Ph.D. Thesis

Studies on characterization of a novel hemicellulase Abf62A-Axe6A from *Ruminiclostridium josui* and development of its host-vector system

(*Ruminiclostridium josui* の新規ヘミセルラーゼ Abf62A-Axe6A の酵素特性と宿主・ベクター系の開発に関する研究)

Graduate School of Bioresources
Mie University

Yayun Wang

September 2018

Abstract

From the viewpoint of reduction of the global warming, the utilization and application of the renewable energies are required. Among various renewable energies, plant biomass is one of the most important energy resources. In the current process of bioconversion of plant biomass to value-added products, plant biomass is hydrolyzed by cellulolytic enzymes from fungi such as *Trichoderma reesei* and the resultant fermentable monosaccharides are fermented to biofuels etc. by yeasts or bacteria. On the other hand, "consolidated bioprocessing (CBP)" is widely recognized as the efficient strategy for simultaneous hydrolysis and fermentation of biomass using cellulolytic anaerobic bacteria. A novel hemicellulase Abf62A-Axe6A from *Ruminiclostridium josui* was biochemically characterized and its host-vector system was established for better use of this bacterium for CBP.

R. josui Abf62A-Axe6A is a modular enzyme consisting of an N-terminal signal peptide, a catalytic module of glycoside hydrolase family 62 (GH62), a family 6 carbohydrate binding module (CBM6), a dockerin module and another catalytic module of carbohydrate esterase family 6 (CE6) in order from the N-terminus. Therefore, three recombinant enzymes, RjAbf62A-Axe6A consisting of 4 modules devoid of the signal peptide, RjAbf62A-CBM6 consisting of GH62 and CBM6 and RjAxe6 consisting of CE6 only, were produced by *Escherichia coli* recombinants and biochemically characterized. RjAbf62A-Axe6 was highly active toward arabinoxylan and moderately active toward sugar beet arabinan, releasing mainly arabinose. Analysis of the hydrolysis products of arabinoxylooligosaccharides indicated that RjAbf62A-Axe6 released not only α-1,2- and α-

1,3-linked arabinofuranoses from singly substituted xylosyl residues but also from doubly substituted xylosyl residues. RjAbf62A-Axe6A efficiently hydrolyzed 3^2-α-L-arabinofuranosyl xylobiose (A^3X) to release arabinose, and the activity toward A^3X was stronger than toward 3^3-α-L-arabinofuranosyl xylotetraose (XA^3XX), 2^3-α-L-arabinofuranosyl xylotriose (A^2XX) and (2^3,3^3-di-α-L-arabinofuranosyl xylotriose) $A^{2,3}XX$. RjAbf62A-Axe6A seemed to prefer Ara*f* of substituted xylosyl residues located at non-reducing end and the 1,3-linked Araf residues in doubly substituted xylosyl residues. The enzyme showed a weak activity toward linear 1,5-α-L-arabinan and arabinooligosaccharides, indicating that it had an exo-α-1,5-L-arabinofuranosidase activity. Surprisingly, RjAbf62A-Axe6 demonstrated an endoxylanase activity toward birchwood and beechwood xylans and xylooligosaccharides. RjAbf62A-Axe6 is the first GH62 enzyme having arabinofuranosidase and endoxylanase activities. Both RjAbf62A-Axe6 and RjAxe6 had an acetylxylan esterase activity toward insoluble wheat arabinoxylan.

R. josui host-vector system has not yet been established because of the existence of a restriction and modification system in *R. josui*. A restriction endonuclease *Rjo*I was purified from *R. josui* cell extract using column chromatography and characterized. The results showed that *Rjo*I is an isoschizomer of *Dpn*I, recognizing the sequence 5'-G^{met}ATC-3', where the A nucleotide is Dam-methylated. *Rjo*I cleaved the recognition sequence between the A and T nucleotides, producing blunt ends. Plasmids prepared from *E. coli* C2925 (*dam*$^-$/*dcm*$^-$) could be introduced into *R. josui* by electroporation. The highest transformation efficiency of 6.6×10^3 transformants/μg of DNA was obtained using a square wave pulse (750 V, 1 ms). When the *R. josui cel48A* gene, devoid of the dockerin encoding region, cloned into newly developed plasmid pKKM801 was introduced into *R.*

josui, a truncated form of Cel48A, RjCel48AΔdoc, was detected in the culture supernatant but not in the intracellular fraction. This is the first report on the establishment of the fundamental technology for molecular breeding of *R. josui*.

Abbreviations

$A^{2,3}XX$: 2^3,3^3-di-α-L-arabinofuranosyl xylotriose

A^2XX: 2^3-α-L-arabinofuranosyl xylotriose

A^3X: 3^2-α-L-arabinofuranosyl xylobiose

ABF: α-L-arabinofuranosidase

Ara*f*: Arabinofuranosyl

AXE: Acetylxylan esterase

AXH: Arabinoxylan arabinofuranohydrolase

CAZy: Carbohydrate-active enzymes

CBB: Coomassie brilliant blue

CBH: Cellobiohydrolase

CBM: Carbohydrate binding module

CBP: Consolidated bioprocessing

CE: Carbohydrate esterase

Cip: Cellulosome-integrating protein

DTT: Dithiothreitol

EG: Endoglucanases

GH: Glycoside hydrolase

HPLC-PAD: High-performance anion-exchange chromatography with a pulsed amperometric detection

LMW: Low molecular weight

MPN: Most probable number

MTase: Methyltransferase

qRT-PCR: Quantitative real-time polymerase chain reaction

REase: Restricion endonuclease

RM: Restriction modification

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SLH: Surface layer homology

TLC: Thin-layer chromatography

XA^3XX: 3^3-α-L-arabinofuranosyl xylotetraose

Contents

Chapter 1: General introduction

1.1 Background

In recent years, with the rapid economic growth in the developing countries, the demand for global energy is rapidly increasing. According to the International Energy Outlook 2016 (1), the global energy consumption in 2000 was about 10.042 billion tons of petroleum, and the energy consumption in 2014 was 13.684 billion tons of petroleum. Furthermore, the global energy consumption in 2040 is expected to increase to 1.78 times higher than in 2000, and it is 17.816 billion tons of petroleum (1). With the rapid economic development of non-Organization for Economic Cooperation and Development (OECD) members including China, India, Middle East countries, African countries and Brazil, the trend of energy demand growth cannot be eased in the short term. With the global consumption of large amounts of fossil fuels, global carbon dioxide emissions reached 35.842 billion tons in 2013, and the average concentration of carbon dioxide in the global atmosphere rose to 395.7 ppm (2). As a result, such global warming and air pollution are threatening the global environment and human health. There are also problems in the exhaustion of fossil fuels, the sharp fluctuations in the prices of petroleum fuels, and the safety issues in the development of nuclear energy, all of which make us try desperately to develop alternative energy sources urgently.

Renewable energy development and utilization is an effective means to deal with global environmental and climate change, while also maintaining a secure energy supply. Renewable energy mainly includes solar, wind, bioenergy, ocean power, geothermal and hydropower (3). The U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA) strongly support research and development of biofuels and bioproducts: first, in order to promote the development of biofuels, alternatives to

petroleum and gas; secondly, to promote the development of agriculture, forestry and rural economy; thirdly, to cultivate new industries, such as manufactures of various biofuels, valuable chemicals and other products. In the 2005 plan report, the United States was expected to use biofuels to replace the current U.S. petroleum consumption by 2030 (4). For new bioenergy research and development, DOE administers four research centers, namely, Great Lakes Bioenergy Research Center, Center for Bioenergy Innovation, Joint BioEnergy Institute, and Center for Advanced Bioenergy and Bioproducts Innovation (https://genomicscience.energy.gov/centers/index.shtml). In Basic Plan for the Promotion of Biomass Utilization by the Ministry of Agriculture, Forestry and Fisheries of Japan (September 2016), it was stated that the utilization of biomass by 2025 will reach about 26 million tons per year in the equivalent of carbon, and the biomass industry plan will reach 500 billion yen (http://www.maff.go.jp/j/shokusan/biomass/attach/pdf/index-4.pdf).

Bioenergy is a generic name of natural renewable energy produced from biomass resources. Biomass refers to all organic matters derived from plants (including algae, trees and crops), microbes, animals, and their organic waste products, where green plants convert carbon dioxide into plant materials using solar energy through photosynthesis. Plant biomass contains a large number of sugar polymers, including starch, cellulose and hemicellulose, which can be decomposed into simple sugars by enzyme mixtures produced by microorganisms and then fermented into value-added products such as bioethanol (5). Among plant biomass, lignocellulose in plant cell wall that does not compete with foodstuffs are the most suitable biological resources for bioenergy.

1.2 Structure of plant cell wall

Lignocellulose is a kind of composite material synthesized by plant cells, mainly composed of polymeric carbohydrates (cellulose and hemicellulose) and aromatic polymer lignin (Fig. 1), followed by a small amount of pectins, inorganic compounds, proteins, waxes and lipids. The content of each component of lignocellulose depends on plant species, tissues, and growth conditions, and plant biomass can be roughly classified into woody (softwood and hardwood) and non-woody (agricultural) biomass according to its physical properties and chemical composition (6).

Cellulose is the simplest polymer among polymeric carbohydrates and a single strand (or molecule) of cellulose has the linear polymeric structure consisting of β-1,4-linked D-glucose units. In natural cellulose, three hydrogen bonds occur per glucosyl unit, in which two intramolecular hydrogen bonds and one intermolecular hydrogen bond are formed to adjacent glycosyl units on the same or neighbor cellulose molecules (7). After a linear cellulose strand is bound to each other through hydrogen bonding, further hydrogen bonding and van der Waals interactions between cellulose chains make them assemble into microfibril up to 20 nanometers in width that are insoluble in water and most organic solvents (8). The matrix structure bundle of microfibrils forms the primary cellulose fibers in the cell wall.

Hemicellulose polymers have more structural diversity, and their classification is usually based on the constituent saccharides consisting of the β-1,4-linked main chains with an equatorial configuration, including xylans, xyloglucans, mannans, glucomannans, and β-1,3-1,4-glucans. The main chain of xylans is formed by polymerization of D-xylose units, while mannans have a backbone consisting of D-mannose and xyloglucans contain D-

glucose as the main constituent (9). The main chains of hemicelluloses tend to be modified with functional substituents such as D-xylose, D-galactose, L-arabinose, D-glucuronic acid, acetic acid and methanol, which enhance the diversity of their structures. Among them, side chains of xylans are particularly diverse, and the diversity of their side chains is also reflected in different parts of the same plant. Their branches are often replaced by different carbohydrate and acid substituents. In addition to the softwood xylan that are heavily acetylated, xylans from the other plants are acetylated to a greater or lesser extent, and in some cases also substituted with phenolic group (10). Structure of xylans is described more in details in the introduction section of Chapter 2.

Lignin is an exceptionally complex, water-insoluble polymer that does not have a definite primary structure and is formed by polymerization of different aromatic alcohols. It provides water-proofing, structural reinforcement and resilience with plant cell walls, giving their resistance to the biological and physical attacks. The composition of lignin is different among softwood, hardwood and grasses. Softwood lignin is composed almost exclusively of guaiacyl units. On the other hand, hardwood lignin contains a large amount of syringyl units as well as guaiacyl units and grass lignin also contain a small amount of *p*-hydroxyphenyl groups (11). These difference in the composition has a significant influence on the delignification chemistry and thus on the biomass deconstruction.

Although the structure of cellulose is simple, its hydrophobic characteristics and tight and hierarchical structure make it resistant to cellulolytic enzymes by preventing them from penetrating it and exerting their full activity. Hemicelluloses play a structural role between cellulose and lignin and are tightly packed on the surface of cellulose, leading to interference of cellulose decomposition caused by cellulolytic enzymes. Besides, because

they have a wide variety of branches, many accessory enzymes are required for their complete degradation. Ultimately, the insolubility of lignin and the diversity of lignocellulose structural units all together play a significant role in inhibiting the efficient degradation of lignocellulose.

1.3 Enzymatic lignocellulose degradation

To depolymerize lignocellulose into carbohydrates for production of biofuels and valuable chemical, environmentally friendly strategies are mostly being used at present, that is, a combination of thermochemical pretreatment and enzymatic hydrolysis (12). Ideal pretreatment can promote the contact of the enzyme with the hydrophobic surface of cellulose to enhance its degradation (13).

Two primary types of microbial cellulolytic systems are well known: the “free” enzyme type and the “cellulosome” type. The mechanism of the deconstruction of the plant cell wall by these two different cellulose degradation systems is significantly different from each other.

In the free enzyme type, each enzyme diffuses as a single catalytic unit, which is usually composed of a catalytic module and a carbohydrate binding module (CBM), covalently linked to each other via a linker sequence, as exemplified by *Trichoderma reesei* (or *Hypocrea jecorina*) enzymes. The catalytic module is mainly a glycoside hydrolase (GH) module, which catalyzes the hydrolysis of glycosidic bonds in complex sugars. Cellulase and hemicellulase are important family members of GHs that typically catalyze the hydrolysis of various kinds of oligosaccharides and polysaccharides in biomass

degradation (14). Non-catalytic CBMs play a crucial role in identifying plant cell wall components and enhancing enzyme activity, that is, CBMs in cellulases and hemicellulases enhance the hydrolytic activity of adjacent catalytic domains by directing the catalytic module to appropriate and readily degradable substrates (15). For more details of GH and CBM, see the website on Carbohydrate-Active enZymes – CAZy (http://www.cazy.org/).

The *T. reesei* system has about 70 years of research history and is the representative of fungal cellulolytic systems (16, 17). In the fungal systems, four enzymes with different modes of action are involved in cellulolysis, that is, cellulase (endo-1,4-β-D-glucanase, often abbreviated as endoglucanase; EC 3.2.1.4), cellulose 1,4-β-cellobiosidase (non-reducing end) (β-1,4-glucan cellobiohydrolase, abbreviated as cellobiohydrolase; EC 3.2.1.91), cellulose 1,4-β-cellobiosidase (reducing end) (abbreviated as cellobiohydrolase; EC 3.2.1.176) and glucan 1,4-β-glucosidase (β-glucosidase; EC 3.2.1.74). Endoglucanase acts on cellulose by cleaving cellulose chains in the amorphous region to supply cellobiohydrolase with new substrates. Cellobiohydrolases attack non-reducing end or reducing end of a cellulose chain in both crystalline and amorphous regions and hydrolyze cellulose chains in a processive manner to exclusively release cellobiose. Cellobiose is hydrolyzed by β-glucosidase to glucose.

About 35 years ago, the cellulosome was first isolated from *Ruminiclostridium thermocellum* (formerly *Clostridium thermocellum*), characterized and defined as multienzyme cellulolytic complex (18-20). After that, many anaerobic bacteria were found to produce similar cellulolytic complexes, cellulosomes. The *R. thermocellum* cellulosome can not only effectively degrade various polysaccharides of plant cell walls but also exhibits strong activity on crystalline cellulose (21). In addition to endoglucanase and

cellobiohydrolase, most cellulosomes were reported to exhibit catalytic activities for the deconstruction of lignocellulosic biomass, including xylanase, arabinofuranosidase, acetyl xylan esterase, pectate lyase, and mannanase (22-25). Cellulosome is mainly composed of two types of subunits: enzymes that contain a dockerin module as well as a catalytic module and/or other accessory modules, and a structural scaffolding protein, called scaffoldin, which contain a number of cohesin modules and a CBM (14). The specific interaction between cohesins of a scaffoldin and dockerins of enzyme subunits determines the assembly of the enzymes and the scaffoldin, resulting in the formation of the cellulosome structure. Some scaffoldins contain an additional dockerin module for interacting with a cohesin of other scaffoldins or surface layer proteins. Besides, the CBM on the scaffoldin guides and binds the cellulosome to the plant cell wall (26). Since many cellulosomal enzymes also have one or more CBM(s) belonging to the different families (27-29), the cellulosome should have an affinity for polysaccharides other than cellulose through CBMs present in enzymic subunits. The cellulosome can be anchored onto the bacterial cell surface through its interaction with surface layer proteins containing surface layer homology (SLH) modules (30), and some cellulosomes are free, called cell-free cellulosome (31).

All cellulolytic enzymes and hemicellulolytic enzymes, regardless of their origins, contain at least a catalytic module, which has different catalytic activities such as glycoside hydrolase (GH), carbohydrate esterase (CE), polysaccharide lyase (PL) and so on. GHs such as cellulases and xylanases are classified into GH families on the basis of the amino acid sequence similarity in the CAZy database (http://www.cazy.org/) (32) . CEs such as acetyl xylan esterase and feruloyl esterase are also classified into CE families. Many of

these enzymes are modular enzymes consisting of at least two modules, catalytic module and dockerin module, and additional modules such as CBM and SLH module. CBMs are also classified into families based on amino acid sequence similarity.

Among two types of microbial cellulolytic systems, the free enzyme type represented by *Trichoderma reesei* is characterized by relatively simple cellulase composition and high protein productivity, while the cellulosome type is featured with high specific activity on crystalline cellulose and quite low protein productivity due to the anaerobic bacterial origin. Because of high protein productivity, fungal cellulase preparations were commercialized for biomass degradation and are now widely being used in the first process of bioconversion of biomass into biofuels, biomass decomposition process.

As for protein productivity, anaerobic cellulosome systems are far inferior to aerobic fungal systems. However, cellulosome systems are quite interesting from the point of view of academic research, e.g., the mechanism of cellulosome assembly and synergism of a large number of cellulosomal enzymes toward various polysaccharides. Furthermore, anaerobic fermentative bacteria having the cellulosome system are promising candidates for consolidated bioprocessing (CBP) (33), which is defined as the conversion of lignocellulose into biofuels in one step without an addition of enzymes, that is, production of cellulolytic and hemicellulolytic enzymes, decomposition of polysaccharides, and fermentation of resultant hexoses and pentoses are carried out by a lignocellulolytic and fermentative microorganism in one bioreactor.

1.4 Restriction enzymes

As early as 1950, with discovery of the host-controlled bacterial virus variations as momentum (34), the restriction endonuclease (REase) and methyltransferase (MTase) activities of Type I (35), Type II (36), Type III (37), and Type IV (38) DNA restriction-modification (RM) systems were gradually discovered. Over the next half-century, a large number of REases were discovered and investigated (39), and Type II REases have become an indispensable tool in modern molecular biology (40). Since the discovery of the first batch of REases *Eco*RI and *Eco*RV, more than 14,000 REases have been assumed so far (http://rebase.neb.com/rebase/rebase.html). Type II REases have already been reported from more than 350 species, each of which has its unique biochemical properties. Type II REases are produced by prokaryotes to fight bacteriophages and foreign nucleic acid sequences to protect their nucleic acid sequences. They all recognize a particular sequence in double-stranded DNA and cleaves internal or nearby fixed sequences position. This type of REase differs from the other types in that they form homodimers, and the recognition sites are palindromic, undivided and 4-8 nucleotides in length (41). To overcome the bacterial restriction systems allows the development of genetic engineering tools for bacteria of interest. (42-45). Therefore, researches on the restriction-modification systems of the bacteria are indispensable to develop the genetic manipulation strategies for them.

1.5 Host-vector systems of *Clostridium* species and related bacteria

Detailed characterization of bacterial growth, metabolism, and physiology is necessary to develop bacterial genetic engineering systems. There have been many reports on genetic engineering strategies for some *Clostridium* and *Ruminiclostridium* species. In

order to develop a genetic engineering tool in a *Clostridium* species, a suitable *Escherichia coli-Clostridium* shuttle vector needs to be selected. For example, *E. coli-Clostridium perfingens* shuttle vector pJIR751 was successfully used for the development of *C. paraputrificum* gene transfer system (46). For the gene expression in *R. thermocellum*, the promoters in the upstream regions of the three genes *gapDH*, *eno*, and *cipA* from *R. thermocellum* are most commonly used (47). The *gapDH* gene encoding glyceraldehyde-3-phosphate dehydrogenase is one of the highest expressed proteins in the proteome (48). The *eno* promoter was used to drive the expression of exogenous pyruvate kinase from *Thermoanaerobacterium saccharolyticum* (49). The *cipA* promoter was used to drive the expression of the scaffolding protein CipA in the *cipA* deletion strain of *R. thermocellum*. Tests of 17 promoters from *R. thermocellum* revealed that in addition to the four commonly used promoters, *gapDH*, *eno*, *cbp*, and *cipA*, for gene expression, additional four promoters of clo1313_2638, clo1313_2926, clo1313_966, and clo1313_815 genes also drove reliable expression of the β-galactosidase (*lacZ*) gene and the NADPH-alcohol dehydrogenase (*adhB*) gene (47).

Genetic manipulation of *Clostridium* and related bacterial species is generally difficult, and it is often necessary to overcome DNA restriction-modification systems, morphological heterogeneity and lower transformation efficiency. According to recent reports on DNA transfer into *Clostridium* species, the transformation by electroporation is the currently preferred method (50, 51). In order to establish the appropriate electroporation procedures, it is first necessary to select the appropriate liquid and solid medium for cultivation and to determine the best optical density of the cells in the log phase for harvest. The presence of REase in a host strain is a large obstructive factor for bacterial genetic

transformation. It was reported that in the transformation of *R. thermocellum* strain DSM1313 by electroporation, a plasmid DNA methylated by *E. coli* Dam methylase notably improved its transformation efficiency but conversely *E. coli* Dcm methylation decreased transformation efficiency (52). However, different bacterial REases are affected in different manners by *E. coli* Dam or Dcm methylation (53, 54).

Several inducible expression systems have also been developed in the gene manipulation of *Clostridium* species. In these inducible expression systems, gene expression is induced by lactose for *Clostridium acetobutylicum* (55), tetracycline for *Clostridium difficile* (56), xylose for *C. perfringens* (57), arabinose for *R. cellulolyticum* (58) and laminaribiose for *R. thermocellum* (59). However, there are still many unexplained mechanisms for the development of suitable gene manipulation tools in *Clostridium* species, and we need to select suitable promoters and achieve the arbitrary regulation of genes of interest.

1.6 The purpose of the present research

As described above, in order to develop efficient bioenergy utilization technologies and develop sustainable and environmentally friendly societies, the understanding of lignocellulose structure and the development of its utilization technologies have made remarkable progress in recent years. Although the diversity and complexity of plant cell wall structure in lignocellulose limit the efficiency of its degradation, we know that there are also a variety of degradation systems in nature, on the basis of deeper understanding of cell wall structure and degradation systems in nature. Therefore, the utilization of lignocellulose will undoubtedly have a massive leap. In this context, our efforts have been

being focused on understanding of the cellulolytic and hemicellulolytic systems: characterization of enzymes responsible for lignocellulose degradation including glycoside hydrolases, in-depth study of the growth and metabolism of various bacteria involved in the lignocellulose degradation, and improvement of the efficiency of the original degradation system by genetic engineering strategies.

Ruminiclostridium (formerly *Clostridium*) *josui* isolated from Thai compost is a strictly anaerobic mesophilic cellulolytic bacterium that degrades plant cell wall polysaccharides by producing multi-enzyme complexes, cellulosomes (60, 61). Although this bacterium was first classified in genus *Clostridium*, it was reclassified as *Ruminiclostridium josui*, along with some cellulolytic bacteria, including *R. thermocellum* and *R. cellulolyticum* (62). Consolidated bioprocessing (CBP) of lignocellulose using cellulolytic anaerobic bacteria offers a promising solution for efficient biofuel production without the requirement to add extra cellulolytic or hemicellulolytic enzymes (33). Therefore, characterization of the enzymatic properties of cellulases and hemicellulases from *R. josui* (23, 24, 29, 33, 63), as well as the establishment of its host-vector system are currently of considerable interest for future biotechnological applications.

In this study, characterization of *R. josui* Abf62A-Axe6A (NCBI Reference Sequence: WP_024831741.1), a modular enzyme containing (in order from the N-terminus) an N-terminal signal peptide, a GH62 module, a CBM6 module, a dockerin module, and a CE6 module (Fig. 2), has been done. Three purified recombinant enzymes derived from the full-length Abf62A-Axe6A: RjAbf62A-Axe6A, a full-length derivative lacking the signal peptide; RjAbf62A-CBM6, consisting of the GH62 and CBM6 modules; and RjAxe6, consisting of the CE6 module only have been constructed, purified and biochemically

characterized. New insights into the specificity and mode of action of the GH62 module were obtained by using a diverse range of xylans and oligosaccharides as substrates,. Besides, a restriction endonuclease, termed *Rjo*I, was identified in the cell-free extract of *R. josui* as an isoschizomer of *Dpn*I. Based on this information, the host-vector system of *R. josui* was established using plasmid vectors by electroporation. Finally, the *R. josui cel48A* gene, devoid of the dockerin-encoding region, was successfully expressed in *R. josui* using a newly developed expression vector plasmid pKKM801.

Chapter 2: Characterization of a novel hemicellulase Abf62A-Axe6A from *Ruminiclostridium josui*

2.1 Summary

R. josui Abf62A-Axe6A is a modular enzyme consisting of five distinct modules: an N-terminal signal peptide, a family-62 GH (GH62) module, a family-6 CBM (CBM6), a dockerin module, and a family-6 CE (CE6) module in order from the N-terminus. The biochemical properties of three purified recombinant enzymes: RjAbf62A-Axe6A, a full-length derivative lacking the signal peptide; RjAbf62A-CBM6, consisting of the GH62 and CBM6 modules; RjAxe6, consisting of the CE6 module only. RjAbf62A-Axe6A and RjAbf62A-CBM6 were highly active toward arabinoxylan and moderately active toward sugar beet arabinan, and released mainly arabinose. Analysis of the arabinoxylooligosaccharide hydrolysis products revealed that RjAbf62A-Axe6A released α-1,2- and α-1,3-linked arabinofuranose from both singly and doubly substituted xylosyl residues. Surprisingly, RjAbf62A-Axe6A also demonstrated an endoxylanase activity toward birchwood and beechwood xylans and xylooligosaccharides. RjAbf62A-CBM6 showed no detectable endo-xylanse activity toward soluble birchwood and beechwood xylans. Furthermore, RjAbf62A-Axe6A and RjAxe6 showed acetylxylan esterase activity toward insoluble wheat arabinoxylan, that is, RjAbf62A-Axe6A is a tri-functional enzyme having α-L-arabinofuranosidase, endo-xylanse and acetylxylan esterase activities.

2.2 Introduction

Plant cell walls consist of structurally intricate networks made up from a variety of carbohydrates and are considered the most abundant and important biological resource for biofuel production. However, the lack of cost-effective and environmentally procedures for the bioconversion of biomass resources to fermentable sugars are major obstacles that still

need to be overcome (64). Xylans, which are major constituents of plant cell-walls, are a diverse group of heteropolysaccharides with complex structures that can differ greatly depending on their origin. The xylan backbone is built of β-1,4-linked D-xylopyranosyl residues with a variable number of side chains consisting of different carbohydrate and acid substituents. Hardwood xylans contain α-1,2-linked 4-*O*-methyl-D-glucuronic acid and acetyl side groups at positions 2 and/or 3. Softwood xylans are similar, but their side chains contain α-1,3-linked L-arabinofuranosyl (Afa*f*) residues instead of acetyl groups. The backbones of graminaceous xylans are singly or doubly substituted with α-1,2- and /or α-1,3-linked L-Afa*f* residues, 4-*O*-methylglucuronic acid residues and acetyl groups (9, 65, 66) (Fig. 3A). Arabinose-containing polysaccharides are the major components of hemicelluloses, including arabinoxylan, arabinan, and arabinogalactan (9, 64). Corn fiber hemicellulose is one of the most complex heteroxylans: it is highly branched with arabinose, xylose, methylglucuronic acid, short glycoside chains, and/or other substituents including acetic acid linked directly to over 70% of the xylose backbone residues (67).

The heterogeneity and complex nature of xylans mean that complete breakdown of their structures requires the cooperative actions of a complex of several hydrolytic enzymes with diverse specificities and modes of action. These include endo-1,4-β-xylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), α-L-arabinofuranosidases (ABFs) (EC 3.2.1.55), α-glucuronidases (EC 3.2.1.139), acetylxylan esterases (AXEs) (EC 3.1.1.72), and feruloyl esterases (EC 3.1.1.73) (10). Among these, the ABFs reported to date have been classified into glycoside hydrolase (GH) families 2, 3, 43, 51, 54, and 62 in the Carbohydrate-Active Enzymes (CAZy) database (http://www.cazy.org/), based on their amino acid sequence similarities (32). ABFs involved in arabinoxylan degradation are divided into two groups

based on their modes of action. For example, the arabinoxylan arabinofuranohydrolase (AXH) from *Bacillus subtilis*, AXH-m2,3, specifically hydrolyzes glycosidic bonds between α-1,2- and α-1,3-linked Ara*f* residues and xylosyl residues singly substituted with Ara*f* at either position 2 or 3, whereas *Humicola insolens* AXHd3 specifically targets 1,3-linked Ara*f* residues on doubly arabinosylated xylans (10).

AXEs catalyze the hydrolysis of *O*-acetyl groups from positions 2 and/or 3 on the β-D-xylopyranosyl residues of acetylated xylans. This reduces steric hindrance and therefore facilitates the action of xylanases and β-xylosidases. AXEs are classified in carbohydrate esterase (CE) families 1, 4, 5, 6, 7 and 16 (10). CE6 esterases possess broad substrate specificities, and do not require a free vicinal hydroxyl group for deacetylation activity.

Cellulases and hemicellulases are generally modular enzymes composed of a catalytic module, one or more carbohydrate binding modules (CBMs), and/or other functional modules (68). CBMs are one of the most important functional modules as they promote catalysis by bringing the associated catalytic module into close and prolonged association with its recalcitrant substrate (15).

2.3 Materials and Methods

2.3.1 Bacterial strains and expression plasmid

Genomic DNA was prepared from. *R. josui* JCM 1788 (61). Plasmid constructs were maintained and propagated in *E. coli* strain DH5α. *E. coli* BL21-CodonPlus (DE3)-RIPL and pET-28α(+) were obtained from Novagen (Madison, WI, USA).

2.3.2 Expression and purification of recombinant RjAbf62A-Axe6A, RjAbf62A-CBM6 and RjAxe6A

DNA encoding Abf62A-Axe6A (NCBI Reference Sequence: WP_024831741.1) minus the signal peptide, and its derivatives RjAbf62A-CBM6 and RjAxe6A were amplified by PCR from *R. josui* genomic DNA using the primer sets described in Supplementary Table 1 and LA *Taq* DNA polymerase (Takara Bio, Kusatsu, Japan). The PCR products, engineered to contain *Nhe*I and *Sal*I sites, were cloned into the corresponding sites of pET-28α(+), creating an in-frame C-terminal 6×Histidine tag (Fig. 2A).

E. coli BL21-CodonPlus (DE3)-RIPL transformed with one of the recombinant plasmids (pET-RjAbf62A-Aex6A, pET-RjAbf62A-CBM6 or pET-RjAex6A) was cultured in 200 ml of Super Broth (3.5% Bacto Tryptone, 2% Bacto yeast extract, 0.5% NaCl, pH 7.5) supplemented with kanamycin (25 μg/ml) and chloramphenicol (40 μg/ml) at 37 °C and 120 rpm to late-exponential growth phase (A_{600}=0.9). Protein expression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside and incubation continued at (20 °C, 80 rpm, 20 h). Cells were harvested, resuspended in 50 mM sodium phosphate buffer pH 7.4, 10 mM imidazole, and disrupted using sonication. Cell debris was removed by centrifugation. The histidine-tagged recombinant proteins were purified from the soluble fraction using a HisTrap HP column (GE Healthcare Japan, Tokyo, Japan), according to the supplier's protocol.

All recombinant proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (69) and Coomassie brilliant blue staining (CBB Stain One, Nacalai Tesque, Kyoto, Japan). Protein concentrations were measured using a BioRad

protein assay kit (Bio-Rad Laboratories, Hercules, CA, USA) with bovine serum albumin as the standard.

2.3.3 Enzyme assays

Rye arabinoxylan, soluble wheat arabinoxylan, insoluble wheat arabinoxylan, sugar beet arabinan and linear 1,5-α-L-arabinan were purchased from Megazyme (Wicklow, Ireland), and oat-spelts xylan, birchwood xylan and beechwood xylan were purchased from Sigma-Aldrich (St. Louis, MN, USA). The α-L-arabinofuranosidase activities of RjAbf62A-Axe6A and RjAbf62A-CBM6 were measured using reaction mixtures consisting of 50 mM sodium phosphate buffer pH 7.0 containing 0.5% (w/v) of substrate. Activities toward rye arabinoxylan, soluble wheat arabinoxylan, or sugar beet arabinan were determined following incubation at 45 °C for 10 min. Activities toward oat-spelt xylan and insoluble wheat arabinoxylan were measured following incubation at 35 °C for 1 h. Activities toward linear 1,5-α-L-arabinan, birchwood xylan and beechwood xylan were determined following a 24h incubation at 35 °C. Reducing sugars released in the reaction mixtures were determined by the Somogyi-Nelson method (70). One unit of activity was defined as the amount of enzyme that released 1 μmol arabinose or xylose equivalent from the substrate per minute.

The acetylxylan esterase activities of RjAbf62A-Axe6A and RjAxe6A was measured using α-D-cellobiose octaacetate (Tokyo Chemical Industry, Tokyo, Japan), cellulose acetate (Sigma-Aldrich), and insoluble wheat arabinoxylan as substrates in 50 mM sodium phosphate buffer pH 7.0. Activities toward 1.5% (w/v) α-D-cellobiose octaacetate were measured following incubation at 45 °C for 10 min. Activities toward 1.5% (w/v) cellulose

acetate and 0.5% (w/v) insoluble wheat arabinoxylan were determined following incubation at 35°C for 24h. Reaction mixtures were then passed through disc syringe filters (0.22 µm, GLCTD-HPTFE 1322, TORAST™ Disc, Shimadzu GLC, Tokyo, Japan), and released acetic acid was quantified using an LC-10A VP of CLASS VP™ HPLC system (Shimadzu, Kyoto, Japan) equipped with a Shim-pack SCR 102 packed column (Shimadzu) and a CDD-10A VP (Shimadzu) conductivity detector. One unit of activity was defined as the amount of enzyme that released 1 µmol of acetic acid equivalent from the substrates per minute. All enzyme assays were performed in triplicate.

2.3.4 Thin-layer chromatography (TLC) analysis of hydrolysis products

Reaction mixtures (20 µl) consisting of purified enzymes (0.2 units) mixed with 0.5% (w/v) polysaccharides (soluble wheat arabinoxylan (Megazyme), insoluble wheat arabinoxylan, sugar beet arabinan, linear 1,5-α-L arabinan, oat-spelt xylan, birchwood xylan, or beechwood xylan) in 50 mM sodium phosphate buffer pH 6.5 were incubated (35 °C, 24 h). Reaction products were separated on silica gel 60 plastic sheets (Merck Japan, Tokyo) using a solvent system of 1-butanol:acetic acid:water (5:2:3, v/v) and visualized by spraying the plate with an aniline-diphenylamine-phosphoric acid reagent and heating the plate at 110 °C.

Reaction mixtures (20 µl) comprising 0.2 units of the purified enzymes mixed with 1,5-α-L-arabinooligosaccharides (arabinobiose to arabinohexaose, 5 µg each), arabinoxylooligosaccharides [3^2-α-L-arabinofuranosyl xylobiose (A^3X), 2^3,3^3-di-α-L-arabinofuranosyl xylotriose ($A^{2,3}XX$), 2^3-α-L-arabinofuranosyl xylotriose (A^2XX), 3^3-α-L-arabinofuranosyl xylotetraose (XA^3XX) (Megazyme) (Fig. 3B), 5 µg each], and 1,4-β-D-

xylooligosaccharides (xylobiose to xylohexaose, 5 µg each) in 50 mM sodium phosphate buffer pH 6.5 were also prepared and incubated (35 °C, 24 h). The hydrolysis products were analyzed by TLC as described above.

2.3.5 Analysis of hydrolysis products using high-performance anion-exchange chromatography with a pulsed amperometric detection system (HPAEC-PAD)

Hydrolysis products from various polysaccharides and oligosaccharides were prepared as described in section 2.4. In some experiments, GH3 β-xylosidase was added to the reaction mixture to confirm that hydrolysis products were xylooligosaccharides as described recently (71). Hydrolysis products were analyzed by HPAEC-PAD (Dionex, Sunnyvale, California, USA) using a CarboPac PA1 column (Dionex). The column was developed at a flow rate of 1 ml/min with 0.1 M NaOH (0–5 min) followed by a linear gradient (5–35 min) of sodium acetate (0–0.4 M).

2.4 Results

2.4.1 Modular structure of R. josui Abf62A-Axe6A, and expression and purification of recombinant enzymes.

R. josui Abf62A-Axe6A is a modular enzyme comprising, in order from the N-terminus, an N-terminal signal peptide followed by GH62, CBM6, dockerin, and CE6 modules (Fig. 2A). Such modular organization is conserved in a few *Ruminiclostridium* species closely related to *R. josui*, including *R. cellulolyticum*, *R. papyrosolvens*, and *R. cellobioparum.* The amino acid sequence of the GH62 module of Abf62A-Axe6A has a

lower homology compared with other biochemically characterized GH62 enzymes, whereas the CE6 module of Abf62A-Axe6A is highly homologous to the CE6 module of *Acetivibrio cellulolyticus* AcGH5 (Fig. 4). Three C-terminally Histidine-tagged recombinant enzymes, RjAbf62A-Axe6A, RjAbf62A-CBM6, and RjAxe6A, were overexpressed in *Escherichia coli* BL21-CodonPlus (DE3)-RIPL and purified by nickel-chelating chromatography to homogeneity as determined by SDS-PAGE (Fig. 2B). The molecular sizes of RjAbf62A-Axe6A, RjAbf62A-CBM6 and RjAxe6A, as estimated from SDS-PAGE, were approximately 85, 53 and 30 KDa, respectively, which was in agreement with the molecular weights (84,570, 51,775, and 29,397, respectively) predicted from their polypeptide sequences.

2.4.2 Characterization of the catalytic activities of purified RjAbf62A-Axe6A and RjAbf62A-CBM6.

With rye arabinoxylan as the substrate, RjAbf62A-Axe6A was optimally active at 55 °C and pH 5.5, and was stable following a 1 hour incubation over a range of pH values from 5.0 to 11.0 at 35 °C. RjAbf62A-CBM6 exhibited maximal activity at 50 °C and pH 6.0, and retained its initial activity after a 1 hour incubation over a pH range from 6.0 to 10.0 at 50 °C.

RjAbf62A-Axe6A exhibited high enzymatic activity toward soluble rye and wheat arabinoxylans; moderate activity toward sugar beet arabinan, insoluble wheat arabinoxylan and oat spelts xylan; and negligible but definite activity toward birchwood xylan, beechwood xylan and linear arabinan (Table 2). Overall, RjAbf62A-CBM6 exhibited a similar substrate preference to RjAbf62A-Axe6A, but RjAbf62A-CBM6 activity was not

detected when birchwood xylan or beechwood xylan were used as substrates (Table 2). Although RjAbf62A-Axe6A exhibited higher hydrolytic activities toward rye and wheat arabinoxylans compared with RjAbf62A-CBM6, RjAbf62A-Axe6A exhibited a lower activity toward insoluble wheat arabinoxylan and oat spelts xylan compared with RjAbf62A-CBM6 (Table 2). These results strongly suggested that the GH62 module of RjAbf62A-Axe6A has an α-L-arabinofuranosidase activity capable of releasing arabinose residues from arabinosylated xylans and branched arabinans, but has a weak xylanase activity.

Next, we determined the hydrolytic activities of both enzymes toward arabinoxylooligosaccharides, including A^3X, $A^{2,3}XX$, A^2XX and XA^3XX. RjAbf62A-Axe6A exhibited high activity toward A^3X, A^2XX and XA^3XX, but low activity toward $A^{2,3}XX$. These results were consistent with those obtained from arabinoxylans, which suggested RjAbf62A-Axe6A has α-L-arabinofuranosidase activity.

2.4.3 Acetylxylan esterase activity of RjAbf62A-Axe6A and RjAxe6A.

When the acetylxylan esterase activities of RjAbf62A-Axe6A and RjAxe6A were assayed using α-D-cellobiose octaacetate, cellulose acetate, and insoluble wheat arabinoxylan as substrates, both enzymes exhibited high activity toward α-D-cellobiose octaacetate and weak activity toward cellulose acetate and insoluble wheat arabinoxylan (Table 3). RjAbf62A-Axe6A exhibited a higher activity toward insoluble wheat arabinoxylan but lower activities toward α-D-cellobiose octaacetate and cellulose acetate compared with RjAxe6A. (Table 3). We determined that RjAxe6A exhibited maximal activity toward α-D-cellobiose octaacetate at 50 °C and pH 6.0.

2.4.5 Analysis of RjAbf62A-Axe6A and RjAbf62A-CBM6 xylan and arabinan hydrolysis products.

To identify the products from xylan and arabinan hydrolysis by RjAbf62A-Axe6A and RjAbf62A-CBM6, reaction mixtures following 24 h incubation were analyzed by TLC and HPAEC-PAD. Both enzymes produced arabinose from rye arabinoxylan, wheat arabinoxylan, insoluble wheat arabinoxylan, sugar beet arabinan, linear arabinan and oat spelts xylan but did not produce arabinose from birchwood or beechwood xylans (Fig. 5A). Surprisingly, RjAbf62A-Axe6A produced some oligosaccharides in addition to arabinose from all xylan variants tested, but only arabinose from arabinans. Conversely, RjAbf62A-CBM6 did not yield any oligosaccharides from the same substrates (Fig. 5B). These results confirmed that RjAbf62A-Axe6A, unlike RjAbf62A-CBM6, had weak but definite activity toward birchwood and beechwood xylans (Table 2). HPAEC-PAD analysis of the RjAbf62A-Axe6A hydrolysis products confirmed the TLC results. As shown in Fig. 6, RjAbf62A-Axe6A produced several oligosaccharides from birchwood xylan and rye arabinoxylan and some were identified as xylotriose (X3), xylotetraose (X4), and xylopentaose (X5) by comparison with standard oligosaccharides. Larger oligosaccharides were also detected (Figs. 5 and 6). When rye arabinoxylan and insoluble wheat arabinoxylan were incubated with RjAbf62A-Axe6A for a range of time periods up to 24 h, only arabinose was identified in the early reaction stages and X3, X4, X5, and other oligosaccharides were detected following prolonged incubations (Fig. 7). The oligosaccharides presented in the reaction mixture following a 24 h incubation diminished upon treatment with GH3 xylosidase. Furthermore, xylose appeared as a new peak, which

strongly indicated that the oligosaccharide products were xylooligosaccharides. RjAbf62A-Axe6A hydrolyzed birchwood xylan to xylooligosaccharides but did not produce any arabinose (Fig. 7), indicating that it had endoxylanase activity in addition to α-L-arabinofuranosidase activity.

Next, we evaluated the enzyme activities against xylooligosaccharides. RjAbf62A-Axe6A produced mainly X2 and X3 from X5, and X2, X3 and X4 from X6. However, RjAbf62A-Axe6A was less active toward X4, and we detected no activity toward X2 and X3 (Fig. 8), suggesting that RjAbf62A-Axe6A hydrolyzed xylooligosaccharides via an endo-mode. Although RjAbf62A-CBM6 exhibited similar specificities toward xylooligosaccharides, its overall activity appeared to be weaker than RjAbf62A-Axe6A (Fig. 8). These results were confirmed by HPAEC-PAD analysis (Fig. 9).

When A^2XX and XA^3XX were separately incubated with RjAbf62A-CBM6 or RjAbf62A-Axe6A for 24 h, both substrates were completely hydrolyzed to release arabinose and xylotriose, and arabinose and xylotetraose, respectively (Fig. 10). These results suggest that A^2XX and XA^3XX were good substrates for RjAbf62A-CBM6 and RjAbf62A-Axe6A. Conversely, $A^{2,3}XX$ was more resistant to enzyme attack: $A^{2,3}XX$ was still detected in reaction mixtures following 24 h incubation with RjAbf62A-CBM6 or RjAbf62A-Axe6A and lower levels of the arabinose and xylotriose products were produced. These results indicated that both enzymes are able to hydrolyze either 1,2- or 1,3-linkage of doubly arabinosylated xylosyl residues in addition to 1,2- and 1,3-linkages of singly arabinosylated xylosyl residues.

Furthermore, RjAbf62A-CBM6 and RjAbf62A-Axe6A were active toward arabinooligosaccharides: both enzymes converted arabinopentaose (A5) and

arabinohexaose (A6) to arabinose (Fig. 11), suggesting that both enzymes were capable of releasing arabinoses from nonreducing ends of α-1,5-linked arabinooligosaccharides.

2.5 Discussion

Xylan is the most abundant hemicellulose and its backbone is decorated with various substituents: for example, graminaceous xylans are singly or doubly substituted with α-1,2- and /or α-1,3-linked L-Afa*f* residues and acetyl groups as well as 4-*O*-methylglucuronic acid residues (24, 33, 60). ABFs and AXEs play important roles in efficient degradation of xylans as they remove xylan sidechain substituents. This reduces steric hindrance and allows xylanases to easily access their substrates.

RjAbf62A-Axe6A contains two different catalytic modules, GH62 and CE6, and was therefore predicted to have ABF and AXE activities. According to the CAZy database (32), many reported ABFs are classified into the GH families GH2, GH3, GH43, GH51, GH54 and GH62. GH62 contains only ABFs and RjAbf62A-Axe6A was mostly active toward soluble rye and wheat arabinoxylans, highly branched arabinoxylans, and moderately active toward oat spelts xylan and sugar beet arabinan (Table 2). Although RjAbf62A-Axe6A showed higher activity toward soluble rye and wheat arabinoxylans than RjAbf62A-CBM6, the former was conversely less active toward insoluble wheat arabinoxylan than the latter. This may be caused by steric hindrance by Axe6A and/or dockerin module, in which accessibility of the GH62 module in this modular enzyme to the substrate is hindered by other modules. Like AnAbf62A-m2,3 from *Aspergillus nidulans*, RjAbf62A-Axe6A is active toward sugar beet arabinan as well as arabinoxylans (72, 73). In contrast, other GH62 ABFs from *Cellvibrio japonicas* (74), *Penicillium chrysogenum* (75), *Penicillium*

funiculosum (72, 73), *Streptomyces coelicolor* (73, 76) and *Streptomyces lividans* (77) exhibit no activity toward sugar beet arabinan.

ABFs responsible for arabinoxylan degradation are divided into two groups: AXHm2,3 and AXHd3 (10). RjAbf62A-Axe6A efficiently hydrolyzed the arabinoxylooligosaccharides A^3X, XA^3XX and A^2XX to release arabinose and xylooligosaccharide moieties. Although such substrate specificity is consistent with that of AXHm2,3 ABFs, RjAbf62A-Axe6A showed an additional weak activity toward $A^{2,3}XX$ (Table 2 and Fig. 10). Like RjAbf62A-Axe6A, *A. nidulans* AnAbf62A-m2,3 also has activities toward A^3X, A^2XX and sugar beet arabinan (72, 73). However, unlike *A. nidulans* AnAbf62A-m2,3 and other ABFs, RjAbf62A-Axe6A is also able to released α-1,2- and α-1,3-linked arabinofuranose from doubly substituted xylosyl residues. From our observations, it is not clear if RjAbf62A-Axe6A initially attacks α-1,3 linked arabinosyl residues, like AXHd3 ABFs. This is because a singly arabinosylated intermediate compound was not detected in the reaction mixture, presumably because of the strong AXHm2,3 activity. It has been reported that some GH51 ABFs with AXHm2,3 activity can release arabinose from doubly substituted xylosyl residues located at the non-reducing end. The enzymes exhibited a preference for 1,3-linked Ara*f* residues in doubly substituted xylosyl residues (78, 79) (80). It is certainly possible that RjAbf62A-Axe6A also has an AXHd3-like activity with preference for 1,3-linked Ara*f* residues. RjAbf62A-Axe6A also displayed different substrate preferences compared with *A. nidulans* AnAbf62A-m2,3: RjAbf62A-Axe6A activity toward rye arabinoxylan was 34-fold higher than that toward oat spelts xylan, whereas AnAbf62A-m2,3 showed similar activities toward rye arabinoxylan and oat spelts xylan (72, 73). Although the oat spelts xylan backbone is less arabinosylated

than in rye arabinoxylan (81, 82), there is no obvious reason for the significant disparity in the activities of RjAbf62A-Axe6A and AnAbf62A-m2,3 toward oat spelts xylan. Furthermore, RjAbf62A-Axe6A showed a clear preference for A^3X (16.1 ± 1.5 U/nmol-protein) over A^2XX (2.84 ± 0.17 U/nmol-protein) whereas conversely, AnAbf62A-m2,3 showed a slightly higher preference for A^2XX (59 ± 0.5 U/mg-protein) over A^3X (37 ± 1.1 U/mg-protein).

Among GH62 ABFs without activity toward sugar beet arabinan, *S. coelicolor* Araf62A was precisely investigated about its tertiary structure and enzyme functions (73). Complex forms of Araf62A with A, X3 and X6 demonstrated five subsites in the catalytic cleft and an L-arabinose-binding pocket at the bottom of the cleft, that is, this active site best fits $XXXA^3X$. Actually, the enzyme preferred XA^3X to A^3X as the substrate. Mutagenesis of Asn462 revealed that this residue played an important role for interacting with the xylosyl residue at the nonreducing-end side of the arabinosylated xylosyl residue and enhancing catalytic activity toward arabinosylated xylans and xylooligosaccharides. In addition, molecular modeling suggested that arabinan was unable to fit into the cleft due to structural issues. It is interesting that RjAbf62A-Axe6A prefers A^3X to XA^3XX and is able to degrade sugar beet arabinan, in contrast to *S. coelicolor* Araf62A, suggesting that their active sites possess different structures.

Surprisingly, RjAbf62A-Axe6A displayed an endoxylanase activity toward arabinoxylans and birchwood xylan: it released X3, X4, X5 and other xylooligosaccharides following a prolonged incubation (Fig. 7). RjAbf62A-Axe6A was also able to efficiently hydrolyze xylooligosaccharides (Fig. 8). RjAbf62A-Axe6A hydrolyzed X5 to form X2 and X3 as major products, suggesting that iRjAbf62A has apparent endo-type activity.

Furthermore, RjAbf62A-Axe6A was active on arabinooligosaccharides. Similar broad specificities have been reported for Axy43A from *Paenibacillus curdlanolyticus* strain B-6, a GH43 enzyme which possesses a single module and exhibits endoxylanase, β-D-xylosidase, and arabinoxylan arabinofuranohydrolase activities . GH43 and GH62 family enzymes are grouped into Clan F in the CAZy database, and, based on common sequence motifs, are predicted to possess a five-bladed β-propeller fold. Therefore, it is likely that RjAbf62A-Axe6A and Axy43A have analogous active site structure that can accommodate various poly- and oligo-saccharides. A recent study introduced endoxylanase activity into the GH43 arabinofuranosidase AXHd3 from *Humicola insolens* by modifying the rim of the active site, with a drastic reduction (about 1/250) of k_{cat} for wheat arabinoxylan hydrolysis (83). This suggests that other GH43 and GH62 enzymes are good targets for modification and improvement to augment their biomass-degradation activities.

Although RjAbf62A-CBM6 exhibited a similar substrate specificity to RjAbf62A-Axe6A, RjAbf62A-CBM6 showed only negligible endoxylanase activity, illustrating the importance of the functional organization of this modular enzyme.

It is well documented that acetylation of the xylosyl residues in the xylan main chain severely affects the susceptibility of xylans to xylanolytic enzymes. Removal of the acetyl groups from acetylated xylans makes them susceptible to xylanolytic enzymes (10, 84, 85). CE6 AXEs have been reported to release acetic acid from both monoacetylated (at position *O*-2 or *O*-3) and di-*O*-acetylated β-1,4-linked xylosyl residues (66, 86). RjAxe6A showed an approximate four-fold higher activity toward the synthetic substrate α-D-cellobiose octaacetate compared with RjAbf62A-Axe6A. However, RjAbf62A-Axe6A exhibited a 22-fold higher activity toward insoluble wheat arabinoxylan compared with RjAxe6A (Table

2). These phenomena may be partly explained by steric hindrance in RjAbf62A-Axe6A and the presence or absence of CBM6 in respective enzymes.

In this study, we have demonstrated that RjAbf62A-Axe6A is a multi-functional enzyme with α-L-arabinofuranosidase, endoxylanase, and acetylxylan esterase activities. To the best of our knowledge, this is the first report of a GH62 family enzyme having α-L-arabinofuranosidase and endoxylanase activities. Although the endoxylanase activity of RjAbf62A-Axe6A was low, the elucidation of the structure of RjAbf62A-Axe6A would be invaluable to aid the design of novel multifunctional enzymes with enhanced biomass-degrading activities. Crystallization of the enzyme is now under study.

Chapter 3: Development of host-vector system of *Ruminiclostridium josui*

3.1 Summary

Although *R. josui*, a strictly anaerobic mesophilic cellulolytic bacterium, is a promising candidate for biomass utilization via consolidated bioprocessing, its host-vector system has not yet been established. It is generally difficult to genetically manipulate bacteria of the biotechnologically important genus *Ruminiclostridium* due to the existence of the restriction and modification (RM) systems. The author purified a restriction enzyme *Rjo*I from *R. josui* by column chromatography and characterized it. To identify the recognition sequence of *Rjo*I, two plasmids pQE-30T and pBluescript SK + Ⅱ were used as substrates. Agarose gel electrophoresis analysis of *Rjo*I-digested plasmid DNAs predicted that the sequence recognized by *Rjo*I is G*met*ATC where A is methylated by Dam methylase. Sequencing the cleaved DNA fragments revealed that *Rjo*I recognized and cleaved the sequence $G^{met}ATC$ between A and T ($G^{met}A \downarrow TC$). Due to *Rjo*I is an obstacle for transformation of *R. josui*, we then successfully introduced plasmids prepared from *Escherichia coli* C2925 (*dam*−/*dcm*−) into *R. josui* by electroporation. The highest transformation efficiency of 6.6×10^3 transformants/μg of DNA was obtained using a square wave pulse (750 V, 1 ms). When the *R. josui cel48A* gene, devoid of the dockerin encoding region, cloned into newly developed plasmid pKKM801 was introduced into *R. josui*, a truncated form of RjCel48A, RjCel48AΔdoc, was detected in the culture supernatant but not in the intracellular fraction.

3.2 Introduction

Anaerobic cellulolytic bacteria are very important for the bioconversion of lignocellulosic biomass into biofuels and commodity chemicals. They produce robust and effective polysaccharide-degrading enzymes that are involved in the fermentation of biomass-derived saccharides, resulting in value-added materials such as ethanol (87, 88). Consolidated bioprocessing (CBP) of lignocellulose to fermentation products using anaerobic bacteria provides a promising solution for efficient lignocellulose conversion without the need for the addition of cellulolytic enzymes (33). Transformation of various cellulolytic and solventogenic anaerobes belonging to the genera *Clostridium* and *Ruminiclostridium* is now possible using electroporation procedures, and these host-vector systems, including *R. thermocellum* (51, 89), *R. cellulolyticum* (90), *C. tyrobutyricum* (91), *C. cellulovorans* (42), and *C. acetobutylicum* (92), have been used for metabolic engineering and other basic studies.

Recently reclassified species *Ruminiclostridium josui*, previously *Clostridium josui* (62), is a strictly anaerobic, mesophilic, cellulolytic bacterium that was originally isolated from compost in Thailand (61). It is also known to form a cellulosome (cellulolytic enzyme complex) (60). The *R. josui* genome contains a gene cluster encoding 11 proteins responsible for cellulosome formation, including (5′–3′) *cipA*, *cel48A*, *cel8A*, *cel9A*, *cel9B*, *orfX*, *cel9C*, *cel9D*, *man5A*, *cel9E*, and *cel5B* (63). CipA, Cel48A, and Cel9A were previously identified as major proteins in the cellulosome (60), suggesting that enzymes encoded by this gene cluster are important for the cellulolytic activity of *R. josui*. Recently, full-length CipA was successfully expressed using an *Escherichia coli* expression system, and was used in the construction of artificial enzyme complexes together with a cellulase

and hemicellulases (23, 24). *R. josui* produces ethanol, acetic acid, and hydrogen in large quantities from cellulosic substrates such as Avicel and rice straw, and is therefore a promising candidate for the conversion of cellulosic biomass into biofuels and commodity chemicals using CBP technologies. However, a host-vector system for *R. josui* has not yet been established.

Restriction endonucleases remain a significant barrier to the introduction of foreign DNA into bacterial hosts. If the nature of the restriction system is known, DNA used for transformation can be protected by methylation (*in vivo* or *in vitro*) (42, 44, 45) or other treatments according to the type of DNA modification. Host restriction-modification (RM) systems are classified into four types (I–IV) (93), with type II RM systems being the most common and best studied. Type II RM systems contain a restriction endonuclease and a methyltransferase, whereby the restriction endonuclease recognizes and cleaves specific sequences, and the methyltransferase modifies adenine or cytosine residues at certain recognition sites. About 14,000 bacterial RM systems have been described in REBASE (http://rebase.neb.com) (94).

3.3 Materials and Methods

3.3.1 Strains and plasmids

Bacterial strains and plasmids used in this study are described in Table 4.

3.3.2 Culture media and growth conditions

R. josui was grown anaerobically at 45°C in modified GS medium (95) in Hungate tubes or pressure-resistant bottles with butyl rubber stoppers. The modified GS medium (1 L) contained: yeast extract (4.5 g), KH_2PO_4 (1.5 g), K_2HPO_4 (2.9 g), urea (2.9 g), cysteine hydrochloride (3.0 g), 3-(N-morpholino)propanesulfonic acid (20.9 g), cellobiose (10.0 g), NaCl (3.0 g), resazurin (2 mg), and mineral solution (1 ml). The mineral stock solution contained: 0.3% cysteine hydrochloride, 25% $MgCl_2 \cdot 6H_2O$, 3.75% $CaCl_2 \cdot 2H_2O$, 0.031% $FeSO_4 \cdot 7H_2O$, and $ZnCl_4 \cdot 7H_2O$ (a slight quantity). The medium was adjusted to pH 7.4, sparged with N_2 gas to anaerobicity, and sterilized by autoclaving at 121°C for 20 min. To prepare solid medium, gellan gum (Wako Pure Chemical Industries, Ltd., Osaka, Japan) was added to give a final concentration of 0.8%. Culture medium for recombinant *R. josui* strains was supplemented with erythromycin (20 μg/ml).

All *E. coli* strains were grown aerobically in Luria-Bertani medium at 37°C with shaking at 100 rpm. The medium was supplemented with ampicillin (50 μg/ml) for growth of *E. coli* strains harboring plasmids pKKM801 or pKKM801-cel48AΔdoc.

3.3.3 Purification of restriction endonuclease *Rjo*I

R. josui cells were harvested from 4 liters of culture at the end of log phase by centrifugation (8,000 × *g* for 5 min), and the pellet was resuspended in ice-cold buffer A (20 mM Tris-HCl, pH 7.5) then frozen at −20°C. Thawed cells were resuspended in ice-cold buffer B (20 mM Tris-HCl, pH 7.5, 1.0 mM dithiothreitol (DTT), 0.05 M NaCl) and then lysed by sonication. Following sonication, the suspension was centrifuged at 20,000 × *g* for 20 min at 4°C to remove cell debris. Streptomycin sulfate was added to the supernatant to a final concentration of 1.0%, and the mixture was stirred at 4°C for 30 min

before being centrifuged at 20,000 × *g* for 30 min to remove nucleic acids. Proteins in the supernatant were precipitated by the addition of ammonium sulfate to 70% saturation. The protein pellet recovered by centrifugation was dissolved in ice-cold buffer B and then dialyzed overnight against the same buffer at 4°C. The dialyzed sample was loaded onto a 5-ml HiTrap Heparin HP column (GE Healthcare, Tokyo, Japan), and the column was washed with buffer B and eluted with a 50-ml linear gradient of 0.05–1.0 M NaCl in the same buffer. Fractions were assayed for restriction endonuclease (*Rjo*I) activity by incubation at 45°C for 1 h in buffer M (Toyobo Co., Ltd., Osaka, Japan), using plasmid pQE-30 (Qiagen, Venlo, The Netherlands, prepared from *E. coli* C2925 (*dam*$^{-}$/*dcm*$^{-}$), as a substrate, followed by agarose gel electrophoresis. Fractions containing *Rjo*I activity were combined, diluted, and applied to a 1-ml HiTrap Q HP column (GE Healthcare) equilibrated with buffer B. Adsorbed proteins were eluted with a 20-ml linear gradient of 0.05–1.0 M NaCl in buffer B. *Rjo*I activity was assayed as described above. Active fractions were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (69).

3.3.4 Characterization of *Rjo*I

Plasmids pQE-30 and pBluesecrpt II SK+, containing 16 and 15 *Sau*3AI, *Dpn*I, and *Mbo*I sites, respectively, were prepared from *E. coli* DH5α (*dam*$^{+}$/*dcm*$^{+}$) and *E. coli* C2925 (*dam*$^{-}$/*dcm*$^{-}$), and used as the Dam-methylated and Dam-unmethylated DNA substrates, respectively. These substrates were treated with *Rjo*I at 45°C for 1 h in buffer M (10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT; Toyobo). Restriction

endonucleases *Sau*3AI, *Dpn*I, and *Mbo*I were also used to digest the same substrates. Digestion profiles were examined by agarose gel electrophoresis.

The recognition and cleavage site of *Rjo*I was determined as follows. Plasmid pQE-30), prepared from *E. coli* DH5α (dam^+/dcm^+), was digested with *Rjo*I and subjected to agarose gel electrophoresis. The 435-bp fragment was recovered from an agarose gel using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). Primers 435bpF and 435bpR (Table 5) were used to sequence both strands of the 435-bp fragment using a DYEnamic ET Terminator Cycle Sequencing Kit (GE Healthcare) and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

3.3.5 Plasmid construction

A scheme of the construction of pKKM801 is illustrated in Fig. 12. Oligonucleotides used as primers or linkers are listed in Table 5. The promoter region (Pcip) of the *R. josui* cellulase gene cluster (63) was amplified from *R. josui* genomic DNA using *LA Taq* DNA polymerase (Takara Bio Inc., Kusatsu, Japan) and primers Pcip2079F and Pcip2622R. The resulting fragment was digested with *Eco*RI and *Bam*HI and ligated into the corresponding sites of plasmid pUC18, yielding pUC18-Pcip. The multiple cloning site of pUC18-Pcip was reconstructed by inserting a linker consisting of pRjHis1 and pRjHis2 between the *Pst*I and *Hin*dIII sites, yielding pUC18-Pcip-His, which included a 6×His-tag-encoding region downstream of the multiple cloning site. A fragment containing the regions responsible for replication (*oriCP* and *rep*) and erythromycin resistance (*ermBP*) in *Clostridium perfringens* was amplified from pJIR751 (96) using primers pJIR-PfoF and

pJIR-EcoO109R. The resulting amplicon was inserted into pUC18-Pcip-His digested with *Pfo*I and *Eco*0109I using an In-Fusion HD Cloning Kit (Takara Bio), yielding pKKM801.

pKKM801-Cel48AΔdoc was constructed as follows. The *cel48A* gene, devoid of the dockerin-encoding region, was amplified using primers Rjcel48AF and Rjcel48AR_2 (Table 5) from *R. josui* genomic DNA. The amplified fragment was then cloned between the *Nco*I and *Sal*I sites of pKKM801 using the In-Fusion HD Cloning Kit, yielding pKKM801-Cel48AΔdoc. The translated protein contained a 6×His-tag at the C-terminus.

3.3.6 Electrotransformation of *R. josui*

R. josui was transformed with pKKM801 or pKKM801-Cel48AΔdoc by electroporation as described for *R. thermocellum* (51), with some modifications. *R. josui* was grown in 10 ml of modified GS medium to an optical density at 595 nm (OD_{595}) of 0.7–0.9. The culture was centrifuged at 10,000 × *g* for 1 min in an anaerobic chamber and the supernatant was discarded. Cells were washed twice with 1 ml of ice cold anaerobic electroporation buffer (5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 500 mM sucrose, pH 6.5), resuspended in 100 µl of the same buffer, and then stored on ice until use. Two microliters of purified pKKM801 (500 ng/µl) were added to 50 µl of the washed cells in a pre-chilled 0.1-cm electroporation cuvette. The washed cells were then subjected to a 750-V, 1-ms square-wave pulse using a Bio-Rad GenePluser XCell (Bio-Rad, Hercules, CA, USA). Cells were immediately transferred into 5 ml of modified GS medium and incubated at 45°C for 4 h. Serial dilutions of the transformed cells were mixed with 20 ml of modified GS medium containing erythromycin (20 µg/ml), and then solidified with gellan gum (final concentration of 0.8%) in plates. Solidified plates were placed in a pouch

of AnaeroPack-Anaero (Mitsubishi Gas Chemical Co., Tokyo, Japan) and incubated at 45°C for up to 1 week, with all resultant colonies then counted. Alternatively, the most probable number (MPN) test was employed for estimation of transformation efficiency according to the United States Food and Drug Administration Bacterial Analytical Manual (Appendix 2: Most Probable Number from Serial Dilutions, available from http://www.fda.gov). Ten-fold serial dilutions of each transformed culture were prepared, and 0.5-ml aliquots of each dilution were inoculated into three separate Hungate tubes containing 10 ml of modified GS medium supplemented with erythromycin (20 µg/ml). The tubes were then incubated at 45°C for 1 week to examine the growth of erythromycin-resistant transformants. MPN counts for each dilution were estimated from the MPN table.

Successful transformation was confirmed by re-isolation of plasmid DNA from transformants as described by Tyurin et al. (97). Plasmid DNA was isolated from erythromycin-resistant *R. josui* transformants using a QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's pretreatment protocol for plasmid DNA isolation from Gram-positive bacteria. The isolated plasmid DNA was used to transform *E. coli* DH5α cells. Plasmid DNA was then re-isolated from ampicillin-resistant *E. coli* transformants and subjected to *Bgl*I or *Sty*I digestion followed by agarose gel electrophoresis analysis to identify the plasmid.

3.3.7 Plasmid copy number

Plasmid copy number in *R. josui* cells was determined using quantitative real-time polymerase chain reaction (qRT-PCR) analysis as described previously (47, 98) using a StepOnePlus Real-time PCR System (Thermo Fisher Scientific Inc., Waltham, MA, USA).

qRT-PCR was performed using the separate primer sets qampF/qampR (Table 5), which is specific for *bla*, and qman5AF/qman5AR (Table 5), which specifically amplifies *man5A*. Total DNA from the *R. josui* transformant, prepared as described by Silhavy et al. [26], was used as a template, and reactions were carried out using THUNDERBID SYBR qPCR Mix (Toyobo) as per the manufacturer's recommendations. Standard curves for the *bla* and *man5A* amplifications were generated using products amplified from purified pKKM801 and wild-type *R. josui* genomic DNA, respectively. As both *bla* and *man5A* are single-copy genes located on pKKM801 and the *R. josui* chromosome, respectively (63), the copy number of pKKM801 in a single *R. josui* cell was estimated from the copy ratio of *bla* to *man5A*.

3.3.8 Detection of recombinant RjCel48AΔdoc

Plasmid pKKM801-Cel48AΔdoc was introduced into *R. josui* by electroporation, and *R. josui*/pKKM801-Cel48AΔdoc was grown anaerobically at 45°C in 20 ml of modified GS medium supplemented with cellobiose and erythromycin. Wild-type *R. josui* was cultivated in the same medium without erythromycin. When the cultures reached an OD_{595} of 1.2–1.4, cells were centrifuged at 12,000 × *g* for 10 min at 4°C. Culture supernatants were concentrated to 200 µl using an Amicon Ultra-2 mL 10K centrifugal filter (cut-off, 10 kDa; MILLIPORE, Billerica, MA, USA) and used as extracellular fractions. Cells were disrupted by sonication, and cell debris was removed by centrifugation at 16,000 × *g* for 30 min. The supernatants were concentrated to 200 µl using an Amicon Ultra-2 mL 10K centrifugal filter and used as intracellular fractions. Aliquots (12 µl each) of the intracellular and extracellular fractions were mixed with 6 µl of loading

buffer for SDS-PAGE, heated at 100°C for 10 min, and subjected to SDS-PAGE analysis (10% polyacrylamide for separating gel). Proteins in the gel were then electrotransferred onto a Hybond-P membrane (GE Healthcare). The membrane was soaked in phosphate-buffered saline (pH 7.2) containing 5.0% bovine serum albumin at 4°C overnight to reduce non-specific binding, and then probed with Anti-His-tag Monoclonal Antibody (Medical & Biological Laboratories Co., Nagoya, Japan). Monoclonal antibody bound to 6×His-tagged Cel48A was detected using anti-mouse horseradish peroxidase conjugate and an Amersham ECL Western Blotting Detection Kit (GE Healthcare). Fluorescence was detected using a LuminoGraph I (ATTO Co., Tokyo, Japan).

3.4 Results

3.4.1 Detection and purification of *Rjo*I from *R. josui*

To determine the presence or absence of restriction enzyme in *R. josui* JCM 17888 cell extract, pQE-30 DNA from *E. coli* DH5α (dam^+/dcm^+) and *E. coli* C2925 (dam^-/dcm^-) was treated with prepared cell extract. Results showed that a restriction enzyme with activity towards Dam-methylated DNA was present in the cell extract of *R. josui*. Therefore, Dam-methylated DNA was used for further characterization of this restriction enzyme, named *Rjo*I. No restriction enzyme activity but *Rjo*I activity in the crude cell extract and during the purification step were detected, suggesting that *Rjo*I is a unique restriction enzyme in *R. josui*.

*Rjo*I was partially purified from crude extract prepared from a 4-liter culture of *R. josui* by ammonium sulfate precipitation (70% saturation), and by HiTrap Heparin HP and

HiTrap Q HP column chromatography. *Rjo*I activity was detected in fractions eluted from the HiTrap Heparin HP column with 0.5–0.65 M NaCl, and in fractions eluted from a HiTrap Q HP column with 0.5–0.65 M NaCl (Fig. 13a). When fractions from the HiTrap Q HP column were subjected to SDS-PAGE (Fig. 13b), several proteins were found in fractions 7 and 8, in which *Rjo*I activity was detected (Fig. 13c). Therefore, unfortunately, the *Rjo*I protein was not identified.

3.4.2 Determination of the recognition sequence and cleavage site of *Rjo*I

Analysis of reaction buffers for assaying *Rjo*I activity showed that all buffers tested produced similar results based on agarose gel electrophoresis of cleaved pQE-30 and pBluescript II SK+ DNA fragments. Therefore, we used buffer M for further analyses.

Migration patterns of *Rjo*I-digested pQE-30 and pBluescript II SK+ DNA in an agarose gel suggested that *Rjo*I was an isoschizomer of *Dpn*I, *Mbo*I, or *Sau*3AI. Fig. 14 shows the migration patterns of pQE-30 and pBluescript II prepared from *E. coli* DH5α (dam^+/dcm^+) and *E. coli* C2925 (dam^-/dcm^-) and digested with *Rjo*I, *Dpn*I, *Mbo*I, or *Sau*3AI. The four restriction enzymes each produced an identical pattern for each of the two plasmids, although the effects of Dam-methylation of the substrates differed for each of the enzymes: *Mbo*I was only active on Dam-unmethylated DNA, *Sau*3AI was active on both Dam-methylated and Dam-unmethylated DNA, and *Rjo*I and *Dpn*I were only active on Dam-methylated DNA. These results suggested that the sequence recognized by *Rjo*I is 5′-G^{met}ATC-3′, where the A nucleotide is Dam-methylated at the N^6 position in the sequence GATC.

To identify the cleavage site of *Rjo*I, plasmid pQE-30 was digested with *Rjo*I, the 435-bp DNA fragment was recovered, and both ends of the fragment were sequenced. As shown in Fig. 15, *Rjo*I cleaved the sequence 5′-G^{met}ATC-3′ between the A and T (5′-G^{met}A↓TC-3′), indicating that *Rjo*I is an isoschizomer of *Dpn*I (99, 100). These results suggested that *Rjo*I is an obstacle for transformation of *R. josui* with Dam-methylated DNA. Therefore, DNA not modified by Dam-methylase should be able to be used for transformation of *R. josui*.

3.4.3 Transformation of *R. josui* by electroporation

An *E. coli* and *R. josui* shuttle vector was constructed from an *E. coli* plasmid pUC18 (101) and a *C. perfringens* plasmid pJIR751 (96). The resulting plasmid, pKKM801, contains the *oriEC* origin for replication in *E. coli*, and the *oriCP* origin for replication in *R. josui* (Fig. 12). The plasmid also contains the erythromycin and ampicillin resistance genes *ermBP* and *bla* for selection of *R. josui* and *E. coli*, respectively. In addition, pKKM801 contains the Pcip promoter region of a large cellulase gene cluster from *R. josui*, multiple cloning sites, and a 6×His-Tag-coding sequence downstream of Pcip to produce recombinant proteins in *R. josui*. The nucleotide sequence of pKKM801 has been deposited in the DNA Data Bank of Japan under accession number LC331101.

This plasmid was successfully introduced into *R. josui* JCM 17888 via electroporation after many attempts. Following electroporation, *R. josui* cells were cultivated in both solid and liquid medium containing erythromycin (20 μg/ml), resulting in growth of erythromycin-resistant transformants in both media. As transformants were more

reproducibly obtained in the liquid medium, the numbers of erythromycin resistant transformants were determined by the MPN method. Using the transformation conditions described above, the highest observed efficiency of *R. josui* transformation with pKKM801 was 6.6×10^3 transformants/μg of DNA. We obtained no transformant when plasmid DNA prepared from *E. coli* C2925 (*dam*$^-$/*dcm*$^-$) was used.

Plasmid copy number in the *R. josui* cells was determined by qRT-PCR analysis. pKKM801 and *R. josui* genomic DNA were detected separately using primer sets specific for *bla* and for *man5A*, respectively, and the plasmid copy number was determined from the copy ratio of *bla* to *man5A*. A copy ratio of 1:14 was determined for pKKM801 in *R. josui* cells.

To determine the integrity of pKKM801 following transformation into *R. josui*, plasmid rescue experiments were carried out as described previously (97). As shown in Fig. 16, migration patterns of re-isolated and digested pKKM801 DNA were identical to those of the original pKKM801 plasmid used for transformation of *R. josui*, confirming the integrity of the transformation.

3.4.4 Production of RjCel48AΔdoc in *R. josui*

To examine the applicability of the *R. josui* host-vector system for the construction of engineered *R. josui*, homologous expression of *cel48A* was attempted using pKKM801 as the expression vector. *cel48A* lacking the dockerin-coding region was cloned into pKKM801, which was then introduced into *R. josui*. Recombinant *R. josui*/pKKM801-RjCel48AΔdoc was cultured in modified GS medium containing cellobiose as a carbon

source, and the intracellular and extracellular fractions were examined for the presence or absence of RjCel48AΔdoc by western blot analysis using an anti-His-tag monoclonal antibody. As shown in Fig. 17, RjCel48AΔdoc was detected in the extracellular fraction of the recombinant clone but not in the intracellular fraction, indicating that RjCel48AΔdoc was secreted into the culture medium.

3.5 Discussion

The construction of host-vector systems in *Ruminiclostridium* species, including *R. thermocellum*, has helped to not only establish and improve CBP but also characterize bacterial genetic systems (33, 47, 59, 89, 102). In this study, we overcame the limitations of the host RM system and constructed a host-vector system for *R. josui* based on a *C. perfringens* shuttle vector and the *cip* promoter of *R. josui*.

RM systems in bacteria are defensive machineries that present strong obstacles for bacterial transformation. Avoiding the RM system is the key factor in achieving transformation in bacteria, including *R. josui*, meaning that it is essential to identify and characterize the restriction endonuclease in *R. josui* to overcome the restriction barriers. To achieve successful transformation, various strategies have been examined, including: 1) disruption of restriction endonuclease genes (103, 104) in the genome using the ClosTron system (105) to accept unmethylated plasmids, 2) the use of plasmids with no, or fewer, restriction sites recognized by restriction enzymes in the host to avoid plasmid degradation (106, 107), 3) temporary inactivation of the RM system by heating (108) to facilitate transformation, 4) methylation of plasmids prior to transformation to enhance the transformation efficiency (42, 45, 109).

Restriction enzyme *Rjo*I from *R. josui* was purified by column chromatography from a 4-liter culture. The molecular size of *Rjo*I was expected to be around 30 kDa from that of *Dpn*I (molecular weight: 29,785) (NCBI Reference Sequence: WP_000418960.1). However, a protein band was not identified corresponding to *Rjo*I on the SDS-PAGE gel, although N-terminal amino acid sequences of some of the proteins were successfully determined in the gel, suggesting that the expression of *rjoI* was quite low. However, *Rjo*I recognizes were determined, where the A nucleotide is methylated by Dam methylase the sequence 5′ -G^{met}ATC-3′ , and cleaves the sequence between the A and T nucleotides (5′ -G^{met}A ↓ TC-3′). These findings confirmed that *Rjo*I is an isoschizomer of *Dpn*I (99), and that *R. josui* does not contain a Dam methylase. In Dam-defective isolates of *Neisseria meningitidis*, *dam* is replaced with *drg* (*dam*-replacing gene), which is flanked by pseudo-transposable small repeated elements. *drg* encodes restriction endonuclease *Nme*BII, an isoschizomer of *Dpn*I (110). The replacement of *dam* with *drg* was related to phase variation via a slippage-like mechanism in some isolates of *N. meningitidis* (110, 111). Although *Rjo*I is also an isoschizomer of *Dpn*I, it is not known whether *rjoI* replaced *dam* in *R. josui*. *R. josui* is taxonomically similar to *R. cellulolyticum* (60), and they have quite similar cellulolytic systems (63). However, the two species have divergent RM systems, with *R. cellulolyticum* shown to contain an isoschizomer of *Msp*I that recognizes the sequence CCGG. Dcm methylation was also reported to decrease the transformation efficiency of some plasmids in *R. thermocellum* (52) and *Clostridium pasteurianum* (112). Thus, all of the plasmids used for transformation of *R. josui* in the current study were prepared from *E. coli* C2925 (*dam*$^{-}$/*dcm*$^{-}$).

As we previously obtained good results using *E. coli*/*C. perfringens* shuttle vector pJIR751 for the transformation of *Clostridium paraputrificum* (46), we used pKKM801, a pJIR751-based plasmid containing the replication origin of *C. perfringens* plasmid pIP404 (96), for transformation of *R. josui*. Because pJIR751 does not contain ampicillin resistance marker *bla*, but does contain two M13rev sequencing primer sites that cannot be used for sequencing of DNA fragments inserted within the MCS, we connected the pIP404 region and pUC18-Pcip-His to yield pKKM801 (Fig. 12). pKKM801 was successfully introduced into *R. josui* at a transformation efficiency of 6.6×10^3 transformants/μg of DNA. Previous reports have shown that plasmids derived from *Enterococcus faecalis* (pAMβ1), *Bacillus subtilis* (pIM13), and *Clostridium butyricum* (pCB1), but not from *C. perfringens* (pIP404), could be successfully transformed into *R. cellulolyticum* (44). Therefore, replication origins other than pIP404 may function efficiently in *R. josui*. The copy number of pKKM801 in *R. josui* was determined to be 14 by qRT-PCR analysis, suggesting that *oriCP* can stably maintain the plasmid in *R. josui*.

Because the *cipA* scaffolding protein-encoding gene is known to be highly expressed in *R. josui* (60), the *cip* promoter (Pcip) was selected during the construction of pKKM801 for the expression of foreign genes. To simplify the purification of recombinant proteins, a 6×His-tag-encoding sequence and multiple cloning sites containing *Nco*I and *Not*I were inserted downstream of Pcip. *cel48A* was chosen for expression in *R. josui* as family 48 cellulases of glycoside hydrolases (32) are recognized as one of the most important enzymes in bacterial cellulolytic systems (89, 113, 114), and are usually a major constituent of bacterial cellulosomes (60, 115, 116). Recombinant Cel48A containing a dockerin module would be integrated into the cellulosome, making its purification difficult.

Therefore, the dockerin-encoding region was removed from *cel48A* for the expression of Cel48A in *R. josui*. 6×His-tagged Cel48AΔdoc was attempted to purify from the culture supernatant using a HisTrap HP column but were unsuccessful (data not shown), probably because the 6×His-tag was not fully exposed to the surface of the recombinant protein. However, we detected 6×His-tagged protein in the culture supernatant by immunological analysis using anti-His-tag monoclonal antibody (Fig. 17). The productivity of Cel48AΔdoc was roughly estimated by comparing the expression levels of Cel48AΔdoc and the purified 6×His-tagged protein RjCE6 (unpublished protein), i.e., about 1 μg per 100 ml of the culture supernatant. These results indicate that Cel48AΔdoc is secreted into the culture medium, but that expression levels are low under the control of Pcip. In recent years, several other plasmids have been constructed and used for CBP studies in *R. cellulolyticum* (90), *C. tyrobutyricum* (91, 117), *C. cellulovorans* (42), and *C. acetobutylicum* (92). In these studies, the strong constitutive promoter of the *C. acetobutylicum* thiolase gene was used for the expression of cellulase genes or genes responsible for n-butanol production. Use of a strong constitutive promoter from *R. josui* may therefore be desirable for high expression of foreign genes. As *R. josui* produces ethanol and acetic acid as fermentation products (61), a promoter for one of the genes responsible for ethanol or acetic acid production may be a good candidate as a strong promoter for construction of an improved expression vector.

In conclusion, the basic technology required for molecular breeding of *R. josui* has been established in this study. This host-vector system will help to better understand the cellulolytic system of *R. josui*, and in the development of CBP technologies using this bacterium.

Chapter 4: General discussion

A novel hemicellulase Abf62A-Axe6A from *Ruminiclostridium josui*

In the past 30 years, this laboratory has made significant progress in the characterization of biomass-degrading enzymes and cellulosome of *R. josui*. After the bacterium was isolated in 1988 (61), endo-β-1,4-glucanases (118-120), an α-galactosidase (121), xylanases (24, 29), a mannanase (63), an arabinoxylan arabinofuranohyrolase (23), an arabinofuranosidase (22), a scaffolding protein (24, 60), some isolated CBMs (24, 29, 122-124), and the structure and function of the cellulosome (23, 24, 122, 125-127) were analyzed. The recognition of the importance of hemicellulases and increasing information about the degradation of natural lignocellulose by hemicellulases may remove the requirement of the pretreatment and detoxification processes during the bioconversion process (128).

RjAbf62A-Axe6A contains two different catalytic modules, GH62 and CE6, and its activities were determined toward various substrates (Tables 2 and 3). Although RjAbf62A-Axe6A showed higher activity toward soluble rye and wheat arabinoxylans than RjAbf62A, the activity of the former was lower than that of the latter when the substrate was insoluble wheat arabinoxylan. The main chain of the insoluble wheat arabinoxylan is substituted singly or doubly by acetyl groups. It is possible, therefore, that the CE6 module in this enzyme has hindered the access of the GH62 module to the substrate, and the CE6 module of RjAbf62A-Axe6A is more active on this substrate than RjAxe6A (Table 3). Although characterizations of the bifunctional enzymes having acetyl xylan esterase and α-L-arabinofuranosidase activities have been reported (129, 130), there is no detailed study on synergistic actions between these two modules. A number of bi/tri-functional hemicellulases have been reported, that is, β-D-xylosidase/α-L-arabinofuranosidase (131-

133), xylanase/β-D-xylosidase/α-L-arabinofuranosidase (134), xylanase/α-L-arabinofuranosidase (135) and xylanase/ferulic acid esterase (136), xylanase/arabinofuranosidase/acetyl xylan esterase (129), acetyl xylan esterase/α-L-arabinofuranosidase (130). In the directed mutagenesis study of xylanase/ferulic acid esterase (Xyn10D-Fae1A), the ferulic acid esterase activity was increased three-fold when a putative catalytic site (Glu) of the xylanase domain was changed to Ser, suggesting that the two domains were functionally coupled (136).

RjAbf62A-Axe6A and RjAbf62A-CBM6 have both AXHm2, 3 and AXHd3 activities judging from their ability to release α-1,2- and α-1,3-linked arabinofuranose from doubly substituted xylosyl residues. Furthermore, no intermediate compound containing arabinose residues was detected in the reaction mixture when $A^{2,3}XX$ was used as the substrate (Fig. 10). It was reported that some GH51 ABFs (78-80) with AXH m2, 3 activity could release arabinose from doubly substituted xylosyl residues located at the non-reducing end. RjAbf62A-Axe6A has substrate specificity similar to such GH51 ABFs and is the first GH62 ABF capable of releasing α-1,2- and α-1,3-linked arabinofuranose from doubly substituted xylosyl residues.

It is surprising that RjAbf62A-Axe6A displayed an endoxylanase activity toward arabinoxylans and birchwood xylan, that is, the enzyme released X3, X4, X5 and other xylooligosaccharides from the substrates (Fig. 6). However, it is puzzling that RjAbf62A-CBM6 does not release X3, X4, X5 and xylooligosaccharides even after prolonged incubation (Fig. 6). RjAbf62A-Axe6A and RjAbf62A-CBM6 were also able to hydrolyze xylooligosaccharides efficiently (Fig. 9). RjAbf62A-Axe6A and RjAbf62A-CBM6 hydrolyzed X5 to form X2 and X3 as major products, confirming that RjAbf62A has an

apparent endo-type activity. Furthermore, RjAbf62A-Axe6A was active on arabinooligosaccharides. The substrate specificity is similar to the broad specificity of Axy43A, consisting of a GH43 module and a CBM6, from *Paenibacillus curdlanolyticus* strain B-6 (134). RjAbf62A-Axe6A and RjAbf62A-CBM6 showed not only different substrate preference as α-L-arabinofuranosidase on different substrates but also different behavior on xylans. The steric hindrance may cause difference in ABF activity between RjAbf62A-Axe6A and RjAbf62A-CBM6 due to the presence of the CE6 module. Furthermore, difference in AXE activity between RjAbf62A-Axe6A and RjAxe6A may be ascribed to the presence or absence of CBM6 in addition to steric hindrance. Unfortunately, relation between the GH62 and CE6 modules in a modular enzyme RjAbf62A-Axe6A remains to be disclosed in this study.

Since RjAbf62A-Axe6A contains a dockerin, it is expected to be a member of catalytic subunits to form an enzyme complex, by its binding strongly to non-catalytic scaffolding protein via interaction of a dockerin module and one of conserved cohesion modules. There are a large number of reports on cellulolytic complexes (cellulosomes) (14, 137-139), but there are only a few reports on xylanolytic complexes (xylanosomes) (23, 24, 140, 141). Recent reports on xylanolytic complexes have shown that the formation of complexes by a variety of different hemicellulases could promote the degradation of hemicellulose in the plant cell wall. Enzymatic hydrolysis tests were performed on wheat arabinoxylan or defatted corn bran using three different bacterial hemicellulases combined into a xylanolytic complex. The results showed that the xylanolytic complex consisting of a xylanase and a bifunctional arabinofuranosidase/xylosidase increased the amount of soluble sugars released from arabinoxylan by 30% compared to the free enzyme combination.

Moreover, the xylanolytic complex consisting of a xylanase and a ferulic acid esterase releases 15% to 20% more ferulic acid from wheat arabinoxylan than the free enzyme combination (142). The synergistic action of the fungal xylanolytic complex formed on a chimeric scaffolding protein toward birchwood xylan was also investigated (140). The complex consisting of a xylanase, β-xylosidase, acetylxylan esterase and the chimeric scaffolding protein (without CBM) showed a 1.6-fold increase in hydrolytic activity as compared to the free enzyme mixture. In addition, the hydrolytic activity of the xylanolytic complex formed on the chimeric scaffolding protein containing xylan binding module increased by 3.3 times compared to the free enzyme combination. Effect of this xylan binding module on xylan hydrolysis was much higher than that of *R. thermocellum* CBM3 (140). However, hemicellulases in the natural cellulosome often contain a CBM with an affinity for xylan (23, 24) and the scaffolding protein usually contains a CBM3. Many studies show that both types of CBMs have unique functions, including polysaccharide preference. Therefore, the effects of these two types of CBMs in the same complex on the hydrolysis of cellulose and hemicellulose remain to be studied. The rod-like cellulose component of the plant cell wall is embedded in a mixture of hemicellulose (Fig. 1), and the main component of hemicellulose is xylan. It has been reported that the addition of xylanolytic complex to the cellulolytic complex facilitates the breakdown of intractable natural cellulosic substrates (141). Due to the complexity of the structure of hemicellulose in natural substrates and the diversity of hemicellulases, there is still much to be understood about the xylanolytic complex.

Host-vector system of *Ruminiclostridium josui*

The construction of a host-vector expression system in *Ruminiclostridium* species not only helps to establish and improve CBP, but also helps to characterize the bacterial genetic systems (48, 58, 59, 89, 143-145). The RM system in bacteria is a defense against foreign DNA and a restrictive barrier when developing genetic manipulation systems. Various strategies have been examined to achieve successful bacterial transformation. *Clostridium acetobutylicum* ATCC 824 contains a restriction enzyme, *Cac*824I, whose recognition sequence is 5'-GCNGC-3', where N can be any nucleotide. The electrotransformation of *C. acetobutylicum* ATCC 824 was successfully achieved by methylation of plasmid DNA by the ϕ3TI methyltransferase in a recombinant *E. coli* strain to protect the plasmid DNA from *Cac*824I (146). *C. acetobutylicum* DSM1731 also had *Cac*824I and its gene was disrupted by the ClosTron system based on group II intron insertion. The resulting strain that lost type II REase activity was successfully transformed with un-methylated DNA (104). For *C. acetobutylicum* SA-1, the bacterial cells were heated at 55 °C for 15 min to inactivate the DNase activity before transformation (108). For overcoming the restrictions by the RM system, however, most efforts had been made to protect the recognition sequences on plasmids by their methylation and disruption of RE genes in the genome using the ClosTron system.

In this study, it was shown that *Rjo*I from *R. josui* recognized the sequence 5'-G^{met}ATC-3', where the nucleotide A was methylated by Dam methylase, and cleaved the sequence between the A and T nucleotides (5'-GmetA↓TC-3'). These findings confirmed that *Rjo*I was an isoschizomer of *Dpn*I [110] and that *R. josui* did not contain a Dam methylase. As the battle between bacteria and their pathogens escalates, the evolution of

modification-dependent restriction enzymes was promoted, that is, unmodified DNA is recognized and treated as "self" and the modified DNA as "non-self" (147). *Dpn*I and *Rjo*I belong to the type IIM (where M represent modification-dependent) in the current nuclease classification (36). Dcm methylation was also reported to decrease the transformation efficiency of some plasmids in *R. thermocellum* [57] and *Clostridium pasteurianum* (123). Thus, all of the plasmids used for transformation of *R. josui* in the current study were prepared from *E. coli* C2925 (*dam*$^-$/*dcm*$^-$).

According to the transformation procedure for *C. paraputrificum*, *E. coli/C. perfringens* shuttle vector plasmid pJIR751 was used for transformation of *R. josui* (46). An expression vector plasmid pKKM801 was also constructed from plasmids pJIR751 and pUC18 (Fig. 12) and successfully introduced into *R. josui* with a transformation efficiency of 6.6×10^3 transformants/μg DNA. Previously, *Enterococcus faecalis* plasmid pAMβ1, *Bacillus subtilis* pIM13, and *Clostridium butyricum* pCB1 but not *C. perfringens* pIP404 could successfully transform *R. cellulolyticum* (44). Therefore, *oriCP* in the *Clostridium-Escherichia coli* shuttle vector pJIR751 may not be suitable for transformation of *R. josui*, although the copy number of pKKM801 in *R. josui* was determined to be 14 by qRT-PCR analysis, suggesting that *oriCP* can stably maintain the plasmid in *R. josui*. Selection of the replication origin may increase transformation efficiency, separation stability, copy number and expression level in *R. josui* (44).

Since the *cipA* gene encoding the scaffolding protein is known to be highly expressed in *R. josui* (60), and the *cip* promoters using the scaffold protein-encoding gene from *C. thermocellum* have been reported (102), the *cipA* promoter (Pcip) was chose to express the foreign gene during the construction of pKKM801 (Fig. 12). The recombinant

protein RjCel48AΔdocwas successfully expressed in *R. josui*, and we roughly estimated the yield of RjCel48AΔdoc by the expression level of purified 6×His-tagged protein RjCE6 (unpublished protein), i.e., about 1 μg per 100 ml of the culture supernatant. These results indicated that RjCel48AΔdoc was secreted into the culture medium but with a low expression level under the control of Pcip. In the evaluation of 17 promoters from *R. thermocellum*, clo1313_2638, clo1313_2926, clo1313_966 and clo1313_815 promoters showed reliable gene expression in *R. thermocellum* (47), which is comparable to the expression level caused by the *gapDH*, *cbp* and *eno* promoters commonly used for *R. thermocellum* gene expression. The *gapDH* gene encodes glyceraldehyde-3-phosphate dehydrogenase, which is one of the most highly expressed proteins in the *R. thermocellum* proteome (48). Other promoters from which high levels of gene expression were reported include: the thiolase (*thl*) gene promoter in *R. cellulolyticum* H10 (103), aldehyde/alcohol dehydrogenase 2 (*adhE2*) in *C. tyrobutyricum* (148), ferredoxin (*fdx*) in *C. pasteurianum* (55) and *Clostridium sporogenes* (149), and pyruvate:ferredoxin oxidoreductase (*pfo*) in *Lachnoclostridium phytofermentans* (150). Therefore, for high expression of foreign genes, it may be necessary to use a strong constitutive promoter of *R. josui*. Since *R. josui* produces ethanol and acetic acid as major fermentation products, the promoter of one of the genes responsible for ethanol or acetic acid production may be a good candidate for a strong promoter for the construction of improved expression vectors. In *R. cellulolyticum* H10, expressed was the heterologous oxygen-independent green fluorescent protein PpFbFPm (103), which can be conveniently used for characterization of protein expression, localization, movement and interaction in cellulolytic bacteria in future studies.

Moreover, the relationship between gut microbiota and human health has been increasingly recognized in recent years (151). These gut microbes acquire energy from plant biomass, which cannot be degraded by mammalian digestive enzymes. Therefore, by studying the degradation of plant biomass in the gut, we can better understand the relationship between the gut microbes and health (152-154).

The prospect of biomass utilization in the future

The development of biomass utilization technology is mainly to improve the energy supply and the environment, and to promote the development of rural areas, increasing the income level of farmers and regulating the fluctuation of supply and demand of food (155).

In two primary types of microbial cellulolytic systems, the “free” enzyme type and the “cellulosome” type, the fungal cellulose decomposition system is the best representative of the “free” enzyme type. The high productivity and significant activity of the fungal system can be well applied in industrial biorefinery applications (156). In the past few decades, the understanding of cellulases has made significant progress, especially in the past 30 years studies in structural biology and basic mechanism of cellulase action have been deeply understood. However, some cellulolytic mechanisms have not been fully understood, and some of the recently discovered cellulolytic mechanisms of lytic polysaccharide monooxygenases (LPMOs) under the aerobic condition are still unclear (157). In addition, the molecular basis of the difference in enzyme activities in the same cellulase family and the synergistic function of the enzyme mixture in cellulose deconstruction is not clear. Current understanding of cellulose deconstruction is based on purified cellulose, but little is known about cellulose deconstruction in natural or pretreated

plant cell walls. With an understanding of fine structure of the plant cell walls, it has been recognized that the degradation of lignin and hemicellulose is also significant in the degradation of natural or pretreated plant cell walls (10, 13), because the cellulose is surrounded by a layer of hemicellulose (11). Due to the many uncertainties in the complex structure of xylans from different origins (158), the study of its structure will undoubtedly stimulate the search for new enzymes with distinct substrate specificities. As more and more studies on hemicellulases and plant cell wall biosynthesis are conducted, a significant understanding of the fine structure of xylans will be achieved (10). In addition, fine studies of the regulation of glycoside hydrolase gene expression in the fungal system can provide new strategies for breeding of industrial fungi, which in turn can improve the production of biomass-degrading enzymes (159).

Furthermore, anaerobic fermentative bacteria including *R. josui* having the cellulosome system are promising candidates for CBP (155). The mechanisms of lignocellulose decomposition into fermentable sugars between cellulosome system and non-complexed fungal cellulase system are significantly different (13). Moreover, the synergistic mechanism between the two types of enzymes, cellulases and hemicellulases, in the complex is still unclear (141). Furthermore, new types of cellulosome-producing bacteria in various life space systems, including terrestrial, aquatic and marine ecosystems, and vertebrate and invertebrate microbiota should be further studied (14). Recently, cellulolytic fiber-degrading bacterium *Ruminococcus champanellensis* has been isolated from human stool samples (152). The cellulosome system of the human gut microbiome *R. champanellensis* is very similar to the rumen bacterium *Ruminococcus flavefaciens*. Therefore, research on the distribution of the human gut microbiome and their relationship

to health and disease is necessary. According to recent researches, gut microbiomes are inextricably linked to human health and disease, and gut microbiomes are inextricably linked to their degradation of biomass as the energy (154).

In CBP, sugars produced by the decomposition of biomass is simultaneously utilized by fermentation of microorganisms. Therefore, the understanding of the metabolic mechanisms in various bacteria involved in CBP and the construction of genetically engineered bacteria by different strategies are particularly important (160).

Figures and Tables

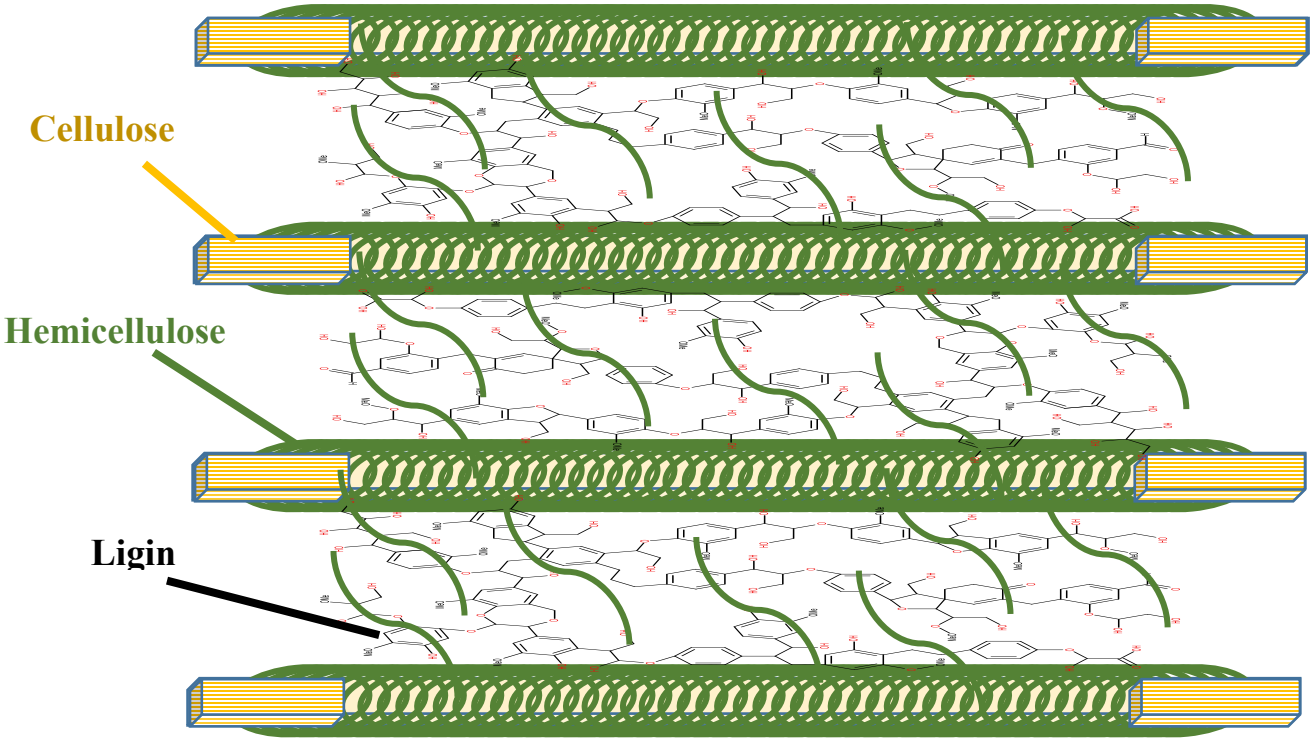


Fig. 1. Spatial arrangement of cellulose, hemicellulose, and lignin in cell wall (11).

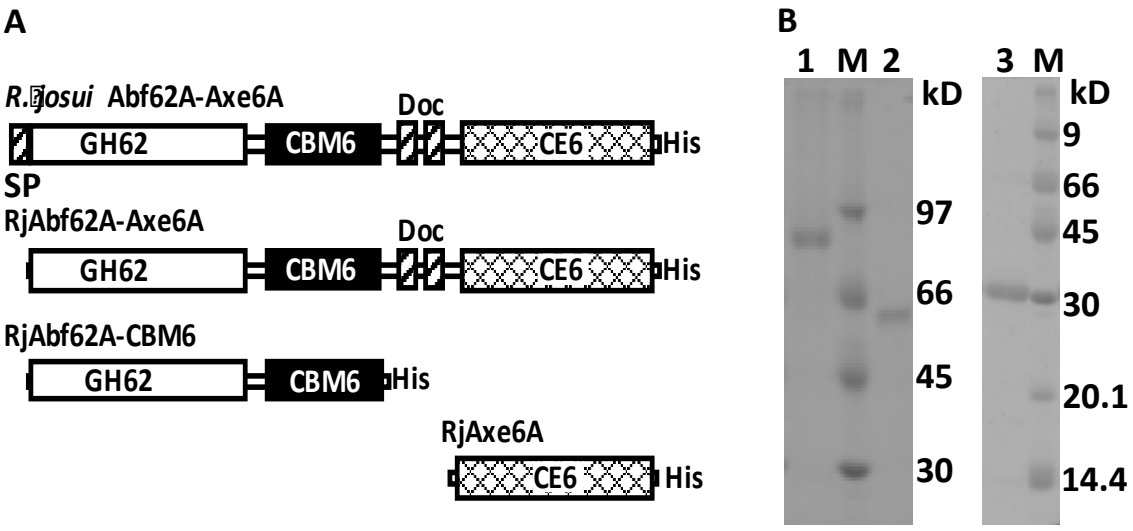


Fig. 2. Module organization of *R. josui* Abf62A-Axe6A and its derivatives used in this study (A) and SDS-PAGE of purified proteins (B). In (A), GH62, family-62 glycoside hydrolase module; CBM6, family-6 CBM; Doc, dockerin module; CE6, family-6 carbohydrate esterase module; His, 6×His tag. In (B), lane 1, RjAbf62A-Axe6A; lane 2, RjAbf62A-CBM6; lane 3, RjAxe6A; lane M, standard proteins.

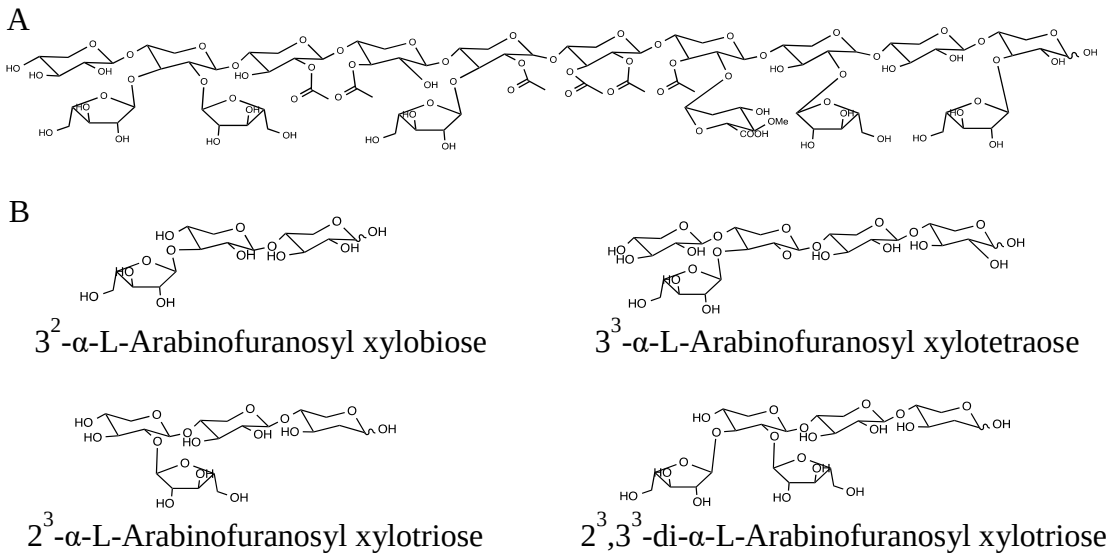


Fig. 3. Scheme of graminaceous xylan (A) and arabinoxylooligosaccharides (Megazyme) used in this study (B).

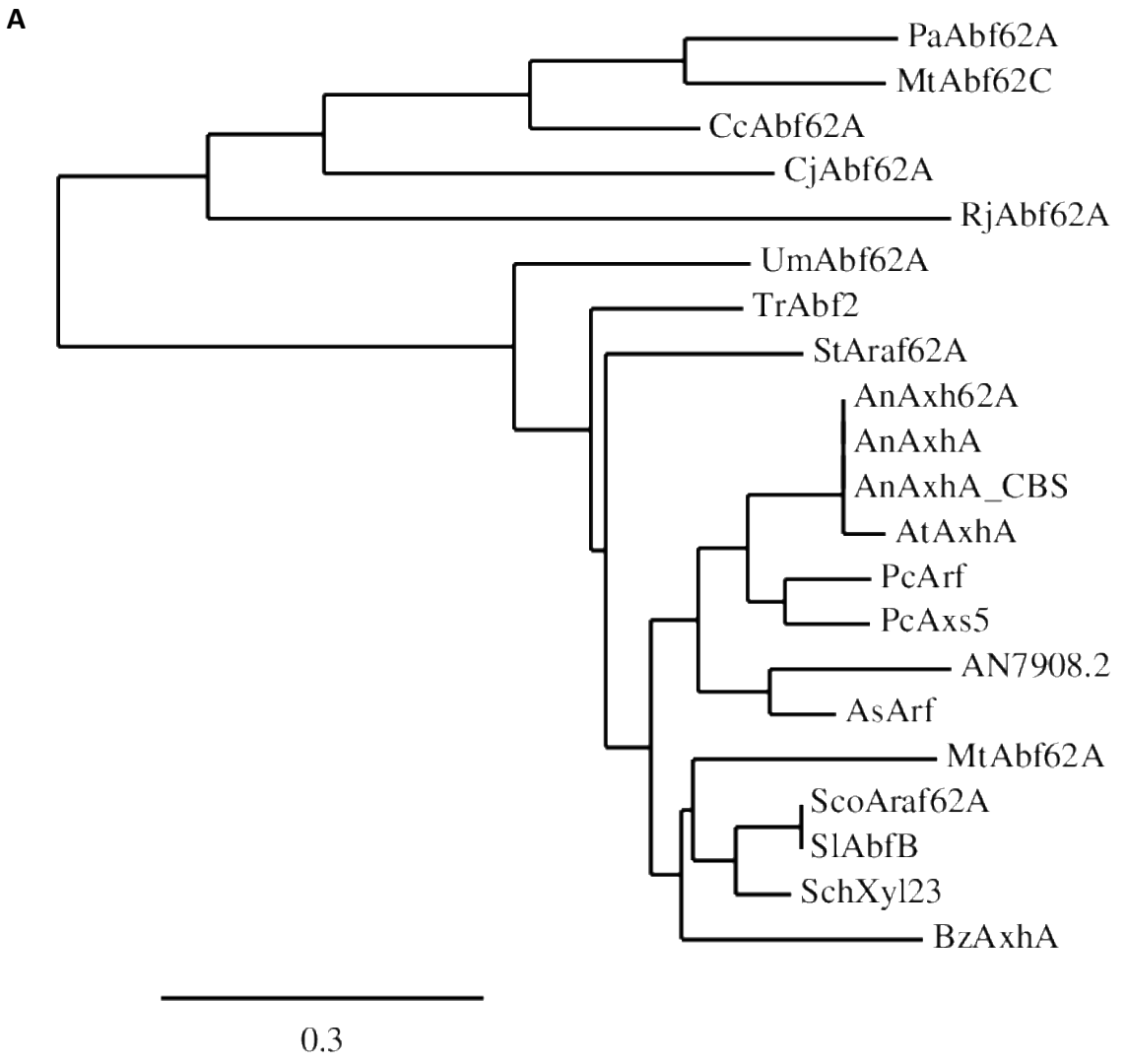

A
PaAbf62A
MtAbf62C
CcAbf62A
CjAbf62A
RjAbf62A
UmAbf62A
TrAbf2
StAraf62A
AnAxh62A
AnAxhA
AnAxhA_CBS
AtAxhA
PcArf
PcAxs5
AN7908.2
AsArf
MtAbf62A
ScoAraf62A
SlAbfB
SchXyl23
BzAxhA
0.3
B
R.17
R.44
AcGH5
RjAxe6A
OspAxeA
NpBnaA
FsAxe6B
FsAxe6A
0.3

Fig. 4. Phylogenetic trees of some GH62 modules (A) and CE6 modules (B) in biochemically characterized enzymes. Phylogenetic tree was analyzed and drawn by using the MUSCLE analysis in Phylogeny.fr (http://www.phylogeny.fr/) (Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol 10:8.). In (A); AnAxh62A, *Aspergillus niger* Axh62A (AIC36734); AnAxhA, *A. niger* AxhA (CAA03158); AnAxhA_CBS, *A. niger* strain CBS AxhA (CAK38069); AN7908.2, *A. niger* hypothetical protein (EAA59562); AsArf, *Aspergillus sojae* Arf (BAA85252); AtAxhA, *Aspergillus tubingensis* AxhA (CAB01408); BzAxhA, *Bipolaris zeicola* AxhA (AF306763_1); CcAbf62A, *Coprinopsis cinerea* Abf62A (BAK14423); CjAbf62A, *Cellvibrio japonicus* Abf62A (ACE85320); MtAbf62A, *Mycothermus thermophilus* Abf62A (AHZ56660); MtAbf62C, *M. thermophilus* Abf62C (AHZ56658); PaAbf62A, *Podospora anserina* Abf62A (CAP62336); PcArf, *Penicillium capsulatum* Arf (CAM07245); PcAxs5, *Penicillium chrysogeum* (BAG71682); RjAbf62A, *R. josui* Abf62A-Axe6A (WP_024831741.1); ScXyl23, *Streptomyces chattanoogensis* Xyl23 (AF121864_1); ScoAraf62A, *Streptomyces coelicolor* Araf62A (CAA16189); ClAbfB, *Streptomyces lividans* AbfB (AAC26524); StAraf62A, *Streptomyces thermoviolaceus* Araf62A (BAD02381); TrAbf2, *Trichoderma reesei* Abf2 (AAP57750); UmAbf62A, *Ustilago maydis* Abf62A (XP_011391032). In (B); AcGH5, *Acetivibrio cellulolyticus* CBP (WP_010246447); FsAxe6A, *Fibrobacter succinogenes* Axe6A (A F180369_1); FsAxe6B, *F. succinogenes* Axe6B (ACX75360); NpBnaA, *Neocallimastix patriciarum* BnaA (AAB69090); OspAxeA, Orpinomyces sp. AxeA

(AAC14690); R.17, unidentified microorganisms (CAJ19109); R.44, unidentified microorganisms (CAJ19130); RjAxe6A, *R. josui* Abf62A-Axe6A (WP_024831741.1).

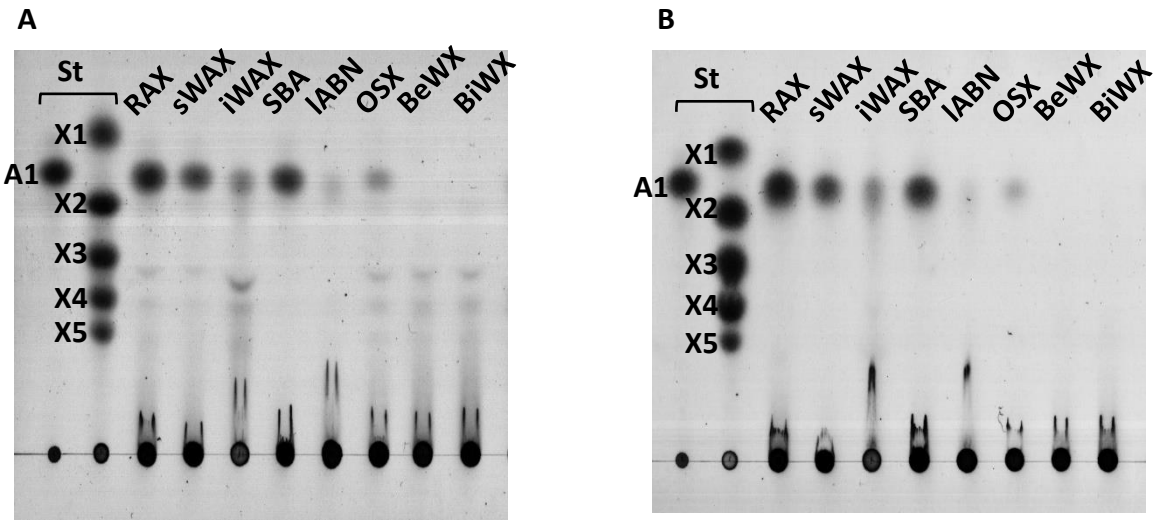


Fig. 5. TLC analysis of hydrolysis products from rye arabinoxylan (RAX), soluble wheat arabinoxylan (sWAX), insoluble wheat arabinoxylan (iWAX), sugar beet arabinan (SBA), linear arabinan (lABN), oat spelts xylan (OSX), beechwood xylan and birchwood xylan (BWX) by RjAbf62A-Axe6A (A) and RjAbf62A-CBM6 (B). Respective polysaccharides were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by TLC. A1, arabinose; X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose; Std, standard.

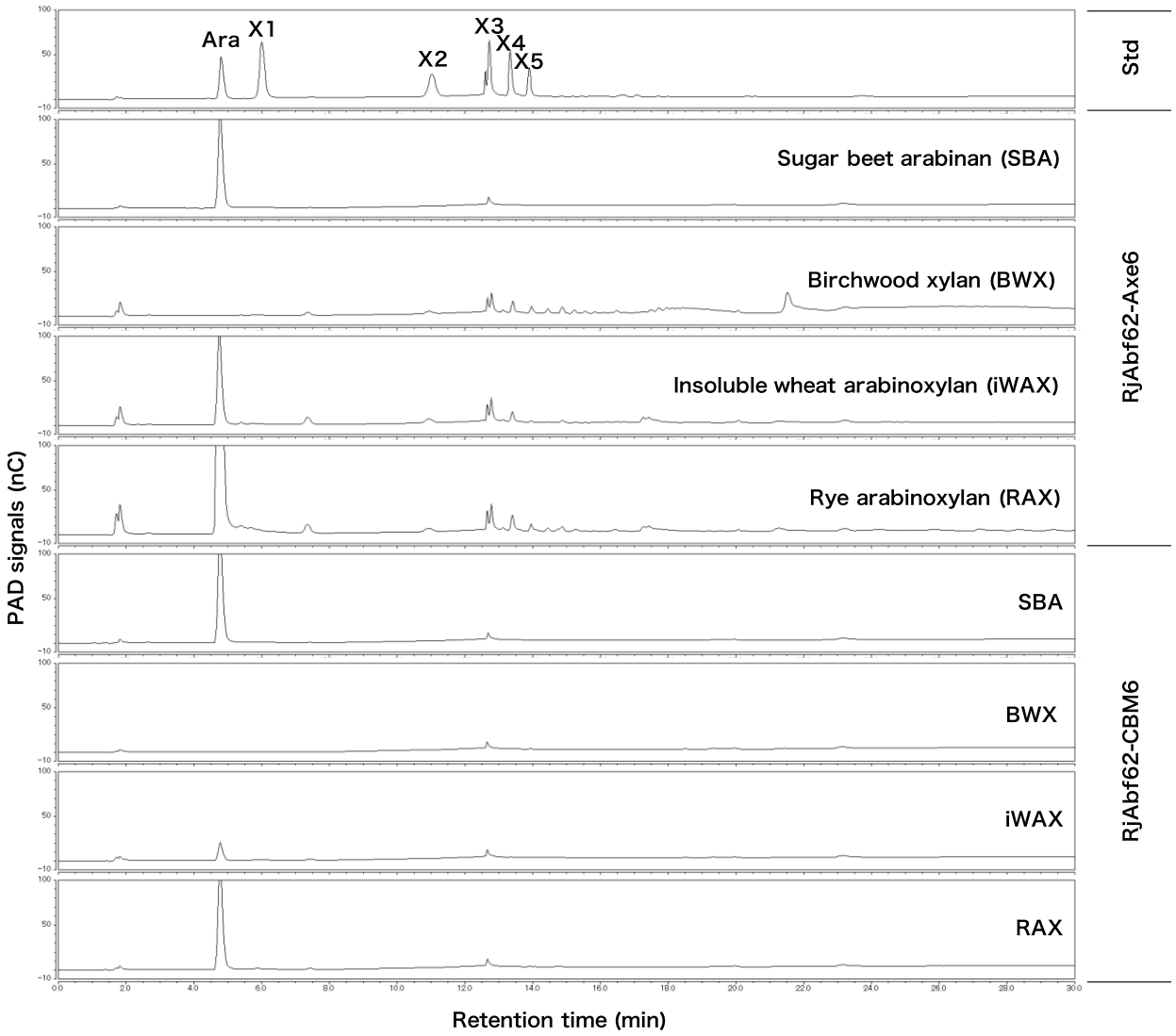


Fig. 6. HPAEC-PAD analysis of hydrolysis products from sugar beet arabinan, birchwood xylan, insoluble wheat arabinoxylan and rye arabinoxylan by RjAbf62A-Axe6A and RjAbf62A-CBM6. Respective polysaccharides were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by HPAEC-PAD. A1, arabinose; X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose.

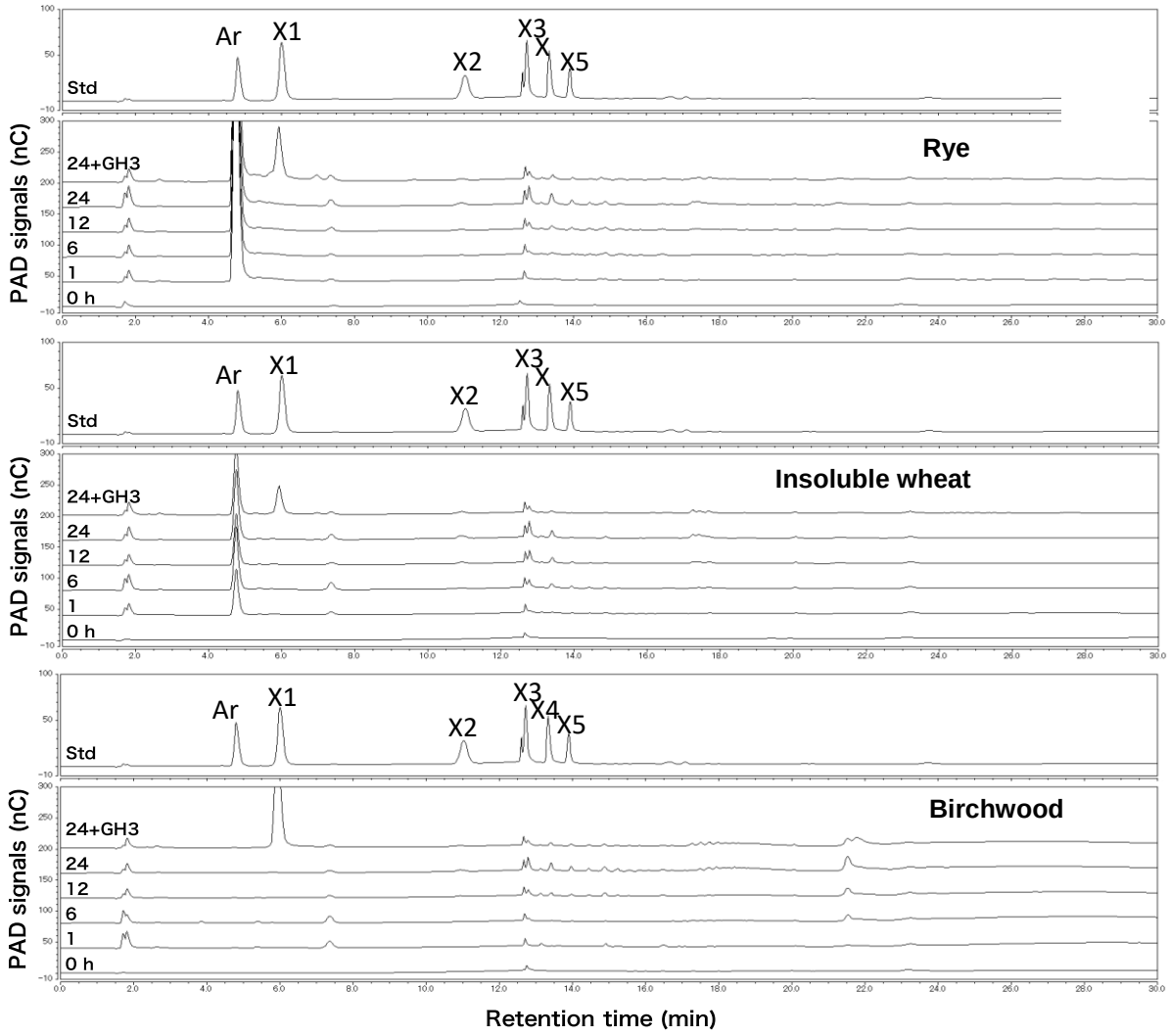


Fig. 7. HPAEC-PAD analysis of hydrolysis products from rye arabinoxylan, insoluble wheat arabinoxylan and birchwood xylan by RjAbf62A-Axe6A. Respective polysaccharides were incubated with RjAbf62A-Axe6A under conditions for indicated periods up to 24 h and hydrolysis products were analyzed by HPAEC-PAD. After 24-h incubation, GH3 β-xylosidase was added to the reaction mixture to degrade xylooligosaccharides (24 + GH3). A1, arabinose; X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose.

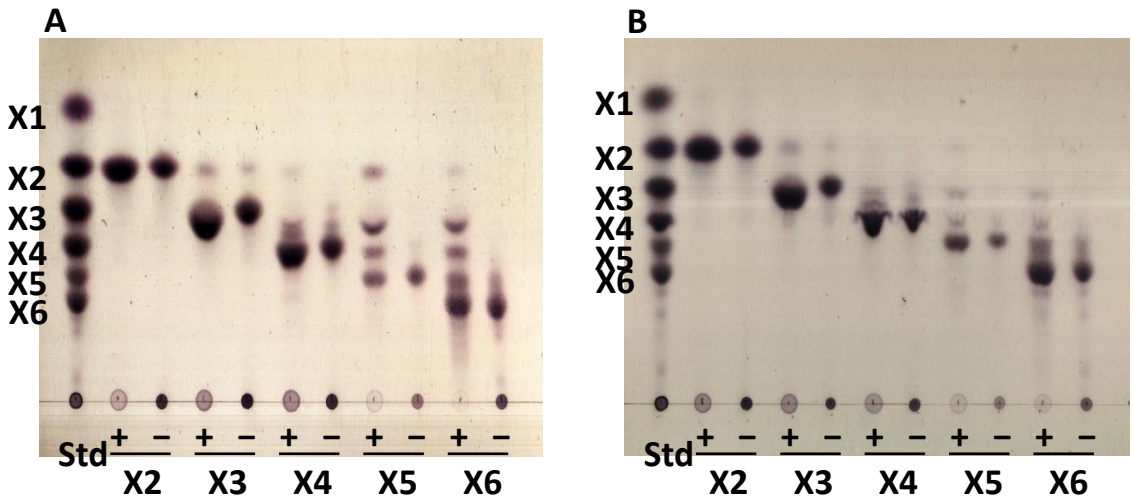


Fig. 8. TLC analysis of hydrolysis products from xylooligosaccharides by RjAbf62A-Axe6A (A) and RjAbf62A-CBM6 (B). Respective xylooligosaccharides (X2 – X6) were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by TLC. X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose; -, incubated without enzymes; +, incubated with enzymes; Std, standard.

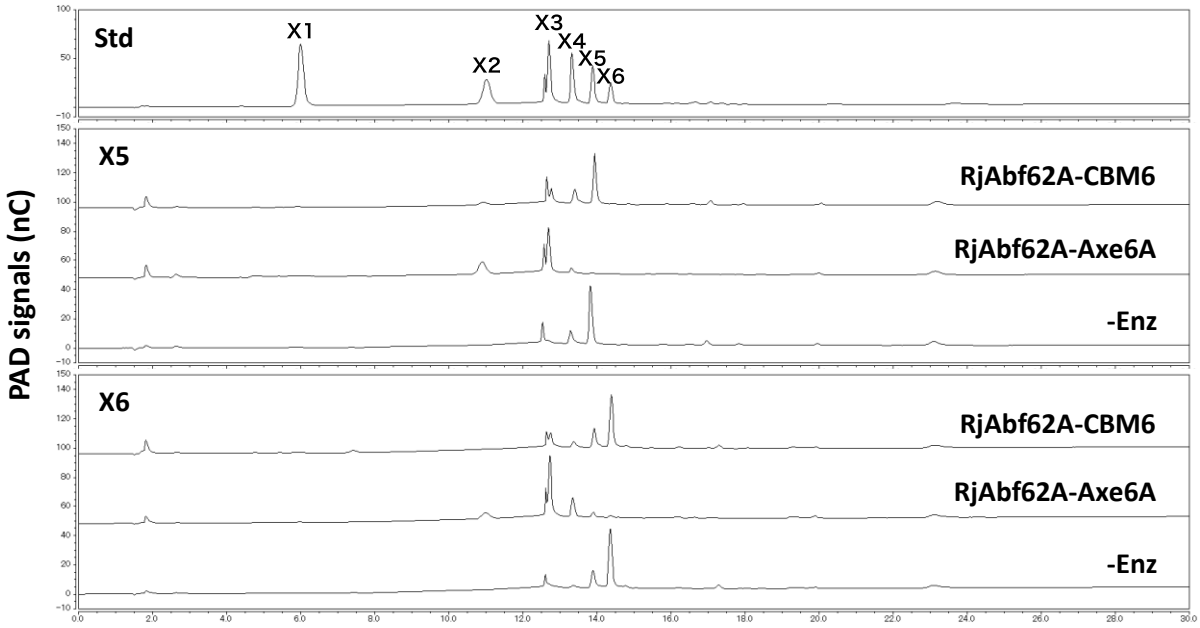


Fig. 9. HPAEC-PAD analysis of hydrolysis products from xylopentaose and xylohexaose by RjAbf62A-Axe6A and RjAbf62A-CBM6. Xylopentaose and xylohexaose were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by HPAEC-PAD. X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose; X5, xylohexaose; -Enz, incubated without enzymes; +, incubated with enzymes; Std, standard.

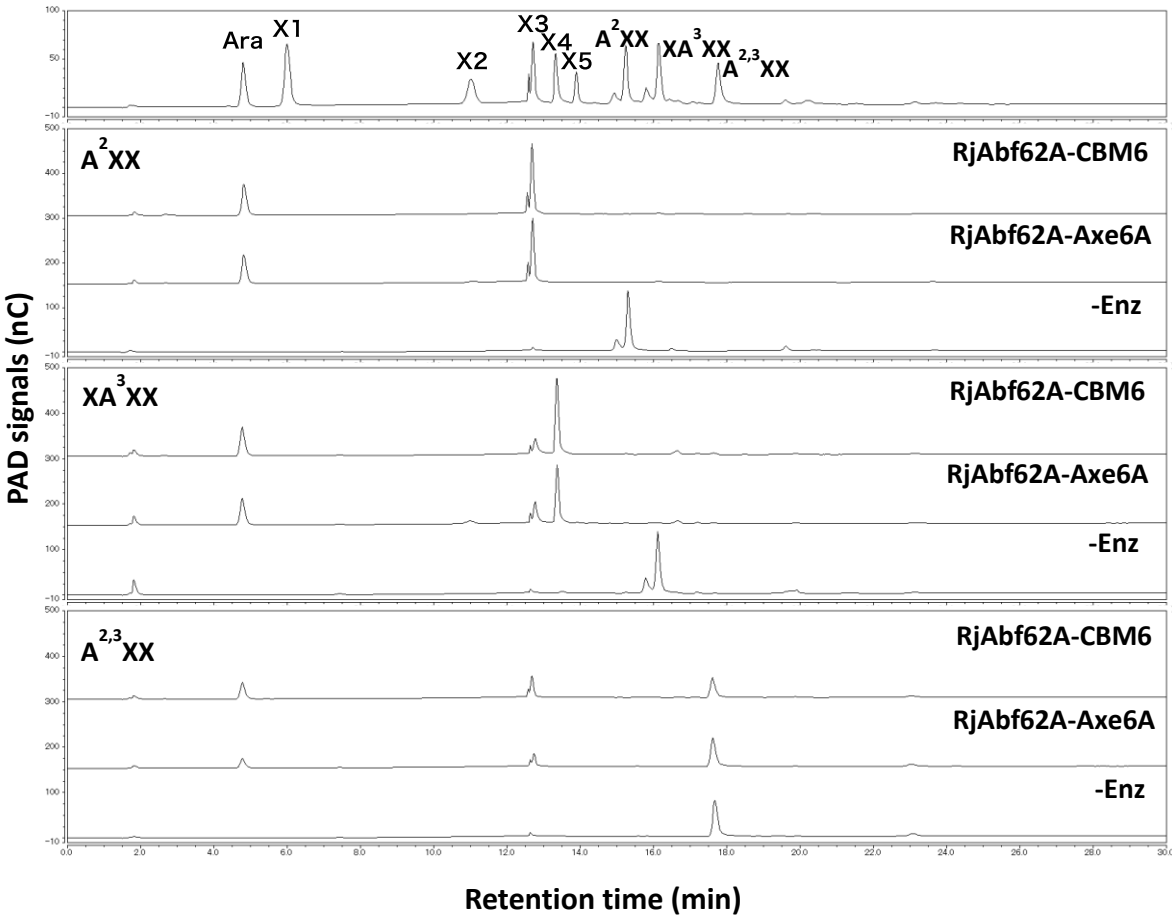


Fig. 10. HPAEC-PAD analysis of hydrolysis products from arabinoxylooligosaccharides by RjAbf62A-Axe6A. Respective xylooligosaccharides (A^2XX, XA^3XX and $A^{2,3}XX$) were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by HPLC. A1, arabinose; X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose; A^2XX, 2^3-α-L-arabinofuranosyl xylotriose; XA^3XX, 3^3-α-L-arabinofuranosyl xylotetraose; $A^{2,3}XX$, 2^3,3^3-di-α-L-arabinofuranosyl xylotriose; -Enz, without enzyme (control).

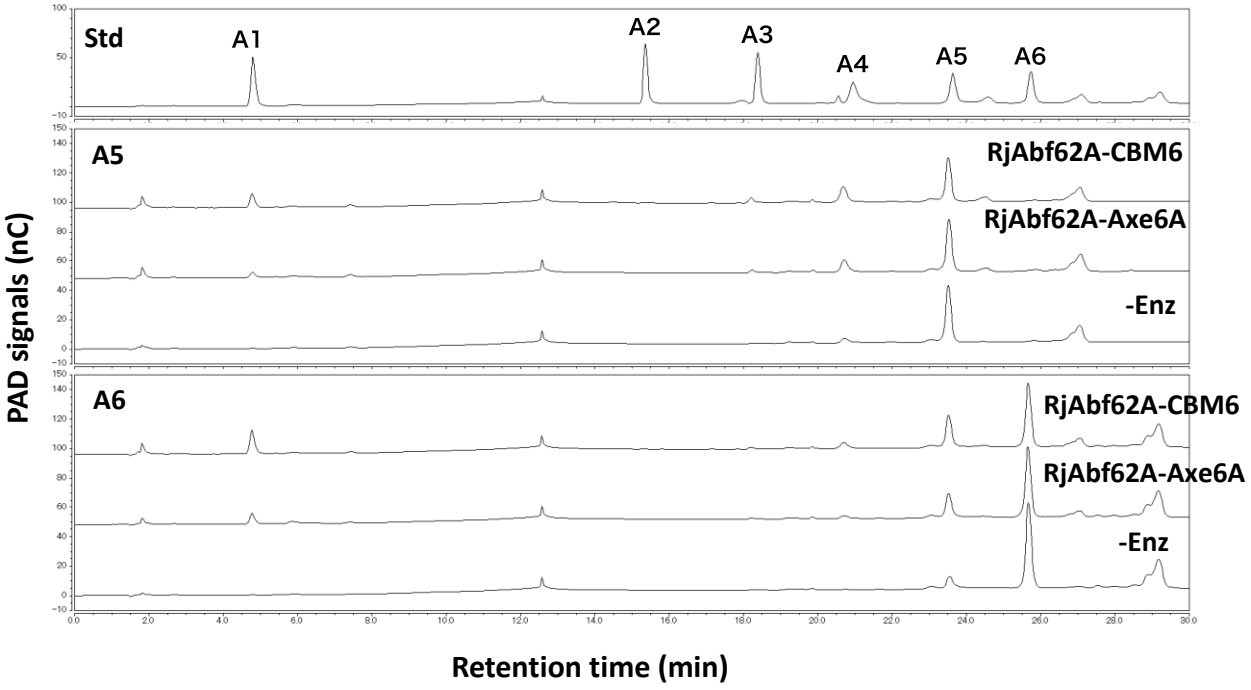
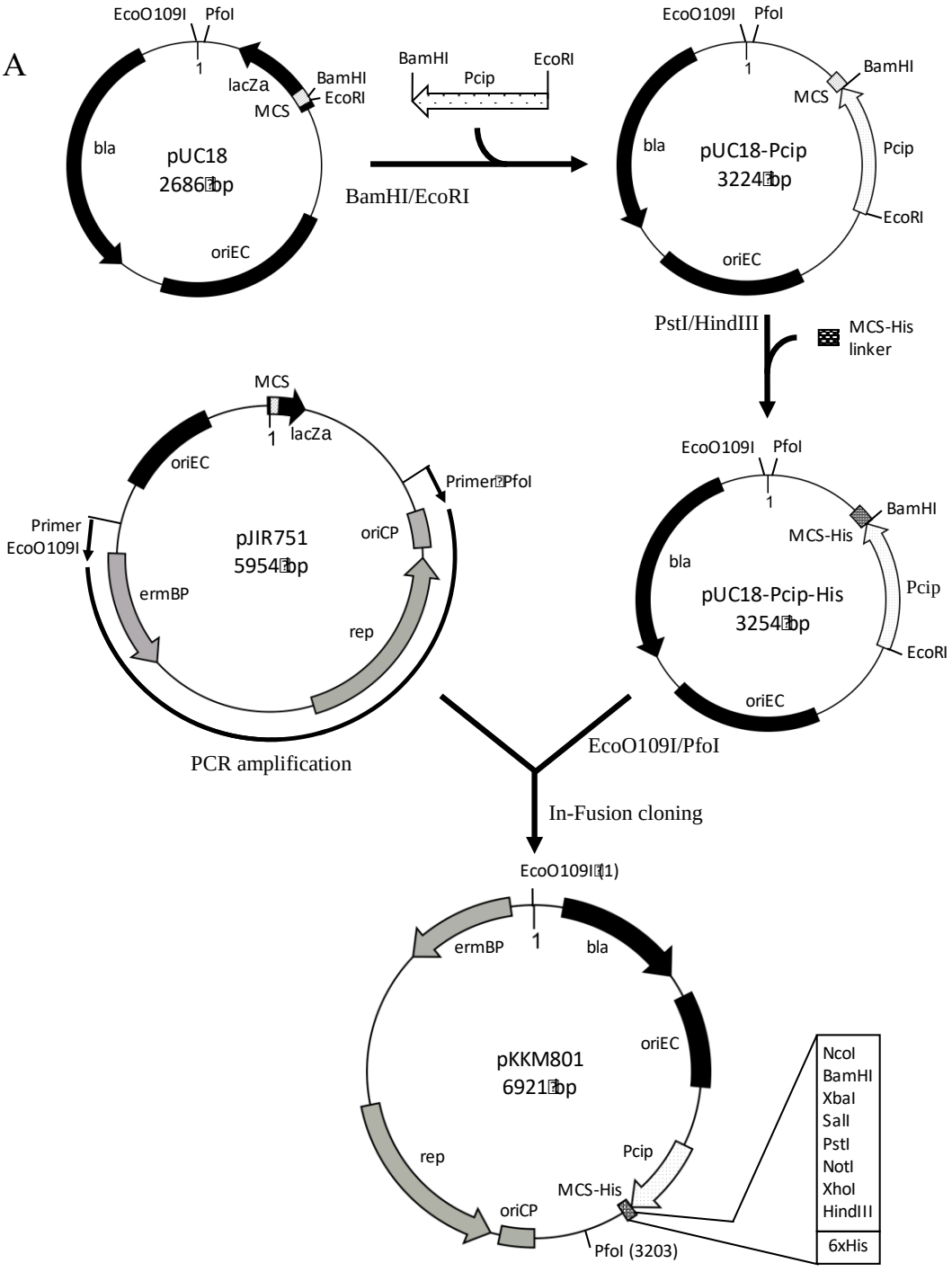


Fig. 11. HPAEC-PAD analysis of hydrolysis products from arabinooligosaccharides by RjAbf62A-Axe6A or RjAbf62A-CBM6. Respective arabinooligosaccharides (A5 and A6) were incubated with RjAbf62A-Axe6A or RjAbf62A-CBM6 under conditions described in the text and hydrolysis products were analyzed by HPAEC-PAD. A1, arabinose; A2, arabinobiose; A3, arabinotriose; A5, arabinotetraose; A5, arabinopentaose; A6, arabinohexaose, Std, standard.

A



B

2761 SD NcoI BamHI XbaI SalI PstI NotI XhoI

AAAT**AGGAGG**TTTACCCATGGCGGGGATCCTCTAGAGTCGACCTGCAGGCGGCCGCACTC

M A G

2860

GAGCACCACCACCACCACCACTGAAAGCTTGGCACTGGCC

E H H H H H H * HindIII

Fig. 12. Construction of pKKM801 (A) and nucleotide sequence of MCS-His-tag region (B). Plasmid construction was described under **Materials and methods**. pKKM801 consists of the replication origin *oriEC* and the β-lactamase gene *bla* derived from *E. coli* plasmid pUC18, the replication origin *oriCP*, replication protein gene *rep* and erythromycin-resistant gene *ermBP* from *C. perfringens* plasmid, the promoter Pcip from *R. josui*, and MCS and 6×His-tag region.

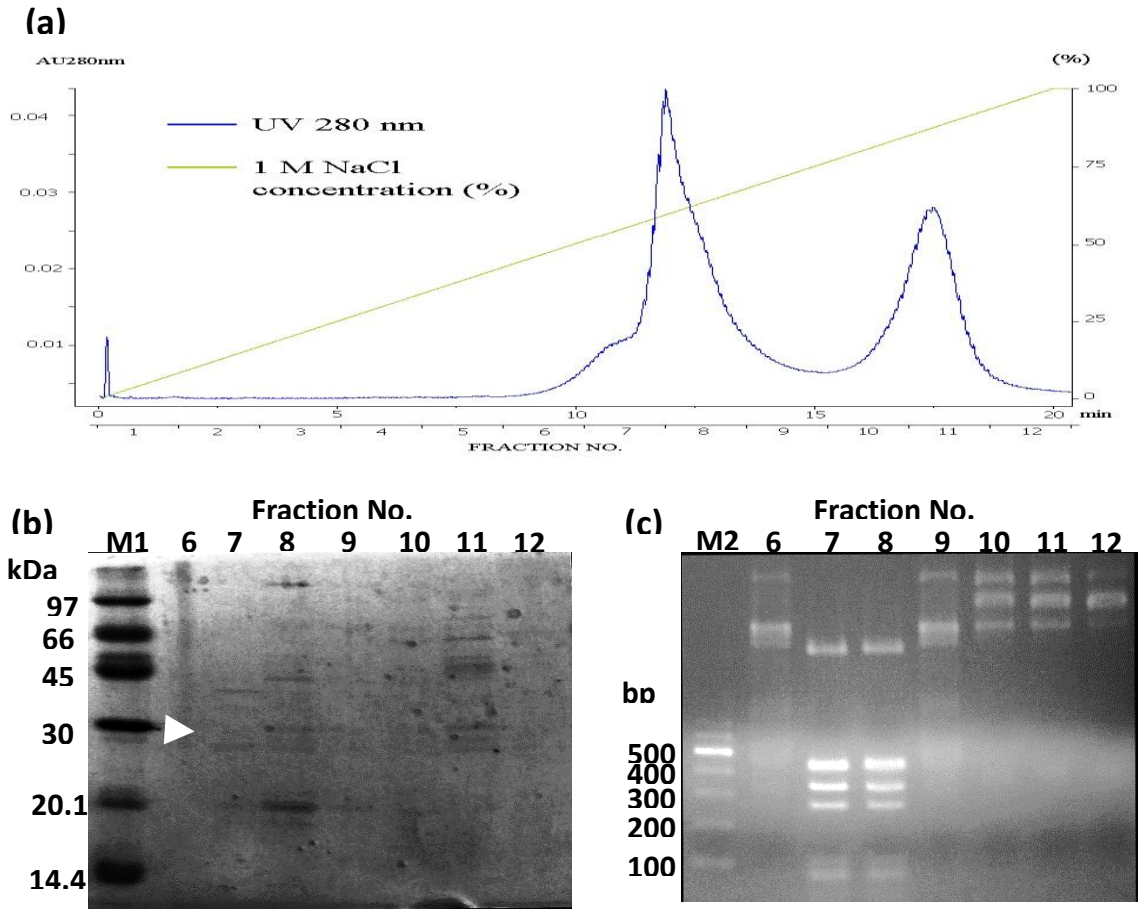


Fig. 13. Elution profile of HiTrap Q HP column chromatography (a), SDS-PAGE of eluted proteins (b) and *Rjo*I activity (c). *Rjo*I-containing fractions eluted from a HiTrap Q HP column were combined and subjected to HiTrap Q HP column chromatography (a). Absorbance at 280 nm (A_{280}) was monitored. Fractions 6 to 12 were subjected to SDS-PAGE analysis using a 15% polyacrylamide for separating gel (b) and the same fractions were examined for *Rjo*I activity by digesting pQE-30 followed by agarose gel electrophoresis (c). A white arrowhead in (b) shows the expected size of *Rjo*I. M1, LMW (low molecular weight) marker (GE Healthcare Japan); M2, 100 bp DNA ladder, New England Biolabs Japan).

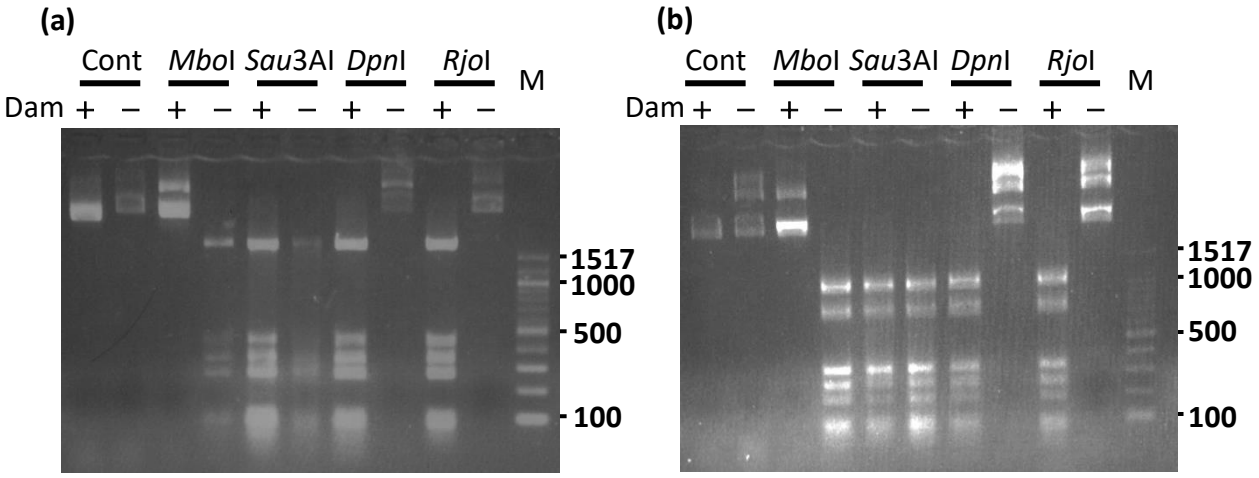


Fig. 14. Agarose gel electrophoresis of plasmids pQE-30 (a) and pBluescript II SK+ (b) digested with *Rjo*I and related REases. Plasmids pQE-30 (a) and pBluescript II SK+ (b) prepared from *E. coli* DH5α (dam^+/dcm^+) and *E. coli* C2925 (dam^-/dcm^-) were treated with *Mbo*I, *Sau*3AI, *Dpn*I and *Rjo*I and subjected to agarose gel electrophoresis. Cont, undigested control plasmids; M, size marker (100 bp DNA ladder); +, plasmid DNA prepared from *E. coli* DH5α (dam^+/dcm^+); -, plasmid DNA prepared from *E. coli* C2925 (dam^-/dcm^-).

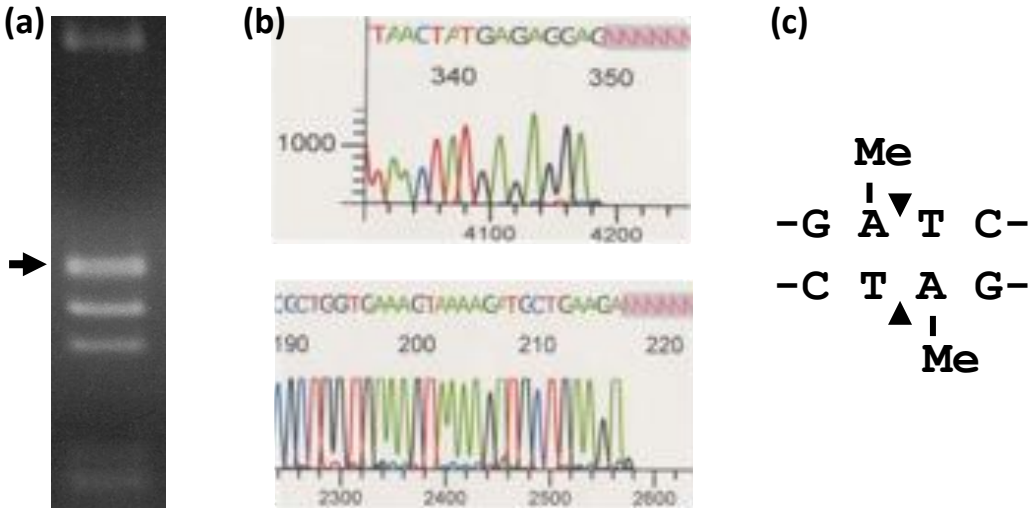


Fig. 15. Identification of *Rjo*I cleavage site. Dam-methylated pQE-30 was digested with *Rjo*I and separated by agarose gel electrophoresis, and the 435-bp fragment shown by an arrow was recovered (a). Sequencing reactions were carried out using the DYEnamic ET Terminator Cycle Sequencing Kit and primers 435bpF and 435bpR. Sequences were analyzed by an ABI PRISM® 3100 Genetic Analyzer (b). Cleavage sites of *Rjo*I are shown by two arrow heads (c).

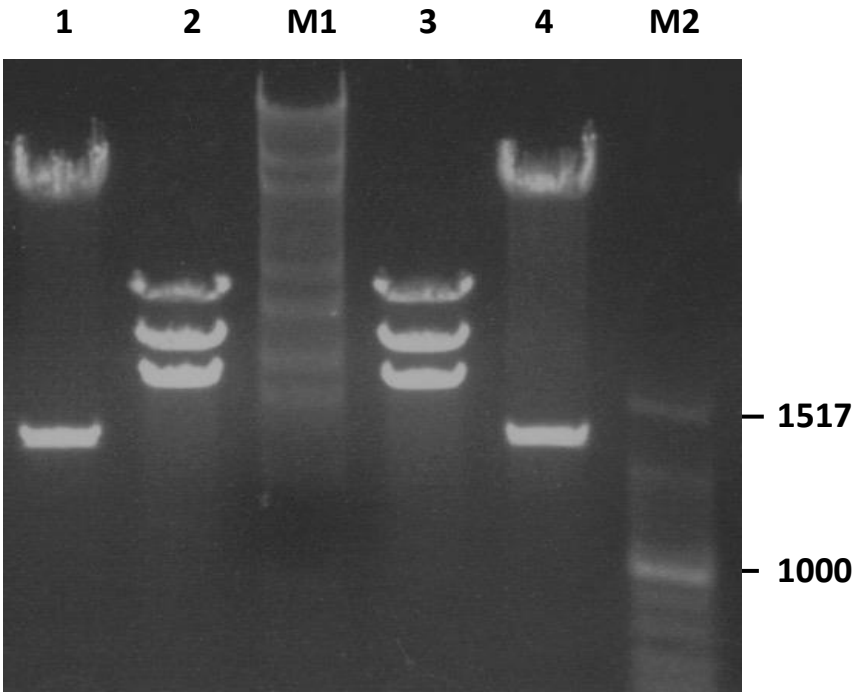


Fig. 16. pKKM801 in *E. coli* DH5α after rescue from plasmid of an erythromycin-resistant clone of *R. josui.* Lanes 1 and 2, pKKM801 used for electrotransformation of *R. josui*; lanes 3 and 4, pKKM801 isolated from *E. coli* transformed with plasmid of an erythromycin-resistant clone of *R. josui*; lanes 1 and 4, *Nde*I digests; lanes 2 and 3, *Bgl*I digests; lane M1, phage λ DNA digested with *Sty*I (marker 6, Nippon gene); lane M2, 100 bp DNA ladder.

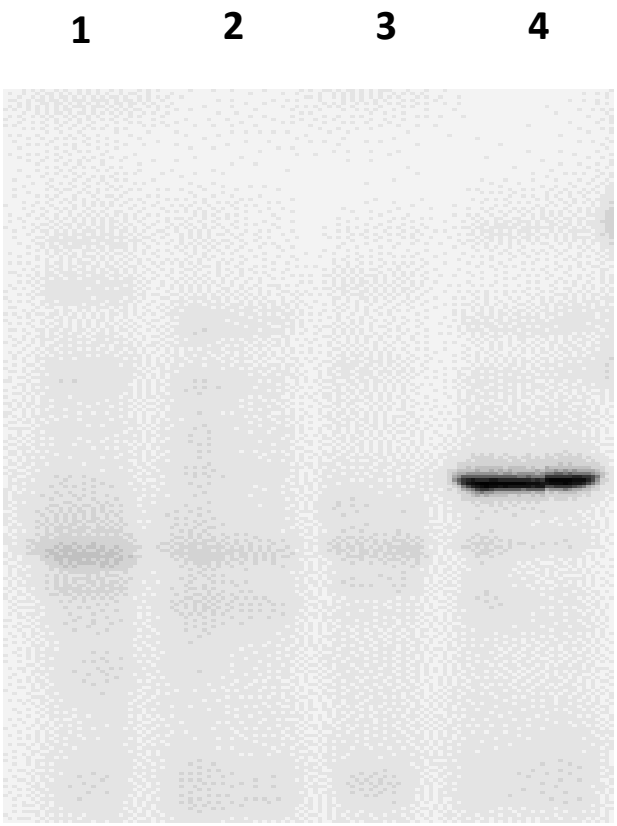


Fig. 17. Production of RjCel48AΔdoc by a recombinant *R. josui*. *R. josui* harboring pKKM801-RjCel48AΔdoc was cultivated in MGS medium containing cellobiose as a carbon source and RjCel48AΔdoc in the intracellular and supernatant fractions were detected by Western blot analysis using an Anti-His-tag mAb. Lane 1, intracellular fraction of a wild type; lane 2, extracellular fraction of wild type; 3, intracellular fraction of a recombinant *R. josui*; lane 4, extracellular fraction of a recombinant *R. josui*.

Table 1. Primer sets used for DNA amplification.

Target protein	Primer	Nucleotide sequence	Restriction site (Underlined)
RjAbf62A-Axe6A	GH62F	5'-CCCGCTAGCGCAGCAAACCCGAATCCATC	*Nhe*I
	CE6R	5'-CCCGTCGACCCATTTCAGCGCCTGAATC	*Sal*I
RjAbf62A-CBM6	GF62F		
	CBM6R	5'-CCCGTCGACCAATTTGCCTGTACTAACAGG	*Sal*I
RjAxe6A	CE6F	5'-CCCGCTAGCGGTATTATAATTGAATTCCCAGG	*Nhe*I
	CE6R		

Table 2. Hydrolytic activities of RjAbf62A-Axe6A and RjAbf62A-CBM6 toward various polysaccharides and oligosaccharides.

Substrate	Specific activity (U/nmol-protein)	
	RjAbf62A-Axe6A	RjAbf62A-CBM6
Rye arabinoxylan	254 ± 23	184 ± 2.4
Wheat arabinoxylan	167 ± 18	106 ± 2.5
Insoluble wheat arabinoxylan	8.45 ± 0.95	17.7 ± 1.2
Oat-spelt xylan	7.48 ± 0.40	9.77 ± 0.58
Birchwood xylan	0.11 ± 0.002	ND[a]
Beechwood xylan	0.15 ± 0.002	ND
Sugar beet arabinan	6.39 ± 0.28	6.14 ± 0.052
Linear arabinan	0.17 ± 0.009	0.037 ± 0.00058
A^3X	16.1 ± 1.5	2.3 ± 0.054
XA^3XX	2.84 ± 0.17	0.41 ± 0.087
A^2XX	6.89 ± 0.57	1.63 ± 0.13
$A^{2,3}XX$	0.037 ± 0.002	0.023 ± 0.0034

[a] Not detected.

Table 3. Acetylxylan esterase activities for RjAbf62A-Axe6A and RjAxe6A

Substrate	Specific activity (U/nmol)	
	RjAbf62A-Axe6A	RjAxe6A
α-D-Cellobiose octaacetate	7.22 ± 0.11	33.6 ± 0.92
Cellulose acetate	0.024 ± 0.002	0.020 ± 0.0003
Insoluble wheat arabinoxylan	0.017 ± 0.0002	0.00079 ± 0.0006

Table 4. Strains and plasmids used in this study.

Strains/plasmids	Relevant characteristics	Source/reference
R. josui JCM 17888	Wild-type	Ref. (61)
R. josui JCM 17888/ pKKM801-cel48AΔdoc	*R. josui* harboring pKKM801-cel48AΔdoc	This study
E. coli DH5α	*dam*$^{+}$, *dcm*$^{+}$	Invitrogen
E. coli C2925	*dam*$^{-}$, *dcm*$^{-}$	New England Biolabs
pBluescript II SK+	A cloning vector for *E. coli*	Stratagene
pQE-30	An expression vector for *E. coli*	Qiagen
pUC18	A cloning vector for *E. coli*	Nippon Gene
pJIR751	*C. perfringens-E. coli* shuttle vector containing erythromycin resistance, replication origins of pUC18 and pIP404	Generous gift of J. I. Rood, Ref. (96)
pKKM801	*R. josui- E. coli* shuttle vector	This study
pKKM801-cel48AΔdoc	pKKM801 containing GH48-encoding region of *cel48A*	This study

Table 5. Oligonucleotides used in this study.

Primer	Sequence (5’→3)	Intended use
435bpF	AAGGGAATAAGGGCGACAC	Sequencing
435bpR	CTTTTCGGGGAAATGTGCG	Sequencing
Pcip2079F	GGGGAATTCGAATTCATATTCATTTCAAATATCTG	PCR
Pcip2622R	CCCGGATCCCCGCCATGGGTAAACCTCCTATTTTTTA	PCR
pJIR-PfoF	TCCGGGACCAATAATCCCACTCTAAGC	PCR
pJIR-EcoO109R	GCGTATCACGAGGCCCTACGAATTGTACCCGGAAACG	PCR
pRjHis1	GGCGGCCGCACTCGAGCACCACCACCACCACCACTGAA	Linker
pRjHis2	AGCCTTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCCTGCA	Linker
Rjcel48AF	GGAGGTTTACCCATGAGTAAGAATTTTAAGAGAG	PCR
Rjcel48AR_2	CGCCTGCAGGTCGACCAAATTAGATCCTTCTGGG	PCR
qampF	TCGCCGCATACACTATTCTCAG	qPCR
qampR	TCTCTTACTGTCATGCCATCCG	qPCR
qman5AF	TTCCACTTTCCGCAATGAGG	qPCR
qman5AR	TTTGCACTGCCGAAGCTTAC	qPCR

*Eco*RI and *Bam*HI sites are underlined.

References

1. Conti J, Holtberg P, Diefenderfer J, LaRose A, Turnure JT, Westfall L. 2016. International Energy Outlook 2016 With Projections to 2040. USDOE Energy Information Administration (EIA), Washington, DC (United States). Office of Energy Analysis.

2. Höök M, Tang X. 2013. Depletion of fossil fuels and anthropogenic climate change—A review. Energy Policy 52:797-809.

3. Bhattacharya M, Paramati SR, Ozturk I, Bhattacharya S. 2016. The effect of renewable energy consumption on economic growth: Evidence from top 38 countries. Applied Energy 162:733-741.

4. Perlack RD, Wright LL, Turhollow AF, Graham RL, Stokes BJ, Erbach DC. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. Oak Ridge National Lab TN.

5. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Jr., Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T. 2006. The path forward for biofuels and biomaterials. Science. 311:484-489.

6. Alvarez C, Reyes-Sosa FM, Diez B. 2016. Enzymatic hydrolysis of biomass from wood. Microbial Biotechnology. 9:149-156.

7. Qian X, Ding SY, Nimlos MR, Johnson DK, Himmel ME. 2005. Atomic and electronic structures of molecular crystalline cellulose Iβ: A first-principles investigation. Macromolecules. 38:10580-10589.

8. Wegner TH, Jones EP. 2009. A fundamental review of the relationships between nanotechnology and lignocellulosic biomass. The Nanoscience and Technology of Renewable Biomaterials. 1:1-41.

9. Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annual Review of Plant Biology. 61:263-289.

10. Biely P, Singh S, Puchart V. 2016. Towards enzymatic breakdown of complex plant xylan structures: State of the art. Biotechnology Advances. 34:1260-1274.

11. Brandt A, Grasvik J, Hallett JP, Welton T. 2013. Deconstruction of lignocellulosic biomass with ionic liquids. Green Chemistry. 15:550-583.

12. Dutta A, Sahir AH, Tan E, Humbird D, Snowden-Swan LJ, Meyer PA, Ross J, Sexton D, Yap R, Lukas J. 2015. Process design and economics for the conversion of lignocellulosic biomass to hydrocarbon fuels: Thermochemical research pathways with In Situ and Ex Situ Upgrading of Fast Pyrolysis Vapors. Pacific Northwest National Lab. Richland, WA (United States).

13. Ding SY, Liu YS, Zeng YN, Himmel ME, Baker JO, Bayer EA. 2012. How Does Plant Cell Wall Nanoscale Architecture Correlate with Enzymatic Digestibility? Science. 338:1055-1060.

14. Artzi L, Bayer EA, Morais S. 2017. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nature Reviews Microbiology. 15:83-95.

15. Gilbert HJ, Knox JP, Boraston AB. 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Current Opinion in Structural Biology. 23:669-677.

16. Bischof RH, Ramoni J, Seiboth B. 2016. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. Microbial Cell Factories. 15:106.

17. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nature Biotechnology. 26:553-560.

18. Lamed R, Setter E, Bayer EA. 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. Journal of Bacteriology. 156:828-836.

19. Lamed R. 1983. The cellulosome : a discrete cell surface organelle of *Clostridium thermocellum* exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnology and Bioengineering Symposium. 13:163-181.

20. Bayer EA, Kenig R, Lamed R. 1983. Adherence of *Clostridium thermocellum* to cellulose. Journal of Bacteriology. 156:818-827.

21. Lu Y, Zhang Y-HP, Lynd LR. 2006. Enzyme–microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. Proceedings of the National Academy of Sciences. 103:16165-16169.

22. Wang Y, Sakka M, Yagi H, Kaneko S, Katsuzaki H, Kunitake E, Kimura T, Sakka K. 2018. *Ruminiclostridium josui* Abf62A-Axe6A: A tri-functional xylanolytic enzyme exhibiting α-L-arabinofuranosidase, endoxylanase, and acetylxylan esterase activities. Enzyme and Microbial Technology. 117:1-8.

23. Orita T, Sakka M, Kimura T, Sakka K. 2017. Characterization of *Ruminiclostridium josui* arabinoxylan arabinofuranohydrolase, RjAxh43B, and RjAxh43B-containing xylanolytic complex. Enzyme and Microbial Technology. 104:37-43.

24. Orita T, Sakka M, Kimura T, Sakka K. 2017. Recombinant cellulolytic or xylanolytic complex comprising the full-length scaffolding protein RjCipA and cellulase RjCel5B or xylanase RjXyn10C of *Ruminiclostridium josui*. Enzyme and Microbial Technology. 97:63-70.

25. Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MWW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiology Reviews. 38:393-448

26. Ferdinand PH, Borne R, Trotter V, Pages S, Tardif C, Fierobe HP, Perret S. 2013. Are cellulosome scaffolding protein CipC and CBM3-containing protein HycP, involved in adherence of *Clostridium cellulolyticum* to cellulose? PLOS One. 8:e69360.

27. Zhao G, Ali E, Araki R, Sakka M, Kimura T, Sakka K. 2005. Function of the family-9 and family-22 carbohydrate-binding modules in a modular β-1,3-1,4-glucanase/xylanase derived from *Clostridium stercorarium* Xyn10B. Bioscience, Biotechnology, and Biochemistry. 69:1562-1567.

28. Ali E, Zhao G, Sakka M, Kimura T, Ohmiya K, Sakka K. 2005. Functions of family-22 carbohydrate-binding module in *Clostridium thermocellum* Xyn10C. Bioscience, Biotechnology, and Biochemistry. 69:160-165.

29. Ali E, Araki R, Zhao G, Sakka M, Karita S, Kimura T, Sakka K. 2005. Functions of family-22 carbohydrate-binding modules in *Clostridium josui* Xyn10A. Bioscience, Biotechnology, and Biochemistry. 69:2389-2394.

30. Lemaire M, Miras I, Gounon P, Beguin P. 1998. Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. Microbiology. 144:211-217.

31. Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, Wilson C, Klingeman DM, Olson DG, Decker SR, Giannone RJ, Hettich RL, Brown SD, Lynd LR, Bayer EA, Himmel ME, Bomble YJ. 2016. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Science Advances. 2:e1501254.

32. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. 42:490-495.

33. Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology. 16:577-583.

34. Bertani G, Weigle JJ. 1953. Host controlled variation in bacterial viruses. Journal of Bacteriology. 65:113-121..

35. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. 2014. Type I restriction enzymes and their relatives. Nucleic Acids Research. 42:20-44.

36. Pingoud A, Wilson GG, Wende W. 2014. Type II restriction endonucleases--a historical perspective and more. Nucleic Acids Research. 42:7489-7527.

37. Rao DN, Dryden DT, Bheemanaik S. 2014. Type III restriction-modification enzymes: a historical perspective. Nucleic Acids Research. 42:45-55.

38. Loenen WA, Raleigh EA. 2014. The other face of restriction: modification-dependent enzymes. Nucleic Acids Research. 42:56-69.

39. Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Research. 42:3-19.

40. Roberts RJ. 2005. How restriction enzymes became the workhorses of molecular biology. Proceedings of the National Academy of Sciences of the United States of America. 102:5905-5908.

41. Pingoud A, Jeltsch A. 2001. Structure and function of type II restriction endonucleases. Nucleic Acids Research. 29:3705-3727.

42. Yang X, Xu M, Yang ST. 2016. Restriction modification system analysis and development of in vivo methylation for the transformation of *Clostridium cellulovorans*. Applied Microbiology and Biotechnology. 100:2289-2299.

43. Lesiak JM, Liebl W, Ehrenreich A. 2014. Development of an in vivo methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. Journal of Biotechnology. 188:97-99.

44. Jennert KC, Tardif C, Young DI, Young M. 2000. Gene transfer to *Clostridium cellulolyticum* ATCC 35319. Microbiology. 146:3071-3080.

45. Davis TO, Henderson I, Brehm JK, Minton NP. 2000. Development of a transformation and gene reporter system for group II, non-proteolytic *Clostridium botulinum* type B strains. Journal of Molecular Microbiology and Biotechnology. 2:59-69.

46. Sakka K, Kawase M, Baba D, Morimoto K, Karita S, Kimura T, Ohmiya K. 2003. Electrotransformation of *Clostridium paraputrificum* M-21 with some plasmids. Journal of Bioscience and Bioengineering. 96:304-306.

47. Olson DG, Maloney M, Lanahan AA, Hon S, Hauser LJ, Lynd LR. 2015. Identifying promoters for gene expression in *Clostridium thermocellum*. Metabolic Engineering Communications. 2:23-29.

48. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R. 2012. Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiology. 12:214.

49. Deng Y, Olson DG, Zhou J, Herring CD, Joe Shaw A, Lynd LR. 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metabolic Engineering. 15:151-158.

50. Pyne ME, Bruder M, Moo-Young M, Chung DA, Chou CP. 2014. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. Biotechnology Advances. 32:623-641.

51. Olson DG, Lynd LR. 2012. Transformation of *Clostridium thermocellum* by electroporation. Methods in Enzymology. 510:317-330.

52. Guss AM, Olson DG, Caiazza NC, Lynd LR. 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnology for Biofuels. 5:30.

53. Gelinas RE, Myers PA, Roberts RJ. 1977. Two sequence-specific endonucleases from *Moraxella bovis*. Journal of Molecular Biology. 114:169-179.

54. Lacks S, Greenberg B. 1975. A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. The Journal of Biological Chemistry. 250:4060-4066.

55. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. Journal of Microbiological Methods. 70:452-464.

56. Fagan RP, Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. The Journal of Biological Chemistry. 286:27483-27493.

57. Nariya H, Miyata S, Kuwahara T, Okabe A. 2011. Development and characterization of a xylose-inducible gene expression system for *Clostridium perfringens*. Applied and Environmental Microbiology. 77:8439-8441.

58. Zhang J, Liu YJ, Cui GZ, Cui Q. 2015. A novel arabinose-inducible genetic operation system developed for *Clostridium cellulolyticum*. Biotechnology for Biofuels. 8:36.

59. Mearls EB, Olson DG, Herring CD, Lynd LR. 2015. Development of a regulatable plasmid-based gene expression system for *Clostridium thermocellum*. Applied Microbiology and Biotechnology. 99:7589-7599.

60. Kakiuchi M, Isui A, Suzuki K, Fujino T, Fujino E, Kimura T, Karita S, Sakka K, Ohmiya K. 1998. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. Journal of Bacteriology. 180:4303-4308.

61. Sukhumavasi, J., Ohmiya, K., Shimizu, S., & Ueno, K. 1988. *Clostridiurn josui* sp. nov. , a cellulolytic, moderate thermophilic species from Thai compost. International Journal of Systematic and Evolutionary Microbiology. 38: 179-182.

62. Yutin N, Galperin MY. 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced *clostridia*. Environmental Microbiology. 15:2631-2641.

63. Sakka M, Goto M, Fujino T, Fujino E, Karita S, Kimura T, Sakka K. 2010. Analysis of a *Clostridium josui* cellulase gene cluster containing the *man5A* gene and characterization of recombinant Man5A. Bioscience, Biotechnology, and Biochemistry. 74:2077-2082.

64. Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant Journal. 54:559-568.

65. Naidu DS, Hlangothi SP, John MJ. 2018. Bio-based products from xylan: A review. Carbohydrate Polymers. 179:28-41.

66. Biely P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnology Advances. 30:1575-1588.

67. Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe EC, Basle A, Morland C, Day AM, Zheng H, Rogers TE, Thompson P, Hawkins AR, Yadav MP, Henrissat B, Martens EC, Dupree P, Gilbert HJ, Bolam DN. 2015. Glycan complexity dictates microbial resource allocation in the large intestine. Nature Communications. 6:7481

68. Mewis K, Lenfant N, Lombard V, Henrissat B. 2016. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. Applied and Environmental Microbiology. 82:1686-1692.

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

70. Smogyi M. 1952. Notes on sugar determination. The Journal of Biological Chemistry. 195:19-23..

71. Maehara T, Yagi H, Sato T, Ohnishi-Kameyama M, Fujimoto Z, Kamino K, Kitamura Y, St John F, Yaoi K, Kaneko S. 2018. GH30 glucuronoxylan-specific xylanase from *Streptomyces turgidiscabies* C56. Applied and Environmental Microbiology. 84.

72. Wilkens C, Andersen S, Petersen BO, Li A, Busse-Wicher M, Birch J, Cockburn D, Nakai H, Christensen HE, Kragelund BB, Dupree P, McCleary B, Hindsgaul O, Hachem MA, Svensson B. 2016. An efficient arabinoxylan-debranching α-L-arabinofuranosidase of family GH62 from *Aspergillus nidulans* contains a secondary carbohydrate binding site. Applied Microbiology and Biotechnology. 100:6265-6277.

73. De La Mare M, Guais O, Bonnin E, Weber J, Francois JM. 2013. Molecular and biochemical characterization of three GH62 α-L-arabinofuranosidases from the soil deuteromycete *Penicillium funiculosum*. Enzyme and Microbial Technology. 53:351-358.

74. Kellett LE, Poole DM, Ferreira LM, Durrant AJ, Hazlewood GP, Gilbert HJ. 1990. Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes. Biochemical Journal. 272:369-376.

75. Sakamoto T, Ogura A, Inui M, Tokuda S, Hosokawa S, Ihara H, Kasai N. 2011. Identification of a GH62 α-L-arabinofuranosidase specific for arabinoxylan produced by *Penicillium chrysogenum*. Applied Microbiology and Biotechnology. 90:137-146.

76. Maehara T, Fujimoto Z, Ichinose H, Michikawa M, Harazono K, Kaneko S. 2014. Crystal structure and characterization of the glycoside hydrolase family 62 α-L-arabinofuranosidase from *Streptomyces coelicolor*. The Journal of Biological Chemistry. 289:7962-72.

77. Vincent P, Shareck F, Dupont C, Morosoli R, Kluepfel D. 1997. New α-L-arabinofuranosidase produced by *Streptomyces lividans*: cloning and DNA sequence of the *abfB* gene and characterization of the enzyme. Biochemical Journal. 322:845-852.

78. Borsenberger V, Dornez E, Desrousseaux ML, Massou S, Tenkanen M, Courtin CM, Dumon C, O'Donohue MJ, Faure R. 2014. A (1)H NMR study of the specificity of α-L-arabinofuranosidases on natural and unnatural substrates. Biochimica et Biophysica Acta. 1840:3106-3114.

79. Ferre H, Broberg A, Duus JO, Thomsen KK. 2000. A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate

preference and specificity by nano-probe NMR. European Journal of Biochemistry. 267:6633-6641.

80. Koutaniemi S, Tenkanen M. 2016. Action of three GH51 and one GH54 α-arabinofuranosidases on internally and terminally located arabinofuranosyl branches. Journal of Biotechnology. 229:22-30.

81. Hespell RB, Cotta MA. 1995. Degradation and utilization by *Butyrivibrio-fibrisolvens* H17c of xylans with different chemical and physical-properties. Applied and Environmental Microbiology. 61:3042-3050.

82. Knudsen KEB, Laerke HN. 2010. Rye Arabinoxylans: Molecular structure, physicochemical properties and physiological effects in the Gastrointestinal Tract. Cereal Chemistry. 87:353-362.

83. Mckee LS, Pena MJ, Rogowski A, Jackson A, Lewis RJ, York WS, Krogh KBRM, Vikso-Nielsen A, Skjot M, Gilbert HJ, Marles-Wright J. 2012. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. Proceedings of the National Academy of Sciences of the United States of America. 109:6537-6542.

84. Biely P, Mastihubova M, Tenkanen M, Eyzaguirre J, Li XL, Vrsanska M. 2011. Action of xylan deacetylating enzymes on monoacetyl derivatives of 4-nitrophenyl glycosides of β-D-xylopyranose and α-L-arabinofuranose. Journal of Biotechnology. 151:137-142.

85. Fang JM, Sun RC, Tomkinson J. 2000. Isolation and characterization of hemicelluloses and cellulose from rye straw by alkaline peroxide extraction. Cellulose. 7:87-107.

86. Mai-Gisondi G, Maaheimo H, Chong SL, Hinz S, Tenkanen M, Master E. 2017. Functional comparison of versatile carbohydrate esterases from families CE1, CE6 and CE16 on acetyl-4-O-methylglucuronoxylan and acetyl-galactoglucomannan. Biochimica et Biophysica Acta-General Subjects. 1861:2398-2405.

87. Kumar R, Singh S, Singh OV. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. Journal of Industrial Microbiology & Biotechnology. 35:377-391.

88. Sanderson K. 2006. A field in ferment. Nature. 444:673-676.

89. Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR. 2010. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. Proceedings of the National Academy of Sciences of the United States of America. 107:17727-17732.

90. Gaida SM, Liedtke A, Jentges AH, Engels B, Jennewein S. 2016. Metabolic engineering of *Clostridium cellulolyticum* for the production of n-butanol from crystalline cellulose. Microbial Cell Factories. 15:6.

91. Yu M, Du Y, Jiang W, Chang WL, Yang ST, Tang IC. 2012. Effects of different replicons in conjugative plasmids on transformation efficiency, plasmid stability, gene expression and n-butanol biosynthesis in *Clostridium tyrobutyricum*. Applied Microbiology and Biotechnology. 93:881-889.

92. Mingardon F, Chanal A, Tardif C, Fierobe HP. 2011. The issue of secretion in heterologous expression of Clostridium cellulolyticum cellulase-encoding genes in *Clostridium acetobutylicum* ATCC 824. Applied and Environmental Microbiology. 77:2831-2838.

93. Tock MR, Dryden DTF. 2005. The biology of restriction and anti-restriction. Current Opinion in Microbiology. 8:466-472

94. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Research. 43:298-299.

95. Johnson EA, Sakajoh M, Halliwell G, Madia A, Demain AL. 1982. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Applied and Environmental Microbiology. 43:1125-1132.

96. Bannam TL, Rood JI. 1993. *Clostridium perfringens-Escherichia coli* shuttle vectors that carry single antibiotic resistance determinants. Plasmid. 29:233-235.

97. Tyurin MV, Desai SG, Lynd LR. 2004. Electrotransformation of *Clostridium thermocellum*. Applied and Environmental Microbiology. 70:883-890.

98. Lee C, Kim J, Shin SG, Hwang S. 2006. Absolute and relative qPCR quantification of plasmid copy number in *Escherichia coli*. Journal of Biotechnology. 123:273-280.

99. Geier GE, Modrich P. 1979. Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of *Dpn* I endonuclease. The Journal of Biological Chemistry. 254:1408-1413.

100. Roberts RJ, Macelis D. 1992. Restriction enzymes and their isoschizomers. Nucleic Acids Research. 20:2167-2180.

101. Azeddoug H, Hubert J, Reysset G. 1992. Stable inheritance of shuttle vectors based on plasmid pIM13 in a mutant strain of *Clostridium acetobutylicum*. The Journal of General and Applied Microbiology. 138:1371-1378.

102. Olson DG, Giannone RJ, Hettich RL, Lynd LR. 2013. Role of the CipA scaffoldin protein in cellulose solubilization, as determined by targeted gene deletion and complementation in *Clostridium thermocellum*. Journal of Bacteriology. 195:733-739.

103. Cui GZ, Hong W, Zhang J, Li WL, Feng YG, Liu YJ, Cui Q. 2012. Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. Journal of Microbiological Methods. 89:201-208.

104. Dong HJ, Zhang YP, Dai ZJ, Li Y. 2010. Engineering *Clostridium* strain to accept unmethylated DNA. Plos One. 5.

105. Kuehne SA, Minton NP. 2012. ClosTron-mediated engineering of *Clostridium*. Bioengineered. 3:247-254.

106. Purdy D, O'Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP. 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Molecular Microbiology. 46:439-452.

107. Lee SY, Mermelstein LD, Bennett GN, Papoutsakis ET. 1992. Vector construction, transformation, and gene amplification in *Clostridium acetobutylicum* ATCC 824. Annals of the New York Academy of Sciences. 665:39-51.

108. Lin YL, Blaschek HP. 1984. Transformation of heat-treated *Clostridium acetobutylicum* protoplasts with pUB110 plasmid DNA. Applied and Environmental Microbiology. 48:737-742.

109. Mermelstein LD, Papoutsakis ET. 1993. In vivo methylation in *Escherichia coli* by the Bacillus subtilis phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. Applied and Environmental Microbiology. 59:1077-1081.

110. Cantalupo G, Bucci C, Salvatore P, Pagliarulo C, Roberti V, Lavitola A, Bruni CB, Alifano P. 2001. Evolution and function of the neisserial dam-replacing gene. FEBS Letters. 495:178-183.

111. Bentley SD, Vernikos GS, Snyder LAS, Churcher C, Arrowsmith C, Chillingworth T, Cronin A, Davis PH, Holroyd NE, Jagels K, Maddison M, Moule S, Rabbinowitsch E, Sharp S, Unwin L, Whitehead S, Quail MA, Achtman M, Barrell B, Saunders NJ, Parkhill J. 2007. Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. PLOS Genetics. 3:230-240.

112. Kolek J, Sedlar K, Provaznik I, Patakova P. 2016. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnology for Biofuels. 9:14.

113. Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, Wang Y, Cui Q, Xu J. 2015. Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. Nature Communications. 6:6900.

114. Vazana Y, Morais S, Barak Y, Lamed R, Bayer EA. 2010. Interplay between *Clostridium thermocellum* family 48 and family 9 cellulases in cellulosomal versus noncellulosomal states. Applied and Environmental Microbiology. 76:3236-3243.

115. Reverbel-Leroy C, Pages S, Belaich A, Belaich JP, Tardif C. 1997. The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: purification and characterization of the recombinant form. Journal of Bacteriology. 179:46-52.

116. Kruus K, Wang WK, Chiu PC, Ching JT, Wang TY, Wu JHD. 1994. Cels - a major exoglucanase component of *Clostridium-Thermocellum* cellulosome. Enzymatic Conversion of Biomass for Fuels Production. 566:84-99.

117. Zhu Y, Liu X, Yang ST. 2005. Construction and characterization of pta gene-deleted mutant of *Clostridium tyrobutyricum* for enhanced butyric acid fermentation. Biotechnology and Bioengineering. 90:154-166.

118. Fujino T, Karita S, Ohmiya K. 1993. Nucleotide sequences of the *celb* gene encoding endo-1,4-B-glucanase-2, Orf1 and Orf2 forming a putative cellulase gene-cluster of *Clostridium josui*. Journal of Fermentation and Bioengineering. 76:243-250.

119. Fujino T, Sasaki T, Ohmiya K, Shimizu S. 1990. Purification and properties of an endo-1,4-β-glucanase translated from a *Clostridium josui* gene in *Escherichia coli*. Applied and Environmental Microbiology. 56:1175-1178.

120. Fujino T, Sukhumavasi J, Sasaki T, Ohmiya K, Shimizu S. 1989. Purification and properties of an endo-1,4-β-glucanase from *Clostridium josui*. Journal of Bacteriology. 171:4076-4079.

121. Jindou S, Karita S, Fujino E, Fujino T, Hayashi H, Kimura T, Sakka K, Ohmiya K. 2002. α-Galactosidase Aga27A, an enzymatic component of the *Clostridium josui* cellulosome. Journal of Bacteriology. 184:600-4.

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

123. Tsukimoto K, Takada R, Araki Y, Suzuki K, Karita S, Wakagi T, Shoun H, Watanabe T, Fushinobu S. 2010. Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module. FEBS Letters. 584:1205-1211.

124. Araki Y, Karita S, Tsuchiya T, Kondo M, Goto M. 2010. Family 17 and 28 carbohydrate-binding modules discriminated different cell-wall sites in sweet potato roots. Bioscience, Biotechnology, and Biochemistry. 74:802-805.

125. Sakka K, Sugihara Y, Jindou S, Sakka M, Inagaki M, Sakka K, Kimura T. 2011. Analysis of cohesin-dockerin interactions using mutant dockerin proteins. FEMS Microbiology Letters. 314:75-80.

126. Kamezaki Y, Enomoto C, Ishikawa Y, Koyama T, Naya S, Suzuki T, Sakka K. 2010. The Dock tag, an affinity tool for the purification of recombinant proteins, based on the interaction between dockerin and cohesin domains from *Clostridium josui* cellulosome. Protein Expression and Purification. 70:23-31.

127. Jindou S, Soda A, Karita S, Kajino T, Beguin P, Wu JH, Inagaki M, Kimura T, Sakka K, Ohmiya K. 2004. Cohesin-dockerin interactions within and between *Clostridium josui* and *Clostridium thermocellum*: binding selectivity between cognate

dockerin and cohesin domains and species specificity. The Journal of Biological Chemistry. 279:9867-9874.

128. Peng X, Qiao W, Mi S, Jia X, Su H, Han Y. 2015. Characterization of hemicellulase and cellulase from the extremely thermophilic bacterium *Caldicellulosiruptor owensensis* and their potential application for bioconversion of lignocellulosic biomass without pretreatment. Biotechnology for Biofuels. 8:131.

129. Shahid S, Tajwar R, Akhtar MW. 2018. A novel trifunctional, family GH10 enzyme from *Acidothermus cellulolyticus* 11B, exhibiting endo-xylanase, arabinofuranosidase and acetyl xylan esterase activities. Extremophiles. 22:109-119.

130. Yang Y, Zhu N, Yang J, Lin Y, Liu J, Wang R, Wang F, Yuan H. 2017. A novel bifunctional acetyl xylan esterase/arabinofuranosidase from *Penicillium chrysogenum* P33 enhances enzymatic hydrolysis of lignocellulose. Microbial Cell Factories. 16:166.

131. Wan CF, Chen CT, Huang L, Li YK. 2007. Expression, purification and characterization of a bifunctional α-L-arabinofuranosidase/β-D-xylosidae from *Trichoderma koningii* G-39. Journal of the Chinese Chemical Society. 54:109-116.

132. Jordan DB, Li XL. 2007. Variation in relative substrate specificity of bifunctional β-D-xylosidase/α-L-arabinofuranosidase by single-site mutations: roles of substrate distortion and recognition. Biochimica et Biophysica Acta. 1774:1192-1198.

133. Sakka K, Yoshikawa K, Kojima Y, Karita S, Ohmiya K, Shimada K. 1993. Nucleotide sequence of the *Clostridium stercorarium xylA* gene encoding a bifunctional protein with β-D-xylosidase and α-L-arabinofuranosidase activities, and properties of the translated product. Bioscience, Biotechnology, and Biochemistry. 57:268-272.

134. Teeravivattanakit T, Baramee S, Phitsuwan P, Waeonukul R, Pason P, Tachaapaikoon C, Sakka K, Ratanakhanokchai K. 2016. A novel trifunctional xylanolytic enzyme Axy43A from *Paenibacillus curdlanolyticus* B-6 exhibiting endo-xylanase, β-D-xylosidase, and arabinoxylan arabinofuranohydrolase activities. Applied and Environmental Microbiology.

135. Yang W, Bai Y, Yang P, Luo H, Huang H, Meng K, Shi P, Wang Y, Yao B. 2015. A novel bifunctional GH51 exo-α-L-arabinofuranosidase/endo-xylanase from *Alicyclobacillus sp.* A4 with significant biomass-degrading capacity. Biotechnology for Biofuels. 8:197.

136. Dodd D, Kocherginskaya SA, Spies MA, Beery KE, Abbas CA, Mackie RI, Cann IKO. 2009. Biochemical analysis of a β-D-xylosidase and a bifunctional xylanase-ferulic acid esterase from a xylanolytic gene cluster in *Prevotella ruminicola* 23. Journal of Bacteriology. 191:3328-3338.

137. Kahn A, Bayer EA, Morais S. 2018. Advanced cloning tools for construction of designer cellulosomes. Methods in Molecular Biology. 1796:135-151.

138. Stern J, Morais S, Lamed R, Bayer EA. 2016. Adaptor Scaffoldins: An 0riginal strategy for extended designer cellulosomes, Inspired from Nature. Mbio. 7:e00083.

139. Bras JL, Pinheiro BA, Cameron K, Cuskin F, Viegas A, Najmudin S, Bule P, Pires VM, Romao MJ, Bayer EA, Spencer HL, Smith S, Gilbert HJ, Alves VD, Carvalho AL, Fontes CM. 2016. Diverse specificity of cellulosome attachment to the bacterial cell surface. Scientific Reports. 6:38292.

140. Srikrishnan S, Chen W, Da Silva NA. 2013. Functional assembly and characterization of a modular xylanosome for hemicellulose hydrolysis in yeast. Biotechnology and Bioengineering. 110:275-285.

141. Morais S, Barak Y, Hadar Y, Wilson DB, Shoham Y, Lamed R, Bayer EA. 2011. Assembly of xylanases into designer cellulosomes promotes efficient hydrolysis of the xylan component of a natural recalcitrant cellulosic substrate. Mbio. 2.

142. McClendon SD, Mao Z, Shin HD, Wagschal K, Chen RR. 2012. Designer xylanosomes: protein nanostructures for enhanced xylan hydrolysis. Applied Biochemistry and Biotechnology. 167:395-411.

143. Liu YJ, Liu S, Dong S, Li R, Feng Y, Cui Q. 2018. Determination of the native features of the exoglucanase Cel48S from *Clostridium thermocellum*. Biotechnology for Biofuels. 11:6.

144. Herman NA, Li J, Bedi R, Turchi B, Liu X, Miller MJ, Zhang W. 2017. Development of a high-efficiency transformation method and implementation of rational metabolic engineering for the industrial butanol hyperproducer *Clostridium saccharoperbutylacetonicum* Strain N1-4. Applied and Environmental Microbiology. 83.

145. Bruder MR, Pyne ME, Moo-Young M, Chung DA, Chou CP. 2016. Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus *Clostridium*. Applied and Environmental Microbiology. 82:6109-6119.

146. Mermelstein LD, Welker NE, Bennett GN, Papoutsakis ET. 1992. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. Biotechnology. 10:190-195.

147. Siwek W, Czapinska H, Bochtler M, Bujnicki JM, Skowronek K. 2012. Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.*Dpn*I. Nucleic Acids Research. 40:7563-7572.

148. Yu M, Zhang Y, Tang IC, Yang ST. 2011. Metabolic engineering of *Clostridium tyrobutyricum* for n-butanol production. Metabolic Engineering. 13:373-82.

149. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. Journal of Microbiological Methods. 80:49-55.

150. Tolonen AC, Chilaka AC, Church GM. 2009. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Molecular Microbiology. 74:1300-1313.

151. O'Toole PW, Jeffery IB. 2015. Gut microbiota and aging. Science. 350:1214-1215.

152. Cann I, Bernardi RC, Mackie RI. 2016. Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. Environmental Microbiology. 18:307-310.

153. Zhang ML, Chekan JR, Dodd D, Hong PY, Radlinski L, Revindran V, Nair SK, Mackie RI, Cann I. 2014. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proceedings of the National Academy of Sciences of the United States of America. 111:3708-3717.

154. White BA, Lamed R, Bayer EA, Flint HJ. 2014. Biomass utilization by gut microbiomes. Annual Review of Microbiology. 68:279-296.

155. Olson DG, McBride JE, Shaw AJ, Lynd LR. 2012. Recent progress in consolidated bioprocessing. Current Opinion in Biotechnology. 23:396-405.

156. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Stahlberg J, Beckham GT. 2015. Fungal cellulases. Chemical Reviews. 115:1308-448.

157. Vaaje-Kolstad G, Forsberg Z, Loose JS, Bissaro B, Eijsink VG. 2017. Structural diversity of lytic polysaccharide monooxygenases. Current Opinion in Structural Biology. 44:67-76.

158. Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P. 2014. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. Plant Journal. 79:492-506.

159. Kunitake E, Kobayashi T. 2017. Conservation and diversity of the regulators of cellulolytic enzyme genes in Ascomycete fungi. Current Genetics. 63:951-958.

160. Olson DG, Sparling R, Lynd LR. 2015. Ethanol production by engineered thermophiles. Current Opinion in Biotechnology. 33:130-141.

Acknowledgements

First of all, I came to Mie University for the first time as an exchange student in April 2012, and spent a wonderful 6.5 years in life at the Applied Microbiology Laboratory of Graduate School & Faculty of Bioresources Mie University. Here, after careful training by the teachers, I have developed a certain academic ability that will allow me to face my work and life alone in the future.

I would like to thank the Pro. Kazuo Sakka giving me a chance to come to Mie University six years ago. I am honored to have spent so much time studying in the Pro. Kazuo Sakka's 36-year coaching career. Thanks to Ph.D. Makiko Sakka for her guidance and help over the years, but I feel that I have not been able to clean the equipment quickly after the experiment. This bad habit has caused you the inconvenience caused in the management of the laboratory. I apologize for this. In the next research activities, I will strictly demand myself, and please rest assured.

I would like to express my deep gratitude to Prof. Kazuo Sakka, Ph.D. Makiko Sakka, Prof. Tetsuya Kimura and Assistant Prof. Emi Kunitake of Applied Microbiology Laboratory of Graduate School & Faculty of Bioresources, Mie University for their love and care for me over the past six years, for the guidance and advice given in research, and for the help of learning and life.

I also would like to thank Ph.D. Candidate Haruka Yagi of The United Graduate School of Agricultural Sciences, Kagoshima University, Japan, Prof. Satoshi Kaneko of Faculty of Agriculture, University of the Ryukyu, Japan, Associate Prof. Hirotaka Katsuzaki of Graduate School & Faculty of Bioresources, Mie University,

Japan, Associate Prof. Naoto Isono of Graduate School & Faculty of Bioresources, Mie University, Japan, and Research Associate Masayuki Fukumura and Shuhei Otsuka of Department of Microbiology and Molecular Genetics, Mie University, Japan for their help and advice.

I would also like to thank my wife for her quiet commitment to the family and understanding and support for me over the years. I also would to thank my parents for their encouragement and support. I am able to successfully complete my studies. I hope that this article will be presented to you as a gift. It is your dedication and support that will enable me to successfully complete my studies.

Finally, even though there have been countless difficulties encountered in these years, the support and encouragement of everyone have enabled me to overcome the difficulties in my life and study, which has made me stronger in my future life. In addition, my experience in studying in Japan has made me gain love and study. It is unforgettable in my life. Youth and sweat are scattered in the laboratory. This is a memorable past.

List of publications

1. **Yayun Wang,** Kei Okugawa, Emi Kunitake, Makiko Sakka, Tetsuya Kimura, Kazuo Sakka. (2018). Development of an efficient host - vector system of *Ruminiclostridium josui*. Journal of basic microbiology, 58, 448-458.

2. Yayun Wang, Makiko Sakka, Haruka Yagi, Satoshi Kaneko, Hirotaka Katsuzaki, Emi Kunitake, Tetsuya Kimura, Kazuo Sakka. (2018). *Ruminiclostridium josui* Abf62A-Axe6A: A tri-functional xylanolytic enzyme exhibiting α-l-arabinofuranosidase, endoxylanase, and acetylxylan esterase activities. Enzyme and Microbial Technology, 117, 1-8.