博士論文

Studies on the enzymatic properties and physiological functions of

fungal glucuronoyl esterases

(糸状菌グルクロン酸エステラーゼの 酵素学的性質と生理機能に関する研究)

HUYNH　Hiep　Hung
ヒン　ヒペ　ホン

博士論文

Studies on the enzymatic properties and physiological functions of fungal glucuronoyl esterases

（糸状菌グルクロン酸エステラーゼの酵素学的性質と生理機能に関する研究）

応用生命工学専攻
平成27年度博士課程入学
氏 名　HUYNH Hiep Hung
指導教員名　吉田 稔

Studies on the enzymatic properties and physiological functions of fungal glucuronoyl esterases

（糸状菌グルクロン酸エステラーゼの酵素学的性質と生理機能に関する研究）

By

HUỲNH Hiệp Hùng

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

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

December 2017

Table of contents   **Paper numbers**

Chapter 3: The role of a conserved Lys on the recognition of the 4-*O*-methyl group on the substrate

Chapter 4: The physiological functions of fungal glucuronoyl esterases

Chapter 5: Final conclution

General introduction

Outline of renewable energy and biofuel

The term biofuel has been mentioned for a long time since the development of the industrial manufacture of cars. Even in the beginning of the 20th century, the model Ts designed by Henry Ford was planned to use ethanol. Moreover, the peanut oil was used early in the diesel engines. Unfortunately, at that time, gasoline and diesel price were quite cheap and more convenient for applying in a large scale due to the constant finding of crude oils reserves. Biofuels, therefore, were nearly abandoned. Nowadays, the global warming is an urgent issue and petroleum consumption is blamed as one of the major reasons for this problem. Another important factor is that the global fossil fuel deposits are predicted to run out in the near future and biofuels will become a "safeguard" for human society.

The fossil fuels require over millions of years with some specific conditions to form, and thus it cannot be sufficiently replaced for the energy consumption by human civilization. In contrast, biofuel production depends on biomass which is naturally produced through the carbon cycle or from the crop farming. For this reason, biomass and its products, the biofuels, are considered as the renewable energy source because it can be replaced every year (Sims et al. 2008).

In the total renewable energy supplement, the highest contribution comes from biofuels. The ratio of the biofuel supply in the global energy consumption is maintained at around 10.5% from 2005 (Eisentraut 2010). Since 2010, the use of biomass for producing bioenergy has steadily increased by around 2.5% per year (Fig. 0-1). Compared to 2015, the total biofuel production grew by 2%, achieving 135 billion liters in 2016. Besides that, in the US, the amount of bioethanol increased by 3.5% to 58 billion liters during the year. In Asia, supported by the government policy to reduce the import of petroleum, the production of bioethanol in India rose to 0.9 billion liters, whereas it reached to 1.2 billion liters in Thailand (REN21 2017).

Until now, the development of biofuels is broken down into several generations. In the first generation, starch, sugar, animal fats, and vegetable oil were used as the primary materials to generate biofuels (http://biofuel.org.uk/). However, there are many arguments risen from scientists, farmers, and even the customers. For example, some biofuel production is highly costly and may release much more carbon in the process than the amount stored in their feedstock so that final greenhouse gases emission is increased. In addition, the competition in the choice between the food and biofuel crops is the most prevalent issue, especially because famine is still a big concern in some countries. Deforestation is also one of the consequences of using farm areas for biofuel crops. For these reasons, the second generation biofuels are suggested to replace the first generation biofuels. The materials such as wood chips, organic

waste, and food crop waste are chosen to produce biofuels. They are plentiful, renewable, and almost costless which make them superior to the first generation biofuels and fossil fuels (Sims et al. 2008).

The second generation biofuels include bioethanol, biodiesel, biomethane, and some higher alcohols. Until now, bioethanol is still the major product. In the first step of biomass conversion to bioethanol, the mixture containing various enzymes is applied to hydrolyze the lignocellulose into sugars (Fig. 0-2). These sugars are then fermented into alcohol by yeast or bacteria (Eisentraut 2010). Currently, almost all the processes of biomass degradation are demonstrated only in the laboratory pilot or small scale due to the recalcitrant and complex structure of lignocellulose which significantly reduces the efficiency of biomass degradation. The commercial procedure still strongly demands the expensive equipment with various pretreatments, thus leading to high cost of the end-products. Further research is required to reduce the price and to improve performance and reliability of the conversion process (Sims et al. 2008).

Lignocellulosic biomass

Biomass is a great renewable source of chemicals and energy. It is relatively cheap and its processing does not increase the percentage of greenhouse gases. The weight of dry biomass on earth is estimated to be $1.85\text{-}2.4 \times 10^{12}$ tons. Every year $150\text{-}180 \times 10^9$ tons of photosynthetic biomass is produced. To obtain more accurate information, the satellites were used to scan the vegetation over the terrestrial terrain. The results showed that around 56.4 billion tons of biomass are generated per year. The amount of produced biomass from the oceans was indirectly estimated through the sea-surface chlorophyll levels and the about 48.5 billion tons of annual cellulose are produced in all over the ocean (Hodásová et al. 2015).

Lignocellulosic biomass is a complex mixture of cellulose, hemicellulose, and lignin (Fig. 0-3). Cellulose is the linear polymer of β-(1,4)-linked D-glucose units which is the core component and accounts for 20–40% of dry weight in the primary cell walls. In the secondary cell walls, it could be increased to be 50% and even 100% in some special cases like in the cottonseed hairs (Watanabe and Tokuda 2010). Therefore, it is considered to be the most abundant organic polymer on Earth, and thus expected to be a highly potential energy source.

Hemicellulose is the hetero-polymer of β-(1,4)-D-xylopyranose, glucopyranose, mannopyranose, and galactopyranose backbone with some side chains such as arabinose. The glucuronoxylan of most plants consists of β-(l,4)-D-xylan main chains with several

substituents, for instance, 4-O-methyl-D-glucuronic acid, acetic acid, and ferulic acid (Jeffries 1994; Rubin 2008). Hemicellulose is believed as the second most abundant renewable source for bioenergy by providing various sugar compounds such as pentose (D-xylose and L-arabinose) and hexoses (D-glucose, D-mannose, and D-galactose).

Lignin constitutes the aromatic components of biomass and occupies about 18–35% of dry weight in the plant cell walls (Watanabe and Tokuda 2010). Three basic monomers of lignin named "monolignols" are p-coumaryl, coniferyl, and sinapyl alcohols which are synthesized from phenylalanine. When these monomers are assembled in the lignin structure, they are called p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units (Fig. 0-3). The ratio of these units within the lignin varies between different plants. For example, H-unit dominates in the grass, while in gymnosperms, it is G-unit that dominates. Lignin plays a role as an adhesive to hold the cellulose and hemicellulose components as main parts of lignocellulose matrix (Fig. 0-3) (Hodásová et al. 2015). In the plant cell walls, lignin can be connected to hemicellulose through chemical linkages that form the macromolecular structures called lignin-carbohydrate complexes (LCCs) (Fig. 0-4) (d'Errico et al. 2015; Bååth et al. 2016). Lignin, as a part of biomass, is also a good renewable source and has relatively stable prices through the years and seasons. In the ecosystem, the quantity of total lignin represents more than 300 billion tons and around 20 billion tons are produced per year. The interesting part of lignin is that it can be broken down into several building blocks and from the reduction of them the primary substances such as benzene, toluene, xylene, and phenol can be produced.

The ester linkage of the lignocellulose

Hemicellulose can associate with lignin through the esters and ethers linkages (Fig 0-4). It is generally accepted that the α-linked esters (benzyl ester types) in LCCs are formed during the lignin biosynthesis of plant cell wall (Watanabe 1995). Recently, the direct evidences of the γ-linked esters in LCCs have been reported by using the advanced 2D NMR spectroscopy in the softwood (pine-*Pinus taeda*) and the hardwoods (white birch-*Betula pendula* and *Populus tomentosa)* (Balakshin et al. 2011; Yuan et al. 2011). In LCCs, lignin is covalently bonded to polysaccharides in the intact plant cell wall, thus reducing the accessible surface area of cellulose and hamper the degradation of lignocellulose (Laureano-Perez et al. 2005). In wood xylans, monomeric side chains consist of 4-O-methyl glucuronic acid units which can be esterified with lignin at a proportion of around 40% (Jeffries 1994). Therefore, effective degradation of this linkage is essential to overcome or circumvent the bottlenecks in the plant

biomass degradation (Fig. 0-5). The use of glucuronoyl esterase (GE) which specially cleaves this ester linkage is supposed to be a new approach to solve this issue without using extreme pre-treatment. It is of note that the GEs can recognize and cleave both the α- and γ-linked esters between glucuronic acid residues of hemicellulose and aromatic alcohols of lignin (Biely 2016; d'Errico et al. 2016).

Glucuronoyl esterase (GE)

GE was first reported in the cellulose-spent culture fluid of the wood-rotting fungus *Schizophyllum commune* (*Sc*GE; Špániková and Biely, 2006). Later another GE, *Trichoderma reesei* (teleomorph *Hypocrea jecorina*) Cip2, was identified and characterized (Li et al. 2007), followed by several other GEs from filamentous fungi, such as GEs from *Phanerochaete chrysosporium* (*Pc*GE1 and *Pc*GE2; Ďuranová et al. 2009b), *Sporotrichum thermophile* (*St*GE2; Topakas et al. 2010), *Podospora anserina* (*Pa*GE1; Katsimpouras et al. 2014), and *Cerrena unicolor* (*Cu*GE; d'Errico et al. 2015). These findings lead to the establishment of a new family of carbohydrate esterases (CE15) in the carbohydrate-active enzymes (CAZy) database (www.cazy.org; Cantarel et al., 2009). Later, based on the alignment of GE sequences in this family at that time and more than 30 GE-like sequences in the NCBI database which exhibit >40% amino acid sequence identity, a novel consensus sequence G-C-S-R-X-G containing the candidate catalytic nucleophile Ser residue was proposed (Topakas et al. 2010). Subsequent elucidation of the crystal structures of two GEs, Cip2 (PDB ID 3PIC; Pokkuluri et al. 2011) and *St*GE2 (4G4G; Charavgi et al. 2013), further unraveled the organization of the catalytic triad that is composed of Ser-His-Glu residues. The GE sequences are unique and phylogenetically distant from other CE families such as CE1 and CE5 of acetyl xylan esterases, CE1 and putative CE3 of feruloyl esterases, and CE8 of pectin methylesterases (Fig. 0-6) (Biely 2016).

Up until now, all the research on GEs depended on the synthetic compounds due to the unavailability of specific natural substrates (Fig. 0-7). By modifying the structure of the synthetic substrate to mimic the ester bond in the plant cell wall, some characteristics of GEs were elucidated. The *O*1 position of 4-*O*-methyl-D-glucuronic acid which can be linked to carbohydrate part of hemicellulose does not play a decisive role in the substrate recognition by *Sc*GE (Špániková et al. 2007). GEs seem to favor the glucuronoyl esters of bulkier alcohols such as benzyl, allyl, and methyl alcohols in the descending order of preference. A possible explanation is that the active site of GEs is located on the surface of the enzyme (Pokkuluri et

al. 2011) and thus providing the access of large substrates (d'Errico et al. 2015; Hüttner et al. 2017). The *C*4 configuration of pyranose ring is important in the enzymatic recognition of GEs. GEs attack exclusively the methyl esters of D-glucuronic acid but do not hydrolyze that of D-galacturonic acid – a C4-epimer of D-glucuronic acid (Ďuranová et al. 2009a; Wong et al. 2012). In addition, the substrates containing 4-*O*-methyl group in the pyranose ring are favored by GEs (Špániková et al. 2007; Wong et al. 2012).

Objective

The utilization of lignocellulosic biomass as a replacing energy source has become an urgent issue in the society due to the unstable price and concern about the depletion of fossil fuel resources. The industrial bio-conversion of lignocellulose into ethanol is built up of three main steps: the pretreatment of the biomass, enzymatic saccharification of LCCs, and microbial fermentation to produce ethanol. However, obtaining fermentable sugars from lignocellulose is not an easy task because of the recalcitrance of covalent linkage between hemicellulose and lignin. The breakdown of this linkage requires several pretreatments which highly consume energy and chemicals. The unique activity of GE which cleaves this ester linkage between 4-O-methyl-D-glucuronic acid residue of glucuronoxylan and aromatic lignin alcohol could provide a promising solution for this issue. In this Thesis, I performed the identification, functional expression, and characterization of two putative fungal GEs, *Nc*GE from *Neurospora crassa* and *Af*GE from *Aspergillus fumigatus*. Molecular modeling and site-directed mutagenesis studies of *Af*GE were performed to elucidate the important role of the conserved Lys residue in the preference for the substrates containing 4-O-methyl group in the glucopyranose ring. Finally, the biological role of GE on the degradation/utilization of plant biomass by *N. crassa* was investigated by using the mutant deleted for the *Nc*GE gene and wood powder from some popular plants in Japan as the sole carbon source. This work would contribute to understanding the characteristics and enzymatic mechanism of GEs. Moreover, it could provide a closer look into the physiological functions of GEs in the fungal growth process, especially in the environment containing the lignocellulosic biomass.

.
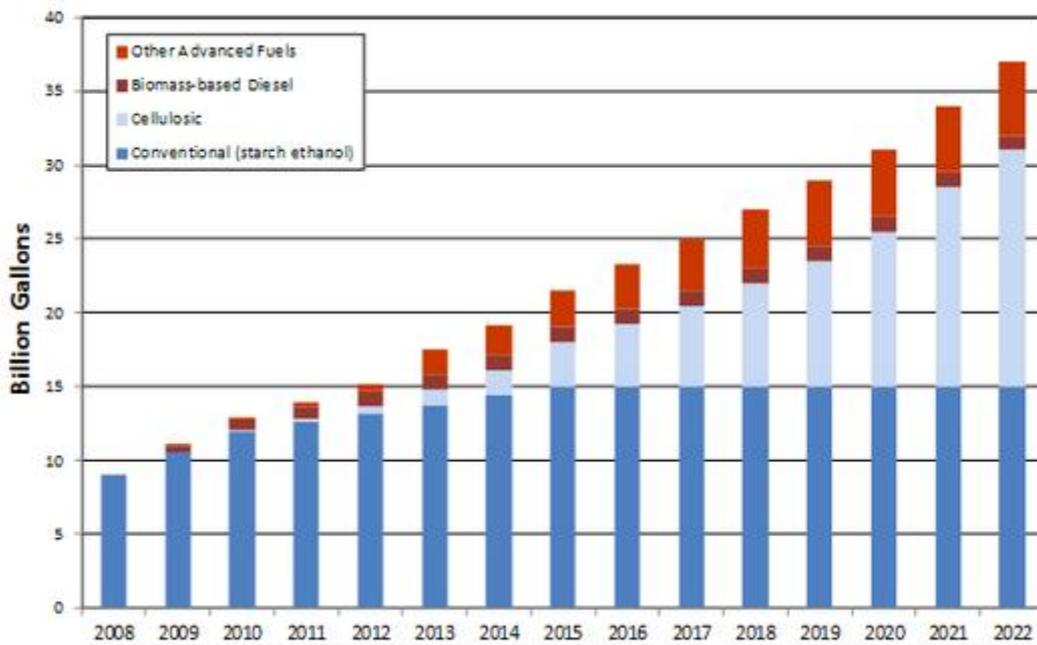
Figure 0-1: Renewable fuel standard volumes by year.

(Source: http://www.hcs.harvard.edu/~res/2014/05/)
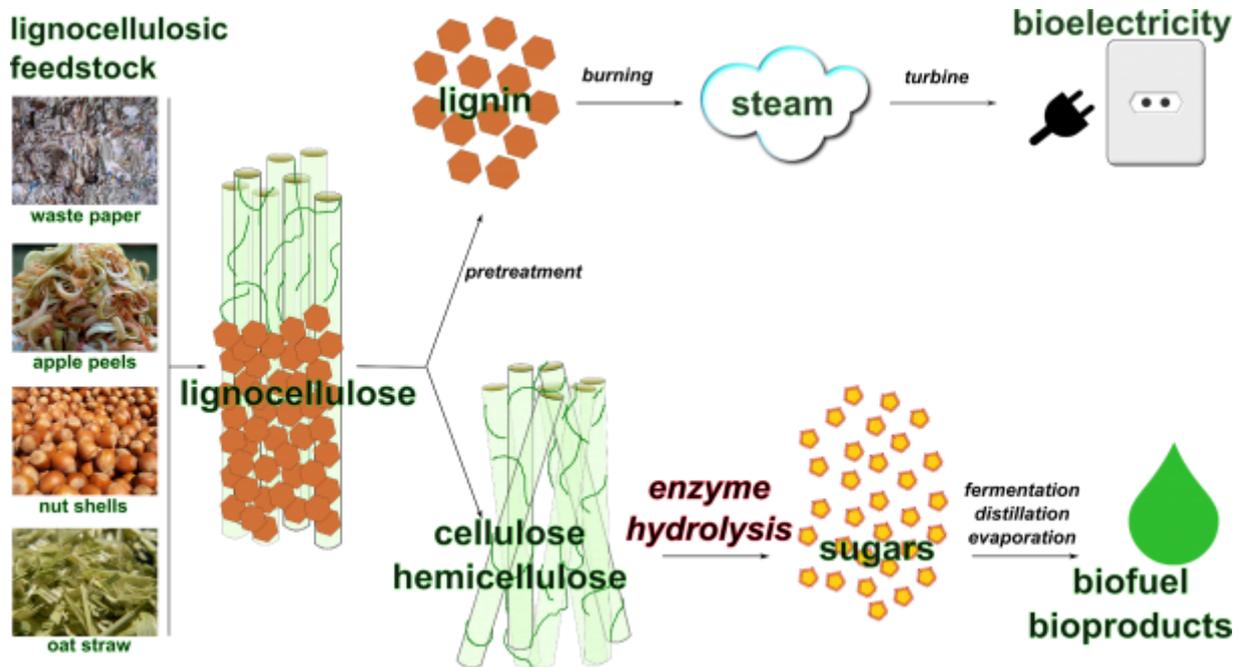
Figure 0-2: The conversion of lignocellulose into fermentable sugars for biofuel production (Source: http://www.semanticsoftware.info/genozymes-project).
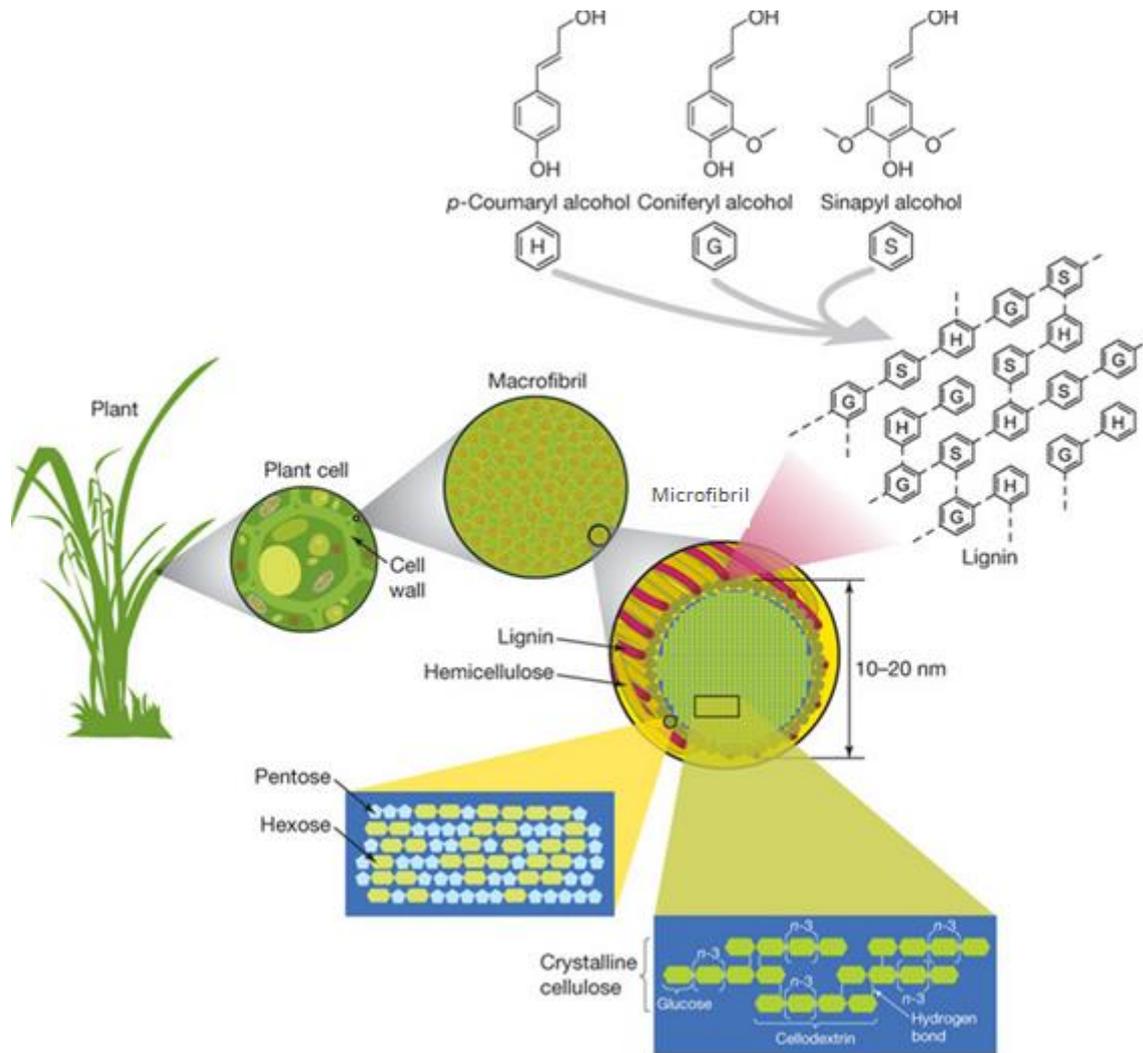
Figure 0-3: Structure of lignocellulose

There are three main components of lignocellulose: cellulose, hemicellulose, and lignin. Cellulose is a polymer of glucose molecules connected by β-(1,4)-linkage. Between different layers of cellulose, hydrogen bonds are formed to create the crystalline region. Hemicellulose is composed of various 5- and 6-carbon sugars like xylose, glucose, mannose, galactose, and arabinose. Lignin consists of three major phenolic compounds, *p*-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S). Cellulose, hemicellulose, and lignin create the matrix called microfibrils, which are collocated into macrofibrils, an intermediate structure in the plant cell wall (Rubin 2008).
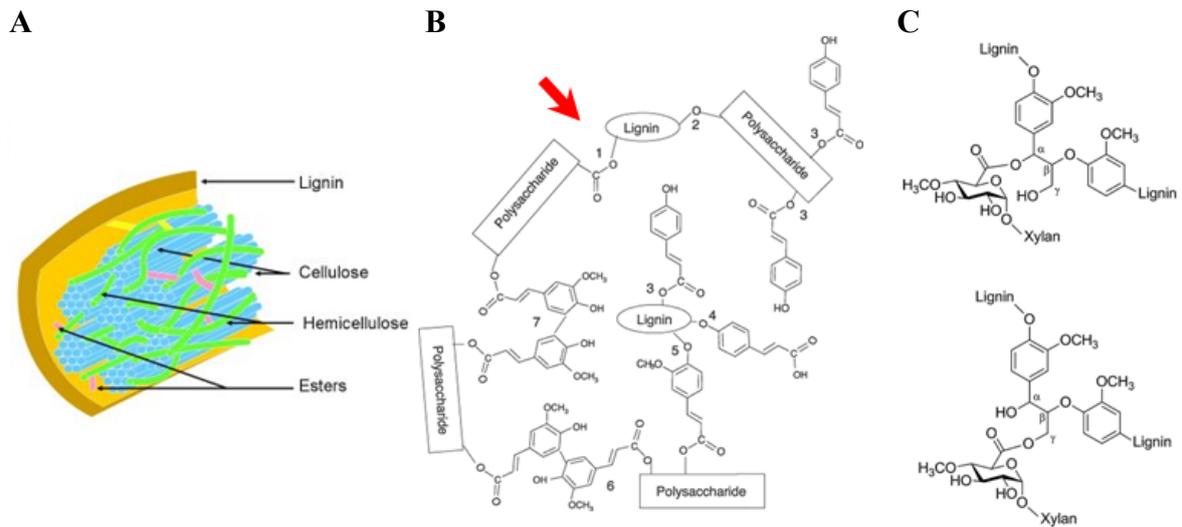
Figure 0-4: The model for the linkages of LCCs in the plant cell walls

(A) Lignocellulose matrix. (B) The possible linkages inside the LCCs. 1, glucuronic acid ester linked to lignin; 2, hydroxycinnamic acid ester; 3, hydroxycinnamic acid ether; 4, ferulic acid bridge; 5, direct ether linkage; 6, dehydrodiferulic acid diester bridge; 7, dehydrodiferulic acid diester–ether bridge. The arrow shows the ester linkage between hemicellulose and lignin which is targeted by GEs (Krause et al. 2003). (C) The simple structure of α-linked esters (upper) and γ-linked esters (lower) in LCCs (d'Errico et al. 2015).
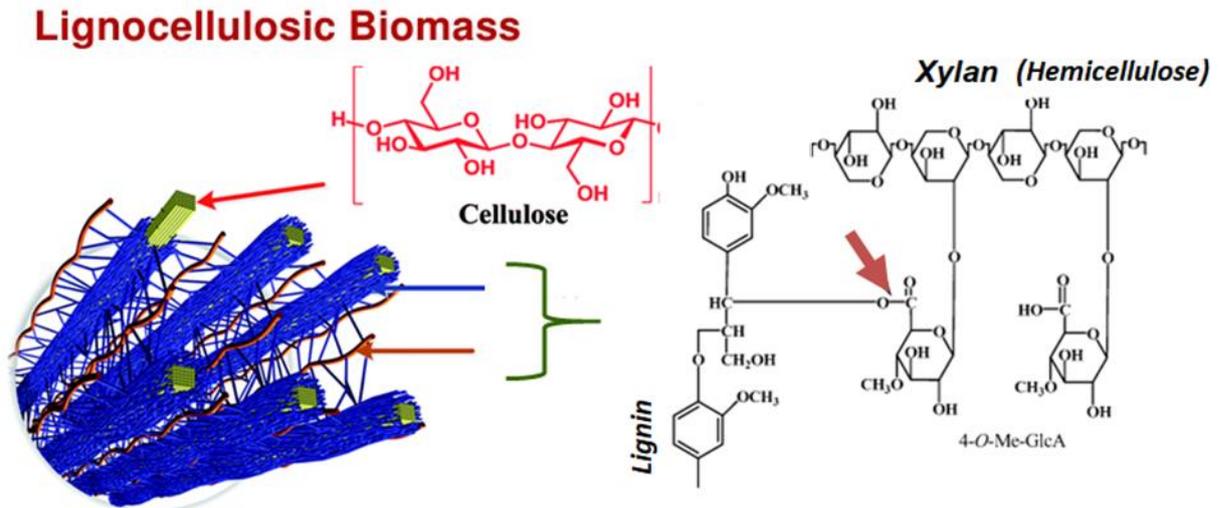
Figure 0-5: The ester linkage between hemicellulose and lignin, which is cleaved by GEs, is indicated by the arrow (Duranova et al., 2009)
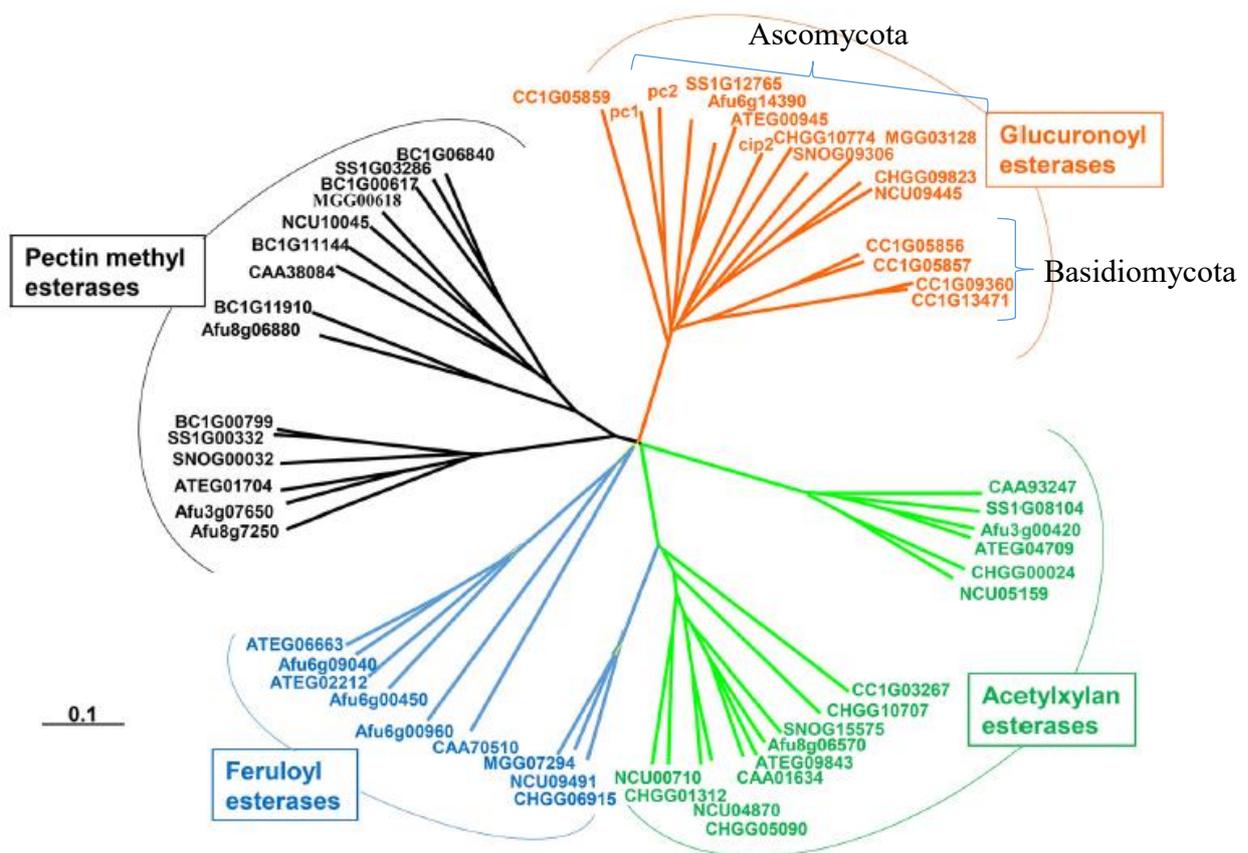
Figure 0-6: The biodiversity of GEs and other esterase families

The phylogenetic tree was built from the data sequences of characterized and putative GEs, acetylxylan esterases, feruloyl esterases, and pectin methyl esterases. The accession numbers were obtained from NCBI, BROAD, and JGI databases presented for the protein sequences (Biely 2016).
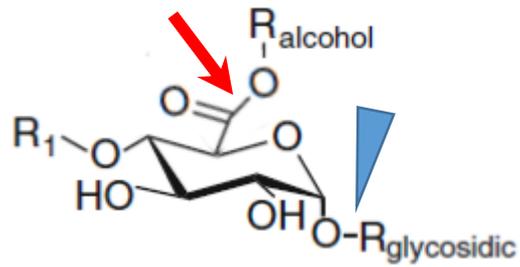
Figure 0-7: General structure of synthetic GE substrates

R_1, H or CH_3 substituent; $R_{alcohol}$, the aromatic groups of lignin; $R_{glycosidic}$, H or D-xylopyranoside. The ester linkage between glucuronic acid and lignin which is cleaved by GEs is shown by the arrow. The glycosidic linkage between glucuronic acid and polysaccharide of hemicellulose is marked by the arrowhead (Hüttner et al. 2017)

1.1. Introduction

Lignocellulose in biomass is a promising renewable energy. However, cellulose is surrounded by hemicellulose and lignin, which hampers complete degradation of cellulose. Hemicellulose and lignin form covalent bonds such as the linkage between 4-*O*-methyl glucuronic acid of hemicellulose and aromatic alcohol of lignin. Glucuronoyl esterase (GE) was recently found to cleave this ester linkage, and thus is expected to aid the decomposition of biomass.

After the discovery of GE, the first crystal structure of GE was reported for Cip2 of *T. reesei*, and its amino acid sequence was often used to search for homologs of GEs. Until now, around 11 GEs were characterized (Hüttner et al. 2017). Most of them are from fungi, but two GEs from bacteria were also characterized. In addition, there are some putative GEs from the symbiotic protist of termites. A recent study about the phylogenetic analysis of GEs by peptide pattern recognition confirmed the abundance of GEs in the Kingdom Fungi containing Basidiomycetous and Ascomycetous species. Surprisingly, GE seems less common in bacteria, although the selected organisms in that study are known as the lignocellulosic degraders in nature (Agger et al. 2017).

Aiming at finding and analyzing more about novel GEs, I performed an initial screening to select some candidate GEs. I paid more attention to the microorganisms, especially filamentous fungi. A BLAST search showed that a significant amino acid sequence homology was found between Cip2 and two putative GE sequences from filamentous fungi, *Nc*GE from *Neurospora crassa* and *Af*GE from *Aspergillus fumigatus*, which were characterized in this study.

1.1.1. *Pichia pastoris* as a host for heterologous expression of *Nc*GE

Pichia pastoris is a single-celled eukaryotic methylotrophic yeast which can use methanol as a carbon source. It is frequently used for heterologous expression of proteins in high yields. As a eukaryotic organism, *P. pastoris* can perform suitable post-translational modifications such as glycosylation which cannot be expected in the *E. coli* expression system. In addition, unlike *Saccharomyces cerevisiae*, *P. pastoris* can grow to very high cell densities in the minimal medium and the expression of recombinant proteins is tightly regulated by the methanol-inducible alcohol oxidase 1 (*AOX1*) promoter (Balbás and Lorence 2004). Another advantage is that the expressed proteins in *P. pastoris* less suffer hyperglycosylation compared to *S. cerevisiae*. Due to the absence of α-1,3-linked mannosyltransferase, glycoproteins

expressed in *P. pastoris* do not have a large number of *N*-linked glycosylation terminated via α-1,3-linked mannose residues, which are considered to be allergenic (Çelik and Çalık 2012). Therefore, *P. pastoris* expression system is suited for the production of heterologous proteins to the culture medium (Balbás and Lorence 2004).

There are two genes, *AOX1* and *AOX2*, encoding the alcohol oxidases in *P. pastoris* which catalyze the first step in the methanol utilization pathway (MUT pathway). There are two phenotypes often used in *P. pastoris* transformation. The first is Mut$^+$ phenotype (<u>m</u>ethanol <u>u</u>tilization <u>plus</u>), in which the cells grow normally in the methanol medium in the presence of both *AOX* genes. The other is Mut^S phenotype (<u>m</u>ethanol <u>u</u>tilization <u>s</u>low) which is observed in the *P. pastoris* KM71H strain (*his4 aox1::ARG4*) used in this study. The *AOX1* gene in this strain is partially deleted and replaced with *S. cerevisiae ARG4* gene. The *AOX2* gene allows this strain to grow in the methanol-containing medium at a slow rate (Cereghino and Cregg 2000). The Mut$^+$ strains have higher growth rate when grown in the methanol medium, which would result in better productivities. However, in the MUT pathway, the oxidation of methanol leads to the reduction of oxygen to hydrogen peroxide which may cause cellular stress. Another important factor is that this metabolism highly demands oxygen and releases heat, which might damage the recombinant proteins, especially thermosensitive proteins. For these reasons, the Mut^S strains have been preferred to be used in the heterologous protein expression (Orman et al. 2009; Krainer et al. 2012).

Some studies also confirmed the high ability of yeast *P. pastoris* together with an *AOX*1 promoter to secrete recombinant protein. *AOX*1 promoter is known to be one of the strongest and tightly regulated eukaryotic promoters. A wide variety of proteins have been expressed using this promoter, with yields up to 14.8 g/l and claims up to 20–30 g/l of recombinant protein (mouse collagen) (Çelik and Çalık 2012; Krainer et al. 2012). In this study, I used *P. pastoris* expression system to produce *Nc*GE in this Chapter. Also in Chapter 3, I used this system to produce wild-type and mutant *Af*GEs.

1.1.2. *Aspergillus oryzae* as a host for heterologous expression of *Af*GE

Filamentous fungi are the common hosts for producing recombinant proteins in the biotechnological application. The main advantages of fungi for heterologous protein expression are their fast growth on various inexpensive media and their high secretory capacity. For example, the concentration of the bovine chymosin achieved around 1 g/l by using *Aspergillus niger* system (Havlik et al. 2017). *Aspergillus oryzae* has long been used in the industrial

fermentation processes including *sake*, *miso*, and soy sauce production. Moreover, together with the development of genetic modification techniques, *A. oryzae* has become a prevalent and safe host for heterologous protein production (Mabashi et al. 2006). In our laboratory, the *A. oryzae* expression system has been successfully used for the production of several heterologous proteins, even including the β-glucosidase from a phylogenetically far distant organism, the termite *Neotermes koshunensis* (Uchima et al. 2011).

In this Chapter, I used the *A. oryzae* expression system to produce *Af*GE, since the gene encoding *Af*GE contained an intron in its coding sequence.

1.1.3. Objective

The discovery of GE has opened a new area for bio-degradation of lignocellulose. To understand more about this promising enzyme, in this Chapter, the Cip2 sequence was used to search for novel GE sequences. Two GEs, *Nc*GE and *Af*GE, were found and expressed in *P. pastoris* and in *A. oryzae*, respectively. Characterization of these GEs would contribute to better understanding of GEs, especially on their enzymatic properties which will be described in the following Chapters.

1.2. Results

1.2.1. Heterologous expression of *Nc*GE in *P. pastoris*

1.2.1.1. Sequence analysis of *Nc*GE

A BLAST search of *N. crassa* genome database (http://fungi.ensembl.org/Neurospora_crassa/Info/Index) using the amino acid sequence of Cip2 from *T. reesei* resulted in the identification of a putative protein, NCU09445.7, displaying significant sequence identity to Cip2 (55%) (Fig. 1-1) and other CE15 GEs. This protein, named *Nc*GE, is comprised of 395 amino acids with the predicted N-terminal signal peptide. The genomic sequence encoding *Nc*GE does not contain any intron and the deduced protein lacks the carbohydrate-binding module (CBM). No *N*-glycosylation site was predicted. The calculated molecular mass of mature polypeptide is around 41.3 kDa with the predicted isoelectric point of about 8.8 (www.expasy.org; Gasteiger et al. 2005). *Nc*GE has the consensus G-C-**S**-R-X-G motif conserved in the CE15 family where the central Ser residue serves as the catalytic nucleophile (Topakas et al. 2010; Pokkuluri et al. 2011).

1.2.1.2. Cloning and expression of *Nc*GE

An *E. coli*/*P. pastoris* shuttle vector pPICZαA was chosen to produce recombinant *Nc*GE. This vector has a strong alcohol oxidase 1 (*AOX1*) promoter which drives the expression of foreign genes when methanol is added to the culture medium (Fig. 1-2). In addition, the prepro-α-factor sequence derived from *S. cerevisiae* directs the secretion of recombinant proteins. Since the linearized plasmid prepared by digestion at the unique *Bgl* II site just upstream of the 5' *AOX1* region was used for transformation, the plasmid would be integrated into the *AOX1* locus of the yeast chromosome (Invitrogen 2010).

1.2.1.3. Expression of *Nc*GE using the pPICZαA vector

After methanol induction, SDS-PAGE analysis of the culture supernatant clearly showed that a protein of about 44 kDa which is nearly equal to the calculated size of c-Myc- and hexahistidine-tagged *Nc*GE was produced in the pPICZαA/*Nc*GE transformant, whereas no protein was detected in the vector-transformed control strain (Fig. 1-3A). Production of *Nc*GE was further confirmed by Western blot analysis using anti-c-Myc antibody (Fig. 1-3B). Smaller-sized doublet bands at around the position of 30 kDa were also detected, suggesting that *Nc*GE was partially degraded. Semi-quantitative measurement of GE activity using the substrate A shown in Fig. 1-4 demonstrated that GE activity was detected in the fractions

15

containing the 44 kDa protein, but not in the culture supernatant of the control strain. The highest amount of *Nc*GE protein was obtained after 2 days of methanol induction.

The synthetic compound 3-(4-methoxyphenyl) propyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate (substrate A) mimics the ester linkage in the lignin-carbohydrate complexes (Fig. 1-4) (Špániková et al. 2007; Sasagawa et al. 2011). This compound has 4-*O*-methyl group in glucuronic acid which is suggested to play an important role in the enzyme-substrate interaction (Špániková et al. 2007; Ďuranová et al. 2009a). Besides, the substrate contains the aromatic UV-absorbing chromophore which can be used to detect the de-esterification of the substrate by high performance liquid chromatography (HPLC).

1.2.2. Heterologous expression of *Af*GE in *A. oryzae*

1.2.2.1. Sequence analysis and cloning of a gene encoding the putative glucuronoyl esterase *Af*GE

Previously, the crystal structures of Cip2 (PDB ID 3PIC) and *St*GE2 (4G4G) suggested the presence of the catalytic triad consisting of Ser-His-Glu residues in GEs (Pokkuluri et al. 2011; Charavgi et al. 2013). By the sequence analysis of *Nc*GE and other characterized GEs, I found that these residues are present in highly conserved novel consensus sequences, VTGCSRXGKGA, HC, and PQESG (catalytic residues are underlined). Two Cys residues in these consensus sequences (Cys277 and Cys412 in Cip2) were suggested to form a disulfide bond and bring the catalytic Ser and His close together. Furthermore, two additional pairs of Cys residues (Cys 96-Cys131 and Cys309-Cys384 in Cip2) are also highly conserved in all characterized fungal GEs and are thought to form disulfide bonds (Fig. 1-5). BLAST search of *Aspergillus* genome database (http://www.aspgd.org; Cerqueira et al. 2014) showed that *A. fumigatus* Afu6g14390 possesses a sequence that displays a high identity (53.9%) with Cip2 and contains these conserved features, including the positions of six Cys residues (Fig. 1-6).

A phylogenic tree was drawn to observe the relationship between the characterized GEs and other carbohydrate esterases (CEs) (Fig 1-7). As shown, GEs are differently grouped from other CE families such as acetyl xylan esterase and feruloyl esterase. In the past, many putative GE sequences were categorized into the acetyl xylan esterase group due to the lack of information about the specific sequence of GEs (Agger et al. 2017). The characterized GEs were separated into two branches: the ascomycota GEs that include *St*GE2 (*S. thermophile*), *Nc*GE, *Pa*GE1 (*P. anserina*), Cip2, and *Af*GE; and the basidiomycota GEs that include *Cu*GE (*C. unicolor*), *Sc*GE (*S. commune*), *Pc*GE2, and *Pc*GE1 (*P. chrysosporium*). *Nc*GE shows

higher identity with *St*GE (72%) than Cip2 (55%) in contrast to *Af*GE (45% and 54%, respectively).

1.2.2.2. Cloning and expression of *Af*GE

Since the gene encoding *Af*GE was predicted containing an intron, *A. oryzae* niaD300 strain was chosen to be the host for the heterologous expression using pUNA vector. The pUNA vector is a traditional expression vector for *A. oryzae* that contains the dextrin-inducible *amyB* promoter, *amyB* terminator, and the selection markers *niaD* and *Amp^r* for *A. oryzae* and *E. coli*, respectively (Fig. 1-8). A gene of interest is inserted to the *Sma* I site located between *amyB* promoter and terminator. *Af*GE was tagged with c-Myc and hexahistidine sequences at the C-terminus. The *A. oryzae* transformant harboring *Af*GE was cultured in 5×DPY (pH 8.0) medium containing dextrin as the carbon source to produce *Af*GE.

CBB staining of the culture supernatant showed that the target protein was secreted into the medium, whereas no corresponding protein was produced in the control strain transformed with the empty vector (Fig. 1-9). Western blot analysis using anti-c-Myc antibody clearly confirmed the expression of *Af*GE which appeared as at least 3 bands at 44-48 kDa. The predicted size of *Af*GE-Myc-His is 44 kDa. Since *Af*GE has two potential *N*-glycosylation sites, two upper bands may be due to heterologous *N*-glycosylation, which will be analyzed in the next Chapter. The highest protein production was achieved after 3 days of culture. The supernatant obtained from the *A. oryzae* transformant showed the ability to hydrolyze the substrate A described above, while the supernatant from the control strain did not (Fig. 1-9).

1.3. Discussion

1.3.1. *Nc*GE expression

N. crassa has the ability to metabolize both cellulose and hemicellulose from the plant cell walls (Znameroski et al. 2012). In this report, I showed that the putative protein NCU09445.7 in *N. crassa* is actually a functional glucuronoyl esterase. CBM involved in the interaction of cellulolytic enzymes with the insoluble substrates was not found in the *Nc*GE sequence. This is similar to *Sc*GE from *S. commune* (Wong et al. 2012), *Pc*GE2 from *P. chrysoporium* (Ďuranová et al. 2009b), and *St*GE2 from *S. thermophile* (Topakas et al. 2010). In contrast, Cip2, *Pc*GE1 from *P. chrysoporium* (Ďuranová et al. 2009b), and *Cu*GE from *C. unicolor* (d'Errico et al. 2015) contain type 1 CBM at the N-terminus of the catalytic domain. CBMs are classified into subfamilies based on the amino acid sequence similarity and most of the fungal CBMs in cellulases belong to the family 1, in which three aromatic residues create a planar structure to bind to the crystallized surfaces of chitin/cellulose (Mello and Polikarpov 2014).

When the recombinant *Nc*GE was expressed in *P. pastoris* using pPICZαA vector, a protein of 44 kDa which is nearly equal to the calculated size of c-Myc- and hexahistidine-tagged *Nc*GE was produced. This agrees well with the absence of potential *N*-glycosylation site in *Nc*GE (http://www.cbs.dtu.dk/services/NetNGlyc/). Although 15 *O*-glycosylation sites were predicted by NetOGlyc 4.0 server (http://www.cbs.dtu.dk/services/NetOGlyc/), it is unlikely that *Nc*GE was highly *O*-glycosylated.

1.3.2. *Af*GE expression

A. fumigatus has a diverse range of lignocellulolytic enzymes. Even though this fungus is a pathogenic species, it has a larger number of biomass-degrading enzymes such as cellulases and hemicellulases than *T. reesei*; a total of 263 glycoside hydrolases (GHs) were found in *A. fumigatus*, while *T. reesei* contains 200 GH-encoding genes (Adav et al. 2015). Indeed, *A. fumigatus* is often found on decaying plants and organic debris (Latgé 1999) and is suggested to play an important role in the carbon cycle. Thus, its lignocellulolytic enzymes would be of potential use if they are heterologously expressed in a safe organism. *A. oryzae* is non-pathogenic and has been used as a host for heterologous protein production due to its prominent ability to secrete a large amount of protein into the culture medium (Uchima et al. 2011). In this study, I therefore used *A. oryzae* as a host for production of *Af*GE. Successful expression

of *C. unicolor Cu*GE in *A. oryzae* further prompted me to use *A. oryzae* as a production host (d'Errico et al. 2015). As expected, recombinant *Af*GE was produced in the culture supernatant of the *A. oryzae* transformant. Moreover, the enzyme clearly showed the activity toward the synthetic substrate A in the GE assay.

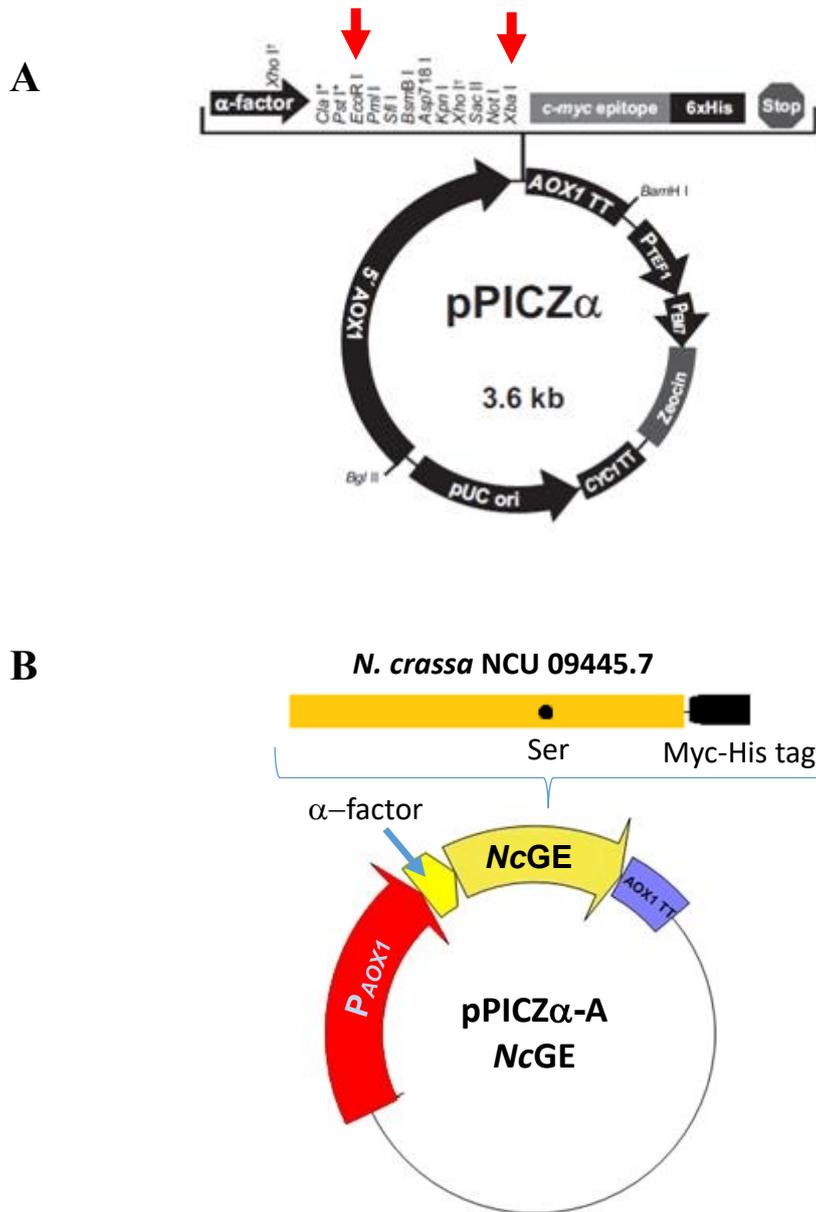**1.3.3. Identification of novel consensus sequences of GEs**

The amino acid sequences of *Nc*GE and other characterized GEs with experimentally-proven GE activity were aligned (Fig. 1-5). It is evident that the motif GC<u>S</u>RXG containing the nucleophile Ser is strongly conserved. This consensus sequence was identified by aligning the sequences of three characterized GEs, *T. reesei* Cip2, *S. thermophile St*GE2, *and P. chrysosporium Pc*GE1, with other putative GE sequences showing more than 40% amino acid sequence identity to *St*GE (Topakas et al., 2010). In this study, using the sequences of seven characterized fungal GEs, I propose that this motif could be expanded to VT<u>GCSRXG</u>KGA (the original motif is underlined). Besides, based on the crystal structures of Cip2 (Pokkuluri et al. 2011) and *St*GE2 (Charavgi et al. 2013), the catalytic triad of GEs was suggested to be composed of two more amino acids, Glu and His, located C-terminal to the nucleophile Ser. The molecular docking study of *St*GE2 with the model substrate (Katsimpouras et al. 2014) also supported this idea, which led to the hypothesis that Ser-Glu-His is the standard catalytic triad for GEs. Taking these into consideration, I propose two more consensus sequences constituting the catalytic triad of GEs: PQ<u>E</u>SG and <u>H</u>C containing the Glu and His residues (underlined; positions 328 and 439 in *Nc*GE, respectively). These sequences are strongly conserved in all GE sequences, suggesting again the importance of these residues as the components of catalytic triad, especially in the fungal GEs.

In support of my proposal, the crystal structures of Cip2 and *St*GE (Pokkuluri et al. 2011; Charavgi et al. 2013) showed that the disulfide bond is formed between the Cys residues equivalent to Cys277-Cys412 in Cip2, linking the consensus sequences VT<u>GCSRXG</u>KGA and <u>H</u>C, hence bringing Ser and His residues close together in the catalytic center (Charavgi et al. 2013; Katsimpouras et al. 2014). Cys309 involved in another disulfide bond with Cys384 in Cip2 is near E328 and both of these Cys residues are also highly conserved, possibly assisting in introducing Glu to create the catalytic triad. In addition, there is also a conserved pair of Cys (96-131) present in the N-terminal region of Cip2 which would assist the correct folding of the protein. Therefore, these three disulfide bonds may contribute to forming the catalytic triad in GEs. Although my proposal needs further verification by additional experiments, it might be informative in identifying novel GE sequences and elucidation of their catalytic mechanism.

The identification of *Af*GE from *A. fumigatus* genome based on these guidelines gives a support for this statement (Fig 1-6).

```
         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               10         20         30         40         50         60         70         80         90        100
Cip2_GE  MASRFFALLL LAIPIQAQSP VWGQCGGIGW SGPTTCVGGA TCVSYNPYYS QCIPSTQASS SIASTTLVTS FTTTTATRTS ASTPPASSTG AGGATCSALP
NcGE     MVHLTPALLL ASAAFAAAAP AS-------- ---------- -----QIFER QCS------- ---------- --------VA GNYPTAAVS- ----------
Clustal Co  *.     ****  : .: * :* .                    :  :  **                    .: .. *.*: :

         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              110        120        130        140        150        160        170        180        190        200
Cip2_GE  GSITLRSNAK LNDLFTMFNG DKVTTKDKFS CRQAEMSELI QRYELGTLPG RPSTLTASFS GNTLTINCGE AGKSISFTVT ITYPSSGTAP YPAIIGYGGG
NcGE     ---------K LPDPFTTAAG QKITTKADFD CRKAEISKIL QQYELGTYPG KPDKVEGSLS GNTLTVRITV GSQTVSFSAS IKKPSSGSGP FPAIIGIGGI
Clustal Co       * * **    * :*:*** .*. **.**:*::: *:***** ** :*.:. .*:* *****:.    ..:::**:.: *. **** .* :***** **

         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              210        220        230        240        250        260        270        280        290        300
Cip2_GE  SLPAPAGVAM INFNNDNIAA QVNTGSRGQG KFYDLYGSSH SAGAMTAWAW GVSRVIDALE LVPGAR-IDT TKIGVTGCSR NGRGAMVAGA FEKRIVLTLP
NcGE     SIPIPSTVAT ITFPNDDFAQ QSGTSSRGRG KFYTLFGSSH SAGALIAWAW GVDRLVDALE QVQSTSGIDP KRLGVTGCSR NGKGAFVAGA LVDRIALTIP
Clustal Co  *:* *: **  *.* **:* .   * .*.***:* *** *:**** ****: **** **.*::**** *  .:   **. .::****** ****  * .**.**:*

         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              310        320        330        340        350        360        370        380        390        400
Cip2_GE  QESGAGGSAC WRISDYLKSQ GANIQTASEI IGEDPWFSTT FNSYVNQVPV LPFDHHSLAA LIAPRGLFVI DNNIDWLGPQ SCFGCMTAAH MAWQALGVSD
NcGE     QESGAGGNAC WRISDSEKSA GKNIQTASQI VTENVWFSPA FNAYTRQTTN IPADHHMLAA LTVPRGLIAF ENDIDWLGPV STTACMQAGR LIYKAYGVSN
Clustal Co  *******:** *****  **  * *******:* : *: ***.: **:*..*..  :* *** *** * .****:.: :*:****** *  .** *.: : ::* ***:

         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ..
              410        420        430        440        450        460
Cip2_GE  HMGYSQIGAH AHCAFPSNQQ SQLTAFVQKF LLGQSTNT-A IFQSDFSANQ SQWIDWTTPT LS
NcGE     HMGFSLVGGH GHCQFPSSQQ SELTSYINYF LLKAGTAPGA VERSSAKVDL KSWAPWDVPA LS
Clustal Co  ***:* :*.* .** ***.** *:**:::: * **  .* . * : :*. ..:  ..* * .*: **
```

Figure 1-1: Amino acid sequence alignment of Cip2 and *Nc*GE

The amino acid sequence alignment of Cip2 (*T. reesei* AAP57749) and *Nc*GE (*N. crassa* NCU09445.7) is shown. The underlines show the signal peptides, and the shaded letters indicate the CBM1 in Cip2. The consensus GCSRXG sequence (Topakas et al., 2010) is boxed.

A

B

Figure 1-2: pPICZαA for protein expression in *P. pastoris*

(A) The structure of pPICZαA plasmid (Invitrogen 2010). (B) The scheme for cloning the *Nc*GE gene into pPICZαA. C-terminally c-Myc- and His$_6$-tagged *Nc*GE sequence was inserted between the *Eco*R I and *Xba* I sites (red arrows) of the plasmid.
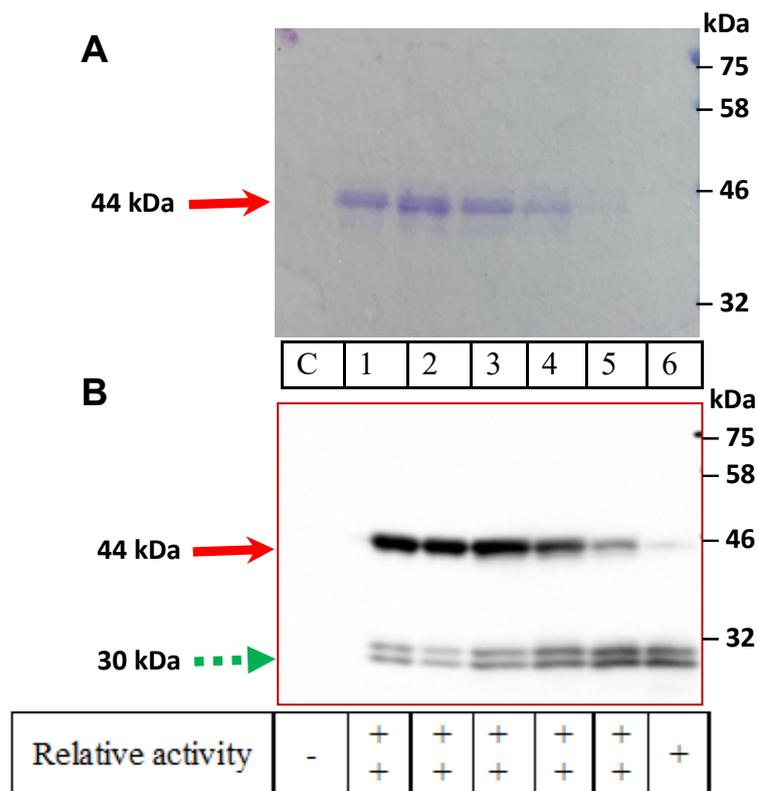
**Figure 1-3: Production of *Nc*GE in *P. pastoris* using pPICZαA vector**

(A) CBB staining and (B) Western blot analysis using anti-c-Myc antibody of the culture supernatants of *P. pastoris* pPICZαA/*Nc*GE transformant. Lanes 1-6 indicate the samples from day 1-6 of methanol induction, respectively. C, control (empty vector) transformant grown for 2 days.
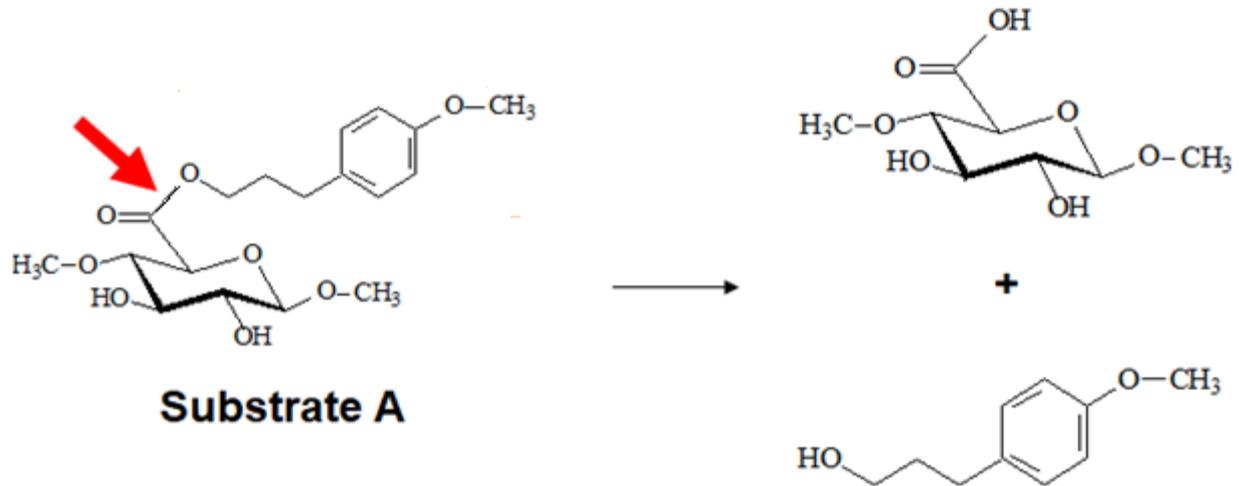
Figure 1-4: Substrate A used for the detection of GE activity

Substrate A, 3-(4-methoxyphenyl) propyl methyl 4-O-methyl-α-D-glucopyranosiduronate, used for the GE activity assay. Two products are generated in the reaction: methyl 4-O-methyl-D-glucuronic acid (right, top) and 3-(4-methoxyphenyl) propyl alcohol (right, bottom). The arrow indicates where the enzyme cleaves the substrate.
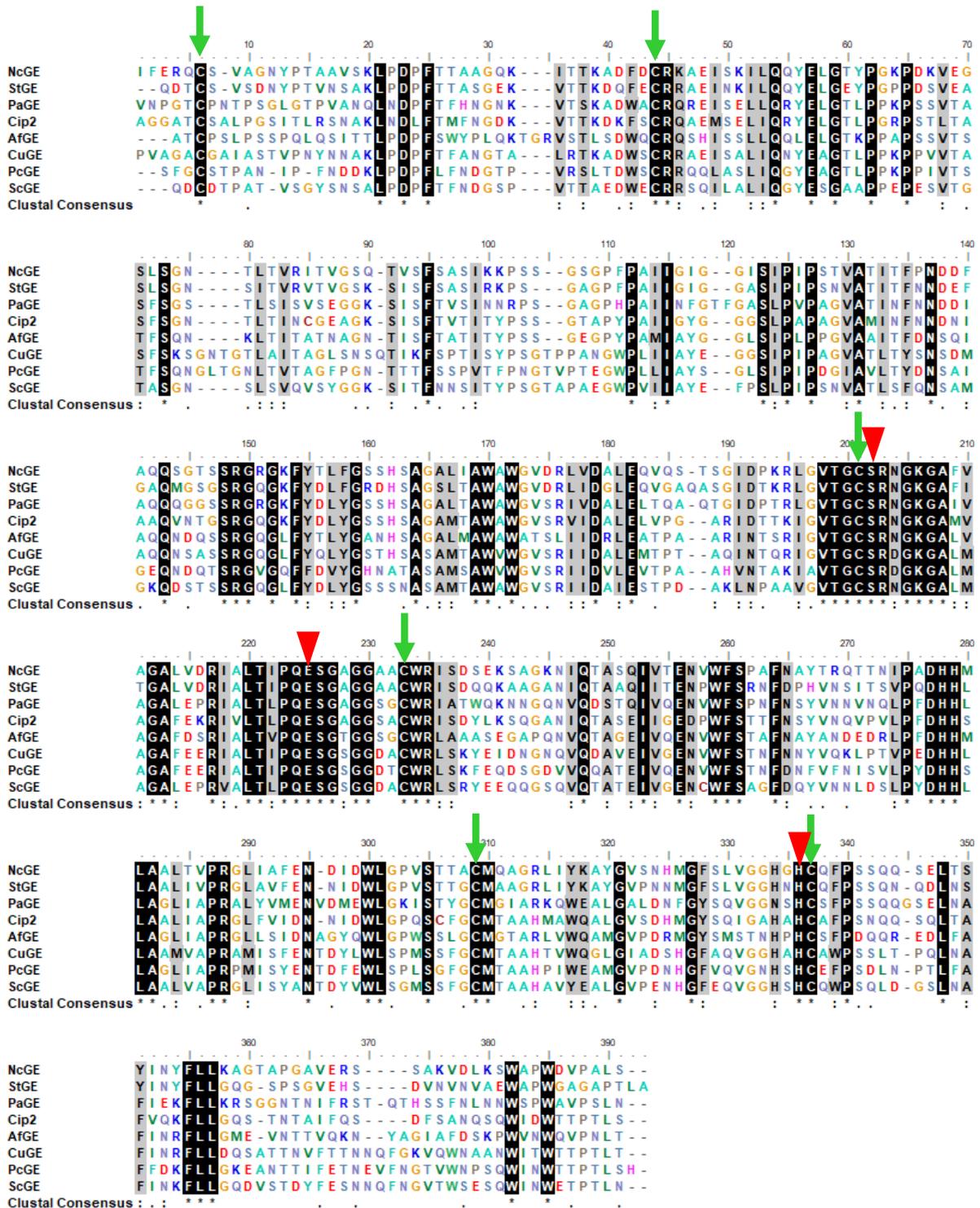
Figure 1-5: Alignment of characterized GEs

The putative catalytic triad residues are marked by arrowheads. Conserved Cys residues linked by disulfide bonds are indicated by the solid arrows.
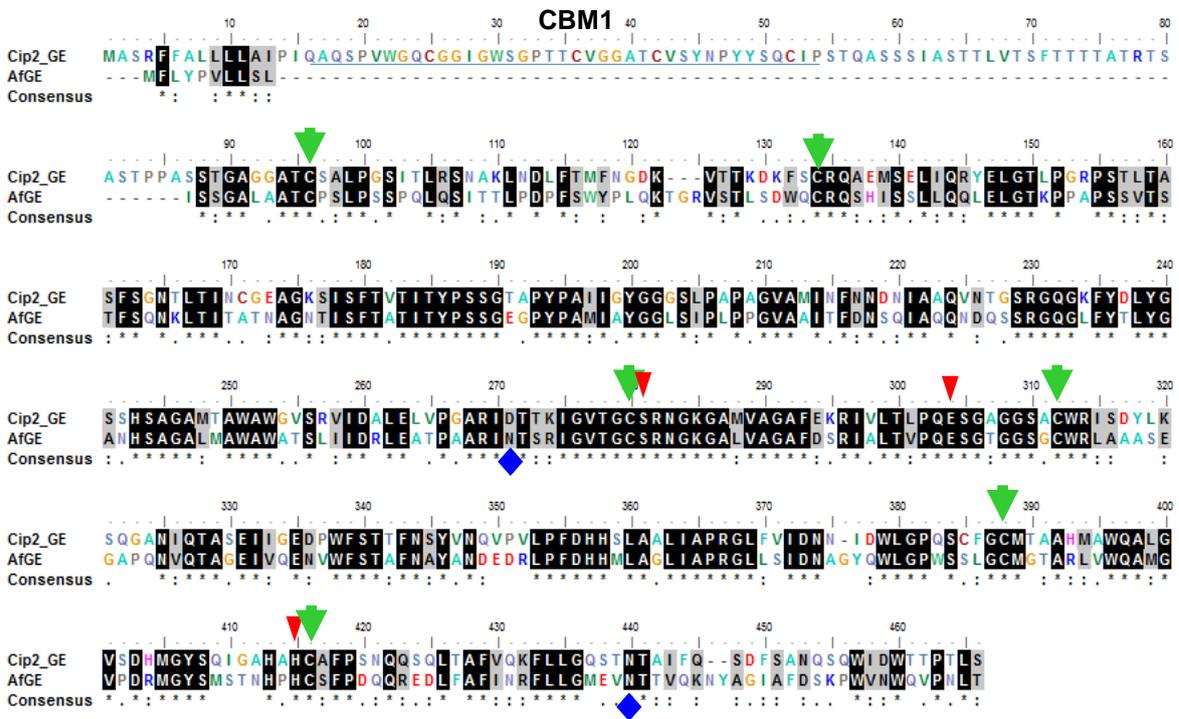
Figure 1-6: Alignment of amino acid sequences of the mature Cip2 and *Af*GE enzyme

Arrowheads (▼) indicate the conserved amino acids forming the catalytic triad of GE; arrows (⬇) indicate the cysteines linked by disulfide bonds. Diamonds (◆) indicate the potential *N*-glycosylation sites in *Af*GE. The sequence of CBM1 in Cip2 was underlined.
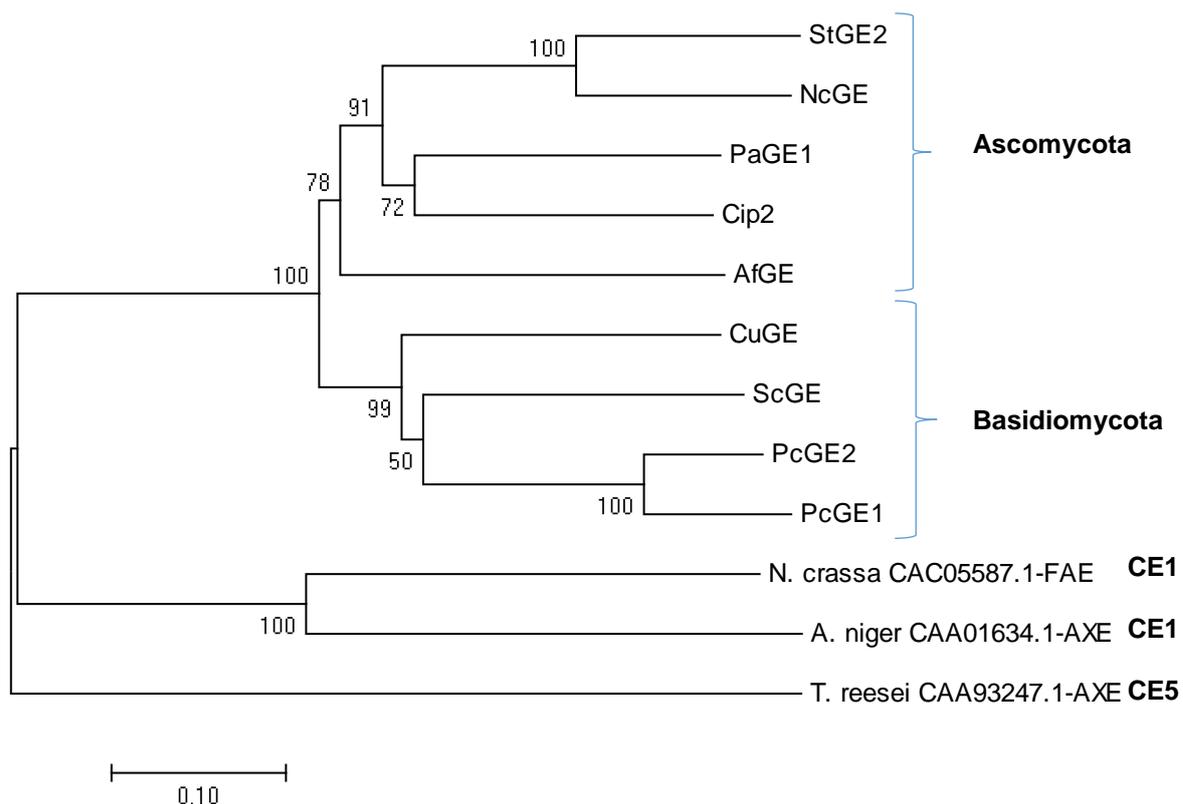
Figure 1-7: Phylogenetic tree of *Nc*GE, *Af*GE, and other fungal GEs

The data on characterized GEs and other enzymes were obtained from the carbohydrate-active enzymes (CAZy) database (www.cazy.org). The phylogeny diagram was built by MEGA6 software with the out-group sequences including two different acetyl xylan esterases (AXEs) from *T. reesei* (CAA93247.1; CE family 5) and *Aspergillus niger* (CAA01634.1; CE family 1), and one ferulic acid esterase (FAE) from *N. crassa* (CAC05587.1; CE family 1).
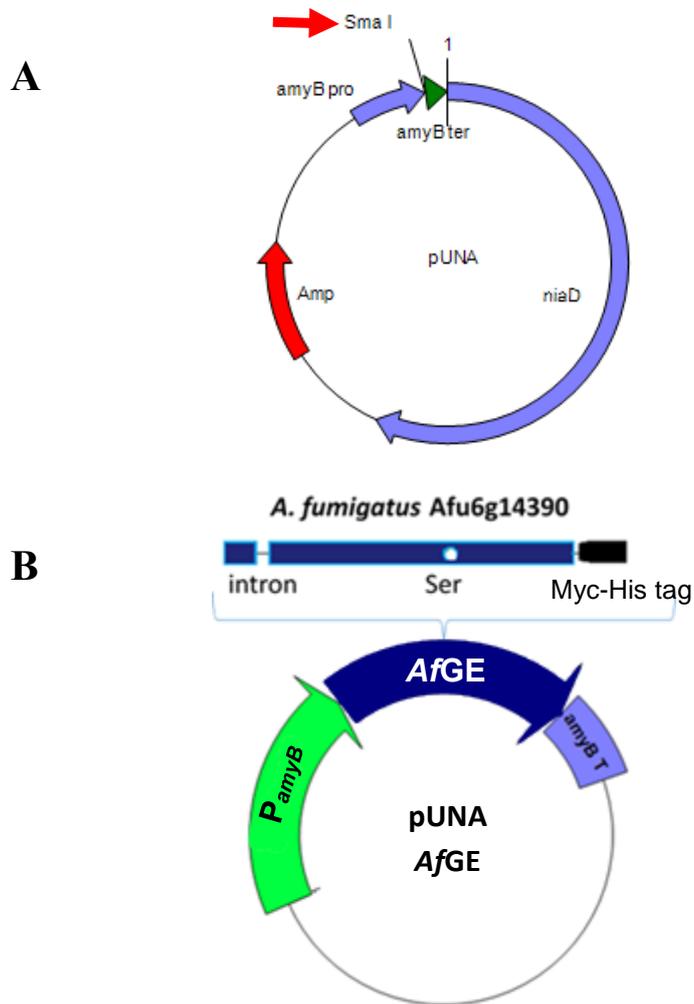
Figure 1-8: pUNA for protein expression in *A. oryzae*

(A) The structure of pUNA plasmid. (B) The scheme for cloning the full-length *Af*GE gene (including signal peptide and intron) into pUNA vector. Myc-His tag was added to the C-terminus of the putative mature region of *Af*GE. The *Af*GE sequence was inserted into the vector at *Sma* I site shown by the red arrow.
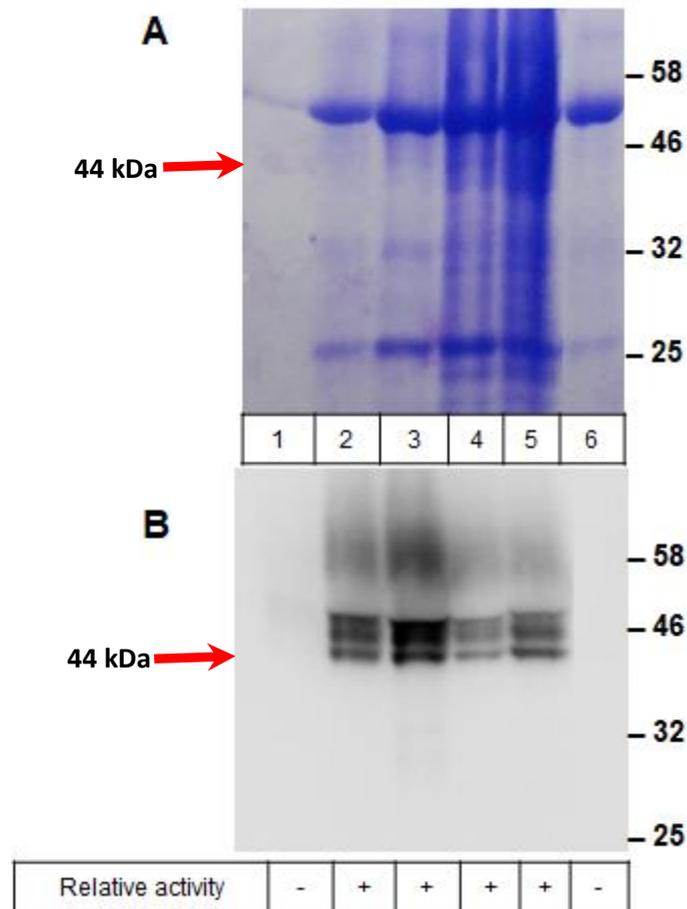
Figure 1-9: Expression of *Af*GE using *A. oryzae* as a host

(A) CBB staining and (B) Western blot analyses of the culture supernatants of pUNA-*Af*GE-Myc-His tag *A. oryzae* transformant. Lanes 1-5 indicate the samples from day 1-5, respectively. Lane 6, control (empty vector) transformant grown for 3 days. For SDS-PAGE, 20 µl of supernatant from each sample was loaded.

2.1. Introduction

After the target enzymes were obtained, the purification process was performed to determine their characteristics. Because the recombinant proteins were tagged with the hexahistidine sequence, the Ni^{2+}-NTA affinity chromatography was applied in the first step. The ion exchange chromatography followed to achieve higher purity. At this point, the enzymes displayed sufficient quality to determine their enzymatic properties.

In order to get a high hydrolysis efficiency, the enzymes need to be operated in the optimum condition. In this study, the effects of some typical factors such as pH, temperature, metal ions, and chemical reagents were investigated. Besides, almost all the studies of GEs relied on the synthetic compounds due to the unavailability of the proper natural substrate. Among these compounds, the esters of glucuronic acid derivatives and aromatic alcohol were more suitable. They mimic better the structure of the ester linkage between hemicellulose and lignin, the latter of which is considered as the phenolic component of biomass. The aromatic group can be easily detected by the HPLC-based assay. The HPLC method provides a highly precise activity measurement, enabling the enzyme kinetics measurement (Sunner et al. 2015). Herein, the synthetic substrate 3-(4-methoxyphenyl) propyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate (substrate A) that mimics the ester linkage between hemicellulose and lignin was used (Fig. 2-1; Špániková et al. 2007; Sasagawa et al. 2011). In addition, a commercially-available compound, benzyl D-glucuronate (substrate B; Sunner et al. 2015), was also exploited to check the GE activity. Substrate A was mainly used to determine the enzymatic properties of two GEs.

2.2. Results

2.2.1. Purification of two fungal GEs

2.2.1.1. Purification of *Nc*GE

*Nc*GE was purified from the culture supernatant of *P. pastoris* transformant firstly by Ni^{2+}-NTA affinity chromatography. The purification was performed using Ni-NTA agarose. It seemed that the target protein did not strongly bind to the column, since a small amount of protein was detected in the flow through (Fig. 2-2, lane 2) and wash (Fig. 2-2, lanes 3-6) fractions. The majority of *Nc*GE was eluted by the buffer containing 100 mM imidazole (Fig. 2-2, lanes 7 and 8).

Secondly, to achieve a higher purity, the affinity-purified *Nc*GE was applied to ion exchange chromatography. Since the predicted pI of *Nc*GE-Myc-His is about 8.8, the purification of *Nc*GE was performed by using HiTrap SP XL cation exchange column at pH 7.0. A band of 44 kDa, which corresponds to the calculated size of c-Myc- and hexahistidine-tagged *Nc*GE, was purified to near homogeneity. The activity of purified fractions were also confirmed (Fig. 2-3).

2.2.1.2. Purification of *Af*GE

Firstly, the supernatant of *A. oryzae* transformant was applied to Ni-NTA agarose column where Tris-HCl buffer (50 mM, pH 8) was used as the binding buffer. Unfortunately, hexahistidine-tagged *Af*GE did not seem to strongly bind to the column, since a significant amount of *Af*GE was detected in the flow-through fraction. Thus, HisTrap FF column and sodium phosphate buffer (20 mM, pH 8) as the binding buffer were selected to be used in the purification as suggested by the manufacturer (GE healthcare, Tokyo, Japan). The bound protein was then eluted by 100 mM imidazole (Fig. 2-4). *Af*GE was further purified by anion exchange chromatography at pH 7.0 (Fig. 2-5), since the predicted pI of *Af*GE-Myc-His was about 6.1. In each step of purification procedure, the substrate A (Fig. 2-1) was used to detect the enzyme activity. The fractions containing *Af*GE displayed the ability to degrade the substrate A, whereas those without *Af*GE did not (data not shown).

*Af*GE-Myc-His has the predicted size of about 44 kDa, but a smear of upper bands was detected by CBB staining and Western blot. After treatment with endoglycosidase H (Endo H), only the 44 kDa band remained and its signal became stronger while the upper bands in the CBB-stained gel and Western blot disappeared (Fig. 2-5, D and E). This result indicates that *Af*GE undergoes heterologous *N*-glycosylation when expressed in *A. oryzae*.

2.2.2. Characterization of two fungal GEs

2.2.2.1. Enzymatic characterization of *Nc*GE

Effect of pH and temperature

The optimum pH and temperature for *Nc*GE were pH 7.0 and 40-50°C, respectively (Fig. 2-6). *Nc*GE was stable in the relatively broad pH range (pH 4-7). The enzyme was stable at 30°C, but it completely lost its activity when pre-incubated at 70°C for 30 min. Thus, the optimal conditions of *Nc*GE lie in the pH range 5-7 and the temperature around 40°C-60°C which are normally observed in the currently described GEs (d'Errico et al. 2015). There are no available data in which the temperature profiles of GEs were examined by employing the same substrate used in this study; in Cip2 and *St*GE2 (both of which were also heterologously expressed in *P. pastoris* using pPICZα vector), the pH and temperature profiles were determined using 4-nitrophenyl 2-*O*-(methyl-4-*O*-methyl-α-ᴅ-glucopyranosyluronate)-β-ᴅ-xylopyranoside as a substrate (Li et al. 2007; Topakas et al. 2010). Both GEs showed higher optimum temperature (40-60°C and 50-55°C for Cip2 and *St*GE2, respectively). In addition, Cip2 was quite stable at 40°C (Li et al. 2007) and *St*GE2 displayed significant stability up to 50°C (Topakas et al. 2010). This is in contrast to *Nc*GE which quickly lost its activity at higher temperatures, especially above 50°C, despite that *Nc*GE is most closely related to *St*GE2 (72% identity) in the phylogenic tree (Fig. 1-7).

Effect of metal ions and other reagents

The effect of cations and chemical compounds on the activity of *Nc*GE is shown in Table 2-1. The metal ions such as Mg^{2+}, Fe^{3+}, and Al^{3+} modestly increased the enzyme activity, and the highest effect was observed with Ca^{2+} and Tween 80. In contrast, significant inhibition (by around 40%) was caused by Cu^{2+} and Ni^{2+}. The enzyme was affected neither by glycerol, ascorbic acid, nor by two potential inhibitors acetic anhydride and dimethyl sulfoxide. As expected, the activity of *Nc*GE was significantly inhibited by the compounds influencing the enzyme structure, such as dithiothreitol, 2-mercaptoethanol, and urea, while the enzyme was completely inhibited by SDS.

As mentioned above, *Nc*GE has the consensus sequence for GEs containing the nucleophile Ser residue. In agreement, *Nc*GE activity was decreased by 60% when the enzyme was incubated with 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and the activity was totally lost in the presence of 1 mM PMSF (data not shown), demonstrating that this enzyme is truly a serine type-esterase. In contrast, *Nc*GE activity was not affected by 5 mM EDTA, suggesting

that it is not a metalloenzyme.

*Kinetic parameter of Nc*GE

Next, the kinetic constants of *Nc*GE were determined using the substrates A and B (Table 2-2). The K_m values were around 15 and 34 mM, respectively. With the substrate A, *Sc*GE, a GE from *S. commune*, K_m was much lower (1.78 mM; Špániková et al. 2007), indicating that *Sc*GE has higher affinity for this substrate, which is reflected in the higher catalytic efficiency (k_{cat}/K_m was reported to be 4.38 mM^{-1}s^{-1} in *Sc*GE). On the other hand, the k_{cat}/K_m ratio of *Nc*GE was similar to that of Cip2 (1.1 mM^{-1}s^{-1}; Li et al. 2007). *Nc*GE showed the lowest catalytic efficiency of around 0.24 mM^{-1}s^1 toward the substrate B compared with *Pa*GE1 and *St*GE2. However, its turnover number (k_{cat} value) was higher than other GEs, indicating a higher speed of the substrate conversion into the product.

2.2.2.2. Enzymatic characterization of *Af*GE

Effect of pH and temperature

*Af*GE showed a high activity and stability from pH 5 to neutral pH with the optimum activity observed at pH 5 (Fig. 2-7A). Its activity was significantly lost after pre-incubation in the pH lower than 4 or higher than 8. The highest activity was observed at 40°C, while at 70°C, the enzyme lost around 50% of its maximum activity. Besides, the enzyme displayed significant stability in the temperatures up to 50°C (more than 90% of its maximum activity was retained) and at 60°C, the residual activity was still around 70% (Fig. 2-7B). These results suggested that *Af*GE is generally a thermophilic enzyme.

Effects of metal ions and other reagents

The effects of cations and chemical compounds on *Af*GE activity are listed in Table 2-3. The activity of *Af*GE was slightly inhibited by Al^{3+}, Cu^{2+}, Li$^+$, Mn^{2+}, and Fe^{2+}, while it increased by Mg^{2+}. The remaining metal cations significantly inhibited *Af*GE (by ≈40%) except Zn^{2+} (no effect). The enzyme activity was not affected by most of the reagents tested. Interestingly, *Af*GE activity was significantly enhanced by Tween 80 (by 2.3 times). Similar enhancement was also observed in *Nc*GE (~1.5 times). Since the substrate A contains a hydrophobic aromatic alcohol group, Tween 80 might have increased its solubility and thus improved the accessibility of the enzymes to the substrate (Zheng et al. 2008). Dithiothreitol and 2-mercaptoethanol partially inhibited *Af*GE activity, suggesting the essential role of three

putative disulfide bonds in *Af*GE.

As the mentioned above, *Af*GE has the consensus sequence containing the putative nucleophile Ser residue. In agreement, *Af*GE activity was inhibited by PMSF, suggesting that this enzyme is truly a serine type-esterase. In contrast, *Af*GE activity was not affected by EDTA, indicating that this enzyme is not a metalloenzyme.

Kinetic parameter of **Af*GE***

Kinetic parameters for *Af*GE were determined using various substrates (Table 2-2). For both substrates A and B, *Af*GE displayed quite high K_m values of 15.8 mM and 16.4 mM, respectively. This means that these compounds were not preferred by *Af*GE. However, the k_{cat} value of *Af*GE was much higher than those of other GEs determined using the same substrates; as to the substrate A, k_{cat} for *Af*GE was 67.1 s^{-1}, which was 9-60 times higher than those for *Nc*GE and *Sc*GE; as to the substrate B, when compared with *Pa*GE1 and *St*GE2, k_{cat} for *Af*GE was 3-7 times higher. Thus, the de-esterifying reaction by *Af*GE proceeds at a significantly high rate, resulting in comparable (compared with *Sc*GE) or higher (compared with *Pa*GE1 and *St*GE2) catalytic efficiency (k_{cat}/K_m) toward the substrates A or B, respectively. Cited from the kinetic parameters of a recent study, *Af*GE also displayed much higher k_{cat} and k_{cat}/K_m values than *Aa*GE1 (a GE from *Acremonium alcalophilum*), *Pc*GE1, and *Wc*GE1 (a GE from *Wolfiporia cocos*) toward the substrate B (Hüttner et al. 2017).

2.3. Discussion

2.3.1. Characterizations of *Nc*GE

Toward the substrate A, *Nc*GE displayed much higher K_m value (15 mM) than *Sc*GE (1.78 mM). It thus seems that *Nc*GE does not prefer this substrate, although it clearly cleaved the ester linkage as a typical GE. The reason may be the synthetic substrate mimics but does not exactly represent the natural bond of the target substrate (Wong et al. 2012). Therefore the natural substrate and physiological role of *Nc*GE are yet to be elucidated. *Nc*GE showed the characteristic of a mesophilic enzyme with the temperature optimum at 40-50°C and displayed the optimal activity at pH 7 and broad pH stability.

2.3.2. Characterizations of *Af*GE

After purification by Ni^{2+}-affinity and anion exchange chromatographies, it was confirmed that *Af*GE is a functional glucuronoyl esterase. Although *Af*GE did not show high preference to the synthetic substrates, it exhibited a significantly high k_{cat} value for the substrate A (67.1 s^{-1}) leading to high catalytic efficiency. As to the thermostability, compared to the mesophilic nature of GEs such as Cip2 (Li et al. 2007) and *Nc*GE which are stable at around 40°C and 30°C, respectively, *Af*GE displayed the feature of a thermophilic enzyme in that the high activity was observed in the temperature range of 30-50ºC with the optimum temperature being around 40ºC. Thus *Af*GE is one of the thermostable GEs such as *Pa*GE1, *Pc*GE1, *Pc*GE2, *St*GE1, and *Cu*GE, maintaining its activity up to 50°C. Addition of *N*-glycans could be the reason for the thermostability of *Af*GE, since some studies have suggested that the glycosylated protein exhibits a higher thermostability than the non-glycosylated form (Han and Lei 1999; Clark et al. 2004; Benoit et al. 2006). In this regard, it is of note that *St*GE1 and *Cu*GE were shown to be *N*-glycosylated (Vafiadi et al. 2009; d'Errico et al. 2015), *Pa*GE1 and *Pc*GE1 were suggested to be *O*-glycosylated (Katsimpouras et al. 2014; Hüttner et al. 2017), and *Pc*GE2 has seven potential *N*-glycosylation sites (Ďuranová et al. 2009), whereas no *N*-glycosylation site was found in the mesophilic enzyme *Nc*GE. Moreover, the indirect comparison with other GEs using different substrates indicated that *Af*GE has the lowest pH optimum at 5.0. Only *Pa*GE1, *Pc*GE1, and *Pc*GE2 were shown to display the optimal activity in the range of pH 5-6 (Ďuranová et al. 2009; Katsimpouras et al. 2014). Interestingly, the enzymatic hydrolysis of lignocellulosic biomass by the commercial cellulase products such as Ultraflo® L or Cellic® CTec is often conducted at 50°C, pH 5.0 (d'Errico et al. 2016). Thus, *Af*GE could collaborate with these cellulases in the industrial degradation of lignocellulose.

2.3.3. Catalytic mechanism of GEs

It is generally accepted that the GE activity involves the function of the catalytic triad Ser-His-Glu. The consensus sequence containing the nucleophilic Ser was first mentioned by Topakas et al. and its important role was demonstrated by mutation studies (Topakas et al. 2010). The crystal structures of Cip2 and *St*GE2 further suggested the roles of His and Glu forming the catalytic triad (Pokkuluri et al. 2011; Charavgi et al. 2013). It was therefore concluded that GE is a serine-type esterase (Topakas et al. 2010; Biely 2016) which belongs to the serine hydrolase superfamily including the well-examined serine protease group (Simon and Cravatt 2010). In Chapter 1, I suggested the presence of expanded consensus sequences surrounding Ser, His, and Glu. The difference from serine proteases is that the catalytic triad of the serine protease group consists of Ser, His, and Asp; however, it is of note that the combination of Ser, His, and Glu also appears in aspartyl dipeptidase and in some other esterases (Håkansson et al. 2000; Hedstrom 2002). In addition, based on the crystal structure of *St*GE2, the catalytic Ser and the next amino acid (Arg) were suggested to form the structure called "oxyanion hole" (Charavgi et al. 2013). Taken all together, guided by the hydrolytic principle of serine proteases, the enzymatic mechanism of GE can be predicted as shown in Fig. 2-8. Since the substrate B has the simpler molecular structure, this substrate was used to demonstrate the predicted enzymatic mechanism of GEs.

Basically, the specific spatial arrangement of Ser-His-Glu is that the imidazole side chain of His is close enough to the hydroxyl group of Ser to form a hydrogen bond. In addition, the Glu residue is also close to His but located on the opposite side from Ser (Neitzel 2010). The reaction can be divided into two stages. In the acylation half of the reaction, Ser residue attacks the carbonyl group of the glucuronic acid ester substrate, assisted by His acting as a general base, to yield a tetrahedral intermediate. The resulting His-H$^+$ is stabilized by the hydrogen bond to the carboxylic acid group of Glu. The oxyanion of the tetrahedral intermediate is stabilized by hydrogen bond interactions with NH$_2$ moiety in the main chain of Ser and Arg residues which form the structure, the oxyanion hole. The tetrahedral intermediate collapses by the expulsion of the aromatic alcohol group. The remaining substrate is still linked to the enzyme, which is called the acyl-enzyme intermediate. The deacylation half of the reaction essentially repeats the above sequence: water assisted by His residue attacks the acyl-enzyme intermediated, which yields a second tetrahedral intermediate. Finally, this intermediate collapses, and thus restoring the original Ser and expelling another reaction product-a glucuronic acid derivative (Hedstrom 2002).
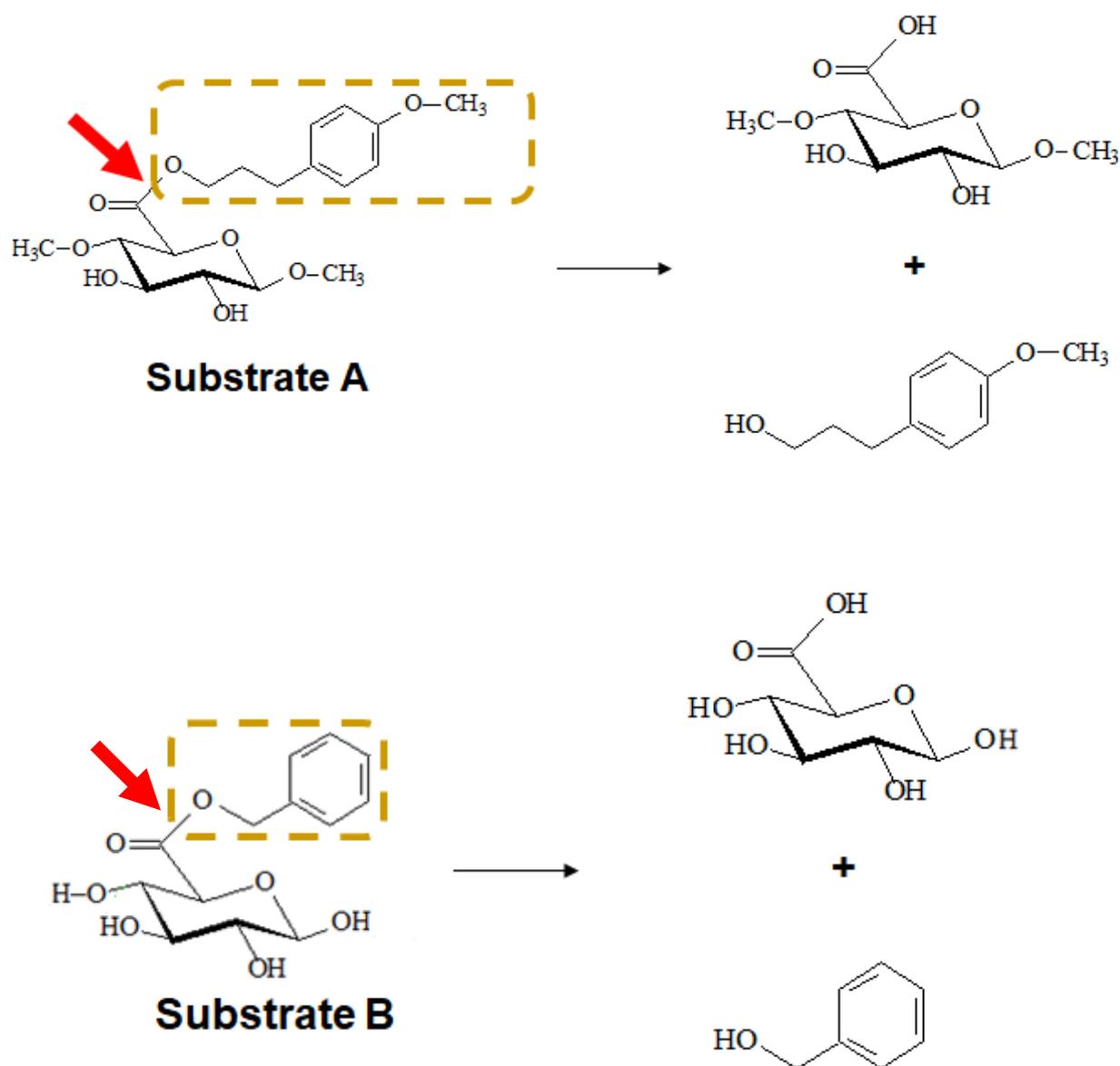
Figure 2-1: Chemical structure of substrates A and B

Substrate A, 3-(4-methoxyphenyl) propyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate; Substrate B, commercially available benzyl D-glucuronate. The aromatic groups detected by HPLC are boxed. The arrows indicate where the enzyme cleaves the substrate.
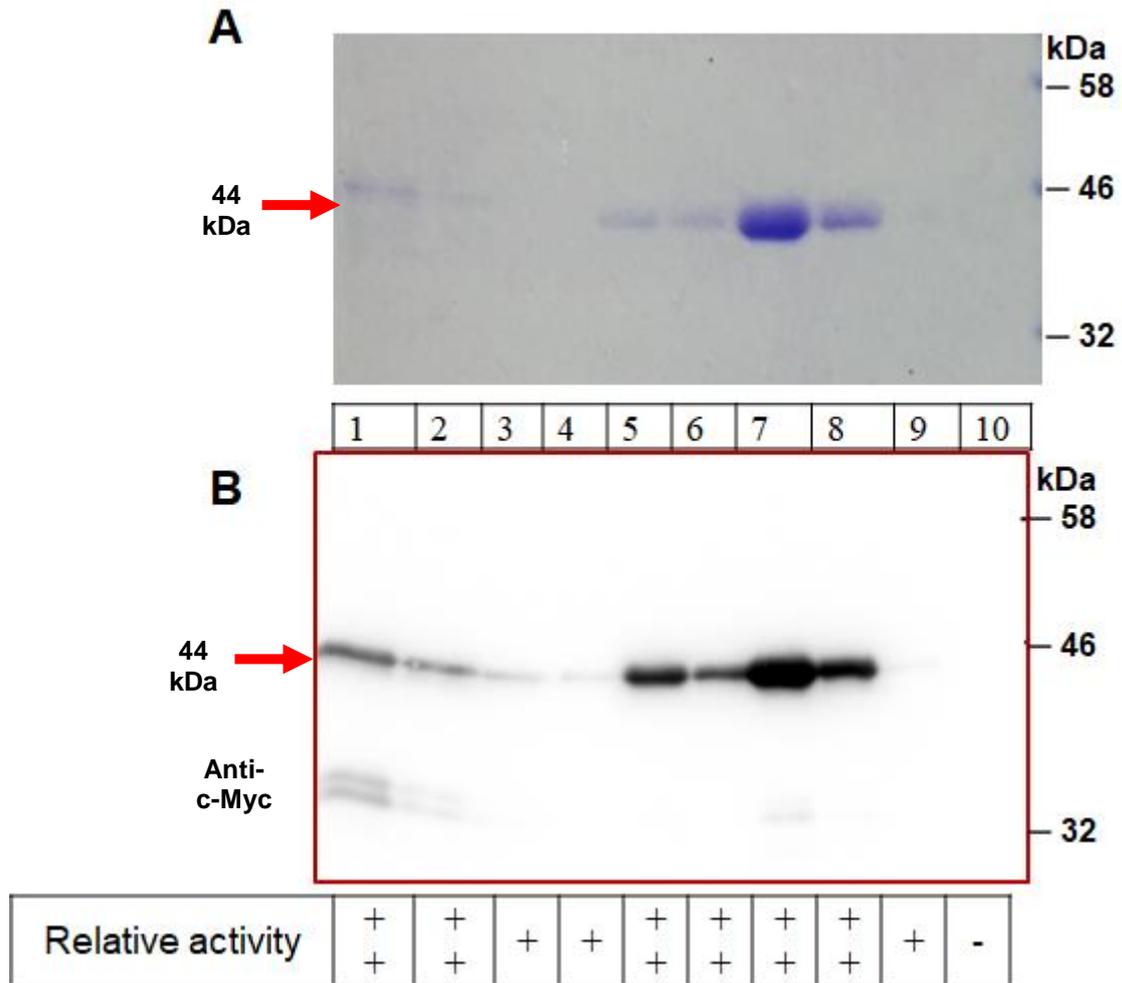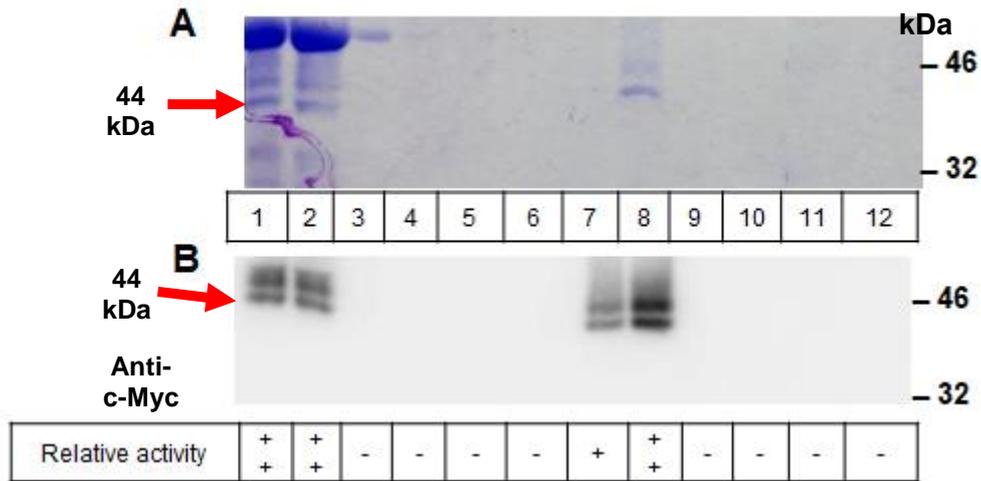
Figure 2-2: Purification of *Nc*GE by Ni^{2+}-NTA column chromatography

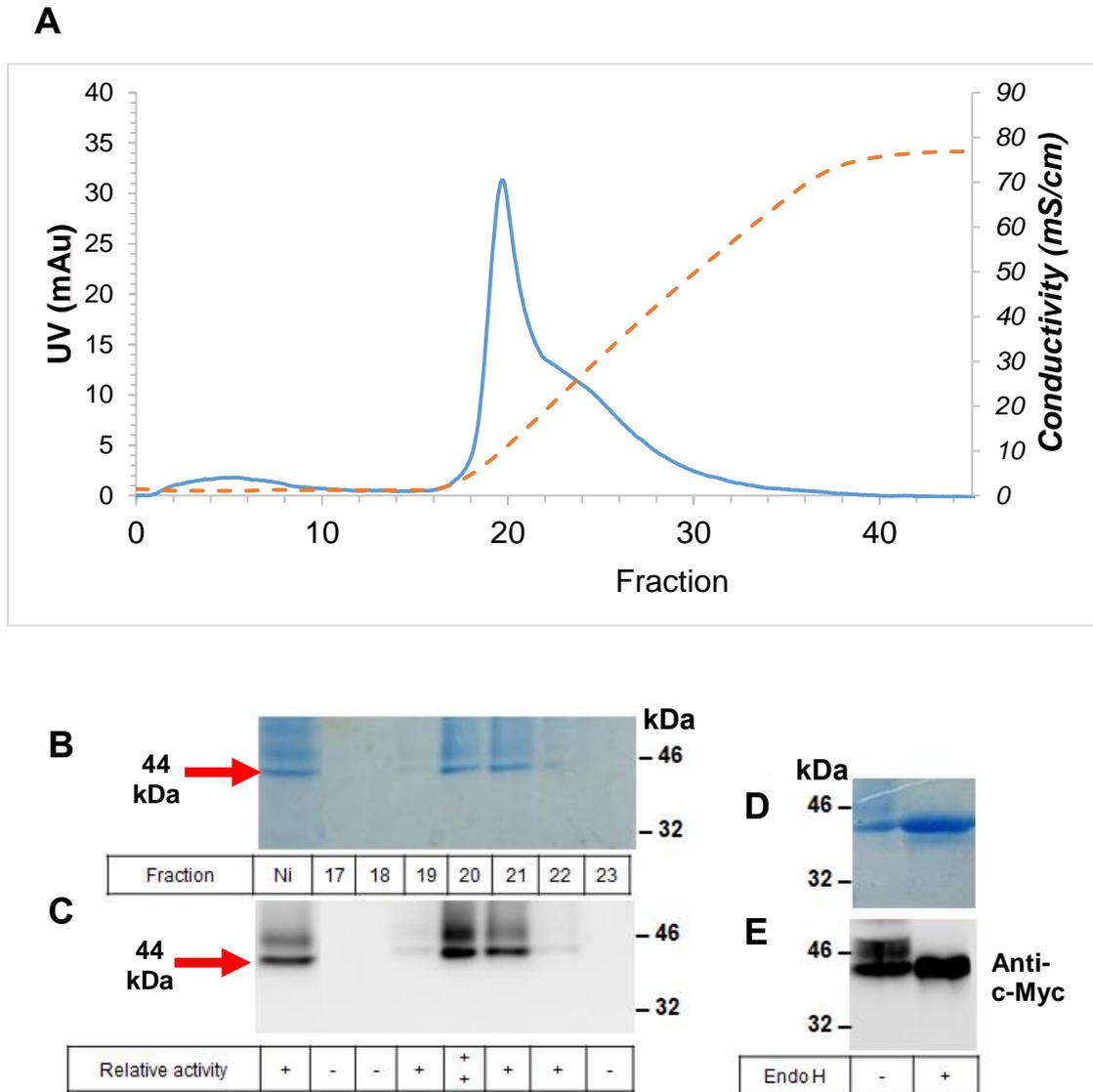(A) CBB staining and (B) Western blot analysis of Ni^{2+}-NTA column chromatography fractions. Lane 1, the culture supernatant of pPICZαA/*Nc*GE transformant; lane 2, flow through, lanes 3 and 4, wash with 0 mM imidazole; lanes 5 and 6, wash with 20 mM imidazole; lanes 7-10, elution with 100 mM imidazole.

Figure 2-3: Cation exchange chromatography of *Nc*GE

(A) A_{280} (solid line), conductivity (dashed line), and GE activity of each fraction of cation exchange chromatography. (B) Lanes 1-4, fractions 18-21; Ni, *Nc*GE sample before loading; M, protein marker.

Figure 2-4: Purification of *Af*GE by Ni^{2+}-NTA chromatography

(A) CBB staining and (B) Western blot analysis of Ni^{2+}-NTA chromatography. Lane 1, the culture supernatant of *A. oryzae* transformant; lane 2, flow through; lanes 3 and 4, wash with 0 mM imidazole; lanes 5 and 6, wash with 10 mM imidazole; lanes 7-10, elution with 100 mM imidazole; lanes 11 and 12, elution with 250 mM imidazole.

Figure 2-5: Anion exchange chromatography and deglycosylation analysis of *Af*GE
(A) A_{280} (solid line), conductivity (dashed line), and GE activity of each fraction of anion exchange chromatography at pH 7.0. (B) CBB staining and (C) Western blot analysis of *Af*GE after anion exchange chromatography. Ni, *Af*GE sample before loading; lanes 17-23 indicate the anion exchange chromatography fractions from 17 to 23, respectively. (D) CBB staining and (E) Western blot analysis of purified *Af*GEs before and after endoglycosidase H (Endo H) treatment.
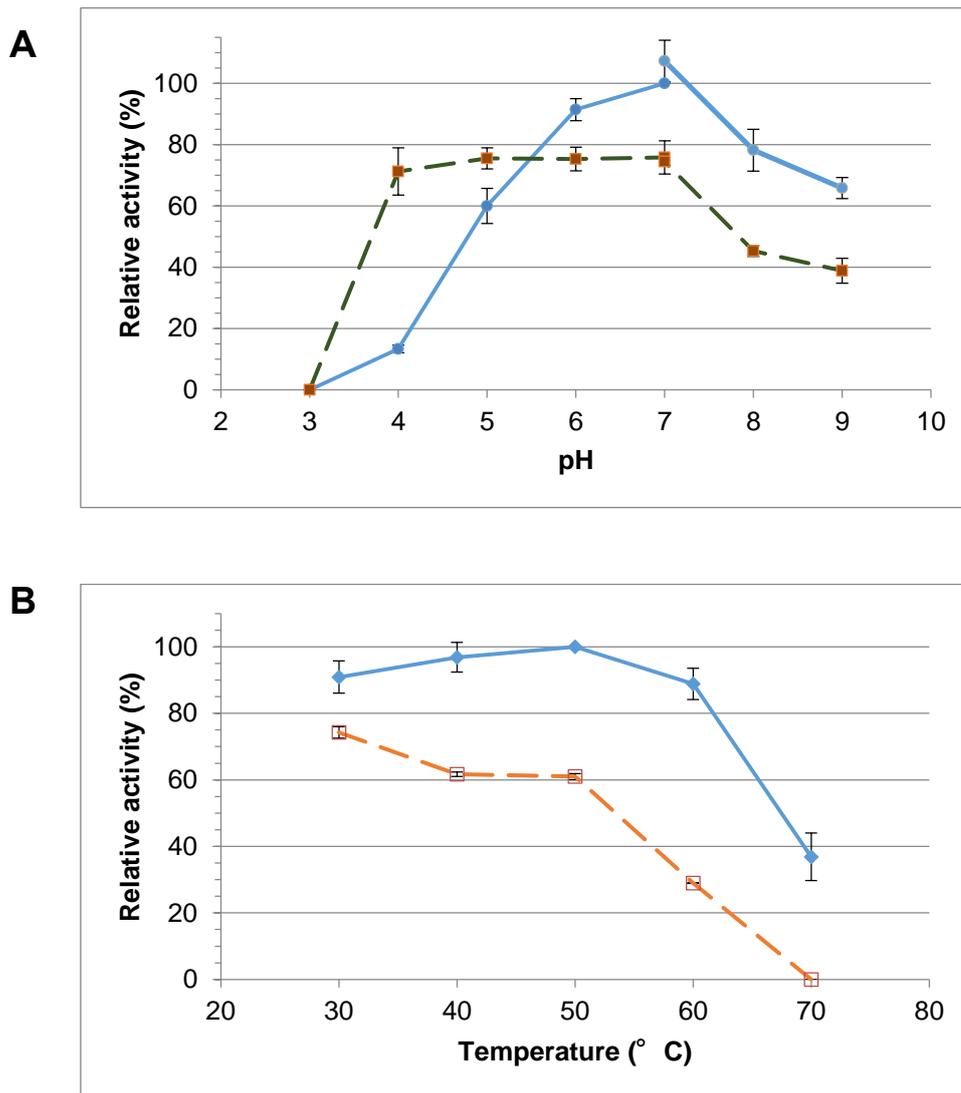
Figure 2-6: Effect of (A) pH and (B) temperature on the activity (solid line) and stability (dashed line) of purified *Nc*GE

The optimum pH was determined by performing the assay at 30°C for 10 min over the pH range 3.0-9.0. The pH stability was evaluated by pre-incubating the enzyme at different pH at 30°C for 30 min and then testing the residual activity at 30°C, pH 7.0. The optimum temperature was determined by performing the assay for 10 min at various temperatures from 30°C to 70°C. For the thermostability, the remaining activity was checked after incubation at different temperatures (30°C-70°C) for 30 min at pH 7.0.

Table 2-1: Effect of metal ions and other reagents on the activity of *Nc*GE

Cation or reagent	Relative activity (%)
Control	100 ± 1.65
$NiSO_4$	59.7 ± 3.52
$CuCl_2$	63.0 ± 1.58
LiCl	75.9 ± 5.65
$ZnSO_4$	84.1 ± 0.31
$MnCl_2$	93.6 ± 2.34
$FeSO_4$	103.6 ± 6.72
$FeCl_3$	114.9 ± 2.37
$AlCl_3$	118.0 ± 3.46
$MgCl_2$	118.4 ± 0.91
$CaCl_2$	127.0 ± 7.11
Urea	85.5 ± 1.59
Dimethyl sulfoxide	86.0 ± 6.70
Acetic anhydride	93.2 ± 5.19
Acid ascorbic	97.5 ± 4.05
EDTA	102.5 ± 1.33
Glycerol	103.8 ± 6.37
Tween 80	149.6 ± 1.38
Dithiothreitol	71.1 ± 2.63
2-Mercaptoethanol	83.9 ± 3.47
PMSF	40.8 ± 5.36
SDS	0.0 ± 00

Table 2-2: Kinetic properties of *Nc*GE, *Af*GE, and other GEs with the substrates A and B

GE	Substrate A			Substrate B			Reference
	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [mM^{-1}s^{-1}]	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [mM^{-1}s^{-1}]	
*Af*GE	15.8	67.1	4.24	16.4	23.8	1.45	This study
*Nc*GE	15	16.8	1.12	34.5	8.4	0.24	This study
Cip2	*nd*	*nd*	1.10	-	-	-	(Li et al. 2007)
*Sc*GE	1.78	7.8	4.38	-	-	-	(Špániková et al. 2007)
*Pa*GE1	-	-	-	12.1	7.8	0.64	(Sunner et al. 2015)
*St*GE2	-	-	-	8.9	3.5	0.39	(Sunner et al. 2015)

nd: not detected (K_m is too high to be determined by the Lineweaver-Burk plots (Li et al. 2007)).
 -: not tested
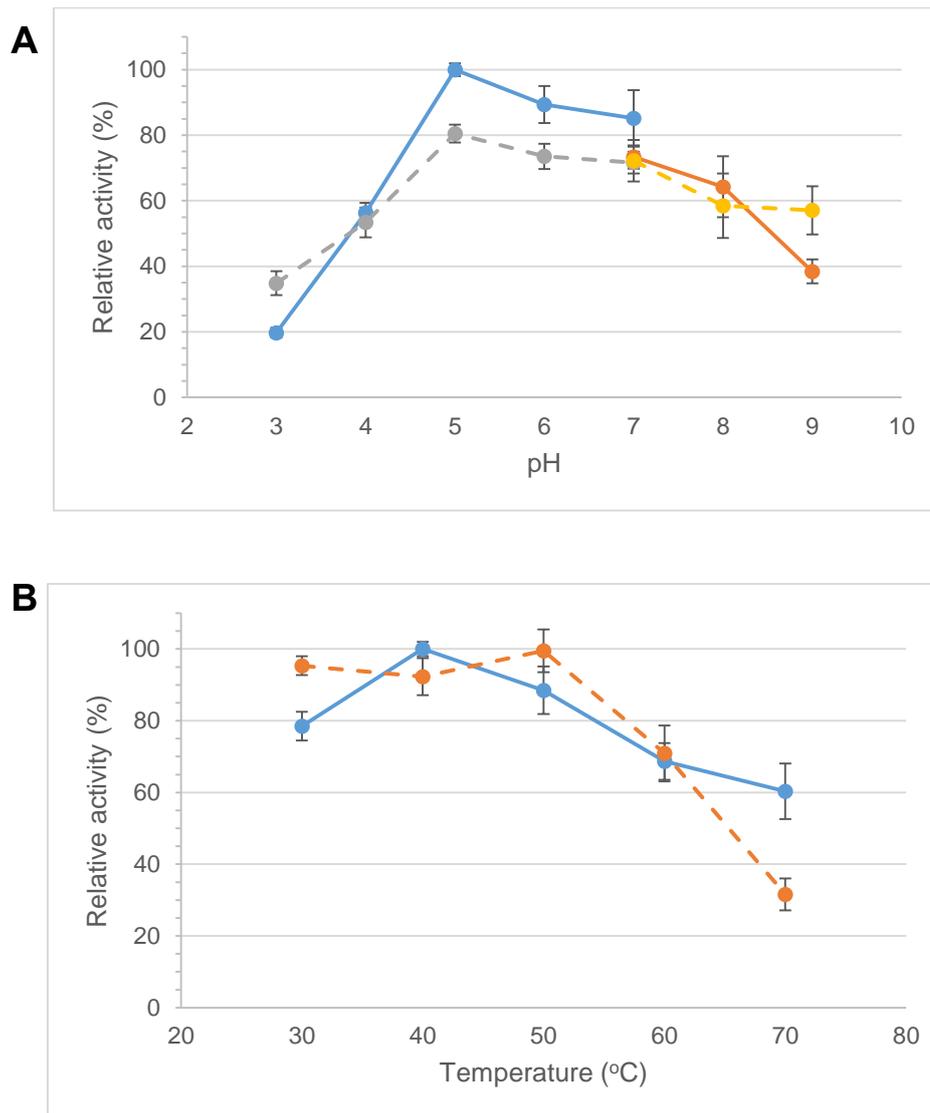
Figure 2-7: Effect of (A) pH and (B) temperature on the activity (solid line) and stability (dashed line) of purified *Af*GE

The optimum pH was determined by performing the assay at 30°C for 5 min over the pH range 3.0-9.0. The pH stability was evaluated by pre-incubating the enzyme at different pH at 30°C for 30 min and then testing the residual activity at 30°C, pH 5.0. The optimum temperature was determined by performing the assay for 5 min at various temperatures from 30°C to 70°C. For the thermostability, the remaining activity was checked after incubation at different temperatures (30°C-70°C) for 30 min at pH 7.0

Table 2-3: Effect of metal ions and other reagents on the activity of *Af*GE

Cation or reagent	Relative activity (%)
Control	100 ± 4.23
$FeCl_3$	51.16 ± 4.68
$CaCl_2$	61.5 ± 6.20
$NiSO_4$	64.4 ± 4.89
$CuCl_2$	72.2 ± 5.15
$AlCl_3$	73.6 ± 8.89
$MnCl_2$	75.8 ± 3.40
$FeSO_4$	89.4 ± 5.98
LiCl	89.4 ± 3.47
$ZnSO_4$	101.0 ± 2.56
$MgCl_2$	113.17 ± 7.25
Glycerol	95.0 ± 7.53
Tween 80	233.1 ± 9.5
Dimethyl sulfoxide	98.9 ± 9.30
Acetic anhydride	106.2 ± 7.52
Urea	106.3 ± 7.20
Ascorbic acid	107.1 ± 3.97
EDTA	99.4 ± 2.08
Dithiothreitol	$71.2 \pm 1.01^{*}$
2-Mercaptoethanol	$93.8 \pm 1.10^{*}$
PMSF	56.7 ± 3.46
SDS	0.0 ± 00

*: reaction was done at 50°C
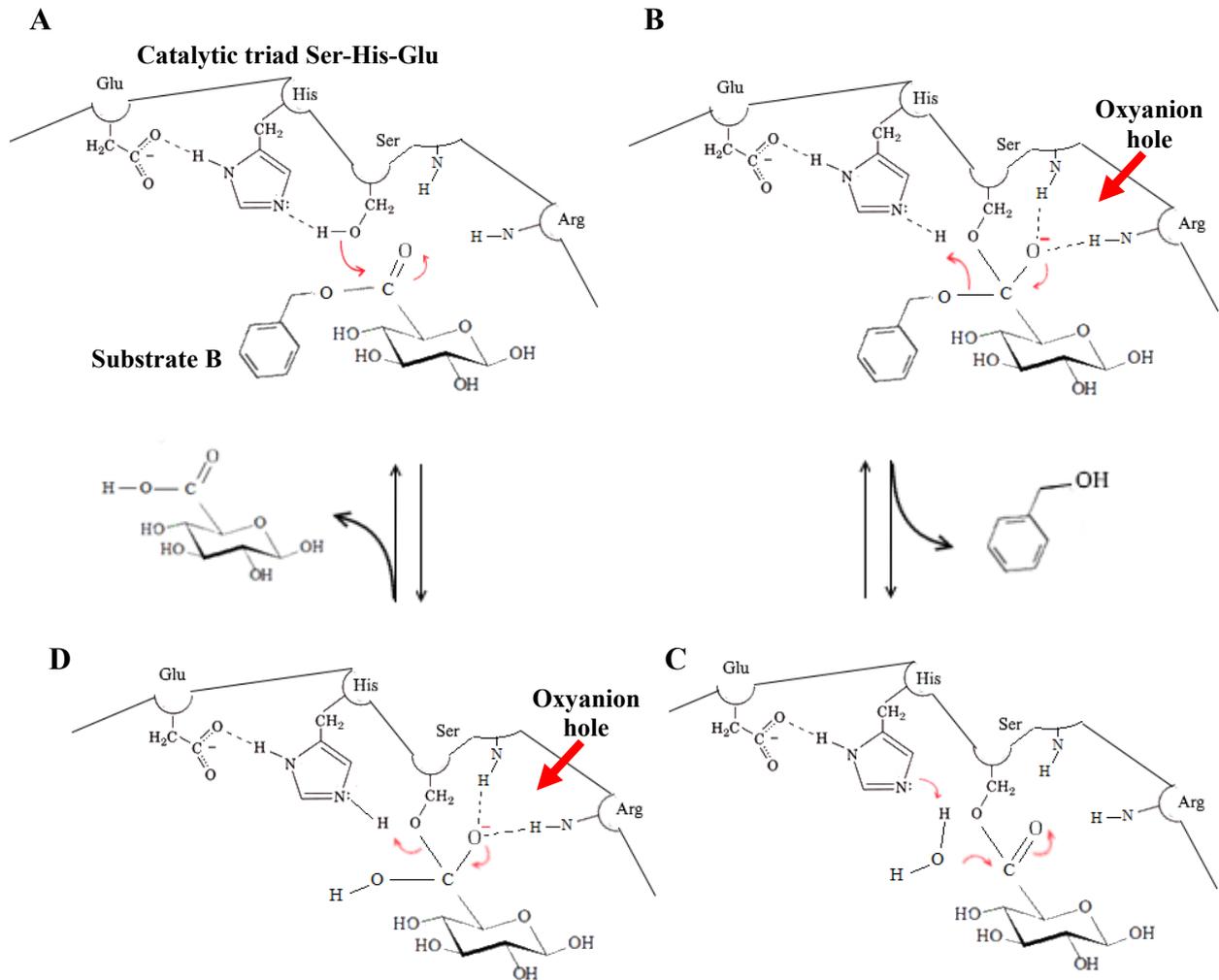
Figure 2-8: The predicted catalytic mechanism of GE based on the scheme of serine proteases

(A) Catalytic Ser performs a nucleophilic attack on the carbon of the carbonyl group in the substrate. (B) The acylation step is finished by the formation of tetrahedral intermediate structure, which collapses with the expulsion of the alcohol group of lignin (benzyl alcohol). (C) The deacylation step begins with the interaction of a water molecule and the catalytic center. (D) The second tetrahedral intermediate is formed, which collapses, releasing carboxylic acid product (glucuronic acid). Thus the catalytic triad is restored to its beginning state.

3.1. Introduction

Currently, the studies of GEs are hampered by the unavailability of specific substrates for GE assay, since only some of the esters of glucuronic acid have been commercialized. Until now, there are two suggested methods to qualify and quantify the GE activity (Biely 2016). In one assay, benzyl glucuronate, the commercially-available synthetic substrate, was applied to detect the GE activity by using HPLC (Sunner et al. 2015). This substrate was employed in Chapter 2 (substrate B in Fig. 2-1). Another one is the β-glucuronidase-coupled assay that uses easily synthesized pro-chromogenic substrates. β-glucuronidases do not hydrolyze the esterified glycosides. Thus, the de-esterification by GEs releases the substrates for β-glucuronidase which then produces the chromophore aglycones (Fig. 3-1). However, this method requires an extra step to synthesize the methyl ester of commercially-available 4-nitrophenyl or 5-bromo-4-chloro-3-indolyl β-glucuronides, which is achieved by cation-exchanger-catalyzed esterification with methanol (Fraňová et al. 2016). Interestingly, in both assays, the lack of the 4-*O*-methyl group in glucuronic acid moiety was shown to reduce the efficiency of hydrolysis by GEs (Biely 2016); in addition, it has been reported that GE activity of Cip2, *Cu*GE, *Pc*GE1, *Pc*GE2, and *Sc*GE is much higher when the substrate contains a methyl group at the *O*4 position (Fig. 0-7) (Ďuranová et al. 2009a; d'Errico et al. 2015). Although this feature is generally accepted, the reason remains unknown.

Glucuronoxylans are the primary components of hemicellulose in the plant cell wall with the backbone polymer of β-(1,4)-linked xylose. Most of the glucuronoxylans have single 4-*O*-methyl-D-glucuronic acid residue attached at the position 2 of xylose (Fig. 3-2; Sárossy et al. 2012; Rogowski et al. 2014). A large proportion of this 4-*O*-methyl-D-glucuronic acid is linked to lignin alcohols. Therefore the natural ester linkage between hemicellulose and lignin involves the ester of 4-*O*-methyl glucuronic acid. Hence, the preference of GEs toward the substrates containing the 4-*O*-methyl group is logical.

In this Chapter, I aimed to elucidate the preference of GE for the substrate containing 4-*O*-methyl group. The identification of amino acid residue(s) contributing to this characteristic would help not only understand the catalytic mechanism but also provide additional information in a search for novel GEs, since it is hypothesized that other putative GEs would show high identity in these residue(s). In addition, the data could be applied to develop a better GE assay using a suitable substrate, the commercialization of which may greatly facilitate the studies of GEs.

3.2. Results

3.2.1. Structural modeling and sequence analysis

It has been reported that GE activity of Cip2, *Cu*GE, *Pc*GE1, *Pc*GE2 and *Sc*GE is much higher when the substrate contains a methyl group at the *O*4 position (Fig. 0-7) (Ďuranová et al. 2009a; d'Errico et al. 2015). The initial GE assay revealed that two fungal GEs characterized in this study, *Nc*GE and *Af*GE, also exhibited this feature (Table 3-1). In this Chapter, *Af*GE was chosen as a template to analyze the mechanism of the preference for 4-*O*-methyl group of the substrate, since *Af*GE showed high activity and thermostability, displaying its potential for industrial application in the degradation of plant biomass.

To elucidate the structural basis whereby GEs prefer the 4-*O*-methylated substrates, molecular modeling of *Af*GE was performed using SwissDock (Grosdidier et al. 2011). The crystal structure of *St*GE2 S213A mutant docked with methyl 4-*O*-methyl-β-D-glucopyranuronate was used as the template (PDB code 4G4J; Charavgi et al. 2013). In this modeled structure, Arg206, Lys209, and Leu304 residues were located nearby the 4-*O*-methyl group with the distance about 4.7, 3.41, and 5.0 Å. Therefore, Lys209 was the nearest and predicted to have an interaction with the *O*4 of the methoxy moiety of the substrate (Fig. 3-3). Moreover, by using PDBePISA to calculate the interfacing residues, the Lys209 displayed the lowest solvation energy effect ($\Delta^i G$) of about -0.23 kcal/mol, suggesting the presence of the highest possible interaction between Lys209 and the *O*4-linked methyl group. It should be noted that this Lys is strongly conserved in all the fungal GEs and included in the consensus sequence VTGC**S**RXG**K**GA, identified in Chapter 1, containing the nucleophile Ser (Huynh and Arioka 2016). Thus it is hypothesized that this conserved Lys is involved in the recognition of 4-*O*-methyl group of the substrate. To examine this hypothesis, Lys209 was mutated to 4 different amino acid residues: alanine (shorter side chain - mutant A), glutamic acid (opposite charge - mutant E), glutamine (without charge - mutant Q), and arginine (similar charge - mutant R) (Fig. 3-4).

3.2.2. *Af*GE used as a model to study the substrate preference

In Chapter 1, *Af*GE was produced in the *A. oryzae* expression system. However, it usually takes at least one month to get the correct transformant strain. In contrast, the procedure is much quicker for *P. pastoris* (about a week). In addition, the production level of *Nc*GE expressed in *P. pastoris* was comparable to or better than that of *Af*GE expressed in *A. oryzae* since the band of *Nc*GE in CBB-stained gel was stronger than that of *Af*GE (Figs. 1-3 and 1-9;

the sample volume of culture supernatants was loaded). Taking these into consideration, the wild-type (WT) and mutated *Af*GEs were cloned and expressed in *P. pastoris* using pPICZαA vector. To remove an intron present at the N-terminal region of the *Af*GE sequence, two exon fragments were linked together by fusion PCR (Fig. 3-5). Then the sequence encoding the mature *Af*GE was introduced into pPICZαA. The WT *Af*GE was produced in the culture supernatant as bands of around 44 kDa (similar to *Af*GE expressed in *A. oryzae*) and the enzyme activity was confirmed by GE assay with the substrate A (data not shown). As expected, the production peaked at day 2 and the amount seemed higher than that expressed in *A. oryzae* (please compare Figs. 1-9 and 3-6).

3.2.3. Identification of the role of conserved Lys (Lys209) in *Af*GE

The WT and mutated *Af*GEs were produced in *P. pastoris* as shown in Fig. 3-7. The proteins were purified by Ni-NTA affinity chromatography, and the activities of imidazole-eluted fractions were tested using the synthetic substrates C and D which differ only in the modification at the *O*4 position (Fig. 3-8). As shown in Table 3-1, WT *Af*GE clearly showed a higher affinity (K_m) and turnover rate (k_{cat}) toward the substrate D than the substrate C, resulting in 2.5-times higher catalytic efficiency (k_{cat}/K_m). This result indicates that, like other fungal GEs, *Af*GE prefers the substrate that contains a methyl group at the *O*4 position; a similar result was also obtained for *Nc*GE. In contrast, the *Af*GE mutants A, E, and Q did not show significant difference in the kinetic parameters toward these substrates, demonstrating that Lys209 is essential for the recognition of *O*4-methyl group. Interestingly, the mutant R, which is considered a conservative substitution, displayed the preference for the substrate D over the substrate C. Collectively, the results obtained in this study demonstrate that Lys209 plays an important role in the recognition of *O*4-methyl group in the glucopyranose ring of the substrate.

3.3. Discussion

The previous studies have confirmed the potential of *P. pastoris* as a host for the expression of fungal GEs such as *St*GE (Topakas et al. 2010), *Pa*GE1 (Katsimpouras et al. 2014), *Nc*GE (Huynh and Arioka 2016), *Aa*GE1, *Pc*GE1, and *Wc*GE1 (Hüttner et al. 2017). Encouraged by these results, I chose *P. pastoris* for the expression of WT and mutated *Af*GEs to investigate their enzymatic properties. It should be noted that the specific activities of WT *Af*GE produced in *A. oryzae* and *P. pastoris* displayed no significant difference (data not shown), although it seems that the latter was less glycosylated (please compare Figs. 1-9, 2-3, 3-6, and 3-7).

Many GEs such as Cip2, *Cu*GE, *Pc*GE1, *Pc*GE2, and *Sc*GE have been shown to prefer the substrates containing 4-*O*-methyl group (Ďuranová et al. 2009a; d'Errico et al. 2015). It is of note that the natural ester linkage between the glucuronic acid group of hemicellulose and lignin known so far contains the 4-*O*-methyl substituent in the glucuronic acid residue (Ďuranová et al. 2009b; Biely et al. 2015; Bååth et al. 2016), which could explain why GEs display low catalytic efficiency toward the substrates without this methyl group (d'Errico et al. 2015; Biely 2016; Hüttner et al. 2017). In this study, I have demonstrated that conserved Lys209 in *Af*GE plays an important role in the recognition of the 4-*O*-methyl group of the substrate. In the molecular modeling analysis, Lys209 was found to be positioned in a close proximity to the 4-*O*-methyl group. Since this Lys is strongly conserved not only in the characterized GEs but also in most GE-like proteins including those of prokaryotic origins (De Santi et al. 2016; Agger et al. 2017), its role was analyzed by site-directed mutagenesis. By changing Lys209 to other amino acids with a shorter side chain (Ala), an opposite charge (Glu), or without a charge (Gln), the preference for the substrate containing the 4-*O*-methyl group was largely lost; these three mutants displayed similar k_{cat}/K_m values to both the substrates C and D. These results are consistent with what was observed in the crystal structure of *St*GE2 bound to the substrate analogue (Charavgi et al. 2013); in this work, the ε-amino group of Lys217 (corresponding to Lys209 in *Af*GE) was shown to be involved in the van der Waals interaction with the 4-*O*-methyl group of the substrate analogue. In fact, this 4-*O*-methyl group forms a total of thirteen van der Waals interactions with the enzyme (7 for *C*4a and 6 for *O*4 atom) and is proposed to enhance binding of the substrate to the enzyme, providing a theoretical support for the preference of GEs toward the substrates containing this 4-*O*-methyl group.

It should be noted, however, that the k_{cat}/K_m values of mutants were far lower than that of WT, mainly due to the decrease in k_{cat} values, suggesting that the Lys209 residue plays an

additional role in the catalysis besides the recognition of 4-*O*-methyl group. Another point to be discussed is the result of Arg mutant. In this mutant, the preference for the substrate D was essentially preserved, albeit at decreased k_{cat} values. This suggests that the charge-mediated interaction might play a role in the recognition of 4-*O*-methyl group. However, additional studies are needed to clarify this feature. An alternative explanation is that, since Arg contains the bulkier side chain compared to Ala, Glu, and Gln, it might have provided van der Waals interactions that mimic Lys.

In addition to the conserved Lys residue, the structural analysis of *St*GE2 has suggested that Glu236 in the catalytic triad is also involved in the van der Waals interaction between the 4-*O*-methyl group of the substrate and the catalytic pocket of the enzyme (Charavgi et al. 2013). In agreement, the modeling study proposed that the SGXGG sequence following Glu228 (corresponding to Glu236 in *St*GE2) forms a concave surface which could create a consistent cavity to keep the 4-*O*-methyl group of the substrate (Fig. 3-9). This was further supported by the interface analysis by PDBePISA suggesting an interaction between this sequence and 4-*O*-methyl group. It should be noted that the SGXGG sequence is also conserved in all the characterized fungal GEs. Thus the combined effects of Lys209 and SGXGG sequence could help maintain the ester of 4-*O*-methyl-D-glucuronic acid in the correct catalytic position. The inability of GEs to cleave the methyl esters of D-galacturonic acid (Ďuranová et al. 2009a; Wong et al. 2012) that contains the 4-hydroxy moiety in the axial position further corroborates the importance of configuration at *C*4.

Until now, most of the studies of GEs have been conducted using the synthetic substrates, among which only 3 are commercially available: benzyl-D-glucuronic acid, allyl-D-glucuronic acid, and methyl-D-glucuronic acid (Hüttner et al. 2017). However, these three substrates do not contain the 4-*O*-methyl group. The preference for this methyl group is a typical characteristic of GEs, which was shown in this study. Therefore, the synthesis of esters of 4-*O*-methyl-D-glucuronic acid and their commercialization are an urgent requirement for the enzymological and biochemical studies of GEs.
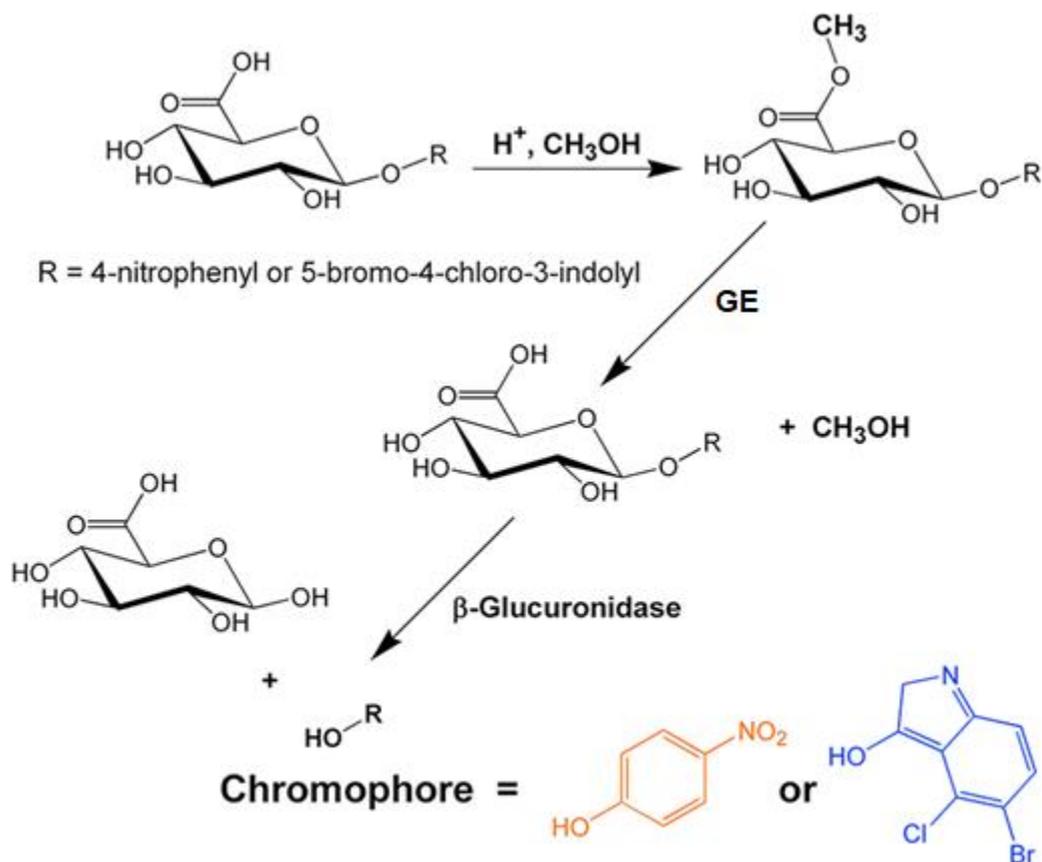
Fig. 3-1: The schematic diagram of β-glucuronidase-coupled assay of GEs

The cooperation of GE and β-glucuronidase releases the chromogenic compounds such as 4-nitrophenol or 5-bromo-4-chloro-3-hydroxyindol which can be detected by spectrophotometry or visualization of the blue product, respectively (Fraňová et al. 2016).
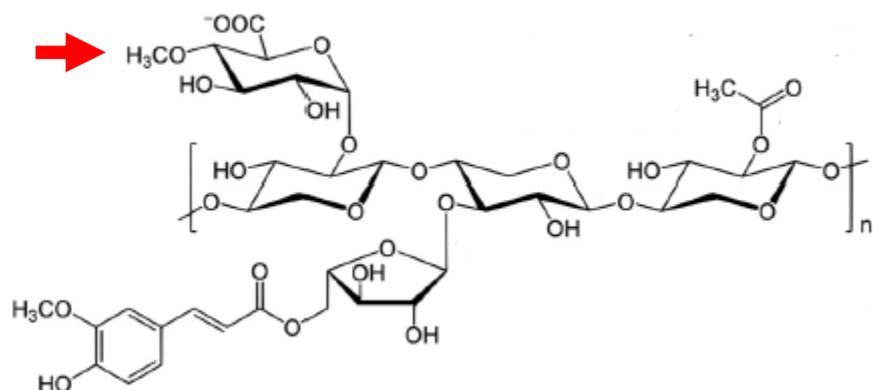
Figure 3-2: Example of one possible xylan structure

The basic structure of glucuronoxylans in the hemicellulose. 4-*O*-methyl-D-glucuronic acid is indicated by the arrow. The acetate group and ferulate group are shown in the upper right and lower left corners, respectively (Rogowski et al. 2014).
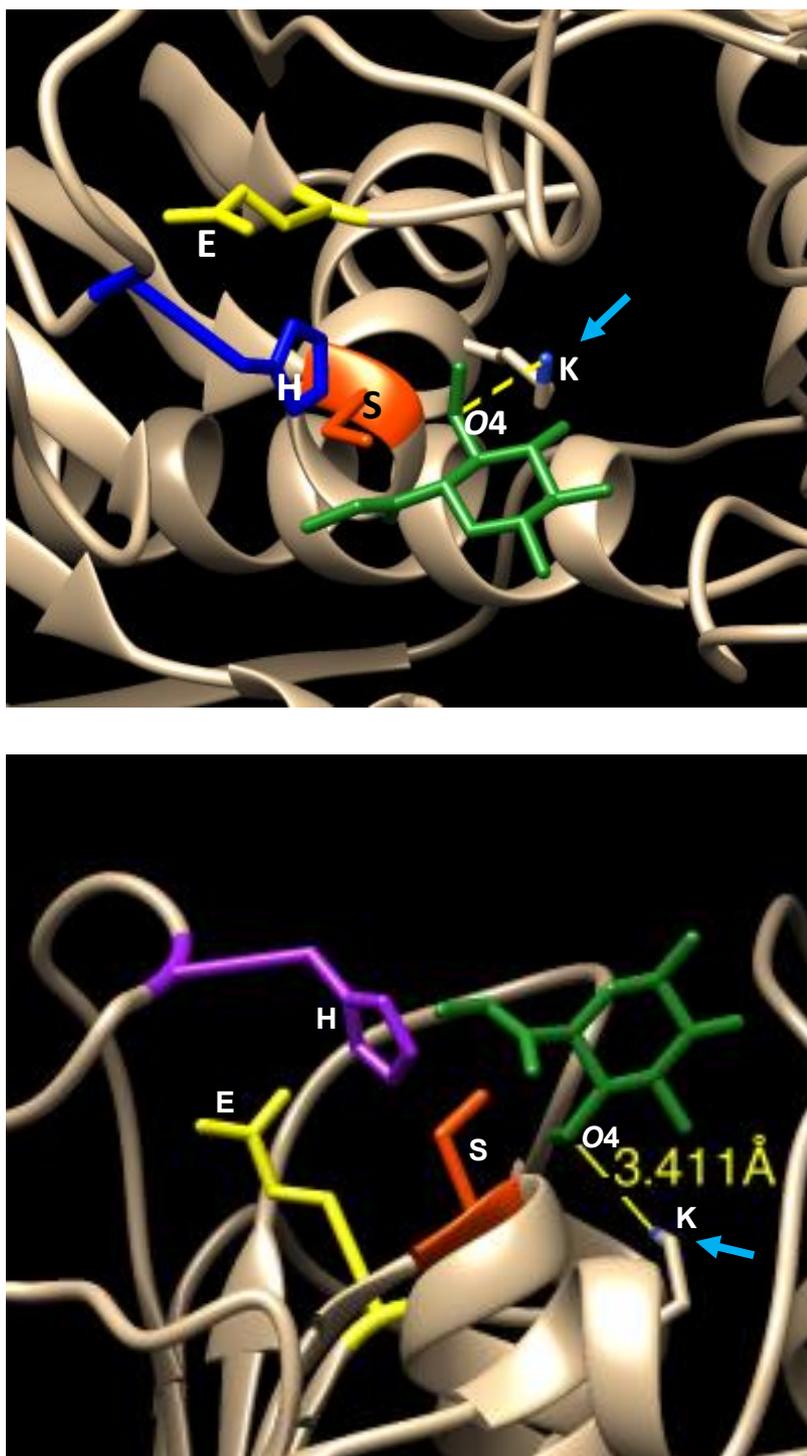
Figure 3-3: Structural modeling of the catalytic domain of *Af*GE

The interaction between Lys209 and 4-*O*-methyl group is indicated by the dashed line.
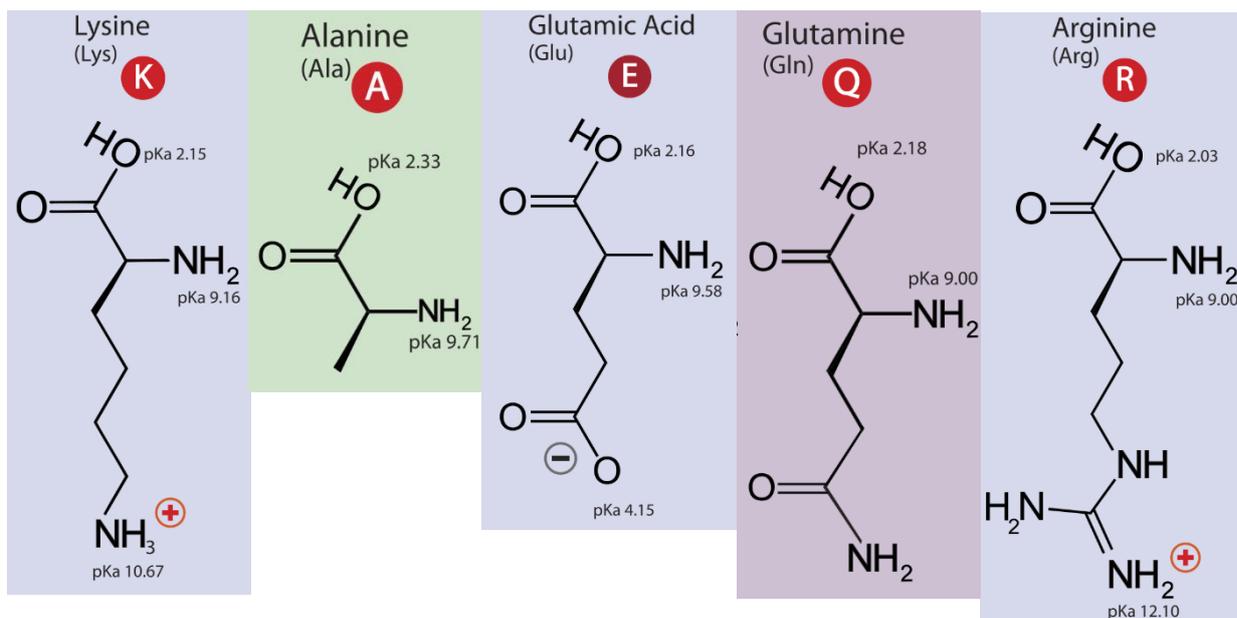
Figure 3-4: Lys and other amino acids selected for creating mutated *Af*GEs

Lys209 residue containing the positive charge was changed to other amino acids with a shorter side chain (Ala), an opposite charge (Glu), without a charge (Gln), or positively-charged Arg. (Source: https://en.wikipedia.org/wiki/Amino_acid).
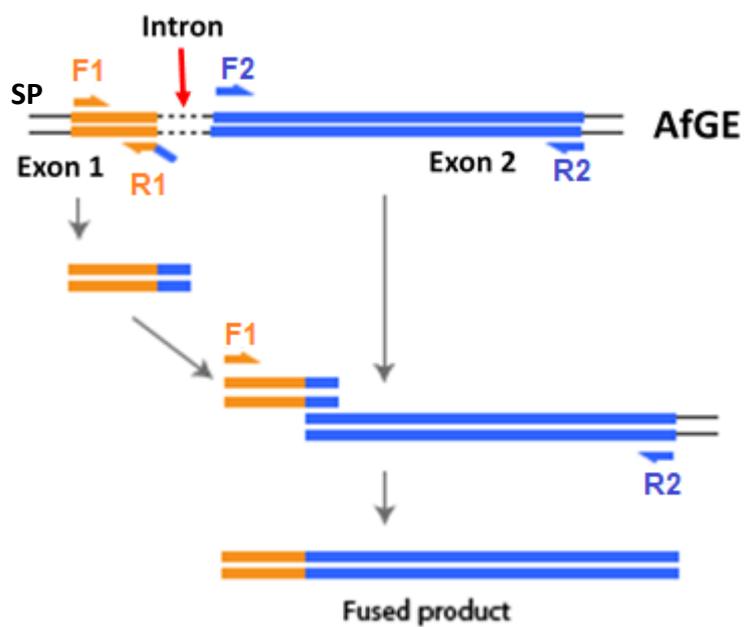
Figure 3-5: Scheme for removing the intron in the *Af*GE gene

SP, signal peptide. F1, R1, F2, and R2 are the primers for amplifying the exons which shown in the materials and methods.
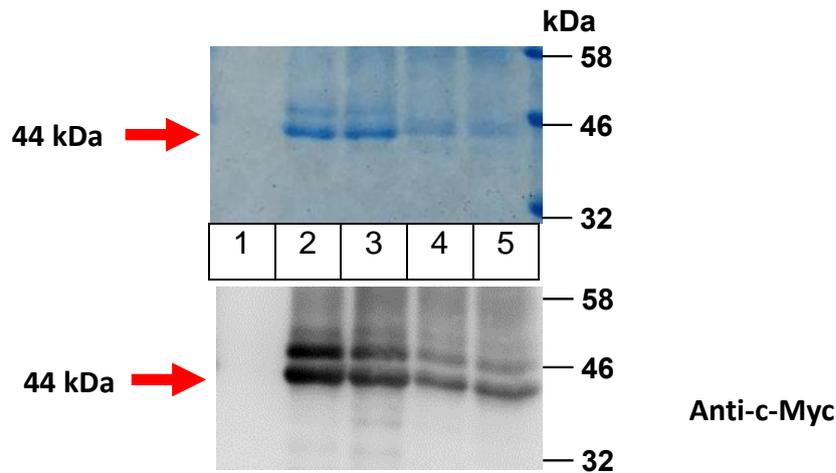
Figure 3-6: Expression of *Af*GE in the culture supernatant using *P. pastoris* as a host

(A) CBB staining and (B) Western blot analyses of the culture supernatants of WT *Af*GE produced in *P. pastoris*. *Af*GE tagged with c-Myc and hexahistidine tags at the C-terminus was expressed in *P. pastoris* by using pPICZαA vector. Lanes 1-5 indicate the samples from day 1-5, respectively. For SDS-PAGE, 20 µl of culture supernatant from each day were loaded.
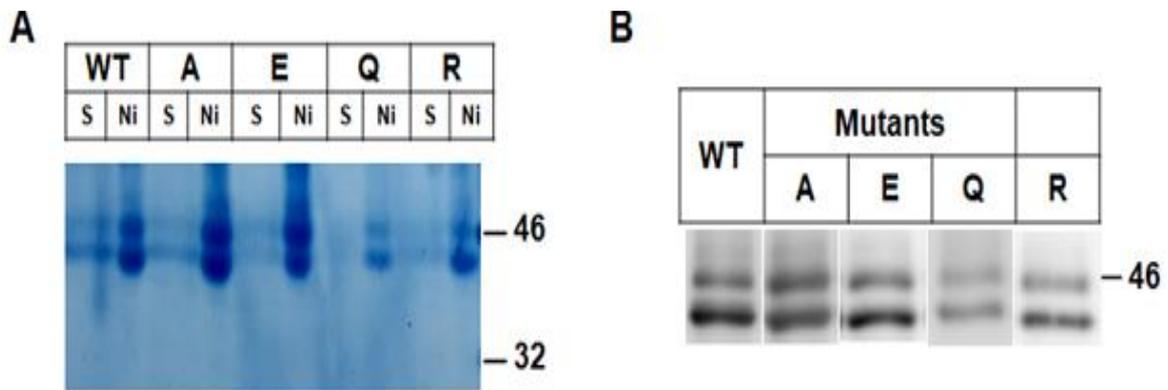
Figure 3-7: Purification of WT and mutant *Af*GEs

(A) CBB staining and (B) Western blot analysis of the enzymes after purification. WT and mutant *Af*GEs tagged with c-Myc and hexahistidine tags at the C-terminus were expressed in *P. pastoris* by using pPICZα vector. S, culture supernatant; Ni, enzyme sample after Ni^{2+}-NTA column chromatography. The proteins were obtained after methanol induction for two days and 20 µl of culture supernatant from each sample was loaded into the SDS-PAGE gel.
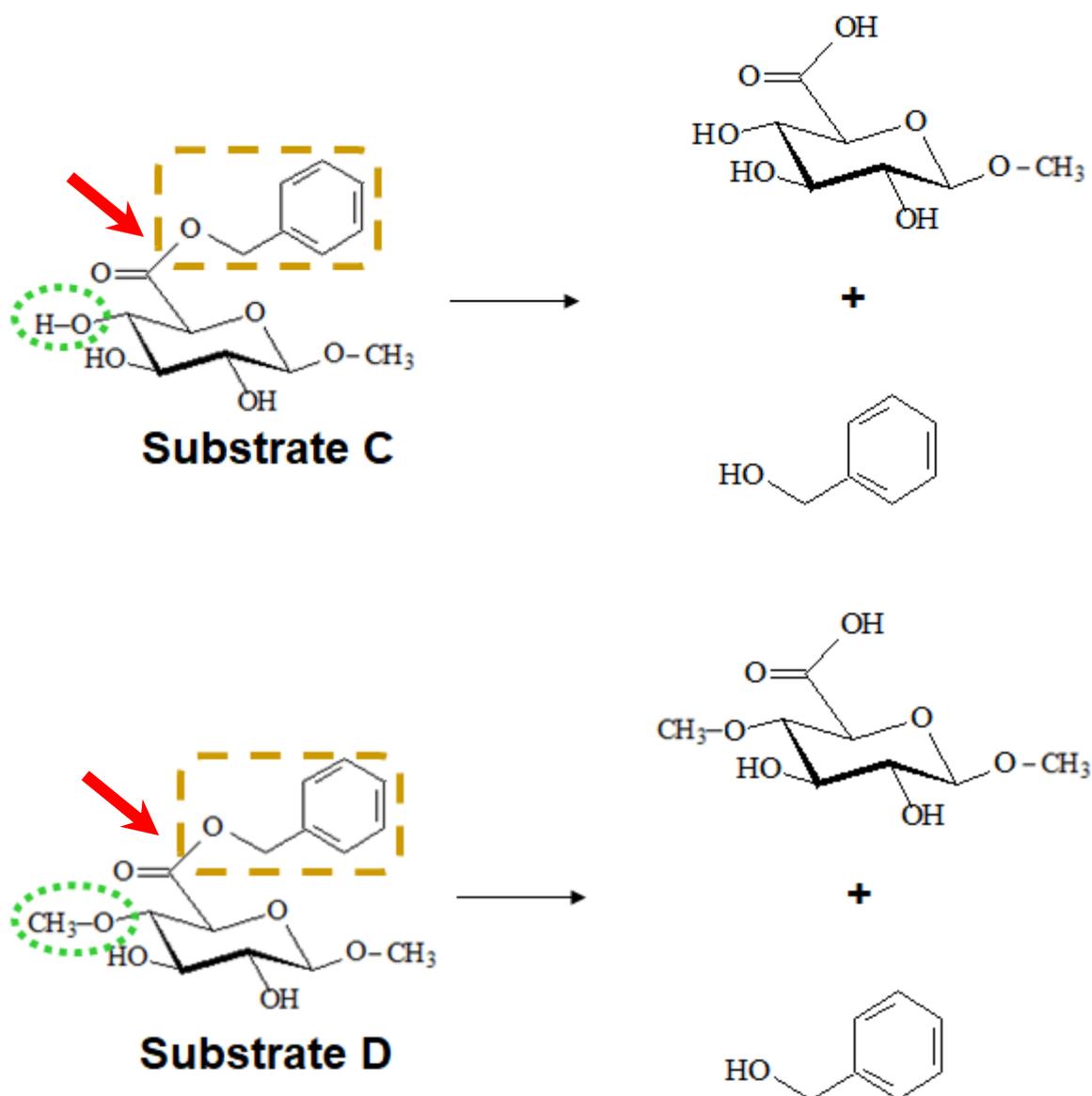
**Figure 3-8: Substrates used for the analysis of the role of Lys209**

Substrate C, benzyl methyl α-D-glucopyranosiduronate and substrate D, benzyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate. The position of *O*4 in the glucopyranose ring of the substrate is circled by the dotted line. The arrows indicate where GEs cleave the substrates. The aromatic groups used for HPLC detection are boxed.

Table 3-1: Kinetic properties of WT and mutated *Af*GEs

GE	Substrate C			Substrate D			Preference for 4-*O*-methyl group	Reference
	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [mM^{-1}s^{-1}]	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [mM^{-1}s^{-1}]		
*Af*GE	38	42.7	1.12	23.7	66.4	2.8	**Yes**	This study
Mut A	40.4	0.75	0.018	40.9	0.84	0.02	*No*	This study
Mut E	71.4	0.86	0.012	65	0.78	0.012	*No*	This study
Mut Q	68.7	1.34	0.019	73.3	1.51	0.02	*No*	This study
Mut R	38.7	0.39	0.01	36.9	1.94	0.05	**Yes**	This study
*Nc*GE	37.3	7.0	0.18	34.3	31	0.9	**Yes**	This study
*Cu*GE	80.0	48.0	0.60	4.6	129	28	**Yes**	(Errico et al. 2015)
*Sc*GE	51.0	64.0	1.20	3.7	118	32	**Yes**	(Errico et al. 2015)
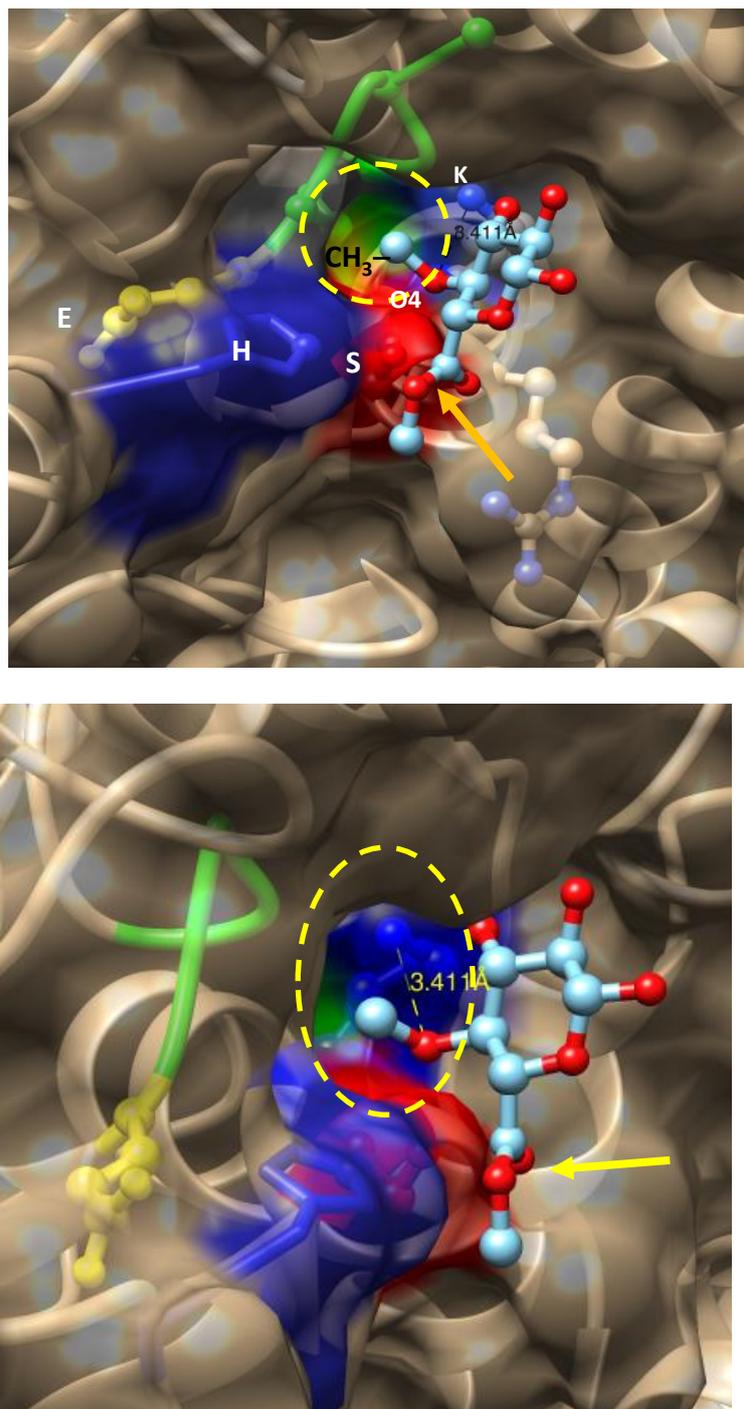
Figure 3-9: The interaction between the catalytic domain of *Af*GE and the 4-*O*-methyl group

The cavity formed by the sequence SGXGG (circled by the dashed line) might be involved in the van der Waals interaction between GE and 4-*O*-methyl group. The catalytic Ser residue stands in front of the ester linkage (arrow) to be cleaved.

4.1. Introduction

From the first discovery in *S. commune*, GEs have been expected to play an important role in decreasing the recalcitrance of lignocellulose (Špániková and Biely 2006). Consequently, this effect can facilitate the biomass saccharification, and thus greatly improve the efficiency of plant biomass utilization. However, these suggestions were based on the GE assay with the synthetic substrates. Recently, GE activity against the natural substrate was examined for the first time using LCCs from free ball-milled woods of birch and spruce. The evidence for the hydrolysis of the natural ester linkage was provided by three observations (Bååth et al. 2016). First, the results of size exclusion chromatography indicated the shift toward lower values in the average molar mass of LCCs after pretreatment with GE. This suggested the expulsion of lignin components from LCCs by GE. Second, the concentration of carboxylic acid content increased in the ^{31}P NMR spectroscopy due to the cleavage of the target ester linkage which generated the glucuronic acid group. Finally, further NMR analysis confirmed that the de-acetylation did not occur in the GE reaction, thus the formation of carboxyl group mostly came from the GE activity.

Until now, there are few reports about the biological actions of GEs. In 2016, d' Errico et al. showed an enhanced degradation of lignocellulosic biomass by adding the purified Cip2 or *Cu*GE to the commercial cellulase mixtures such as Ultraflo® and Cellic® CTec from Novozymes (d'Errico et al. 2016). The other indirect proofs were presented by the heterologous expression of *Pc*GE1 in *Arabidopsis* or *Populus* (Tsai et al. 2012; Gandla et al. 2015). The alteration of cell walls of these plants was attributed to the deduction of the ester linkage between hemicellulose and lignin. Consequently, the amount of xylan extracted and its saccharification were improved in these transgenic plants.

4.1.1. Hardwood and softwood

Generally, wood is separated into two groups: hardwood and softwood. These terms are based on the taxonomical difference of the plants and do not reflect the actual hardness measured in the wood samples. Hardwood is the wood from angiosperms which produces the mature seeds within an enclosure; in other words, a fruiting plant. The leaves of the hardwood trees are often broad and are replaced once per year. Birches, maples, oaks, and fruit trees are the typical examples of hardwood group. Softwood, on the other hand, is often referred as gymnosperms based on the unenclosed condition of their seeds. Softwoods trees have the needle-sharp leaves which means that they do not lose their leaves at the end of the growing season as the evergreen plant such as firs, pines, and spruces (Table 4-1) (Monlau et al. 2013). The important difference

between hardwoods and softwoods is their cellular structure (Fig. 4-1). The vessel elements to transport the water thoroughly the body only appear in hardwoods. By using the microscope, it has been shown that the vessel elements present as the pores with various sizes and numbers depending on the difference in the part or species of plants. In contrast, water is delivered by the longitudinal tracheids system in softwoods, and thus the pore structures do not exist (Bond and Hamner 2002).

4.1.2. Molecular structures of hardwood and softwood

Wood is mainly constructed by three components: cellulose, hemicellulose, and lignin, which is known as the lignocellulose matrix. In hardwood, the amount of cellulose, hemicellulose, and lignin is normally 40-50%, 15-35%, and 17-25%, respectively, whereas softwood contains 40-50% of cellulose, 20-30% of hemicellulose, and 25-35% of lignin (Bodîrlău et al. 2007; Wahab et al. 2013). The cellulose structure is quite similar in both types of woods. However, there are some differences in the hemicellulose and lignin formation. Hardwood xylans contain less 4-*O*-methyl glucuronic acid groups with the ratio about one per ten xylose and have a small amount of arabinose (Fig. 4-2). In softwood xylans, the arabinose side groups often attach along with the 4-*O*-methyl glucuronic acid groups which appear after every six xylose unit (Fig. 4-3) (Holtzapple 2003).

4.1.3. Objective

The recent studies have elucidated the activity of GEs against the natural substrate, and thus GEs can help improve the sugar recovery when combined with the commercial cellulase mixture (d'Errico et al. 2016). It is noteworthy that some commercial cellulase preparations such as Cartazyme HS 10 (Sandoz, Holzkirchen, Germany), Celluclast® 1.5L (Novo Nordisk, Bagsvaerd, Denmark), Depol™ (AB Enzymes, Darmstadt, Germany), and Ultraflo™ L (Novozymes, Bagsvaerd, Denmark) display the trace amount of GE activity (Fraňová et al. 2016). Hence, it would be of interest to further clarify the role of GEs in the degradation of lignocellulosic biomass, especially hardwood and softwood (Biely 2016). However, as far as I know, there has been no study that presented the evidence that GEs contributed to the degradation of lignocellulose in the living (micro) organism. For these reasons, in this Chapter, I intended to demonstrate the physiological function(s) of *Nc*GE in its native host, the filamentous fungus *N. crassa*.

4.2. Results

4.2.1. *N. crassa* strains

To elucidate the role of *Nc*GE in *N. crassa* grown in the lignocellulosic biomass, two *N. crassa* strains were obtained from the Fungal Genetics Stock Center (FGSC). The wild-type (WT) *N. crassa* (FGSC 9717) has the *mus-51*::*bar his-3 mat A* genotype. The other one is the *Nc*GE-knockout (KO) strain (FGSC 21356) in which the gene encoding *Nc*GE (NCU09445.2) is replaced by the hygromycin-resistant gene as the selectable marker (Colot et al. 2006).

4.2.2. Phenotypic analysis

At first, some basic phenotypes of two *N. crassa* strains were examined such as the growth rate of mycelia, the production of conidia, and the morphology of hyphae. The growth rate of two trains in the Vogel's minimal medium (VM) containing sucrose as the carbon source was similar during 4 days of incubation (Fig. 4-4). The generation of pigment and conidiation were clearly seen at day 3. The numbers of conidia obtained from both strains were similar (Fig. 4-5). Student's *t*-test showed no statistical difference between the two samples. In addition, the appearance of basal hyphae under the microscope was quite similar (Fig. 4-6). These results indicate that there is no significant difference between WT and KO strains when they are grown in the sucrose-VM medium, and the deletion of *Nc*GE gene generally does not affect the development of the KO strain. Both strains were used in the next step to further study the role of *Nc*GE.

4.2.3. Preparation of LCCs extract

GEs are expected to facilitate the biomass degradation by breaking down the stubborn ester linkage between hemicellulose and lignin. To analyze this hypothesis, the raw materials containing this linkage were prepared from wood powders. In this study, the LCCs were extracted from the wood powders of Cedar (softwood) and Pasania (hardwood). Briefly, the thermal treatment by autoclaving was used to extract the LCCs from the wood powder (particles size ≤ 100 µm) as suggested (Ko et al. 2015; Bååth et al. 2016). During the treatment, the wood powders are shattered into smaller particles by high temperature and pressure, and thus some small particles are suspended into the upper layer containing water (Fig. 4-7). This suspension fraction was collected and used as the sole carbon source for culturing *N. crassa*. The particles in the LCCs extract could be observed under the microscope (Fig. 4-8).

4.2.4. The role of *Nc*GE in *N. crassa* grown in LCCs extract from wood

To elucidate the biological function of *Nc*GE, WT and KO strains were cultured in the VM media containing LLCs extract as the sole carbon source (Fig. 4-9) under five different conditions as shown below:

- WT: WT strain was cultured in the LCCs-VM medium (10 ml) for 7 days.
- KO: KO strain was cultured in the LCCs-VM medium (10 ml) for 7 days.
- KO GE: KO strain was cultured in the LCCs-VM medium (10 ml) with the daily addition of purified *Nc*GE (0.8 μg) for 7 days.
- KO GE-inactive: KO strain was cultured in the LCCs-VM medium (10 ml) with the daily addition of heat-denatured (boiled for 10 min) *Nc*GE (0.8 μg) for 7 days.
- KO GE-pretreat: The LCCs-VM medium (10 ml) had been pre-treated by the daily addition of the purified *Nc*GE (0.8 μg) for 5 days at first. Then this pre-treated medium was used to grow KO strain for 7 days. Purified *Nc*GE was not added during the culture.

4.2.4.1. Culturing *N. crassa* in LCCs extract from Cedar wood

Cells were grown in the medium containing LCCs from Cedar wood, and the concentration of reducing sugars in the culture supernatant from each day was measured. The amount of reducing sugars liberated in the culture supernatant due to the degradation of plant biomass by endogenous enzymes was used as a hallmark of cell growth (Rao et al. 1985; Hildebrand et al. 2015). When the cells are grown in the wood extract medium, they release endogenous cellulases/hemicellulases to hydrolyze the lignocellulose and produce reducing sugars. The increase in sugar concentration would stimulate the cells to grow, which in return contributes to the lignocellulose degradation by secretion of more enzyme. If the amount of possibly hydrolyzable substrate is limited, the sugar concentration would stop rising and gradually decrease through the consumption by cells in the later stage of the culture. Hence, in my experiments, a small amount of conidia (10^5 conidia/10 ml culture medium) was added to the medium to prevent the quickly reaching carbon limitation. The result is shown in Fig. 4-10. A noticeable difference was displayed after 3 days of culture. The highest amount of reducing sugars was released in the case of KO GE (KO strain added with pure *Nc*GE). KO GE-pretreat (KO strain grown in GE-pre-treated medium) and WT also released a higher amount of reducing sugars than the KO strain without the addition of *Nc*GE or with the addition of denatured *Nc*GE. These results suggest that *Nc*GE contributed to the degradation of Cedar wood and enhanced the amount of reducing sugars released in the culture medium.

4.2.4.2. Culturing *N. crassa* in LCCs extract from Pasania wood

When the Pasania wood extract was used as the sole carbon source in the medium, the similar result was obtained as shown in Fig. 4-11. The conditions in which *Nc*GE was present (WT, KO GE, and KO GE-pretreat) exhibited the higher amount of reducing sugars in the medium than the conditions in which *Nc*GE was absent or inactivated (KO and KO GE-inactive). Again, *Nc*GE was suggested to play an important role in the deconstruction of lignocellulose.

4.2.4.3. Culturing *N. crassa* in the medium containing CMC or Avicel as the sole carbon source

It is suggested that GEs work only on the substrate carrying the particular ester linkage and therefore cannot affect the state of the substrates such as carboxymethyl cellulose (CMC) and Avicel, both of which are the polymers of glucose, that lack the target ester linkage. I then checked the effect of *Nc*GE on these two simple substrates.

As expected, there was no difference in the production of reducing sugars among the three culture conditions (WT, KO, and KO GE) when the cells were grown in the CMC or Avicel medium. The highest amount of reducing sugars was obtained after 2 days of incubation in CMC-VM medium (Fig. 4-12). In contrast, it took 6 days to reach the highest concentration of reducing sugars in the Avicel-VM medium (Fig. 4-13). Moreover, the amount of reducing sugars released was much higher in CMC-VM than in Avicel-VM (16 mM vs 0.8 mM), indicating that CMC was much more easily degraded by the fungus than Avicel.

4.2.4.4. Reverse transcription – PCR (RT-PCR) analysis

The expression of *Nc*GE and some other key enzymes for cellulose hydrolysis was examined. Normally, for efficient cellulose hydrolysis, the cooperation of three main enzymes, i.e. endoglucanase, cellobiohydrolase, and β-glucosidase, is required (Chandel et al. 2012). In this experiment, the expression of three major cellulases which are highly expressed in the biomass degradation of *N. crassa*, CBH-1 (cellobiohydrolase), GH 5-1 (endoglucanase), and GH 3-4 (β-glucosidase) was examined in the WT and KO strains (Figs. 4-14, 15, 16, and 17). Three selected cellulases are the common enzymes used in the transcriptomic analysis due to their high expression level when the cells were grown in CMC, Avicel, or straw (Phillips et al. 2011; Znameroski et al. 2012; Wang et al. 2015).

The same amounts of total RNA were used for the RT-PCR reaction, which was confirmed by the similar level of the product of the control gene - actin. In order to utilize wood

powder, CMC, and Avicel, the enzymes involved the hydrolysis of cellulose are needed to be secreted into the culture medium. In agreement, the gene expression of GE and other cellulases was higher in the VM medium containing wood powder, CMC, and Avicel than in the sucrose-VM medium. This result agrees well with the observation in the previous studies which indicated the increase in *Nc*GE and cellulase gene expression by CMC, Avicel, and straw (Benz et al. 2014; Wang et al. 2015). Naturally, the expression of *Nc*GE was only detected in WT. The expression of cellulase genes was quite similar between the WT and KO strains when they were cultured in the medium containing the same sole carbon source. These results suggest that the release of a larger amount of reducing sugars by the WT strain than the KO strain was mostly due to the effect of *Nc*GE itself, not by the enhanced expression of cellulases.

4.2.4.5. GE activity in the culture supernatant of *N. crassa*

The GE activities in the culture supernatant were also examined using the substrate D. The supernatants of WT grown in CMC or Avicel showed some trace activity when reacted with the substrate overnight while those of the KO strain did not (data not shown). However, GE activity of the WT strain was not detected when grown in both wood extract media. A recent study has suggested that the VM medium is the nitrogen-deficient medium which may reduce the protein production and promote the secreted protease activity (Havlik et al. 2017). The authors recommended that the supplementation of complex nitrogen sources such as yeast extract and tryptone from casein could overcome this issue. Therefore, WT and KO strains were grown for 5 days in the Cedar and Pasania wood extract-VM media with the addition of yeast extract and tryptone (complex medium). Interestingly, some trace GE activities (corresponding to around 4.4 and 3.8 ng of purified *Nc*GE per ml culture medium, respectively) were detected from the culture supernatant of the WT strains after incubating overnight at 30ºC. Together with the identification of gene expression by RT-PCR, I could conclude that the endogenous *Nc*GE was expressed and exhibited the real GE activity when the WT strain was grown in the medium containing lignocellulose as the sole carbon source.

4.3. Discussion

The fungus *N. crassa* has a long history of research as a model organism. Therefore, many culture processes and molecular tools have been established for *N. crassa* such as the data for the transcriptome and proteome analyses in different growth conditions and the library of gene knockout strains. *N. crassa* is also one of the filamentous fungi displaying the fast and robust growth rate. Moreover, this fungus is considered as a non-toxic organism and can be used in many applications (Havlik et al. 2017). Therefore, *N. crassa* was chosen to identify the physiological role of GE in this study.

I could not obtain the data for the exact amount of *Nc*GE released in the wood medium. However, based on the proteomic analysis of *N. crassa* grown on the pectin medium, I estimate that 0.03-0.07 μg/ml of *Nc*GE was released after 5 days of culture (Benz et al. 2014). Hence, the addition of 0.08 μg/day/ml culture medium of *Nc*GE was much higher than the level of endogenously-produced *Nc*GE and might have saturated the requirement of *Nc*GE to degrade the ester linkage in the wood powder-VM medium at the highest efficiency. In agreement, the KO GE condition produced the highest amount of reducing sugars, which was followed by the KO-pretreat condition.

When the WT *N. crassa* was grown in the Cedar and Pasania wood as the sole carbon source, higher amount of reducing sugars was released than the KO strain (Figs. 4-10 and 11). The presence of endogenous *Nc*GE only in the WT samples was confirmed by RT-PCR analysis. The addition of the purified *Nc*GE, but not heat-inactivated enzyme, to the culture of KO strain stimulates the production of reducing sugars. Thus, both endogenous and exogenously-added *Nc*GE assisted the performances of endogenous cellulase/hemicellulase. Moreover, when WT and KO strains were cultured in the same medium, the expression levels of three major cellulase genes were similar (Figs. 4-14, 15, 16, and 17), suggesting the *Nc*GE contributed mostly to this enhancement. On the other hand, the presence/absence of *Nc*GE did not affect the reducing sugar concentration when CMC or Avicel (not containing the target ester linkage) was used (Figs 4-13 and 13). It should be noted that during the pretreatment (KO GE-pretreat), the *Nc*GE itself did not increase the amount of reducing sugars. Hence, the effect of GE was considered as the ability to separate hemicellulose and lignin by removing their ester linkage and thus support the actions of other carbohydrate-acting enzymes.

In this Chapter, *in vivo* role of GE in the degradation of wood extracts was demonstrated for the first time. *Nc*GE is suggested to contribute to the enhanced release of reducing sugars when the fungus was grown on the wood extract medium. It is hypothesized that by cleaving the ester linkage between hemicellulose and lignin, GEs would reveal the cellulose core inside

the matrix structure of lignocellulose, and open the space for cellulases, hemicellulases, and other enzymes for efficient hydrolysis of the biomass (Fig. 4-18). In turn, the actions of cellulases/hemicellulases might remove the layer of cellulose and hemicellulose outside and reveal more ester linkages buried inside. Thus the combined effect of *Nc*GE and the wood-decaying enzymes in *N. crassa* would have resulted in the efficient degradation of lignocellulosic biomass *in vivo*.

Table 4-1: Comparison between hardwood and softwood (Jasuja et al. 2017)**.**

	Hard wood	Softwood
Structure	The vessel elements or pores help to deliver water in the wood.	The water was conducted by tracheids system which is not considered as pores
Use	High-quality or long last furnitures such as deck and flooring	A wide range of applications (take ~ 80% of timber products) such as windows, doors, and fiberboard.
Density	Higher than softwood	Lower than hardwood
Cost	Generally higher	Generally lower
Growth	Slower growth rate.	Higher growth rate
Shedding of leaves	Normally, in the autumn and winter.	Evergreen
Fire-resistant	Higher	Slower
Examples	Aspen, balsa, beech, cherry, maple, oak, teak, and walnut.	Cedar, cypress, juniper, pine, redwood, spruce, and yew.
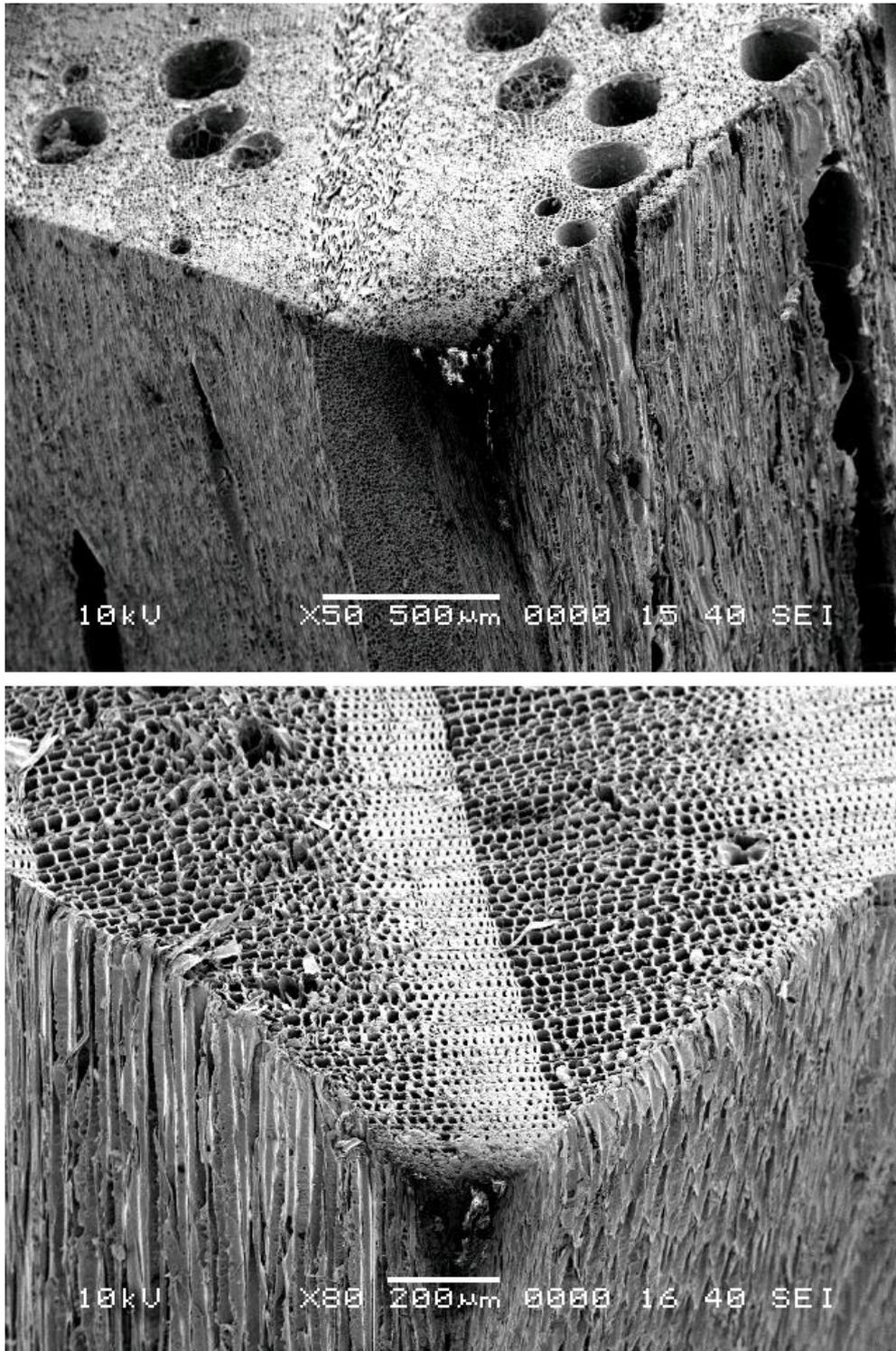
Figure 4-1: The images of hardwood and softwood under scanning electron microscope
The upper image displays the various pores in hardwood (Oak); the lower image presents the tracheid matrix in softwood (Jasuja et al. 2017).

Xyl*p*, xylopyranose; Glc*p*A, glucuronic acid; Ara*f* , arabinofuranose (Di Donato et al. 2015).
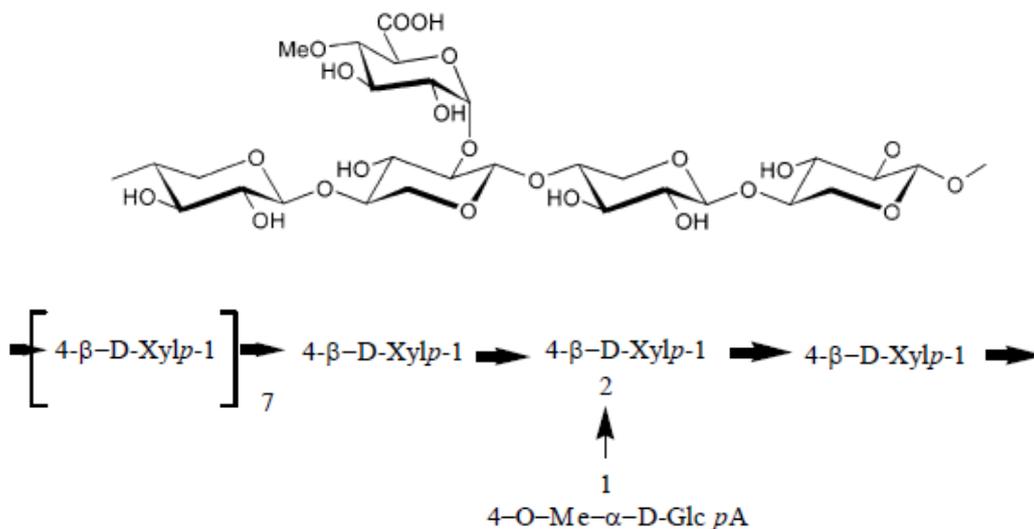
Figure 4-2: Structure of glucuronoxylan in hardwood

Xyl*p*, xylopyranose; Glc*p*A, glucuronic acid; Ara*f* , arabinofuranose (Di Donato et al. 2015).
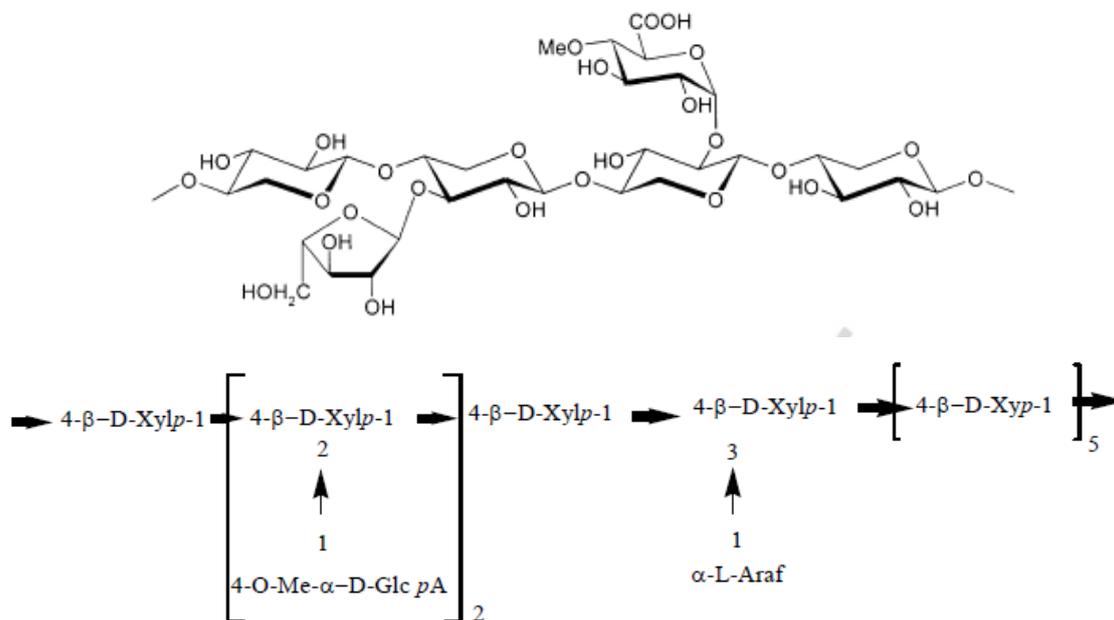
Figure 4-3: Structure of arabinoglucuronoxylan in softwood (Di Donato et al. 2015).

72
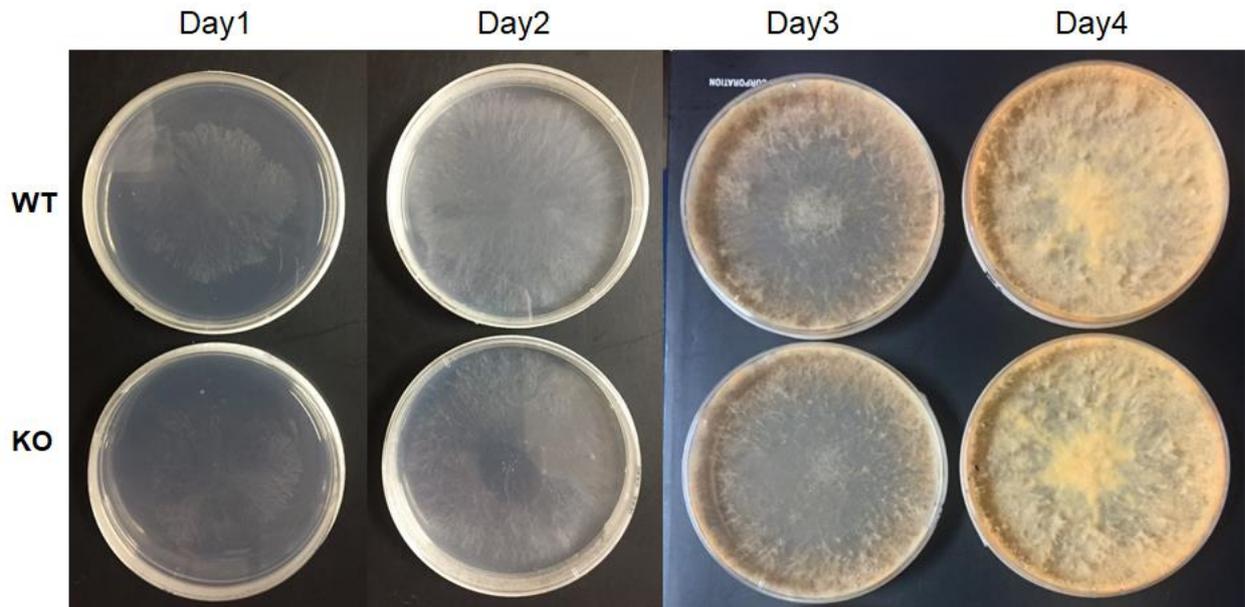
Figure 4-4: The growth of WT and KO strains in the sucrose-VM medium

WT and *Nc*GE KO strains of *N. crassa* (1×10^3 conidia in 5 µl) were spotted onto the center of the plate and grown in the sucrose-VM plates at 30°C.
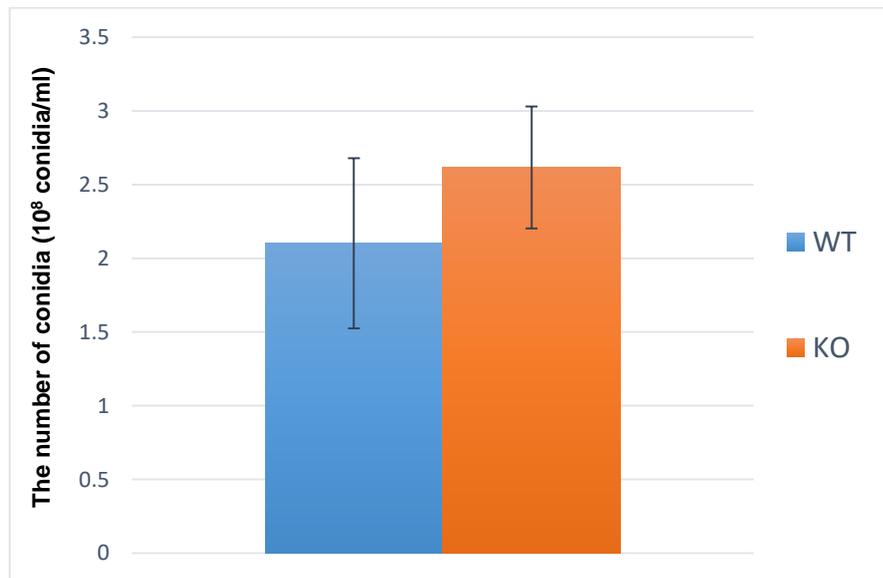
Figure 4-5: The numbers of conidia from WT and KO strains grown in the sucrose-VM plates for 4 days

Data presented the mean of at least three independent experiments with error bars indicating SD. *, $p \leq 0.05$; **, $p \leq 0.01$ (*t*-test).
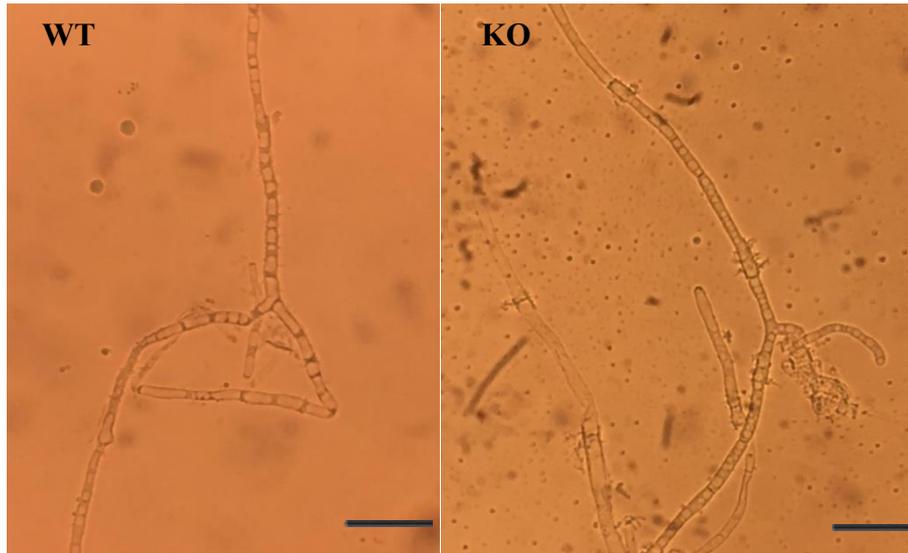
Figure 4-6: The hyphae of WT and KO strains under the microscope

Scale bar presents 10 µm.
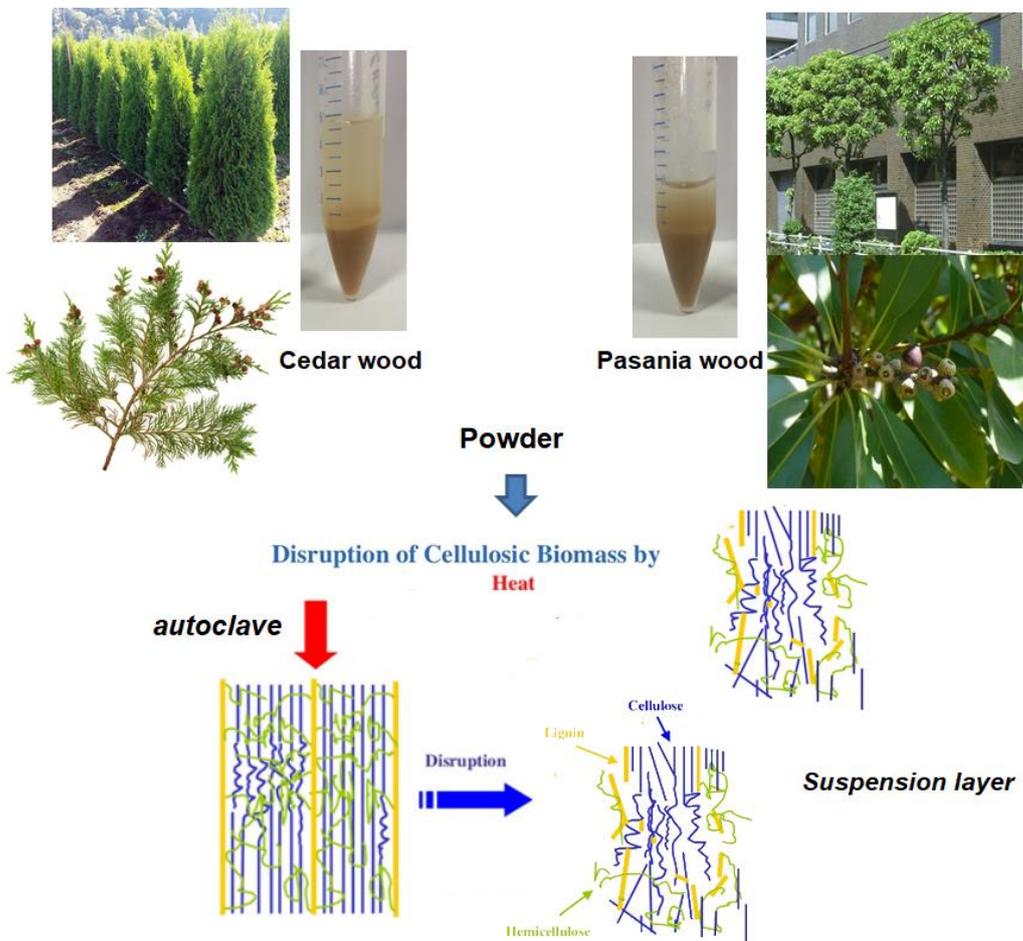
Figure 4-7: Preparation of the LCCs extract from Cedar and Pasania wood
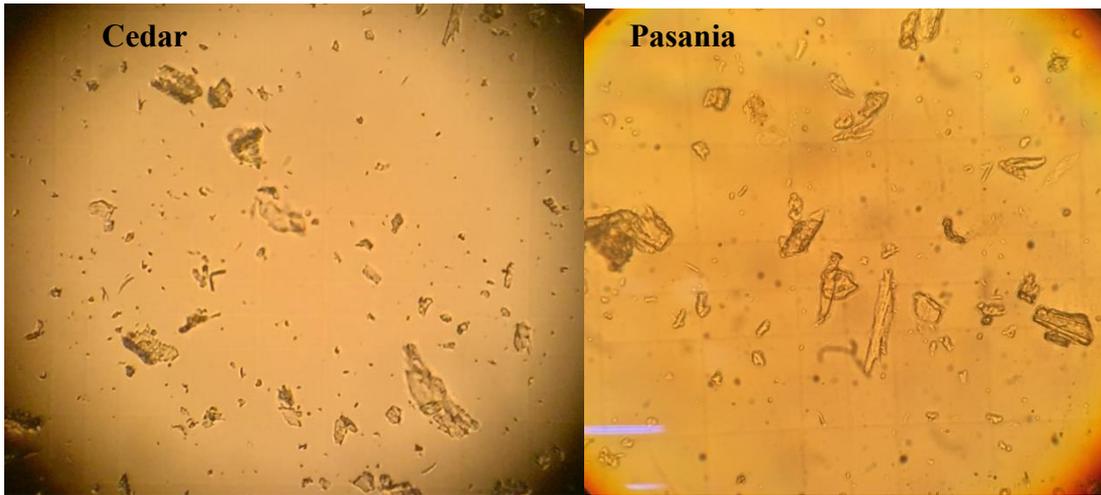
Figure 4-8: The micro-particles from the LCCs fractions of Cedar and Pasania wood
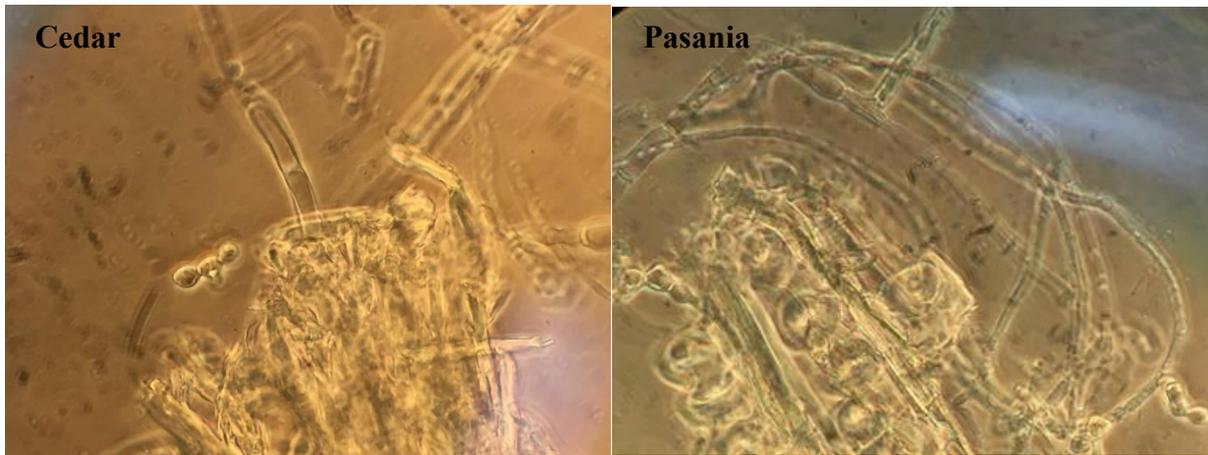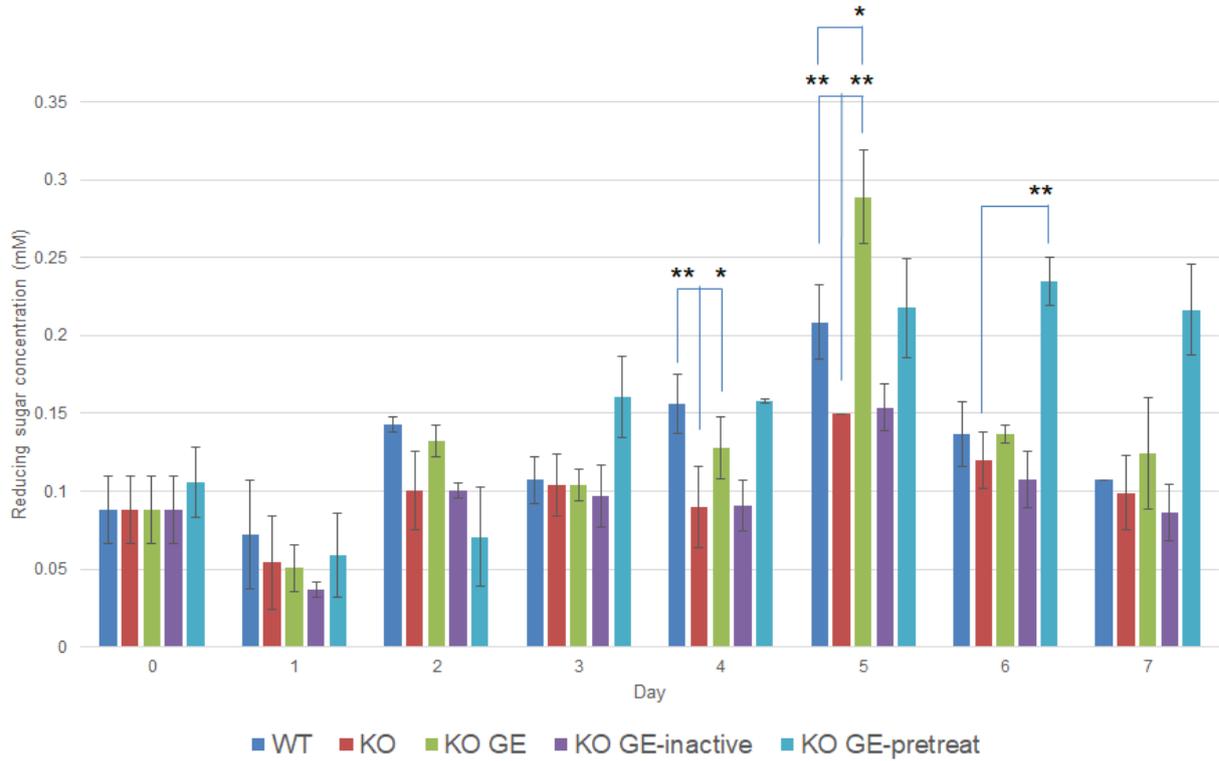

Figure 4-9: *N. crassa* **WT strain grown in the wood extract medium for 4 days.**
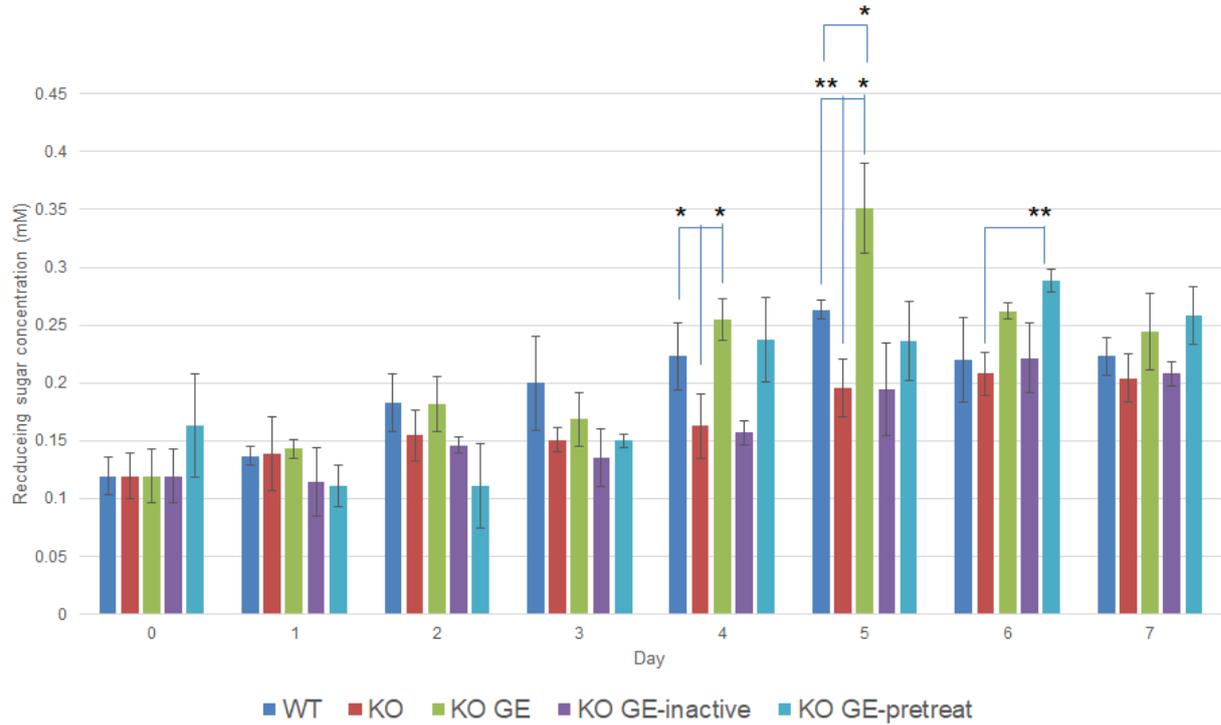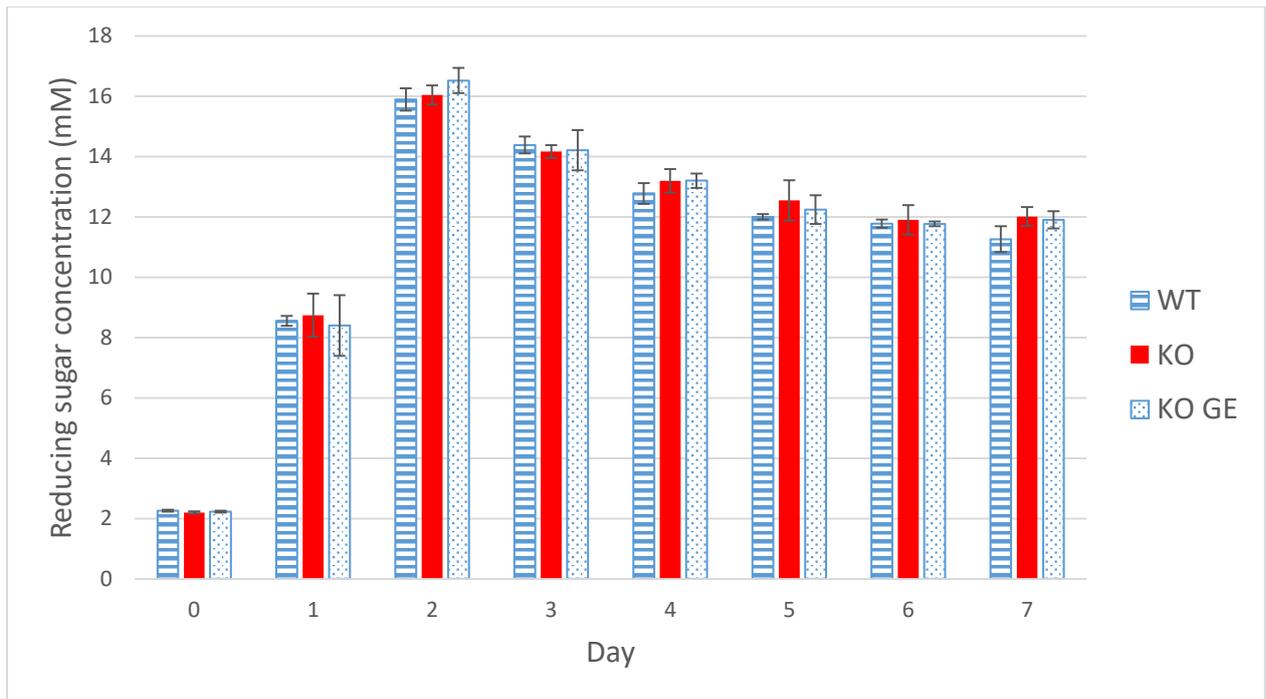
Figure 4-10: The release of reducing sugars in the liquid medium containing Cedar wood as the sole carbon source.

Data presented the mean of at least three independent experiments with error bars indicating SD. *, $p \leq 0.05$; **, $p \leq 0.01$ (*t*-test).
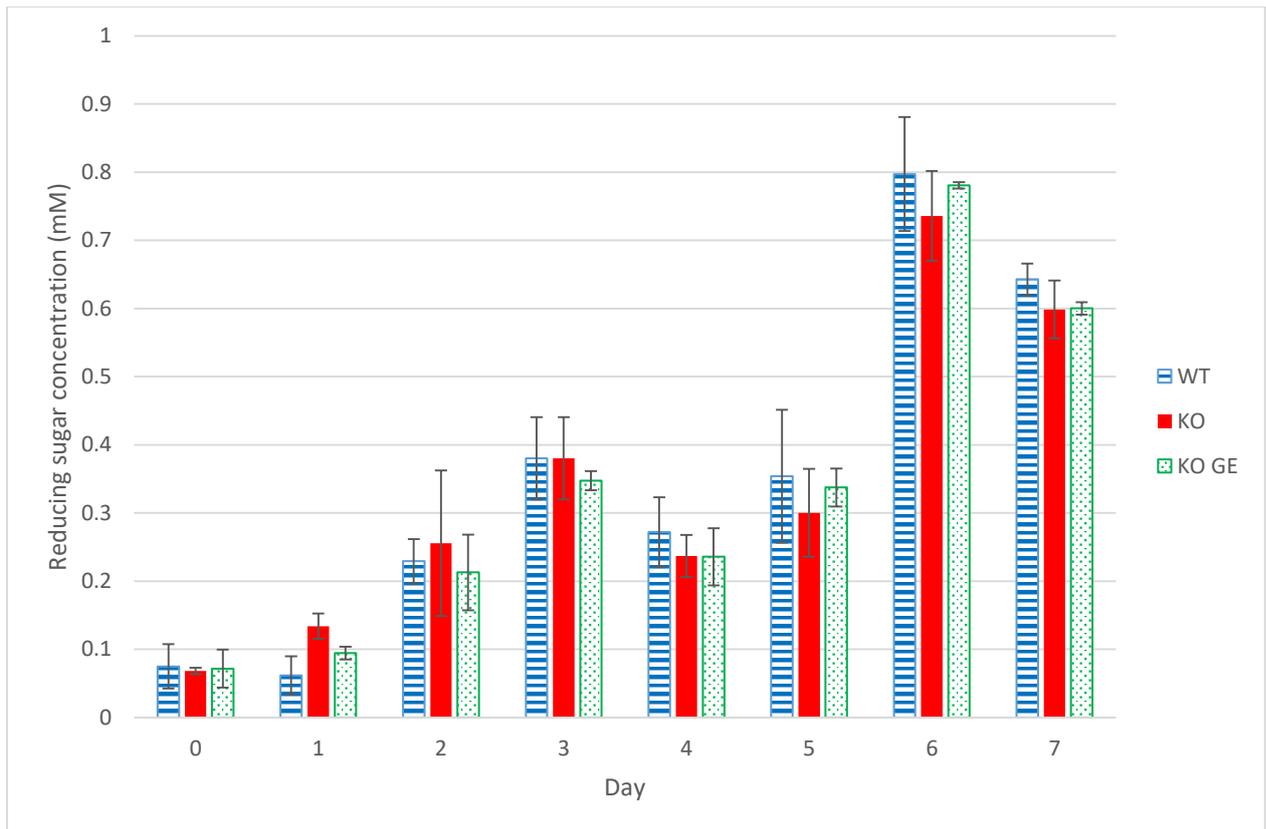
Figure 4-11: The release of reducing sugars in the liquid medium containing Pasania wood as the sole carbon source.

Data presented the mean of at least three independent experiments with error bars indicating SD. *, $p \leq 0.05$; **, $p \leq 0.01$ (t-test).

Figure 4-12: The release of reducing sugars in the liquid medium containing CMC as the sole carbon source

Data presented the mean of at least three independent experiments with error bars indicating SD. *, $p \leq 0.05$; **, $p \leq 0.01$ (*t*-test).

Figure 4-13: The release of reducing sugars in the liquid medium containing Avicel as the sole carbon source

Data presented the mean of at least three independent experiments with error bars indicating SD. *, $p \le 0.05$; **, $p \le 0.01$ (t-test).
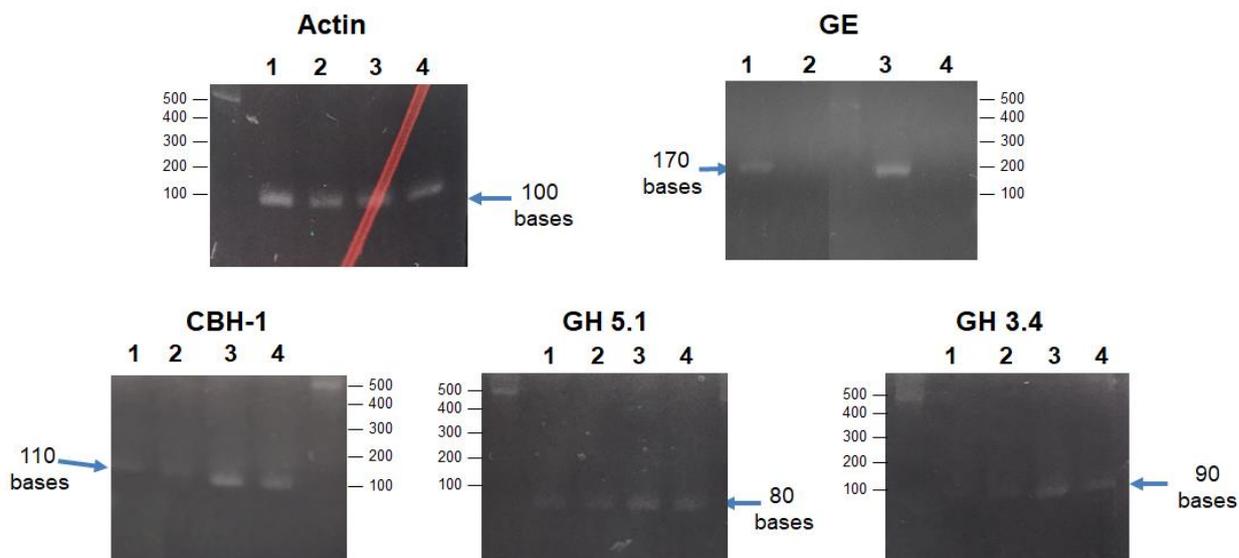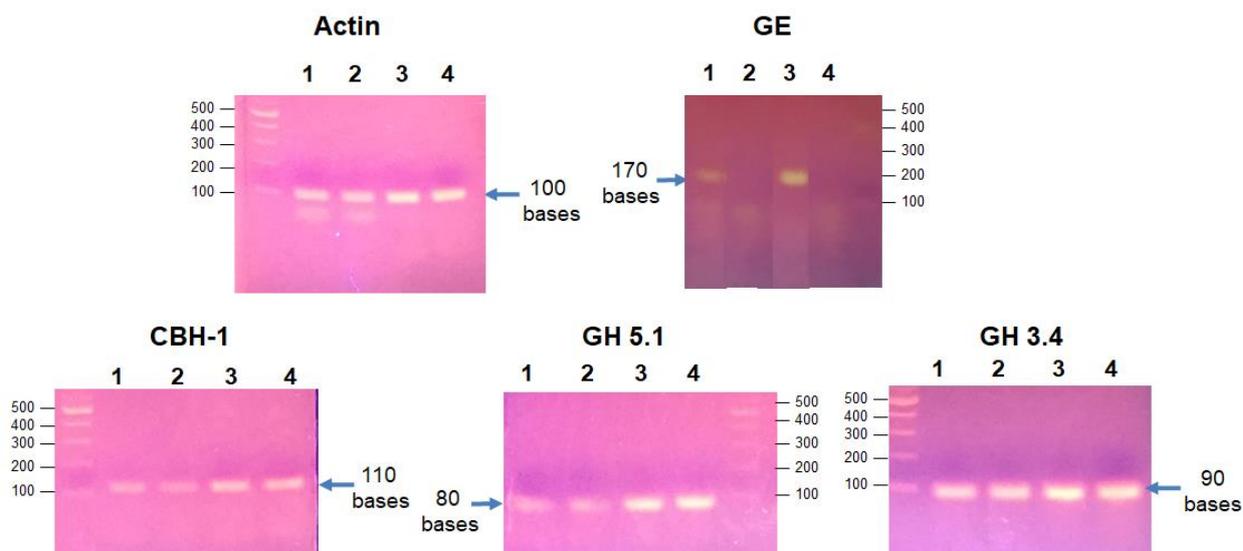
Figure 4-14: RT-PCR analysis of WT and KO strains grown in the VM medium containing sucrose or cedar wood

1, WT strain was cultured in the sucrose-VM medium; 2, KO strain was cultured in the sucrose-VM medium; 3, WT strain was cultured in the cedar-VM medium; 4, KO strain was cultured in the cedar-VM medium
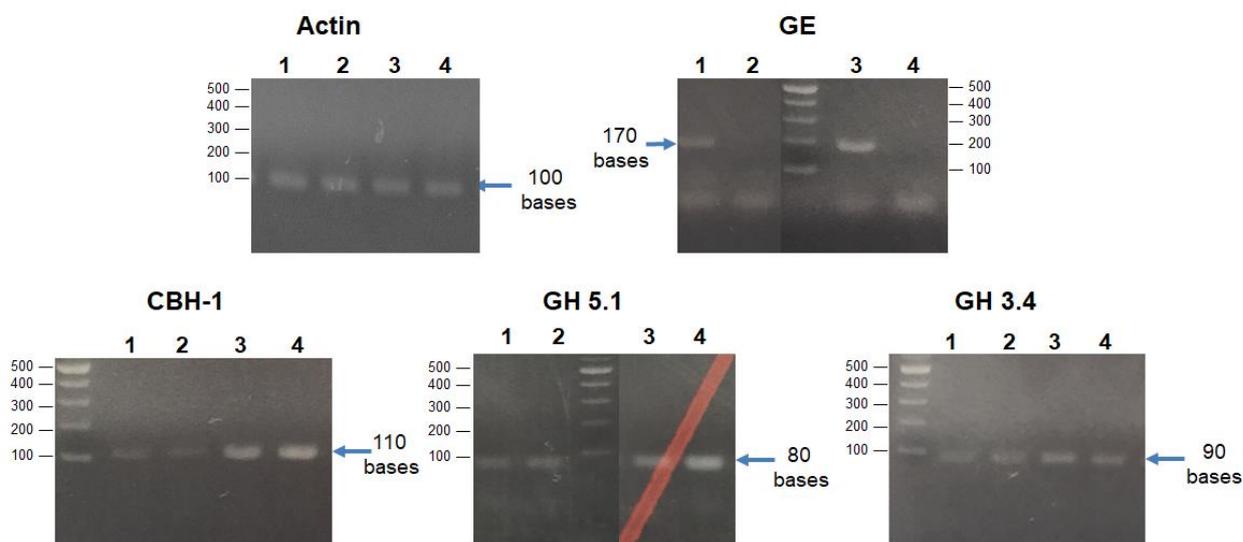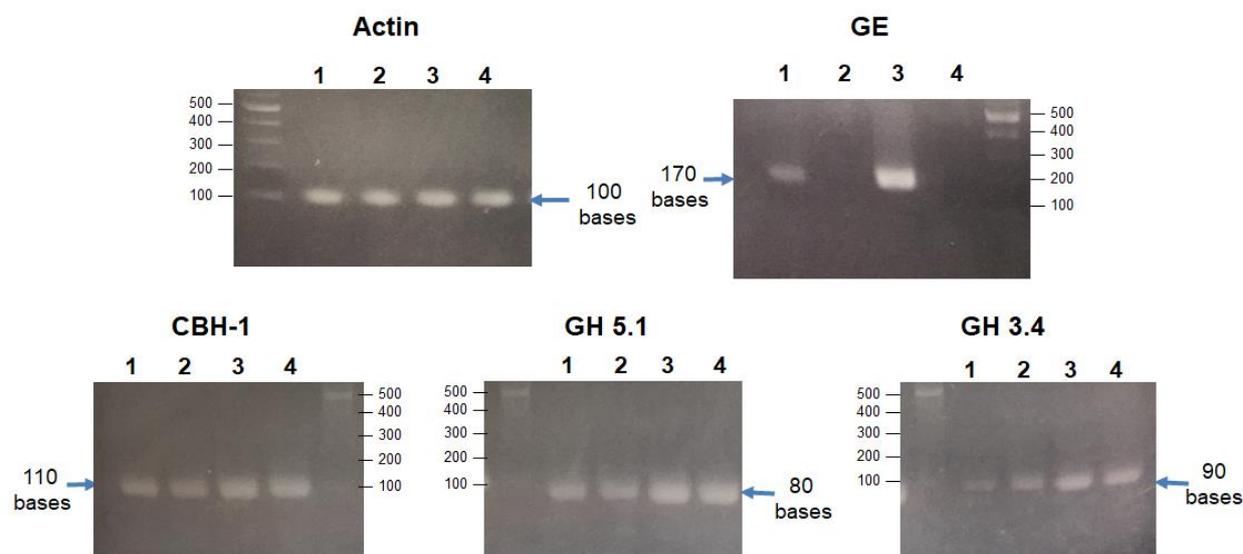

Figure 4-15: RT-PCR analysis of WT and KO strains grown in the VM medium containing sucrose or pasania wood

1, WT strain was cultured in the sucrose-VM medium; 2, KO strain was cultured in the sucrose-VM medium; 3, WT strain was cultured in the pasania-VM medium; 4, KO strain was cultured in the pasania-VM medium

Figure 4-16: RT-PCR analysis of WT and KO strains grown in the VM medium containing sucrose or CMC

1, WT strain was cultured in the sucrose-VM medium; 2, KO strain was cultured in the sucrose-VM medium; 3, WT strain was cultured in the CMC-VM medium; 4, KO strain was cultured in the CMC-VM medium

Figure 4-17: RT-PCR analysis of WT and KO strains grown in the VM medium containing sucrose or Avicel

1, WT strain was cultured in the sucrose-VM medium; 2, KO strain was cultured in the sucrose-VM medium; 3, WT strain was cultured in the Avicel-VM medium; 4, KO strain was cultured in the Avicel-VM medium
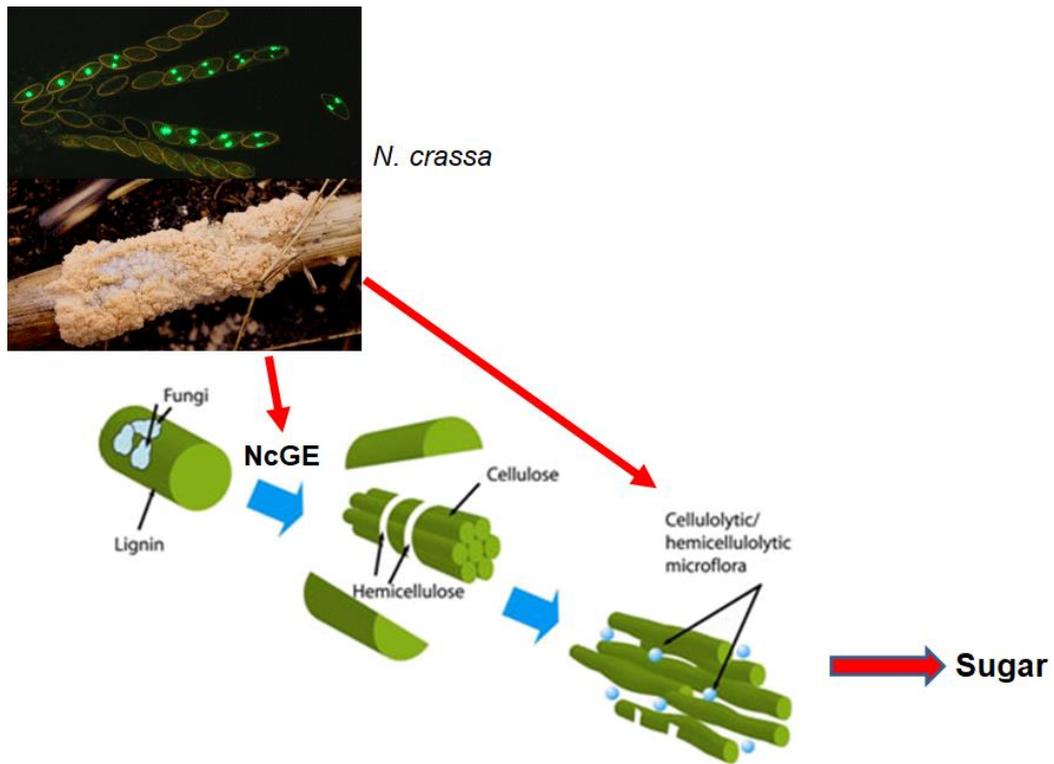
Figure 4-18: The suggested role of *Nc*GE in the lignocellulose degradation (Chaucheyras-Durand et al. 2012)**.**

## 5.1. The importance of studying glucuronoyl esterases

The world's major energy source, fossil fuel, is depleting which lead to the search for new alternative and renewable energy sources. Biofuels, especially bioethanol, have already used in many fields in the society. For example, it was blended into gasoline and petroleum as the transportation fuels in Brazil, US, and many countries. Lignocellulose, the most abundant organic material on earth, is expected to further improve the usage of biofuels. Therefore, many cellulases and hemicellulases have been searched for to establish a better system for enzymatic saccharification of lignocellulose. However, the production cost of cellulosic ethanol still does not satisfy the demand of commercial application. The ester linkage between hemicellulose and lignin in the matrix structure of lignocellulose is considered as one of the major obstacles for its enzymatic degradation. Hence, the unique ability of GEs to cleave this ester linkage has attracted more and more attention, as evidenced by the recent increase in the publication of GEs. A project "Optimized esterase biocatalysts for cost-effective industrial production" granted by the European Union (EU) (OPTIBIOCAT, www.optibiocat.eu) also shows a great interest in the CE15 family enzymes (Dilokpimol et al. 2017).

GEs have been classified to the novel CE15 family with only 10 members having been characterized. Although there are significant interests in this enzyme group, the number of studies dedicated to new GEs remains limited and therefore more investigation is needed. In 2017, at least two attempts of genome mining of GEs were performed not only to find new putative GEs but also to provide the overview about the structure and diversity in the sequence of the CE15-like proteins. These results indicated that a large number of GE sequences were mistakenly classified into other enzyme groups such as acetyl xylan esterases or sialidases due the lack of available information about GE sequence (Agger et al. 2017). In fact, *Af*GE was also classified as an acetyl xylan esterase at first. Moreover, the newly established Carboxylic Ester Hydrolases (CEH) database by Iowa State University (CASTLE - Carboxylic Ester Hydrolases, http://castle.cbe.iastate.edu/) currently puts GEs and acetyl xylan esterase in the same group, CEH8. Therefore, the identification of consensus sequences containing the catalytic triad as well as the role of conserved Lys209 in *Af*GE may provide more evidence for selecting new putative GEs, especially in the kingdom Fungi. The alignment of some newly identified GEs from both ascomycota and basidiomycota further confirmed this idea (Fig. 5-1). *Wc*GE1 from *Wolfiporia cocos* (Hüttner et al. 2017) was characterized, while other GEs such as *Ai*GE1 (from *Ascobolus immerses*), *Bd*GE1 (from *Botryosphaeria dothidea*), *Hs*GE1 (from *Hypholoma sublateritium*), *Pir*GE1 (from *Piromyces* sp. E2), and *Sh*GE1 (from *Stereum hirsutum*) were expressed in *P. pastoris* and the GE activities were only determined in the

culture supernatants. Three consensus sequences containing the catalytic triad (VTGC<u>S</u>RXGKGA, <u>H</u>C, and PQ<u>E</u>SG) are found in these GEs, except for *Pir*GE1 in which Arg is replaced by Tyr and *Sh*GE1 in which Pro is replaced by Ala. It is of note that *Piromyces* belongs to the phylum Chytridiomycota which is diverged in the beginning of the fungal lineages (Fig. 5-2). This suggests that GEs may have appeared early in evolution.

5.2. Future prospect

The cooperation of *Nc*GE and the cellulase/hemicellulase system to improve the lignocellulose degradation was suggested in this study. However, more research are needed to further clarify their physiological roles and the cooperation between GEs and other enzymes, especially the wood-decaying enzyme. Transcriptional profiling of GE and carbohydrate-active enzymes in *N. crassa* grown in different carbon sources may provide more information about the cooperation between them in biomass degradation. *Nc*GE showed the ability to enhance the release of reducing sugars from the wood extracts of both Cedar (softwood) and Pasania (hardwood). Three wood powders from beech wood (hardwood), mulberry (hardwood), and red pine (softwood) are available in our laboratory, thus they could be good candidates in the subsequent studies. Besides that, the addition of *Nc*GE in the KO GE condition resulted in the release of higher amount of reducing sugars compared to WT condition. Since the amount of *Nc*GE added to the culture was estimated to be much higher than that produced endogenously, overexpression of *Nc*GE might also provide a similar effect. This would prompt researchers to examine the effect of heterologously-expressing GEs in other wood-decaying organisms to enhance their ability to hydrolyze plant cell walls. Another viewpoint is that the separation of hemicellulose and lignin by GEs could facilitate the actions of lignin-modifying enzymes. Thus, the identification of cooperation between GE and other wood-decaying enzymes would help improve the enzymatic pretreatment of lignocellulose instead of using the extreme conditions such as alkaline treatment.
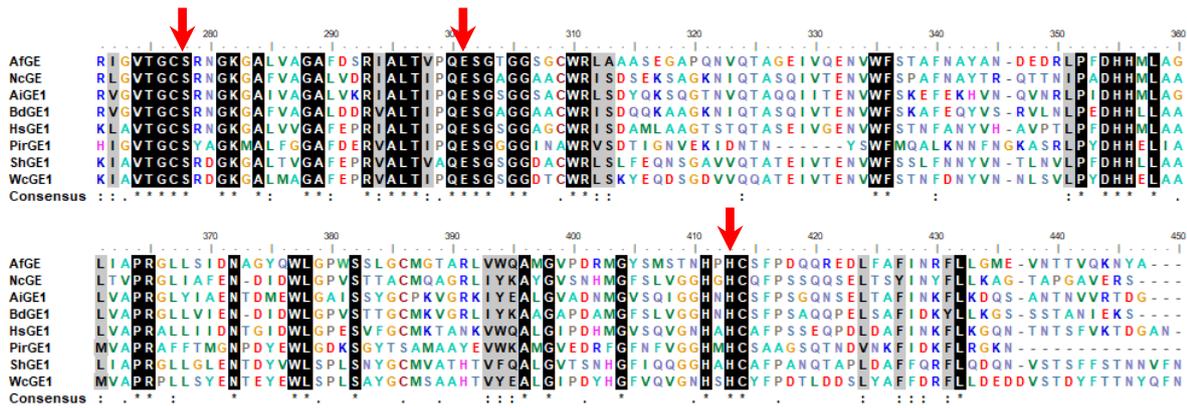
Figure 5-1. Amino acid sequence alignment of *Af*GE, *Nc*GE, and other newly identified GEs

The amino acid sequence alignment of *Af*GE, *Nc*GE, the Ascomycota GEs (*Ai*GE1 and *Bd*GE1), and the Basidiomycota GEs (*Hs*GE1, *Sh*GE1, and *Wc*GE1) is shown. The residues forming the catalytic triad Ser-His-Glu are marked by the arrows.
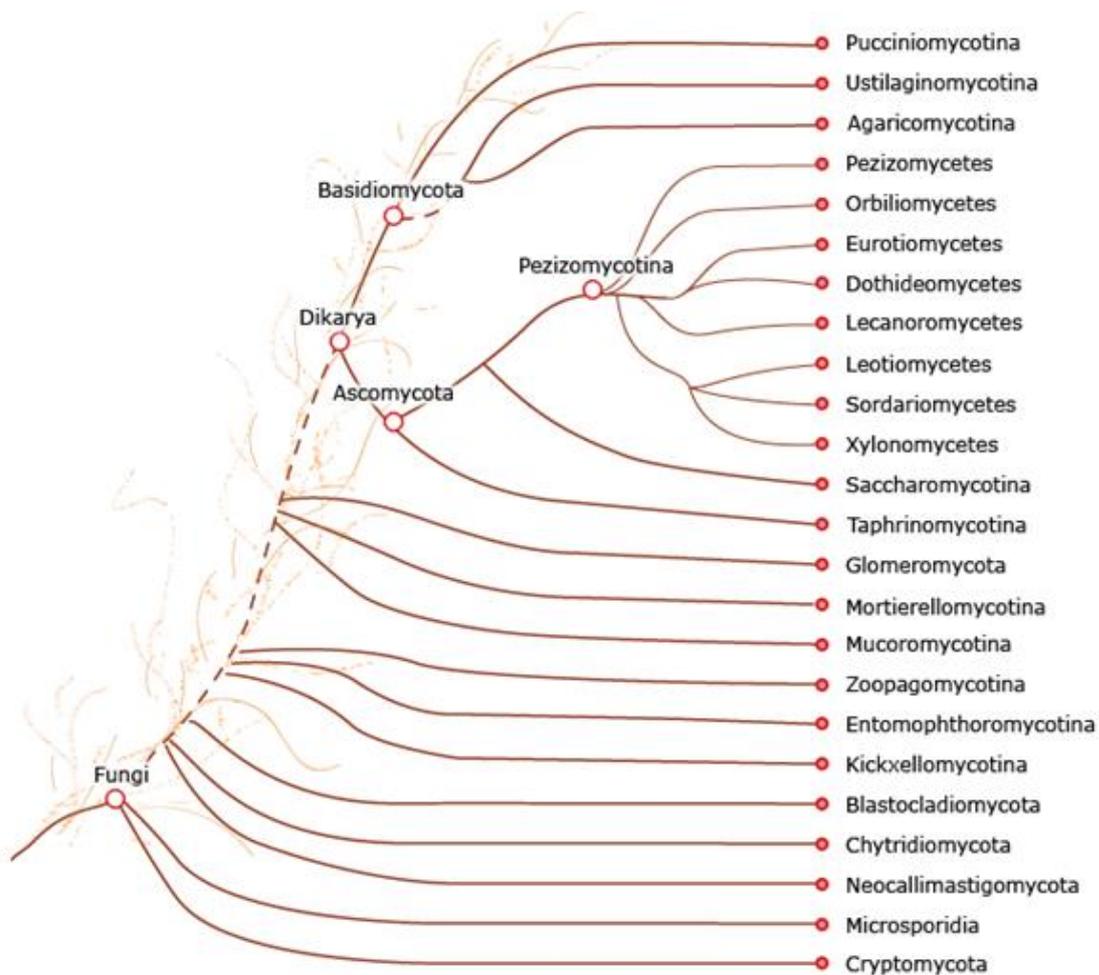
Figure 5-2. The phylogeny of the kingdom Fungi

(Source: https://genome.jgi.doe.gov/programs/fungi/index.jsf)

1. Strains and growth media

Escherichia coli DH5α was cultured in LB medium (1% peptone, 0.5% yeast extract, and 0.5% NaCl) containing the proper selection marker such as zeocin (40 μg/ml) or ampicillin (100 μg/ml) for pPICZα vector, and ampicillin (100 μg/ml) for pUNA vector.

Pichia pastoris KM71H strain (*his4 aox1*::*ARG4*) was used to express *Nc*GE. For methanol induction of *P. pastoris*, YPG medium containing 1% yeast extract, 2% peptone, 1% glycerol, and zeocin (100 μg/ml) was used, whereas glycerol and zeocin were omitted in YP medium.

A. oryzae niaD300 strain (*niaD⁻*), derived from *A. oryzae* RIB40 strain (ATCC42149), was chosen for the expression of *Af*GE. Conidia or mycelia of *A. oryzae* transformants were collected by growing the cells on the PDA agar medium (Potato Dextrose Agar; Nissui Pharmaceutical, Tokyo, Japan). To induce *Af*GE production, DPY medium containing 2% dextrin, 1% polypeptone, 0.5% yeast extract, 0.5% KH$_2$PO$_4$, and 0.05% MgSO$_4$ was used.

2. Construction of expression plasmids

2.1. Construction of *Nc*GE-pPICZα vector

The gene encoding *Nc*GE (NCU09445.7) without the signal peptide was amplified from the genomic DNA of *N. crassa* strain FGSC987 by polymerase chain reaction using PrimeSTAR DNA polymerase (Takara, Japan). Two PCR primers used to amplify the putative mature region (amino acids 21-395) were designed as follows: 5'-AAGCA<u>GAATTC</u>GCCCCTGCTTCTCAAATCT-3' (*Eco*R I site is underlined) and 5'-GC<u>TCTAGA</u>TACGACAAAGCCGGCACATC-3' (*Xba* I site is underlined). The PCR cycle includes denaturation (98°C, 10 sec), annealing (55°C, 10 sec), and extension (72°C, 2 min). The amplified fragment was digested with the restriction enzymes and then gel-purified to be cloned into pPICZαA vector. After verification of the nucleotide sequence, this recombinant plasmid pPICZαA/*Nc*GE was used in the subsequent studies.

2.2. Construction of *Af*GE-pUNA vector

Vector pUNA containing the dextrin-inducible *amyB* promoter and *amyB* terminator was used for the transformation of *A. oryzae* niaD300 strain. The target gene was inserted at the *Sma*I site of pUNA located at the downstream of the *amyB* promoter. First, the DNA fragment encoding Myc-His tag containing the stop codon was amplified by PCR from the

vector pPICZα (Thermo Fisher Scientific, Yokohama, Japan) using PrimeSTAR DNA polymerase (Takara Bio, Shiga, Japan) and the primers (Table S1). Second, the *Afu6g14390* gene encoding *Af*GE without the stop codon was obtained by PCR from the genomic DNA of the *A. fumigatus* Af293 strain (kindly provided by Dr. D. Hagiwara, Chiba University, Japan) using the reverse primer containing 18 bases complementary to the 5' sequence of Myc-His tag fragment. Then, a Myc-His tag was added to the C-terminal end of *Af*GE by the fusion PCR. The *Af*GE-Myc-His fragment was amplified by two primers containing the overlapping sequences (19 bases) flanking the *Sma*I site of the pUNA vector which was constructed using MultiSite Gateway^TM Technology (Mabashi et al. 2006). The recombinant vector pUNA-*Af*GE-Myc-His tag was generated by a novel cloning technique named Seamless Ligation Cloning Extract (SLiCE) (Motohashi 2015). The ligated plasmids were verified by sequencing.

3. Transformation of the expression plasmid

3.1. Transformation of *P. pastoris*

3.1.1. Preparation of *P. patoris* competent cells

KM71H *P. pastoris* strain was grown in 10 ml YPD medium at 30°C, 150 rpm/min for 24 hours. Then, 200 and 400 μl of the pre-culture solution were added to 200 ml YPD medium in 500 ml flask and culture at 30°C, 150 rpm/min until A_{600} of the culture achieved to 1.3-1.5 (it took around 13-15 hours). Next, the culture medium was divided into 4 falcon tube (50 ml). The cells were harvested by centrifuging at $1700 \times g$ for 7 min at room temperature. The cell pellet was washed two times with 40 ml MiliQ, and then two times with 10 ml sorbitol 1 M. After that, the cell was suspended in 0.5 ml sorbitol 1 M and ready for transformation at this time.

3.1.2. Electrotransformation of *P. patoris*

The plasmid pPICZαA/*Nc*GE was linearized with *Bgl* II and transformed to *P. pastoris* by electroporation according to the instruction manual of EasySelect *Pichia* expression kit (Invitrogen). Transformation of *P. pastoris* KM71 strain was performed by electroporator as the suggestion of the standard protocol (Invitrogen 2010). Briefly, 5 μl of the linearized plasmid containing 2.5-5 μg of DNA was mixed with 50 μl of competent cells and keep on ice for 5 min. Then, the cell suspension with plasmid was loaded into the iced-cuvette (Bio-Rad, catalog No.165-2086) and pulsed at the Pichia mode as suggested in the user's manual of Bio-Rad. After that, 1 ml of the sorbitol 1 M was added and incubated at 30°C for 1h with vigorous

shaking. After incubation, the cell suspension was centrifuged at 1700 × g for 5 min at room temperature. 800-900 µl sorbitol solution (upper layer) was removed and the pellet was spread on YPDZS (containing 100 µg/ml zeocin) agar medium. The transformants could be obtained after 3 or 4 days. Colony PCR using KOD FX Neo (Toyobo, Japan) was applied to screen the correct transformants. Positive clones were kept frozen in 25% glycerol at -80°C.

3.2. Transformation of *A. oryzae*

3.2.1. Preparation of *A. oryzae* protoplast

Transformation of *A. oryzae* was performed as described previously (Uchima et al. 2011). Conidia of *A. oryzae* niaD300 host strain (*niaD-*) were inoculated into 100 ml DPY liquid medium and grown at 120-150 rpm, 30°C for 18-20 h. The TF solution I was filtered through 0.45 µm sterilizing filter and kept in the L-shape tube. The mycelia were collected using Miracloth (Calbiochem, CA, USA) and washed with sterile distilled water. Then, the mycelia were incubated in 10 ml TF solution above at 30°C, 50 rpm for 3 h. Next, the protoplasts in the supernatant were collected by Miracloth (the remaining mycelia were disposed together with the Miracloth. 10 ml TF solution II were added to the protoplasts solution and centrifuged at 4°C, 2,000 rpm for 8 min. The protoplasts were re-suspended in TF solution II with a dropper. The amount of the protoplasts were counted by hemocytometer and adjust to the concentration around 5×10^7 protoplasts/ml.

3.2.2. Transformation of *A. oryzae*

200 µl of the protoplasts (above) was transferred into falcon tube (15 ml), gently mixed with 1-5 µg plasmid DNA and kept into the ice for 30 min. TF solution III was added to the DNA-protoplast mixture in three consecutive steps (250, 250, and 850 µl). Each time, the mixture needed to suspend carefully and then incubated at room temperature for 20 min. It was followed by adding 5 ml of TF solution II and centrifuged at 4°C, 2000 rpm for 6 min. Finally, the treated protoplasts were resuspended in 500 µl of TF solution II. At this point, it was mixed with 5 ml top agar (CD medium containing 1.2 M sorbitol and 0.8% agar) and spread on the agar plate containing the same medium as above. The transformant could be seen after 3-5 days. Transformants were transferred to a new selective medium twice for generation of the homokaryotic strains. Colony PCR using KOD FX Neo (Toyobo, Tokyo, Japan) was applied to confirm the correct transformants. These strains were then stored in 25% glycerol at -80°C for a long-term storage.

4. The heterologous expression of target protein

4.1. *P. pastoris* expression system

The *P. pastoris* transformant was pre-cultured in 10 ml YPG medium containing zeocin at 30°C for 24 h with strong shaking. Then the cells were transferred to the main culture containing 190 ml YPG medium in 500 ml flask and grown for another 24 h with shaking at 150 rpm. The cells were harvested by centrifugation (1,500 g, 5 min) and re-suspended in 40 ml YP medium for methanol induction. The medium was supplemented daily with 0.75% (v/v) methanol and the culture was continued for 4 days at 26°C. The supernatant was collected to analyze the protein expression by SDS-PAGE and Western blotting

4.2. *A. oryzae* expression system

4.2.1. Collection of conidia of *A. oryzae*

The *A. oryzae* strains grown in PDA plates in 5 days to fully produce conidia. 10 ml of sterile 0.01% Tween 80 was poured into the plate. The dropper was used to mix and collect the Tween 80 solution together with conidia. This suspension was filtrated using Miracloth and then centrifuged at 3,500 rpm for 5 min. The supernatant was discarded to wash with 5 ml of Tween 80, and centrifuge again. The supernatant was removed and 1 ml of sterile distilled water (DW) was added to become the suspension "A"). Hemocytometer was used to measure the conidia concentration in suspension "A".

For the calculation, 10 μl of diluted suspension "A" was put on the hemocytometer. The suspension "A" could be diluted if there are too many conidia. The number of conidia inside of 16 squares (X) was counted and applied to the next formula for calculating the concentration of conidia: X ÷ 4 × 62,500. For making a glycerol stock, 500 μl of the conidial solution "A" was mixed with 500 μl of sterile 80% glycerol, and stored at -80°C.

4.2.2. Culturing *A. oryzae* for protein expression

*Af*GE was expressed under the control of a strong promoter of α-amylase (*amyB*). To induce *Af*GE production, the conidia of *A. oryzae* transformants were grown in 5×DPY liquid medium (pH 8.0) with an approximate number of 1×10^7 conidia per 100 ml. After incubation at 30°C for 1-5 days at 150 rpm, the culture supernatant was collected by filtrating with Miracloth for further analysis.

5. Purification of the fungal GEs

5.1. Ni^{2+}-NTA affinity chromatography

5.1.1. Purification of *Nc*GE

A sufficient amount of Ni^{2+}-nitriloacetic acid (NTA) agarose was added to the column for binding of the secreted *Nc*GE protein in the supernatant. Purification of *Nc*GE was performed based on the instruction of QIAGEN manual. The contaminant proteins were removed by washing the column with 50 mM Tris-HCl buffer (pH 8), 200 mM NaCl containing 0 or 10 mM imidazole. The target protein was eluted with the same buffer containing 100 mM imidazole. To elute *Nc*GE, one ml of elution buffer was used in the first fraction, and 0.5 ml each of elution buffer was applied in the subsequent three fractions to obtain all *Nc*GE enzyme. Then, the elution fractions were dialyzed to keep at 10 mM Tris-HCl buffer (pH 7).

5.1.2. Purification of *Af*GE

Firstly, the supernatant of *A. oryzae* transformant was applied to Ni^{2+}-NTA affinity chromatography. The purification of *Af*GE was performed with HisTrap FF column (1 ml) based on the instruction of the manufacturer (GE Healthcare Japan, Tokyo, Japan). Sodium phosphate buffer (20 mM, pH 8) was used as the binding buffer. Contaminant proteins were washed with this buffer containing 0 and 10 mM imidazole. The target protein was eluted with the same buffer containing 100 mM imidazole. After that, the elution fractions were kept at 10 mM Tris-HCl buffer (pH 7) by dialysis.

5.2. Ion exchange chromatography

5.2.1. Purification of *Nc*GE using cation exchange column

To achieve higher purity, *Nc*GE partially purified by Ni^{2+}-NTA chromatography was applied to ÄKTA chromatography system (GE Healthcare) with PrimerView 5.0 program. Enzyme purification was carried out using HiTrap SP XL column (1 ml) with a flow rate of 1 ml/min at room temperature. The starting buffer was 10 mM Tris-HCl, pH 7, while the elution buffer contained 1 M NaCl. A linear gradient of 0% to 100% elution buffer was applied to collect the target protein. Purified *Nc*GE was kept in 10 mM Tris-HCl (pH 7) containing 20% glycerol for storage.

5.2.2. Purification of *Af*GE using anion exchange column

To achieve higher purity, the affinity-purified *Af*GE was applied to ÄKTA purifier chromatography system (GE Healthcare Japan, Tokyo, Japan) at room temperature.

Purification of *Af*GE was performed by HiTrap Q FF anion-exchange column (1 ml) and preserved in starting buffer (10 mM Tris-HCl, pH 7). Purification was carried out using a linear gradient of NaCl (from 0 to 1 M) at a flow rate of 1 ml/min. The protein concentration was determined by the Bradford assay (Bio-Rad, Hercules, CA, USA) using the bovine serum albumin as the standard (Wako, Tokyo, Japan).

6. Substrates for glucuronoyl esterase assay

In order to check the enzyme activity, the synthetic substrate 3-(4-methoxyphenyl) propyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate (substrate A) that mimics the ester linkage between lignin and hemicellulose was used (Špániková et al. 2007; Sasagawa et al. 2011). In addition, three other substrates, commercially available benzyl D-glucuronate (substrate B; Carbosynth, Berkshire, UK) (Sunner et al. 2015), benzyl methyl α-D-glucopyranosiduronate (substrate C), and benzyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate (substrate D) were also used. The latter two substrates were used to examine the preference of *Af*GE for 4-*O*-methyl glucuronoyl esters and synthesized by the collaborators (Ms. Nozomi Ishii and Dr. Ichiro Matsuo at Gunma University) as described below following the method reported by d'Errico et al. with slight modifications (d'Errico et al. 2015). The structures were confirmed by ^{1}H NMR spectroscopy and MALDI-TOF MS analysis. These four substrates contain the aromatic UV-absorbing chromophore which is suitable to detect GE activity by high-performance liquid chromatography (HPLC).

For the synthesis of substrate C, a mixture of methyl α-D-glucopyranoside (1.0 g, 5.2 mmol), NaBr (25 mg, 0.24 mmol), and NaClO (200 mg, 2.69 mmol) in H_2O (4 ml) was added 1 M NaOH (1 ml) and TEMPO (2.5 mg, 0.02 mmol). The reaction mixture was stirred for 8 h and the product was purified by silica gel column chromatography ($CHCl_3$/methanol = 50/1 ~ 1/2) to afford methyl α-D-glucuronic acid (0.7 g, 60%). The glucuronic acid derivative (100 mg, 0.48 mmol) and K_2CO_3 (49 mg, 0.36 mmol) in dry *N,N*-dimethylformamide (DMF; 500 μl) were stirred at 0°C for 30 min, then added benzyl bromide (86 μl, 0.71 mmol). The reaction mixture was stirred at room temperature for 8 h. The product was purified by silica gel column chromatography ($CHCl_3$/methanol = 10/1 ~ 2/1) to afford the compound C (31 mg, 22%).
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.36-7.26 (m, 5H, Ar), 5.24 (s, 2H), 4.80 (d, 1H, J = 3.6 Hz), 4.14 (d, 1H, J = 9.6 Hz), 3.76 (m, 2H), 3.58 (m, 1H), 3.41 (s, 3H); MS (MALDI-TOF) Calcd for $C_{14}H_{18}O7Na$ [M+Na]$^+$: 321.10, found: 321.26.

For the synthesis of substrate D, to a stirred solution of methyl 4-*O*-methyl α-D-

91

glucopyranoside (33 mg, 0.16 mmol), NaBr (11 mg, 0.10 mmol), NaClO (175 mg, 1.06 mmol) in H_2O (1.5 ml) was added 1 M NaOH (300 µl) and TEMPO (2.7 mg, 0.02 mmol). The mixture was stirred for 42 h. The product was purified by silica gel column chromatography ($CHCl_3$/methanol = 50/1 ~ 2/1) to afford methyl 4-*O*-methyl α-D-glucuronic acid (21 mg, 59%). The glucuronic acid derivative (21 mg, 0.09 mmol) was dissolved in dry DMF (350 µl). The mixture was added K_2CO_3 (40 mg, 0.29 mmol) and benzyl bromide (23 µl, 0.19 mmol) at 0°C and then stirred at room temperature for 3 h. The reaction mixture was diluted with $CHCl_3$ and washed with brine, dried (Na_2SO_4), and evaporated *in vacuo*. The residue was purified by silica gel column chromatography (hexane/ethyl acetate = 1/1 ~ 1/3) to afford the compound 2 (D) (8.4 mg, 28%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.41-7.32 (m, 5H, Ar), 5.26 (s, 2H), 4.81 (d, 1H, J = 4.0 Hz), 4.11 (d, 1H, *J* = 10.0 Hz), 3.78 (t, 1H, J = 9.6 Hz), 3.59 (m, 1H), 3.45 (s, 3H), 3.37 (t, 1H, *J* = 9.4 Hz), 3.37 (s, 3H); MS (MALDI-TOF) Calcd for $C_{15}H_{20}O_7Na$ [M+Na]$^+$: 335.11, found: 335.45.

7. Glucuronoyl esterase assays

Because of the self-degradation of the substrate under high alkaline condition (Špániková and Biely 2006; Li et al. 2007), enzyme reaction was conducted in 100 mM sodium phosphate buffer, pH 5.5 at 30°C. The reaction was stopped at an appropriate time 5-30 min by adding glacial acetic acid (1:5 (v/v)). One unit of GE activity was defined as the amount of the enzyme required to release 1 µmol of alcohol product per min at 30ºC. The de-esterified products were analyzed by HPLC system (Hitachi High-Tech Science, Tokyo, Japan) at room temperature. The mixture components were separated by using the reverse-phase column (SSC-2300, Senshu Scientific, Tokyo, Japan) with the mobile phase solvent acetonitrile:water (2:1 (v/v) for substrate A, 5:4 (v/v) for substrate B and C, and 3:1 (v/v) for substrate D) at a flow rate of 1 ml/min.

8. Enzymatic characterization

8.1. Effects of pH and temperature

The optimum pH was analyzed over the pH range 3.0-9.0 using 0.1 M citrate-phosphate buffer (pH 3.0-7.0) and 0.1 M Tris-HCl buffer (pH 7.0-9.0) at 30°C. The pH stability was evaluated by pre-incubating the enzyme with these buffers at 30°C for 30 min and then checking the residual activity in citrate-phosphate buffer at pH 5.0. The optimum temperature

was determined by performing the assay at various temperatures from 30°C to 70°C in 0.1 M citrate-phosphate buffer (pH 5.0). For the thermostability, the remaining activity was detected after incubation at different temperatures (30°C-70°C) for 30 min at pH 5.0.

8.2. Effects of metal ions and other reagents

Purified AfGE was pre-incubated at 30°C for 10 min in the presence of metal ions (1 mM each of Mg^{2+}, Mn^{2+}, Ca^{2+}, Fe^{2+}, Fe^{3+}, Cu^{2+}, Zn^{2+}, Al^{3+}, Ni^{+}, Li^{+}) or reagents (10 mM glycerol, Tween 80, acetic anhydride, ascorbic acid, dimethyl sulfoxide, dithiothreitol, 2-mercaptoethanol, SDS, urea; 0.5-1 mM phenylmethylsulfonyl fluoride (PMSF); 5-10 mM ethylenediaminetetraacetic acid (EDTA)). Next, the residual GE activity was measured; the control enzyme activity without cations/reagents was taken as 100%. The experiments were done in triplicates and the statistical analysis was performed using Student's t test.

8.3. Kinetic parameter analysis

Kinetic parameters were determined by the Lineweaver-Burk plot analysis using different concentrations of substrate A (0.2 to 10 mM) or substrate B, C, and D (1 to 20 mM). To quantify the enzyme activity, the amount of the reaction product was measured by the standard curve drawn using their corresponding commercial compound 3-(4-methoxyphenyl)-1-propanol (Sigma, MO, USA) and benzyl alcohol (Nacalai, Kyoto, Japan). All experiments were repeated three times.

9. Homology analysis

The amino acid sequences of GEs were collected from NCBI database and aligned by ClustalX program (Larkin et al., 2007). Evolutional analyses were conducted by MEGA6 (Tamura et al., 2013) using the Neighbor-Joining method. The bootstrap consensus tree was inferred from 1000 replicates. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The data were computed using the partial deletion and p-distance algorithm with out-group sequences including two different acetyl xylan esterases (AXEs) from *H. jecorina* (CAA93247.1; CE family 5) and *Aspergillus niger* (CAA01634.1; CE family 1), and one ferulic acid esterase (FAE) from *N. crassa* (CAC05587.1; CE family 1).

10. Site-directed mutagenesis

Wild-type and mutant AfGEs were produced using the methylotrophic yeast *Pichia*

pastoris KM71H strain as a host. The DNA fragment coding for the WT *Af*GE without the sequence for the signal peptide and an intron was obtained as follows. To remove the only one intron of *Af*GE, the DNA sequences carrying the first and second exons (without the stop codon) were amplified by PrimeSTAR PCR at first. In case of the first exon, the forward primer starts after the end of signal peptide sequence and the reverse primer contains 19 bases complementary to 5' sequence of the second exon. Then, two exon fragments were linked by fusion PCR using the primers (A and B) containing the homologous sequence (18 bases) with two ends of the pPICZα vector (cut with *Eco*RI and *Xba*I). In the next step, this mature WT-*Af*GE sequence was used as a template to create the mutated *Af*GEs by PCR-driven overlap extension. This method requires two internal primers containing desired mutations in the middle of primers (9 bases on either side) and hybridizing to the region to be altered (Table S1). These nested primers of two flanking primers A and B (above) and two internal primers were used in the first round of PCR to create two fragments containing the mutated internal codon inside the overlapping sequence region. In the second round of PCR, these DNA products were mixed with two primers A and B to amplify the full-length product that contains the desired mutation. Next, the recombinant plasmids pPICZα-(WT and mutated) *Af*GE were generated by SLiCE protocol (Motohashi 2015). After that, the plasmid harboring the WT and desired mutations were transformed into *P. pastoris* KM71H (Thermo Fisher Scientific, Yokohama, Japan) and expressed under methanol induction as instructed by Invitrogen's user manual. The WT and mutated *Af*GEs were purified for testing their activities by Nickel column as described above.

11. Molecular modeling

The 3D structure of *Af*GE was constructed using SWISS-MODEL via ExPASy web server. Visualization and analysis of structural models were performed by Chimera software. Protein-ligand accessible surface areas were calculated using PDBePISA server (http://www.ebi.ac.uk/pdbe/pisa/).

12. SDS-PAGE and gel electrophoresis analysis

The protein samples were kept in $1 \times$ loading buffer (mixed from $5 \times$ buffer solution). Then, the mixture was boiled for 103 min to denature protein completely. The electrophoresis was done using 12% polyacrylamide gels.

12.1. Coomassie Brilliant Blue (CBB) staining

The gel after electrophoresis was washed twice with DW in 10 min. Then it was stained with CBB solution overnight. After that, the gel was put into 10% (v/v) acetic acid to wash the remaining CBB inside the gel.

12.2. Western blot analysis

Western blot analysis of the culture supernatants was done as follows. Proteins were fractioned by SDS-PAGE (12% acrylamide) as mentioned above, which were then transferred to membrane Immobilon-P membranes (0.45 μm; Merck Millipore, Tokyo, Japan). Anti-c-Myc mouse monoclonal antibody and peroxidase-labeled anti-mouse IgG antibody was used as the primary and secondary antibodies, respectively. Western Lightning Plus system (PerkinElmer, Waltham, MA, USA) was used to detect c-Myc-tagged fusion proteins. The target protein bands were visualized through luminescent image analyzer (LAS-4000 mini; Fujifilm, Tokyo, Japan).

12.3. Endoglycosidase H (Endo H) treatments

The process was done according to the manufacturer's guideline (New England Biolabs). The sample containing around 1-20 μg glycoprotein was mixed with the denaturing buffer and DW to make 10 μl reaction volume. This mixture was boiled for 10 min. The final 20 μl deglycosylation sample was made by adding 2 μl of $10 \times$ G5 reaction buffer, DW, and 1-5 μl Endo H. The solution was incubated at 37 °C for 1 hour. SDS-PAGE could be applied to analyze the protein sample after treatment

13. Protein assay

Bradford assay was used to determine the protein concentration using the standard reagent of Bio-Rad. The reagent was diluted 5 times before use and bovine serum albumin (Takara, Japan) was prepared to make the standard with different concentration (0.2, 0.4, 0.6, and 0.8 mg/ml of BSA). 20 μl of protein sample was incubated with 1 ml Bio-Rad protein assay solution and measured the absorbance at 595 nm (A_{595}). The reaction mixture was kept at room temperature for 5 min before checking with spectrophotometers.

14. Reducing sugar assay

The tetrazolium blue reagent (0.1 % tetrazolium blue in 0.05 M NaOH) was used to identify the amount of reducing sugar in the medium. 10 μl sample was added to 1 ml tetrazolium blue reagent. The mixture was boiled for 10 min and cool down at room

temperature. The reducing sugar was measured at A_{660}. A standard curve from the series of glucose concentration (0.1, 0.2, 0.4, 0.7, 1 mM) was made to calculate the reducing sugar concentration of the samples.

15. SLiCE protocol

15.1. Preparation of SLiCE solution (from *E. coli* DH5α extract)

The DH5α strain was pre-cultured in 10 ml LB, incubate at 37°C, overnight (14-16h). Then, the main-culture was set up by transferring 1 ml pre-culture to 50 ml 2xYT in 500 ml flask and incubate at 37°C, 150 rpm until OD_{600} reaches a value of 2-3 (about 6-7 h). The cells were harvest by centrifugation at 5,000g for 10min at 4°C and washed with 50ml ice-cold sterilized Milli-Q. The wet cells were gently resuspended in 1.2 ml 3% (w/v) Triton X-100 in Tris-HCl (pH 8.0) and kept at room temperature for 10 min. After that, the solution was centrifuged at 20,000g for 2 min at 4°C. The supernatant was collected and kept on ice, then added 1 volume of ice-cold 80% (v/v) glycerol. Aliquot 40µl of each SLiCE extract into the 0.2mL-PCR tubes. Snap-freeze in a bath of liquid nitrogen and maintain this stock solution at -80°C. For short-term storage, SLiCE extract can be stored at -20°C for 3 months without significant loss of activity.

15.2. Standard protocol for SLiCE reaction

The principle of SLiCE is similar to In-Fusion cloning but the overlap sequence is about 19 base pairs (bp) instead of 15 bp. If the vector was cut by the restriction enzyme producing sticky end, the sequence of restriction sited is not counted for the overlap sequence. The vector DNA (about 50 ng) was mixed with the insert DNA at the ratio vector: insert (in concentration) around 1:3. Then 1 µl of $10\times$ SLiCE buffer, 1 µl SLiCE extract and DW was added to become 10 µl solution. The mixture was incubated at 37°C for 15-20 min. After that, all 10 µl was used to transform to 100 µl competent *E. coli*

16. Culturing *N. crassa*

16.1. Preparing the wood extract medium

To prepare 100 ml of the medium, 2 g of wood powder was transferred to 100 ml flask, added with 40 ml of DW, and autoclaved at 121ºC for 15 min. Then, all of the sample was move to the 50 ml conical tube and centrifuged at 3,000 rpm for 3 min. The supernatant was collected as the wood extract. The pellet was added with 40 ml of DW, vortexed, and

centrifuged to obtain a total of 80 ml wood extract. This wood extract was supplied with 2 ml of Vogel's salt and DW to fill up to 100 ml.

16.2. Culturing *N. crassa*

The conidia of *N. crassa* were collected as described in the *A. ozyzae* section. After counting the number of conidia, they were added to the medium with an approximate number of 1×10^5 conidia per 10 ml. The conidia were cultured in 10 ml of the wood extract medium as described above in 50 ml conical tube at 30°C, 135 rpm for 7 days. Each day 100 µl of culture supernatant was collected to be analyzed afterwards. When *N. crassa* was grown in CMC or Avicel, 2% of the designated substrate was put into the medium instead of 2% of sucrose in the standard Vogel's medium.

For culturing in the KO GE condition, the purified *Nc*GE (0.8 µg) was added to the culture (10 ml) daily. For KO GE-inactive condition, the similar process was done with the heat-inactivated *Nc*GE which shows no GE activity. The heat-inactivated *Nc*GE was made by boilinig the purified *Nc*GE for 10 min. For KO GE-pretreat condition, at first the same amount (0.8 µg) of purified *Nc*GE was added daily to the culture medium (10 ml) and incubated for 5 days at 30°C. Then the pretreated medium was used to culture the fungus; the purified *Nc*GE was not added to culture medium after inoculation.

16.3. Preparation of the purified *Nc*GE

The purified *Nc*GE was obtained from the *P. pastoris* transformant (Chapter 1). The expression of *Nc*GE was induced by methanol in 40 ml YP medium for 2 days. Then, the culture medium was collected and centrifuged at 5,000 rpm for 5 min. The clear supernatant was loaded to the Ni-NTA-agarose column and *Nc*GE was eluted with 2 ml of 50 mM Tris-HCl buffer (pH 8), 200 mM NaCl containing 100 mM imidazole. The eluted sample was dialyzed in 1 litter of 10 mM Tris-HCl buffer (pH 7) for overnight. The dialyzed samples were aliquoted (30 µl) into each of the 0.2ml-PCR tube and stored in -20°C.

16.4. Statistical analysis

The experiments were essentially done in triplicates and the statistical analysis was performed using Student's *t* test.

17. Preparation of total RNA

The mycelia in the culture medium were collected Miracloth and wash three times with the same volume of distilled water. The sample was kept at room temperature for 10 min to briefly dry. 100 mg mycelia were frozen in liquid nitrogen and crushed by the multi-bead shocker machine (Yasui Kikai, Japan). 1 ml ISO-GEN (Nippon Gene, Toyama, Japan) was added and transfer to a new 1.5 ml tube. From this step, the equipment needs to ensure RNase-free. Then 200 μl of chloroform was added to the ISO-GEN solution, stir with a vortex for 15 sec, and keep at room temperature for 2-3 min. The solution was centrifuged at 15,000 rpm for 15 min at 4°C. The upper layer was carefully collected and then mixed with 500 μl 2-propanol. The sample was left at room temperature for 7 min before centrifuged at 15,000 rpm for 10 min at 4°C. The supernatant was discarded and 1 ml 70% ethanol was added to the pellet. The sample was centrifuged at 7,500 rpm for 5 min at 4°C. The supernatant was removed and kept at the room temperature with the opened cap to dry the remaining ethanol. Finally, the pellet was dissolved in around 50-100 of diethylpyrocarbonate (DEPC)-treated water. The concentration of RNA was measured at A_{260} and the samples with the ratio of A_{260}:A_{280} higher than 1.8 are chosen for analysis in next step. The concentration of RNA (Y) was calculated as follows: Y (μg/ml) = Absorbance at A_{260} (unit of single-stranded RNA) × 40 μg/ml × 50 (the dilution value in this experiment)

17.1. DNase treatment

The following process is based on the user's manual of PrimeSript^{TM} RT reagent Kit (Takara, Shiga, Japan). Around 1 μg of total RNA was mixed with 2 μl of 5 × gDNA Eraser buffer, 1 μl of gDNA Eraser, and RNase free dH$_2$O to become total 10 μl. The solution was incubated at 42°C for 2 min or room temperature for at least 5 min.

17.2. RT-PCR analysis

After that, 4μl of 5 × PrimeScript buffer 2, 1μl of PrimeScript RT Enzyme Mix I, 1μl of RT Primer Mix, and RNase free dH$_2$O were added to 10 μl DNase treatment solution above to total 20 μl. The reverse transcription was performed with 1 cycle of 37°C for 15 min, 85°C for 5 sec, and the sample was kept at 4°C. For amplification the cDNA after reverse transcription, r*Taq* polymerase was used.

18. Polymerase Chain Reaction (PCR)

18.1. KOD FX Neo (Toyobo)

2 × KOD FX Neo buffer	10 µl
2 mM dNTPs	4 µl
10 µM Primer F	0.4 µl
10 µM Primer R	0.4 µl
KOD FX Neo DNA polymerase	0.4 µl
TE	2 µl
DW	2.8 µl
Total	20 µl

For colony PCR, a small amount of mycelia taken by the bamboo stick was firstly stirred in 100 µl TE. Then 2 µl this mixture was added to PCR solution.

Procedure:

1. 94°C, 2 min (initialization)

2. 98°C, 10 s (denaturation)

3. Tm°C, 30 s (annealing)

4. 68°C, 1 kb/30 s (elongation)

5. 16°C (final hold)

(The process from step 2 to 4 was repeated 30-35 times)

18.2. PrimeSTAR DNA polymerase (Takara)

5 × PrimeSTAR buffer	10 µl
2.5 mM dNTPs	4 µl
10 µM Primer F	1 µl
10 µM Primer R	1 µl
Template	1 µl (< 200 ng)
PrimeSTAR DNA polymerase	0.3 µl
DW	32.7 µl
Total	50 µl

Procedure:

1. 94°C, 2 min (initialization)

2. 98°C, 10 s (denaturation)

3. Tm°C, 30 s (annealing)

4. 68°C, 1 kb/30 s (elongation)

5. 16°C (final hold)

(The process from step 2 to 4 was repeated 30-35 times)

18.3. r*Taq* DNA polymerase (Takara)

Composition	For colony PCR	For RT-PCR
10 × r*Taq* buffer	2 μl	2 μl
2.5 mM dNTPs	2 μl	2 μl
10 μM Primer F	2 μl	1 μl
10 μM Primer R	2 μl	1 μl
Template	*	1ul
PrimeSTAR DNA polymerase	0.1 μl	0.1 μl
DW	11.9 μl	12.9 μl
Total	20 μl	20 μl

*: in colony PCR, a small amount of mycelia taken by the bamboo stick was stirred in PCR solution.

Procedure:

1. 94°C, 5 min (initialization)

2. 98°C, 30 s (denaturation)

3. Tm°C, 30 s (annealing)

4. 72°C, 1 kb/min (elongation)

5. 72°C, 5 min (final elongation)

6. 16°C (final hold)

(The process from step 2 to 4 was repeated 30-40 times)

19. Primers list

19.1. Cloning the *Nc*GE gene

Primer name	Sequence (5′ to 3′)
*Nc*GE F	AAGCA<u>GAATTC</u>GCCCCTGCTTCTCAAATCT
*Nc*GE R (non-stop)	GC<u>TCTAGA</u>TACGACAAAGCCGGCACATC

Underlined: *Eco*RI site and *Xba*I site, respectively.

19.2. Cloning the *Af*GE gene

Primer name	Sequence (5´to 3´)
*Af*GE F	*GAATTCGAGCTCGGTA***CCC**ATGTTCCTTTACCCTGTCCT
*Af*GE R (non-stop)	*TGAGTTTTTGTTCTAGAAA*AGTCAAATTGGGAACTTGCC
Myc-His F	*TTTCTAGAACAAAAACTC*AT
Myc-His R	*CGAGCTACTACAGATC***CCC**TCAATGATGATGATGATGAT
*Af*GE exon1 pPIC F	*AAAAGAGAGGCTGAAGCT*GAATTCGCCACCTGCCCGAGTCT
*Af*GE exon1 pPIC R	*GGGCAGGTGGCAGCCAGTG*CTCCACTGGAGATAAGGGA
*Af*GE exon2 pPIC F	CACTGGCTGCCACCTGC
*Af*GE exon2 pPIC R	*GAGATGAGTTTTTGTTCT*AGACCAGTCAAATTGGGAACTTGCC
*Af*GE-pI K209A F	CGGAACGGC<u>GCT</u>GGCGCATTA
*Af*GE-pI K209A R	TAATGCGCC<u>AGC</u>GCCGTTCCG
*Af*GE-pI K209E F	CGGAACGGC<u>GAG</u>GGCGCATTA
*Af*GE-pI K209E R	TAATGCGCC<u>CTC</u>GCCGTTCCG
*Af*GE-pI K209Q F	CGGAACGGC<u>CAA</u>GGCGCATTA
*Af*GE-pI K209Q R	TAATGCGCC<u>TTG</u>GCCGTTCCG
*Af*GE-pI K209R F	CGGAACGGC<u>AGA</u>GGCGCATTA
*Af*GE-pI K209R R	TAATGCGCC<u>TCT</u>GCCGTTCCG

Italic: overlapping sequence

Bold: a part of *Sma*I in pUNA vector

<u>Underline:</u> mutated sequence

19.3. RT-PCR analysis

Primer name	Sequence (5´to 3´)
*Nc*GE RT F	GCCCCTGCTTCTCAAATCT
*Nc*GE RT F	GTACTGCTGGAGGATCTTG
Nc Actin1 F	TGATCTTACCGACTACCT
Nc Actin1 R	CAGAGCTTCTCCTTGATG
NCU07340 cbh-1 F	ATCTGGGAAGCGAACAAAG
NCU07340 cbh-1 R	TAGCGGTCGTCGGAATAG
NCU00762 gh5-1 F	GAGTTCACATTCCCTGACA
NCU00762 gh5-1 R	CGAAGCCAACACGGAAGA
NCU04952 gh3-4 F	GACGCAGCCTATTCTCAG
NCU04952 gh3-4 R	CTTATTCCAGCCGACTCC

List of recipes

A. Medium

Czapex Dox medium (selective media for *A. oryzae*)

0.3% NaNO$_3$, 0.2% KCl, 0.1% KH$_2$PO$_4$, 0.05% MgSO$_4$.7H$_2$O, 0.002% of 2% FeSO$_4$.7H2O, and 2% glucose. Adjust to pH 5.5.

→ Top agar is the selective media containing 1.2 M sorbitol and 0.8% agar.

5 × DPY liquid medium (for protein expression of *A. oryzae*)

- Solution A (pH 8.0)

- 10% dextrin
- 5% polypeptone

- Solution B (pH 8.0)

- 2.5% yeast extract
- 0.5% K$_2$HPO$_4$
- 0.05% MgSO$_4$.7H$_2$O

Autoclave solution A and B separately. Mix after the solutions cool down.

DPY liquid medium (for growth of *A. oryzae*)

2% dextrin, 1% polypeptone, 0.5% yeast extract, 0.5% KH$_2$PO4, and 0.05% MgSO$_4$, 7H$_2$O, Adjust to pH 5.5.

Luria Bertani (LB - growth of *E. coli*)

1% bacto peptone, 0.5% NaCl, 1.5% agar, Ampicillin 100 µg/ml (if required). Add 10 ml of LB into L-shape tube and sterilize.

Potato Dextrose Agar (PDA, growth of fungi in general)

PDA was made according to the instructions given by the manufacturer (Nissui Pharmaceuticals; 39 g/l).

VM medium (for culturing *N. crassa*)

2 ml of 50 × Vogel's salt, 2% sucrose, 98 ml H$_2$O, 100 µg/ml L-histidine (if required). Add hygromycin (300 µg/ml) after the solution cools down (if required). For complex media, add 0.5% yeast extract and 0.5% tryptone (from casein).

50 × Vogel's salt (per 2 litter)

253.6 g sodium citrate.$2H_2O$, 500 g KH_2PO_4, 200 g NH_4NO_3 (omit if using for selecting media), 20 g $MgSO_4.7H_2O$, 10 g $CaCl_2$, 10 ml biotin solution, and 10 ml trace element.

Biotin solution

5 mg biotin in 100 ml 50% ethanol (v/v). Filter sterilize and keep at 4^oC

Trace element

5 g citric acid, 5 g $ZnSO_4.7H_2O$, 1 g $Fe(NH_4)2SO_4.6H2O$, 0.25 g $CuSO_4.5H_2O$, 0.05 g $MnSO_4.H_2O$, 0.05 g H_3BO_3, 0.05 g $NaMoO_4.2H_2O$, and H_2O up to 100 ml. Filter sterilize and keep at 4^oC.

YPDS (recovery medium for *P. pastoris*)

1% yeast extract, 2% peptone, 2% glucose, 1 M sorbitol

YPDS agar medium (growth of *P. pastoris* host strain)

1% yeast extract, 2% peptone, 2% glucose, 1 M sorbitol, 2% agar

YPDS + zeocin (selective medium for *P. pastoris*)

1% yeast extract, 2% peptone, 2% glucose, 100 μg/ml zeocin. Add zeocin after the solution cools down.

YPDS + zeocin (selective medium for *P. pastoris*)

1% yeast extract, 2% peptone, 2% dextrose, 1 M sorbitol, 2% agar, 100 μg/ml zeocin. Add zeocin after the solution cools down and plate. Keep plates in dark by covering with aluminum foil at 4°C. Use plates within two weeks.

B. Chemicals and reagents

APS (for SDS-PAGE analysis)

10% w/v. Dissolve 100 mg in 1 ml water and store at 4^oC. Keep it fresh.

EDTA 0.5 M, pH 8.0

Add 18.6 g of EDTA to 80 ml of distilled water and stir vigorously. Adjust the pH to 8.0 by NaOH (the EDTA will not completely dissolve until the pH is adjusted to pH 8.0.

SDS 10%

Dissolve 10 g of SDS in 100 ml distilled water and store at room temperature.

SDS-PAGE gel

- Separation gel

- 30% acrilamide 10 ml

- 2 x Tris-SDS (pH 8.8) 15 ml

- H_2O 5 ml

- 10% APS 0.3 ml

- TEMED 30 µl

- Stacking gel

- 30% acrylamide 2 ml

- 2 x Tris-SDS (pH 6.8) 6 ml

- H_2O 4 ml

- 10% APS 100 µl

- TEMED 10 µl

10 × SLiCE buffer

500 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$, 10 mM ATP, 10 mM DTT

Skim Milk

Add 5 g of Skim Milk in 100 ml TBS and dissolve by vigorous stirring

SOC (recovery medium for *E. coli*)

2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

Tetrazolium blue reagent

- Solution A:

- Tetrazolium blue (Wako) 0.2g

- 1M sodium hydroxide 10 ml

- H_2O up to 100 ml

Heat at 60°C with vigorous stirring

- Solution B:

- Sodium potassium tartrate 28.22 g

- H_2O up to 100 ml

Both solutions were run through 0.2 µm filter and then mixed together.

TF solution I (for transformation of *A. oryzae*)

1% Yatalase (TaKaRa), 0.6 M $(NH_4)_2SO_4$, 50 mM maleate buffer. Adjust to pH 5.5.

TF solution II (for transformation of *A. oryzae*)

50 mM CaCl$_2$, 35 mM NaCl, 10 mM Tris-HCl, and 1.2 M sorbitol. Adjust to pH 7.5.

TF solution III (for transformation of *A. oryzae*)

60% PEG4000, 50 mM CaCl$_2$, 10 mM Tris-HCl. Adjust to pH 7.5.

Appendixes

Abbreviation list

a.a.	amino acid
AOX	alcohol oxidase promoter
ATP	Adenosine triphosphate
BSA	bovine serum albumin
BG	β-glucosidase
^{o}C	degree Celsius
CBB	coomassie brilliant blue
CBH	cellobiohydrolase
cDNA	complementary cDNA
CMC	carboxymethyl cellulose
Da	Dalton
DEAE	diethylaminoethyl
DEPC	diethylpyrocarbonate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DW	distilled water
EDTA	ethylenediaminetetraacetic acid
EG	endoglucanase
Endo H	endoglycosidase H
FT	flow-through
g	gram
g	unit of gravity
GE	glucuronoyl esterase
HPLC	High performance liquid chromatography
LCCs	lignin-carbohydrate complexes
min	minutes
ml	milliliter
mol	mole
mRNA	messenger RNA
OD	optical density

ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNA	ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
SDS	sodium dodecyl sulfate
Tm	melting temperature
Tris	tris(hydroxymethyl) aminomethane
UV	ultra violet
v/v	volume/volume
wt	weight

HPLC chromatogram of substrates and their products in GE assay
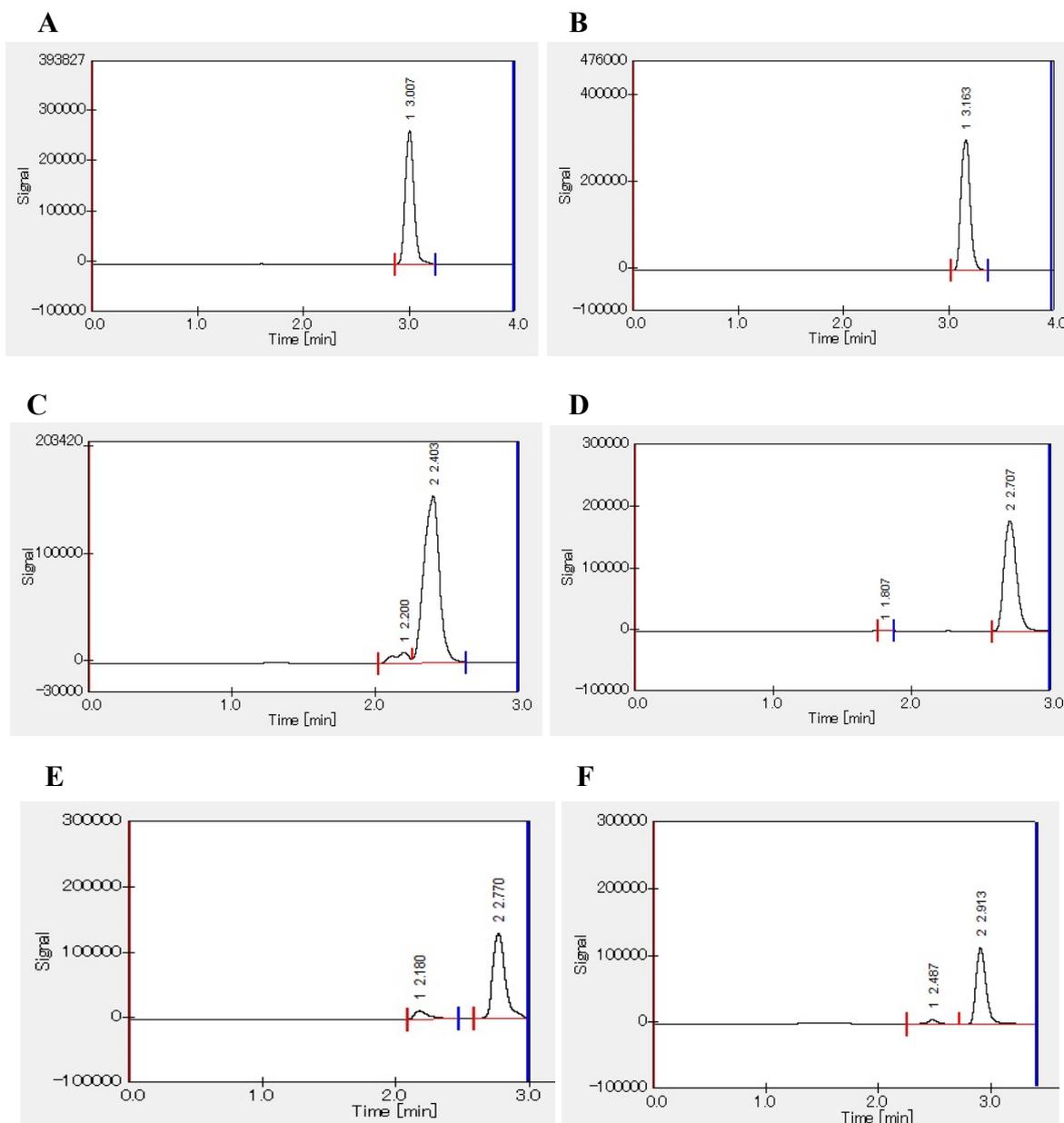
(A) substrate A (3-(4-methoxyphenyl) propyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate); (B) aromatic product of substrate A after cleaved by GE (3-(4-methoxyphenyl) propanol); (C) substrate B (benzyl D-glucuronate); substrate C, (D) substrate C (benzyl methyl α-D-glucopyranosiduronate); (E) substrate D (benzyl methyl 4-*O*-methyl-α-D-glucopyranosiduronate); and (F) aromatic product of substrate B,C, and D after cleaved by GE (benzyl alcohol).

Adav SS, Ravindran A, Sze SK (2015) Quantitative proteomic study of *Aspergillus fumigatus* secretome revealed deamidation of secretory enzymes. J Proteomics 119:154–168. doi: 10.1016/j.jprot.2015.02.007

Agger JW, Busk PK, Pilgaard B, Meyer AS, Lange L (2017) A new functional classification of glucuronoyl esterases by peptide pattern recognition. Front Microbiol 8:1–8. doi: 10.3389/fmicb.2017.00309

Bååth J, Giummarella N, Klaubauf S, Lawoko M, Olsson L (2016) A glucuronoyl esterase from *Acremonium alcalophilum* cleaves native lignin-carbohydrate ester bonds. FEBS Lett 590:2611–2618. doi: 10.1002/1873-3468.12290

Balakshin M, Capanema E, Gracz H, Chang H min, Jameel H (2011) Quantification of lignin-carbohydrate linkages with high-resolution NMR spectroscopy. Planta 233:1097–1110. doi: 10.1007/s00425-011-1359-2

Balbás P, Lorence A (2012) Recombinant Gene Expression. Humana Press, Totowa, NJ. doi: 10.1007/978-1-61779-433-9

Benoit I, Asther M, Sulzenbacher G, Record E, Marmuse L, Parsiegla G, Gimbert I, Asther M, Bignon C (2006) Respective importance of protein folding and glycosylation in the thermal stability of recombinant feruloyl esterase A. FEBS Lett 580:5815–5821. doi: 10.1016/j.febslet.2006.09.039

Benz PJ, Chau BH, Zheng D, Bauer S, Glass NL, Somerville CR (2014) A comparative systems analysis of polysaccharide-elicited responses in *Neurospora crassa* reveals carbon source-specific cellular adaptations. Mol Microbiol 91:275–299. doi: 10.1111/mmi.12459

Biely P (2016) Microbial glucuronoyl esterases: ten years after discovery. Appl Environ Microbiol 82:7014–7018. doi: 10.1128/AEM.02396-16

Biely P, Malovíková A, Uhliariková I, Li XL, Wong DWS (2015) Glucuronoyl esterases are active on the polymeric substrate methyl esterified glucuronoxylan. FEBS Lett 589:2334–2339. doi: 10.1016/j.febslet.2015.07.019

Bodîrlău R, Spiridon I, Teacă CA (2007) Chemical investigation of wood tree species in temperate forest in east-northern Romania. BioResources 2:41–57. doi: 10.15376/biores.2.1.41-57

Bond B, Hamner P (2002) Wood Identification for Hardwood and Softwood Species Native to Tennessee. Agricultural Extension Service, PB1692

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The

Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–D238. doi: 10.1093/nar/gkn663

Çelik E, Çalık P (2012) Production of recombinant proteins by yeast cells. Biotechnol Adv 30:1108–1118. doi: 10.1016/j.biotechadv.2011.09.011

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev 24:45–66. doi: S0168-6445(99)00029-7 [pii]

Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, Miyasato SR, Binkley J, Orvis J, Shah P, Wymore F, Sherlock G, Wortman JR (2014) The *Aspergillus* Genome Database : multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. 42:705–710. doi: 10.1093/nar/gkt1029

Chandel AK, Chandrasekhar G, Silva MB, Silvério da Silva S (2012) The realm of cellulases in biorefinery development. Crit Rev Biotechnol 32:187–202. doi: 10.3109/07388551.2011.595385

Charavgi MD, Dimarogona M, Topakas E, Christakopoulos P, Chrysina ED (2013) The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst. Acta Crystallogr Sect D Biol Crystallogr 69:63–73. doi: 10.1107/S0907444912042400

Chaucheyras-Durand F, Chevaux E, Martin C, Forano E (2012) Use of Yeast Probiotics in Ruminants : Effects and Mechanisms of Action on Rumen pH , Fibre Degradation , and Microbiota According to the Diet. In: Probiotic in Animals. Intech, pp 119–152

Clark SE, Muslin EH, Henson CA (2004) Effect of adding and removing *N*-glycosylation recognition sites on the thermostability of barley α-glucosidase. Protein Eng Des Sel 17:245–249. doi: 10.1093/protein/gzh028

Colot H V., Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. PNAS 103:10352–10357. doi: www.pnas.org/cgi/doi/10.1073/pnas.0601456103

d'Errico C, Börjesson J, Ding H, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN (2016) Improved biomass degradation using fungal glucuronoyl esterases-Hydrolysis of natural corn fiber substrate. J Biotechnol 219:117–123. doi: 10.1016/j.jbiotec.2015.12.024

d'Errico C, Jørgensen JO, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN (2015) Enzymatic degradation of lignin-carbohydrate complexes (LCCs): Model studies using a fungal glucuronoyl esterase from *Cerrena unicolor*. Biotechnol Bioeng 112:914–922. doi:

10.1002/bit.25508

De Santi C, Willassen NP, Williamson A (2016) Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine Arctic metagenome. PLoS One 11:1–22. doi: 10.1371/journal.pone.0159345

Di Donato P, Poli A, Taurisano V, Nicolaus B (2015) Polysaccharides from Bioagro-Waste for New Biomolecules. In: Polysaccharides. Springer International Publishing, Cham, pp 603–637

Dilokpimol A, Mäkelä MR, Cerullo G, Zhou M, Varriale S, Gidijala L, Brás JLA, Jütten P, Piechot A, Verhaert R, Faraco V, Hilden KS, de Vries RP (2017) Fungal glucuronoyl esterases: genome mining based enzyme discovery and biochemical characterization. N Biotechnol 1–6. doi: 10.1016/j.nbt.2017.10.003

Ďuranová M, Hirsch J, Kolenová K, Biely P (2009a) Fungal Glucuronoyl Esterases and Substrate Uronic Acid Recognition. Biosci Biotechnol Biochem 73:2483–2487. doi: 10.1271/bbb.90486

Ďuranová M, Špániková S, Wösten H a B, Biely P, De Vries RP (2009b) Two glucuronoyl esterases of *Phanerochaete chrysosporium*. Arch Microbiol 191:133–140. doi: 10.1007/s00203-008-0434-y

Eisentraut A (2010) Sustainable Production of Second-Generation Biofuels. OECD Publishing. doi: 10.1787/9789264084247-en

Fraňová L, Puchart V, Biely P (2016) β-Glucuronidase-coupled assays of glucuronoyl esterases. Anal Biochem 510:114–119. doi: 10.1016/j.ab.2016.07.023

Gandla ML, Derba-Maceluch M, Liu X, Gerber L, Master ER, Mellerowicz EJ, Jönsson LJ (2015) Expression of a fungal glucuronoyl esterase in *populus*: Effects on wood properties and saccharification efficiency. Phytochemistry 112:210–220. doi: 10.1016/j.phytochem.2014.06.002

Gasteiger E, Hoogland C, Gattiker a, Duvaud S, Wilkins M, Appel R, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc Handb 571–607. doi: 10.1385/1-59259-890-0:571

Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 39:270–277. doi: 10.1093/nar/gkr366

Håkansson K, Wang a H, Miller CG (2000) The structure of aspartyl dipeptidase reveals a unique fold with a Ser-His-Glu catalytic triad. Proc Natl Acad Sci U S A 97:14097–14102. doi: 10.1073/pnas.260376797

Han Y, Lei XG (1999) Role of glycosylation in the functional expression of an *Aspergillus*

niger phytase (phyA) in Pichia pastoris. Arch Biochem Biophys 364:83–90. doi: 10.1006/abbi.1999.1115

Havlik D, Brandt U, Bohle K, Fleißner A (2017) Establishment of *Neurospora crassa* as a host for heterologous protein production using a human antibody fragment as a model product. Microb Cell Fact 16:128. doi: 10.1186/s12934-017-0734-5

Hedstrom L (2002) Serine protease mechanism and specificity. Chem Rev 102:4501–4523. doi: 10.1021/cr000033x

Hildebrand A, Szewczyk E, Lin H, Kasuga T, Fan Z (2015) Engineering *Neurospora crassa* for improved cellobiose and cellobionate production. Appl Environ Microbiol 81:597–603. doi: 10.1128/AEM.02885-14

Hodásová L, Jablonsky M, Skulcova A, Haz A (2015) Lignin , potential products and their market value. WOOD Res 60:973–986.

Holtzapple M (2003) HEMICELLULOSES. In: Encyclopedia of Food Sciences and Nutrition (Second Edition). Academic Press, pp 3060–3071

Hüttner S, Klaubauf S, de Vries RP, Olsson L (2017) Characterisation of three fungal glucuronoyl esterases on glucuronic acid ester model compounds. Appl Microbiol Biotechnol 101:5301–5311. doi: 10.1007/s00253-017-8266-9

Huynh H., Arioka M (2016) Functional expression and characterization of a glucuronoyl esterase from the fungus *Neurospora crassa*: identification of novel consensus sequences. J Gen Appl Microbiol 62:217–224. doi: https://doi.org/10.2323/jgam.2016.03.004

Invitrogen (2010) Pichia expression vectors for selection on Zeocin ™ and purification of secreted, recombinant proteins. Life Technol. 48.

Jasuja N, Gauri S, Sehgal P, Bansal R (2017) Hardwood vs. Softwood. In: Diffen.com. Diffen LLC, Web. https://www.diffen.com/difference/Hardwood_vs_Softwood.

Jeffries TW (1994) Biodegradation of lignin and hemicelluloses. In: Biochemistry of microbial degradation. Kluwer Academic Publisher, pp 233–277

Katsimpouras C, Bénarouche A, Navarro D, Karpusas M, Dimarogona M, Berrin JG, Christakopoulos P, Topakas E (2014) Enzymatic synthesis of model substrates recognized by glucuronoyl esterases from *Podospora anserina* and *Myceliophthora thermophila*. Appl Microbiol Biotechnol 98:5507–5516. doi: 10.1007/s00253-014-5542-9

Ko JK, Kim Y, Ximenes E, Ladisch MR (2015) Effect of liquid hot water pretreatment severity on properties of hardwood lignin and enzymatic hydrolysis of cellulose. Biotechnol Bioeng 112:252–262. doi: 10.1002/bit.25349

Krainer FW, Dietzsch C, Hajek T, Herwig C, Spadiut O, Glieder A (2012) Recombinant protein

expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microb Cell Fact 11:22. doi: 10.1186/1475-2859-11-22

Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT, McSweeney CS (2003) Opportunities to improve fiber degradation in the rumen: Microbiology, ecology, and genomics. FEMS Microbiol Rev 27:663–693. doi: 10.1016/S0168-6445(03)00072-X

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. doi: 10.1093/bioinformatics/btm404

Latgé J-P (1999) *Aspergillus fumigatus* and *Aspergillosis*. Clin Microbiol Rev 12:310–350.

Laureano-Perez L, Teymouri F, Alizadeh H, Dale BE (2005) Understanding factors that limit enzymatic hydrolysis of biomass. Appl Biochem Biotechnol 124:1081–1099. doi: 10.1007/978-1-59259-991-2_91

Li X-L, Špániková S, de Vries RP, Biely P (2007) Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett 581:4029–4035. doi: 10.1016/j.febslet.2007.07.041

Mabashi Y, Kikuma T, Maruyama J, Arioka M (2006) Development of a Versatile Expression Plasmid Construction System for *Aspergillus oryzae* and Its Application to Visualization of Mitochondria Construction System. Biosci Biotechnol Biochem 70 (8):1882–1889. doi: 10.1271/bbb.60052

Mello BL, Polikarpov I (2014) Family 1 carbohydrate binding-modules enhance saccharification rates. AMB Express 4:36. doi: 10.1186/s13568-014-0036-9

Monlau F, Barakat A, Trably E, Dumas C, Steyer JP, Carrère H (2013) Lignocellulosic materials into biohydrogen and biomethane: Impact of structural features and pretreatment. Crit Rev Environ Sci Technol 43:260–322. doi: 10.1080/10643389.2011.604258

Motohashi K (2015) Seamless ligation cloning extract (SLiCE) method using cell lysates from laboratory *Escherichia coli* strains and its application to slip site-directed mutagenesis. Methods Mol Biol 1498:349–357. doi: 10.1007/978-1-4939-6472-7_23

Neitzel JJ (2010) Enzyme Catalysis : The Serine Proteases. Nat Educ 3(9):21.

Orman MA, Çalık P, Özdamar TH (2009) The influence of carbon sources on recombinant-human- growth-hormone production by Pichia pastoris is dependent on phenotype: a comparison of Muts and Mut+ strains. Biotechnol Appl Biochem 52:245. doi: 10.1042/BA20080057

Phillips CM, Iavarone AT, Marletta MA (2011) Quantitative proteomic approach for cellulose degradation by *Neurospora crassa*. J Proteome Res 10:4177–4185. doi:

10.1021/pr200329b

Pokkuluri PR, Duke NEC, Wood SJ, Cotta M a., Li XL, Biely P, Schiffer M (2011) Structure of the catalytic domain of glucuronoyl esterase Cip2 from *Hypocrea jecorina*. Proteins Struct Funct Bioinforma 79:2588–2592. doi: 10.1002/prot.23088

Rao M, Mishra C, Keskar S, Srinivasan MC (1985) Production of ethanol from wood and agricultural residues by *Neurospora crassa*. Enzyme Microb Technol 7:625–628. doi: 10.1016/0141-0229(85)90033-X

REN21 (2017) Renewables 2017: global status report. Renewable and Sustainable Energy Reviews 1-302. doi: 10.1016/j.rser.2016.09.082

Rogowski A, Baslé A, Farinas CS, Solovyova A, Mortimer JC, Dupree P, Gilbert HJ, Bolam DN (2014) Evidence that GH115 α-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. J Biol Chem 289:53–64. doi: 10.1074/jbc.M113.525295

Rubin EM (2008) Genomics of cellulosic biofuels. Nature 454:841–845. doi: 10.1038/nature07190

Sárossy Z, Plackett D, Egsgaard H (2012) Carbohydrate analysis of hemicelluloses by gas chromatography-mass spectrometry of acetylated methyl glycosides. Anal Bioanal Chem 403:1923–1930. doi: 10.1007/s00216-012-6038-z

Sasagawa T, Matsui M, Kobayashi Y, Otagiri M, Moriya S, Sakamoto Y, Ito Y, Lee CC, Kitamoto K, Arioka M (2011) High-throughput recombinant gene expression systems in *Pichia pastoris* using newly developed plasmid vectors. Plasmid 65:65–69. doi: 10.1016/j.plasmid.2010.08.004

Simon GM, Cravatt BF (2010) Activity-based proteomics of enzyme superfamilies: Serine hydrolases as a case study. J Biol Chem 285:11051–11055. doi: 10.1074/jbc.R109.097600

Sims R, Taylor M, Jack S, Mabee W (2008) From 1st to 2nd Generation Bio Fuel Technologies: An overview of current industry and RD&D activities. IEA Bioenergy 1–124. doi: 10.1128/AAC.03728-14

Špániková S, Biely P (2006) Glucuronoyl esterase – Novel carbohydrate esterase produced by *Schizophyllum commune.* FEBS Lett 580:4597–4601. doi: 10.1016/j.febslet.2006.07.033

Špániková S, Poláková M, Joniak D, Hirsch J, Biely P (2007) Synthetic esters recognized by glucuronoyl esterase from *Schizophyllum commune.* Arch Microbiol 188:185–189. doi: 10.1007/s00203-007-0241-x

Sunner H, Charavgi M-D, Olsson L, Topakas E, Christakopoulos P (2015) Glucuronoyl Esterase Screening and Characterization Assays Utilizing Commercially Available Benzyl

Glucuronic Acid Ester. Molecules 20:17807–17817. doi: 10.3390/molecules201017807

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi: 10.1093/molbev/mst197

Topakas E, Moukouli M, Dimarogona M, Vafiadi C, Christakopoulos P (2010) Functional expression of a thermophilic glucuronyl esterase from *Sporotrichum thermophile:* identification of the nucleophilic serine. Appl Microbiol Biotechnol 87:1765–72. doi: 10.1007/s00253-010-2655-7

Tsai AYL, Canam T, Gorzsás A, Mellerowicz EJ, Campbell MM, Master ER (2012) Constitutive expression of a fungal glucuronoyl esterase in *Arabidopsis* reveals altered cell wall composition and structure. Plant Biotechnol J 10:1077–1087. doi: 10.1111/j.1467-7652.2012.00735.x

Uchima CA, Tokuda G, Watanabe H, Kitamoto K, Arioka M (2011) Heterologous expression and characterization of a glucose-stimulated β-glucosidase from the termite *Neotermes koshunensis* in *Aspergillus oryzae*. Appl Microbiol Biotechnol 89:1761–1771. doi: 10.1007/s00253-010-2963-y

Vafiadi C, Topakas E, Biely P, Christakopoulos P (2009) Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. FEMS Microbiol Lett 296:178–184. doi: 10.1111/j.1574-6968.2009.01631.x

Wahab R, Gigantochloa B, Levis G, Scortechinii G, Wrayi G, Mustafa MT, Sudin M, Mohamed A, Rahman S, Samsi HW, Khalid I (2013) Extractives, Holocellulose, α-Cellulose, Lignin and Ash Contents in Cultivated Tropical. Curr Res J Biol Sci 5:266–272.

Wang B, Cai P, Sun W, Li J, Tian C, Ma Y (2015) A transcriptomic analysis of *Neurospora crassa* using five major crop residues and the novel role of the sporulation regulator rca-1 in lignocellulase production. Biotechnol Biofuels 8:21. doi: 10.1186/s13068-015-0208-0

Watanabe H, Tokuda G (2010) Cellulolytic Systems in Insects. Annu Rev Entomol 55:609–632. doi: 10.1146/annurev-ento-112408-085319

Watanabe T (1995) Important Properties of Lignin-Carbohydrates Complexes (LCCs) in Environmentally Safe Paper Making. Trends Glycosci Glycotechnol 7:57–68.

Wong DWS, Chan VJ, McCormack A a, Hirsch J, Biely P (2012) Functional Cloning and Expression of the *Schizophyllum commune* Glucuronoyl Esterase Gene and Characterization of the Recombinant Enzyme. Biotechnol Res Int 2012:951267. doi:

10.1155/2012/951267

Yuan TQ, Sun SN, Xu F, Sun RC (2011) Characterization of lignin structures and lignin-carbohydrate complex (LCC) linkages by quantitative 13C and 2D HSQC NMR spectroscopy. J Agric Food Chem 59:10604–10614. doi: 10.1021/jf2031549

Zheng Y, Pan Z, Zhang R, Wang D, Jenkins B (2008) Non-ionic surfactants and non-catalytic protein treatment on enzymatic hydrolysis of pretreated creeping wild ryegrass. Appl Biochem Biotechnol 146:231–248. doi: 10.1007/s12010-007-8035-9

Znameroski EA, Coradetti ST, Roche CM, Tsai JC, Iavarone AT, Cate JHD, Glass NL (2012) Induction of lignocellulose-degrading enzymes in *Neurospora crassa* by cellodextrins. Proc Natl Acad Sci 109:6012–6017. doi: 10.1073/pnas.1118440109

論文の内容の要旨

応用生命工学専攻
平成27年度博士課程入学
氏 名　HUYNH Hiep Hung
指導教員名　吉田 稔

論文題目

Studies on the enzymatic properties and physiological functions of fungal glucuronoyl esterases

（糸状菌グルクロン酸エステラーゼの酵素学的性質と生理機能に関する研究）

Introduction

Plant biomass is a complex mixture of cellulose, hemicellulose, and lignin. Cellulose and hemicellulose, and even lignin are considered to be the most abundant organic polymers on earth, and thus they are expected to be highly potential sources for renewable energy. However, these components are weaved in a complex structure which greatly contributes to the rigidity of plant cell walls, and thus hampers the enzymatic degradation. One of the intimate covalent linkages is found between the aromatic alcohol of lignin and the 4-O-methyl-D-glucuronic acid residue of hemicellulose. The enzyme glucuronoyl esterase (GE) that constitutes a novel family, CE15, of carbohydrate esterases cleaves this ester linkage and is thus expected to play a key role in the degradation/utilization of lignocellulose. In this study, I identified, expressed, purified, and characterized two novel GEs from filamentous fungi, NcGE from *Neurospora crassa* and AfGE from *Aspergillus fumigatus*. In the course of characterization, I found that both GEs display preference for the substrate containing the methyl moiety at the O4 position of glucuronic acid. Although similar behavior was previously reported for other fungal GEs, and glucuronic acid is commonly modified by the 4-O-methyl group in lignocellulose, the mechanism of this substrate preference has remained elusive. Herein, molecular modeling and site-directed mutagenesis analyses demonstrated that Lys209 in AfGE plays an important role in this substrate preference. Besides, the physiological role of GE was investigated by growing the wild-type and NcGE deletion strains of *N. crassa* in the medium containing wood powder extract. The results showed that GE contributes to the degradation/utilization of lignocellulose *in vivo*.

Chapter 1. Heterologous expression of NcGE and AfGE

The expression system of the methylotrophic yeast *Pichia pastoris* was employed to express NcGE. The culture supernatant of NcGE transformant degraded the ester bond of the synthetic substrates

that mimic the linkage between hemicellulose and lignin, demonstrating that *Nc*GE is a functional GE. The alignment of *Nc*GE and other characterized fungal GEs lead to the identification of novel consensus sequences surrounding the catalytic triad, VTGC<u>S</u>RXGKGA, <u>H</u>C, and PQ<u>E</u>SG (catalytic residues are underlined). The positions of six Cys residues were also conserved, which would bring the catalytic residues close together by forming three disulfide bonds. These features were also found in another uncharacterized GE, *Af*GE. Because of the presence of an intron, *Af*GE was heterologously expressed using the expression system of *Aspergillus oryzae*. As expected, *Af*GE produced in the culture supernatant exhibited the activity as a typical GE. Together with *Nc*GE, the enzymatic characteristics of *Af*GE were analyzed in the subsequent Chapters.

Chapter 2. Purification and characterization of *Nc*GE and *Af*GE

Since both fungal GEs contain the hexahistidine-tag at the C-terminus, Ni^{2+}-NTA affinity chromatography was performed. Higher purity was achieved by ion exchange chromatography. With the pure proteins obtained, the enzymatic characteristics were studied. GE is considered as a serine-type esterase; in agreement, both *Nc*GE and *Af*GE were strongly inhibited by PMSF. Among other chemicals tested, Tween 80 was found to significantly enhance the activity of both enzymes. While the optimal pH for *Nc*GE was neutral, *Af*GE preferred more acidic condition of pH 5. *Nc*GE behaved as a mesophilic enzyme, showing the activity and stability at around 30°C. In contrast, *Af*GE was more thermostable, maintaining more than 90% and 70% of its maximum activity at 50°C and 60°C, respectively. Regarding the kinetic parameters, *Af*GE displayed high $k$cat values and catalytic efficiencies toward the synthetic substrates. Since the enzymatic hydrolysis of biomass in industries is most commonly conducted at around 50°C and pH 5, *Af*GE fulfilled these requirements and thus expected to be industrially applicable.

Chapter 3. The role of conserved Lys residue in the recognition of 4-*O*-methyl group in the substrate

Glucuronoxylan is one of the primary components of hemicellulose in which the backbone polymer, β-(1,4)-linked xylose (xylan), is attached by glucuronic acid (GlcA) and 4-*O*-methyl-D-glucuronic acid (MeGlcA). Generally, around 10-20% of xylose are substituted by MeGlcA, which is also esterified to lignin alcohol at a high ratio such that around 30% and 40% are esterified to lignin in beechwood and birchwood, respectively. In nature, the presence of these lignin-carbohydrate complexes (LCCs) is considered as the major cause of recalcitrance of lignocellulose. It has been known that the synthetic esters of MeGlcA are much more preferred by fungal GEs than those of GlcA, and two fungal GEs characterized in this study also exhibited this feature. However, the reason for this remains unknown. Aiming to elucidate the molecular basis of this feature, the modeling analysis of *Af*GE was conducted, especially to elucidate the interaction of 4-*O*-methyl group in the substrate and the active site of *Af*GE. The conserved Lys209 was suggested to be involved in the recognition of 4-*O*-methyl group due to its proximity. The analysis of model interfaces by PDBePISA also pointed to this

interaction. Then, site-directed mutagenesis was performed to introduce four different mutations at Lys209 in *Af*GE. Wild-type and K209R mutant displayed a clearly higher activity to the substrate D containing 4-*O*-methyl substituent than to its counterpart, substrate C, lacking this modification. Other mutants (K209A, K209E, and K209Q) did not show any difference in the preference between these substrates. These results indicate that the positive charge of Lys209 could contribute to the interaction of GEs and the 4-*O*-methyl group of the substrates.
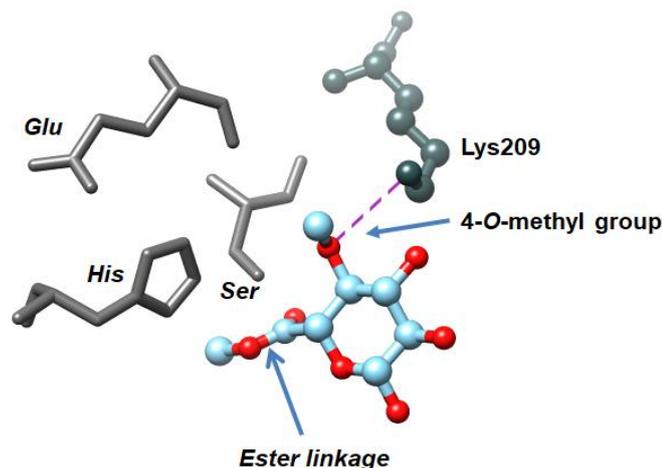
Modeled structure of *Af*GE catalytic domain

The catalytic triad Ser-His-Glu which stands in front of the target ester linkage is shown on the left. The interaction between Lys209 and 4-*O*-methyl group is indicated by the dashed line.

Chapter 4. Physiological functions of GE

From the first discovery in 2006, GEs have been suggested to facilitate the enzymatic hydrolysis of plant cell walls; however, this statement has just been confirmed recently. In the degradation of corn fiber using the commercial cellulase mixture, the amount of released reducing sugar was increased by adding GEs. Later, the direct evidence of the action of GEs on the breakdown of the ester linkage in LCCs was demonstrated by using NMR spectroscopy. In this study, for the first time, I attempted to elucidate the physiological role of GE in its host microorganism. The wild-type (WT) and *Nc*GE deletion (KO) strains of *N. crassa* were cultured in the medium containing LCCs extracted from the wood chips as the sole carbon source in different conditions. The amount of reducing sugars liberated in the culture supernatant due to the degradation of plant biomass by endogenous enzymes was used as a hallmark of cell growth. I found that the KO strain produced less amount of reducing sugars compared to WT. When purified *Nc*GE, but not heat-inactivated enzyme, was added to the culture of KO strain, the amount of reducing sugars recovered, suggesting that exogenously-added *Nc*GE assisted the actions of endogenous cellulases/hemicellulases. Similar enhancement of reducing sugar production was observed when the *Nc*GE-pretreated medium was used to grow the KO strain; it is of note that *Nc*GE itself did not increase the amount of reducing sugars during the pretreatment. RT-PCR analysis demonstrated that the transcript of *Nc*GE was only detected in WT, and the expression level of three

major cellulases CBH-1 (cellobiohydrolase), GH 5.1 (endoglucanase), and GH 3.4 (β-glucosidase) was similar in WT and KO strains. Thus both endogenous and exogenously-added *Nc*GE contributed to the enhanced production of reducing sugars. The absence/addition of *Nc*GE did not affect the production of reducing sugars when the cells were grown in carboxymethyl cellulose or Avicel (crystalline cellulose), consistent with the lack of target ester linkage of GE in these polysaccharides. Taken together, it was concluded that the cooperation between GE and other carbohydrate-acting enzymes is necessary for the efficient degradation of lignocellulosic biomass *in vivo*.

Conclusion

The increasing demand for finding the renewable energy sources lead to higher interest in the biomass degradation. Genome mining of microorganisms resulted in the finding of a variety of enzymes that hydrolyze cellulose and hemicellulose. However, this approach surely has a limitation since the complete deconstruction of plant biomass requires the cleavage of all the linkages inside lignocellulose. Therefore, much attention has recently been paid to the roles of GEs. In this study, two fungal GEs, *Nc*GE and *Af*GE, were analyzed. Through their characterization, I found that *Af*GE displays the properties suited for its biotechnological application, such as its activity/stability at pH 5 and 50ºC. Structure prediction together with site-directed mutagenesis analysis elucidated the role of Lys209 in the recognition of 4-*O*-methyl group in the glucopyranose ring of the substrate. Finally, the role of GE *in vivo* was demonstrated for the first time by growing WT and *Nc*GE KO strains of *N. crassa* in the medium containing lignocellulose. Additional study is surely needed to understand more clearly their physiological roles and the cooperation between GEs and other wood-decaying enzymes, which would be an interesting task for improving the efficiency of hydrolysis of lignocellulose and achieving the ultimate goal---complete degradation of biomass.

Publications

1. Huynh, H.H. and Arioka, M. (2016) Functional expression and characterization of a glucuronoyl esterase from the fungus *Neurospora crassa*: identification of novel consensus sequences. J. Gen. Appl. Microbiol. <u>62</u>, 217–224. doi: https://doi.org/10.2323/jgam.2016.03.004

2. Huynh, H.H., Ishii, N., Matsuo, I., and Arioka, M. A novel glucuronoyl esterase from *Aspergillus fumigatus* - the role of conserved Lys residue in the preference for 4-*O*-methyl glucuronoyl esters. Appl. Microbiol. Biotechnol., in press.

Acknowledgement