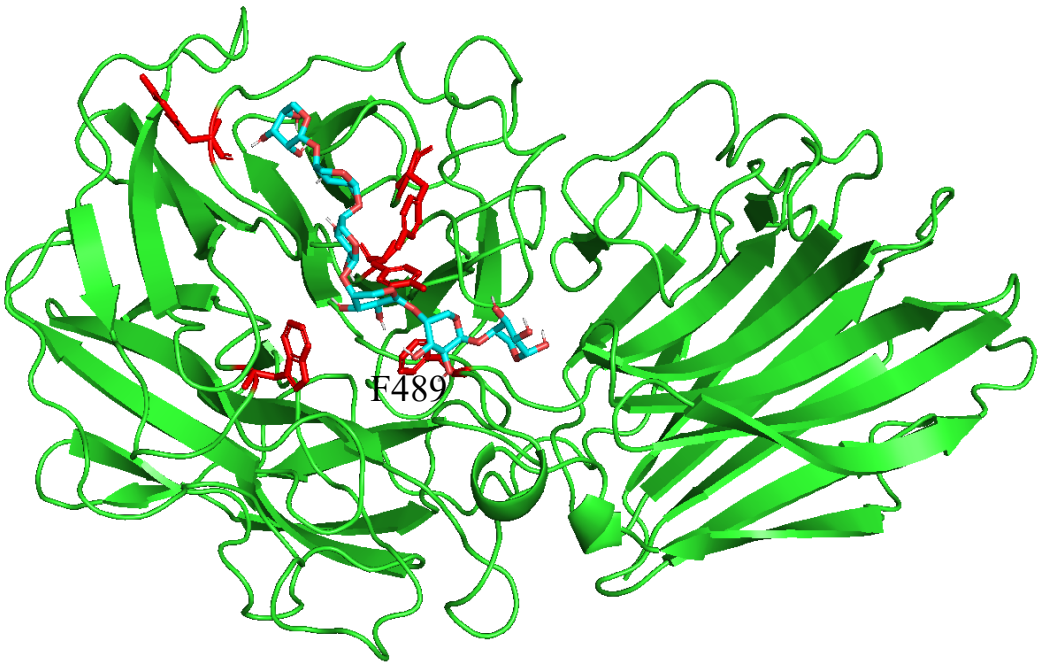


Fig. 2-9. Structure resulted from docking test of *Px*Xyl43A and xylohexaose. CB-Dock2 was used to perform a docking test between *Px*Xyl43A (green) and xylohexaose (cyan). Aromatic amino acids which involved in binding to xylohexaose are colored in red. The side chains were shown as sticks.

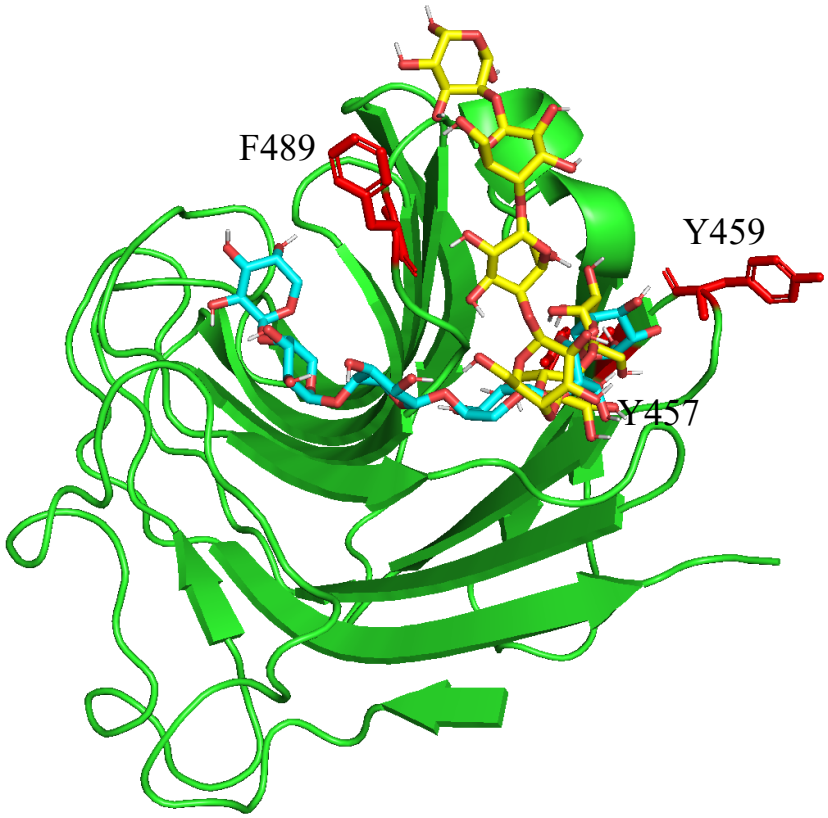


Fig. 2-10. Structure resulted from docking test for CBM91 and hexaoligosaccharides. CB-Dock2 was used to perform a docking test between CBM91 (green) and xylohexaose (cyan) or cellohexaose (yellow). Aromatic amino acids involved in binding to substrates are colored in red. The side chains were shown as sticks. A side chain of tyrosine Y457 is involved in binding to both substrates.

Table 2-2. Specific activities and kinetic parameters of *Px*Xyl43A

substrates	spesific activity (mU/mg)	K_m (mM)	V_{max} (mM/min)	K_{cat} (/min)	K_{cat}/K_m (min/mM)
*p*NP-β-D-xylopyranoside	250 ± 37	1.2 ± 0.068	0.75 ± 0.039	170 ± 8.8	144 ± 0.91
*p*NP-α-L-arabinofuranoside	306 ± 29	—	—	—	—
*p*NP-β-D-glucopyranoside	No actvity	—	—	—	—
*p*NP-β-D-galactopyranoside	No actvity	—	—	—	—
Oat-spelt xylan	No actvity	—	—	—	—

Values are the means of triplicate experiments ± the standard deviation.

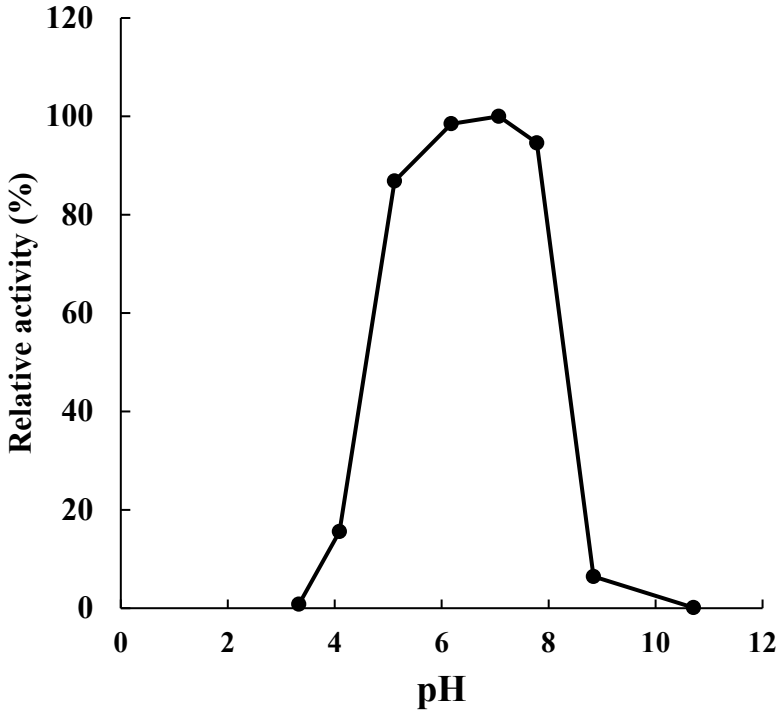


Fig. 2-11. Optimal pH of *Px*Xyl43A. Error bars represent standard deviation.

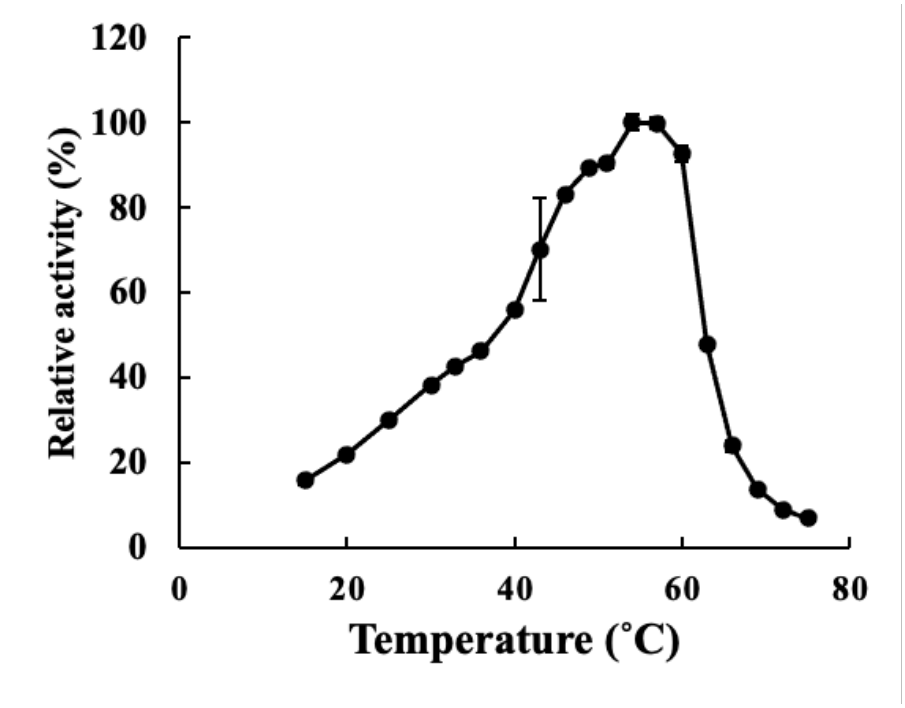


Fig. 2-12. Optimal temperature of *Px*Xyl43A. Error bars represent standard deviation.

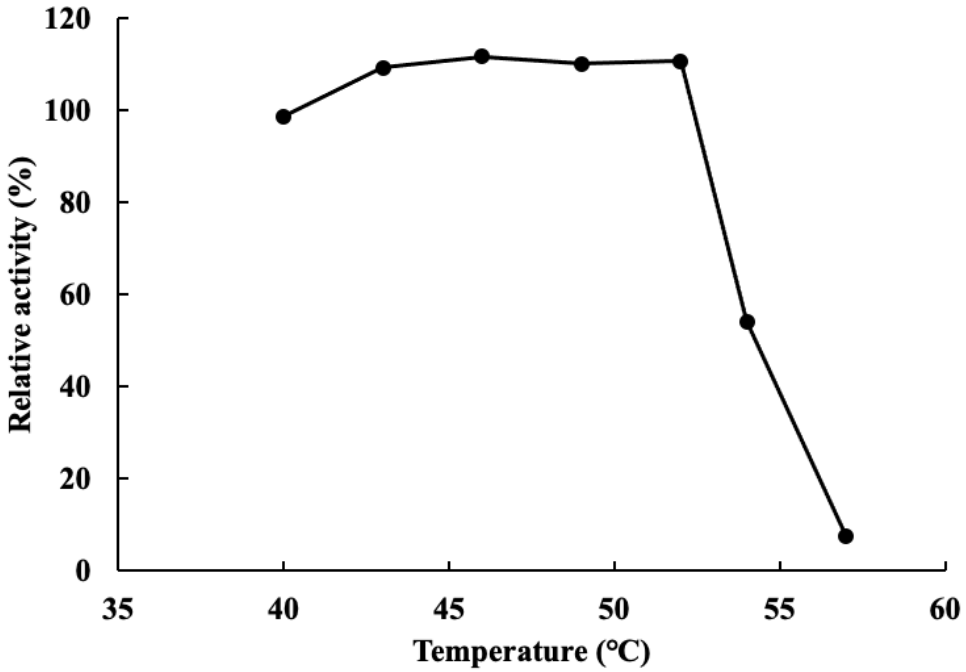


Fig. 2-13. Thermostability of *Px*Xyl43A. Error bars represent standard deviation.

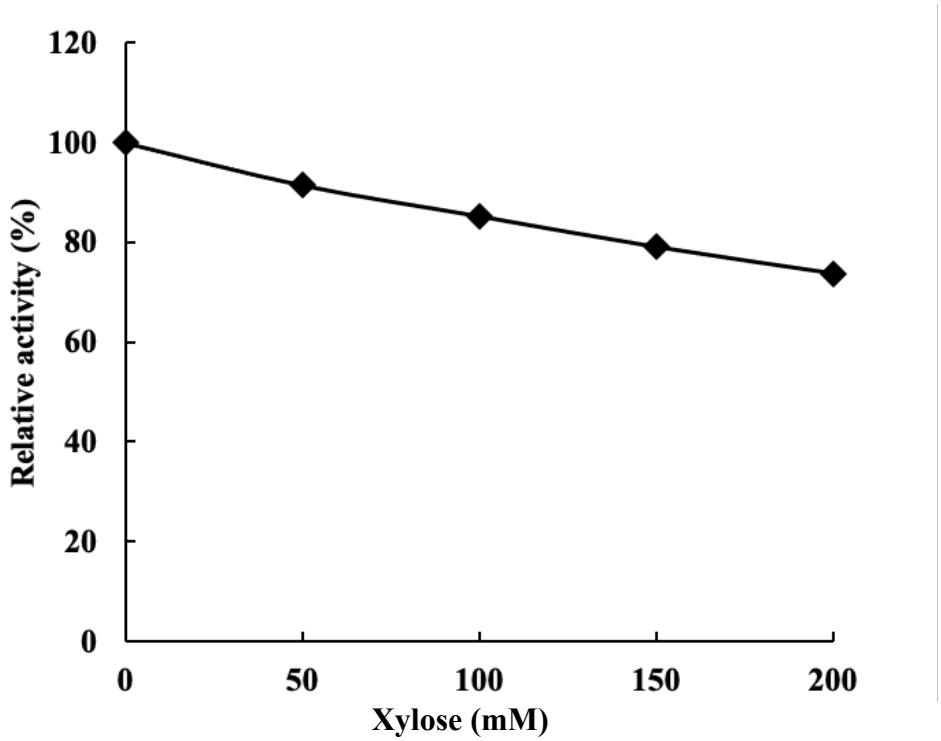


Fig. 2-14. Effect of xylose concentration on the *Px*Xyl43A activity. Error bars represent standard deviation.

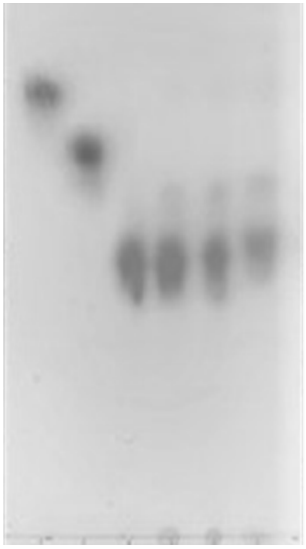
(a)

X1 1 2 3 4

(b)

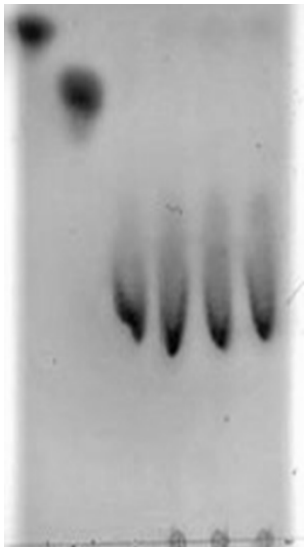
X1 1 2 3 4

(c)



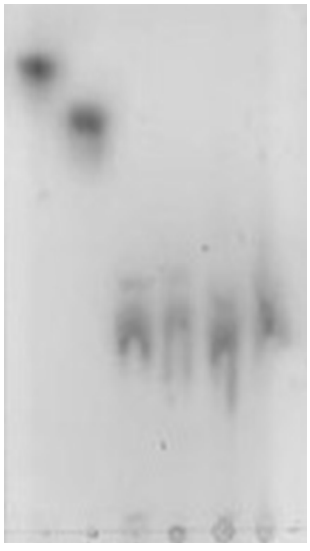
X1 X2 1 2 3 4

(d)



X1 X2 1 2 3 4

(e)



X1 X2 1 2 3 4

Fig. 2-15. Thin-layer chromatography of hydrolysis products released from xylooligosaccharides, xylobiose (a), xylotriose (b), xylotetraose (c), xylopentaose (d) and xylohexaose (e). Lane X1, xylose; lane X2, xylobiose; lane 1-4, hydrolysis of xylooligosaccharides after 1, 2, 3 and 4 hours, respectively.

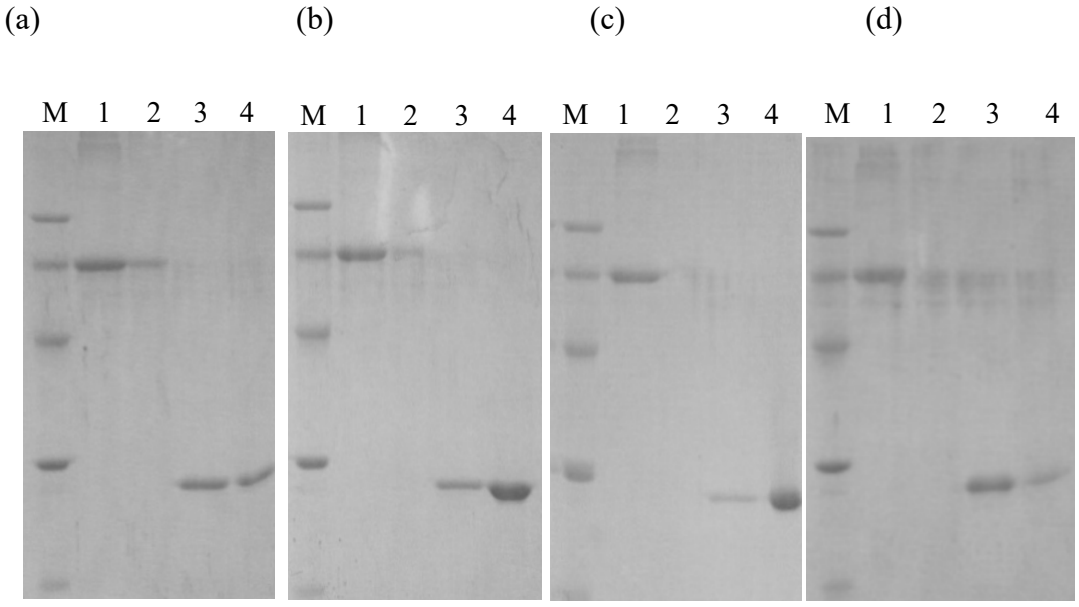


Fig. 2-16. Binding of *Px*Xyl43A-UM to insoluble polysaccharides, BMS (a), OSX, (b) BWX (c) and lichenan (d). Lane M, marker; lane 1 bound fraction (BSA); lane 2, unbound fraction (BSA); lane 3, bound fraction (each polysaccharide); lane 4, unbound fraction (each polysaccharide).

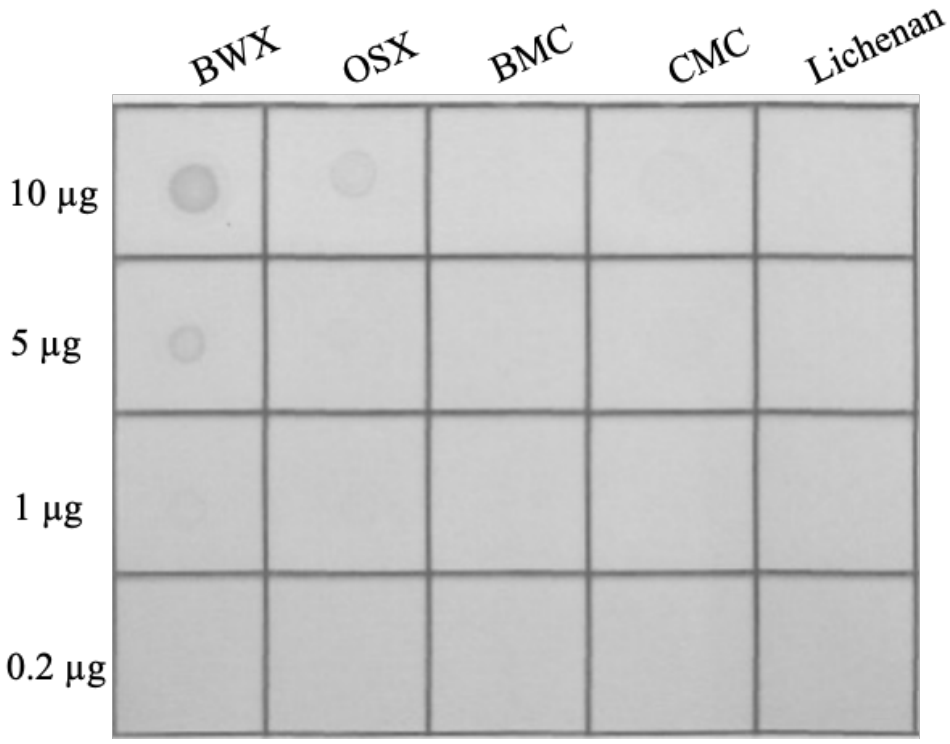
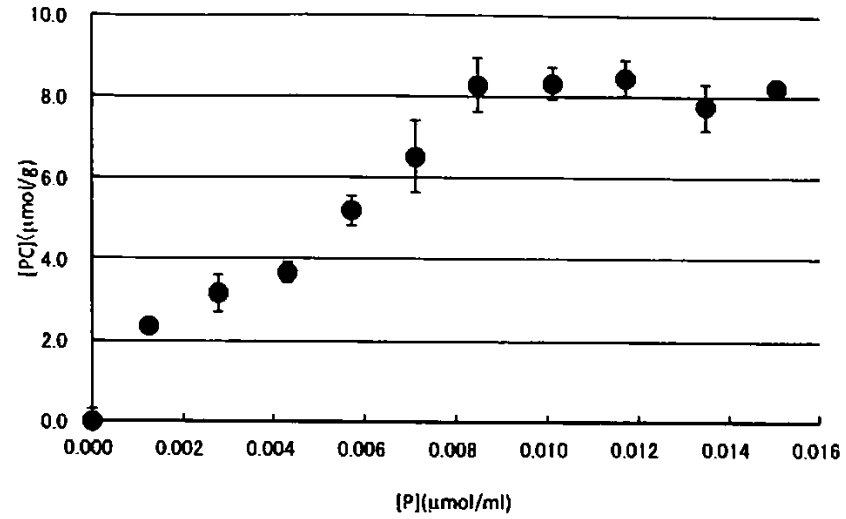


Fig. 2-17. Macroarray assays of the binding of *Px*Xyl43A-UM.

(a)



(b)

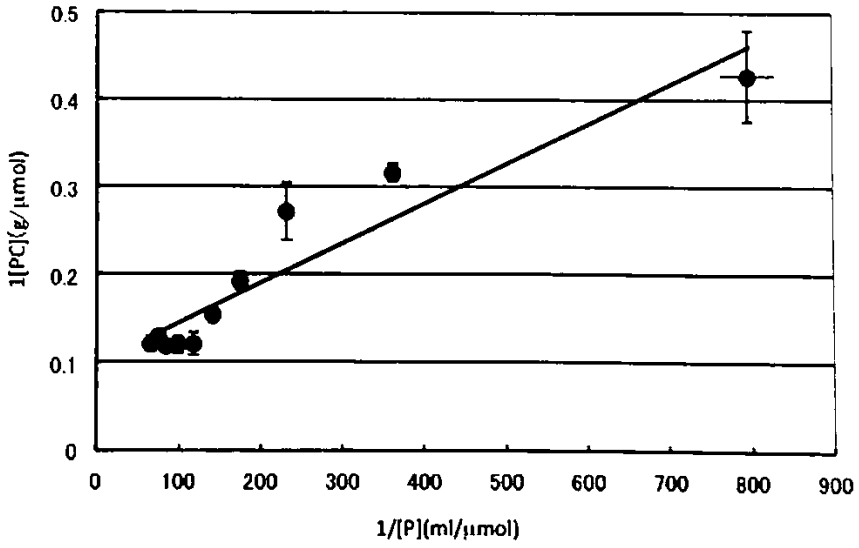
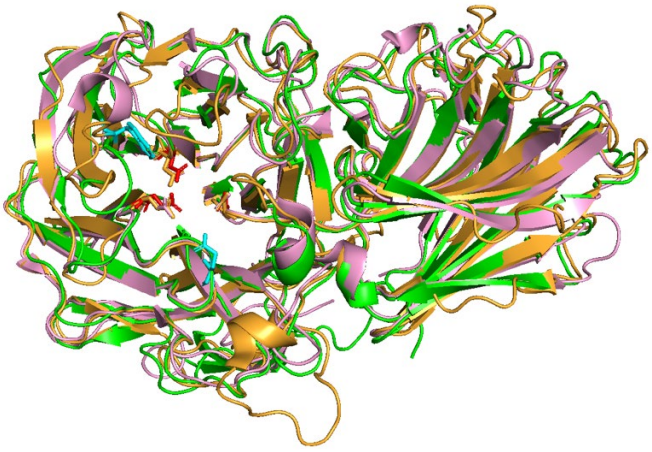


Fig. 2-18. Double-reciprocal plot of the binding of *Px*Xyl43A-UM to oat spelt xylan. The results are shown in adsorption isotherms (a) and double reciprocal plots (b). Error bars represent standard deviation.

(a)



(b)

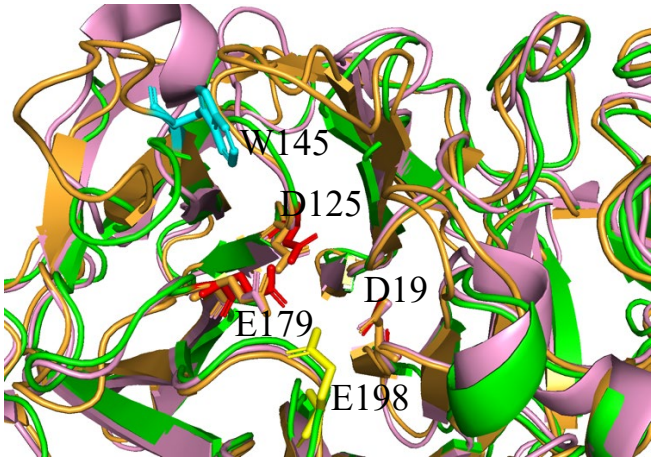


Fig. 2-19. Structures of *Px*Xyl43A, Xyl from *G. thermoleovorans* and Xynb2 from S. *ruminantium*. (a) *Px*Xyl43A (green), Xyl from *G. thermoleovorans* (pink) and Xynb2 from S. *ruminantium* (orange) were overlaid. (b) An enlarged view of a catalytic center of three GH43 enzymes. Each amino acid residue for a catalytic center corresponded. W145 in β-xylosidase SXA from *S. ruminantium* GA192 (cyan) and E198 in β-xylosidase Xyl from *G. thermoleovorans* IT-08 (yellow).

Table 2-3. β-Xylosidases from bacteria of inhibition by D-xylose (Rohman *et al*. 2019).

Bacteria	GH43_subfamily	D-xylose concentration (mM)	Inhibition (%)
Paenibacillus xylaniclasticus TW1	GH43_12	200	30
Corynebacterium alkanolyticum ATCC 21511	GH43_11	20	45
Lactobacillus brevis ATCC14869	GH43_11	100	20
Selenomonas ruminantium GA192	GH43_11	40	57
Lactobacillus brevis ATCC14869	GH43_12	100	66
Paenibacillus woosongensis KCTC 3953	GH43_35	100	25

Table 2-4. K_a values for CBMs from bacteria.

Bacteria	CBM family	K_a	$[PC]_{max}$ (μmol/g)	Ref
		($\times 10^6\ M^{-1}$)		
Clostridium cellulovorans	CBM3	1.7	2.2	Goldstein *et al*. 1993
Clostridium thermocellum	CBM3	3.7	1.9	Zverlof *et al*. 2002
		($\times 10^5\ M^{-1}$)		
Clostridium jossui	CBM3	8.6	12.04	Ichikawa 2011
Clostridium jossui	CBM17	7.2	4.8	Araki *et al*. 2009
Clostridium jossui	CBM28	3.5	5.3	Araki *et al*. 2009
Paenibacillus xylaniclasticus TW1	CBM91	2.0	10.1	
		($\times 10^4\ M^{-1}$)		
Clostridium thermocellum	CBM30	6.4		Arai *et al*. 2003
Clostridium thermocellum	CBM11	7.8		Carvalho *et al*. 2004

Chapter3

Characterization of an AA10 Lytic Polysaccharide Monooxygenase

A C1/C4-oxidizing AA10 Lytic Polysaccharide Monooxygenase from *Paenibacillus xylaniclasticus* strain TW1 (Ito *et al*. in press)

3-1. Summary

Lytic polysaccharide monooxygenases (LPMO) are one of the metalloenzymes, requiring a copper ion in the active site, and oxidoreductases producing oxidized sugars. These are key enzymes to completely degrade lignocellulose biomass composed of cellulose and hemicellulose, boosting lignocellulose degradation with synergy effects cooperating with other cellulases, especially exoglucanases and cellobiohydrolases. *Paenibacillus xylaniclasticus* strain TW1 is one of the promising bacteria for lignocellulose degradation and has many genes encoding cellulolytic enzymes and a gene of LPMO, named *Px*AA10A. An LPMO which was homologous to *Px*AA10A was multifunction and had a boosting effect on lignocellulose by other bacterial cellulases. But, concerning *Px*AA10A, C1/C4-oxidizing selectivity which is an essential feature to characterize LPMOs has yet to be elucidated. In this study, the C1/C4-oxidizing selectivity of *Px*AA10A was analyzed and investigated a boosting effect with a cellulase cocktail from fungi. Both C1-oxidized products and C4-oxidized products were detected as a result of analysis for products with MALDI-TOF/MS, indicating that *Px*AA10A is a C1/C4-oxidizing LPMO. And *Px*AA10A demonstrated a boosting effect with a cellulase cocktail. Therefore, it was suggested that *Px*AA10A produced new ends on cellulose for other cellulases to attack substrates by C1/C4-oxidation, and as well as other LPMOs, boosted lignocellulose degradation.

3-2. Introduction

Lytic polysaccharide monooxygenases (LPMO) are one of the metalloenzymes, requiring a copper ion in the active site, and oxidoreductases producing oxidized sugars. These are key enzymes to completely degrade lignocellulose biomass composed of cellulose and hemicellulose, boosting lignocellulose degradation with synergy effects cooperating with other cellulases, especially exoglucanases and cellobiohydrolases.

P. xylaniclasticus strain TW1 has an LPMO gene encoding *Px*AA10A that belongs AA10 family. It has a similarity to an LPMO gene from *P. curdlanolyticus* strain B-6 (Limsakul *et al*. 2020). The LPMO from *P. curdlanolyticus* strain B-6, named *Pc*AA10A, consisted of an AA10 catalytic domain (CD), a CBM and Fn III domain and cleaved cellulose, xylan and chitin. And, *Pc*AA10A boosted cellulose degradation with other enzymes (Limsakul *et al*. 2020). However, selectivity for the oxidized position of this LPMO was not characterized. In this study, characteristics of *Px*AA10A including C1/C4 selectivity and the products were revealed.

3-3. Material and methods

Strain, plasmid and media.

P. xylaniclasticus strain TW1 was isolated previously from the wastes of a pineapple processing factory in Thailand (Tachaapaikoon *et al*. 2012). The *Escherichia coli* strains ME9806 (iVEC3) (NBRP, Japan) were used as cloning hosts while *E. coli* JM109 (TOYOBO, Japan) were used as protein expression hosts (Nozaki *et al*. 2019). The pMAL-c2 plasmid (New England BioLabs Inc., MA, USA) was used for cloning and expression of recombinant proteins with maltose-binding protein. Transformed *E. coli* JM109 was cultivated in Luria-Bertani broth (LB) supplemented with ampicillin (50 μg/ml). The recombinant proteins of the full length of *Px*AA10A and the AA10A catalytic domain at the *C*-terminus of *Px*AA10A (*Px*AA10A-CD) were expressed using the plasmids pMAL-c2-*Px*AA10A and pMAL-c2-*Px*AA10A-CD, respectively.

Cloning, expression and purification of recombinant proteins of *Px*AA10A.

DNA fragments of *Px*AA10A (1731 bp) and *Px*AA10A-CD (495 bp) were obtained from genomic DNA of *P. xylaniclasticus* strain TW1 by PCR using combinations of primers #1 and #2, and #3 and #4, respectively (Table 3-1). The both DNA fragments were amplified again using primers #5 and #6, and #7 and #8, respectively (Table 3-1). The linear pMAL-c2 for iVEC cloning was amplified using primers #9 and #10 (Table 3-1). The DNA fragments of *Px*AA10A and *Px*AA10A-CD were inserted into pMAL-c2 by in vivo cloning using *E. coli* strain ME9806 (Nozaki *et al*. 2019). Resulted plasmid pMAL-c2-*Px*AA10A and pMAL-c2-*Px*AA10A-CD were sequenced to confirm the absence of mutations.

Transformed *E. coli* strain JM109 cells having pMAL-c2-*Px*AA10A or pMAL-c2-*Px*AA10A-CD were grown overnight at 37°C in LB medium supplemented with ampicillin (50 μg/ml). After several hours of incubation, cells were collected by centrifugation at 12,000 rpm and disrupted with sonication in 20 mM Tris-HCl buffer (including 200 mM NaCl and 100 μM $CuCl_2$: pH 7.4). The recombinant proteins were purified from the cell-free extracts with the pMAL Protein Fusion and Purification System (New England BioLabs Inc., MA, USA) according to the manufacturer's instructions. The purified proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis and the protein concentration was determined using a Protein Assay Kit (Bio-Rad Laboratories, CA, USA) with bovine serum albumin (BSA) as the standard.

After purification, dialysis was performed to remove maltose in the elution buffer for pMAL fusion protein purification. The elution solution was poured into a cellulose tube which was sealed on both ends. The tube was soaked in 50 mM potassium phosphate buffer (pH 7.0) for a day.

Enzyme assay

The activities of *Px*AA10A and *Px*AA10A-CD were measured by incubating with 1% of Avicel as substrates and 100 μM of ascorbic acid or H_2O_2 as the electron donors, in 50 mM potassium phosphate buffer (pH 7.0) at 37˚C for 3 days. The concentration of *Px*AA10A and *Px*AA10A-CD in the reaction mixtures was 1.0 μM and Meicelase was added to the mixtures at 0.4 mg/ml. After enzymatic reactions, reducing sugars were quantified using the 3,5-dinitrosalicylic acid method measured by the absorbance at 570 nm (Miller *et al*. 1959). All enzyme assays were repeated at least three times to confirm

reproducibility. The synergy effect was calculated by dividing the value of the simultaneous reaction by the sum of the value of LPMO and Meicelase.

A measurement using 2,6-dimethoxyphenol (DMP) was performed. Phosphate/succinate buffer (100 mM) was used as a reaction buffer. The reaction buffer (including 1 mM DMP, 100 μL H_2O_2, and 1 μM $CuCl_2$) was prepared and incubated at 37°C for 15 minutes. Then, 20 μL of LPMO solution (final concentration 0.4 μM) was added to the reaction buffer. The increase in absorbance at 469 nm was measured after 5 minutes.

Analysis of products of *Px*AA10A

Cellohexaose (2 mg/mL) (Seikagaku Biobissines, Japan) was incubated with 0.2 μM of the purified *Px*AA10A-CD in 50 μl of 50 mM potassium phosphate buffer (pH7.0) at 37℃ for 3 days. Separation of oxidized products in the reaction mixtures was performed by thin-layer chromatography (TLC) on a DC-Fertigplatten SIL G-25 plate (Macherey-Nage, Germany) developed with butanol-acetic acid-H_2O (1:2:1, v/v/v) solvent mixture. The products were visualized by incubating at 130℃ for several minutes after soaking in methanol (95%)-sulfuric acid (5%) solution.

MALDI-TOF/MS analysis was performed to analyze oxidized products. The products in the corresponding spot on the TLC plate were extracted in dilute water without soaking in methanol-sulfuric acid solution. 0.4 μL of sample solution was mixed with an equal volume of 10 mg/mL α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% trifluoroacetic acid and spotted on Opti-TOF 384 Well Insert plate (Applied Biosystem, US). After air-drying, oxidizing products in the sample solution were analyzed by mass spectrometry on a 4800 Plus MALDI-TOF/TOF Analyzer (Applied Biosystem, US).

3-4. Result

Sequence and Phylogenetic Analysis

Sequence analysis showed that the open reading frame including *Px*AA10A gene consists of 1,845 nucleotides encoding a protein of 615 amino acids (Accession number, WP_246027971). The carbohydrate-active enzyme annotation of *Px*AA10A with dbCAN meta server revealed that *Px*AA10A was composed of 160 amino acids AA10 (40-200) and 80 amino acids CBM3 (474-553) (Fig 3-1, 3-2, 3-3). It was predicted that the signal peptide was 39 amino acids at *N*-terminus with SignalP-5.0. BlastN homology analysis revealed that the amino acids sequence of *Px*AA10A corresponded to that of the AA10 from *P. curdlanolyticus* strain B-6 and had high homology with 68% similarity to the LPMO from *P. woosongensis* (WP_230873677) and 67% similarity to the LPMO from *P. ihumii* (WP_074048692).

Of the AA10 enzymes, 32 enzymes assigned EC numbers were used for phylogenetic analysis (Fig 3-4). The resulting phylogenetic tree indicated that they were divided into two main groups, cellulose-active and chitin-active. The chitin-active group contained mostly chitin-active LPMOs (EC. 1.14.99.3), while the cellulose-active group contained cellulose-active LPMOs (EC. 1.14.99.54 and EC. 114. 99.56) and some chitin-active LPMOs. No cellulose-active LPMOs were distributed around *Px*AA10A.

Cloning, expression and purification of *Px*AA10A.

The amplified DNA encoding *Px*AA10A and *Px*AA10A-CD were 1731 bp and 495 bp, respectively. The molecular sizes of the recombinant proteins with maltose-binding protein of *Px*AA10A and *Px*AA10A-CD were estimated at 103.9 kDa and 61.1 kDa, respectively by SDS-PAGE, corresponding to the predicted molecular sizes (data not

shown). The concentration of purified recombinant proteins of *Px*AA10A and *Px*AA10A-CD were 1.2 mg/ml and 0.9 mg/ml, respectively. After purification for proteins, dialysis was performed to remove maltose mixed into an elution buffer. After dialysis, the activity of *Px*AA10A was measured by TLC. As a result of TLC, *Px*AA10A did not show the activity. When an LPMO with Cu (I) does not bind to substrates, the reduced LPMO generate H_2O_2 by reducing O_2. A reduced LPMO forms reactive oxygen species without substrate, and they damage and inactivate LPMOs (Hegnar *et al*. 2019). Inactivating LPMOs is caused by oxidative modifications of histidine residues in a catalytic center (Bissaro *et al*. 2019). The inactivation would have occurred in *Px*AA10A during dialysis. In subsequent experiments, the protein solution including 10mM maltose was used as *Px*AA10A solution.

LPMO activity and synergistic effect between *Px*AA10A with cellulases

To measure LPMO activity of *Px*AA10A, 2,6-DMP was used as substrates, but the activity of *Px*AA10A was not detected using this assay. To study the activity of *Px*AA10A, Avicel as a substrate and Meicelase as cooperative enzymes, were used. In cases of reaction LPMO only, the reducing sugars were no detective. When LPMO and Meicelase were added, the activity increased (Fig. 3-5, 3-6). The synergy effects for *Px*AA10A-CD and *Px*AA10A were 1.1-fold and 1.3-fold, respectively when ascorbic acids were used as a reductant (Fig. 3-5). When H_2O_2 was used as a reductant, the synergy effects for *Px*AA10A-CD and *Px*AA10A were 1.1-fold and 1.4-fold, respectively (Fig. 3-6). The effect for *Px*AA10A was higher than that for *Px*AA10A-CD. *Px*AA10A is full length and has a CBM which binds to substrates, suggesting that a CBM would affect the synergy effect.

Analysis of products of *Px*AA10A

To investigate oxidized products by *Px*AA10A-CD, the reaction mixtures were analyzed by TLC and mass spectrometry using a 4800 Plus MALDI-TOF/TOF Analyzer. In the result of TLC, when cellohexaose was used as the substrate, one of the products appeared in a position corresponding to cellotetraose. Therefore, *Px*AA10A recognized cellohexaose and cleaved it into cellotetraose and cellobiose (Fig. 3-7). To determine the C1/C4 selectivity of *Px*AA10A-CD, MALDI-TOF/MS analysis of the resulting cellotetraose was performed. As a result, Both C1-oxidized products, aldonic acid, and C4-oxidized products, cellotetraose, were detected (Fig. 3-8).

3-5. Discussion

Sequence and phylogenetic analysis of *Px*AA10A

In this study, one of the LPMO genes named *Px*AA10A from *P. xylaniclasticus* strain TW1 was focused on. An amino acids sequence analysis revealed that the amino acid sequence of *Px*AA10A corresponded to that of *Pc*AA10A from *P. curdlanolyticus* strain B6, suggesting that *Px*AA10A also had the same characteristics (Limsakul *et al*. 2020). *Pc*AA10A has broad substrate specificity and is active against cellulose, xylan and mannnan. LPMOs which have similarities to *Px*AA10A and *Pc*AA10A also would exhibit multifunction. *Px*AA10A was also active against cellulose, confirming a synergistic effect with the hydrolytic enzyme, as well as *Pc*AA10A. But, The C1/C4-oxidizing selectivity for *Px*AA10A has not been analyzed in previous studies, so it was analyzed in this study.

To analyze AA10 phylogenetically for *Px*AA10A, 32 enzymes including these AA10 family enzymes were used for phylogenetic analysis (Fig. 3-4). The phylogenetic tree indicated that they were divided into two groups: Clade A and Clade B. Most LPMOs in clade A are chitin-active (EC 1.14.99.53), and most LPMOs in clade B are cellulose-active (EC 1.14.99.54 and EC 1.14.99.56). Some LPMOs, however, are active on both chitin and cellulose. An LPMO (*Bl*AA10A) in Clade A (the chitin active group) was active against cellulose. *Pc*AA10A was also active against chitin and cellulose. *Px*AA10A is homologous to *Pc*AA10A, and the substrate specificity of *Px*AA10A would be consistent with *Pc*AA10A. In this study, *Px*AA10A was active against cellulose, but not against chitin (data not shown). This suggests that the maltose-binding protein added to the *N*-terminus of r*Px*AA10A may have affected this enzyme activity. However, some

chitin-active LPMOs are active against cellulose, and it is reasonable to assume that *Px*AA10A is active against cellulose.

C1/C4-selectivity of *Px*AA10A

*Px*AA10A cut cellohexaose and produced cellotetraose as confirmed by TLC (Fig. 3-7). In a binding simulation using an LPMO and chitohexaose, *Sm*AA10A which was phylogenetically close to *Px*AA10A had six substrate-binding subsites ranging from -4 to +2 (Bissaro *et al*. 2018). Even when cellulosic substrates are used for *Sm*AA10A, *Sm*AA10A is likely to bind cellohexaose, then cut it into cellobiose and cellotetraose. One of the major products of *Px*AA10A was cellotetraose, which is consistent with the result of a binding simulation of *Sm*AA10A.

There are three possible sugars produced from C1/C4-oxidizing LPMOs; aldonic acids whose reducing ends are oxidized, gem-diol whose non-reducing ends are oxidized, and double oxidized products with an aldonic acid in one end and a gem-diol in the other end (Fig. 3-9). It was considered that the double-oxidized products would not be produced since the products analyzed in this study would be produced by cleavage at once. Therefore, the products detected in this analysis would be aldonic acids from C1-oxidizing and/or cellotetraose from C4-oxidizing. *Px*AA10A produced C1-oxidizing and C4-oxidizing ends because aldonic acids and cellotetraose were detected when the products were analyzed with MALDI-TOF/MS (Fig 3-8). Most bacterial LPMOs are classified into C1-oxidizing or C1/C4-oxidizing LPMOs, and no C4-oxidizing LPMOs from bacteria were known yet (Zhou *et al*. 2019). It is reasonable to consider *Px*AA10A as a C1/C4-oxidizing LPMO.

LPMO activity and synergistic effect of *Px*AA10A with cellulases

To measure LPMO activity of *Px*AA10A, 2,6-DMP was used as substrates. In 2,6-DMP oxidation, LPMOs oxidize 2,6-DMP to the corresponding phenoxy radical. The active-site Cu (II) is reduced by 2,6-DMP which generates the 2,6-DMP radical. Two formed 2,6-DMP radicals dimerize rapidly to hydrocoerulignone, which again is quickly converted to coerulignone by LPMO. (Breslmayr *et al*. 2018). This method was known as a fast and sensitive assay for LPMO activities. In the present study, the assay was used to measure LPMO activity of *Px*AA10A. The activity of *Px*AA10A was not detected using this assay. When the activity of *Px*AA10A was measured, Cu (II) had an effect on it. This assay is sensitive and only Cu (II) would oxidize 2,6-DMP to coerulignone without LPMO. It was reported that very low amounts of free copper are extremely sensitive because reductants reduce free Cu (II) ions (Stepnov *et al*. 2021). 2,6-DMP method was not adopted in this study, then synergy effects were used to measure the activity of *Px*AA10A.

For LPMOs, the C1/C4-oxidizing selectivity is important concerning the synergy effects with cellobiohydrolases (CBH) and/or exoglucanases that occur after an LPMO oxidizing reaction (Keller *et al*. 2012). One of the C4-oxidized ends of cellulose chains is a reducing end, while the C1-oxidized ends of these are aldonic acid (Forsberg *et al*. 2014). The reducing ends that C4-oxidizing LPMOs produced are recognized by typically reducing end-specific cellulases, such as CBH I from *T*. *reesei*. C1-oxidized ends with aldonic acid inhibit a synergistic effect with CBH I (Keller *et al*. 2021). The LPMO itself did not impede CBH I activity, but the C1-oxidized end produced by an LPMO affects the synergistic effect (Keller *et al*. 2021). Therefore, insights into C1/C4-oxidizing selectivity would provide a clue for promoting substrate degradation activity in combination with other enzymes.

In the present study, Meicelase was used as a representative of enzyme cocktails including a β-glucosidase and a CBH from *T. viride*. The major element in Meicelase is considered to be a CBH. Generally, LPMOs boost the degradation of cellulosic substrates with exoglucanases and CBH (Keller *et al*. 2021). As a result of a synergy effect between *Px*AA10A and Meicelase, the synergy effects for *Px*AA10A-CD and *Px*AA10A were 1.1-fold and 1.3-fold, respectively when ascorbic acids were used as a reductant. When H_2O_2 was used as a reductant, the synergy effects for *Px*AA10A-CD and *Px*AA10A were 1.1-fold and 1.4-fold, respectively. (Fig. 3-5, 3-6). the synergy effects were higher when H_2O_2 was used than when ascorbic acids were used. H_2O_2 boosts LPMO activity and cellulose degradation (Stepnov *et al*. 2021). Therefore, Similarly for *Px*AA10A, H_2O_2 was suitable for LPMO activity.

In cases of other bacterial LPMOs, *Tf*AA10A and *Tf*AA10B from *Thermobifida fusca* XY increased cellobiose production and reduced sugars production with exoglucanases by up to approximately 20% to 60% (Arfi *et al*. 2014; Moser *et al*. 2008). LPMOs from *Streptomyces coelicoler* and *Hahella chejuensis* boosted cellulosic substrate degradation (Forsberg *et al*. 2011; Ghatge *et al*. 2015).

LPMOs have been added to commercial enzymes and have contributed to boosting substrate degradation. Most of the LPMOs that were added to commercial enzymes were AA9 which was derived from filamentous fungi, for example, *T. reesei*. Also, in experiments using a combination of commercial enzymes and LPMOs, AA9 is often used (Corrêa *et al*. 2015). On the other hand, a bacterial LPMO *Hc*AA10-2 from *H. chejuensis* enhanced the hydrolysis of crystalline cellulose with Celluclast, one of the commercial cellulase cocktails (Ghatge *et al*. 2015). *Kp*LPMO10A from *Kitasatospora papulosa* increased the release of reducing sugars with Celluclast (Corrêa *et al*. 2019). In this study, it was indicated that *Px*AA10A boosted the degradation

with a commercial enzyme cocktail Meicelase, containing cellulases, as well as other bacterial AA10 LPMOs.

Figures and Tables

Table 3-1. List of primers.

#1	CATGGATGGATTGAAGAGAGC
#2	TTAGCGCGTCAAGTTAACATCAATCG
#3	CATGGATGGATTGAAGAGAGC
#4	TTAAGGAGTCGTTCCCCACACCAGCG
#5	CAACAACCTCGGGATCGAGGGAAGGCATGGATGGATTGAAGAGAG
#6	TCTGTATCAGGCTGAAAATCTTATCTTAGCGCGTCAAGTTAACAT
#7	TCTGTATCAGGCTGAAAATCTTATCTTAGCGCGTCAAGTTAACAT
#8	TCTGTATCAGGCTGAAAATCTTATCTTAAGGAGTCGTTCCCCACA
#9	CCTTCCCTCGATCCCGAGGTTG
#10	GATAAGATTTTCAGCCTGATACAGA

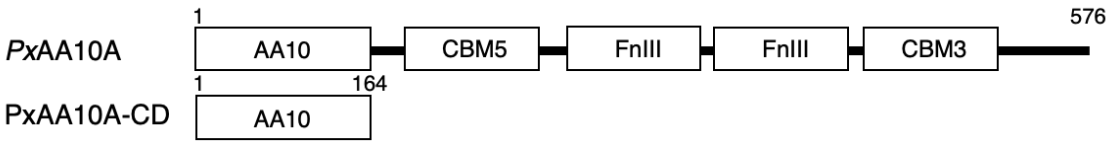


Fig. 3-1. The domain structure of *Px*AA10A and the phylogenetic analysis with AA10 enzymes. Schematic diagram of *Px*AA10A and *Px*AA10A-CD. The numbers represent the number of amino acids. *Px*AA10A was composed of 160 amino acids AA10 (1-160), 43 amino acids CBM5 (186-228), two Fn III domains (242-331 and 336-426) and 80 amino acids CBM3 (434-513).

Fig. 3-2. Multiple protein aliment of *Px*AA10A, *Pc*LPMO10A from *P. curdlanolyticus* (AIY28331), *Lm*LPMO10 from *Listeria monocytogenes* (AEO07443) and *Jd*LPMO10A from *Jonesia denitrificans* (ACV09037). Each aliment was obtained by using NCBI and aligned by using Clustal Omega.

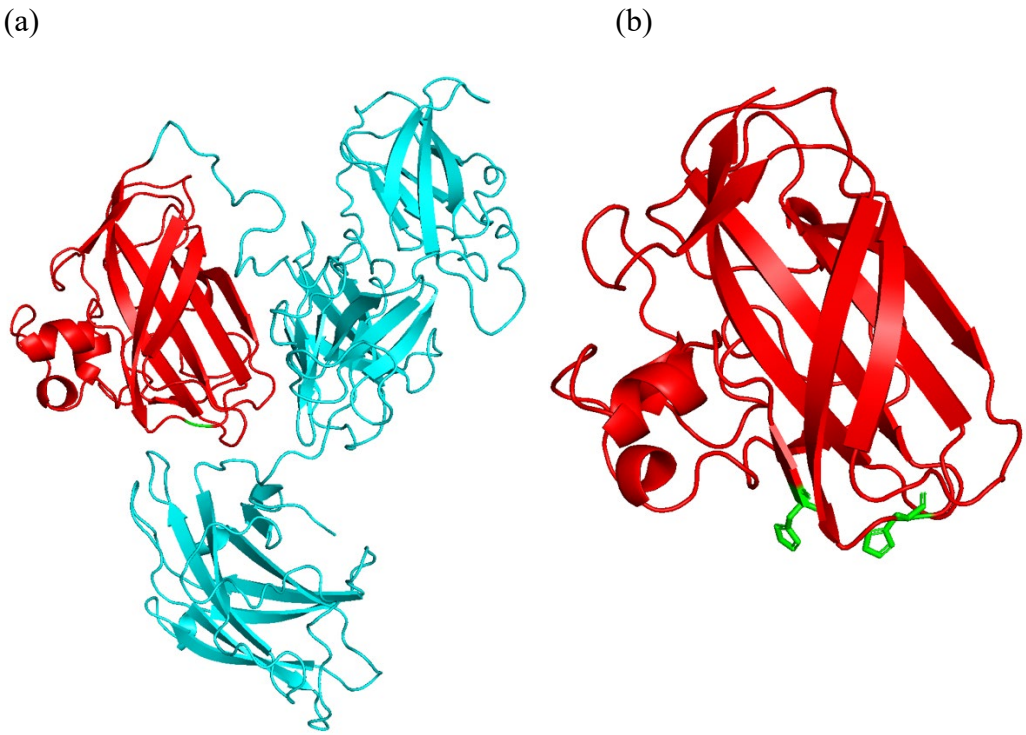


Fig. 3-3. Structure of *Px*AA10A. The prediction structure was performed by using Alpha Fold 2 and the structure was built by Pymol. (a) *Px*AA10A is composed of the catalytic domain of AA10 (red), CBM5, two Fn III domains and CBM3 (cyan). (b) The catalytic domain of AA10 (red) and amino acid residues of the catalytic center (green).

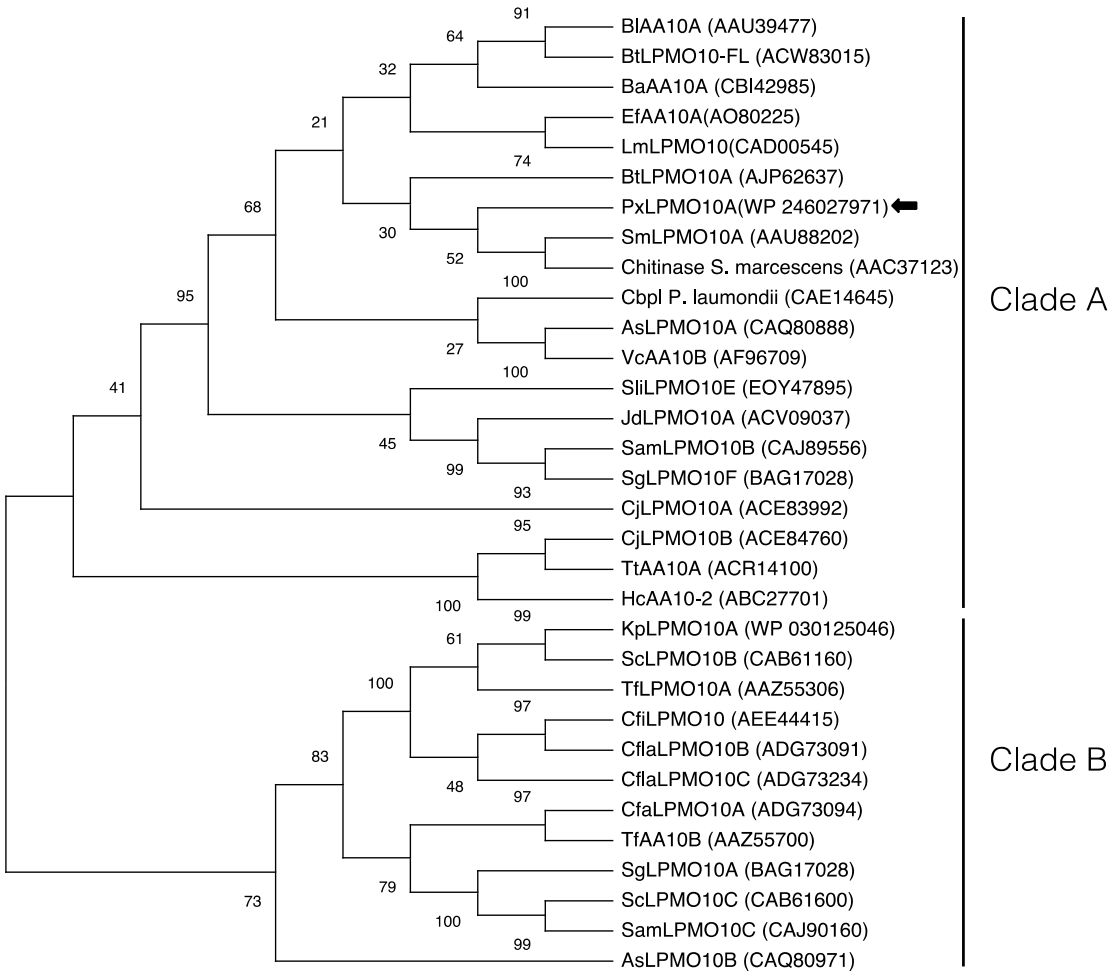


Fig. 3-4. The phylogenetic tree of bacterial LPMOs. Thirty-two LPMOs characterized for substrate specificity were analyzed. The members in clade A and clade B are chitin-active and cellulose-active, respectively. The arrow points out *Px*AA10A. In a cellulose-active group, C1-oxidizing (EC 3.14.99.54) and C4-oxidizing (EC 3.14.99.56) LPMO were included.

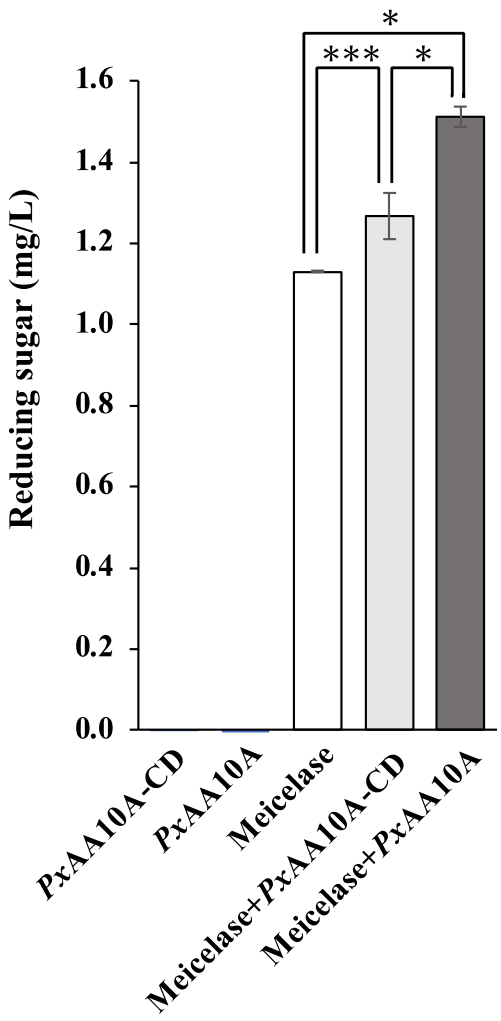


Fig. 3-5. Synergy effects between Meicelase and *Px*AA10A with ascorbic acids. To study synergy effects, Meicelase was added with *Px*AA10A in reaction mixtures. After reactions, released reducing sugars from Avicel were measured using the DNS method. Values are the means of triplicate experiments. Error bars represent standard deviation. Statistical analysis was performed using Tukey's test. Asterisks indicate a significant difference (* : $p < 0.01$, *** : $p < 0.10$).

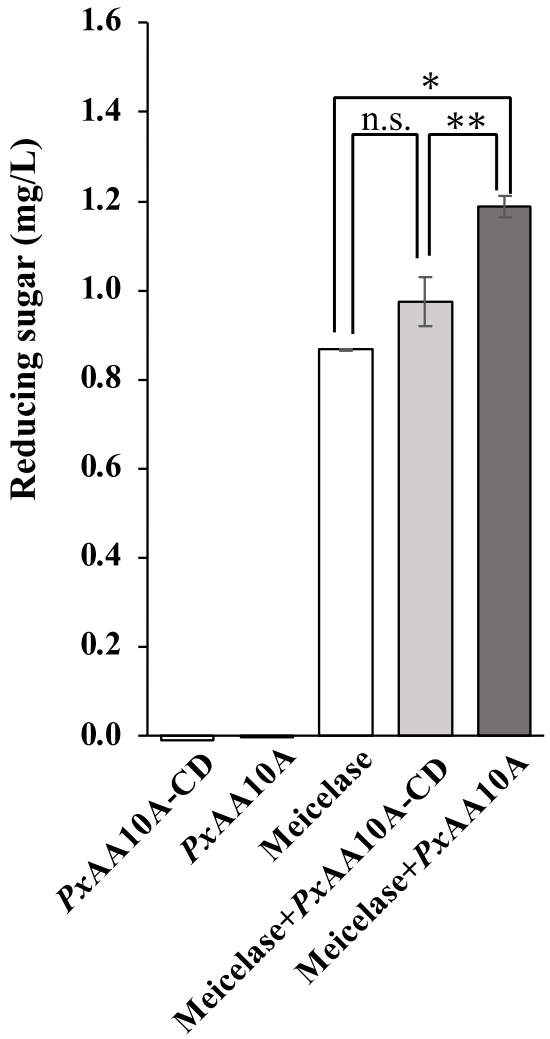


Fig. 3-6. Synergy effects between Meicelase and *Px*AA10A with H_2O_2. To study synergy effects, Meicelase was added with *Px*AA10A in reaction mixtures. After reactions, released reducing sugars from Avicel were measured using the DNS method. Values are the means of triplicate experiments. Error bars represent standard deviation. Statistical analysis was performed using Tukey's test. Asterisks indicate a significant difference (* : $p < 0.01$, ** : $p < 0.05$), and n.s. indicates no significant difference.

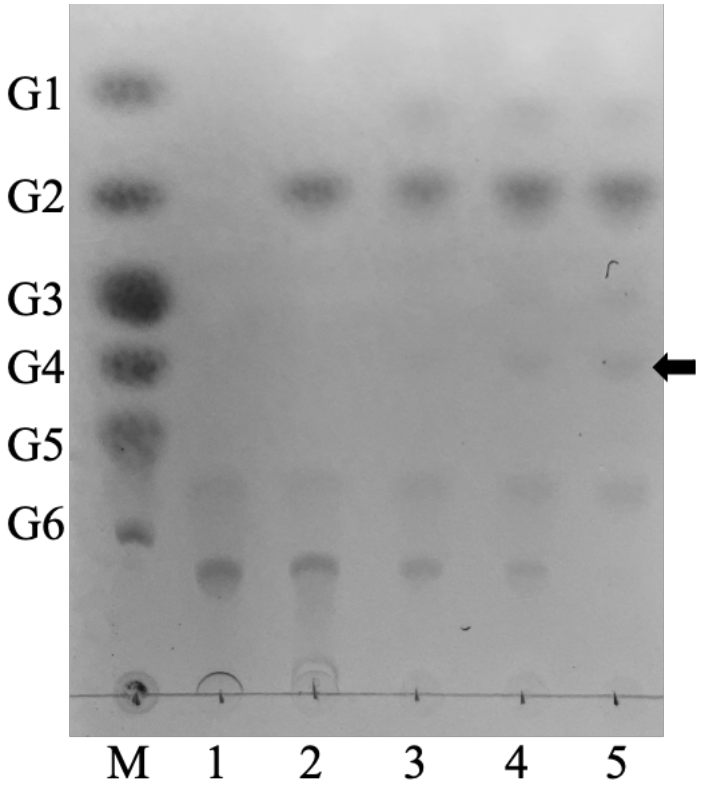


Fig. 3-7. TLC of hydrolysis products of *Px*AA10A released from cellohexaose. Lane M, standard Glucose to cellohexaose (G1 - G6); lane 1, control contained no *Px*AA10A; lane 2, reaction after 0h; lane 3, reaction after 1d; lane 4, reaction after 2d; lane 5, reaction after 3d. The arrow points out the C4 products.

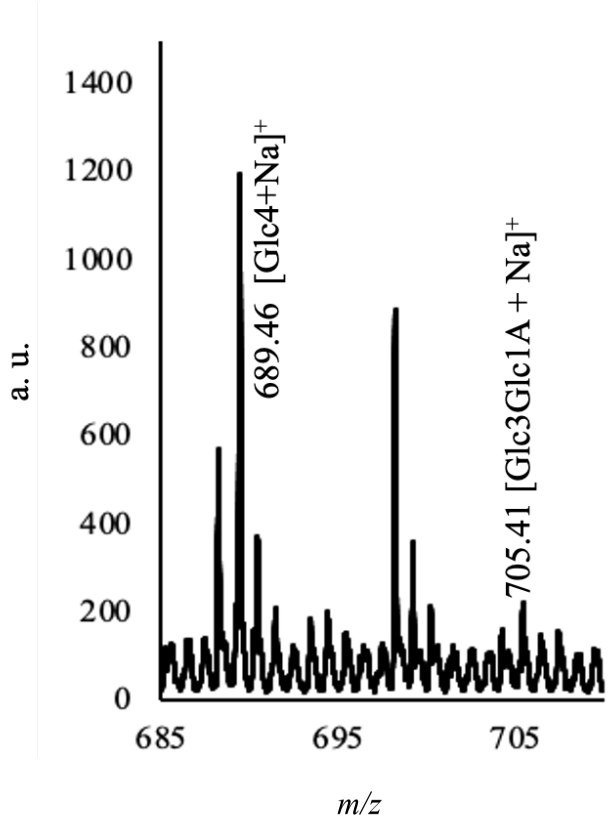


Fig. 3-8. Mass spectra of aldonic acid cellotetraose and cellotetraose. Each product was identified by the major peak of $[M+Na]^+$ adducts. Glc4 represents cellotetraose from a C4-oxidizing reaction, and Glc3Glc1A represents aldonic acid cellotetraose.

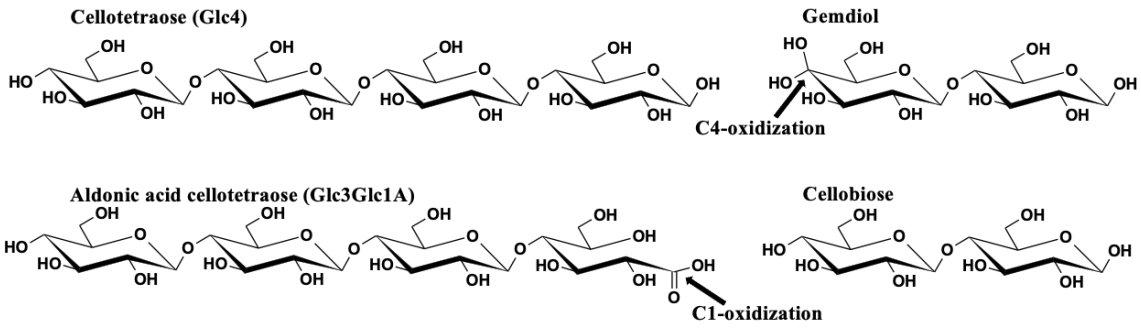


Fig. 3-9. Schematic diagram of the products from C1- and C4-oxidization. Cellohexaose was separated into cellotetraose (Glc4) and gem-diol on C4-oxidization, aldonic acid cellotetraose (Glc3Glc1A) and cellobiose on C1-oxidization. In this study, cellotetraose (Glc4) and aldonic acid cellotetraose (Glc3Glc1A) were detected by MALDI-TOF/MS analysis.

Chapter 4
General Discussion

P. xylaniclasticus strain TW1, a Gram-positive facultative anaerobic bacterium, was isolated as a xylanolytic microorganism from the wastes of a pineapple-processing factory. TW1 grows on xylan as the sole carbon source and has many genes of carbohydrate-active enzymes, including genes for at least 29 glycoside hydrolases (GH), 11 carbohydrate esterases, a polysaccharide lyase, a lytic polysaccharide monooxygenase and 24 carbohydrate-binding modules (CBMs) according to the draft genome sequence. In this study, one of the β-xylosidases and a lytic polysaccharide monooxygenase from *P. xylaniclasticus* strain TW1 were characterized.

A β-xylosidase from *P. xylaniclasticus* strain TW1, named *Px*Xyl43A, was composed of a GH43 subfamily 12 catalytic module and an unknown module (UM). *Px*Xyl43A exhibited hydrolysis activity against both *p*-nitrophenyl-β-D-xylopyranoside (*p*NPX) and *p*-nitrophenyl-α-L-arabinofuranoside. The optimal reaction parameters for *p*NPX hydrolysis were pH 7.1 and 54˚C. At pH 7.0 and 54˚C, the K_m and k_{cat} for *p*NPX were 1.2 mM and 2.8 ± 0.15 s^{-1}, respectively. The unknown module of *Px*Xyl43A (*Px*Xyl43A-UM) could bind to insoluble xylans such as birchwood xylan and oat spelt xylan. The binding constant value K_a of *Px*Xyl43A-UM for oat spelt xylan was 2.0×10^{-5} M^{-1}. These results suggest that *Px*Xyl43A-UM is a novel xylan-specific CBM. Therefore, I proposed this module as a novel CBM in the international CAZy database. This proposal was accepted and created CBM family 91.

Lytic polysaccharide monooxygenases (LPMO) are key enzymes to completely degrade lignocellulose biomass with synergistic effects cooperating with other cellulases. *P. xylaniclasticus* strain TW1 has a gene for LPMO, named *Px*AA10A. The C1/C4-

oxidizing selectivity of *Px*AA10A was determined using MALDI TOF spectrometry of the reaction products and the synergistic effect was investigated with a cellulase cocktail, Meicelase from *T. viride*. The product analysis revealed that both C1-oxidized products and C4-oxidized products were generated, indicating that *Px*AA10A was a C1/C4-oxidizing LPMO. And *Px*AA10A demonstrated a synergistic effect with Meicelase. Therefore, it was suggested that *Px*AA10A degraded the substrates by C1/C4-oxidation, and boosted plant cell wall degradation.

Much research on cellulose degradation has been conducted over the past decades, and the mechanisms have been gradually elucidated. First, it was found that *T. reesei* had a strong ability for cellulose degradation. From this discovery, it was expected that two elements would be needed to degrade cellulose. This hypothesis, called the "C_1-C_x hypothesis," became an established theory. C_x would be a glycoside hydrolase, and C_1 would be an element that does not have hydrolase activity but that hydrates cellulose (Reese *et al*. 1950). However, although decades have passed since the C_1-C_x hypothesis was introduced, C_1 has not been discovered. Subsequently, it became widely accepted that the synergistic effects of endo- and exo-type cellulases are essential for cellulose degradation (Wood *et al*. 1972). In this mechanism, endo-type cellulases first cut the cellulose chain, and then exo-type cellulases recognize the end which is produced by endo-type cellulases and begin to decompose cellulose from the end. These two types of cellulases work efficiently and create a synergistic effect. Following on the acceptance of this mechanism, the "endo-exo hypothesis" was put forward to explain the process of cellulose degradation. This hypothesis was also boosted by a key discovery in the field of cellulose degradation (Eriksson *et al*. 1974; Vaaje-Kolstad *et al*. 2010)—namely, the discovery of an oxidoreductase, LPMO. LPMO was identified based on the observation that cellulose degradation was more enhanced under aerobic conditions than under

anaerobic conditions. LPMO, like endo-type cellulases, promotes cellulose degradation by cleaving the cellulose chain to create new ends. Moreover, a recent study revealed that LPMOs degrade cellulose chains, loosening the overall structure of the cellulose crystals, and that this loosening promotes their degradation (Uchiyama *et al*. 2022). To summarize the research on cellulose degradation to date, three types of enzymes are essential for cellulose degradation: endo-type cellulases, exo-type cellulases, and LPMOs. C_1, in the C_1-C_x hypothesis, can be thought of as an LPMO, and LPMOs also work as one of the endo-type enzymes. C_x can be thought of as endo-type cellulase and exo-type cellulase, each of which works in a coordinated manner. This mechanism can be considered a function of many cellulose-degrading microorganisms, and *P. xylaniclasticus* strain TW1 would be no exception.

P. xylaniclasticus strain TW1 has many cellulolytic enzyme genes as well as xylan-degrading enzymes, and it would use them to efficiently degrade cellulosic biomass. More importantly, natural substrates have hemicellulose intertwined with the outer cellulose, and their degradation must begin with the degradation of hemicellulose. Here, the strategy of cellulose biomass decomposition in *P. xylaniclasticus* strain TW1 is expected (Fig. 4-1). First, xylanolytic enzymes degrade hemicellulose. Xylanases belonging to GH8, GH10 and GH11 cleave xylan, and GH74 xyloglucanases remove xylan units from xyloglucan. GH43 β-xylosidases, moreover, separate xylobiose, which is released from xylan by xylanases into xylose units. Then, once the xylan is removed and the cellulose is exposed, cellulases and LPMOs synergistically degrade the cellulose. The cellulases from *P. xylaniclasticus* strain TW1 are GH5, GH9 and GH48 cellulases. GH48 cellulases would work as exo-type cellulases, and GH5 and GH9 cellulases would function as endo-type cellulases.

In fact, the mechanism may be even more complex, which would be due to the diverse modifications of hemicellulose. Enzymes belonging to GH43 have not only β-D-xylopyranoside activity, but also α-L-arabinofranosidase activity, and some exhibit multiple functions. This feature would be essential for removing hemicellulose, degrading cellulose, and solubilizing cellulosic biomass. Ruminants that feed on cellulosic biomass have large numbers of cellulolytic bacteria in their rumen. The degradation rate in the rumen is very high, and cellulosic biomass is degraded very efficiently by cellulolytic bacteria in the rumen. Metagenomic analysis shows that genes for hemicellulolytic enzymes are highly expressed in the rumen (Wang *et al*. 2013). Enzymes belonging to GH43 are also highly expressed in the rumen and would play an important role in the degradation of cellulosic biomass therein.

In this study, two plant cell wall-degrading enzymes from *P. xylaniclasticus* strain TW1 were characterized. *Px*Xyl43A would be an important xylanase for *P. xylaniclasticus* strain TW1 to saccharify xylan to xylooligosaccharides for their carbon source. Moreover, a novel CBM, CBM91, from *Px*Xyl43A was found. This provided new insights into protein-carbohydrate interaction on how xylanolytic enzymes recognize carbohydrates and revealed that cellulolytic microorganisms produce a variety of modular enzymes and one of the mechanisms that they utilize these enzymes.

Investigation into plant cell wall degradation is still at a challenging stage and studies should continue. LPMOs are expected to promote plant cell wall degradation, but there is still little knowledge about LPMOs from bacteria. In the present study, *Px*AA10A was characterized and found to promote cellulose degradation, suggesting that *Px*AA10A probably contributes profoundly to the plant cell wall degradation in *P. xylaniclasticus* strain TW1.

This study has elucidated part of the mechanism for the plant cell wall-degrading capacity of *P. xylaniclasticus* strain TW1 and provided new enzymatic insights into the field of plant cell wall degradation.

Figure

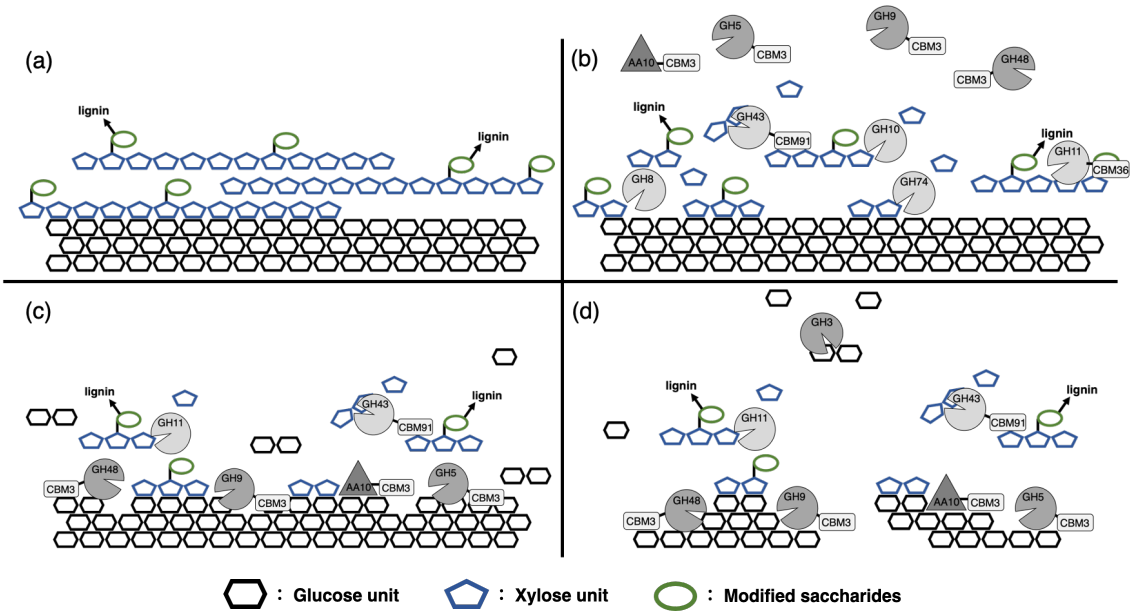


Fig. 4-1. The strategy of plant cell wall biomass degradation for *P. xylaniclasticus* strain TW1. (a) Plant cell wall include cellulose, hemicellulose and lignin. (b) GH8, GH10 and GH11 which have xylanase activities cleave xylan and a xyloglucanase GH74 removes xylan units from xyloglucan. GH43, moreover, separates xylobiose which is released from xylan by xylanases into xylose units. (c, d) Once the xylan is removed and the cellulose is exposed, cellulase and LPMO synergistically degrade the cellulose.

Reference

Abbott, D. W., A. L. van Bueren: Using structure to inform carbohydrate-binding module function. *Curr*. *Opin*. *Struct*. *Biol*., **28**, 32-40 (2014).

Alhassid, A., A. Ben-David, O. Tabachnikov, D. Libster, E. Naveh, G. Zolotnitsky, Y. Shoham, G. Shoham: Crystal structure of an inverting GH 43 1,5-α-L-arabinanase from *Geobacillus stearothermophilus* complexed with its substrate. *Biochem*. *J*., **422**, 73–82 (2009).

Arai T., R. Araki, A. Tanaka, S. Karita, T. Kimura, K. Sakka, K. Ohmiya: Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. *J*. *Bacteriol*. **185**, 504-512 (2003).

Araki, Y., S. Karita, A. Tanaka, M. Kondo, M. Goto: Characterization of family 17 and family 28 carbohydrate-binding modules from *Clostridium josui* Cel5A. *Biosci*. *Biotechnol. Biochem*., **73**, 1028–1032 (2009).

Arfi, Y., M. Shamshoum, I. Rogachev, Y. Peleg, E. A. Bayer: Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc. Natl. Acad. Sci. US*A, **111**, 9109-9114 (2014).

Bischof, R. H., J. Ramoni, B. Seiboth: Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. *Microb. Cell Fact.*, **15**, 106 (2016).

Bissaro, B., Å. K. Røhr, G. Müller, P. Chylenski, M. Skaugen, Z. Forsberg, S. J. Horn, G. Vaaje-Kolstad, V. G. H. Eijsink: Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2. *Nat*. *Chem*. *Biol*., **13**, 1123-1128 (2017).

Bissaro, B., I. Isaksen, G. Vaaje-Kolstad, V. G. H Eijsink, Å.K. Røhr: How a lytic polysaccharide monooxygenase binds crystalline chitin. *Biochemistry*, **57**, 1893-1906 (2018).

Boraston, A. B., D. N. Bolam, H. J. Gilbert, G. J. Davies: Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem*. *J*., **382**, 769–781 (2004).

Breslmayr, E., M. Hanžek, A. Hanrahan, C. Leitner, R. Kittl, B. Šantek, C. Oostenbrink, R. Ludwig: A fast and sensitive activity assay for lytic polysaccharide monooxygenase. *Biotechnol*. *Biofuels*, **11**, 79 (2018).

Brüx, C., A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, Y. Shoham, D. Schomburg: The structure of an inverting GH43 β-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J*. *Mol*. *Bio*l., **359**, 97–109 (2006).

Carvalho, A.L., A. Goyal, J. A. Prates, D. N. Bolam, H. J. Gilbert, V. M. Pires, L. M. Ferreira, A. Planas, M. J. Romão, C. M. Fontes: The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *J. Biol. Chem.*, **279**, 34785-34793 (2004).

Charnock, S. J., D. N. Bolam, D. Nurizz, L. Szabó, V. A. McKie, H. J. Gilbert, D. J. Davies: Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc*. *Natl*. *Acad*. *Sci*. *USA*, **99**, 14077-14082 (2002).

Cicortas Gunnarsson, L., C. Montanier, R. B. Tunnicliffe, M. P. Williamson, H. J. Gilbert, E. Nordberg Karlsson, M. Ohlin: Novel xylan-binding properties of an engineered family 4 carbohydrate-binding module. *Biochem*. *J*., **406**, 209–214 (2007).

Couturier, M., S. Ladevèze, G. Sulzenbacher, L. Ciano, M. Fanuel, C. Moreau, A. Villares, B. Cathala, F. Chaspoul, K. E. Frandsen, A. Labourel, I. Herpoël-Gimbert, S. Grisel, M. Haon, N. Lenfant, H. Rogniaux, D. Ropartz, G. J. Davies, M. N. Rosso, P. H. Walton, B. Henrissat, J. G. Berrin: Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nat. Chem. Biol*., **14**, 306-310 (2018).

Corrêa, T. L. R., L. V. dos Santos, G. A. Pereira: AA9 and AA10: from enigmatic to essential enzymes. *Appl. Microbiol. Biotechnol.*, **100**, 9-16 (2016).

Corrêa, T. L. R., A. T. Júnior, L. D. Wolf, M. S. Buckeridge, L. V. Dos Santos, M. T. Murakami: An actinobacteria lytic polysaccharide monooxygenase acts on both cellulose and xylan to boost biomass saccharification. *Biotechnol. Biofuels*, **10**, 117 (2019).

Chen, Z., Y. Liu, Q. Yan, S. Yang, Z. Jiang: Biochemical characterization of a novel endo-1,5-α-L-arabinanase from *Rhizomucor miehei*. *J. Agric*. *Food Chem*., **63**, 1226–1233 (2015).

Das, S. P., R. Ravindran, S. Ahmed, D. Das, D. Goyal, C. M. G. A Fontes, A. Goyal: Bioethanol production involving recombinant *C. thermocellum* hydrolytic hemicellulase and fermentative microbes. *Appl*. *Biochem*. *Biotechnol*., **167**, 1475–1488 (2012).

de Camargo, B. R., N. J. Claassens, B. F. Quirino, E. F. Noronha, S. W. M. Kengen: Heterologous expression and characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Clostridium thermocellum* B8. *Enzyme Microb*. *Technol*., **109**, 74–83 (2018).

Drula, E., M. L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon: The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.*, **50**, D571–577 (2022).

Eriksson, K. E., B. Pettersson, U. Westermark: Oxidation: an important enzyme reaction in fungal degradation of cellulose. *FEBS Lett*., **49**, 282-285 (1974).

Fan, Z., L. Yuan, D. B. Jordan, K. Wagschal, C. Heng, J. B. Braker: Engineering lower inhibitor affinities in β-D-xylosidase. *Appl*. *Microbiol*. *Biotechnol*., **86**, 1099–1113 (2010).

Filiatrault-Chastel, C., D. Navarro, M. Haon, S. Grisel, I. Herpoël-Gimbert, D. Chevret D, M. Fanuel, B. Henrissat, S. Heiss-Blanquet, A. Margeot, J. G. Berrin: AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnol. Biofuels*, **12**, 55 (2019).

Forsberg, Z., G. Vaaje-Kolstad, B. Westereng, A. C. Bunæs, Y. Stenstrøm, A. MacKenzie, M. Sørlie, S. J. Horn, V. G. Eijsink: Cleavage of cellulose by a CBM33 protein: *Protein Sci*., **20**, 1479-1483 (2011).

Forsberg, Z., A. K, Mackenzie, M. Sørlie, A. K. Røhr, R. Helland, S. V. Arvai, G. Vaaje-Kolstad, V. G. Eijsink: Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. USA,* **111**, 8446-8451 (2014).

Furtado, G. P., M. R. Lourenzoni, C. A. Fuzo, R. Fonseca-Maldonado, M. E. Guazzaroni, L. F. Ribeiro, R. J. Ward: Engineering the affinity of a family 11 carbohydrate-binding module to improve binding of branched over unbranched polysaccharides. *Int. J. Biol. Macromol.*, **120**, 2509–2516 (2018).

Ghatge, S. S., A. A. Telke, T. R. Waghmode, Y. Lee, K. W. Lee, D. B. Oh, H. D. Shin, S. W. Kim: Multifunctional cellulolytic auxiliary activity protein *Hc*AA10-2 from *Hahella chejuensis* enhances enzymatic hydrolysis of crystalline cellulose. *Appl. Microbiol. Biotechnol.*, **99**, 3041-3055 (2015).

Gilbert, H. J., L. P. Knox, A. B. Boraston: Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.*, **23**, 669-677 (2013).

Gilkes, N. R., E. Jervis, B. Henrissat, B. Tekant, R. C. Miller, R. A. J. Warren, D. G. Kilburn: The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *J. Biol. Chem.*, **267**, 6743- 6749 (1992).

Goldstein, M. A., M. Takagi, S. Hashida, O. Shoseyov, R. H. Doi, I. H. Segel: Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein A. *J. Bacteriol.*, **175**, 5762-5768 (1993).

Guo, J., J. M. Catchmark: Binding specificity and thermodynamics of cellulose-binding modules from *Trichoderma reesei* Cel7A and Cel6A. *Biomacromolecules*, **14**, 1268–1277 (2013).

Hamelinck, C. N., G. van Hooijdonk, A. P. C. Faaij: Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass Bioenergy*, **28**, 384-410 (2005).

Harris, P. V., D. Welner, K. C. McFarland, E. Re, J. C. Navarro Poulsen, K. Brown, R. Salbo, H. Ding, E. Vlasenko, S. Merino, F. Xu, J. Cherry, S. Larsen, L. Lo Leggio: Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry*, **49**, 3305-3316 (2010).

Hegnar, O. A., D.M. Petrovic, B. Bissaro, G. Alfredsen, A. Várnai, V. G. H. Eijsink: pH-Dependent relationship between catalytic activity and hydrogen peroxide production shown via characterization of a lytic polysaccharide monooxygenase from *Gloeophyllum trabeum*. *Appl*. *Environ*. *Microbiol*., **85**, e02612-18 (2019).

Hemsworth, G. R., E. J. Taylor, R. Q. Kim, R. C. Gregory, S. J. Lewis, J. P. Turkenburg, A. Parkin, G.J. Davies, P. H. Walton: The copper active site of CBM33 polysaccharide oxygenases. *J. Am. Chem. Soc.*, **135**, 6069-6077 (2013).

Henrissat B.: A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem*. *J*., **280**, 309–316 (1991).

Himmel, M. E., S. Ding, D. K. Johnson, W. S. Adney: Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, **315**, 804–808 (2007).

Ichikawa, S., S. Karita, M. Kondo, M. Goto: Cellulosomal carbohydrate-binding module from *Clostridium josui* binds to crystalline and non-crystalline cellulose, and soluble polysaccharides. *FEBS Lett*., **588**, 3886–3890 (2014).

Ichinose, H., M. Yoshida, Z. Fujimoto, S. Kaneko: Characterization of a modular enzyme of exo-1,5-α-L-arabinofuranosidase and arabinan binding module from *Streptomyces avermitilis* NBRC14893. *Appl*. *Microbiol*. *Biotechnol*., **80**, 399–408 (2008).

Inácio, J. M., I. de Sá-Nogueira: Characterization of abn2 (yxiA), encoding a *Bacillus subtilis* GH43 arabinanase, Abn2, and its role in arabino-polysaccharide degradation. *J. Bacteriol*., **190**, 4272–4280 (2008).

Ito, D., E. Nakano, S. Karita, M. Umekawa, K. Ratanakhanokchai, C. Tachaapaikoon: Characterization of a GH family 43 β-xylosidase having a novel carbohydrate-binding module from *Paenibacillus xylaniclasticus* strain TW1. *J*. *Appl*. *Glycosci*., **69**, 65-71 (2022) (10.5458/jag.jag.JAG-2022_0001).

Ito, D., S. Karita, M. Umekawa: A C1/C4-oxidizing AA10 lytic polysaccharide monooxygenase from *Paenibacillus xylaniclasticus* strain TW1. *J*. *Appl*. *Glycosci*., in press (10.5458/jag.jag.JAG-2022_00011).

Jam, M., E. Ficko-Blean, A. Labourel, R. Larocque, M. Czjzek, G. Michel: Unraveling the multivalent binding of a marine family 6 carbohydrate-binding module with its native laminarin ligand. FEBS J., 283, 1863–1879 (2016).

Jordan, D. B., X. L. Li, C. A. Dunlap, T. R. Whiteheada, M. A. Cotta: β-D-Xylosidase from Selenomonas ruminantium of glycoside hydrolase family 43. Appl. Biochem. Biotechnol., 137–140, 93–104 (2007).

Jamal-Talabani, S., A. B. Boraston, J. P. Turkenburg, N. Tarbouriec, V. M. A. Ducros, G.J. Davies: Ab initio structure determination and functional characterization of CBM36:A new family of calcium-dependent carbohydrate binding modules. Structure. 12, 1177–1187 (2004).

Karita, S.: Carbohydrate-binding modules in plant cell wall-degrading enzymes. *Trends Glycosci. Glycotechnol*., **28**, E49–E53 (2016).

Keller, M. B., S. F. Badino, N. Røjel, T. H. Sørensen, J. Kari, B. McBrayer, K. Borch, B. M. Blossom, P. Westh: A comparative biochemical investigation of the impeding effect of C1-oxidizing LPMOs on cellobiohydrolases. *J. Biol. Chem.*, **296**, 100504 (2021).

Ketudat Cairns, J. R., A. Esen: β-Glucosidases. *Cell Mol. Life Sci.*, **67**, 3389-405 (2010).

Kim, S., J. Ståhlberg, M. Sandgren, R. S. Paton, G. T. Beckham: Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci. USA*, **111**, 149-154 (2014).

Kim, Y. A., K. H. Yoon: Characterization of a *Paenibacillus woosongensis* β-xylosidase/ α-arabinofuranosidase produced by recombinant *Escherichia coli*. *J. Microbiol. Biotechnol*., **20**, 1711–1716 (2010).

Kracher, D., S. Scheiblbrandner, A. K. Felice, E. Breslmayr, M. Preims, K. Ludwicka, D. Haltrich, V. G. Eijsink, R. Ludwig: Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science*, **352**, 1098-1101 (2016).

Kraulis, J., G.M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, A. M. Gronenborn: Determination of the three-dimensional solution structure of the *C*-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*, **28**, 7241-7257 (1989).

Lagaert, S., A. Pollet, J. A. Delcour, R. Lavigne, C. M. Courtin, G. Volckaert: Characterization of two β-xylosidases from *Bifidobacterium adolescentis* and their contribution to the hydrolysis of prebiotic xylooligosaccharides. *Appl*. *Microbiol*. *Biotechnol*., **92**, 1179–1185 (2011).

Lamed, R., E. Setter, R. Kenig, E. A. Bayer: The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol*. *Bioeng*. *Symp*., **13**, 163–181 (1983).

Langston, J. A., T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko, M. D. Sweeney: Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.*, **77**, 7007-7015 (2011).

Levasseur, A., E. Drula, V. Lombard, P. M. Coutinho, B. Henrissat: Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol*. *Biofuels*, **6**, 41 (2013).

Li, X., X. Xie, J. Liu, D. Wu, G. Cai, J. Lu: Characterization of a putative glycoside hydrolase family 43 arabinofuranosidase from *Aspergillus niger* and its potential use in beer production. *Food Chem*., **305**, 125382 (2020).

Liu, J., D. Sun, J. Zhu, C. Liu, W. Liu: Carbohydrate-binding modules targeting branched polysaccharides: overcoming side-chain recalcitrance in a non-catalytic approach. *Bioresour*. *Bioprocess*, **8**, 28 (2021).

Liu, Y., X. Yang, L. Gan, S. Chen, Z. X. Xiao, Y. Cao: CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res*., **50**, W159–164 (2022).

Limsakul, P., P. Phitsuwan, R. Waeonukul, P. Pason, C. Tachaapaikoon, K. Poomputsa, A. Kosugi, M. Sakka, K. Sakka, K. Ratanakhanokchai: A novel AA10 from *Paenibacillus curdlanolyticus* and its synergistic action on crystalline and complex polysaccharides. *Appl. Microbiol. Biotechnol.*, **104**, 7533-7550 (2020).

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat: The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, **42**, D489-5 (2014).

Lowry, O. H., N. J. Rosebrough, A. L. Farr, R. J. Randall: Protein measurement with the Folin phenol reagent. *J. Biol. Chem*., **193**, 265- 275 (1951).

Michlmayr, H., J. Hell, C. Lorenz, S. Böhmdorfer, T. Rosenau, W. Kneifel: Arabinoxylan oligosaccharide hydrolysis by family 43 and 51 glycosidases from *Lactobacillus brevis* DSM 20054. *Appl*. *Environ*. *Microbiol*., **79**, 6747–6754 (2013).

Miller, G. L.: Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem*., **31**, 426-428 (1959).

Mirdita, M, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger: ColabFold: making protein folding accessible to all. *Nat*. *Methods*, **19**, 679-682 (2022).

Mewis, K., N. Lenfant,V. Lombard, B. Henrissat: Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Appl*. *Environ*. *Microbiol*., **82**, 1686–1692 (2016).

Moser, F., D. Irwin, S. Chen, D.B. Wilson: Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng*., **100**, 1066-1077 (2008).

Moraïs, S., O. Salama-Alber, Y. Barak, Y. Hadar, D.B. Wilson, R. Lamed, Y. Shoham, A.E. Bayer: Functional association of catalytic and ancillary modules dictates enzymatic activity in glycoside hydrolase family 43 β-xylosidase. *J. Biol*. *Chem*., **287**, 9213–9221 (2012).

Marana S.R.: Structural and mechanistic fundamentals for designing of cellulases. *Comput. Struct. Biotechnol. J.*, **2**, e201209006 (2012).

Mroueh, M., M. Aruanno, R. Borne, P. de Philip, H.P. Fierobe, C. Tardif, S. Pagès: The xyl-doc gene cluster of *Ruminiclostridium cellulolyticum* encodes GH43- and GH62-α-L-arabinofuranosidases with complementary modes of action. *Biotechnol*. *Biofuel.*, **12**, 144 (2019).

Notenboom, V., A.B. Boraston, D.G. Kilburn, D.R. Rose: Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry*, **40**, 6248–6256 (2001).

Nozaki, S., H. Niki: Exonuclease III (XthA) enforces *in vivo* DNA cloning of *Escherichia coli* to create cohesive ends. *J. Bacteriol*., **201**, e00660-18 (2019).

Okuyama, M.: Function and structure studies of GH family 31 and 97 α-glycosidases. *Biosci. Biotechnol. Biochem.*, **75**, 2269-2277 (2011).

Prates, J.A.M., H.J. Gilbert, L.M.A. Ferreira, A.S. Luís, A. Rogowski, J. Xue, A. Baslé, S. Najmudin, I. Venditto, J.P. Knox, M.J. Temple, C.M.G.A Fontes: Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan. *J. Biol. Chem.*, **288**, 4799–4809 (2012)

Peterson, R., H. Nevalainen: *Trichoderma reesei* RUT-C30 - thirty years of strain improvement. *Microbiology*, **158,** 58–68 (2012).

Quinlan, R.J., M.D. Sweeney, L. Lo Leggio, H. Otten, J.C. Poulsen, K.S. Johansen, K.B. Krogh, C.I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C.P. Walter, P. Dupree, F. Xu, G.J. Davies, P.H. Walton: Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. USA*, **108**, 15079-15084 (2011).

Rabinovich, M.L., M.S. Melnick, A.V. Bolobova: The structure and mechanism of action of cellulolytic enzymes. *Biochemistry (Mosc)*, **67**, 850-871 (2002).

Ratanakhanockchai, K., C. Tachaapaikoon, K.L. Kyu, P. Pason: A novel multienzyme complex from a newly isolated facultative anaerobic bacterium, *Paenibacillus* sp. TW1. *Act*. *Biol*. *Hung*., **63**, 288–300 (2012).

Reese, E.T., R.G. Siu, H.S. Levinson: The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J. Bacteriol*., **59**, 485-497 (1950).

Robak K., M. Balcerek: Review of second-generation bioethanol production from residual biomass. *Food Technol. Biotechnol..*, **56**, 174–187 (2018).

Rohman, A., N. van Oosterwijk, N.N.T. Puspaningsih, B.W. Dijkstra: Structural basis of product inhibition by arabinose and xylose of the thermostable GH43 β-1,4-xylosidase from *Geobacillus thermoleovorans* IT-08. *PLoS One*, **13**, e0196358 (2018).

Rohman, A., B.W. Dijkstra, N.N.T. Puspaningsih: β-Xylosidases: Structural diversity, catalytic mechanism, and inhibition by monosaccharides. *Int. J. Mol. Sci.*, **20**, 5524 (2019).

Sabbadin, F., G.R. Hemsworth, L. Ciano, B. Henrissat, P. Dupree, T. Tryfona, R.D.S. Marques, S.T. Sweeney, K. Besser, L. Elias, G. Pesante, Y. Li, A.A. Dowle, R. Bates, L.D. Gomez, R. Simister, G.J. Davies GJ, P.H. Walton, N.C. Bruce, S.J. McQueen-Mason: An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat. Commun.*, **9**, 756 (2018).

Sabbadin, F., S. Urresti, B Henrissat, A.O. Avrova, L.R.J Welsh, P.J. Lindley, M. Csukai, J.N. Squires, P.H. Walton, G.J. Davies, N.C. Bruce, S.C. Whisson, S.J. McQueen-Mason: Secreted pectin monooxygenases drive plant infection by pathogenic oomycetes. *Science*, **373**, 774-779 (2021).

Sakka, K., M. Nakanishi, M. Sogabe, T. Arai, H. Ohara, A. Tanaka, T. Kimura, K. Ohmiya: Isothermal titration calorimetric studies on the binding of a family 6 carbohydrate-binding module of *Clostridium thermocellum* XynA with xylooligosaccharides. *Biosci. Biotechnol. Biochem.*, **67**, 406–409 (2003).

Shallom, D., M. Leon, T. Bravman, A. Ben-David, G. Zaide, V. Belakhov, G. Shoham, D. Schomburg, T. Baasov, Y. Shoham: Biochemical characterization and identification of the catalytic residues of a family 43 β-D-xylosidase from *Geobacillus stearothermophilus* T-6. *Biochemistry*, **44**, 387–397 (2005).

Shi, P., X. Chen, K. Meng, H. Huang, Y. Bai, H. Luo, P. Yang, B. Yao: Distinct actions by *Paenibacillus* sp. strain E18 α-L-arabinofuranosidases and xylanase in xylan degradation. *Appl*. *Environ. Microbiol*., **79**, 1990–1995 (2013).

Srivastava, N., R. Rathour, S. Jha, K. Pandey, M. Srivastava, V.K. Thakur, R.S. Semger, V.K. Gupta, P.B. Mazumder, A.F. Khan, P.K. Mishra: Microbial beta glucosidase enzymes: Recent advances in biomass conversation for biofuels application. *Biomolecules*, **9**, 220 (2019).

Stepnov A.A., Z. Forsberg, M. Sørlie, G.S. Nguyen, A. Wentzel, Å.K. Røhr, V.G.H Eijsink: Unraveling the roles of the reductant and free copper ions in LPMO kinetics. *Biotechnol*. *Biofuels*, **14**, 28 (2021).

Suzuki, K., M. Suzuki, M. Taiyoji, N. Nikaidou, T. Watanabe: Chitin binding protein (CBP21) in the culture supernatant of *Serratia marcescens* 2170: *Biosci. Biotechnol. Biochem*., **62**, 128-135 (1998).

Szabó, L., S. Jamal, H. Xie, S.J. Charnock, D.N. Bolam, H.J. Gilbert, G.J. Davies: Structure of a family 15 carbohydrate-binding module in complex with xylopentaose: Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.*, **276**, 49061–49065 (2001).

Tan, H., K.R. Korbin, G.B. Fincher: Emerging technologies for the production of renewable liquid transport fuels from biomass sources enriched in plant cell walls. *Front. Plant Sci.*, **7**, 1854 (2016).

Tachaapaikoon, C., S. Tanasupawat, P. Pason, S. Sornyotha, R. Waeonukul, K.L. Kyu, K. Ratanakhanockchai: *Paenibacillus xylaniclasticus* sp. nov., a xylanolytic-cellulolytic bacterium isolated from sludge in an anaerobic digester. *J. Microbiol.*, **50**, 394–400 (2012).

Teeravivattanakit, T., S. Baramee, P. Phitsuwan, R. Waeonukul, P. Pason, C. Tachaapaikoon: Novel trifunctional xylanolytic enzyme Axy43A from *Paenibacillus curdlanolyticus* strain B-6 exhibiting endo-xylanase, β-D-xylosidase, and arabinoxylan arabinofuranohydrolase activities. *Appl. Environ. Microbiol.*, **82**, 6942–6951 (2016).

Tindall, B.J.: The names *Hungateiclostridium* Zhang *et al*. 2018, *Hungateiclostridium thermocellum* (Viljoen *et al*. 1926) Zhang *et al*. 2018, *Hungateiclostridium cellulolyticum* (Patel *et al*. 1980) Zhang *et al*. 2018, *Hungateiclostridium aldrichii* (Yang *et al*. 1990) Zhang *et al*. 2018, *Hungateiclostridium alkalicellulosi* (Zhilina *et al*. 2006) Zhang *et al*. 2018, *Hungateiclostridium clariflavum* (Shiratori *et al*. 2009) Zhang *et al*. 2018, *Hungateiclostridium straminisolvens* (Kato *et al*. 2004) Zhang *et al*. 2018 and *Hungateiclostridium saccincola* (Koeck *et al*. 2016) Zhang *et al*. 2018 contravene rule 51b of the international code of nomenclature of prokaryotes and require replacement names in the genus *Acetivibrio* Patel *et al*. 1980. *Int J. Syst. Evol. Microbiol.*, **69**, 3927-3932 (2019).

Uchiyama, T, T. Uchihashi, T, Ishida, A. Nakamura, J.V. Vermaas, M.F. Crowley, M. Samejima, G.T. Beckham, K. Igarashi: Lytic polysaccharide monooxygenase increases cellobiohydrolases activity by promoting decrystallization of cellulose surface. *Sci. Adv.*, **8**, eade5155 (2022).

van Bueren, A.L., C. Morland, H.J. Gilber, A.B. Boraston: Family 6 carbohydrate binding modules recognize the non-reducing end of β-1,3-linked glucans by presenting a unique ligand binding surface. *J. Biol. Chem.*, **280**, 530–537 (2005).

Vaaje-Kolstad, G, B. Westereng, S.J. Horn, Z. Liu, H. Zhai, M. Sørlie, V.G. Eijsink: An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, **330**, 219-222 (2010).

Viborg, A.H., K.I.Sørensen, O. Gilad, D.B.Steen-Jensen, A. Dilokpimol, S. Jacobsen, B. Svensson: Biochemical and kinetic characterisation of a novel xyloolosaccharide-upregulated GH43 β-D-xylosidase/α-L-arabinofuranosidase (BXA43) from the probiotic *Bifidobacterium animalis* subsp. *lactis* BB-12. *AMB Express*, **3**, 56 (2013).

Vu, V.V., W.T. Beeson, E.A. Span, E.R. Farquhar, M.A. Marletta: A family of starch-active polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. USA*, **111**, 13822-13827 (2014).

Wagschal, K., C. Heng, C.C. Lee, D.W.S. Wong: Biochemical characterization of a novel dual-function arabinofuranosidase/ xylosidase isolated from a compost starter mixture. *Appl*. *Microbiol*. *Biotechnol*., **81**, 855–863 (2009).

Wang, L., A. Hatem, U.V. Catalyurek, M. Morrison, Z. Yu: Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One*, **8**, e78507 (2013).

Wood, T.M., S.I. McCrae: The purification and properties of the C_1 component of *Trichoderma koningii* cellulase. *Biochem*. *J*., **128**, 1183-1192 (1972).

Yang, X., P. Shi, H. Huang, H. Luo, Y. Wang, W. Zhang, B. Yao: Two xylose-tolerant GH43 bifunctional β-xylosidase/α-arabinosidases and one GH11 xylanase from *Humicola insolens* and their synergy in the degradation of xylan. *Food Chem*., **148**, 381–387 (2014).

Yang, X., P. Shi, R. Ma, H. Luo, H. Huang, P. Yang, B. Yao: A new GH43 α-arabinofuranosidase from *Humicola insolens* Y1: Biochemical characterization and synergistic action with a xylanase on xylan degradation. *Appl*. *Biochem*. *Biotechnol*., **175**, 1960–1970 (2015).

Yoshida, S., C.W. Hespen, R.L. Beverly, R.I. Mackie, I.K. Cann: Domain analysis of a modular α-L-arabinofuranosidase with a unique carbohydrate binding strategy from the fiber-degrading bacterium *Fibrobacter succinogenes* S85. *J*. *Bacteriol*., **192**, 5424–5436 (2010).

Zanphorlin, L.M., M.A.B. de Morais, J.A. Diogo, M.N. Domingues, F.H.M. de Souza, R. Ruller, M.T. Murakami: Structure-guided design combined with evolutionary diversity led to the discovery of the xylose-releasing exo-xylanase activity in the glycoside hydrolase family 43. *Biotechnol*. *Bioeng*., **116**, 734–744 (2019).

Zhang, R., N. Li, Y. Liu, X. Han, T. Tu, J. Shen, S. Xu, Q. Wu, J. Zhou, Z. Huang: Biochemical and structural properties of a low-temperature-active glycoside hydrolase family 43 β-xylosidase: Activity and instability at high neutral salt concentrations. *Food Chem*., **301**, 125266 (2019).

Zhou, J., L. Bao, L. Chang, Y. Zhou, H. Lu: Biochemical and kinetic characterization of GH43 β-D-xylosidase/ α-L-arabinofuranosidase and GH30 α-L-arabinofuranosidase/ β-D-xylosidase from rumen metagenome. *J*. *Ind*. *Microbiol*. *Biotechnol*., **39**, 143–152 (2012).

Zhou, X., X. Qi X, H. Huang, H. Zhu: Sequence and structural analysis of AA9 and AA10 LPMOs: an insight into the basis of substrate specificity and regioselectivity. *Int. J. Mol. Sci.*, **20**, 4594 (2019).

Zverlov, V. V., G. A. Velikodvorskaya, W. H. Schwarz: A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology*, **148**, 247-255 (2002).

ACKNOWLEDGEMENTS

I would like to my deepest sense of gratitude to Dr. Shuichi Karita and Dr. Midori Umekawa, Laboratory of Fermentation Biology, for their intent guidance and supervision.

I am grateful to Dr. Tetsuya Kimura, Laboratory of Applied Microbial Genetics, Dr. Minoru Inagaki, Laboratory of Biofunctional Chemistry, and Dr. Hideo Miyake, Laboratory of Molecular Bioinformatics, for their valuable comments and kind encouragement.

I am greatly appreciated Dr. Naoto Isono, Laboratory of Food Chemistry, Dr. Hirotaka Katsuzaki and Ms. Yu Kawamura, Laboratory of Bioorganic Chemistry, and Mr. Koki Taniguchi, Laboratory of Fermentation Biology, for their technical assistance and valuable comments.

My note of gratitude also goes to Dr. K. Ratanakhanockchai and Dr. C. Tachaapaikoon, King Mongkut's University of Technology, Thailand, for the provision of *P. xylaniclasticus* strain TW1 and useful discussions.

I wish to express my unfailing gratitude to Ms. Rina Shirasaki and Ms. Emiri Nakano for their experimental support.

I am grateful to Ms. Tomiko Chikada and Ms. Mina Noborikawa, Center for Molecular Biology and Genetics, Mie university, for gene technical assistance and encouragement.

I am also extremely grateful to the past and present members of the Laboratory of Fermentation Biology.

Lastly, I would be remiss in not mentioning my family, especially my parents, my friends, and the participants of “The Academic Night” for emotional support and keeping my spirits and motivation high. Thank you all.

This work was supported by JST SPRING, Grant Number JPMJSP2137.