DISCOVERY AND FUNCTIONAL CHARACTERIZATION OF MULTIPLE HEMICELLULOSE-RESPONSIVE TRANSCRIPTIONAL REGULATORS IN THE CELLULOLYTIC *STREPTOMYCES* SP. SIREXAA-E.

A Dissertation

Submitted to the graduate school of Global Food Resources

of Hokkaido university

by

Keisuke Ohashi

In Partial Fulfillment of

the Requirements for the Degree

of

Doctor of Philosophy.

January 2022

Hokkaido university

Sapporo, Japan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1. GENERAL INTRODUCTION

Recent world energy situation and biomass refinery.

The world's energy demands are expected to increase because of the ever-lasting growth of the world population and the standard of living rises in the developing world. The U.S. Energy Information Administration reported that the global annual total energy consumption in 2018 (600 quadrillion British thermal units) increased two times than that in 1980 (293 quadrillion British thermal units) (U.S. Energy Information Administration, https://www.eia.gov accessed in Oct 2021). In 2020, Fossil resources accounted for almost 70% of primary energy consumption, followed by renewable resources (～25%). These fossil resources are used not only for energy production but also for the starting materials of a wide range of fine chemicals and synthetic polymers in large-scale manufacture. However, they are limited since the current fossil resources were made very slowly from the remains of prehistoric dead plants and animals (1, 2) Furthermore, the overconsumption of fossil resources causes serious environmental damage, for instance, global warming, and water and air pollution (3, 4) .To date, many researchers have been searching for alternative resources that are eco-friendly and can utilize continuously.

Among potentially sustainable energy and chemicals resources, one of the promising resources is plant biomass. The plant biomass is mainly separated into edible crops and lignocellulose biomass (5–7) .The former includes corn and sugarcane, which provide fermentable sugars such as glucose and other soluble sugars. Fermentable sugars can be converted to alchols, lactic acid, and others by fermenter microbes, and used for heat generation and motive power of transportation (8, 9) .More importantly to our modern life, these products are also used as starting materials for the production of bioplastics (10). Although bio-based products from edible crops are available worldwide, their utilization put pressure on the food supplies (5, 11) .Recently, lignocellulose biomass derived from plant cell walls has been receiving a lot of attention as an alternative resource, which is the most abundant biomass on terrestrial, and its usage does not compete against the food demands. Additionally, a part of waste organic materials from plant origins or agricultural industry can be used to produce bio-based products (12, 13) .The conversion from lignocellulosic materials to bio-based products requires one more processing step, called saccharification followed by fermentation, compared to the conversion of edible

biomass. Saccharification is a process whereby polysaccharides are broken down into fermentable sugars, including glucose, xylose, and other monosaccharides. In the following section, the major composition of lignocellulosic materials and the saccharification process for downstream bioproduction will be described.

Lignocellulose composition

Lignocellulose from plant cell walls composes of polysaccharides (65-75% of total weight), lignin (18-35%), ash (usually 4-10%), and other minor materials including minerals and salts depending largely on plant species (14, 15). Cellulose, hemicellulose, and lignin are present in the largest amount in the terrestrial environment. As shown in Fig. 1.1A, cellulose, the main polysaccharide of plant biomass, is a polymer of glucose units joined by β-1,4-glycosidic bonds and form a highly crystalline structure due to the presence of abundant intramolecular hydrogen bridges. Because of the crystalline structure of cellulose, cellulose fibre is considerably more resistant to enzymatic hydrolysis than the starch, in which glucose molecules are linked by α-glycosidic bands (16). Thus, efficient enzymatic hydrolysis of cellulose requires various enzymatic reactions, including endo- and exo- glycoside hydrolase activities and lytic polysaccharide monooxygenase activity, and others, as described later. Hemicelluloses, including xylan, mannan, and others, have a backbone composed of various types of pentose and hexose and occasional side chains (17). Hence, complete decomposition of hemicellulose requires various types of glycoside hydrolases (18). Among hemicelluloses, xylan is the second largest natural polysaccharide next to cellulose, and different amounts of xylan can be present, depending on the types of plant species. For example, xylan contents in hardwood (10-17%), softwood (5-7%), and non-woody plant such as perennial grasses (～20%) have been reported (19). Xylose polymers include glucuronoxylan, arabinoxylan, and glucuronoarabinoxylan (Fig. 1.1B) (20). Xylan has the main chain of β-1,4-linked xylopyranose and contains a wide variety of substituted side groups such as acetyl, L-arabinofuranosyl, and 4-*O*-methylglucuronyl residues at the position of O-2 and/or O-3 (21). Mannan is another type of representative hemicellulose, found in hardwood (1-3%), softwood (10-12%), and non-woody plants (～0.4%). Mannose polymers can be found as a form of pure β-1,4-mannose polysaccharide, β-mannan, a mixture of β-1,4-linked glucose and mannose linear polysaccharide, glucomannan, mannan with occasional α-1,6-galactose polymer, galactomannan, and galactoclucomannan (Fig.

1.1C). Mannan and glucomannan may also link with various types of branched sugars, including xylose, arabinose, and other monosaccharides (22) .

Lignin is aromatic polymer that is present mainly in the secondarily plant cell walls, and the second major polymers after cellulose, among the terrestrial biopolymers, accounts for approximately 30% of the organic carbon in the biosphere (23) .The lignin structure is known to contribute to mechanical strength of the plant cell walls and potentially improving defenses against pathogens (24) .The linkages between its building blocks called monolignols are heterogenous, resulting in recalcitrant structure to chemical and enzymatic breakdown (25, 26) .

Polysaccharide degradation

Although lignocellulosic materials are promising renewable resources, their recalcitrant characteristic and the high processing costs have hindered their utilization to date (27, 28). Thus, further studies on lignocellulose-degrading enzymes are essential to address this challenge, and achieving the efficient breakdown of lignocellulosic materials is considered to be urgent demand in order to solve fuel problems and global warming issues (29). Most lignocellulose-degrading enzymes are cataloged into the Carbohydrate Active enZymes (CAZymes) based on their amino acid sequence similarities and protein structure and function, except for lignin-decomposing enzymes such as peroxidases and laccases (30). Currently, CAZymes include glycoside hydrolases (GHs), glycoside transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs). As of today, 172 families of GHs, 114 families of GTs, 42 families of PLs, 19 families of CEs, and 17 families of AAs are detailed in the CAZy database (31).

Based on the traditional idea, a complete enzymatic hydrolysis of cellulose requires three types of glycoside hydrolases, including endoglucanases (EGs), cellobiohydrolases (CBHs), and β-glucosidases. EGs randomly cut internal β-1,4-glycosidic bonds, resulting in a rapid decrease in chain length and creating free ends. CBHs attack both reducing and non-reducing ends of cellulose chains to release short cello-oligosaccharides such as cellobiose and cellotriose. β-glucosidases hydrolyze soluble cello-oligosaccharides into glucose. Thus, for efficient cellulose degradation, the synergistic involvement of these enzymes is thought to be essential (32–34). Besides the above three enzyme functions, recent studies reported the new class of enzyme for efficient cellulose-degradation, lytic polysaccharide monooxygenases

(LPMOs) from AA families, which oxidatively cleave crystalline-arranged cellulose chain in the presence of divalent metal ions and an electron donor (35). Domain structures of cellulases usually show a catalytic domain and one or more associated carbohydrate binding modules (CBMs), which have an important role in substrate-enzyme interaction (36). Hemicellulose-degradation requires various CAZymes due to heterogeneous sugar composition and sugar linkages. In xylan degradation, the backbone of xylan is broken down by multiple endo-1,4-xylanases (37, 38). Acetyl xylan esterase catalyzes the hydrolysis of *O*-linked acetyl groups form O-acetylxylose residues, resulting in improvement of the accessibility of xylan to the endo-1,4-xylanases (39). The synergistic action between xylanases and acetyl xylan esterases has been reported to increase xylan degradation efficiency for industrial setups (40–42). In addition, β-xylosidases depolymerize xylo-oligosaccharides into their constituents, mainly xylose. Additionally, α-L-arabinofuranosidase and α-glucronidase are required to remove side chain sugars that are branched at various points of xyose molecules on xylans (43, 44). Other than the above xylan substrates, several β-xylanases have also been reported to decompose transglycosylate xylo-oligosaccharides such as xylohexaose (45). The decomposition of plant mannans can be routinely executed by reacting with endo-1,4-β-mannanases, which cleave randomly within the β-1,4-mannan main chain of β-mannan, galactomannan, glucomannan, and galactoglucomannan (46). Like a β-xylosidases in the xylan decomposition, β-mannosidases are essential for a complete hydrolysis of mannans. Endo-1,4-β-mannanases and β-mannosidases synergistically break down mannan-containing polysaccharides (22, 47). Additionally, β-glucosidases and α-galactosidases are required to remove side chain sugars that are branched at various points of mannose molecules on mannans. Other than the above mannan substrates, several β-mannanases have also been reported to decompose transglycosylate manno-oligosaccharides, which structure is galactomannan and galactoglucomannan (48, 49).

Lignin is a complex aromatic heteropolymer formed by radical polymerization of p-hydroxy- phenyl (H), guaiacyl (G), and syringyl (S) units linked by β-aryl ether β-linkages, biphenyl bonds and heterocyclic linkages (26). The composition of three main components considerably varies depending on the plant species. Hardwood lignin composes of similar levels of G and S units, and tiny amount of H units (50). In contrast, hardwood lignin consists mostly of G units and very low levels of H units. Generally, depolymerization of lignin requires four enzymes including lignin

peroxidases (LiPs), manganese peroxidases (MnPs), versatile peroxidases (VPs) and laccases (51). These enzymes catalyze the oxidation of lignin substructure model compounds. LiPs oxidize side chain- and aromatic ring cleavage products of both phenolic and non-phenolic substrates. MnPs oxidize phenolic and non-phenolic moieties with some mediators. Laccase is a copper phenoloxidase that uses oxygen as an electron acceptor and oxidizes phenolic compounds. Enzymatic decomposition of lignin polymer in the plant cell wall is very challenging due to its high heterogeneity and complex structure (26). Thus, current strategy for plant biomass degradation is to separate lignin from lignocellulose fractions by chemical pretreatment such as ammonia fiber expansion (AFEX) and IL(Ionic liquid) (52, 53). Delignified plant biomass can be hydrolyzed by commercially available enzyme cacktails, which contain a set of cellulases and hemicellulases (54, 55).

In the delignified plant biomass degradation, hemicellulases and cellulases show synergistic effects on biomass. For example, removal and disruption of hemicellulose by hemicellulases help access cellulases to cellulose. On the other hand, cellulases reduce the crystallinity of the cellulose and reduce the degree of polymerization of cellulose, increasing hemicellulolytic activities (56–58). Therefore, various types of CAZymes are required for the efficient and complete decomposition of plant biomass. In the next section, I will describe microorganisms that produce polysaccharide-degrading enzymes, especially cellulolytic bacterium.

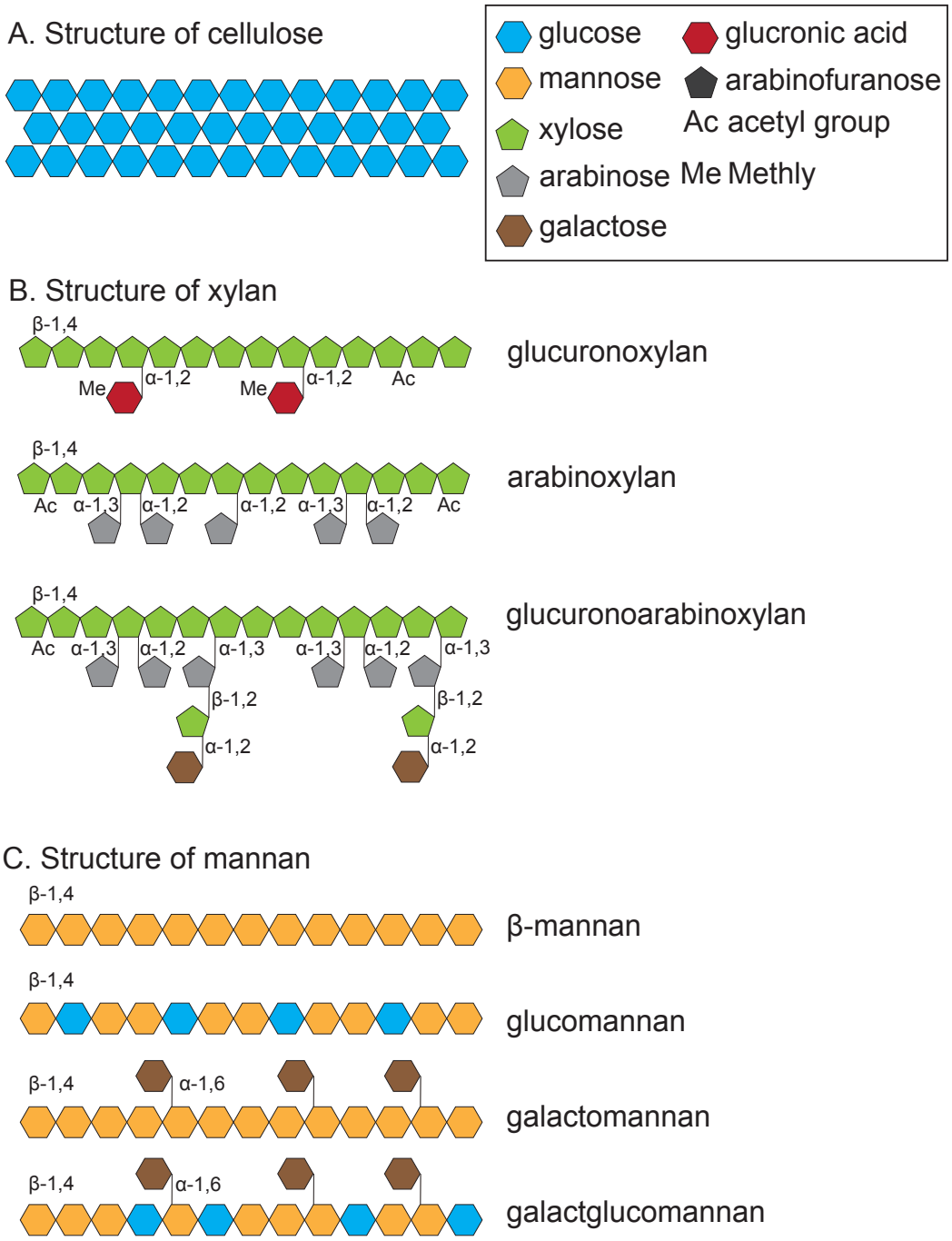


Figure 1.1 General structure of three plant polysaccharides.

Hexose (glucose, blue; mannose, orange; glucronic acid, red; galactose, brown) and pentose (xylose, green; arabinose, grey; arabinofuranose, black) are represented by hexagons and pentagons, respectively. Two symbols, Ac and Me, indicate acetylation and methylation, respectively.

Biomass-degrading microorganisms

In the field of cellulolytic microbes, fungi have been investigated rather than bacterial species. Especially, *Aspergillus* spp. and *Trichoderma* spp. are well described in their production of extracellular enzymes that break down cellulose, hemicellulose, and lignin polymers (59–61). However, a knowledge of bacteria for lignocellulosic degradation had been very limited since the plant cell wall capacity of bacteria was considered not to outcompete with reported fungi. The cellulolytic capabilities of bacteria are just beginning to be explored and have received research attention in the past several years. Nevertheless, there is no reported lignin-degrading ability in bacteria, the bacteria-based lignocellulolytic enzymes production has the following advantages. First, it is generally considered that bacteria grow more rapidly than fungi, allowing the large enzyme production in a shorter time (62). Second, bacteria tolerate diverse environmental stress such as high temperature and extremely high/ low pH conditions (63). As an example of cellulolytic bacteria, anaerobic bacteria, *Clostridium* spp. were well known to produce unique enzyme complex, called cellulosome, for cellulose and hemicellulose decomposition (64, 65). For these reasons, bacteria are still thought to be important sources for biomass-degrading enzyme producers. Another example of cellulolytic bacteria is *Streptomyces* spp. *Streptomyces* is a gram-positive bacteria and a member of the genera actinobacteria. *Streptomyces* spp. have been studied extensively for their production of secondary metabolites, including antibiotics and one of the industrial targets for decades, thusfar genetics and genome manipulation technologies have been widely available (66–68). Recently, several report described the cellulose-degrading potential of *Streptomyces* that were isolated originally from wood-devastating insects such as wood wasp and burk beetle (Fig. 1.2). Among such cellulose-degrading bacteria, *Streptomyces* sp. SirexAA-E (SirexAA-E), a symbiont of the wood-devastating wood wasp, was reported to grow in the liquid media containing cellulose or xylan as a sole carbon source (69). Unlike other *Streptomyces* that are known to produce antibiotics, SirexAA-E was shown to possess a high cellulose-degrading capability, comparable to the reported cellulolytic fungi (Fig. 1.2). SirexAA-E secretes EGs, CBHs, and AAs that altogether break down cellulose into cellobiose and cellotriose, then transported into the cell and converted into glucose by intracellular β-glucosidases. In contrast, a minimum set of xylan-degrading enzymes and mannan-degrading enzymes is used to decompose hemicellulose polymers by this bacterium. For example, a few endoxylanases and acetyl esterase were determined in

the culture supernatant derived from xylan grown medium, and two mannosidases were determined in the culture supernatant derived from the mannanoligosacchride-containing medium. To note, it has not been determined no lignin-degrading enzyme nor activity by SirexAA-E. In nature, its host wood wasp, *Sirex noctilio*, makes a pole on the pine and lays eggs with symbiotic fungi and bacteria that feed on the larvae until they hatch out (70). A series of biochemical analyses demonstrated that the specific activities of cellulose degradation in the culture supernatant of SirexAA-E, when grown in the cellulose as a sole carbon source, showed around 40% of the specific activity of the commercial enzyme cocktail, Spezyme (71). Importantly, when the SirexAA-E culture supernatant was prepared from the xylan grown condition, the specific activity of hemicellulose degradation was 20% higher than the Spezyme. In addition, transcriptome and proteome analyses of SirexAA-E showed that SirexAA-E secreted a specialized set of enzymes depending on the carbon sources used to grow the cells, and enzymes can be GH5, 6, 9, 48 families for cellulose degradation and another GH5, 10, 11 families for hemicellulose degradation (71–73). The knowledge about how the SirexAA-E controls gene expression of the above enzymes was limited except for the well-described cellobiose responsive transcriptional regulator, CebR, in *S. griseus*, and more recently in SirexAA-E (71, 73). In the absence of cellobiose, which is the main end-product of cellulose degradation, the CebR binds to a conserved 14 bp palindromic promoter sequence, 5'-TGGGAGCGCTCCCA-3', and blocks RNA polymerase from transcribing the downstream genes encoding cellulases. On the other hands, when cellobiose is available in the growth condition, CebR is released from the DNA, resulting in the expression of downstream genes. However, hemicellulose-specific responses of SirexAA-E are elusive to date.

In my doctoral research, I have aimed to understand how the SirexAA-E responds to available carbon sources and control hemicellulose-degrading enzymes. Especially, I have been focusing on the two major hemicelluloses, mannan and xylan, response of SirexAA-E.

In chapter 2, molecular mechanisms of how SirexAA-E responds to mannan is described. Using a combination of genomic and biochemical approaches, mannose and mannobiose specific regulator, SsManR, was determined.

In chapter 3, The response of SirexAA-E to xylan is detailed by introducing several xylooligosaccharide-responsive regulators. Using proteomic, genomic, and biochemical approaches, I proposed the model of the transcriptional regulation of SirexAA-E to xylooligosaccharides.

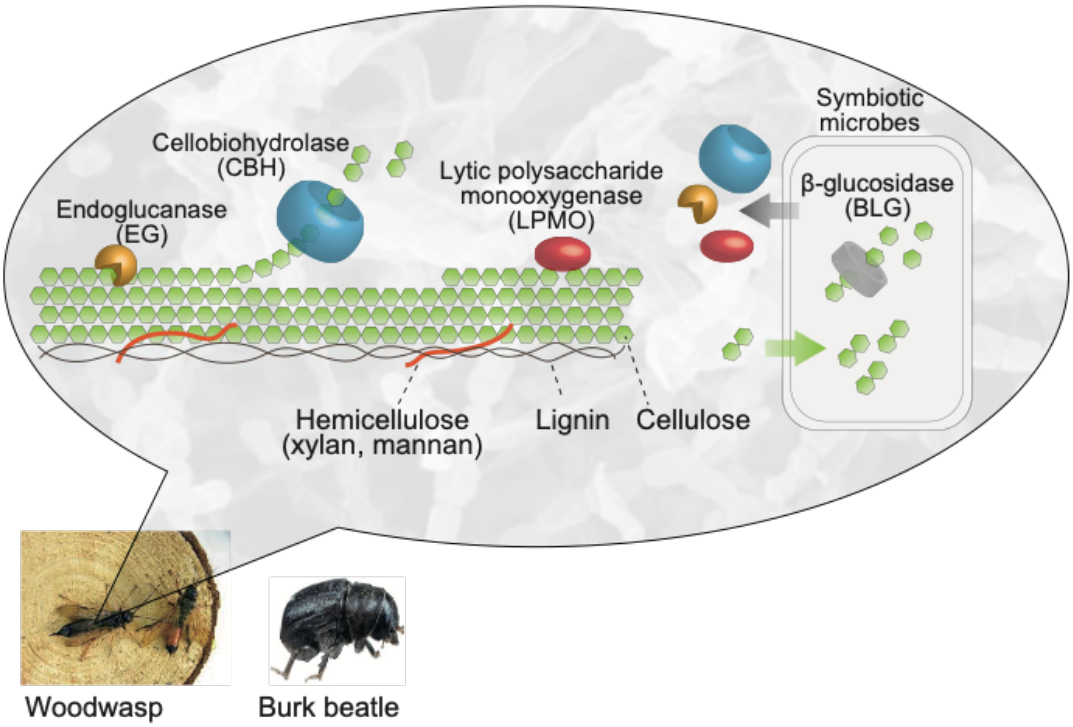


Figure 2. Polysaccharide degradation of insect symbiotic microbes.

CHAPTER 2. MANNOSE AND MANNOBIOSE SPECIFIC RESPONSES OF INSECT ASSOCIATED CELLULOLYTIC *STREPTOMYCES*

Abstract

The cellulolytic insect symbiont bacterium, *Streptomyces* sp. SirexAA-E (SirexAA-E) secretes a suite of Carbohydrate Active enZymes (CAZymes), which are involved in the degradation of a various polysaccharides in the plant cell wall, in response to the available carbon sources. Here, we examined a poorly understood response of this bacterium to mannan, one of the major plant cell wall components. SirexAA-E grew well on mannose, carboxymethyl cellulose (CMC), and locust bean gum (LBG) as sole carbon sources in the culture medium. The secreted proteins from each culture supernatant were tested for their polysaccharide-degrading ability, and the composition of secreted CAZymes in each sample was determined by LC-MS/MS. The results indicated that mannose, LBG, and CMC induced the secretion of mannan and cellulose-degrading enzymes. Interestingly, two α-1,2-mannosidases were abundantly secreted during growth on mannose and LBG. By genomic analysis, we found a unique 12 bp palindromic sequence motif at 4 locations in the SirexAA-E genome, two of which were found upstream of the above-mentioned α-1,2-mannosidase genes, along with a newly identified mannose and mannobiose-responsive transcriptional regulator, SsManR. Furthermore, the previously reported cellobiose-responsive repressor, SsCebR, was determined to also use mannobiose as an effector ligand. To test whether mannobiose induces the sets of genes under the control of the two regulators, SirexAA-E was grown on mannobiose, and the secretome composition was analyzed. As hypothesized, the composition of the mannobiose secretome combined sets of CAZymes found in both LBG and CMC secretomes, and so are likely under the regulation of both SsManR and SsCebR.

Introduction

Streptomyces is one of the largest genera of Actinobacteria, and thousands of isolates from both free-living and eukaryotic symbionts have been reported (73–75). Because of the great potential for use of their secondary metabolites for human health and other applications, the gene regulation, metabolic pathways, and physiology of *Streptomyces* have been extensively studied for decades (76–78). Recently, several insect-associated *Streptomyces* were characterized for their high cellulolytic capabilities, with a promise to advance the technology of biofuels and biochemicals production in the emerging bioeconomy. One such *Streptomyces*, *Streptomyces* sp. SirexAA-E (SirexAA-E) was originally isolated from the pine boring wood wasp, *Sirex noctilio* (69), and its cellulolytic potential to degrade plant biomass was shown to be comparable to the industrial fungal strain, *Trichoderma reesei* RutC-30 (71). In the presence of plant cell wall materials, SirexAA-E secretes a specific set of Carbohydrate Active enZymes (CAZymes), which are involved in the degradation, transport, and metabolism of carbohydrates. For example, a suite of cellulose-degrading enzymes was found in the culture supernatants when cells were grown on cellulose and cellobiose, and the latter dimeric sugar is the major end product in the cellulose hydrolysate. In contrast, a limited number of enzymes were identified in the medium containing glucose as a sole carbon source (72). The response of *Streptomyces* to available carbon sources has been reported, and our previous studies identified the transcription regulator of SirexAA-E, SsCebR (encoded by the SACTE_2285 gene), which regulates transcription of itself and a set of genes encoding cellulose-degrading enzymes (71). SsCebR belongs to the LacI-like repressor family, and a homolog in the *S. griseus*, SgCebR, was shown to bind inverted repeat DNA motif (CebR-box: 5'-TGGGAGCGCTCCCA-3') in the *S. griseus* genome (79). By *in vitro* study, SgCebR was shown to be dissociated from the genomic DNA when the effector ligands cellobiose, cellotriose, and other soluble cellooligosacharides were present, which in turns activates genes involved in glucose catabolism in the cells. While the identical DNA motif was found in the genome of SirexAA-E, and SsCebR was demonstrated to be a repressor of itself and the above-mentioned cellulose-degrading enzymes by deletion of the *sscebr* gene, yet an *in vitro* study has not been reported (73). It is currently thought that SsCebR senses available cellooligosaccharides from the surrounding environment and lets SirexAA-E synthesize cellulose-degrading enzymes for the plant cell wall utilization via a derepression mechanism.

In addition to abundant secretion of cellulose-degrading enzymes, SirexAA-E can degrade other polysaccharides in the plant cell wall, including at least xylan, mannan, and pectin. However, the molecular mechanisms of how SirexAA-E regulates genes encoding hemicellulose-degrading enzymes are mostly elusive to date. Furthermore, in pinewood, mannan is one of the major hemicelluloses, yet the responses of *Streptomyces* spp. to mannan and mannooligosaccharides have not been described (80).

In the current study, a mannose and mannooligosaccharide-responsive novel repressor was identified by a combination of proteomics, biochemistry, and genomic approaches. Furthermore, the previously reported cellobiose-responsive repressor was shown to also use mannobiose as an effector ligand. These findings let us draw a transcriptional network model for mannose and mannobiose-dependent gene expression in SirexAA-E.

Materials and Methods

General reagents

Polysaccharides, including carboxymethyl cellulose (CMC) (Sigma-Aldrich, MO, USA), locust bean gum (LBG) (Wako Pure Chemical Industry, Ltd., Osaka, Japan), D-mannan (Megazyme, Ireland), and monosaccharides, including D-glucose (Sigma-Aldrich, MO, USA), D-mannose, D-galactose (Wako Pure Chemical Industry, Ltd., Osaka), and D-mannobiose (Megazyme, Ireland) were used to prepare the culture medium. All primers used in this study were purchased (Eurofins genomics, Tokyo, Japan) (Table 2.1).

HPLC analysis of soluble sugars and LBG materials.

Soluble sugars in glucose, mannose, galactose, cellobiose, and LBG were determined using an HPLC equipped with a Chromaster 5450 refractive index (RI) detector (Hitachi, Tokyo, Japan) hooked up with a Shodex SUGAR SP0810 column (Shodex, Tokyo, Japan). Degassed Milli-Q water was used as the mobile phase, with the flow rate at 0.6 mL/min at 70˚C.

Growth on different carbon sources

SirexAA-E was pre-cultured in YME medium (4 g yeast extract, 10 of malt extract and 4 g glucose in 1L of distilled water) for 3 days at 30˚C. For the growth assay, 2.0 mL cultures containing 0.5% each of the following carbon sources, including glucose, mannose, CMC, LBG, or mannobiose, with three biological replicas in M63 (10.72 g K_2HPO_4, 5.24 g KH_2PO_4, 2 g of $(NH_4)_2SO_4$ in 1L of distilled water) supplemented with 1mM $MgSO_4$ and thiamin (10 mg/mL) was grown for 3 days at 30 ˚C. The average and standard deviation of dry cell weights were estimated based on three biological replicates. For enzyme activity measurements and proteomic analysis in the culture supernatant, 10% (v/v) of pre-cultured cells in the presence of glucose as a sole carbon source was inoculated into 50.0 mL of M63 minimal medium containing 0.5% of each of the following substrates, glucose, mannose, CMC, LBG, or mannobiose as a carbon source, and cultivated for 3 days at 30˚C. The average and standard deviation of protein concentration in the culture supernatants were estimated based on three biological replicates.

Preparation of the culture supernatant for biochemical analyses

The culture supernatants were collected by centrifugation for 15 min at 4,000 x *g* at 4˚C to remove insoluble polysaccharides and cells. The supernatant fraction was then filter-sterilized using a 0.45-μm glass fiber filter (AS ONE Co., Osaka, Japan). To remove residual soluble sugars from the monosaccharide containing culture medium and also to adjust the protein final concentration to be ~0.1 mg/mL, the supernatants were buffer exchanged then concentrated with 20 mM sodium phosphate (pH 6.0) using a centrifugal ultrafiltration (VIVASPIN 20, Sartorius Stedim, Goettingen, Germany). The concentration of proteins in each culture supernatant was determined by the BCA protein assay (Bio-Rad, Hercules, CA).

Enzyme activity measurements

The specific activity of the culture supernatant from three biological replicates of SirexAA-E from different carbon sources was determined for CMC, xylan, and mannan. 2 μg of secreted proteins from each culture supernatant was incubated with 10 mg/mL substrate in 10 mM phosphate buffer, pH 6.0, for 20 h at 40˚C, in technical triplicate. The reactions were halted by heating for 5 min at 95˚C. The amount of soluble reducing sugar release was measured by the dinitrosalicylic acid (DNS) assay (81). Glucose, xylose, and mannose were used as reducing sugar standards to quantify the amount of reducing end products. The enzyme activity in each solution (μmol reducing sugar produced/L/h) was calculated and reported as the average and standard deviation.

Preparation of extracellular proteomic samples

For the proteome preparation, 30 μg equivalent of protein from the culture supernatant from each culture (three biological replicates) was precipitated by adding 25% (v/v) trichloroacetic acid (TCA), then washed two times with ice-cold acetone, and suspended in 8 M urea. Solubilized samples were diluted to 1 M urea by adding 25 mM ammonium bicarbonate followed by a reduction with 50 mM dithiothreitol for 30 min at 50˚C, and alkylation with 500 mM iodoacetamide for 30 min in the dark at 25˚C. Tryptic digestion was carried out for 10 h at 37˚C by using a 1:100 trypsin/secretome ratio (proteomic grade; Roche, Mannheim, Germany), and peptides were purified by using C18 ziptip pipette tips (Millipore, Ireland).

Proteome analyses

Mass spectra were obtained using an Easy nLC1000 liquid chromatography system with Q-exactive plus Orbitrap mass spectrometer (ThermoFisher Scientific, IL, USA), and Xcalibur software (ver. 3.1, Thermo Fisher Scientific, IL). The peptides were separated on a C18 capillary tip column (NTCC-360/75-3-125, Nikkyo Techno, Japan) by linear gradient from 5 to 30 % with two solutions: 0.1% formic acid in water (solution A) and 0.1% formic acid in acetonitrile (solution B) in 120 min. Full scan mass spectra were obtained in the Orbitrap with a scan range of 300.0 to 2,000.0 *m*/*z* with a resolution of 70,000. Proteins were identified from the acquired MS/MS spectra using Proteome Discoverer 2.1 (Thermo Fisher Scientific, IL) with all CDSs of *S.* sp. SirexAA-E reported in the previous study (71). The peptide mass tolerance was set at 10 ppm, and fragment mass tolerance was set at 0.8 Dalton. The peptide charge was set at +2, +3 and +4. The accuracy and sensitivity of peptide identification was optimized using the automatic decoy and percolator functions built into the Proteome Discoverer software. All proteomic data sets obtained in this study can be accessed via the PRIDE proteomics identification database (https://www.ebi.ac.uk/pride/) with the accession number of 456149. The secreted proteins were identified by SignalP version 4.0 (http://www.cbs.dtu.dk/services/SignalP/) (82), and the CAZyme functions were annotated by dbCAN (http://bcb.unl.edu/dbCAN2/index.php) (83).

Cloning and protein expression of transcriptional regulators

A polymerase incomplete primer extension (PIPE) method was used for cloning (84, 85), and all primers used in this study are shown in Table 2.1. SACTE_0503, SACTE_0504, and SACTE_2285 genes were amplified from the SirexAA-E genomic DNA by using a set of primer pairs shown in Table 2.1. To create the C-terminal His tag, the reverse primer was designed to encode a His_6-coding sequence (5'-TAGTAGTAGTAGTAGTAGTAG-3'), and the forward primer was designed to skip the N-terminal His_6-tagged maltose-binding protein (MBP)-coding sequence of the original pVP68K plasmid (Center for Eukaryotic Structural Genomics, Madison, WI) so that the product should possess a tag-free N-terminus and a C-terminus of 6 His residues. The amplified vector and each of the three genes were mixed and subsequently transformed into *E. coli* JM109 competent cells (Takara, Shiga, Japan). Transformants were screened by a colony PCR using the primer pairs, colony PCR_fw, and colony PCR_rv (Table 2.1), then the sequence was verified (Eurofins genomics, Tokyo,

Japan). pVP68K plasmids harboring SACTE_0503, SACTE_0504, and SACTE_2285 genes were individually transformed into *E. coli* strain Rosseta2 (DE3) (Merck Millipore, Burlington, Massachusetts, USA) for recombinant expression.

Protein expression and purification were carried out accordingly to the previous study (86). Briefly, each transformant was pre-cultured in 5.0 mL of LB medium (5 g Tryptone, 10 g yeast extract, 5 g NaCl in 1L distilled water) supplemented with 50 µg/mL of kanamycin for 16 h at 37 °C on a rotating shaker. The precultures were added to 250 mL of the LB medium supplemented with 50 µg/mL kanamycin and cultivated for 3 h at 30 °C with shaking at 250 rpm. To induce expression, isopropyl thiogalactoside was added to a final concentration of 1 mM, and incubated for 20 h at 25 °C with shaking at 250 rpm. The cells were harvested by centrifugation at 4000 x *g* for 15 min at 4˚C, then washed with ice cold 20 mM phosphate buffer (pH 7.4) two times. Cells were resuspended in 20 mL of 20 mM phosphate buffer (pH 7.4) containing 1.5 µL of Benzonase (Novagen, Wisconsin, USA) and 20 µL of 100 mg/mL of lysozyme, followed by sonication with a cycle of 10 s on and 10 s off for 10 min on ice (Branson, Danbury, CT). Cell lysates were centrifuged at 10,000 x *g* for 30 min at 4°C, and supernatants were filtered through a 0.45-µm filter (AS ONE Co., Osaka, Japan). One step Ni-affinity chromatography was performed by using an AKTA prime protein purification system (GE Healthcare, Freiburg, Germany) equipped with a 1 mL HisTrap-FF column (GE Healthcare, Freiburg, Germany) and a linear 50 mL gradient from buffer A (20 mM phosphate buffer, pH 7.4, containing 500 mM NaCl, 25 mM imidazole) to buffer B (20 mM phosphate buffer, pH 7.4, containing 500 mM NaCl, 500 mM imidazole). The fractionated proteins were concentrated and buffer-exchanged into 20 mM phosphate buffer, pH 7.4, using a VIVASPIN 20 (Sartorius Stedim, Goettingen, Germany Sartorius AG, Germany) at 4000 x *g* to a final concentration of ~1 mg/mL at 4˚C. The amount of purified protein was quantified by Bradford assay and store at -80°C until use.

Table 2.1. Primers used in this study.

Name	Sequcence (5'-3')
pVP68K_vector for CHis fw[a]	CATCATCATCATCATCATTGAGTTTAAACGAATTC
pVP68K_vector for CHis rv[a]	GGTTAATTTCTCCTCTTTAATGAATTCTGTGTG
SACTE0503_fw[a]	ATTAAAGAGGAGAAATTAACCATGCAAACCGACCTCGTCGCC
SACTE0503_rv[a]	TCAATGATGATGATGATGATGCTGCGCCGCCACCGCCCGTAC
SACTE0504_fw[a]	ATTAAAGAGGAGAAATTAACCATGGCCGAGACCGCCCGCCATC
SACTE0504_rv[a]	TCAATGATGATGATGATGATGGGCGGGCGGGAGTTCGCCG
SACTE2285_fw[a]	CTGTACTTCCAGTCCATGGCGGCAGCGCGAGTACG
SACTE2285_rv[a]	CGAATTCGTTTAAACTCAGGACGACTCGCGGACCA
colony PCR_CHis_fw[a]	ACGAGGCCCTTTCGTCTTCACC
colony PCR_rv[a]	CGGTGGCAGCAGCCAACTC
Prom_SACTE0383_fw[c]	GCGAGGGCGCCGGCGAGGGCTAC
Prom_SACTE0383_rv[c]	AAGGCGTACGTCCAGCGAATCAC
Prom_SACTE0504_fw[c]	ATCTTGGTGCCGCCGATGTC
Prom_SACTE0504_rv[c]	CCGGGTTCGCTGTTCACCACC
Prom_SACTE5235_fw[c]	TAGCCGTGTCTCTTTCGGGAGTC
Prom_SACTE5235_rv[c]	GCCGAACGCTTGATCACATCGCG
Prom_SACTE6282_fw[c]	CGAGTGAGCCCGCCAGGAGCAC
Prom_SACTE6282_rv[c]	GACGCCCGTGGACGACGGGTGAC
Prom_SACTE2285_fw[c]	GTACGGTTCCCCGCGAGC
Prom_SACTE2285_rv[c]	GCGCGTCCGCCTGCCCTTC

[a]Primers used for cloning plasmids.

[b]Primers used for clone verification.

[c]Primers used for amprification of DNA fragments.

Electrophoresis mobility shift assay (EMSA) and effector assay

Prom_0383, Prom_0504, Prom_2285, Prom_5235, and Prom_6282 were amplified from the SirexAA-E genomic DNA using a set of primers shown in Table 2.1. The DNA concentration was estimated by the DS-11 nano liter spectrophotometer (DeNovix, IL, USA). The binding assay was performed accordingly to the previous study with some modifications (79). Briefly, 10 ng of each DNA fragment was mixed with a purified transcriptional regulator in 10 µL of binding buffer (100 mM Tris-HCl, pH 7.4, containing 5 mM EDTA, 50 mM $(NH_4)_2SO_4$, 5 mM DTT, 150 mM KCl, 1.0% (v/v) Tween-20, and 5.0% (v/v) glycerol), and incubated for 15 min at 25°C. After the reaction, the samples were run on 2.0% TAE agarose at 50 V for 3 h at 5˚C, and the gels were stained with ethidium bromide. The amount of bound and unbound fractions on the gel was quantified by using a Gel Doc™ EZ imager system (Bio-Rad, Hercules, CA, USA). In each assay, bovine serum albumin (BSA) was used as a negative control. The average and standard deviation were estimated based on three independent experiments. The apparent K_D values of SACTE_0504p for each DNA fragment were estimated based on the DNA concentration used in the reaction, which gives approximately 50% of free DNA to the complex by EMSA. The formula used for calculation of the apparent K_D value is as follows (87):

[complex] = [protein]*[free DNA]/ (K_D +[free DNA]),

where [complex], [protein], and [free DNA] indicate the concentration of the DNA-protein complex, protein, and free DNA, respectively.

Result

Growth of SirexAA-E on medium containing mannose and galactomannan (LBG)

SirexAA-E can grow on plant cell wall polysaccharides and produce various extracellular enzymes depending on the available carbon sources present in the growth medium. We initially tested whether SirexAA-E utilizes mannose and mannose-containing polysaccharides, LBG, together with glucose and CMC as sole carbon sources (Fig. S2.1). As previously reported, SirexAA-E grew on glucose and CMC, and here we also show growth on mannose and LBG. After 3 days cultivation at 30˚C, the dry cell weight (mg/mL) was measured for growth on glucose (115±25 mg/mL), mannose (65±4 mg/mL), CMC (33±5 mg/mL), and LBG (93±15 mg/mL), respectively. To test the possibility that there might be free sugars present in the original LBG material as received, we analyzed soluble sugars in the LBG by HPLC and found ~25 µg of glucose in 50 µg of LBG; no other sugar derivatives were found (Fig. S2.2). Thus, the growth of SirexAA-E in the LBG culture was initially due to the utilization of glucose. To note, when the SirexAA-E was grown on pure mannan, no sign of growth was determined (data not shown). In the later section, we will discuss how SirexAA-E responds to LBG in the growth medium.

Biochemical analysis of the secretomes of SirexAA-E

To test if SirexAA-E secretes enzymes that decompose defined polysaccharides in response to available carbon sources present in the growth media, culture supernatants were prepared after growth on the following carbon sources: glucose; mannose; CMC; and LBG for 3 days at 30˚C. The highest concentration of secreted proteins was obtained from SirexAA-E grown on CMC (15.1±0.9 µg/mL), followed by LBG (12.2±5.1 µg/mL), mannose (10.5±2.1 µg/mL), and glucose (1.8±0.6 µg/mL) (n=3) (Table S2.1). In the presence of glucose, the amount of secreted proteins was significantly lower than other conditions, consistent with the our previous report (71). As reported earlier, the secreted proteins obtained from the CMC culture were fairly abundant, and interestingly the secreted proteins from mannose and LBG cultures contained an elevated level of proteins compared to the glucose culture. To test whether each culture supernatant possessed unique polysaccharide-degrading capabilities, the specific activity of each culture supernatant was measured using pure CMC, xylan, and mannan as substrates (Fig. S2.3). The CMC culture supernatant showed the highest polysaccharide-degrading activities for all tested substrates tested.

In contrast, the mannose and LBG supernatants showed only ~35% of the CMCase and ~30% of xylanase activities relative to the CMC culture supernatant were found in the mannose and LBG samples, respectively. Compared to the CMC sample, and around 7% and 10% mannan-degrading activities. The glucose culture supernatant showed only minimal reactivities relative to the CMC, xylan, and mannan culture supernatants.

LC-MS/MS analysis of secretomes

To determine and compare the proteins secreted in each culture supernatant, the secretome composition of SirexAA-E grown on mannose, CMC, and LBG was analyzed by LC-MS/MS (Fig. 2.1). In Table 2.2 and corresponding Table S2.2, the number of identified proteins, and exponentially modified protein abundance index (emPAI) values (n=3) are shown (88). To note, the previously reported composition of secreted proteins in the glucose secretome was used as a benchmark, with only 7 CAZymes found among 32 secreted proteins in a total of 164 detected proteins (71). In the mannose secretome, we determined 305 secreted proteins (97 CAZymes) in 1409 identified proteins, which account for 21.6% of the total proteome. In the CMC and LBG secretomes, 172 secreted proteins (78 CAZymes) out of 805 total proteins (21.4%), and 190 secreted proteins (90 CAZymes) out of 838 total proteins (22.7%) were identified, respectively (Table 2.1, and Table S2.2). Due to a large number of secreted proteins determined by our proteomic analysis, the top 100 extracellular CAZymes in each secretome were selected based on the calculated emPAI value from 3 biological replicas except for the glucose secretome (Table S2.3).

The 7 reported CAZymes in the glucose secretome included a putative dehydrogenase (GH109: SACTE_0604), β-xylosidase (GH43: SACTE_5267), β-1,3-glucanase (GH64: SACTE_4755), β-N-acetylhexosaminidase (GH3: SACTE_2232), and 3 chitinases (GH23s: SACTE_1422, SACTE_2558 and SACTE_4281), all of which were also found in the mannose secretome dataset (Table S2.2 and Table S2.3A). In the mannose secretome, in addition to the 7 constitutively secreted CAZymes, 13 secreted CAZymes were determined, including 2 predicted α-1,2-mannosidase (GH92: SACTE_0383, and SACTE_5235), cellobiohydrolase (GH6: SACTE_0237), 3 lytic polysaccharide monooxygenases (LPMOs) (AA10s: SACTE_0080, SACTE_2313, and SACTE_6428), and 7 other CAZymes (Fig. 2.1, Table S2.2, and S3A). In the CMC secretome, 46 CAZymes were determined, and all the proteins identified in the

previous study were found in our dataset (72), which included prominent endo- and exo-type cellulases (GH48: SACTE_0236 and GH6: SACTE_0237), characterized mannanase (GH5: SACTE_2347) (89), and 4 LPMOs (AA10s: SACTE_0080, SACTE_2313, SACTE_3159, and SACTE_6428) (Fig. 2.1, Table S2.2, and S3A). In the LBG secretome, we found 52 CAZymes in the top 100 secreted proteins, and interestingly most of the CAZymes, 44 out of 52 proteins, were identical as those found in the CMC secretome (Fig. 2.2.1 and Table S2.3B).

Fig. 2.1 compares the CAZymes determined in the mannose, CMC, and LBG secretomes. Between the CMC and LBG secretomes, 2 CAZymes (CE4: SACTE_0357, and GH9: SACTE_3717) and 8 CAZymes (GH23: SACTE_1422, GH18s: SACTE_2290 and SACTE_5764, GH16: SACTE_2876, GH64: SACTE_2877, GH33: SACTE_5371, GH13: SACTE_5685, and CBM13: SACTE_6512) were found to be unique and in common for the CMC and LBG secretomes, respectively (Fig. 2.1). Among the 44 CAZymes found in both CMC and LBG secretomes, the following 8 CAZymes were thought to be regulated by the SsCebR repressor (71, 73), including the above-mentioned cellulases and mannanase (GH48: SACTE _0236, GH6: SACTE_0237 and GH5: SACTE_2347), β-1,4-endoglucanase (GH5: SACTE_0482), LPMOs (AA10: SACTE_3159 (90) and SACTE_6428), cellulose-binding protein (GH74: SACTE_0562), and β-1,4-xylanase (GH10: SACTE_0265) (Blue filled stars in the grey category in Fig. 2.1). This result suggested that both cellobiose and mannobiose play roles as an effector ligand of SsCebR since there was no detectable cellobiose in the LBG used in this study (Fig. S2.2). Between the mannose and LBG secretomes, 15 CAZymes were found in both secretomes (Fig. 2.1 and Table S2.3B). Moreover, two putative α-1,2-mannosidases (GH92: SACTE_0383 and SACTE_5235) were found in the mannose, CMC and LBG secretomes, but not in glucose secretome, yet the amount of both enzymes in the CMC secretome was low. Thus, we thought that both GH92s are regulated by the same transcriptional regulator.

Table 2.2. Summary of secretome analysis for SirexAA-E after growth on different carbon sources. The number of identified proteins, secreted proteins, percentage of secreted proteins in the total identified proteins, the numer of total CAZymes and the number of CAZymes in the top 100 secreted proteins are shown. *The data of proteome of glucose-grown culture is quoted from previous study (71).

Secretome	No. of identified proteins	No. of secreted proteins	%secreted protein	No. of CAZymes	No. of extracellular CAZymes	No. of CAZymes in the top 100 secreted proteins
Glucose secretome*	164	32	19.5	7	7	7
Mannose secretome	1409	305	21.6	96	72	20
CMC secretome	805	172	21.4	78		46
LBG secretome	838	190	22.7	90		52
Mannobiose secretome	1048	184	17.6	80		34

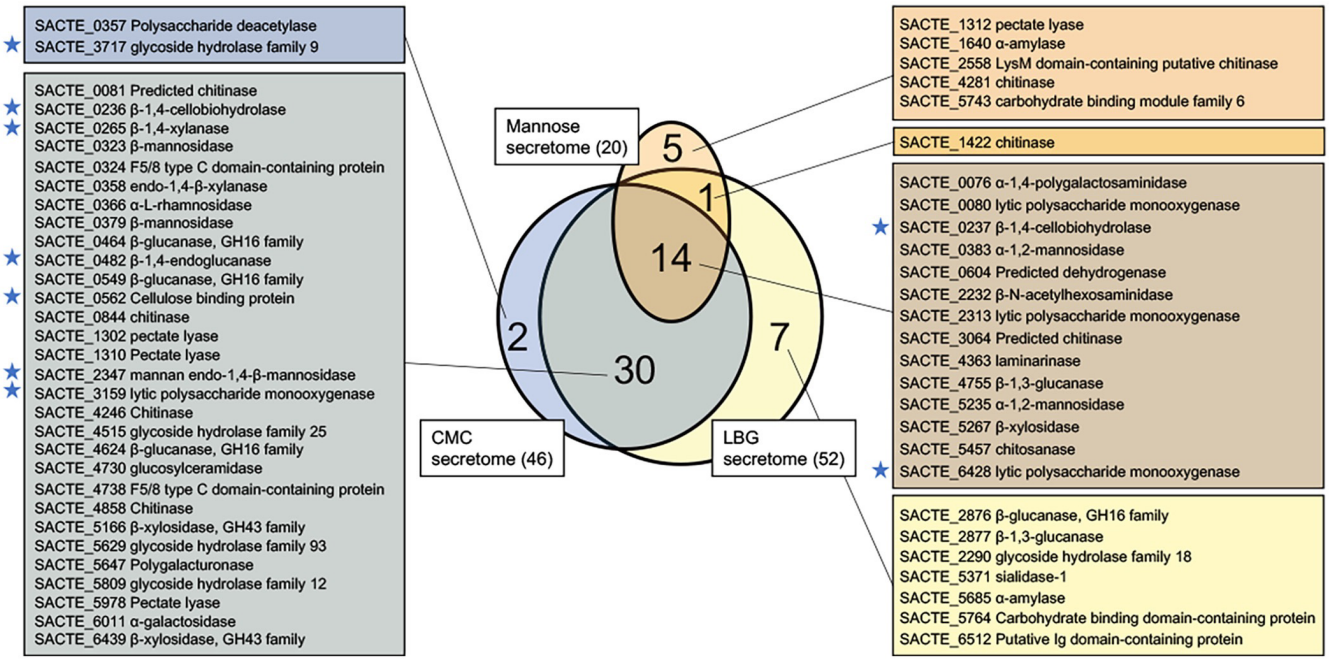


Figure 2.1. Venn diagram of three SirexAA-E secretomes: mannose; LBG; and CMC. The number in each category indicates the number of CAZymes determined by secretome analysis. The list of CAZymes determined in each category is shown with a locus number and predicted function. The blue filled stars indicated the CAZymes that are thought to be underregulated by SsCebR.

Potential regulators for mannanase activity of SirexAA-E

The two genes encoding α-1,2-mannosidases, SACTE_0383, and SACTE_5235, were most likely regulated by the same repressor protein in a mannose- or mannooligosaccharide-dependent manner. Thus, we looked at the predicted promoter regions of two α-1,2-mannosidase gene loci and found the 12 bp palindromic sequence, 5'-GACAACGTTGTC-3' from the 20 to 30 bp upstream from the start codon of two genes (Fig. S2.4A and B). Moreover, when we searched the SirexAA-E genome, two additional copies of this sequence were found, one of which was flanked by two regulator coding genes, SACTE_0503 and SACTE_0504 (Fig. S2.4C), while the other was found at the upstream from the third putative α-1,2-mannosidase coding gene (GH92: SACTE_6282) (Fig. S2.4D). The promoter region flanked by two transcription regulator families (ROK: SACTE_0503 and LacI: SACTE_0504) was particularly interesting since transcriptional regulators from these families often regulate their own gene expression (15–18). Consequently, we wanted to determine whether either or both of these might function as a mannose-responsive transcriptional regulator.

Functional characterization of two putative regulators for mannosidases

To test whether two putative regulator proteins, SACTE_0503p and SACTE_0504p, bind to the 12 bp predicted motif *in vitro*, each gene was amplified from the genome of SirexAA-E, then cloned into pVP68K vector, overexpressed in *Escherichia coli,* and purified by affinity purification methods using a designed C terminal histidine tag (Table 2.3 and Fig. S2.5). The following four promoter-containing regions were prepared from the genomic DNA: a 309 bp region (Prom_0383) located upstream from SACTE_0383 (Fig. S2.4A); a 626 bp region (Prom_5235) located upstream from SACTE_5235 (Fig. S2.4B); a 354 bp region (Prom_0504) located between SACTE_0503 and SACTE_0504 loci (Fig. S2.4C); and a 250 bp region (Prom_6282) located upstream from SACTE_6282 (Fig. S2.4D). A 350 bp of DNA fragment, including promoter region of SACTE_2285 gene (Prom_2285), was used as a negative control. The result of electromobility shift assay (EMSA) of SACTE_0504p to five promoter-containing regions is shown in Fig. 2.2 (a full image of the gel is shown in Fig. S2.6.) and summarized in Table 2.4. We did not detect apparent interaction between SACTE_0504p and the negative control Prom_2285. However, mobility shifts were seen with the four other fragments when the amount of SACTE_0504p was increased, and apparent K_D was estimated to be ~90 nM (Prom_0383), ~80 nM (Prom_0504*)*,

~150 nM (Prom_5253), and ~110 nM (Prom_6282), respectively. In contrast, SACTE_0503p did not bind any tested fragment, even when potential effector ligands, glucose, mannose, cellobiose, and mannobiose, were added at high concentrations (Fig. S2.7 and Table 2.4). Thus, we concluded that at least SACTE_0504p binds to promoter regions that contain the predicted 12 bp sequence motif, 5'-GACAACGTTGTC-3'. To ask whether the observed interaction between SACTE_0504p and the tested promoter regions can be disrupted by effector ligand, the following sugars, glucose, cellobiose, mannose, and mannobiose, were tested. In Fig. 2.3A and Table 2.4, the binding of SACTE_0504p to the Prom_0504 was disrupted by 50-300 mM of mannose and mannobiose, while glucose and cellobiose did not affect the interaction even at a high concentration of ligand (300 mM). These results indicated that SACTE_0504p, named herein SsManR, acts as a mannose/mannobiose responsive repressor that transcriptionally controls the production of GH92 mannosidases depending on the presence of mannose and mannobiose.

Table 2.3. List of proteins prepared in this study. [a]Expected molecular weight was calculated based on amino acid sequence of each protein. [b]Either one step affinity purification (His-tag purification) or affinity purification followed by immobilized metal affinity chromatography was used.

Protein	MW (kDa)[a]	Purification[b]
SACTE_0503p	32.6	C terminal His tag purification
SACTE_0504p	37.5	C terminal His tag purification
SACTE_2285p	37.4	C terminal His tag purification

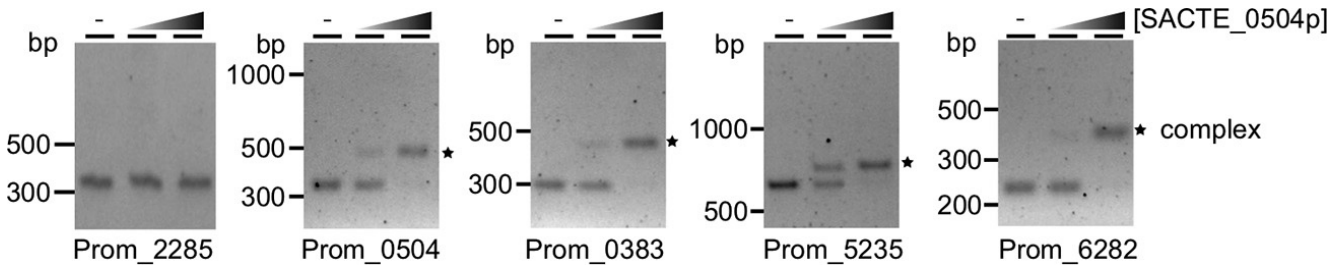


Figure 2.2. Electromobility shift assay of the SACTE_0504p for four putative promoter regions. Five PCR-amplified DNA fragments, Prom_2285, Prom_0504, Prome_0383, Prom_5235, and Prom_6282, were incubated with increasing amounts of SACTE_0504p. 0, 21, and 300 ng of SACTE_0504p were mixed with each promoter region shown in lane 1, 2, and 3, respectively in each gel. The black filled star indicated the presence of SACTE_0504p-DNA complex.

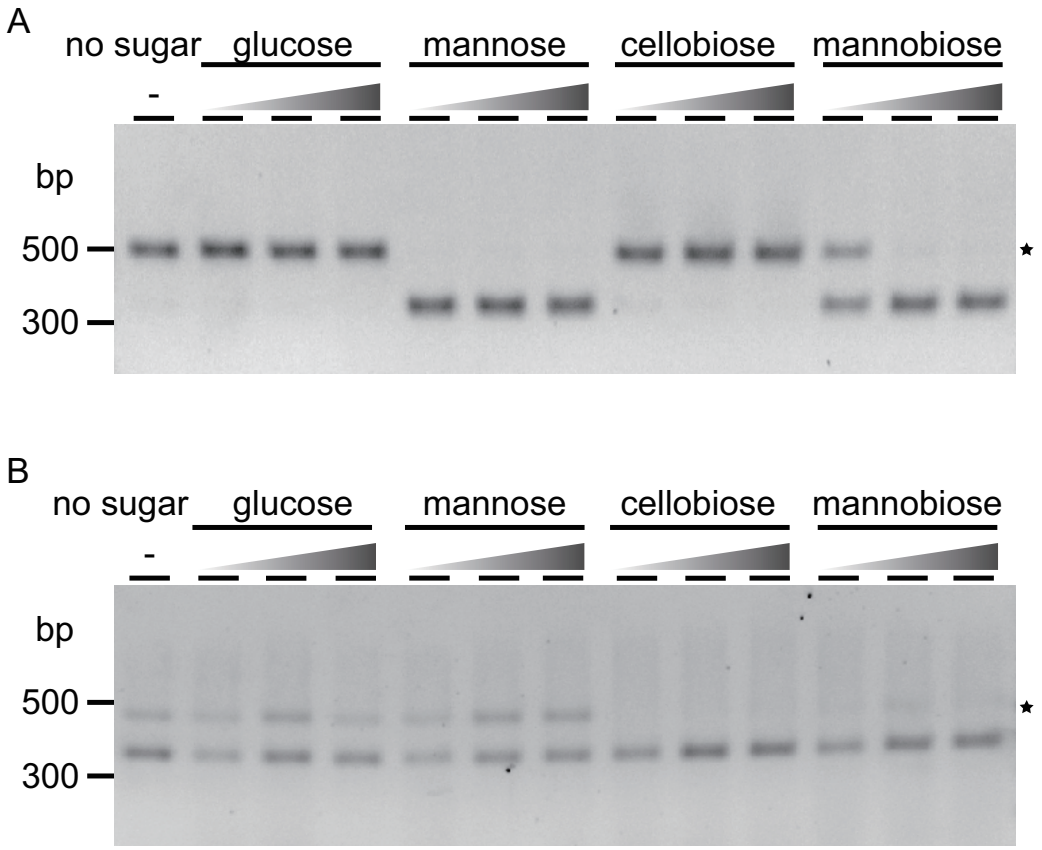


Figure 2.3. Effector assay of the SACTE_0504p and SACTE_2285p.
The effect of four different sugars, glucose, mannose, cellobiose, and mannobiose, on SACTE_0504p-Prom_0504 complex and SACTE_2285p-Prom_2285 complex, was tested. (A) 140 ng of the purified SACTE_0504p was mixed with 10 ng of Prom_0504, and 50, 150, and 300 mM indicated sugar was added. (B) 140 ng of the purified SACTE_2285p was mixed with 10 ng of Prom_2285, and 50, 150, and 300 mM indicated sugar was added. The black filled star indicated the formation of a protein-DNA complex.

Table 2.3. Summary of EMSA.

	DNA fragments					
Protein	Prom_2285	Prom_0504	Prom_0383	Prom_5235	Prom_6282	Effector ligands
SACTE_0503p	-	-	-	-	-	n.d.
SACTE_0504p	-	+	+	+	+	mannose/ mannobiose
SACTE_2285p	+	-	-	-	-	cellobiose/ mannobiose

+: the binding was obserbed., -: no interaction; ND: not determined

Mannobiose, a potential effector of SsCebR repressor

In the previous study, SsCebR (SACTE_2285) was suggested to be the transcriptional repressor that interacts with the unique 14 bp palindrome sequence (5'-TGGGAGCGCTCCCA -3'), located in the promoter regions of 15 cellulase genes (71). This DNA-protein binding activity was thought to be inhibited explicitly by cello-oligosaccharides in a dose-dependent manner (79). Our analysis of the LBG secretome suggested that mannobiose also induced SsCebR-regulated CAZymes hence mannobiose was thought to be a potential effector ligand of SsCebR. To test this, SsCebR (SACTE_2285p) was prepared (Table 2.3 and Fig. S2.5), and 350 bp of Prom_2285, containing the SSCebR-box, was also prepared for the EMSA and effector assays (Fig. 2.3B and Table 2.4). In the absence of the effector ligand, the mobility shift of Prom_2285 DNA was observed, consistent with the previous observation (73). Neither glucose nor mannose affected the interaction between SsCebR and Prom_2285. However, in the presence of 50 mM cellobiose, the mobility shift disappeared. Interestingly, in the presence of mannobiose in the EMSA reaction, 50 mM mannobiose disrupted the interaction between SsCebR and Prom_2285, suggesting that mannobiose is also an effector ligand for SsCebR.

Mannobiose secretome analysis

In the previous studies, only cellobiose and longer cellooligosaccharides were thought to be the effector molecules of SgCebR in *S. griseus* (10). Additionally, in SirexAA-E, the SsCebR was indicated to regulate a set of CAZyme coding genes downstream from the SSCebR-box in a cellobiose-responsive manner (71). However, our current finding suggested that mannobiose also influences the binding of SsCebR to SSCebR-box. To assess whether mannobiose induces expression of the SsCebR regulated genes *in vivo*, SirexAA-E was grown on mannobiose as the sole carbon source. In the presence of mannobiose, SirexAA-E grew well (32±4 mg/mL dry cell weight), and the culture supernatant was collected for the proteomic analysis (n=3). 184 secreted proteins were determined out of 1048 total proteins identified, and 34 CAZymes were included in the top 100 proteins (Table 2.1 and Table S2.2). With the mannobiose secretome, a heatmap was constructed in order to compare the secreted CAZymes from cells grown on 5 different carbon sources (Fig. 2.4). The hierarchical tree on the top, accounting for the overlap of secreted CAZymes in each dataset, indicated that the LBG secretome and CMC secretome were the closest in composition, followed by the

mannobiose, and then mannose secretomes. The CAZymes regulated by SsManR were found in all the secretomes except for the glucose secretome. Furthermore, the SsCebR regulated CAZymes were detected in the mannobiose, LBG, and CMC secretomes. By comparing between mannose and mannobiose secretomes, 12 CAZymes were found to be present in both datasets, and α-1,2-mannosidases (SACTE_0383, and SACTE_5235) were included (Fig. 2.4). In Fig. S2.8, the comparison between the mannobiose, LBG secretome, and CMC secretome is shown, where 31 secreted CAZymes are shared by all three secretomes, including the two α-1,2-mannosidases (SACTE_0383, and SACTE_5235) regulated by SsManR, and the five CAZymes (SACTE_0237, SACTE_0482, SACTE_2347, SACTE_3159, and SACTE_6428) regulated by SsCebR. Altogether the expression of the above CAZymes was shown to be regulated by the two transcriptional regulators described in this study.

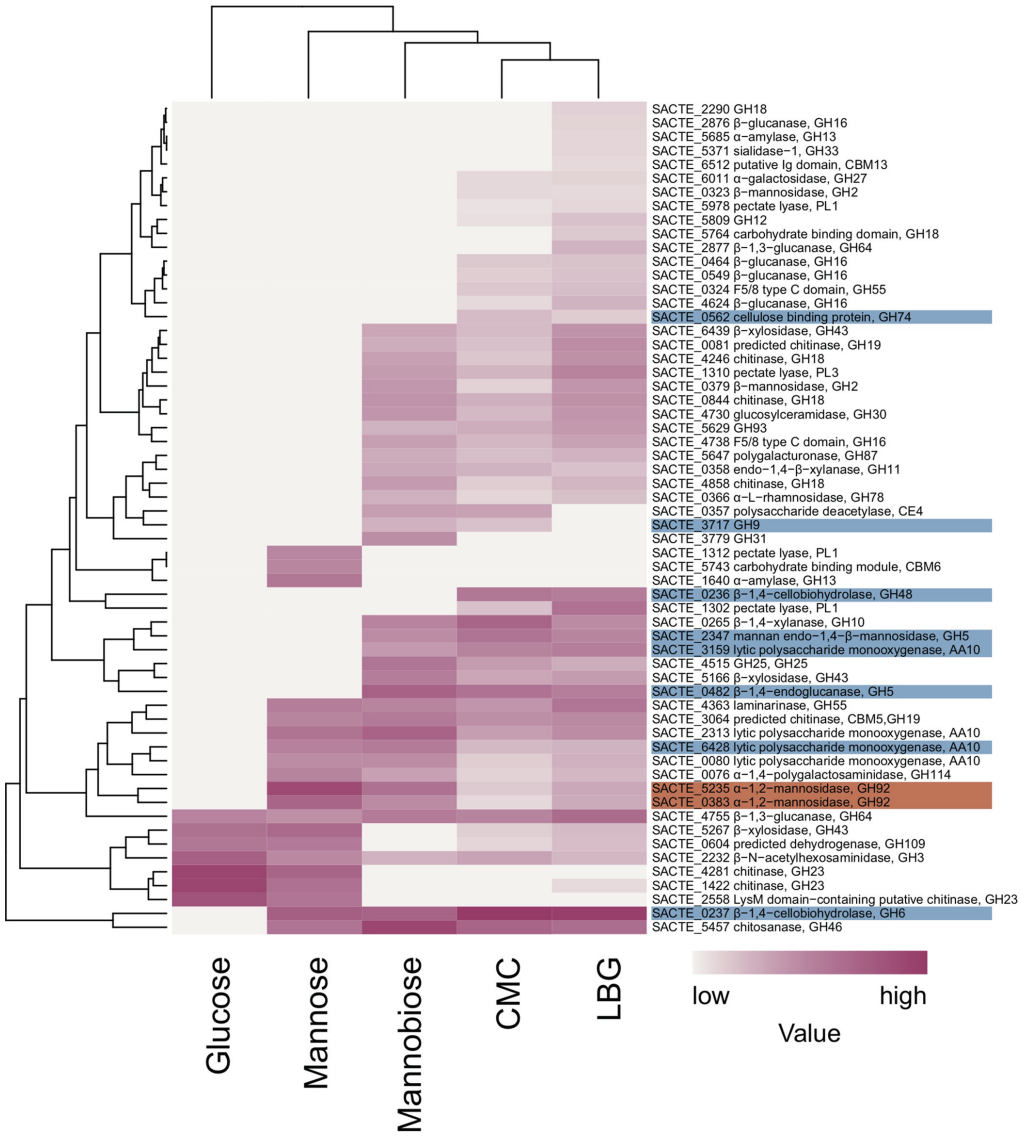


Figure 2.4. Hierarchical clustering of CAZymes in the top 100 secreted proteins identified in the glucose, mannose, mannobiose, CMC, and LBG culture supernatants of SirexAA-E. A total number of 60 CAZymes were clustered based on their similarities in protein abundance (emPAI value) indicated as color code. Possible regulations either by SsCebR or SsManR of determined CAZymes were orange or blue highlighted with corresponding locus number and putative function of each protein shown on the right.

Discussion

Prior to this work, the response of SirexAA-E to mannan, one of the most abundant hemicelluloses in the plant cell wall, particularly softwood (conifer), had not been described (95). We found CMC-, xylan-, and mannan-degrading activities in the mannose and LBG culture supernatants, suggesting that both mannose and LBG induce enzymes that depolymerize the tested substrates. The compositions of mannose, CMC, and LBG secretomes were compared (Fig. 2.1), and 14 CAZymes were found in all three secretomes. Among the 14 CAZymes, two α-1,2-mannosidases (GH92: SACTE_0383 and SACTE_5235) were of interest since their predicted functions were mannan utilization and they were also abundant in the mannose-induced secretome. Furthermore, the enzyme composition between the CMC and LBG-induced secretomes was highly overlapped, and thus we investigated whether the same transcriptional regulation occurred in both carbon source media.

With respect to the mannan response of SirexAA-E, we hypothesized two possibilities: 1) there is a novel regulator that functions in a mannose- and mannobiose- (the end product of LBG) responsive manner and regulates the expression of two α-1,2-mannosidases; 2) mannobiose is an effector ligand that prevents the binding of SsCebR to the target regulatory elements (SSCebR-box) like reported for cellobiose in the genome of SirexAA-E.

To test the first possibility, we performed genome analysis and *in vitro* binding assays (Fig. 2.2, Fig. 2.3, and Fig. S2.4) and identified a novel transcriptional regulator, SsManR (SACTE_0504p), that recognizes the motif upstream from the three α-1,2-mannosidase-coding genes. The apparent K_D of SsManR to the motif was comparable to the reported K_D of *E. coli* LacI and operator complex (96, 97) and also to that of the SgCebR and CebR box in *S. griseus* (79). Note that the apparent K_D of SsManR estimated in this study ranged from 80-150 nM, suggesting that the interaction between SsManR and the same motif in four different genome locations might depend on the surrounding sequences of the motif in each promoter (Fig. 2.2). Furthermore, we determined that both mannose and mannobiose are the effector ligands that disrupt the interaction between SsManR and Prom_0504, while mannose disrupts the interaction more effectively than the mannobiose (Fig. 2.3). SACTE_0503p did not bind to any promoter region, even in the presence of a series of mono- or di-saccharides as a potential effector ligand of activator function (Fig. S2.7).

Next, we tested the effect of mannobiose on the SsCebR function. CebR repressors in the family Actinobacteria, including *S. griseus, S. reticuli, and Thermobifida fusca*, were shown to be regulated by cellooligosaccharides, but the effect of mannooligosaccharides on these CebR repressors was not described (79, 98, 99). As shown in Figure 2.3B, both cellobiose and mannobiose disrupted the interaction between SsCebR and Prom_2285, suggesting that SsCebR does not differentiate between the two molecules, as mannobiose is a C-2 epimer of cellobiose.

With the reported genomic and *in vitro* studies, we anticipated that if mannobiose is used as a sole carbon source in the medium, the set of genes regulated under both transcriptional regulators should be upregulated. To test this, cells were grown on mannobiose as a sole carbon source, and secretomic analysis was performed. Indeed, two α-1,2-mannosidases were found, which suggested that mannobiose de-repressed their regulation under SsManR. Furthermore, a set of SsCebR regulated CAZymes was also detected in the mannobiose secretome. Thus, we proved that both transcriptional regulators respond to mannobiose, allowing an expanded set of CAZymes to be expressed.

In the mannose secretome, cellobiohydrolase (GH6: SACTE_0237), endoglucanase (GH5: SACTE_0482), and LPMO (AA10: SACTE_6428) were also found. The leaky control of expression of these genes would be driven by an extremely strong activator (100, 101). In *S. rimosus*, a member of the SARP family transcription regulator, OtcR, was shown to be an extremely strong activator for the oxytetracycline biosynthesis gene cluster (102). However, such activators for CAZyme genes in *Streptomyces* have not been determined to date. From the heatmap, 3 LPMOs (AA10s: SACTE_0080, SACTE_2313, SACTE_6428) were also found in the mannobiose, LBG, and CMC secretomes, however, the transcription factors that regulate these genes are also still elusive (Fig. 2.4).

With these new findings, we can propose a transcriptional network model that enables SirexAA-E to respond to LBG (Fig. 2.5). As was determined, LBG used in this study contained a large amount of glucose (Fig. S2.2), which lets cells grow and secrete a basal level of CAZymes that could subsequently hydrolyze LBG, including β-mannanase (GH5: SACTE_2347) (71, 89). Alternatively, in a natural symbiont environment, a series of oligosaccharides such as mannooligosaccharides would be provided by symbiont microbes that can decompose plant cell wall components (103). When cells are exposed to mannose or mannobiose, they are imported into the cell by

a putative mannose transporter (SACTE_5236, SACTE_5237, and SACTE_5238) (104), and displaces SsManR from the 4 promoter regions (Prom_0383, Prom_0504, Prom_5235, and Prom_6282). This turns on the transcription of a set of genes. At Prom_0383, at least two mannose utilization genes would be transcribed, α-1,2-mannosidase (GH92: SACTE_0383) and β-mannosidases (GH2: SACTE_0382), with both enzymes predicted to function in the hydrolysis of α, and β-linked mannose in the LBG, respectively (Fig. S2.4A) (105). Prom_0504 is considered as a bidirectional promoter, which may regulate both ROK-like SACTE_0503 and LacI-like SACTE_0504-coding genes. Although we were able to show autoregulation of SsManR (SACTE_0504) via Prom_0504, a function of SACTE_0503 is still unknown (Fig. S2.4C). At Prom_5235, three genes, α-1,2-mannosidase (GH92: SACTE_5235), the domain of the unknown function (DUF: SACTE_5324), and peptidase (SACTE_5233) (106) are in the same orientation, and the putative mannose transporter genes (SACTE_5236, SACTE_5237, and SACTE_5238) are in the opposite direction, offering the potential that these will be regulated together (Fig. S2.4B). In the proposed network model, a predicted peptidase (SACTE_5233) is drawn with a potential function, which recognizes and cleaves the signal peptide of extracellular enzymes, i.e., SACTE_5235 α-1,2-mannosidase (Fig. 2.5). Although the third α-1,2-mannosidase (GH92: SACTE_6282) present in the SirexAA-E was not detected in our secretomic analysis, this gene is also thought to be regulated by SsManR through Prom_6282 based on the binding study (Fig. S2.4D). Overall, the described set of enzymes except for SACTE_6282 are produced under the presence of mannose via derepression of genes under the control of a newly identified transcriptional regulator, SsManR. Once mannobiose becomes available in the medium, an additional set of CAZymes under the SsCebR repression is concomitantly expressed. Importantly, the previously characterized β-mannanase (GH5: SACTE_2347) is also regulated under the SsCebR, and this enzyme actively hydrolyzes the mannan interchain (89).

In summary, we determined the novel response of SirexAA-E to the mannan, which was not known. A novel transcriptional regulator, SsManR, was determined to be the repressor that regulates a set of mannosidases and other mannose utilizing genes. Furthermore, we determined that mannobiose is a new effector ligand of the previously described cellooligosaccharide responsive regulator, SsCebR. The described mannose and mannobiose responses of SirexAA-E help us rationalize how this bacterium

responds to the naturally occurring galactomannan in hardwood. This linked regulatory system allows SirexAA-E to effectively respond to the presence of both cellulose and mannan, the majority polysaccharides found in pine wood, through detection of the major disaccharide products derived from their enzymatic hydrolysis.

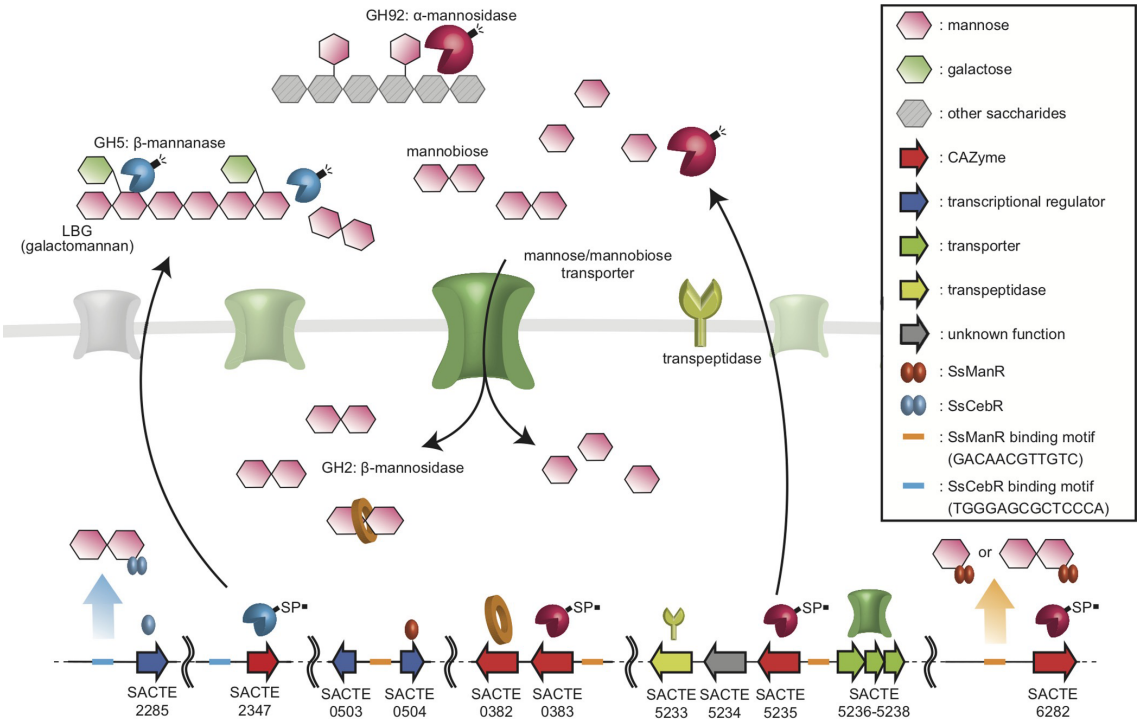


Figure 2.5. The transcriptional network circuit of SirexAA-E in response to mannose-containing polysaccharides via two LacI-repressors, SsManR and SsCebR, described in this study. The red, blue, green, and grey arrows in the bottom indicated genes encoding CAZymes, transcriptional regulators, transporters, and currently unknown function, respectively. Light blue and orange boxes indicate the SsCebR and SsManR binding motifs, respectively. Pink, green and gray hexagons are mannose, galactose, and other saccharides.

CHAPTER 3. MOLECULAR MECHANISMS OF MULTIPLE HEMICELLULOSE-RESPONSIVE TRANSCRIPTIONAL REGULATORS IN THE CELLULOLYTIC *STREPTOMYCES* SP. SIREXAA-E

Abstract

We have focused on the transcriptional networks of the insect-symbiont s*treptomyces* sp. SirexAA-E (SirexAA-E) to understand the regulation of genes that encode for plant biomass-degrading enzymes for potential biofuels applications. Unlike other actinobacteria, well known antibiotic producers, i.e. *S. griseus*, the transcriptional regulations of SirexAA-E to environmental nutrients are poorly described. We recently showed that SirexAA-E utilizes SsCebR and SsManR to regulate a set of genes encoding cellulose- and mannan-degrading enzymes in response to available carbon sources. However, a little is known about how SirexAA-E might regulate genes encoding enzymes that degrade xylan, the second abundant polysaccharide in the plant cell wall.

In this study, we cultivated SirexAA-E with xylose-, xylobiose- and xylan-containing media, and proteomic analysis was performed on the culture supernatants. In xylobiose and xylan culture supernatants, three xylanases and one xylose isomerase were determined to be highly secreted, compared to the xylose culture. To test if those four genes (two are in the same operon) are regulated by the same transcription regulators, pull-down proteomic analysis was conducted. Three biotinylated DNA sequences, upstream from the four genes, were incubated with the cell lysate of SirexAA-E, followed by proteomic analysis. Results showed that four predicted transcriptional regulators were pull-downed with all tested biotinylated DNA sequences. These proteins were overexpressed in *E. coli* then purified, and each protein was confirmed to bind above biotinylated sequences by electrophoresis mobility shift assay. Moreover, DNase I footprinting was carried out to determine the sequence motifs for each transcriptional regulator. In conclusion, we were able to identify a set of transcriptional regulators that regulates genes involved in xylan degradation and utilization of SirexAA-E.

Introduction

Streptomyces sp. strain SirexAA-E (SirexAA-E), a symbiont of the wood-devastating wood wasp, is a potentially useful strain for biofuels application with the reported high plant biomass-degrading capability. A series of biochemical analyses demonstrated that the specific activities of cellulose degradation in the culture supernatant of SirexAA-E, when cellulose was used as a sole carbon source in the growth medium, showed around 40% of the specific activity of the commercial enzyme cocktail, Spezyme CP (71). Importantly, when the SirexAA-E culture supernatant was prepared from the xylan grown condition, the specific activity of xylan degradation was shown to be ~20% higher than the Spezyme CP. Moreover, the xylan-specific induction of genes encoding three xylanases and one xylose isomerase was determined by the proteomic analysis. Thusfar, overproduction of xylan-degrading enzymes by this bacterium is of great interest, and to do so, the transcriptional network that regulates gene expression for a set of xylanases is core to understand for forthcoming genome-editing.

Xylan, the largest hemicellulose of plant cell wall materials, has the main chain of β-1,4-linked xylose and contains a variety of substituted side groups such as acetyl, L-arabinosyl, and 4-*O*-methylglucuronyl residues at the position of *O*-2 and/or *O*-3 of D-xylopyranosyl residues (15, 95). It represents an important renewable resource of bio-based products such as ethanol, lactic acid, and xylitol for a wide range of industrial applications (107–109). Xylan polysaccharides in nature are mainly present in the forms of acetyl xylan and arabinoxylan in hardwoods and softwoods (110). Considering the heterogeneous sugar composition and sugar linkages, complete and rapid hydrolysis of xylan polysaccharides require various types of enzymes and they are classified in the Carbohydrate-Active Enzymes (CAZymes) (31, 111, 112). For example, glycoside hydrolase family 10 and 11 (GH10 and 11) endo-1,4-xylanases (EC 3.2.1.8) break down the backbone of xylan. GH3 β-xylosidases (EC 3.2.1.37) release xylose monomer from non-reducing ends of xylooligosaccharides. GH62 α-L-arabinofuranosidases (EC 3.2.1.55) and GH115 α-glucuronidases (EC 3.2.1.131) remove the side chain sugars that are branched at various points of xylose molecules on xylan polymers. Furthermore, carbohydrate esterase family 4 (CE4) acetyl xylan esterases (EC 3.1.1.72) catalyze the hydrolysis of *O*-linked acetyl groups from the *O*-acetyl xylose residues. It is known that the synergistic action among the above enzymes, at least, is needed to realize efficient xylan hydrolysis (41, 43, 113).

A vast majority of reported xylan-degrading enzymes are originally found in microbes such as bacteria (*Streptomyces* and *Bacillus*) and fungi (*Aspergillus* and *Trichoderma*) (114). Among such xylolytic microorganisms, *Streptomyces* species, gram-positive bacteria, produce various xylan-degrading enzymes, most of which show high thermostability and alkaline tolerance, **compared to other microbe originated ones** (115, 116). Hence *Streptomyces* **species** can be important source**s** of xylan-degrading enzymes. With biochemistry and structural studies of xylan-degrading enzymes from *Streptomyces*, great details of the functions in xylan hydrolysis have been described (117, 118)**.** However, knowledge of the molecular mechanisms by which the *Streptomyces* respond to the substrate and adapt their enzyme production to xylan metabolism are elusive to date, whilst it will provide us the idea to create genome-edited strain for xylan utilization, especially based on the reported high cellulolytic SirexAA-E.

The LacI family regulator of *S. thermoviolaceus* OPC-520, BxlR, was reported to act as a repressor of xylan degradation and uptake system of xylo-oligosaccharides (92). This regulator was shown to control the expression of the five characterized xylan-degrading enzymes including two GH10 and GH11 endo-1,4-xylanases, one GH3 β-xylosidase, GH62 α-L-arabinofuranosidase, and CE4 acetyl xylan esterase in response to available xylobiose *in vivo* and *in vitro* (119–122). BxlR binds to the four inverted repeat, 5'-CGAA-N_3-TTCG-3', in the promoter regions of genes encoding the xylo-oligosaccharide transporters and the above five xylan-degrading enzymes. Thus, analogous gene regulation to the *S. thermoviolaceus* OPC-520 was thought to occur in SirexAA-E, which transcriptional circuits for cellulose- and mannan-degrading enzymes were previously described (71, 73, 123).

In this study, we sought to determine the already known and potentially new xylan responsive transcriptional regulators in SirexAA-E. Three xylan responsive transcriptional regulators, SACTE_0535p, SACTE_5479p, and SACTE_5759p, were determined and characterized with the corresponding new effector ligands, which likely regulate the downstream genes in a xylooligosaccharide-responsive manner. Furthermore, novel DNA sequence motifs for SACTE_0535p and SACTE_5479p were determined by DNase I footprinting assay. In conclusion, we identified a set of transcriptional regulators that regulates genes involved in xylan degradation and metabolism of SirexAA-E.

Materials and Methods

General reagents

Xylan (Tokyo Chemical Industry, Tokyo, Japan) and D-glucose (Wako Pure Chemical Industry, Ltd., Osaka, Japan) were purchased for the defined carbon source-containing culture media. Carbohydrates, including D-xylose, D-cellobiose (Wako Pure Chemical Industry, Ltd., Osaka, Japan), D-xylobiose (Megazyme, Ireland), allantoin (Tokyo Chemical Industry, Tokyo, Japan), allantoic acid (Tokyo Chemical Industry, Tokyo, Japan), and glyoxylate (Tokyo Chemical Industry, Tokyo, Japan), used in EMSA experiments were all purchased.

Growth on different carbon sources

Yeast-malt extract (YME) (4 g of yeast extract, 10 g of malt extract, and 4 g of glucose for 1L) and YME agar were used for routine cultivation of SirexAA-E as described in the previous study (123).To prepare cell lysate, the cells were pre-cultured in 2.0 mL M63 (10.72 g of K_2HPO_4, 5.24 g of KH_2PO_4, 2 g of $(NH_4)_2SO_4$, 1 mM MgSO4, and 10 g of Thiamin for 1L) medium containing 0.5% (w/v) of glucose for 3 days at 30 ˚C with shaking at 250 rpm. 10% (v/v) of pre-cultured cells were inoculated into 50.0 mL of M63 minimal medium containing 0.5% of either glucose or xylan as sole carbon source, and cultivated for 3 days at 30˚C with shaking at 250 rpm. All cultures were performed in triplicate.

Preparation of the intracellular proteins

The cell pellets were collected by centrifugation for 15 min at 4,000 *x g* at 4˚C, and washed with 20 mM phosphate buffer (pH 7.4) twice. The cell fraction was resuspended in 500 µL of 20 mM phosphate buffer (pH 7.4) and sonicated (Branson, Danbury, CT) with output control at 3, 30% of a duty cycle for 15 min on ice. Cell lysates were centrifuged at 10,000 x *g* for 30 min at 4°C, and the supernatants were collected as the fraction containing intracellular proteins. To remove soluble sugars, intracellular proteins were dialyzed against 50 mM NaCl, 20 mM phosphate buffer (pH7.4) for 1 hr using Ultracel® 10 kDa Ultrafiltration Discs (Millipore, Ireland). The concentration of proteins in each intracellular protein was determined by the BCA protein assay (Bio-Rad Laboratories, Hercules, CA, USA).

Preparation of DNA fragments for pulldown assay

350-550 bp of biotinylated DNA fragments containing three promoter regions were amplified using 5'-biotinylated forward primers (Eurofins genomics, Tokyo, Japan) and non-labeled reverse primer pairs (Fasmac, Kanagawa, Japan) listed in Table 3.1. These primer pairs were used to PCR amplify three promotor regions using the SirexAA-E genomic DNA as a template and KOD FX NEO PCR kit (TOYOBO, Osaka, Japan). To prepare competitor DNA fragments for pulldown assay, the SirexAA-E genomic DNA was physically fragmented by the sonication under the following setting, output control at 3 and 30% of a duty cycle for 30 min on ice, which resulted in an average size of 500 bp sheared DNA fragment.

Pulldown assay

Biotin-based pulldown proteomics analysis was carried out as described in the previous study with some modifications (124). 50 ug of the intracellular proteins extracted from either glucose- or xylan-grown cells were mixed with 250 ng of a biotinylated DNA fragment in 500 µL binding buffer, containing 200 mM NaCl, 0.1% Tween20, 1 mM EDTA, and 20 mM Tris-HCl (pH 7.4). 50 ug of BSA and 2.5 mg of SirexAA-E genome fragments were also added to the reaction mixture to avoid any non-specific binding of proteins to the biotinylated DNA. The mixtures were incubated for 1 hr at 15˚C with moderate rotation, followed by adding 50 ug of streptoavidin magnetic beads (Magnosphere^{TM}, JSR Life Sciences Co, ibaraki, Japan), and incubated for 15 min at 15˚C to allow biotinylated DNA and streptavidin beads interaction. The streptoavidin magnetic beads were then washed 10 times with 200 µL of the binding buffer, and the intracellular proteins that remained bound during extensive wash were suspended in 10 µL of 8 M urea.

Preparation of proteomic samples

Solubilized samples in 8M urea were diluted to 1 M urea by adding 25 mM ammonium bicarbonate buffer followed by a reduction reaction with 50 mM dithiothreitol for 30 min at 50˚C, and alkylation reaction with 500 mM iodoacetamide for 30 min in the dark at room temperature. Tryptic digestion was carried out for 10 hrs at 37˚C by treating with 0.5 ug of proteomic grade trypsin (Roche, Mannheim, Germany), and peptides were purified using C18 ziptip pipette tips (Millipore, Ireland).

Proteome analyses

Mass spectra were separated using an Easy nLC1000 liquid chromatography system with Q-Exactive plus Orbitrap mass spectrometer (ThermoFisher Scientific, City, IL, USA), and Xcalibur software (ver. 3.1, Thermo Fisher Scientific, IL). The trypsin digested peptides were separated on a C18 capillary tip column (NTCC-360/75-3-125, Nikkyo Techno, Japan) by a linear gradient from 5 to 30 % with two solutions: 0.1% formic acid in water and 0.1% formic acid in acetonitrile in 2 hrs. Full scan mass spectra were obtained in the Orbitrap with a scan range of 300.0 to 2,000.0 *m*/*z* with a resolution of 70,000. Proteins were identified from the acquired MS/MS spectra using Proteome Discoverer 2.1 (Thermo Fisher Scientific, IL) with all CDSs of *S.* sp. SirexAA-E reported in the previous study (71). The peptide mass tolerance was set at 10 ppm, and fragment mass tolerance was set at 0.8 Dalton, respectively. The peptide charge was set at +2, +3 and +4. The accuracy and sensitivity of peptide identification were optimized using the automatic decoy and percolator functions under Proteome Discoverer software built-in function. All proteomic data sets obtained in this study can be found in the PRIDE proteomics identification database (https://www.ebi.ac.uk/pride/) with the accession number of xxxxxx.

Cloning and protein expression of transcriptional regulators

All plasmids used in this study were cloned using a sequence and ligation–independent cloning (SLIC) method (125). All primers were purchased (Fasmac, Kanagawa) and listed in Table 3.1. The SACTE_0535, SACTE_5479, and SACTE_5759 genes were cloned into the *Escherichia coli* expression vector pVP68K (Center for Eukaryotic Structural Genomics, Madison, WI), which is designed to produce the following construct, an N-terminal 8-histidine tag, maltose binding protein (MBP), tobacco etch virus (TEV) protease cleavable sequence, followed by the protein of interest. The SACTE_0535, SACTE_5479, and SACTE_5759 genes were PCR-amplified using the SirexAA-E genomic DNA as a template and the KOD FX NEO polymerase (TOYOBO, Osaka). The amplified insert DNA is designed to contain 15 bp overlapping sequence in both ends, which are complement with the pVP68K vector sequence, necessary for the SLIC method (125). The backbone of the pVP68K vector was amplified using a pair of primers listed in Table xx. The standard SLIC reaction contains 50 ng of a vector backbone PCR product and corresponding insert DNA with the molar ratio of 1:05 to 1:4, 0.25 (3 units/μL) of T4 DNA polymerase (NEB, Ipswich, Massachusetts, USA) and

1 µL of NEBuffer r2.1 in 10 µL reaction. The SLIC reaction mixture was incubated for 5 min at 26˚C, then chilled in ice for 10 min. The ligated plasmids were transformed into *E. coli* JM109 competent cells (Takara Bio Inc., Shiga, Japan) by a standard heat-shock method, and plated onto an LB medium supplemented with 50 ng/mL kanamycin for antibiotic selection. Transformants were screened by a colony PCR using the primer pairs, colony PCR_fw and colony PCR_rv, then the sequence was verified (Eurofins genomics, Tokyo). Sequence-verified plasmids were transformed into *E. coli* strain Rosseta2 (DE3) (Merck Millipore, Burlington, Massachusetts, USA).

Protein overexpression and purification were carried out accordingly to the previous study (126). Briefly, cells transformed with the plasmid were precultured in 20 mL of non-inducing MDAC media (127) supplemented with 50 µg/mL of kanamycin overnight at 37 °C with shaking at 250 rpm. The precultures were added to 1 L of TB+G autoinduction media (12 g of tryptone, 24 g of yeast extract, 6.4 g of glycerol, 0.12 g of glucose, 4 g of alpha-lactose, 3.75 g of aspartate, 12.54 g of K_2HPO_4, 2.31 g of KH_2PO_4, 0.24 g of $MgSO_4$ for 1L) supplemented with 50 µg/mL kanamycin and cultivated overnight at 30°C with shaking at 250 rpm. The cell pellets were harvested by centrifugation at 4,000 x *g* for 15 min at 4˚C, and washed twice with ice-cold 20 mM phosphate buffer (pH 7.4). Cell pellets were resuspended in 20 mL of 20 mM phosphate buffer (pH 7.4) containing 20 µL of 100 mg/mL of lysozyme, and sonicated by the following settings, output control at 3 and 30% of a duty cycle for 15 min on ice. Cell lysates were centrifuged at 10,000 x *g* for 60 min at 4°C, and supernatants were filtered through a 0.45-µm filter (AS ONE Co., Osaka, Japan). The Ni-affinity chromatography was performed by using an AKTA prime protein purification system (GE Healthcare, Freiburg, Germany) equipped with a 1 mL HisTrap-FF column (GE Healthcare, Freiburg) and a linear 50 mL gradient from equilibration buffer (20 mM phosphate buffer, pH 7.4, containing 500 mM NaCl, 25 mM imidazole) to equilibration buffer supplemented with 0.5 M imidazole. The purified fractions were concentrated and buffer-exchanged against XX mM NaCl, 20 mM phosphate buffer, pH 7.4, using a VIVASPIN 20 concentrator (Sartorius Stedim, Goettingen, Germany Sartorius AG, Germany) at 4,000 x *g* for 2 h at 4°C. The N-terminal His-tag and MBP were removed by adding 1/100 (w/w) TEV protease for 16 hrs at 4˚C and the protein product lacking MBP protein was purified by collecting the flow-through of Ni-column. The tag-free protein was concentrated and buffer-exchanged against 20 mM phosphate buffer, pH

7.4, using a VIVASPIN 20 concentrator. The purified protein concentration was quantified by BCA assay and stored at -80°C until use.

Electrophoresis mobility shift assay (EMSA) and effector assay

The six DNA fragments containing the predicted promoter regions were amplified from the SirexAA-E genomic DNA using primers shown in Table xx and KOD FX NEO polymerase (TOYOBO, Osaka). The DNA concentration was estimated by the DS-11 nano liter spectrophotometer (DeNovix Inc., Wilmington, Delaware, USA) and adjusted to 10 μg/uL. The EMSA reactions were carried out according to the previous study with modifications (123). 10 ng of each DNA fragment was mixed with varying the amount of the purified protein in the binding buffer consisting with 100 mM Tris-HCl, pH 7.4, containing 5 mM EDTA, 50 mM $(NH_4)_2SO_4$, 5 mM DTT, 150 mM KCl, 1.0% (v/v) Tween-20, and 5.0% (v/v) glycerol), and incubated for 15 min at 25°C. For the effector assay, different amounts of a series of effector ligands were added to the reaction mixtures, in which DNA-protein complex formation was determined, and incubated for 10 min at 25°C. The samples were then loaded onto 2% TAE agarose gels, and the free and bound forms of DNA fragments were separated by electrophoresis at 50 V for 3 hrs at 5˚C. The gels were stained with ethidium bromide, and the amount of free and unbound fractions on the gel was quantified by using a Gel Doc™ EZ imager system (Bio-Rad Laboratories, Hercules, CA, USA). 0.5 μg/uL BSA was used as a negative control in each assay, and all reactions were performed in triplicate. The apparent K_D values of SACTE_0535p, SACTE_5479p and SACTE_5759p for the indicated DNA fragments were estimated based on the DNA concentration used in the reaction, which gives approximately 50% of free DNA to the DNA-protein complex by EMSA. The formula used for calculation of the apparent K_D value is shown below (87):

$$[complex] = [protein]*[free\ DNA]/(K_D + [free\ DNA]),$$

where [complex], [protein], and [free DNA] indicated the concentration of the DNA-protein complex, protein, and free DNA, respectively.

DNA footprinting assay.

DNase I footprinting was performed as described elsewhere (128). The same binding condition used in the EMSA assay was carried out to obtain the DNA-protein complex. ~250 bp of DNAs were PCR-amplified by using the following 5'-biotinylated forward primers (SACTE_0535_fp_fw and SACTE_5479_fp_fw) and corresponding non-biotinylated reverse primers (SACTE_0535_fp_rv and SACTE_5479_fp_rv), shown in Table XX. The formation of the biotinylated DNA-protein complex was performed using 5 ng biotinylated DNA with 0.025/ 0.05 µg/µL SACTE_0535p or 0.028 µg/µL SACTE_5479p in the 15 µL reaction for 15 min at room temperature, followed by 3 units/µL DNaseI treatment for 60 min at 15°C. The sample was ethanol precipitated by adding 50 µL 100% ethanol and 1.5 µL 5 M NaCl, then suspended in the 3 µL loading dye (1 ug of bromophenol blue in 95% of formamide and 5% of glycerol for 1 mL). The nucleotide standard for the footprinting was prepared using the Thermo Sequenase™ Dye Primer Manual Cycle Sequencing Kit (Thermo Fisher Scientific, IL, USA) accordingly to the manufacture's instruction. Right before running on the sequencing gel, samples were heat-denatured for 5 min at 95°C, and run on the 6% (19:1) polyacrylamide gel containing 7 M urea, 1xTBE, and 1% formaldehyde, for 25 min at 300 V at room temperature following pre-running for 30 min at 300 V at room temperature. The DNAs on the gel was transfered to the Zeta-Probe® Blotting Membranes (Bio-Rad Laboratories, Hercules, CA, USA), accordingly to the previous study (129) . The transfered DNAs were UV-crosslinked to the membrane using a UVP Crosslinker CL-3000 (Analytik Jena, Jena, Germany), which takes approximately 2.5 min under default setting (1000 mJ/cm^2). To detect the DNA, the DNA-hybridized membrane was washed for 5 min with blocking solution (5.0 g SDS, 0.73 g NaCl, 0.24 g Na_2HPO_4, and 0.11 g $NaH_2PO_4.H_2O$ for 100 mL), then the membrane was soaked for 5 min in the 20 mL blocking solution containing 1 mg/mL of streptavidin, and washed twice with 20 mL of 0.1xblocking solution. The membrane was soaked for 10 min in the 20 mL of blocking solution containing 0.5 mg/mL of biotinylated alkaline phosphatase reagent. Subsequently, the membrane was wash twice for 5 min with 25 mL Triton X-100 buffer (5.0 mL 1 M Tris-HCl, pH 8.0, 5.0 mL 20% Triton X-100, 0.20 mL 0.5 M Na_2EDTA, 0.73 g NaCl, for 100 mL), then wash twice for 5 min with 25 mL pH 9.5 buffer (0.88 g NaCl, 1.87 g Tris-base, 1.0 mL 1 M $MgCl_2$, adjust pH to 9.5 using HCl, for 200 mL). Finally, the membrane was placed on a clean glass plate, with the DNA-side facing up. 5 mL of 100 ng/ mL NBT and 5 mL of 75 mg/mL BCIP were added to

1.0 mL pH 9.5 buffer and quickly mixed by vortexing. The NBT/BCIP solution was poured on the top of the membrane, and the membrane was immediately covered with a clean glass plate. The membrane was reacted in the dark overnight to develop the signals, then washed briefly with Triton X-100 solution for 5 min, and rinsed with distilled water.

Table 3.1. Primers used in this study.

Name	Sequcence (5'-3')
pVP68K_vector for _fw[a]	GGAGGTGGTCAGCCAGAGCCC
pVP68K_vector for_ rv[a]	GTTCTACGCGGCGCTCGAGGC
SACTE0535_fw[a]	GTGGTGGCGGCCTTGGTCACCGG
SACTE0535_rv[a]	GGGCCGCACAGCACGATCGTTC
SACTE5479_fw[a]	CATGGTCCCGGCCTCCCCGCAG
SACTE5479_rv[a]	AACTTGGCGTTGCCCAGTCCGG
SACTE5759_fw[a]	GTCGGACCGTTTCCGCGCCAAC
SACTE5759_rv[a]	GTTCAACACCCAGGCGGCACCC
colony PCR_fw[a]	TGGTAGGCCAGGTTCAGTGCGAGC
colony PCR_rv[a]	ACGTCCTGGACACCAGCATCGGC
Prom_SACTE0265_fw[c]	CGCCGCACGTCACGGTCGTTG
Prom_SACTE0265_rv[c]	CGGTCTGTGGACCGTGGGCTGG
Prom_SACTE0357_fw[c]	CAGGTCGTCGACATGTGGCAGGC
Prom_SACTE0357_rv[c]	AGTGCTCGCCGGGGGTGACGG
Prom_SACTE5230_fw[c]	CCGCGTCGTCAAGATCGTGCGG
Prom_SACTE5230_rv[c]	TTGCCGTCGGAGGTGGCGATGCG
Prom_SACTE0535_fw[c]	CGCGTCCTCGAAGTCGGCGAGCC
Prom_SACTE0535_rv[c]	GAGGTCCTGACCGATGCGGCCC
Prom_SACTE5479_fw[c]	CGTCGAGCAGGGCCATGTTCGCC
Prom_SACTE5479_rv[c]	TGCAGTCGCTCGAGCGCGCCTTC
Prom_SACTE5759_fw[c]	GTCGTCACCCCGGACGGGACC
Prom_SACTE5759_rv[c]	CCGCTTGGTGCGCAGGTGGCC

[a]Primers used for cloning plasmids.

[b]Primers used for clone verification.

[c]Primers used for amprification of DNA fragments.

Result

Pulldown assay for xylan-responsive transcriptional regulators

In order to search for xylan-specific responsive transcriptional regulators of SirexAA-E, we first performed pulldown experiments with three promoter regions of xylan-utilizing enzyme genes. These DNA fragments, a 500 bp region (Prom_0265) located upstream from SACTE_0265 gene (Fig. S9A), a 550 bp region (Prom_0357) located upstream from SACTE_0357 and SACTE_0358 genes (Fig. S9B), and a 350 bp region (Prom_5230) located between SACTE_5230 and SACTE_5231 genes (Fig. S9C), were PCR-amplified using 5'-biotinylated primers. Proteins bound to each DNA fragment were pulled down from the intracellular proteome in the either of xylan- or glucose-grown cell lysate. According to the result of semi-quantitative proteomic analysis, a total of 77 proteins of SirexAA-E that predicted to be transcriptional regulators were identified in the elute fraction of at least one sample (Table. 3.2). To focus on xylan-responsive regulators, we set up two criteria, 1) transcriptional regulators that were pulled down from xylan-grown cell lysate using all three DNA fragments, 2) transcriptional regulators that were not found in the sample using glucose-grown cell lysate using any DNA fragments. Among them, 3 transcriptional regulators including SACTE_0535, SACTE_5479, and SACTE_5759 met this criterion (Table 3.2.). Therefore, we characterized these three proteins by biochemical and genomic approaches for further experiments.

Table 3.2. LC-MS/MS identification of transcriptional regulators. Only proteins that are predicted as transcriptional regulators are shown. Protein abundance was estimated based on spectral count. Out of 83 detected regulators, 3 regulators, SACTE_0535, SACTE_5479, and SACTE_5759 were bound to all of the three DNA fragments. In order to prove the relevance of the results from the pulldown assay, they were chosen for further studies to assess their DNA binding capabilities.

		glucose intracellular protein			xylan intracellular protein		
Locus	Description	P0265	P0357	P5230	P0265	P0357	P5230
SACTE_0535	LacI-like transcriptional regulator				12	12	8
SACTE_5479	IclR-like transcriptional regulator				3	3	3
SACTE_5759	LacI-like transcriptional regulator				6	5	2
SACTE_1266	transcriptional regulator	1	2	3			
SACTE_3650	MarR-like transcriptional regulator	2	3	1			
SACTE_1606	GntR-like transcriptional regulator	3	1	2	4	4	3
SACTE_3019	CrpA-like transcriptional regulator	34	27	31	62	56	51

Sequence analysis of three predicted transcriptional regulators

To predict the function of three putative xylan-responsive transcriptional regulators, we searched for each homolog based on their amino acid sequence similarities using BLAST. As shown in Table 3.3., three homologs of other *Actinomyces* with 32%-70% of amino acid sequence similarities were found. As to the homolog of SACTE_0535, BxlR of *S. thermoviolaceus* OPC-520, it bound its binding motif, 5'-CGAA-N3-TTCG-3' and repressed the transcription of several xylan-degrading enzymes. These xylanases included homologs of SACTE_0265, SACTE_0357, and SACTE_0358 that were induced by xylan in the SirexAA-E (71, 73). This would support the result of the pulldown-based proteomic analysis. Moreover, xylobiose, xylo-disaccharide, was reported to inhibit the interaction between the BxlR and DNA. On the other hands, the homolog of SACTE_5479, AllR of *S. coelicolor*, were reported to act as a repressor of allantoin degradation pathway depending on the presence/ absence of allantoate and glyoxylate. The third one, SACTE_5759, RbsR of *C. glumamicum*, were reported to act as a repressor of allantoin degradation pathway depending on the presence/ absence of allantoate and glyoxylate.

Table 3.3. A summary of three homologs of the *Actinomyces*.

Three proteins, BxlR, AllR, and RbsR, were characterized in these previous studies (92, 130, 131).

Characterized proteins	Identities (homolog)	Repressed proteins	Effector ligands	Binding motifs
BxlR of *S. thermoviolaceus* OPC-520[3]	70% (SACTE_0535)	xylanases	xylobiose	CGAA-N_3-TTCG
AllR of *S. coelicolor*[4]	89% (SACTE_5479)	allantoin degrading enzymes	allantoate glyoxylate	TTCCNC-N_3-GNGGAA
RbsR of *C. glutamicum*[5]	32% (SACTE_5759)	ribose transporters ribose kinase	ribose-5P	A-N_3-AANCGNTT-N_3-T

Functional characterization of three xylan-responsive regulators

To test whether three putative regulator proteins, SACTE_0535p, SACTE_5479p, and SACTE_5759p, bind to the promoter regions used in the pulldown assay, each gene was amplified from the genome of SirexAA-E, then cloned into pVP68K vector, overexpressed in *Escherichia coli,* and purified by affinity tag methods (Table 2 and Fig. S5). Along with three DNA fragments used in the pulldown assay, we prepared the following three DNA fragments containing promoter region of genes encoding the transcriptional regulators, since we suspected their autoregulation. These DNA sequences including a 200 bp region (Prom_0535) located between SACTE_0534 and SACTE_0535 genes (Fig. S9D); a 350 bp region (Prom_5479) located between SACTE_5479 and SACTE_5480 (Fig. S3.4E); a 500 bp region (Prom_5759) located between SACTE_5758 and SACTE_5759(Fig. S3.5E);, were amplified from genome DNA. The result of electromobility shift assay (EMSA) of SACTE_0535p to six promoter-containing regions is shown in Fig. 3.1. SACTE_0535p was found to bind neither Prom_5479 nor Prom_5759 even at a large amount of the protein. In contrast, the mobility shift of other four fragments were seen when the amount of SACTE_0535p is increased, and apparent K_D was estimated to be ~78 nM (Prom_0265), ~116 nM (Prom_0357*)*, ~143 nM (Prom_5230), and ~ 88 nM (Prom_0535), respectively. The previously determined consensus sequence of the homolog of SACTE_0535p, BxlR in *S. thermoviolaceus* OPC-520, 5'-CGTT-N_3-AACG-3', were observed in the Prom_0265, Prom_0357, and Prom_0535, but not in the Prom_5230. Thus, we attempted to determined currently unknown SACTE_0535p-binding region on the Prom_5230 using DNaseI footprinting assay. The 30 bp sequence, 5'-TGATGTTTCGGTCACATGGCGCAAACATGG-3', was determined as a novel binding region of SCATE_0535p. (Fig. 3.3.). Although we searched for the 30 bp sequence on the SirexAA-E genome, we didn't find another region. The result of electromobility shift assay (EMSA) of SACTE_5479p to six promoter-containing regions is shown in Fig. 3.1. SACTE_5479p was found to bind neither Prom_0535 nor Prom_5759 even at a large amount of the protein. In contrast, the mobility shift of other four fragments were seen when the amount of SACTE_5479p is increased, and apparent K_D was estimated to be ~17 nM (Prom_5479), and ~21 nM (Prom_0357*)*, respectively. The result of electromobility shift assay (EMSA) of SACTE_5759p to six promoter-containing regions is shown in Fig. 3.1. SACTE_5759p was found to bind only its promoter region, Prom_5759 (K_D was estimated to be ~89 nM) under the tested condition.

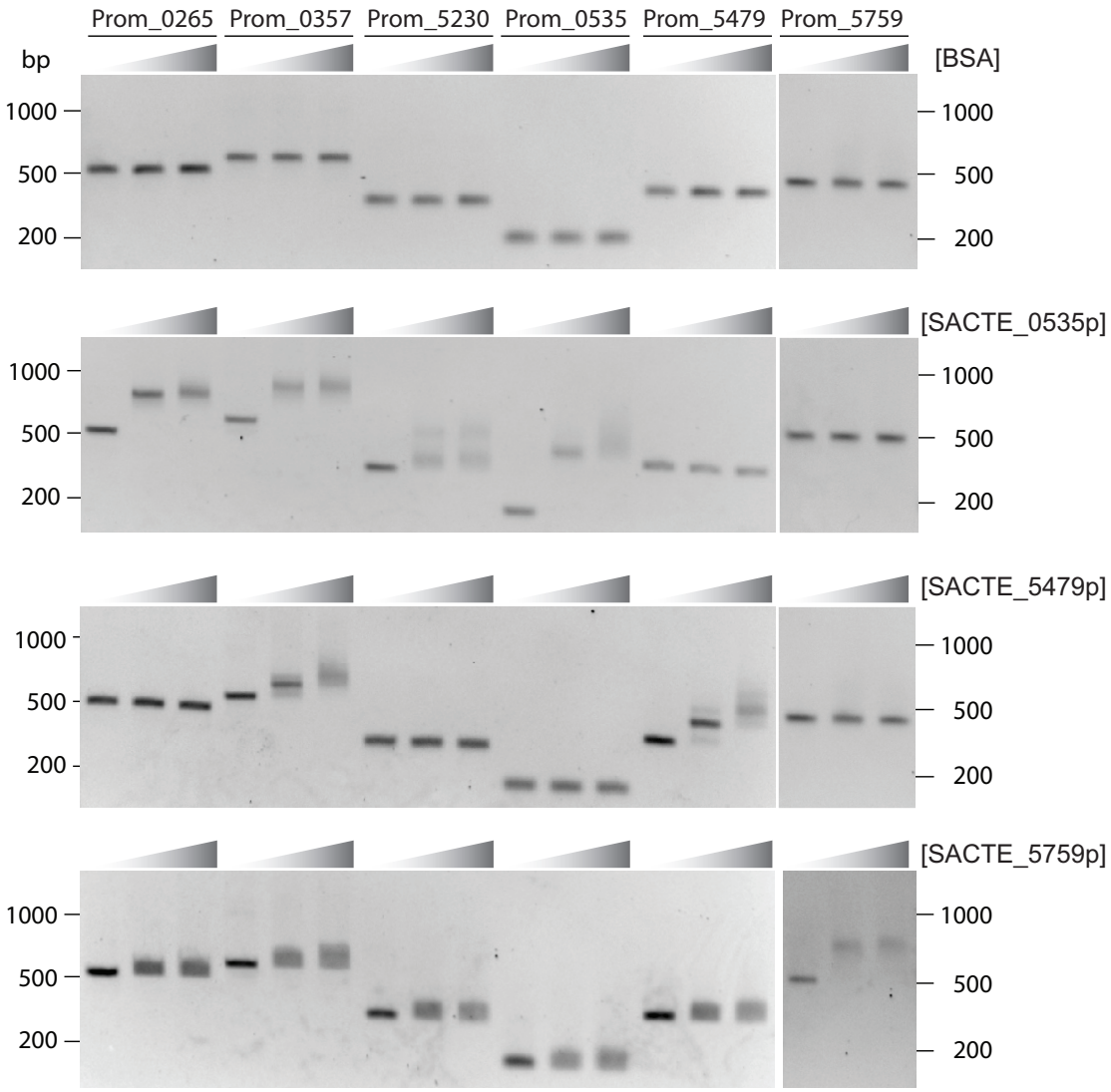


Figure 3.1. Electromobility shift assay of SACTE_0535p, 5479p and 5759p for the upstream region of xylan-induced genes and their upstream regions. The recombinant expressed and Ni-purified proteins were used for the *in vitro* binding assay. The following fragments, Prom_0535, Prom_5479, and Prom_5759 were tested to assess autoregulation. The black filled triangles indicate apparent DNA-protein complex formation. BSA used as a negative control in this assay and did not bind to all tested DNA fragments (data was not shown).

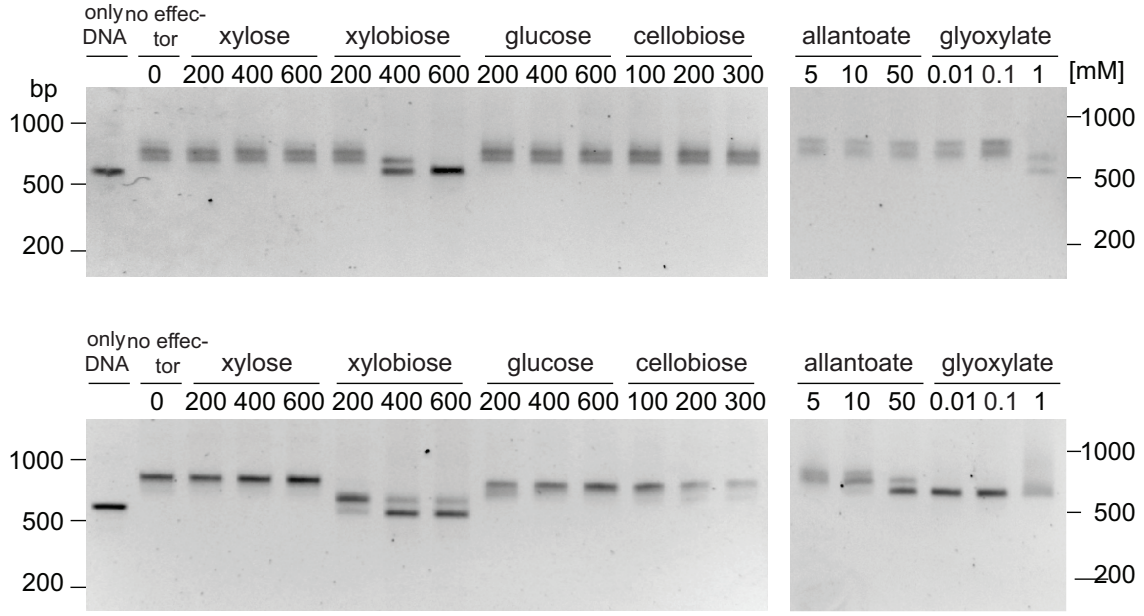


Figure 3.2. Effector assay of SACTE_0535p or only 5479p to the Prom_0357 by six possible effectors ligands. Since mono- and di-saccharides were known to associate with transcriptional regulators, especially from LacI family protein, we tested inhibitory effects for xylose, xylobiose, glucose and cellobiose, which are major end-products of xylan and cellulose degradation. Allantoate and glyoxylate, products of allantoin degradation, are the reported effectors of the AllR, homolog of SACTE_5479 in *S. coelicolor*. Xylobiose and glyoxylate acted as negative effectors of both SACTE_0535p and 5479p. Allantoate also inhibited the binding of SACTE_5479p to the DNA.

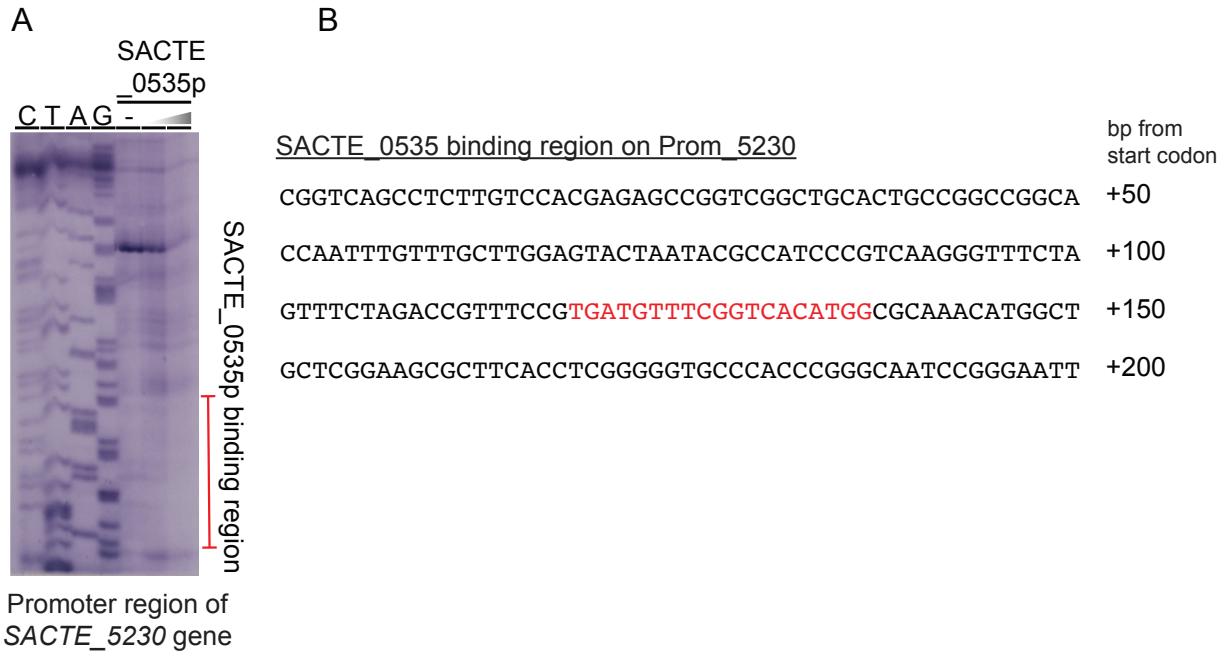


Figure 3.3. DNA footprinting assay of SACTE_0535p.

A. Gel images of DNA footprinting assay. B. DNA sequence of Prom_5230 with determined binding regions. Protein binding regions on two promoter regions are shown in red and blue-colored bars (A) and DNA sequences (B), respectively. Red boxes indicated previously reported binding sites of the SACTE_0535 homolog.

Discussion

Prior to this work, the response of SirexAA-E to xylan, one of the most abundant hemicelluloses in the plant cell wall, particularly coniferous woods had not been described (95). To identify transcription regulators that controlled three xylan-induce genes, pull-down proteomic analysis was conducted. Three predicted transcriptional regulators (SACTE_0535, SACTE_5479, and SACTE_5759) were detected. These proteins were overexpressed in *Escherichia coli* then purified, and two of them were confirmed to bind above biotinylated sequences by electrophoresis mobility shift assay. With these new findings, we can propose a transcriptional network model that enables SirexAA-E to respond to xylan polysaccharides (Fig. 3.4). Additionally, xylobiose would be produced by the decomposition of xylan by other symbionts in nature. When cells are exposed to xylobiose, the disaccharide are imported into the cell by a putative xylobiose transporter (SACTE_0532, SACTE_0533, and SACTE_0534) (92), and displaces SACTE_0535 from the 4 promoter regions (Prom_0265, Prom_0357, Prom_0535, and Prom_5230). This turns on the transcription of a set of genes. At Prom_0357, at least two xylan degradation genes would be transcribed, acethyl xylan esterase (CE4: SACTE_0357) and endo-1,4-xylanase (GH11: SACTE_0358), with both enzymes predicted to function in the xylan degradation, respectively (Fig. S9A). At Prom_0265, endo-1,4-xylanase (GH10: SACTE_0265) would be transcribed. (SACTE_0503 and SACTE_0504), with currently unknown function and autoregulation of its transcription, respectively (Fig. S9C). At Prom_0535, LacI family transcriptional regulator would be transcribed. (SACTE_0535), with of its transcription, respectively (Fig. S9C). At Prom_5230, at least two xylose metabolizing genes, xylose isomerase and α-D-xylohydrolase (SACTE_5230 and SACTE_5231), and the putative xylose transporter genes (SACTE_5227, SACTE_5228, and SACTE_5229) are in the same direction, offering the potential that these will be regulated together (Fig. S9B). For xylan degradation, xylan-degrading enzymes (encoded by SACTE_0265, 0357 and 0358 genes) convert xylan into xylooligosaccharides including xylobiose. They are transported into the cell through a putative xylooligosaccharide transporter (SACTE_0532-0534). Xylooligosacchrides can be converted into monomer and xylulose by β-xylosidase (SACTE_0531) and xylose isomerase (SACTE_5230), respectively, followed by phosphorylation and processed in a pentose phosphate pathway. On the other hands, the SACTE_5479 regulates two gene clusters that contain genes function in xylan and allantoin degradation. Allantoin are thought to

release from other organisms as an end-product of purine metabolism. The allantoinase (SACTE_5480) converts intracellular allantoin into allantoate. Allantoate are thought to be degraded into glyoxylate by two enzymes (EC 3.5.3.4 and EC 4.3.2.3). The former was predicted to be SACTE_2075 which is constitutively transcribed, according to the previous transcriptome analysis (71). In contrast, the latter was found to be allA in *E. coli*, but not present in the SirexAA-E genome. The transcription of two genes (SACTE_0357 and 0358) are also controlled by the SACTE_5479 in response to the xylobiose, allantoate and glyoxylate.

In summary, we determined LacI- and IclR-like transcriptional regulators, which were determined to be the repressor that regulates a set of xylan degradation and metabolizing genes. Furthermore, Moreover, DNase I footprinting was carried out to determine the sequence motifs for two transcriptional regulators. The described xylan and xylooligosaccharide responses of SirexAA-E, help us rationalize how this bacterium responds to the naturally occurring xylan polysaccharide in the plant cell walls. We identified key genes/proteins involved in the hemicellulose deconstruction of SirexAA-E. We proposed the model of for molecular mechanisms of SirexAA-E to response and utilize xylan, which could be helpful for future genetically engineering strategies.

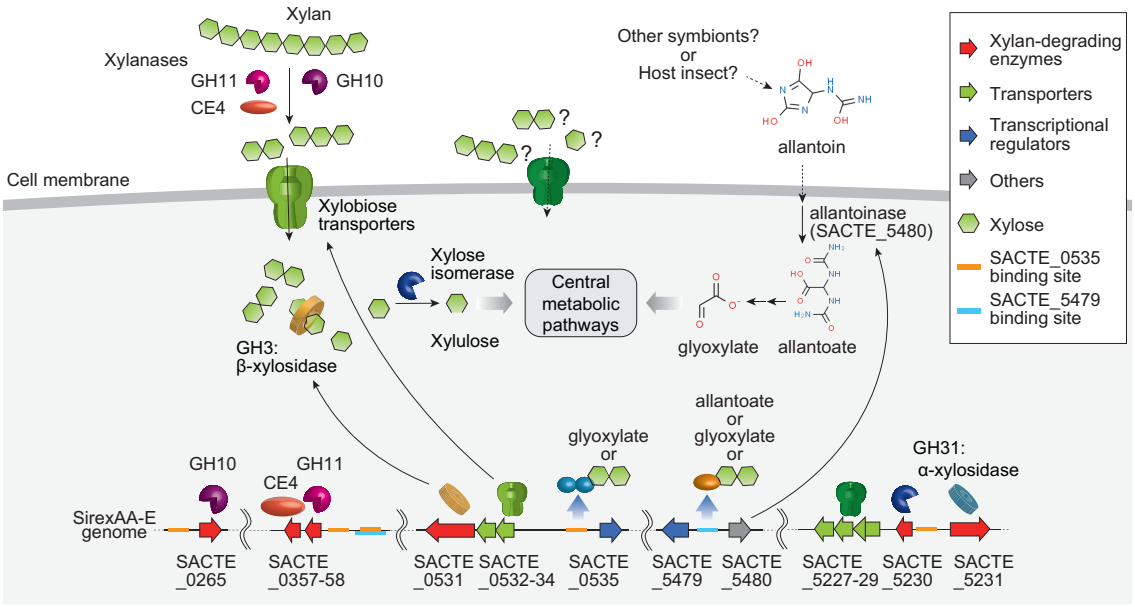


Figure 3.4. The transcriptional network circuit of SirexAA-E in response to xylan and allantoin via multiple repressors, SACTE_0535 and 5479, described in this study. The red, green, blue and gray arrows at the bottom indicate genes encoding xylan-degrading enzymes, transporters, transcriptional regulators and currently unknown function, respectively. Orange and light blue boxes indicate the SACTE_0535 and 5479 binding sites, respectively. Green hexagons are xlyose molecules. The SACTE_0535 regulates four gene clusters that contain genes function in xylan degradation/utilization.

LIST OF REFERENCES

1. Chapman RE. 1983. Chapter 10 Origin and Migration of Petroleum: Geological and Geochemical Aspects, p. 209–229. *In* .
2. Hollis D. Hedberg (2). 1964. Geologic Aspects of Origin of Petroleum. Am Assoc Pet Geol Bull 48.
3. Coady D, Parry I, Sears L, Shang B. 2017. How Large Are Global Fossil Fuel Subsidies? World Dev 91:11–27.
4. BARBIR F, VEZIROGLU T, PLASSJR H. 1990. Environmental damage due to fossil fuels use. Int J Hydrogen Energy 15:739–749.
5. Searchinger T, Heimlich R. 2015. Avoiding bioenergy competition for cood crops and land. World Resour Inst Work Pap 44.
6. Miret C, Chazara P, Montastruc L, Negny S, Domenech S. 2016. Design of bioethanol green supply chain: Comparison between first and second generation biomass concerning economic, environmental and social criteria. Comput Chem Eng 85:16–35.
7. Naik SN, Goud V V., Rout PK, Dalai AK. 2010. Production of first and second generation biofuels: A comprehensive review. Renew Sustain Energy Rev 14:578–597.
8. Thangavelu SK, Ahmed AS, Ani FN. 2016. Review on bioethanol as alternative fuel for spark ignition engines. Renew Sustain Energy Rev 56:820–835.
9. Kukoyi TO, Muzenda E, Mashamba A. 2015. Biomethane and Bioethanol as Alternative Transport Fuels 2012.
10. Karamanlioglu M, Preziosi R, Robson GD. 2017. Abiotic and biotic environmental degradation of the bioplastic polymer poly(lactic acid): A review. Polym Degrad Stab 137:122–130.
11. Karp A, Richter GM. 2011. Meeting the challenge of food and energy security. J Exp Bot 62:3263–3271.
12. Mumtaz T, Yahaya NA, Abd-Aziz S, Abdul Rahman N, Yee PL, Shirai Y, Hassan MA. 2010. Turning waste to wealth-biodegradable plastics polyhydroxyalkanoates from palm oil mill effluent – a Malaysian perspective. J Clean Prod 18:1393–1402.
13. Jesse TW, Ezeji TC, Qureshi N, Blaschek HP. 2002. Production of butanol from starch-based waste packing peanuts and agricultural waste. J Ind Microbiol Biotechnol 29:117–123.

14. Zhao X, Zhang L, Liu D. 2012. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels, Bioprod Biorefining 6:465–482.
15. Heredia A, Jiménez A, Guillén R. 1995. Composition of plant cell walls. Z Lebensm Unters Forsch 200:24–31.
16. Zeeman SC, Kossmann J, Smith AM. 2010. Starch: its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol 61:209–34.
17. Puls J. 1997. Chemistry and biochemistry of hemicelluloses: Relationship between hemicellulose structure and enzymes required for hydrolysis. Macromol Symp 120:183–196.
18. Chukwuma OB, Rafatullah M, Tajarudin HA, Ismail N. 2020. Lignocellulolytic enzymes in biotechnological and industrial processes: A review. Sustain 12:1–31.
19. Álvarez C, Reyes-Sosa FM, Díez B. 2016. Enzymatic hydrolysis of biomass from wood. Microb Biotechnol 9:149–156.
20. Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe EC, Baslé 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. Nat Commun 6.
21. Thomson JA. 1993. Molecular biology of xylan degradation. FEMS Microbiol Lett 104:65–82.
22. Moreira LRS, Filho EXF. 2008. An overview of mannan structure and mannan-degrading enzyme systems. Appl Microbiol Biotechnol 79:165–178.
23. Li S-H, Liu S, Colmenares JC, Xu Y-J. 2016. A sustainable approach for lignin valorization by heterogeneous photocatalysis. Green Chem 18:594–607.
24. Lee M, Jeon HS, Kim SH, Chung JH, Roppolo D, Lee H, Cho HJ, Tobimatsu Y, Ralph J, Park OK. 2019. Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. EMBO J 38.
25. dos Santos AC, Ximenes E, Kim Y, Ladisch MR. 2019. Lignin–Enzyme Interactions in the Hydrolysis of Lignocellulosic Biomass. Trends Biotechnol 37:518–531.
26. Chio C, Sain M, Qin W. 2019. Lignin utilization: A review of lignin depolymerization from various aspects. Renew Sustain Energy Rev 107:232–

249.

27. Ezeilo UR, Zakaria II, Huyop F, Wahab RA. 2017. Enzymatic breakdown of lignocellulosic biomass: the role of glycosyl hydrolases and lytic polysaccharide monooxygenases. Biotechnol Biotechnol Equip 1–16.
28. Woo HL, Hazen TC, Simmons BA, DeAngelis KM. 2014. Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. Syst Appl Microbiol 37:60–67.
29. Talebi A, Razali YS, Ismail N, Rafatullah M, Azan Tajarudin H. 2020. Selective adsorption and recovery of volatile fatty acids from fermented landfill leachate by activated carbon process. Sci Total Environ 707:134533.
30. Kuhad RC, Kuhar S, Sharma KK, Shrivastava B. 2013. Microorganisms and Enzymes Involved in Lignin Degradation Vis-à-vis Production of Nutritionally Rich Animal Feed: An Overview, p. 3–44. *In* Biotechnology for Environmental Management and Resource Recovery. Springer India, India.
31. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res 37:233–238.
32. Väljamäe P, Sild V, Nutt A, Pettersson G, Johansson G. 1999. Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. Eur J Biochem 266:327–334.
33. Fox JM, Jess P, Jambusaria RB, Moo GM, Liphardt J, Clark DS, Blanch HW. 2013. A single-molecule analysis reveals morphological targets for cellulase synergy. Nat Chem Biol 9:356–361.
34. Liu Y, Nemmaru B, Chundawat SPS. 2020. Thermobifida fusca Cellulases Exhibit Increased Endo-Exo Synergistic Activity, but Lower Exocellulase Activity, on Cellulose-III. ACS Sustain Chem Eng 8:5028–5039.
35. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. 2012. Novel enzymes for the degradation of cellulose. Biotechnol Biofuels. Biotechnol Biofuels 5:45.
36. Linder M, Teeri TT. 1997. The roles and function of cellulose-binding domains. J Biotechnol 57:15–28.
37. Subramaniyan S, Prema P. 2002. Biotechnology of microbial xylanases: Enzymology, molecular biology, and application. Crit Rev Biotechnol 22:33–64.
38. Wong KK, Tan LU, Saddler JN. 1988. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. Microbiol Rev 52:305–317.

39. Biely P, MacKenzie CR, Puls J, Schneider H. 1986. Cooperativity of Esterases and Xylanases in the Enzymatic Degradation of Acetyl Xylan. Nat Biotechnol 4:731–733.
40. Lei Z, Shao Y, Yin X, Yin D, Guo Y, Yuan J. 2016. Combination of Xylanase and Debranching Enzymes Specific to Wheat Arabinoxylan Improve the Growth Performance and Gut Health of Broilers. J Agric Food Chem 64:4932–4942.
41. Zhang J, Siika-Aho M, Tenkanen M, Viikari L. 2011. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnol Biofuels 4.
42. Kumar R, Singh S, Singh O V. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J Ind Microbiol Biotechnol 35:377–391.
43. Rhee MS, Sawhney N, Kim YS, Rhee HJ, Hurlbert JC, St. John FJ, Nong G, Rice JD, Preston JF. 2017. GH115 α-glucuronidase and GH11 xylanase from Paenibacillus sp. JDR-2: potential roles in processing glucuronoxylans. Appl Microbiol Biotechnol 101:1465–1476.
44. Shi P, Chen X, Meng K, Huang H, Bai Y, Luo H, Yang P, Yao B. 2013. Distinct Actions by Paenibacillus sp. Strain E18 α-l-Arabinofuranosidases and Xylanase in Xylan Degradation. Appl Environ Microbiol 79:1990–1995.
45. SUGIMURA M, NISHIMOTO M, KITAOKA M. 2006. Characterization of Glycosynthase Mutants Derived from Glycoside Hydrolase Family 10 Xylanases. Biosci Biotechnol Biochem 70:1210–1217.
46. Mccleary B V., Matheson NK. 1987. Enzymic Analysis of Polysaccharide Structure, p. 147–276. *In* .
47. Malgas S, van Dyk JS, Pletschke BI. 2015. A review of the enzymatic hydrolysis of mannans and synergistic interactions between β-mannanase, β-mannosidase and α-galactosidase. World J Microbiol Biotechnol 31:1167–1175.
48. Rosengren A, Reddy SK, Sjöberg JS, Aurelius O, Logan DT, Kolenová K, Stålbrand H. 2014. An Aspergillus nidulans β-mannanase with high transglycosylation capacity revealed through comparative studies within glycosidase family 5. Appl Microbiol Biotechnol 98:10091–10104.
49. Schroder R, Wegrzyn T, Bolitho K, Redgwell R. 2004. Mannan transglycosylase: a novel enzyme activity in cell walls of higher plants. Planta 219.
50. Duval A, Lawoko M. 2014. A review on lignin-based polymeric, micro- and nano-structured materials. React Funct Polym 85:78–96.

51. Pollegioni L, Tonin F, Rosini E. 2015. Lignin-degrading enzymes. FEBS J 282:1190–1213.
52. Balan V, Bals B, Chundawat SPS, Marshall D, Dale BE. 2009. Lignocellulosic Biomass Pretreatment Using AFEX, p. 61–77. *In* .
53. Tadesse H, Luque R. 2011. Advances on biomass pretreatment using ionic liquids: An overview. Energy Environ Sci 4:3913.
54. Bhardwaj N, Kumar B, Agrawal K, Verma P. 2021. Current perspective on production and applications of microbial cellulases: a review. Bioresour Bioprocess 8:95.
55. Guo H, Wang X-D, Lee D-J. 2018. Proteomic researches for lignocellulose-degrading enzymes: A mini-review. Bioresour Technol 265:532–541.
56. Van Dyk JS, Pletschke BI. 2012. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. Biotechnol Adv 30:1458–1480.
57. Hu J, Arantes V, Pribowo A, Saddler JN. 2013. The synergistic action of accessory enzymes enhances the hydrolytic potential of a “cellulase mixture” but is highly substrate specific. Biotechnol Biofuels 6:112.
58. Selig MJ, Knoshaug EP, Adney WS, Himmel ME, Decker SR. 2008. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. Bioresour Technol 99:4997–5005.
59. Ong LGA, Abd-Aziz S, Noraini S, Karim MIA, Hassan MA. 2004. Enzyme Production and Profile by Aspergillus niger During Solid Substrate Fermentation Using Palm Kernel Cake as Substrate. Appl Biochem Biotechnol 118:073–080.
60. Saloheimo M, Lehtovaara P, Penttilä M, Teeri TT, Ståhlberg J, Johansson G, Pettersson G, Claeyssens M, Tomme P, Knowles JKC. 1988. EGIII, a new endoglucanase from Trichoderma reesei: the characterization of both gene and enzyme. Gene 63:11–21.
61. Jagadeeswaran G, Gainey L, Prade R, Mort AJ. 2016. A family of AA9 lytic polysaccharide monooxygenases in Aspergillus nidulans is differentially regulated by multiple substrates and at least one is active on cellulose and xyloglucan. Appl Microbiol Biotechnol 100:4535–4547.
62. Lazcano C, Gómez-Brandón M, Revilla P, Domínguez J. 2013. Short-term effects of organic and inorganic fertilizers on soil microbial community structure and function. Biol Fertil Soils 49:723–733.

63. Maki M, Leung KT, Qin W. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. Int J Biol Sci 5:500–516.
64. Lamed R, Bayer EA. 1988. The Cellulosome of Clostridium thermocellum, p. 1–46. *In* .
65. Bégum P, Lemaire M. 1996. The Cellulosome: An Exocellular, Multiprotein Complex Specialized in Cellulose Degradation. Crit Rev Biochem Mol Biol 31:201–236.
66. Malla S, Prasad Niraula N, Singh B, Liou K, Kyung Sohng J. 2010. Limitations in doxorubicin production from Streptomyces peucetius. Microbiol Res 165:427–435.
67. Lee N, Hwang S, Kim J, Cho S, Palsson B, Cho B-K. 2020. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in Streptomyces. Comput Struct Biotechnol J 18:1548–1556.
68. Ramachander TVN, Rawal SK. 2005. PHB synthase from Streptomyces aureofaciens NRRL 2209. FEMS Microbiol Lett 242:13–18.
69. Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp Sirex noctilio. ISME J 5:1323–1331.
70. Thompson BM, Grebenok RJ, Behmer ST, Gruner DS. 2013. Microbial Symbionts Shape the Sterol Profile of the Xylem-Feeding Woodwasp, Sirex noctilio. J Chem Ecol 39:129–139.
71. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. 2013. Aerobic deconstruction of cellulosic biomass by an insect-associated Streptomyces. Sci Rep 3:1–10.
72. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, Blodgett JA V, Clardy J, Raffa KF, Fox BG, Currie CR. 2014. Cellulolytic Streptomyces strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. Appl Environ Microbiol 80:4692–4701.
73. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, Doering DT, Suh S, Raffa KF, Fox BG, Currie CR. 2016. Evolution of High Cellulolytic Activity in Symbiotic Streptomyces through Selection of Expanded Gene Content and Coordinated Gene Expression. PLoS Biol 14:1–21.

74. Poulsen M, Oh DC, Clardy J, Currie CR. 2011. Chemical analyses of wasp-associated Streptomyces bacteria reveal a prolific potential for natural products discovery. PLoS One 6.
75. Lewin GR, Carlos C, Chevrette MG, Horn HA, McDonald BR, Stankey RJ, Fox BG, Currie CR. 2016. Evolution and Ecology of *Actinobacteria* and Their Bioenergy Applications. Annu Rev Microbiol 70:235–254.
76. Demain AL, Sanchez S. 2009. Microbial drug discovery: 80 Years of progress. J Antibiot (Tokyo) 62:5–16.
77. Clardy J, Fischbach MA, Walsh CT. 2006. New antibiotics from bacterial natural products. Nat Biotechnol 24:1541–1550.
78. Alvarez A, Saez JM, Davila Costa JS, Colin VL, Fuentes MS, Cuozzo SA, Benimeli CS, Polti MA, Amoroso MJ. 2017. Actinobacteria: Current research and perspectives for bioremediation of pesticides and heavy metals. Chemosphere 166:41–62.
79. Marushima K, Ohnishi Y, Horinouchi S. 2009. CebR as a master regulator for cellulose/cellooligosaccharide catabolism affects morphological development in Streptomyces griseus. J Bacteriol 191:5930–5940.
80. Brennan MSM, Chauhan SS. 2015. Wood quality assessment of Pinus radiata (radiata pine) saplings by dynamic mechanical analysis. Wood Sci Technol 49:1239–1250.
81. Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem 31:426–428.
82. Nielsen H. 2017. Predicting Secretory Proteins with SignalP. Methods Mol Biol 1611:59–73.
83. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445-51.
84. Klock HE, Lesley SA. 2009. The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis, p. 91–103. *In* .
85. Bianchetti CM, Takasuka TE, Deutsch S, Udell HS, Yik EJ, Bergeman LF, Fox BG. 2015. Active site and laminarin binding in glycoside hydrolase family 55. J Biol Chem 290:11819–11832.
86. Zhang Q, Huang Q, Fang Q, Li H, Tang H, Zou G, Wang D, Li S, Bei W, Chen

H, Li L, Zhou R. 2020. Identification of genes regulated by the two-component system response regulator NarP of Actinobacillus pleuropneumoniae via DNA-affinity-purified sequencing. Microbiol Res 230:126343.

87. Mizugishi K, Aruga J, Nakata K, Mikoshiba K. 2001. Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins. J Biol Chem 276:2180–2188.
88. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4:1265–72.
89. Takasuka TE, Acheson JF, Bianchetti CM, Prom BM, Bergeman LF, Book AJ, Currie CR, Fox BG. 2014. Biochemical properties and atomic resolution structure of a proteolytically processed β-mannanase from cellulolytic Streptomyces sp. SirexAA-E. PLoS One 9.
90. Book AJ, Yennamalli RM, Takasuka TE, Currie CR, Phillips GN, Fox BG. 2014. Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases. Biotechnol Biofuels 7:109.
91. Choi J, Klingeman DM, Brown SD, Cox CD. 2017. The LacI family protein GlyR3 co-regulates the celC operon and manB in Clostridium thermocellum. Biotechnol Biofuels 10:156.
92. Tsujibo H, Kosaka M, Ikenishi S, Sato T, Miyamoto K. 2004. Molecular Characterization of a High-Affinity Xylobiose Transporter of Streptomyces thermoviolaceus OPC-520 and Its Transcriptional Regulation 186:1029–1037.
93. Dubeau MP, Guay I, Brzezinski R. 2011. Modification of genetic regulation of a heterologous chitosanase gene in Streptomyces lividans TK24 leads to chitosanase production in the absence of chitosan. Microb Cell Fact 10:1–10.
94. Hoa TT, Tortosa P, Albano M, Dubnau D. 2002. Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK. Mol Microbiol 43:15–26.
95. Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annu Rev Plant Biol 61:263–289.
96. Chen J, Matthews KS. 1994. Subunit Dissociation Affects DNA Binding in a Dimeric Lac Repressor Produced by C-Terminal Deletion. Biochemistry 33:8728–8735.
97. Whitson PA, Matthews KS. 1986. Dissociation of the Lactose Repressor—

Operator DNA Complex: Effects of Size and Sequence Context of Operator-Containing DNA. Biochemistry 25:3845–3852.

98. Schlösser A, Aldekamp T, Schrempf H. 2000. Binding characteristics of CebR, the regulator of the ceb operon required for cellobiose/cellotriose uptake in Streptomyces reticuli. FEMS Microbiol Lett 190:127–132.
99. Spiridonov NA, Wilson DB. 1999. Characterization and cloning of CelR, a transcriptional regulator of cellulase genes from Thermomonospora fusca. J Biol Chem 274:13127–13132.
100. Devine CS, Rangwala SH. 1988. The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically. Gene 73:227–235.
101. Belitsky BR, Sonenshein AL. 1999. An enhancer element located downstream of the major glutamate dehydrogenase gene of Bacillus subtilis. Proc Natl Acad Sci U S A 96:10290–10295.
102. Yin S, Wang W, Wang X, Zhu Y, Jia X, Li S, Yuan F, Zhang Y, Yang K. 2015. Identification of a cluster-situated activator of oxytetracycline biosynthesis and manipulation of its expression for improved oxytetracycline production in Streptomyces rimosus. Microb Cell Fact 14:46.
103. Hori C, Song R, Matsumoto K, Matsumoto R, Minkoff BB, Oita S, Hara H, Takasuka TE. 2020. Proteomic Characterization of Lignocellulolytic Enzymes Secreted by the Insect-Associated Fungus Daldinia decipiens oita, Isolated from a Forest in Northern Japan. Appl Environ Microbiol 86.
104. Saito A, Kaku H, Minami E, Fujii T, Ando A, Nagata Y, Schrempf H, Miyashita K. 2006. An enzyme-linked immunosorbent assay (ELISA) to determine the specificity of the sugar-binding protein NgcE, a component of the ABC transporter for N-acetylglucosamine in Streptomyces olivaceoviridis. Biosci Biotechnol Biochem 70:237–242.
105. Zhu Y, Suits MDL, Thompson AJ, Chavan S, Dumon C, Smith N, Moremen KW, Xiang Y, Williams SJ, Gilbert HJ, Davies GJ. 2014. Mechanistic insights into a Ca2+ -dependent family of α- mannosidases in a human gut symbiont 6:125–132.
106. Duong A, Capstick DS, Di Berardo C, Findlay KC, Hesketh A, Hong HJ, Elliot MA. 2012. Aerial development in Streptomyces coelicolor requires sortase activity. Mol Microbiol 83:992–1005.
107. Galbe M, Zacchi G. 2002. A review of the production of ethanol from softwood.

Appl Microbiol Biotechnol 59:618–628.

108. Abdel-Rahman MA, Tashiro Y, Sonomoto K. 2011. Lactic acid production from lignocellulose-derived sugars using lactic acid bacteria: Overview and limits. J Biotechnol 156:286–301.
109. Albuquerque TL de, da Silva IJ, de Macedo GR, Rocha MVP. 2014. Biotechnological production of xylitol from lignocellulosic wastes: A review. Process Biochem 49:1779–1789.
110. Bastawde KB. 1992. Xylan structure, microbial xylanases, and their mode of action. World J Microbiol Biotechnol 8:353–368.
111. Biely P. 1985. Microbial xylanolytic systems. Trends Biotechnol 3:286–290.
112. Malgas S, Mafa MS, Mkabayi L, Pletschke BI. 2019. A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation. World J Microbiol Biotechnol 35:1–13.
113. Raweesri P, Riangrungrojana P, Pinphanichakarn P. 2008. α-l-Arabinofuranosidase from Streptomyces sp. PC22: Purification, characterization and its synergistic action with xylanolytic enzymes in the degradation of xylan and agricultural residues. Bioresour Technol 99:8981–8986.
114. Popa A, Israel-Roming F, Cornea CP. 2016. Microbial xylanase: A review.
115. Beg QK, Bhushan B, Kapoor M, Hoondal GS. 2000. Production and characterization of thermostable xylanase and pectinase from Streptomyces sp. QG-11-3. J Ind Microbiol Biotechnol 24:396–402.
116. Techapun C, Poosaran N, Watanabe M, Sasaki K. 2003. Thermostable and alkaline-tolerant microbial cellulase-free xylanases produced from agricultural wastes and the properties required for use in pulp bleaching bioprocesses: a review. Process Biochem 38:1327–1340.
117. Beg QK, Bhushan B, Kapoor M, Hoondal GS. 2000. Enhanced production of a thermostable xylanase from Streptomyces sp. QG-11-3 and its application in biobleaching of eucalyptus kraft pulp. Enzyme Microb Technol 27:459–466.
118. Techapun C, Charoenrat T, Poosaran N, Watanabe M, Sasak K. 2002. Thermostable and alkaline-tolerant cellulase-free xylanase produced by thermotolerant Streptomyces sp. Ab106. J Biosci Bioeng 93:431–433.
119. TSUJIBO H, TAKADA C, WAKAMATSU Y, KOSAKA M, TSUJI A, MIYAMOTO K, INAMORI Y. 2003. Cloning and Expression of an α- L -Arabinofuranosidase Gene (stxIV) from Streptomyces thermoviolaceus OPC-520, and

Characterization of the Enzyme . Biosci Biotechnol Biochem 66:434–438.

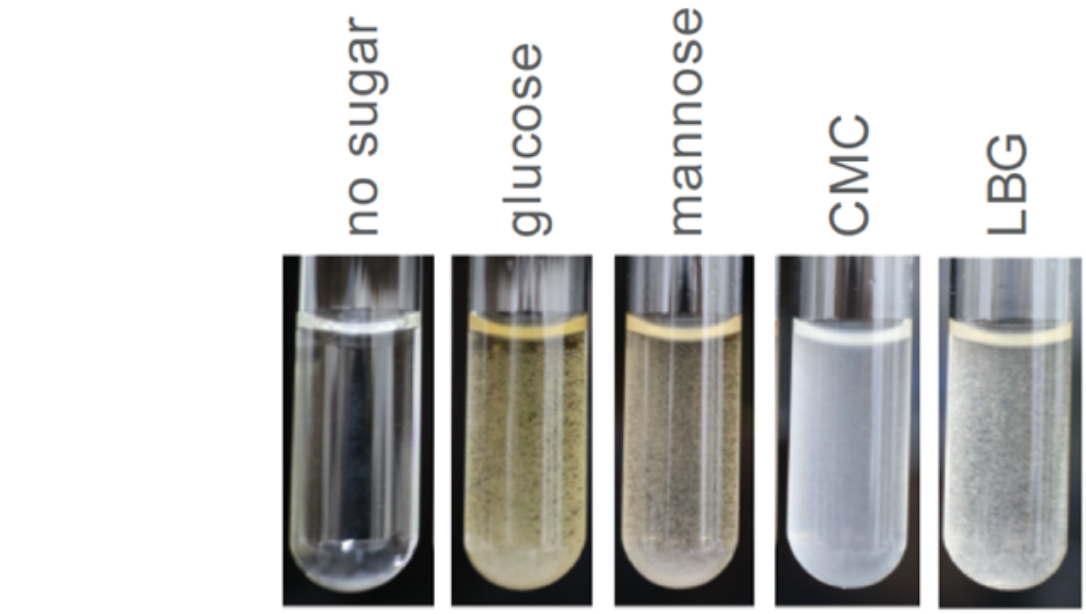
120. Tsujibo H, Ohtsuki T, Iio T, Yamazaki I, Miyamoto K, Sugiyama M, Inamori Y. 1997. Cloning and sequence analysis of genes encoding xylanases and acetyl xylan esterase from Streptomyces thermoviolaceus OPC-520. Appl Environ Microbiol 63:661–664.
121. TSUJIBO H, TAKADA C, TSUJI A, KOSAKA M, MIYAMOTO K, INAMORI Y. 2002. Cloning, Sequencing, and Expression of the Gene Encoding an Intracellular β-D-Xylosidase from Streptomyces thermoviolaceus OPC-520. Biosci Biotechnol Biochem.
122. Tsujibo H, Miyamoto K, Kuda T, Minami K, Sakamoto T, Hasegawa T, Inamori Y. 1992. Purification, properties, and partial amino acid sequences of thermostable xylanases from Streptomyces thermoviolaceus OPC-520. Appl Environ Microbiol 58:371–375.
123. Ohashi K, Hataya S, Nakata A, Matsumoto K, Kato N, Sato W, Carlos-Shanley C, Beebe ET, Currie CR, Fox BG, Takasuka TE. 2021. Mannose- and Mannobiose-Specific Responses of the Insect-Associated Cellulolytic Bacterium Streptomyces sp. Strain SirexAA-E. Appl Environ Microbiol 87.
124. Bekiesch P, Franz-Wachtel M, Kulik A, Brocker M, Forchhammer K, Gust B, Apel AK. 2016. DNA affinity capturing identifies new regulators of the heterologously expressed novobiocin gene cluster in Streptomyces coelicolor M512. Appl Microbiol Biotechnol 100:4495–4509.
125. Li MZ, Elledge SJ. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods 4:251–256.
126. Takasuka TE, Bianchetti CM, Tobimatsu Y, Bergeman LF, Ralph J, Fox BG. 2014. Structure-guided analysis of catalytic specificity of the abundantly secreted chitosanase SACTE_5457 from Streptomyces sp. SirexAA-E. Proteins Struct Funct Bioinforma 82:1245–1257.
127. Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234.
128. Li J, Wang J, Ruiz-Cruz S, Espinosa M, Zhang J-R, Bravo A. 2020. In vitro DNA Inversions Mediated by the PsrA Site-Specific Tyrosine Recombinase of Streptococcus pneumoniae. Front Mol Biosci 7.
129. Takasuka TE, Hsieh Y-J, Stein A. 2014. Miniaturized Sequencing Gel System for Quick Analysis of DNA by Hydroxyl Radical Cleavage. Appl Biochem

Biotechnol 172:1–8.

130. Navone L, Macagno JP, Licona-Cassani C, Marcellin E, Nielsen LK, Gramajo H, Rodriguez E. 2015. AllR controls the expression of Streptomyces coelicolor allantoin pathway genes. Appl Environ Microbiol 81:6649–6659.

131. Nentwich SS, Brinkrolf K, Gaigalat L, Hüser AT, Rey DA, Mohrbach T, Marin K, Pühler A, Tauch A, Kalinowski J. 2009. Characterization of the LacI-type transcriptional repressor RbsR controlling ribose transport in Corynebacterium glutamicum ATCC 13032. Microbiology 155:150–164.

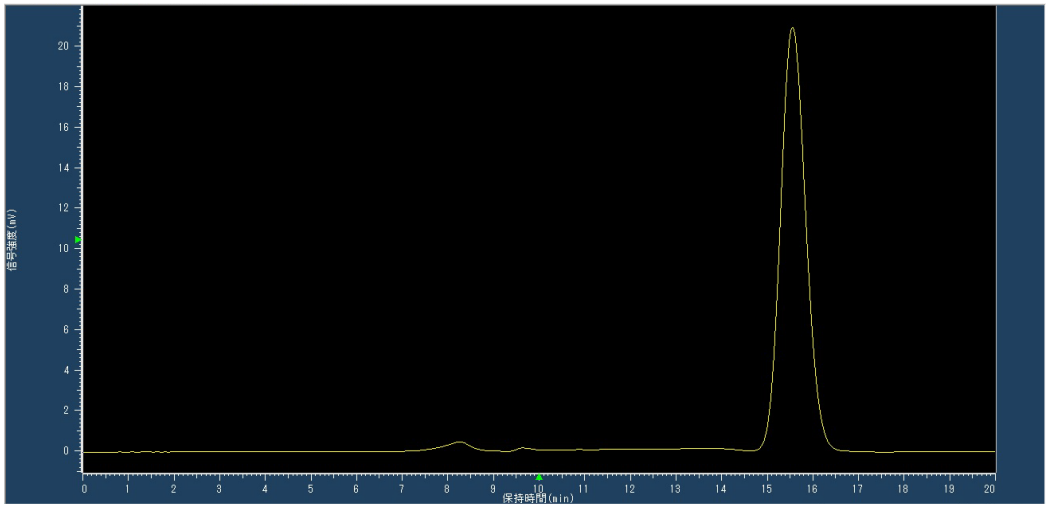
APPENDICES



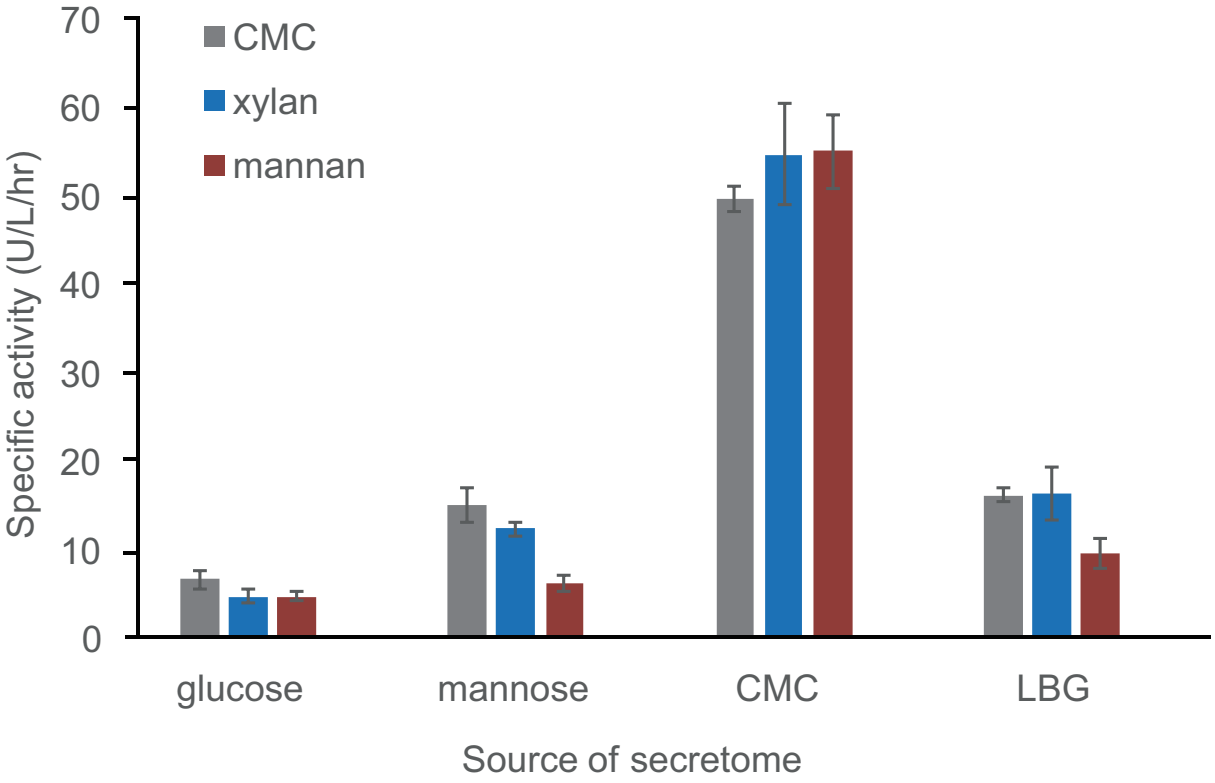
Appendix Fig. S1. Appearance of SirexAA-E cultures after growth on different carbon sources. SirexAA-E was grown in minimal (M63) media containing 0.5% D-glucose, D-mannose, CMC and LBG as the sole carbon source. The culture was incubated for 3 days at 30˚C. The dry cell weight from each culture was measured and shown (n=3).

The HPLC analysis of contaminated mono/disaccharide in 50 μg of the LBG

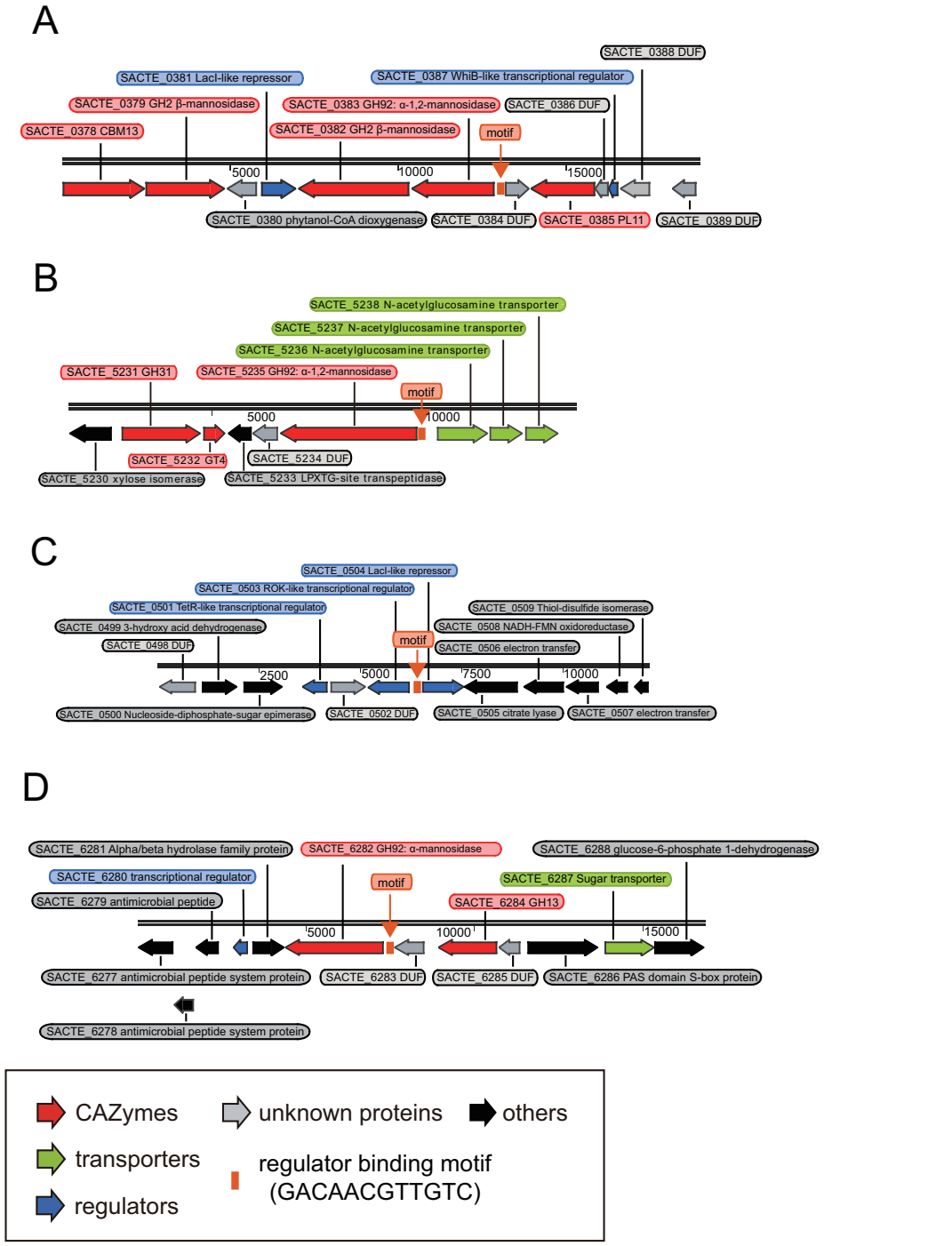
	retention time [min]	amount [μg]
cellobiose	13.3	not detected
glucose	15.6	24.7 ± 3.3
mannose	18.0	not detected
galactose	19.7	not detected



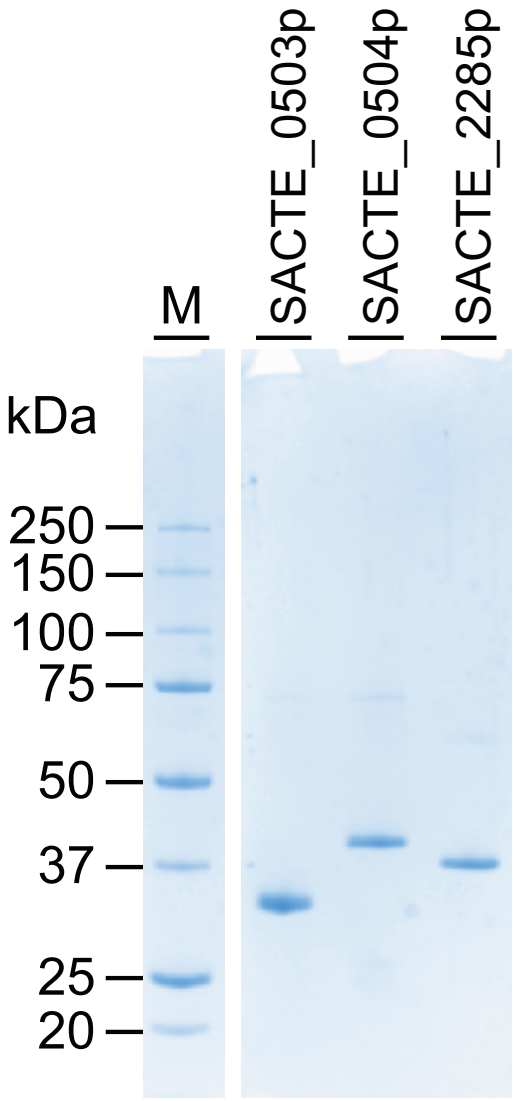
Appendix Fig. S2. HPLC analysis of the sugar composition in the commercial LBG.



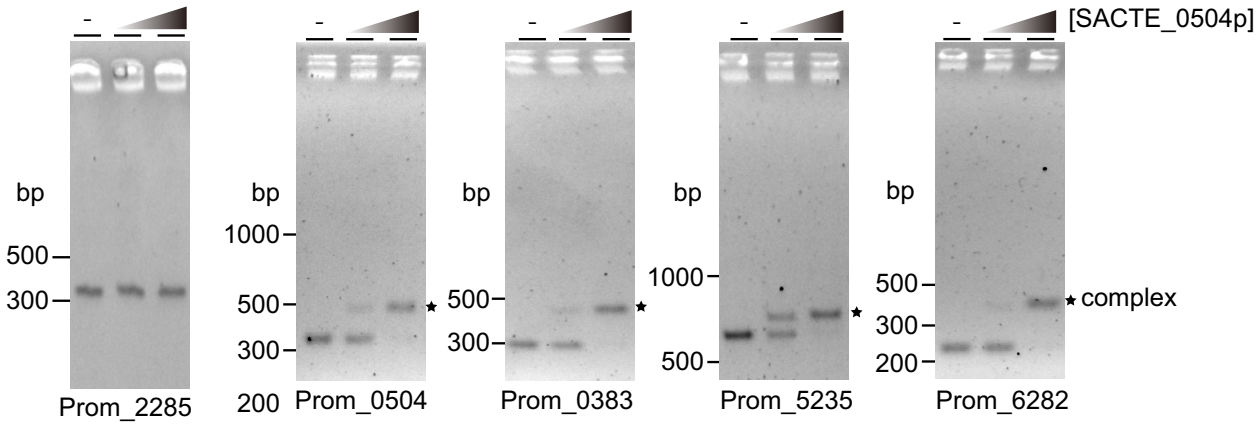
Appendix Fig. S3. Specific activity (U/L/hr) of the culture supernatants of glucose, mannose, CMC, and LBG media for polysaccharide degradation is shown. (A) Cellulase, xylanase, and mannanase activities from SirexAA-E culture supernatants were measured with three different polysaccharides, CMC, xylan, and mannan for 20 h at 40˚C. Error bars indicate the standard deviation from three independent experiments ($P<0.05$).



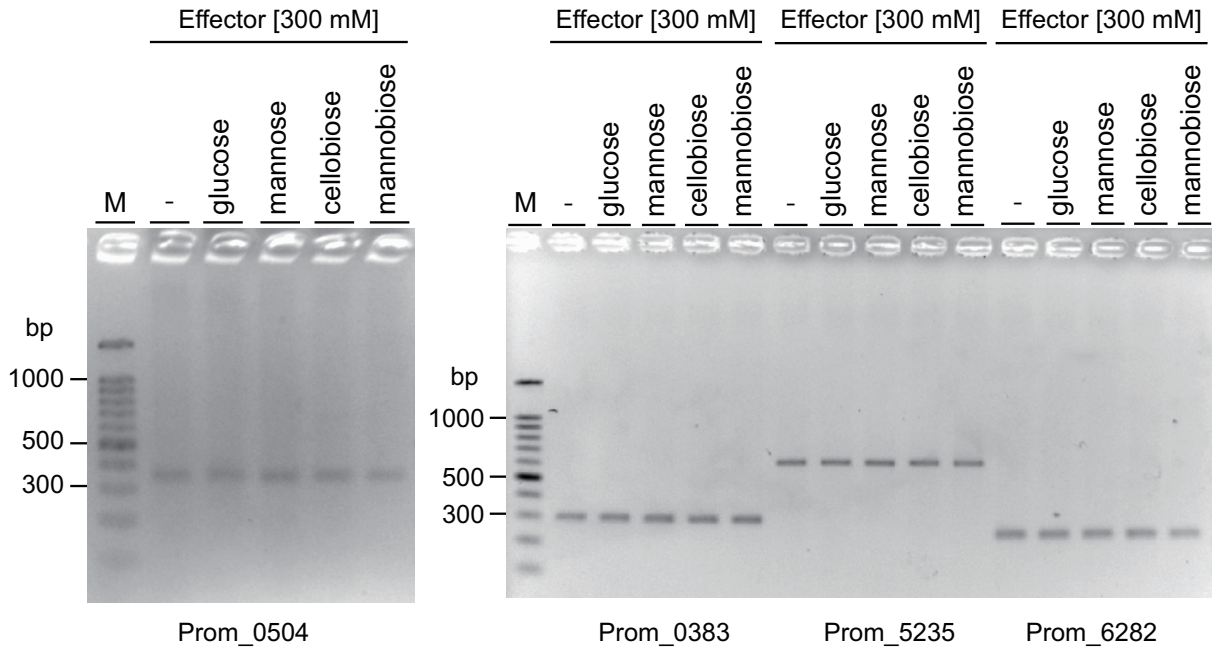
Appendix Fig. S4. Putative repressor binding motifs in the SirexAA-E genome. Four genome locations (A to D) are shown with the potential binding motif to mannose responsive repressor, indicated by the orange filled box. The red, blue, and green filled arrows denote the genes annotated as CAZymes, transcriptional regulators, and transporters, respectively. The black and gray filled arrows indicated genes encoding the proteins with other functions or the proteins with unknown functions (DUF).



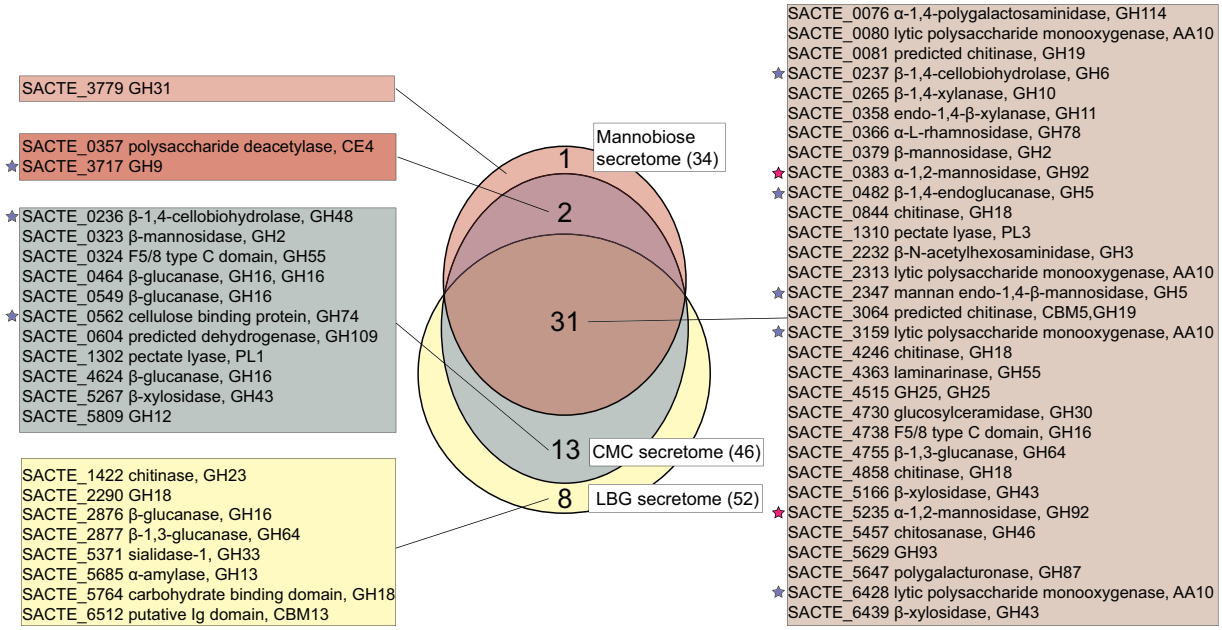
Appendix Fig. S5. SDS-PAGE of recombinant proteins. Three proteins, SACTE_0503p (32.6 kDa); 2, SACTE_0504p (37.5 kDa); 3, SACTE_2285p (37.4 kDa), were overexpressed and his-tag purified, and each purified protein was found at the around estimated size.



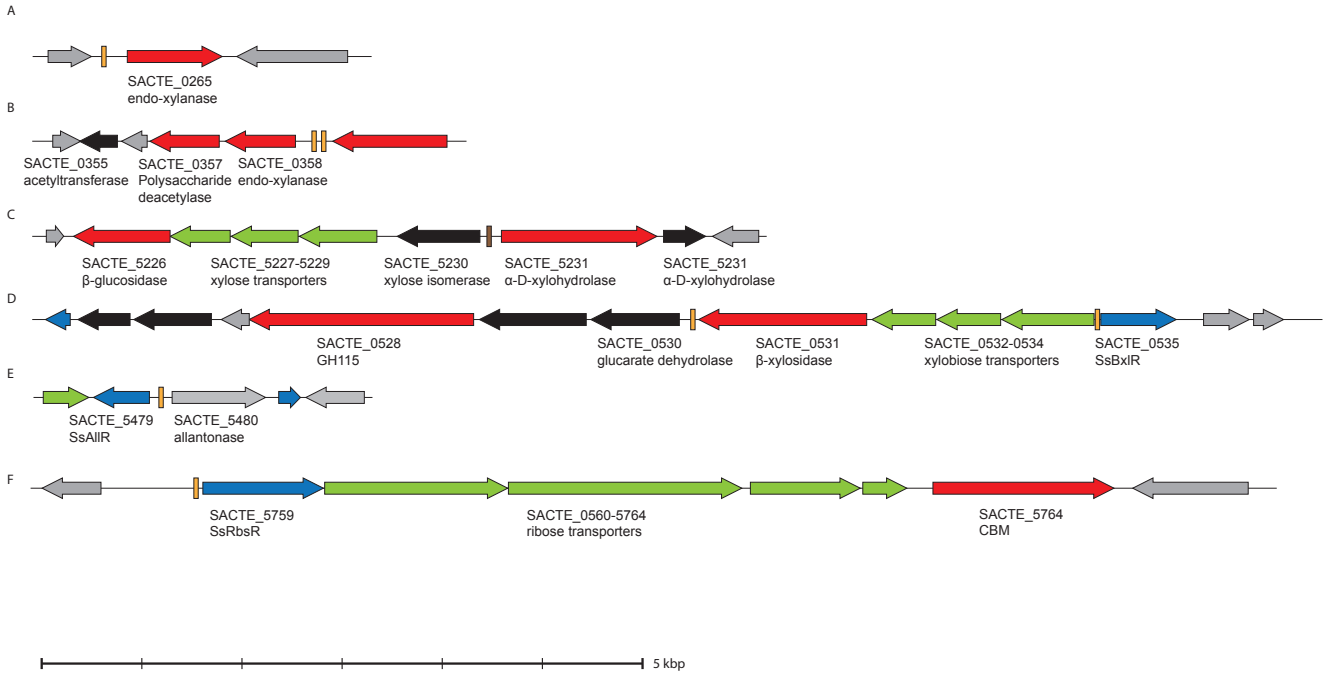
Appendix Fig. S6. A full gel image of EMSA analysis shown in Fig. 2.2.



Appendix Fig. S7. EMSA analysis and sugar effector assays of SACTE_0503p to the possible binding sites. SACTE_0503p was incubated with 10 ng of four promoter regions, Prom_0504, Prom_0383, Prom_5235, and Prom_6282 in the presence and absence of four soluble sugars, including 300 mM glucose, mannose, cellobiose, and mannobiose.



Appendix Fig. S8. The Venn diagram of CAZymes in the top 100 secreted proteins in the SirexAA-E secretomes identified in the mannobiose (34 CAZymes), CMC (46 CAZymes), and LBG (52 CAZymes) culture supernatants is shown. Each category indicated the number of CAZymes determined in either one, two, or all secretomes and listed with locus ID and putative functions with corresponding filled color boxes. The red and purple filled stars indicated the CAZymes under the regulation of SsManR, and SsCebR, respectively.



Appendix Fig. S9. Putative repressor binding motifs in the SirexAA-E genome. Four genome locations (A to F) are shown with the potential binding motif to three transcriptional regulators, indicated by the orange filled box. The red, blue, and green filled arrows denote the genes annotated as CAZymes, transcriptional regulators, and transporters, respectively. The black and gray filled arrows indicated genes encoding the proteins with other functions or the proteins with unknown functions (DUF).

Appendix TableS1. Protein concentration in the culture supernatant

Table S1. Protein concentration in the culture supernatant.

Source of secretome	SirexAA-E [µg/mL]
Glucose secretome	1.8 ± 0.6
Mannose secretome	10.5 ± 2.1
CMC secretome	15.1 ± 0.9
LBG secretome	12.2 ± 5.1

Appendix Table S2. Proteome analysis of SirexAA-E secretomes on the growth with CMC, LBG, glucose, mannose, and mannobiose.

Appendix Table S3. Comparison of secreted CAZymes among the culture supernatant prepared from different carbon sources.