Studies on molecular recognition and degradation mechanism of plant cell wall polysaccharides-assimilating *Clostridium cellulovorans* using proteome analysis

Shunsuke ABURAYA

2019

Contents

Abbreviations

AICc	Akaike’s information criterion with a correction for finite sample sizes
AGC	Target value of auto gain control
CAZy	Carbohydrate-active enzymes
CE	Carbohydrate esterase
CID	Collision induced dissociation
EC	Enzyme commission number
ESI	Electrospray ionization
FDR	False discovery rate
GH	Glycoside hydrolases
GT	Glycosyltransferase
HCA	Hierarchical cluster analysis
HCD	Higher-energy c-trap dissociation
ID	Internal diameter
IT	Maximum ion injection time
KEGG	Kyoto encyclopedia of genes and genomes
LBG	Locust bean gum
LC	Liquid chromatography
MS/MS	Tandem mass spectrometry
NCBI	National center for biotechnology information
PBS	Phosphate buffered saline
PL	Polysaccharide lyase
SDS	Sodium dodecyl sulfate
TCS	Two component system
TEAB	Triethyl ammonium bicarbonate
TMT	Tandem mass tag
TP	Time point

Introduction

Humanity's lifestyle relies on fossil fuels, but the overconsumption of such fuels gives rise to too much CO_2 emission and environmental pollution (Mood et al., 2013). Furthermore, the supply of fossil fuels is limited (Mood et al., 2013). Therefore, new resources that can act as alternatives to fossil fuels are required. Such alternative resources have some requirements; that is, they should be renewable and carbon neutral to overcome the limitations in fossil fuels, and they should be able to manufacture products that are usually made from fossil fuels. Plant biomass has attracted much attention as an alternative resource to satisfy these requirements, and can be used to produce bioethanol and useful chemical products.

Problems with plant biomass

Plant biomass has the potential to overcome the limitation of fossil fuels because it is a renewable and carbon neutral resource (Lee, 1997; Menon and Rao, 2012; Mood et al., 2013). Nowadays, in Brazil, bioethanol production is being conducted using the starch from edible polysaccharides, such as corn and sugarcane. However, bioethanol production from edible polysaccharides remains problematic. First, expanding edible biomass production for fuel creates a poor energy balance (energy output from the biofuel/energy inputs used for production) and has negative impacts on the regional water resources, biodiversity, and soil quality (Groom et al., 2008). Second, a vast amount of agrowastes is released from edible biomass-based biofuel production (Bjerre et al., 1996). Third, the use of limited agricultural land for producing biomass for fuel competes with food production. Therefore, new biomass sources are required to overcome these dilemmas (Sun and Cheng, 2002).

Currently, lignocellulosic biomass has gained much attention as an alternative and renewable resource (Lynd et al., 2005). This biomass type is a cheap, abundant, and renewable carbon source, being obtained as agricultural byproducts and industrial garbage (Lynd et al., 1999). Lignocellulosic biomass has a highly complex and robust structure, being made up of cellulose, hemicellulose, pectin, and lignin, making its

hydrolysis to simple sugars difficult (Agbor et al., 2011; Mood et al., 2013). The major polysaccharide components of some lignocellulosic biomass materials are shown in Table 1. Although lignocellulosic biomass is composed mainly of cellulose, some materials contain mostly hemicellulose. Therefore, all polysaccharide components of lignocellulosic biomass have to be degraded.

Table 1 Composition of the polysaccharides in major lignocellulosic biomass materials

Lignocellulosic biomass	Cellulose	Hemicellulose				Others	Ref.
	Glucan	Xylan	Arabinan	Galactan	Mannan		
Corn cob	34.3	31.1	3.0	-	-	31.6	(Garrote et al., 2007)
Corn stover	38.2	21.0	2.7	2.1	-	36.0	(Li et al., 2010c)
Rice straw	31.1	18.7	3.6	-	-	46.6	(Chen et al., 2011)
Switchgrass	39.5	20.3	2.1	2.6	-	35.5	(Li et al., 2010a)
Sugarcane bagasse	43.3	22.4	2.3	0.6	0.4	31.0	(Girisuta et al., 2013)
Wheat straw	39.8	34.5	2.8	-	-	22.9	(Kristensen et al., 2008)

All values are percentages [%] based on dry weight

Structural polysaccharides of the plant cell wall

Cellulose

Cellulose, a major component of the plant cell wall, is the most abundant polysaccharide in lignocellulosic biomass. It is a linear and unbranched homopolymer of ß-1,4-glycosidic bonds. Each cellulose chain is bundled together to construct cellulose microfibrils that are linked by intramolecular or intermolecular hydrogen bonds (Li et al., 2010b) and van der Waals forces (Agbor et al., 2011). This structure makes cellulose difficult to degrade with enzymes.

Hemicellulose

Hemicellulose has a complex structure made up of pentose (xylose and arabinose), hexose (mannose, glucose, and galactose), and acetylated sugars. The main component of hemicellulose in hard wood, straw, and grass is xylan, whereas it is glucomannan in soft wood (Agbor et al., 2011). Hemicellulose is thought to crosslink cellulose and lignin, with the crosslinked structure forming the entire structure of the plant cell wall (Cosgrove, 2005; Hayashi, 1989). Hemicellulose has a branched and amorphous structure and a low molecular weight, and because it is sensitive to enzymes, it degrades more easily than cellulose does (Li et al., 2010b; Saha, 2003). Therefore, the degradation of hemicellulose increases the degradation efficiency of lignocellulosic biomass (Agbor et al., 2011).

Pectin

Pectin is a family of galacturonic acid-rich polysaccharides composed of homogalacturonan, rhamnogalacturonan I, and rhamnogalacturonan II (Mohnen, 2008). Pectin has the most complex polysaccharide structure in the living world (Willats et al., 2001), and these polysaccharides are thought to be crosslinked together via covalent bonds (Cumming et al., 2005; Vincken et al., 2003). Furthermore, it was suggested that pectins may serve to hold some hemicelluloses in the cell wall (Cumming et al., 2005).

Homogalacturonan is a linear homopolymer of galacturonic acid, and is methylesterified at the C-6 carboxyl residues and *O*-acetylated at O-2 and O-3 (O'Neill et al., 1990).

Rhamnogalacturonan I is composed of galacturonic acid, rhamnose, α-L-arabinofuranosyl, and β-D-galactopyranosyl residues (Mohnen, 2008). The type and number of sugars are quite different according to the developmental stages and cell types (Ridley et al., 2001).

Rhamnogalacturonan II has 12 different sugar residues in its main structure (Mohnen, 2008), and it forms a dimer with the borate ester (Kobayashi et al., 1996).

Cellulosome

Cellulosomes, which are localized at the bacterial cell surface, are valuable multienzymatic complexes composed of scaffoldins and enzymes for the degradation of

lignocellulosic biomass (Figure 1) (Artzi et al., 2017; Shoham et al., 1999). The scaffoldins play a role in the assembly of other cellulosomal proteins and in binding to carbohydrates (Artzi et al., 2017; Shoham et al., 1999). The scaffoldins are made up of cohesin domains that interact with a dockerin domain, and are the carbohydrate-binding modules that bind to and arrange enzymes at appropriate sites on the plant cell wall (Artzi et al., 2017; Shoham et al., 1999). The dockerin domain is usually fused with carbohydrate-active enzymes (cellulosomal proteins) (Artzi et al., 2017; Shoham et al., 1999). In the cellulosome, these carbohydrate-active enzymes with dockerin domains are proximally positioned by the consecutive cohesin domains (Artzi et al., 2017; Shoham et al., 1999). Therefore the assembled carbohydrate-active enzymes react synergistically, and cellulosomes have a higher activity than free carbohydrate-active enzymes do for the degradation of polysaccharides (Murashima et al., 2003).

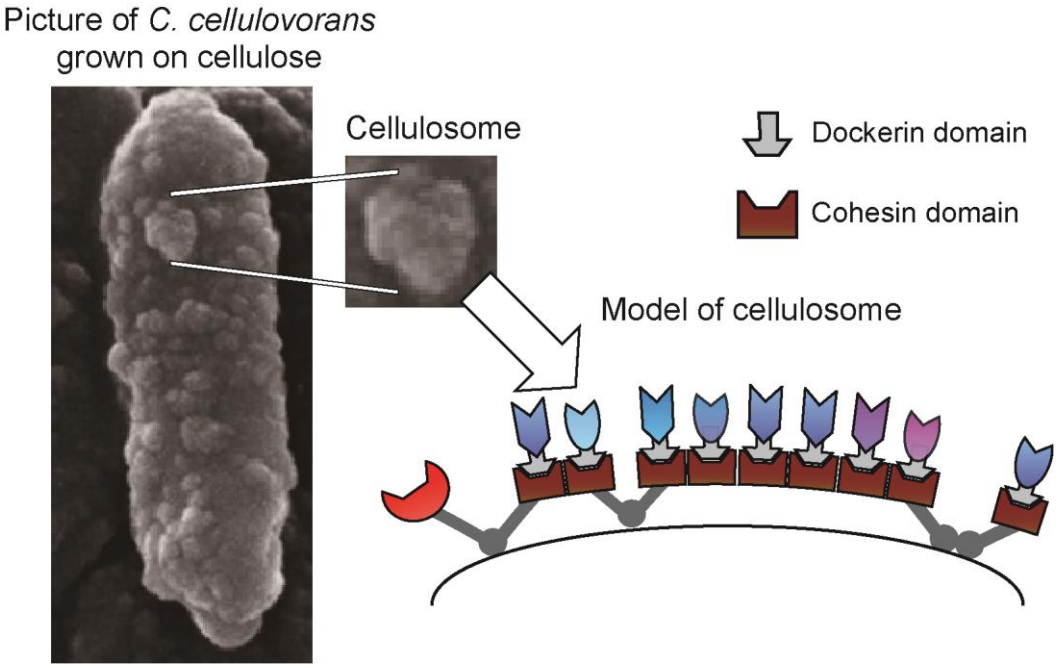


Figure 1 Schematic diagram of the cellulosome of *Clostridium cellulovorans*
(Modified from Blair and Anderson, 1999)

With regard to their structure, there are two types of cellulosomes: simple cellulosomes and highly structured cellulosomes. Simple cellulosomes are produced by mesophilic bacteria, such as *Clostridium cellulolyticum* (Gal et al., 1997), *Clostridium cellulovorans* (Doi and Tamaru, 2001), *Ruminococcus albus* (Ohara et al., 2000), and *Ruminococcus bromii* (Cann et al., 2016). A simple cellulosome is constructed from a single primary scaffoldin and does not have a complex structure (Figure 2a) (Artzi et al., 2017). Highly structured cellulosomes are produced by thermophilic and mesophilic bacteria, such as *Clostridium thermocellum* (Bayer et al., 1985) and *Acetivibrio cellulolyticus* (Xu et al., 2003). A highly structured cellulosome is constructed from multiple scaffoldins, and a single complex has many scaffoldins and carbohydrate-degrading enzymes (Bayer et al., 1998). The mechanism by which simple cellulosomes bind to the cell surface has not been elucidated. On the other hand, the mechanism behind the binding of highly structured cellulosomes to the cell surface has been suggested (Lemaire et al., 1995). First, one of the dockerins in the primary scaffoldin binds with cohesin at the cell surface. This cohesin molecule then connects either with the peptidoglycan-binding S-layer homology domain by noncovalent binding, or with the sortase motif by covalent binding (Lemaire et al., 1995).

Recently, it was revealed that some cellulosomes were not localized at the cell surface. Some highly structured cellulosomes produced by *C. thermocellum* and *Clostridium clariflavum* do not localize at the cell surface, but act as secreted protein complexes instead (Artzi et al., 2015; Xu et al., 2016). These cellulosomes form a complex with cellulosomal proteins without binding to the cell surface, and are known as “cell-free cellulosomes” (Figure 2b). In *C. cellulolyticum*, the enzymatic activities from cell-free cellulosomes have been determined (Mohand-Oussaid et al., 1999). Furthermore, it is known that cellulosomes at the cell surface detach from the surface at the late stationary phase in *C. thermocellum* (Bayer and Lamed, 1986).

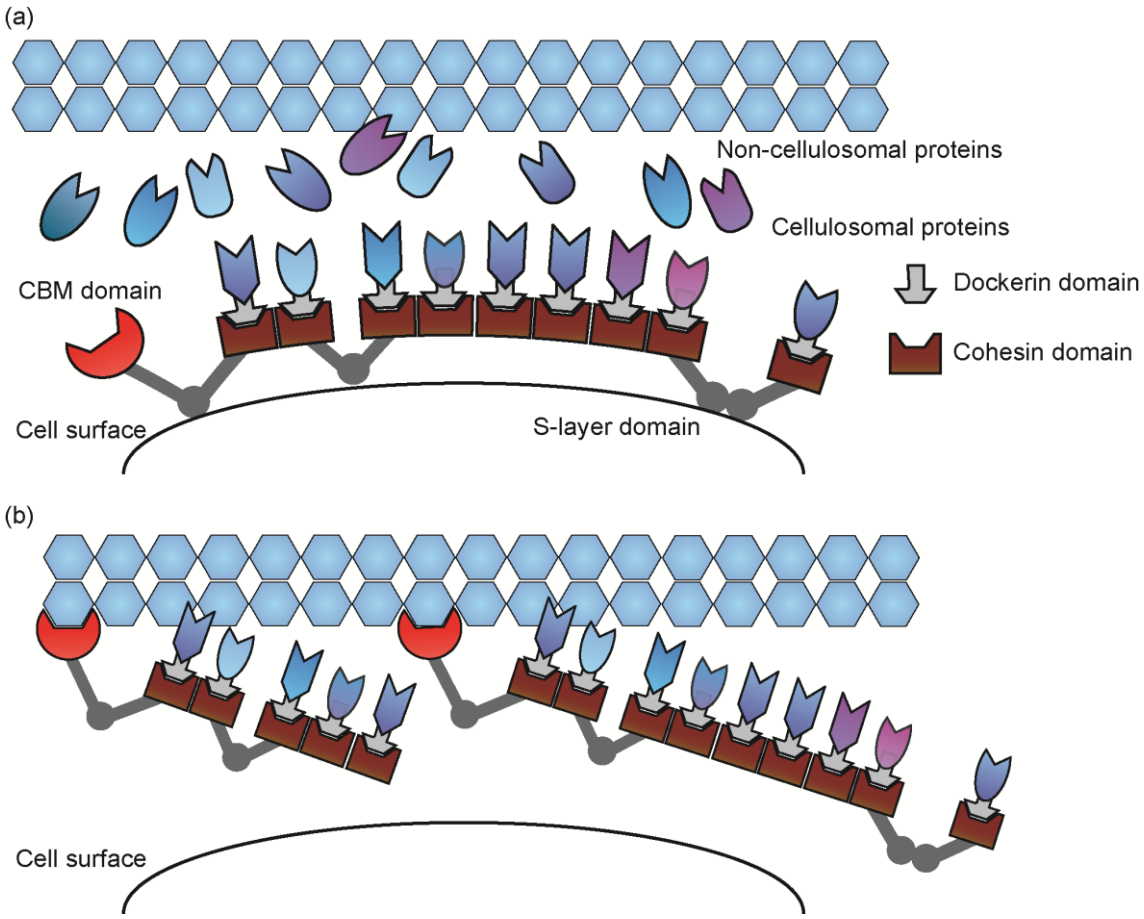


Figure 2 (a) Model of a simple cellulosome. (b) Model of a cell-free cellulosome. CBM: carbohydrate-binding module

Clostridium cellulovorans

C. cellulovorans, which was isolated from fermented wood biomass in 1984 (Sleat et al., 1984), is a mesophilic and anaerobic gram-positive bacterium (Figure 3). Recently, *C. cellulovorans* has garnered much attention as a cellulosome producer (Blair and Anderson, 1999; Sleat et al., 1984). It can degrade and metabolize the major plant cell-wall polysaccharides of lignocellulosic biomass, such as cellulose, hemicellulose, and pectin (Sleat et al., 1984). *C. cellulovorans* produces simple cellulosomes, the scaffoldin of which is constructed from nine cohesin domains (Blair and Anderson, 1999; Doi et al., 1994). Furthermore, the bacterium degrades plant cell-wall polysaccharides with both cellulosomal and non-cellulosomal proteins (Doi et al., 1998; Morisaka et al., 2012), the latter of which are secreted carbohydrate-degrading proteins without a dockerin domain (Doi et al., 1998).

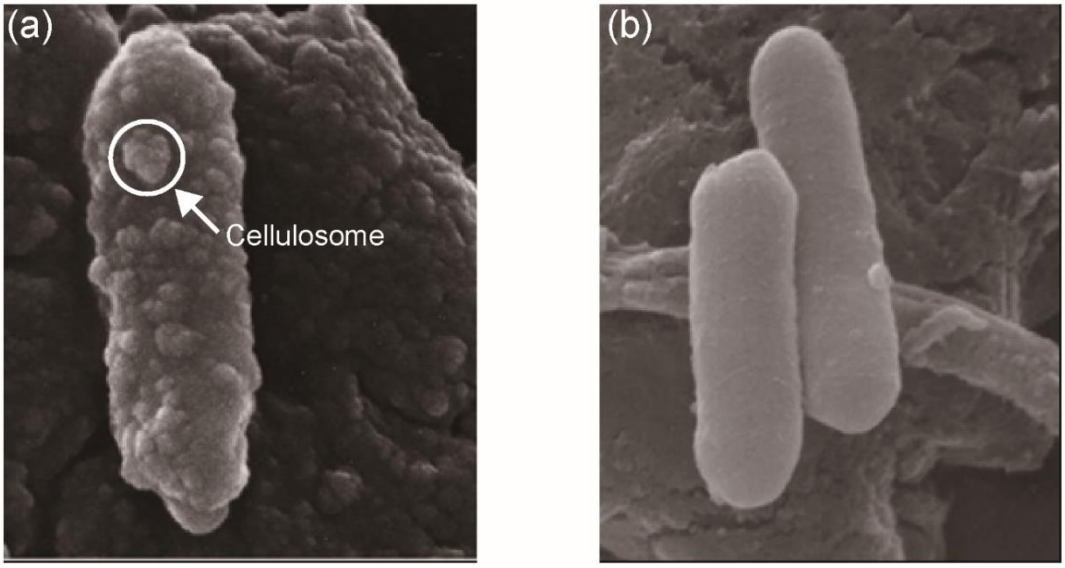


Figure 3 (a) Cell surface structure of *C. cellulovorans* grown on cellulose (b) Cell surface structure of *C. cellulovorans* grown on cellobiose (Modified from (Blair and Anderson, 1999))

In 2010, Tamaru *et al.* clarified the genome sequence of *C. cellulovorans*, offering new insight into the bacterium's efficient mechanism of lignocellulosic biomass degradation, (Tamaru et al., 2011; Tamaru et al., 2010). *C. cellulovorans* has 57 cellulosomal genes in its genome, which include those coding for 25 glycoside hydrolase (GH) family proteins, 2 carbohydrate esterase (CE) family proteins, and 4 polysaccharide lyase (PL) family proteins. In addition, the bacterium has 168 non-cellulosomal proteins, which comprise 89 GH family proteins, 19 CE family proteins, 9 PL family proteins, and 38 glycosyltransferase (GT) family proteins (Matsui et al., 2013). In *C. cellulovorans*, the ratio of cellulosomal GH and PL family proteins to non-cellulosomal GH and PL family proteins is lower than that of other cellulosome-producing *Clostridium* species, whereas the total number of GH and PL family proteins is higher (Table 2) (Tamaru et al., 2011; Tamaru et al., 2010). This implies that *C. cellulovorans* can degrade more types of plant cell-wall polysaccharides than other *Clostridium* species can, and emphasizes the importance of non-cellulosomal proteins for the degradation of plant cell-wall polysaccharides in this species. To reveal the mechanisms by which the cellulosomal and non-cellulosomal proteins degrade polysaccharides, we should analyze the changes of these proteins between different polysaccharides and time points.

Table 2 Number of cellulosomal and non-cellulosomal genes in cellulosome-producing *Clostridium* species (Tamaru et al., 2011)

Organism	Cellulosomal GHs + PLs	Non-cellulosomal GHs + PLs	Cellulosomal GHs + PLs/ Non-cellulosomal GHs + PLs
C. cellulovorans	29	98	0.30
C. cellulolyticum	47	42	1.1
C. thermocellum	53	14	3.8

GHs: Glycoside hydrolase family proteins; PLs: Polysaccharide lyase family proteins

Proteome analysis

Proteome analysis is a valuable tool for analyzing the dynamics of cellulosomal and non-cellulosomal proteins. For analyzing the role of cellulosomal proteins in cellulosome-producing *Clostridium* species, the enzymatic activities of the cellulosomal proteins have been analyzed; for example, EngK (Arai et al., 2006), PelA (Tamaru and Doi, 2001), and XynA (Kosugi et al., 2002) in *C. cellulovorans*. However, these analyses could not clarify the difference between the distinct compositions of cellulosomal and non-cellulosomal proteins. For such elucidation, we require a comprehensive and less-biased proteome analysis, including that of non-cellulosomal proteins.

Previously, the analysis of cellulosomal proteins was mainly performed using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (Blouzard et al., 2010; Cho et al., 2010). However, we could not detect a lot of proteins with 2D-PAGE because antibodies or cohesin markers were used for detecting cellulosomal proteins; hence, we could only identify cellulosomal proteins (Blouzard et al., 2010; Cho et al., 2010). Furthermore, previous studies had concentrated proteins with the carbohydrate-binding domain by phosphoric acid-swollen cellulose, and therefore proteome analysis of non-cellulosomal proteins could not be performed. The genome analysis of *C. cellulovorans* (Tamaru et al., 2010) enabled the non-targeted proteome analysis of the proteins secreted by this bacterium (Morisaka et al., 2012). Furthermore, ultrafiltration of the cell-free culture medium was applied for enrichment of the secreted proteins, thereby permitting us to perform a less-biased analysis of the cellulosomal and non-cellulosomal proteins (Morisaka et al., 2012). As a result, Matsui *et al*. showed that the variance of non-cellulosomal proteins has an important role in the degradation of xylan and pectin (Matsui et al., 2013). Ever since this study, however, proteome analysis of the non-cellulosomal proteins of other cellulosome-producing bacteria has not been carried out.

In *C. thermocellum* and *C. cellulolyticum*, proteome analysis of the cellular proteins using liquid chromatography–tandem mass spectrometry (LC-MS/MS) and transcriptome analysis has been performed for analyzing the cellular dynamics (Rydzak

et al., 2012; Xu et al., 2013). For example, *C. thermocellum* was cultured in cellobiose, and the differences in proteins for core metabolic functions between the different growth phases were clarified by proteome analysis of the cellular proteins using LC-MS/MS (Rydzak et al., 2012). In *C. cellulolyticum*, transcriptome analysis revealed the sensing mechanism of carbon sources in the culture medium, and the regulatory component for expressing the proteins involved in carbohydrate degradation and metabolism under different carbon sources (Xu et al., 2013). In contrast to these cellulosome-producing bacteria, neither proteome analysis of the cellular proteins nor transcriptome analysis has been conducted in *C. cellulovorans*, and identification of the proteins or genes involved in the mechanisms for the metabolism of various carbon sources, mechanisms for sensing carbon sources in the culture media, and regulatory mechanisms of the proteins involved in the metabolism and degradation of each carbon source have not been performed for this bacterium.

Optimization of LC-ESI-MS/MS in proteome analysis

High sensitivity and highly efficient separation are required for performing high-performance proteome analysis with LC-electrospray ionization (ESI)-MS/MS. Proteome samples have a high complexity and a broad dynamic range (Michalski et al., 2011). Therefore, the good separation and detection of a complex peptide mixture are important for proteome analysis. Recently, the development of mass spectrometry, as represented by Orbitrap mass spectrometry, has increased the number of peptides identified per minute by about tenfold (Eliuk and Makarov, 2015; Hebert et al., 2014). However, the separation power in the LC column has not increased in like. The high separation power in a particle-packed LC column is achieved by minimizing the diameter of the packed particle, but doing so increases the column back pressure (Giddings, 1991). Therefore, high separation efficiency in a particle-packed column is technically difficult to achieve beyond the upper-pressure limit of the LC instrument, necessitating new separation material for the LC column. To solve this problem, the monolithic column was developed as a new separation medium (Tanaka et al., 2002), enabling the preparation of a meter-long capillary column from which highly efficient

separation and sensitivity were obtained for proteome analysis (Iwasaki et al., 2010; Morisaka et al., 2012).

Optimization of the instruments is required for proteome analysis with a higher sensitivity. Although recent developments in instruments have enabled high sensitivity and highly efficient separation, we could still not perform highly sensitive proteome analysis without optimization of the instruments. Optimization has crucial effects; for example, in LC, the appropriate selection of a column and gradient method is necessary to maximize the number of peptides eluted at a single time (Vollmer et al., 2004). In ESI, an appropriate spray voltage is required for a high ionization efficiency and low background. In MS, several parameters should be appropriately set for identifying the eluting peptides (Vollmer et al., 2004). Similar to these examples, the parameters of LC, ESI, and MS affect the number of proteins identified and the peak area of MS1 (Andrews et al., 2011; Kalli and Hess, 2012). Hence, we attempted to optimize such parameters to identify larger numbers of proteins in non-targeted proteome analysis.

The design of experiments (DoE) could be a powerful tool for optimizing the combination of LC, ESI, and MS parameters. In several studies, the optimization of LC or MS has been performed by changing the parameters one by one (Kalli and Hess, 2012; Xu et al., 2009). In fact, although these studies optimized LC or MS parameters, such optimization procedure could not detect two-factor interactions, which can be important for determining an optimal parameter set. To solve this problem, a recent study used DoE for optimizing MS parameters for non-targeted proteome analysis (Andrews et al., 2011). That study adopted a fractional factorial design, which could detect two-factor interactions from approximately three parameters. Using the fractional factorial design, MS parameters were systematically optimized and the number of proteins identified was increased. However, LC parameters and two-factor interactions between LC and MS were not analyzed. For optimizing LC, ESI, and MS parameters simultaneously, other statistical tools are required. One such new statistical tool, “definitive screening design”, was recently developed (Jones and Nachtsheim, 2011). This tool might enable the identification of main factors and two-factor interactions from several parameters with less experiments.

Cellulosome-producing *Clostridium* species

(a) Sensing of carbon sources and regulation of carbohydrate-degrading enzymes in cellulosome-producing *Clostridium* species

Cellulosome-producing *Clostridium* species change their expression pattern of carbohydrate-degrading enzymes depending on the polysaccharides present in the culture medium. Therefore, some studies have attempted to reveal the mechanisms responsible for sensing the carbon sources in the culture media and for the changing expression pattern of carbohydrate-degrading enzymes. In *C. thermocellum*, sigma factor and anti-sigma factor are used for changing the expression pattern of carbohydrate-degrading enzymes (Figure 4a) (Nataf et al., 2010). Sigma factors are released in the cell when polysaccharides in the culture medium bind to the carbohydrate-binding module domain. Then, the sigma factors interact with RNA polymerase and promote the transcription of carbohydrate-degrading enzymes. In *C. cellulolyticum*, a two-component regulatory system is used for sensing polysaccharides in the culture medium and regulating the carbohydrate-degrading enzymes (Figure 4b) (Xu et al., 2013). In the two-component regulatory system of this bacterial species, hydrolysates of polysaccharides in the culture medium bind to sugar-binding proteins, thereby activating histidine kinase and expression of the carbohydrate-degrading enzymes.

In contrast to these cellulosome-producing bacteria, the mechanisms for sensing polysaccharides in the culture medium and for changing the expression pattern of carbohydrate-degrading enzymes have not been clarified in *C. cellulovoran*s. This bacterium is known to degrade and metabolize more types of carbon sources than the other cellulosome-producing bacteria do. Therefore, the elucidation of these mechanisms in *C. cellulovorans* is an attractive research subject.

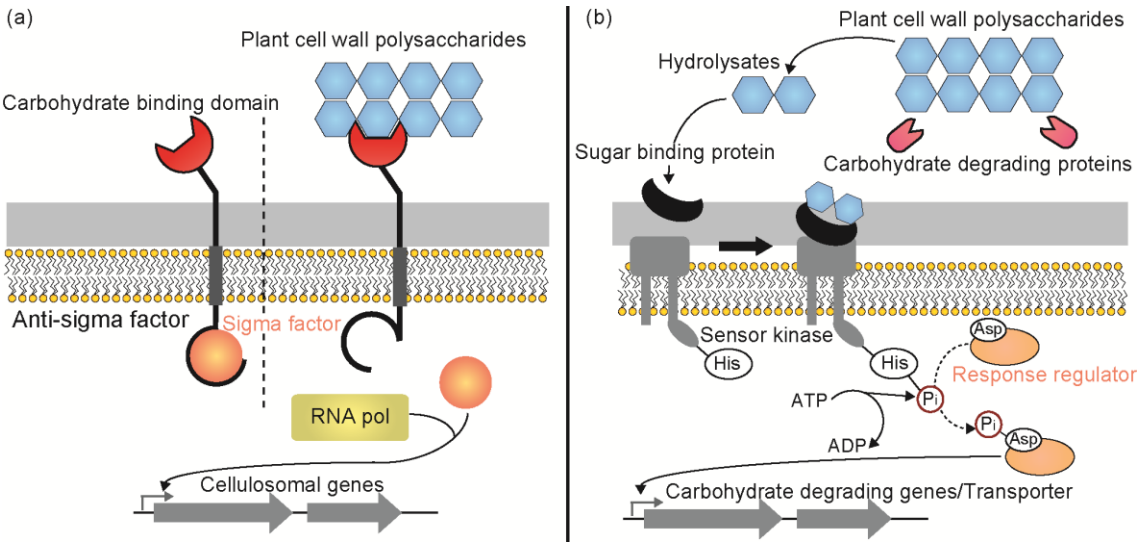


Figure 4 Plant cell-wall polysaccharide-sensing model in cellulosome-producing *Clostridium* species. (a) Plant cell-wall polysaccharide-sensing model in *C. thermocellum*. (b) Plant cell-wall polysaccharide-sensing model in *C. cellulolyticum*

(b) Degradation and metabolism of carbon sources in cellulosome-producing *Clostridium* species

Cellulosome-producing *Clostridium* species can degrade most of the major plant cell-wall polysaccharides. However, each bacterium has different types of carbohydrate-degrading enzymes and metabolizes different carbon sources (Table 3). For example, *C. thermocellum* grows well on cellooligosaccharides, ranging from cellobiose to cellohexaose (Zhang and Lynd, 2005), but cannot metabolize xylan and other major plant cell-wall polysaccharides (Ng et al., 1977). *C. thermocellum* utilizes highly structured cellulosomes for degrading plant cell-wall polysaccharides, but this bacterium does not have many genes coding for non-cellulosomal proteins (Tamaru et al., 2011).

C. cellulovorans shows a distinct system for degrading and metabolizing major plant cell-wall polysaccharides compared with the other cellulosome-producing *Clostridium* species. *C. cellulovorans* can metabolize cellulose, hemicellulose, and pectin (Sleat et al., 1984). Furthermore, it encodes 97 non-cellulosomal proteins in its genome, and degrades the plant cell wall efficiently through a combination of

cellulosomal and non-cellulosomal proteins (Matsui et al., 2013; Tamaru et al., 2011). However, the metabolic pathways of each polysaccharide in this bacterial species are not well understood on the protein level, and the temporal changes of carbohydrate-degrading enzymes and proteins related to polysaccharide metabolism have not been analyzed. Therefore, the mechanisms of polysaccharide degradation in *C. cellulovorans* are not well understood.

Table 3 Degradation of carbon sources or growth in carbon sources in cellulosome-producing *Clostridium* species

Clostridium species	Function	Cellulose	Xylan	Galactomannan	Pectin
C. cellulovorans	Degradation	○	○	○	○
	Growth	○	○	○	○
C. thermocellum	Degradation	○	○	○	○
	Growth	○	×	×	×
C. cellulolyticum	Degradation	○	○	○	○
	Growth	○	○	△	×

(○: Ability for full degradation of each polysaccharide or high growth; △: Ability for partial degradation or little growth; ×: No ability for degradation or no growth)

In the research conducted for this thesis, after improving the coverage of proteome analysis with DoE, the *C. cellulovorans* model of sensing polysaccharides in the culture medium and regulating the carbohydrate-degrading enzymes could be clarified. Furthermore, temporal cellular and secreted proteome analyses should be performed simultaneously, and the dynamics of the proteins involved in carbohydrate degradation and metabolism in *C. cellulovorans* should be also clarified.

References

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A., and Levin, D.B. (2011). Biomass pretreatment: fundamentals toward application. Biotechnol Adv *29*, 675-685.

Andrews, G.L., Dean, R.A., Hawkridge, A.M., and Muddiman, D.C. (2011). Improving proteome coverage on a LTQ-Orbitrap using design of experiments. J Am Soc Mass Spectrom *22*, 773-783.

Arai, T., Kosugi, A., Chan, H., Koukiekolo, R., Yukawa, H., Inui, M., and Doi, R.H. (2006). Properties of cellulosomal family 9 cellulases from *Clostridium cellulovorans*. Appl Microbiol Biotechnol *71*, 654-660.

Artzi, L., Bayer, E.A., and Moraïs, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol *15*, 83.

Artzi, L., Morag, E., Barak, Y., Lamed, R., and Bayer, E.A. (2015). *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. MBio *6*, e00411-00415.

Bayer, E., Setter, E., and Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. J Bacteriol *163*, 552-559.

Bayer, E.A., and Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J Bacteriol *167*, 828-836.

Bayer, E.A., Shimon, L.J.W., Shoham, Y., and Lamed, R. (1998). Cellulosomes—Structure and ultrastructure. J Struct Biol *124*, 221-234.

Bjerre, A.B., Olesen, A.B., Fernqvist, T., Plöger, A., and Schmidt, A.S. (1996). Pretreatment of wheat straw using combined wet oxidation and alkaline hydrolysis resulting in convertible cellulose and hemicellulose. Biotechnol Bioeng *49*, 568-577.

Blair, B.G., and Anderson, K.L. (1999). Regulation of cellulose-inducible structures of *Clostridium cellulovorans*. Can J Microbiol *45*, 242-249.

Blouzard, J.C., Coutinho, P.M., Fierobe, H.P., Henrissat, B., Lignon, S., Tardif, C., Pagès, S., and de Philip, P. (2010). Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. Proteomics *10*, 541-554.

Cann, I., Bernardi, R.C., and Mackie, R.I. (2016). Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. Environ Microbiol *18*, 307-310.

Chen, W.-H., Pen, B.-L., Yu, C.-T., and Hwang, W.-S. (2011). Pretreatment efficiency and structural characterization of rice straw by an integrated process of dilute-acid and steam explosion for bioethanol production. Bioresour Technol *102*, 2916-2924.

Cho, W., Jeon, S.D., Shim, H.J., Doi, R.H., and Han, S.O. (2010). Cellulosomic profiling produced by *Clostridium cellulovorans* during growth on different carbon sources explored by the cohesin marker. J Biotechnol *145*, 233-239.

Cosgrove, D.J. (2005). Growth of the plant cell wall. Nat Rev Mol Cell Biol *6*, 850-861.

Cumming, C.M., Rizkallah, H.D., McKendrick, K.A., Abdel-Massih, R.M., Baydoun, E.A., and Brett, C.T. (2005). Biosynthesis and cell-wall deposition of a pectin–xyloglucan complex in pea. Planta *222*, 546-555.

Doi, R.H., Goldstein, M., Hashida, S., Park, J.-S., and Takagi, M. (1994). The *Clostridium cellulovorans* cellulosome. Crit Rev Microbiol *20*, 87-93.

Doi, R.H., Park, J.-S., Liu, C.-c., Malburg Jr, L.M., Tamaru, Y., Ichiishi, A., and Ibrahim, A. (1998). Cellulosome and noncellulosomal cellulases of *Clostridium cellulovorans*. Extremophiles *2*, 53-60.

Doi, R.H., and Tamaru, Y. (2001). The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. Chem Rec *1*, 24-32.

Eliuk, S., and Makarov, A. (2015). Evolution of Orbitrap mass spectrometry instrumentation. Annu Rev Anal Chem *8*, 61-80.

Gal, L., Pages, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C., and Belaich, J.-P. (1997). Characterization of the cellulolytic complex (cellulosome) produced by *Clostridium cellulolyticum*. Appl Environ Microbiol *63*, 903-909.

Garrote, G., Falqué, E., Domínguez, H., and Parajó, J.C. (2007). Autohydrolysis of agricultural residues: study of reaction byproducts. Bioresour Technol *98*, 1951-1957.

Giddings, J.C. (1991). Unified separation science (Wiley New York, NY).

Girisuta, B., Dussan, K., Haverty, D., Leahy, J.J., and Hayes, M.H.B. (2013). A kinetic study of acid catalysed hydrolysis of sugar cane bagasse to levulinic acid. Chem Eng J

217, 61-70.

Groom, M.J., Gray, E.M., and Townsend, P.A. (2008). Biofuels and biodiversity: principles for creating better policies for biofuel production. Conserv Biol *22*, 602-609.

Hayashi, T. (1989). Xyloglucans in the primary cell wall. Annu Rev Plant Biol *40*, 139-168.

Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014). The one hour yeast proteome. Mol Cell Proteomics *13*, 339-347.

Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., and Ishihama, Y. (2010). One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. Anal Chem *82*, 2616-2620.

Jones, B., and Nachtsheim, C.J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. J Qual Technol *43*, 1.

Kalli, A., and Hess, S. (2012). Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. Proteomics *12*, 21-31.

Kobayashi, M., Matoh, T., and Azuma, J. (1996). Two chains of rhamnogalacturonan II are cross-linked by borate-diol ester bonds in higher plant cell walls. Plant Physiol *110*, 1017-1020.

Kosugi, A., Murashima, K., and Doi, R.H. (2002). Xylanase and acetyl xylan esterase activities of XynA, a key subunit of the *Clostridium cellulovorans* cellulosome for xylan degradation. Appl Environ Microbiol *68*, 6399-6402.

Kristensen, J.B., Thygesen, L.G., Felby, C., Jørgensen, H., and Elder, T. (2008). Cell-wall structural changes in wheat straw pretreated for bioethanol production. Biotechnol Biofuels *1*, 5.

Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. J Biotechnol *56*, 1-24.

Lemaire, M., Ohayon, H., Gounon, P., Fujino, T., and Béguin, P. (1995). OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol *177*, 2451-2459.

Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H.V., Auer, M., Vogel, K.P., Simmons, B.A., and Singh, S. (2010a). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. Bioresour Technol *101*, 4900-4906.

Li, M.-F., Fan, Y.-M., Xu, F., Sun, R.-C., and Zhang, X.-L. (2010b). Cold sodium hydroxide/urea based pretreatment of bamboo for bioethanol production: Characterization of the cellulose rich fraction. Ind Crops Prod *32*, 551-559.

Li, X., Kim, T.H., and Nghiem, N.P. (2010c). Bioethanol production from corn stover using aqueous ammonia pretreatment and two-phase simultaneous saccharification and fermentation (TPSSF). Bioresour Technol *101*, 5910-5916.

Lynd, L.R., Wyman, C.E., and Gerngross, T.U. (1999). Biocommodity engineering. Biotechnol Prog *15*, 777-793.

Lynd, L.R., Zyl, W.H.v., McBride, J.E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol *16*, 577-583.

Matsui, K., Bae, J., Esaka, K., Morisaka, H., Kuroda, K., and Ueda, M. (2013). Exoproteome profiles of *Clostridium cellulovorans* grown on various carbon sources. Appl Environ Microbiol *79*, 6576-6584.

Menon, V., and Rao, M. (2012). Trends in bioconversion of lignocellulose: biofuels, platform chemicals & biorefinery concept. Prog Energy Combust Sci *38*, 522-550.

Michalski, A., Cox, J., and Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC−MS/MS. J Proteome Res *10*, 1785-1793.

Mohand-Oussaid, O., Payot, S., Guedon, E., Gelhaye, E., Youyou, A., and Petitdemange, H. (1999). The extracellular xylan degradative system in *Clostridium cellulolyticum* cultivated on xylan: evidence for cell-free cellulosome production. J Bacteriol *181*, 4035-4040.

Mohnen, D. (2008). Pectin structure and biosynthesis. Curr Opin Plant Biol *11*, 266-277.

Mood, S.H., Golfeshan, A.H., Tabatabaei, M., Jouzani, G.S., Najafi, G.H., Gholami, M., and Ardjmand, M. (2013). Lignocellulosic biomass to bioethanol, a comprehensive

review with a focus on pretreatment. Renew Sust Energ Rev *27*, 77-93.

Morisaka, H., Matsui, K., Tatsukami, Y., Kuroda, K., Miyake, H., Tamaru, Y., and Ueda, M. (2012). Profile of native cellulosomal proteins of *Clostridium cellulovorans* adapted to various carbon sources. AMB Express *2*, 37.

Murashima, K., Kosugi, A., and Doi, R.H. (2003). Synergistic effects of cellulosomal xylanase and cellulases from *Clostridium cellulovorans* on plant cell wall degradation. J Bacteriol *185*, 1518-1524.

Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E.A., Sonenshein, A.L., and Shoham, Y. (2010). *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci U S A *107*, 18646-18651.

Ng, T., Weimer, P., and Zeikus, J. (1977). Cellulolytic and physiological properties of *Clostridium thermocellum*. Archives Microbiol *114*, 1-7.

O'Neill, M., Albersheim, P., and Darvill, A. (1990). 12 - The pectic polysaccharides of primary cell walls. In Methods in Plant Biochemistry, P.M. Dey, ed. (Academic Press), pp. 415-441.

Ohara, H., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2000). Characterization of the cellulolytic complex (cellulosome) from *Ruminococcus albus*. Biosci Biotechnol Biochem *64*, 254-260.

Ridley, B.L., O'Neill, M.A., and Mohnen, D. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry *57*, 929-967.

Rydzak, T., McQueen, P.D., Krokhin, O.V., Spicer, V., Ezzati, P., Dwivedi, R.C., Shamshurin, D., Levin, D.B., Wilkins, J.A., and Sparling, R. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiol *12*, 214.

Saha, B.C. (2003). Hemicellulose bioconversion. J Ind Microbiol Biotechnol *30*, 279-291.

Shoham, Y., Lamed, R., and Bayer, E.A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol *7*, 275-281.

Sleat, R., Mah, R.A., and Robinson, R. (1984). Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. Appl Environ Microbiol *48*, 88-93.

Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour Technol *83*, 1-11.

Tamaru, Y., and Doi, R.H. (2001). Pectate lyase A, an enzymatic subunit of the *Clostridium cellulovorans* cellulosome. Proc Natl Acad Sci U S A *98*, 4125-4129.

Tamaru, Y., Miyake, H., Kuroda, K., Nakanishi, A., Matsushima, C., Doi, R.H., and Ueda, M. (2011). Comparison of the mesophilic cellulosome-producing *Clostridium cellulovorans* genome with other cellulosome-related clostridial genomes. Microb Biotechnol *4*, 64-73.

Tamaru, Y., Miyake, H., Kuroda, K., Ueda, M., and Doi, R.H. (2010). Comparative genomics of the mesophilic cellulosome-producing *Clostridium cellulovorans* and its application to biofuel production via consolidated bioprocessing. Environ Technol *31*, 889-903.

Tanaka, N., Kobayashi, H., Ishizuka, N., Minakuchi, H., Nakanishi, K., Hosoya, K., and Ikegami, T. (2002). Monolithic silica columns for high-efficiency chromatographic separations. J Chromatogr A *965*, 35-49.

Vincken, J.-P., Schols, H.A., Oomen, R.J., McCann, M.C., Ulvskov, P., Voragen, A.G., and Visser, R.G. (2003). If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. Plant Physiol *132*, 1781-1789.

Vollmer, M., Hörth, P., and Nägele, E. (2004). Optimization of two-dimensional off-line LC/MS separations to improve resolution of complex proteomic samples. Anal Chem *76*, 5180-5185.

Willats, W.G., McCartney, L., Mackie, W., and Knox, J.P. (2001). Pectin: cell biology and prospects for functional analysis. Plant Mol Biol *47*, 9-27.

Xu, C., Huang, R., Teng, L., Wang, D., Hemme, C.L., Borovok, I., He, Q., Lamed, R., Bayer, E.A., and Zhou, J. (2013). Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. Biotechnol Biofuels *6*, 73.

Xu, P., Duong, D.M., and Peng, J. (2009). Systematical optimization of reverse-phase

chromatography for shotgun proteomics. J Proteome Res *8*, 3944-3950.

Xu, Q., Gao, W., Ding, S.-Y., Kenig, R., Shoham, Y., Bayer, E.A., and Lamed, R. (2003). The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol *185*, 4548-4557.

Xu, Q., Resch, M.G., Podkaminer, K., Yang, S., Baker, J.O., Donohoe, B.S., Wilson, C., Klingeman, D.M., Olson, D.G., and Decker, S.R. (2016). Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Sci Adv *2*, e1501254.

Zhang, Y.-H.P., and Lynd, L.R. (2005). Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc Natl Acad Sci U S A *102*, 7321-7325.

Chapter I

Definitive screening design enables optimization of LC–ESI–MS/MS parameters in proteomics

Proteome samples have a high complexity and broad dynamic range (Beck et al., 2011; Nagaraj et al., 2011; Taniguchi et al., 2010). Therefore, the optimization of each parameter of LC-ESI-MS/MS is essential for exerting the best performance to identify larger numbers of proteins (Andrews et al., 2011). The Design of experiments (DoE) could be a powerful tool to optimize combination of LC, ESI and MS parameters. However, previous DoE method is not suitable for optimizing many combination of LC, ESI and MS parameters, and I could not optimize LC, ESI and MS parameters at the same time. Therefore, a novel DoE method is required for optimization of LC, ESI and MS parameters.

In chapter I, I aimed to systematically optimize LC, ESI and MS parameters using a novel DoE method. I adopted a definitive screening design to optimize monolithic capillary LC-ESI-MS/MS parameters for the systematic optimization of all LC, ESI, and MS parameters. The definitive screening design was recently developed by Jones et al. (Jones and Nachtsheim, 2011), and it enables the identification of main factors and two-factor interactions from several parameters with less experiments. Here I report the optimization of monolithic capillary LC-ESI-MS/MS parameters using the definitive screening design.

Materials and methods

Sample preparation

Human whole K562 cell lysate was obtained from an MS compatible human protein extract, Intact (Promega, WI, USA). Reduction and alkylation were performed as described in the instruction manual, with a slight modification. Samples (400 µL of 2.5 µg/µL human protein extract) were reduced with 20 mM dithiothreitol (Nacalai Tesque,

Kyoto, Japan) and incubated at 37°C for 30 min. Further, the reduced proteins were alkylated with 60 mM iodoacetamide (Wako, Osaka, Japan) and incubated at room temperature for 30 min in the dark. Following the addition of 500 µL of 50 mM triethylammonium bicarbonate (TEAB) (Sigma-Aldrich, MO, USA), the proteins were digested to peptides by 20 µg sequencing grade trypsin (Promega) at 37°C overnight. The protein concentration was adjusted to 1 µg/µL by the addition of 400 µL TEAB. The samples were desalted using MonoSpin C18 (GL Science, Tokyo, Japan) prior to analysis.

Monolithic capillary LC-ESI-MS/MS analysis

The parameters that were not considered as variables are shown in this section. All analyses were performed using an LC (Ultimate 3000®; Thermo Fisher Scientific Waltham, USA) equipped with a long monolithic silica capillary column (4,200 mm, ID 0.1 mm), and ESI (XYZ-stage; AMR, Tokyo, Japan)-MS/MS (LTQ Orbitrap Velos Mass Spectrometer®; Thermo Fisher Scientific) system. A meter-scale monolithic silica capillary column has a higher separation capacity than a particle-packed column (Morisaka et al., 2012; Yamana et al., 2013) . A gradient was achieved by changing the ratio of the following two eluents: eluent A, 0.1% (v/v) formic acid (Wako); eluent B, 80% acetonitrile (Wako) containing 0.1% (v/v) formic acid. The separated peptides were directly electrosprayed into the mass spectrometer using a Fortis tip (AMR).

The LTQ Orbitrap Velos was operated in a data-dependent analysis mode with a full scan range of 350–1,500 *m/z* in the orbitrap and subsequent CID MS/MS scans in the linear ion trap. CID was performed at a normalized collision energy of 35%. Blank analyses were included between different settings.

Data analysis

Data analysis was performed using Proteome Discoverer 2.1 (Thermo Fisher Scientific). Protein identification was performed using the Mascot algorithm against the *Homo sapiens* protein database (20,172 sequences, version 2015-11-11) from UniProt, with a precursor mass tolerance of 100 ppm and a fragment ion mass tolerance of 0.8 Da. Carbamidomethylation of cysteine was set as a fixed modification, and the oxidation of

methionine and acetylation of protein N-terminus were set as dynamic modifications. The data were then filtered with a cut-off criteria of *q*-value ≤0.01, corresponding to a 1% false discovery rate (FDR) on a spectral level. Proteins that had more than one unique peptide were accepted as identified proteins.

In calculating the peak area, I used PSMs based on Proteome Discoverer 2.1 (Lemeer et al., 2012). I used the Top 3 method for calculating the peak area (Silva et al., 2006). In this method, peptides with the three highest signal intensities for each protein were summed, and the summed value was used as a peak area of MS1. The expected mass deviation was set as 4 ppm when the same peaks were repeatedly measured in consecutive scans.

Selection of parameters to be optimized

In previous papers, the number of identified proteins has been used as a response variable to optimize non-targeted proteomic analysis (Andrews et al., 2011; Kalli and Hess, 2012; Xu et al., 2009). In this study, the average peak area of MS1 was used as a response variable in addition to the number of identified proteins. Because the peak area of MS1 is used for label-free quantitative proteome analysis, I thought that maximization of this value is required for high sensitivity analysis in label-free quantitative proteome analysis. To screen the possible parameters affecting the number of identified proteins and the average peak area of MS1, I chose 14 parameters as explanatory variables. These parameters were selected by referring to previous papers (Kalli and Hess, 2012; Kalli et al., 2013; Xu et al., 2009), which optimized LC or MS instruments. In LC parameters, the linear velocity, gradient, injection amount, and gradient time were chosen as explanatory variables. In ESI parameters, the ESI spray voltage was chosen. In MS parameters, the maximum ion injection time (IT) of MS1/MS2, target value of auto gain control (AGC) in MS1/MS2, dynamic exclusion, resolution, number of MS2 scans, minimum count threshold, and capillary temperature were chosen. AGC is the upper limit of ions introduced into MS. IT is the higher limit of time for accumulating ions in MS. Dynamic exclusion is the time interval between selecting the same *m/z* for isolation and fragmentation. Minimum count threshold is the lower limit of ion intensity applied to

MS2. I selected standard parameter from previous proteome analysis using monolithic capillary LC-ESI-MS/MS [11].

Definitive screening design

The definitive screening design was adopted using JMP 11(SAS Institute Inc., Cary, NC.) (Jones and Nachtsheim, 2011). The number of identified proteins and average peak area of MS1 were used to generate the outcomes of the definitive screening design. The sequence of analysis was randomized to reduce the experimental error caused by the consecutive sequence of similar settings. After obtaining the analysis results, data analysis was performed. I initially explored the effects of each explanatory variable and two-factor interactions by minimizing Akaike's information criterion with a correction for finite sample sizes (AICc) (Hurvich and Tsai, 1989) in a stepwise manner. By applying AICc for model selection, I could balance the complexity of the model and model fitness. Further, effects of the selected parameters were calculated using multiple regression analysis, and explanatory variables with significantly affecting effects (p-value < 0.05) were identified.

Results and discussion

Experimental design

To prepare an experimental design for optimizing monolithic capillary LC, ESI and MS parameters, I adopted the definitive screening design based on JMP 11 (Jones and Nachtsheim, 2011). The settings used are shown in Table 1. As response variables, the number of identified proteins and average peak area of MS1 were chosen. The number of identified proteins has been generally used as a response variable to optimize non-targeted proteomic analysis in previous papers (Andrews et al., 2011; Kalli and Hess, 2012; Kalli et al., 2013). In contrast, the average peak area of MS1 has not been utilized as a response variable in non-targeted proteomics (Andrews et al., 2011; Kalli and Hess, 2012; Kalli et al., 2013). However, it is considered that optimization of the average peak area of MS1 is needed for current non-targeted proteomics. Recently, label-free quantification has been mainly employed for quantification in non-targeted proteomics (Cox et al., 2014; Feltcher et al., 2015; Silva et al., 2006). In label-free quantification of protein abundance, MS peak intensities are summed, as described in the Materials and methods. Therefore, I tried to maximize the average peak area of MS1 to minimize ion suppression and reduce experimental error between analyses.

Table 1 Used parameters and values

	Parameters	Low level	Median level	High level
1	Linear velocity [mm/s]	0.8	1.0	1.2
2	Gradient [%]	5–30	5–45	5–60
3	Injection amount [μg]	1	3	5
4	Gradient time [min]	450	600	750
5	ESI spray voltage [kV]	1.9	2.1	2.3
6	Capillary temperature [°C]	260	280	300
7	Resolution	30000	-	60000
8	AGC in MS1	1.0×10^6	5.5×10^6	1.0×10^7
9	AGC in MS2	5.0×10^3	2.8×10^4	5.0×10^4
10	IT of MS1 [ms]	500	750	1000
11	IT of MS2 [ms]	50	75	100
12	Dynamic exclusion [s]	30	90	150
13	Minimum count threshold	500	750	1000
14	Number of MS2 scans	10	15	20

AGC, target value of auto gain control; IT, maximum ion injection time

Fourteen parameters in Table 1 were used to optimize combination of monolithic capillary LC, ESI, and MS to improve the number of identified proteins and average peak intensity. Among these parameters, resolution was considered as a categorical variable and other parameters were considered as continuous variables. In continuous variables, three levels (low, median, and high) were used, and in categorical variables, two levels (low and high) were used. Using these parameters, I adopted the definitive screening design for preparing the experimental design.

Parameters affecting response variables

I conducted all analyses, and the number of identified proteins and average peak area of MS1 of each setting are shown in Table 2.

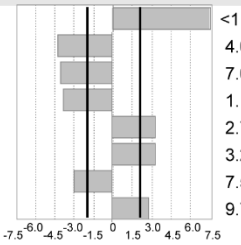
Table 2 Number of identified proteins and average peak area of MS1

Injection order	Number of identified proteins	Average peak area of MS1	Injection order	Number of identified proteins	Average peak area of MS1	Injection order	Number of identified proteins	Average peak area of MS1
1	2105	5.00×10^5	11	2896	1.16×10^6	21	1885	1.40×10^6
2	3159	1.08×10^6	12	2376	2.77×10^6	22	2344	3.45×10^6
3	2239	3.62×10^5	13	2271	1.73×10^6	23	1528	4.65×10^6
4	3403	7.31×10^5	14	2385	1.27×10^6	24	3400	2.41×10^6
5	3377	6.47×10^5	15	2256	8.19×10^5	25	3096	9.08×10^5
6	2981	8.37×10^5	16	2411	9.49×10^5	26	2778	4.09×10^6
7	2735	8.00×10^5	17	1925	7.88×10^5	27	3210	2.56×10^6
8	2977	5.86×10^5	18	1807	1.33×10^6	28	2789	2.40×10^6
9	3680	2.27×10^6	19	2038	1.01×10^6	29	2426	2.54×10^6
10	2942	6.56×10^5	20	2143	1.24×10^6	30	2147	1.30×10^6

Along with the procedures mentioned in the Materials and methods, I explored explanatory variables and two-factor interactions, which affected response variables, by minimizing AICc in a stepwise manner. Following this, I calculated effects of the selected parameters using multiple regression analysis and acquired explanatory variables that significantly affected the response variables (Figure 1). Furthermore, two-factor interactions between the explanatory variables selected in a stepwise manner are shown (Figure 2). In Figure 2, two-factor interaction plots with maximum response variables are shown.

(a)

Parameters	Coefficient of regression	S.E.	*t*-value	Bar chart of *t*-value	*p*-value
Gradient [%]	3.16×10^{2}	41.3	7.67		$<1.00 \times 10^{-4}$**
AGC of MS2	-1.77×10^{2}	41.3	-4.30		4.00×10^{-4}**
Resolution	-1.58×10^{2}	38.9	-4.08		7.00×10^{-4}**
Ion injection time of MS1 [ms]	-1.60×10^{2}	41.3	-3.90		1.10×10^{-3}**
Dynamic exclusion [s]	1.43×10^{2}	41.3	3.47		2.70×10^{-3}**
Gradient [%] - Dynamic exclusion [s]	1.66×10^{2}	49.0	3.40		3.20×10^{-3}**
Gradient time [min] - Resolution	-1.36×10^{2}	45.3	-3.01		7.50×10^{-3}**
Ion injection time of MS2 [ms] - AGC of MS2	1.46×10^{2}	50.6	2.89		9.70×10^{-3}**



(b)

Parameters	Coefficient of regression	S.E.	*t*-value	Bar chart of *t*-value	*p*-value
AGC of MS1	-5.36×10^{5}	7.99×10^{3}	-67.16		9.50×10^{-3}**
Dynamic exclusion [s]	-3.40×10^{5}	7.99×10^{3}	-42.64		1.49×10^{-2}*
Injection amount [μL]	3.18×10^{5}	7.99×10^{3}	39.88		1.60×10^{-2}*
Linear velosity [mm/s]	-2.59×10^{5}	7.99×10^{3}	-32.42		1.96×10^{-2}*
Ion injection time of MS1 [ms]	-2.26×10^{5}	7.99×10^{3}	-28.37		2.24×10^{-2}*
Number of MS2 scans	1.56×10^{5}	7.99×10^{3}	19.55		3.25×10^{-2}*
Ion injection time of MS2 [ms]	1.43×10^{5}	7.99×10^{3}	17.98		3.54×10^{-2}*
Ion injection time of MS2 [ms] - Minimum count threshold	2.83×10^{5}	2.18×10^{4}	13.00		4.89×10^{-2}*

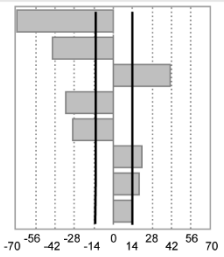


Figure 1 Effects of parameters on (a) the number of identified proteins and (b) average peak area of MS1.

Hyphen (Parameter A-Parameter B) indicates two-factor interactions between parameters A and B. Single asterisk (*) indicates *p*-value < 0.05, and double asterisk (**) indicates *p*-value < 0.01. A positive *t*-value indicates that variation is observed at a high level of the parameter, and a negative *t*-value indicates that variation is observed at a low level of the parameter. Black bold line indicates a significant *t*-value. *p*-value was calculated with the *t*-test. S.E. indicates standard error, AGC indicates the target value of auto gain control, and IT indicates the maximum ion injection time.

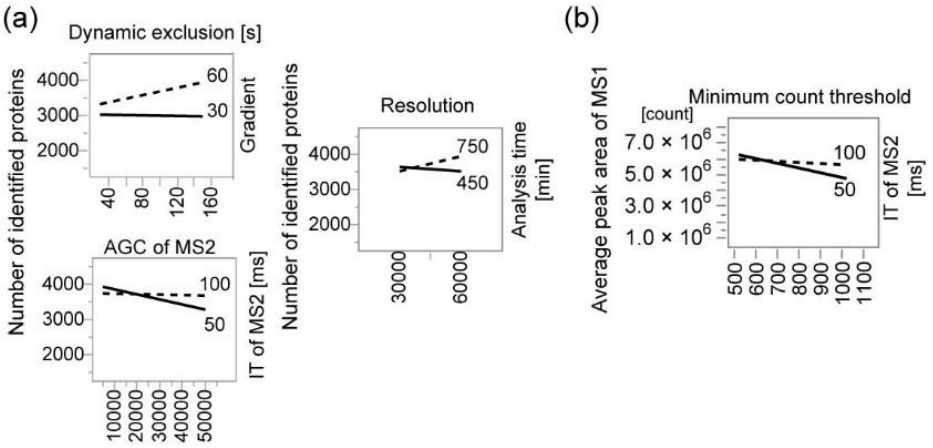


Figure 2 Two-factor interaction plots of (a) the number of identified proteins and (b) average peak area of MS1.

Dotted line shows a larger setting value of the parameter shown on the right side of the cell, and solid line shows a smaller setting value. It implies that how changes number of identified proteins and average peak area of MS1 when LC-ESI-MS/MS setting changes. AGC indicates the target value of auto gain control and IT indicates the maximum ion injection time.

Contribution of possible effects of the parameters on increase in number of identified proteins

When the number of identified proteins was selected as a response variable, five main effects (gradient, AGC of MS2, resolution, ion injection time of MS1, and dynamic exclusion) and three two-factor interactions (gradient-dynamic exclusion, gradient time-resolution, and ion injection time of MS2-AGC of MS2) significantly affected the response variable.

"Gradient" had the most significant effect on the number of identified proteins. Larger numbers of proteins were identified in a 5%–60% gradient than in other gradients. As shown in the previous report, the peak capacity became higher in the 5%–60% gradient (Horie et al., 2012; Köcher et al., 2011) and more peptides eluted per unit time. Furthermore, high concentrations of acetonitrile increased the ionization efficiency (Horie et al., 2014). These previous observations explained the reason why a high level of gradient increased the number of identified proteins.

“Smaller AGC of MS2” significantly increased the number of identified proteins. No difference in the quality of the spectrum should be observed between AGCs at 5000 and 50000. Therefore, I could detect more peptides with faster scan speeds of AGC of MS2 at 5000, as described previously (Kalli et al., 2013).

Further, a “higher resolution” significantly increased the number of identified proteins. A high separation capacity of monolithic columns may not be sufficient for separating peptides with minor differences in the sequence and composition of amino acids. Then, a “high resolution” may be required to detect more peptides with a high mass accuracy. In a previous report using *Escherichia coli* lysate as the analyte (Kim et al., 2010), larger numbers of proteins were identified at a resolution of 30000 than at a resolution of 60000. In our analysis using human cell lysates, I identified larger number of proteins at a resolution of 60000. Human produces more proteins than *E. coli* and therefore sample complexity become lesser. Thus, I can identify more proteins in resolution of 30000.

“Shorter IT of MS1” increased the number of identified proteins, probably because of a shorter scan time. Prior to the analysis, I had expected that a longer ion injection time made it possible to detect peptides around the limit of detection; however, a longer ion injection time did not increase the number of identified proteins because peptides with small ion intensities may not produce detectable MS spectra, as described previously (Kalli and Hess, 2012).

A “longer dynamic exclusion” increased the number of identified proteins. In a shorter dynamic exclusion time, peptides with the same *m/z* were detected multiple times in one analysis and the number of identified proteins became lesser, as described in previous reports (Kalli et al., 2013).

“Dynamic exclusion” and “gradient” had a two-factor interaction. In the larger gradient, I detected more proteins with a longer dynamic exclusion. In the smaller gradient, the number of identified proteins did not change between longer and shorter dynamic exclusion. Previous analysis suggests relationship between monolithic capillary LC and MS, but in this analysis I confirmed a two-factor interaction between both monolithic capillary LC and MS parameters (Pirmoradian et al., 2013). As shown in the

above gradient section, I obtained a high peak capacity in the larger gradient. A larger gradient increased peptide elution per unit time; therefore, a longer dynamic exclusion time had a larger effect than low peak capacity with a smaller gradient.

Further, “resolution” and “gradient time” had a two-factor interaction. A longer gradient time was necessary for a higher resolution. However, in a shorter gradient time, the effect of resolution became lesser. In a shorter gradient time, many peptides may co-elute; therefore, the number of identified proteins did not increase because of the limited scan speed in higher resolution. On the other hand, a longer gradient time would reduce co-eluting peptides and increase the margin for scan speed with high resolution. Hence, a high resolution could separate the proximal peaks and the number of identified proteins greatly increased. When multiple peptides co-elute in one scan, a high resolution may not contribute to an increase in the number of identified proteins because of a limited scan speed.

“IT of MS2” and “AGC of MS2” had a two-factor interaction. In 100 ms of IT of MS2, ions were maximally introduced into MS in our AGC setting (Kalli et al., 2013). The scan number and number of identified proteins did not change significantly between AGC of MS2 at 5000 and 50000. In 50 ms of IT of MS2, the amount of introduced ions did not reach the upper limit with AGC at 5000, as shown in the previous report (Kalli et al., 2013). Therefore, the scan number and number of identified proteins increased.

Contribution of possible effects of parameters on increase in average peak area of MS1

When the average peak area of MS1 was considered as a response variable, it was significantly affected by seven main effects (AGC of MS1, dynamic exclusion, injection amount, linear velocity, ion injection time of MS1, number of MS2 scans, and ion injection time of MS2) and one two-factor interaction (ion injection time of MS2-minimum count threshold) (Figure 1).

“Smaller AGC of MS1” and “shorter IT of MS1” significantly increased the average peak area of MS1. As the amount of ions increases, mass deviation increases (Kalli and Hess, 2012). Therefore, many ions may have a negative effect on the average

peak area of MS1.

In a “shorter dynamic exclusion time,” the average peak area of MS1 increased. I could identify stronger ions in the MS2 spectrum by shorter dynamic exclusion. Then, the average peak area of MS1 significantly increased. In this analysis, the average peak area of MS1 was calculated by integrating PSMs of Top3 peptides. Therefore, the average peak area of MS1 became higher when peptides with higher ion intensities were identified in the MS2 spectrum.

A “higher injection amount of the sample” increased the average peak area of MS1. When a higher amount of peptides was introduced into MS, a higher amount of ions was introduced to MS. Therefore, the amount of ions, which reached the detector, increased and the average peak area of MS1 significantly increased, as described in the previous report (Xu et al., 2009).

A “slow linear velocity” significantly increased the average peak area. A slow linear velocity leads to a high ionization efficiency and reduction in ion suppression(Schmidt et al., 2003). Furthermore, a small flow rate enriches the analyte on the column and leads to a high ion intensity (Meiring and der Heeft, 2002). These effects were critical for increasing the ion intensity because of a high S/N ratio.

A “higher number of MS2 scans” increased the average peak area of MS1. I could detect peptides with high ion intensities within each scan by a higher number of MS2 scans. Therefore, the average peak area of MS1 increased, as described in the section of dynamic exclusion.

A “longer IT of MS2” increased the average peak area of MS1. A longer IT of MS2 leads to a slightly increased number of identified proteins. A larger number of identified proteins implied that a longer IT of MS2 enables to obtain a better spectrum. In addition, PSMs assigned to peak areas may increase, as described in the above section on dynamic exclusion.

The “minimum count threshold” and “IT of MS2” had a two-factor interaction on the average peak area of MS1. In IT of MS2 at 100 ms, the average peak area of MS1 did not change depending on the minimum count threshold. In IT of MS2 at 50 ms, the average peak area of MS1 increased at a smaller minimum count threshold. In a longer

IT of MS2 at a small minimum count threshold, IT of MS2 became longer to introduce ions maximally to MS. Therefore, the scan speed may become slower in IT of MS2 at 100 ms than in IT of MS2 at 50 ms. Hence, the average peak area of MS1 may increase, as described in the above section on dynamic exclusion.

Compared with the two-factor interaction between the number of identified proteins and average peak area of MS1, the average peak area of MS1 had a two-factor interaction only between MS parameters. To maximize the peak area of monolithic capillary LC-ESI-MS/MS, I may not need to optimize the parameters of LC and MS simultaneously.

Optimization of LC, ESI and MS parameters

The optimized parameter sets for maximizing the number of identified proteins or average peak area of MS1 are shown in Table 3. I performed proteome analysis with human cell lysate in triplicate to verify that an increase of the number of identified proteins and the average peak area of MS1 with these parameter sets is not due to run-to-run variance. As a result, the number of identified proteins was significantly increased by 8.1% and the average peak area of MS1 was increased by 67% (Figure 3). In these analyses, I could optimize LC, ESI, and MS parameters depending on the target response variables with a high reproducibility. I should mention that these parameters could be local maximum. In further research, I will improve our method to find global maximum by using wide range of values for each parameter. These parameters cannot be readily adopted to other MS instruments and LC column, but this DoE methodology can be adopted to other MS instruments and LC column.

Table 3 Optimized parameters for increasing number of identified proteins or average peak area of MS1

Parameters	Optimized parameters for maximum number of identified proteins	Optimized parameters for maximum average peak area of MS1	Standard parameters
Linear velocity [mm/s]	1.0	0.8	1.0
Gradient [%]	5–60	5–30	5–45
Injection amount [μg]	5	5	1
Gradient time [min]	750	450	600
ESI spray voltage [kV]	2.1	2.3	2.1
Capillary temperature [°C]	280	260	280
Resolution	60000	30000	60000
AGC in MS1	1.0×10^6	1.0×10^6	1.0×10^6
AGC in MS2	5.0×10^3	5.0×10^4	5.0×10^3
IT of MS1 [ms]	500	500	500
IT of MS2 [ms]	50	50	50
Dynamic exclusion [s]	150	30	90
Minimum count threshold	500	500	500
Number of MS2 scans	10	20	10

AGC, target value of auto gain control; IT, maximum ion injection time

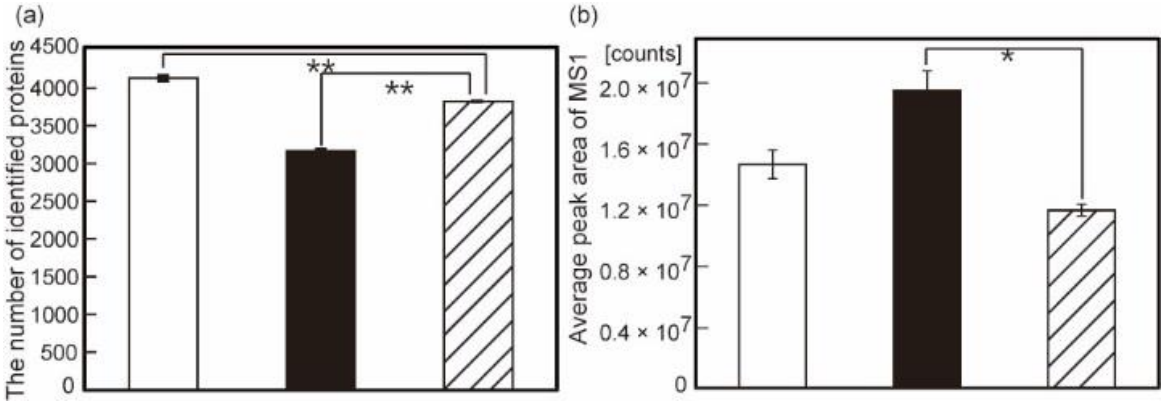


Figure 3 (a) The number of proteins and (b) average peak area of MS1 obtained with optimized parameter sets.

Blank bar implies analysis with optimized parameters for maximizing the number of identified proteins. Filled bar implies analysis with optimized parameters for maximizing the average peak area of MS1. Hatched bar implies analysis with standard parameters. In both sets, objective response variables were significantly increased. Error bars indicate standard error (N = 3). Asterisk (*) indicates *p*-value < 0.05, and double asterisks (**) indicate *p*-value < 0.01.

With respect to the parameter set for maximizing the number of identified proteins, the number of identified proteins increased by 8.1% and gradient time increased by 25%, i.e., the increasing rate of gradient time was larger than that of the number of identified proteins. Long gradient times would not be ideal for increasing the number of identified proteins per minute. To identify a higher number of proteins in one-shot analysis with high time efficiency, a high scan speed with high resolution is required, particularly in human proteome analysis (Shishkova et al., 2016). In our settings, the scan speed of MS is approximately 2–3 s per MS1 scan and subsequent ten MS2 scans (Kim et al., 2010). Shishkova et al. (Shishkova et al., 2016) recently described that a high peak capacity with a longer separation time or highly efficient chromatography did not lead to an increase in the number of identified proteins without a high scan speed. In addition, even if the scan speed increases, the number of identified proteins does not readily increase unless the peak capacity increases. The scan speed of the most recent MS, e.g., Orbitrap Fusion, is approximately five times faster than that of another recent MS, e.g., Orbitrap Velos. To

utilize such cutting-edge MS, I should simultaneously optimize LC, ESI and MS parameters such as scan speed, resolution, and peak capacity.

Summary

With a definitive screening design, I could optimize LC, ESI and MS parameters for maximizing the number of identified proteins and average peak area of MS1 in non-targeted proteomics. It is required to optimize instruments with a definitive screening design depending on the samples, column length, inner diameter, and MS instruments. By referring to this study, such parameters can be optimized with fewer experiments.

Previous papers have only focused on LC or MS parameters individually for optimizing LC-ESI-MS/MS (Andrews et al., 2011; Kalli and Hess, 2012; Kalli et al., 2013; Xu et al., 2009). In my analysis, I could detect two-factor interactions between LC and MS for the first time, to the best of our knowledge. I should now take into consideration two-factor interactions between both LC and MS for optimizing LC, ESI and MS. Optimizing parameters using a definitive screening design will be a versatile method for any LC-ESI-MS/MS.

References

Andrews, G.L., Dean, R.A., Hawkridge, A.M., and Muddiman, D.C. (2011). Improving proteome coverage on a LTQ-Orbitrap using design of experiments. J Am Soc Mass Spectrom *22*, 773-783.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. Mol Sys Biol *7*, 549.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., and Nagaraj, N. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cel Proteomics *13*, 2513-2526.

Feltcher, M.E., Gunawardena, H.P., Zulauf, K.E., Malik, S., Griffin, J.E., Sassetti, C.M., Chen, X., and Braunstein, M. (2015). Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall. Mol Cel Proteomics *14*, 1501-1516.

Horie, K., Kamakura, T., Ikegami, T., Wakabayashi, M., Kato, T., Tanaka, N., and Ishihama, Y. (2014). Hydrophilic interaction chromatography using a meter-scale monolithic silica capillary column for proteomics LC-MS. Anal Chem *86*, 3817-3824.

Horie, K., Sato, Y., Kimura, T., Nakamura, T., Ishihama, Y., Oda, Y., Ikegami, T., and Tanaka, N. (2012). Estimation and optimization of the peak capacity of one-dimensional gradient high performance liquid chromatography using a long monolithic silica capillary column. J Chromatogr A *1228*, 283-291.

Hurvich, C.M., and Tsai, C.L. (1989). Regression and time series model selection in small samples. Biometrika, 297-307.

Jones, B., and Nachtsheim, C.J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. J Qual Technol *43*, 1.

Köcher, T., Swart, R., and Mechtler, K. (2011). Ultra-high-pressure RPLC - an LTQ-Orbitrap celos reveals a linear relation between peak capacity and number of identified peptides. Anal Chem *83*, 2699-2704.

Kalli, A., and Hess, S. (2012). Effect of mass spectrometric parameters on peptide and

protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. Proteomics *12*, 21-31.

Kalli, A., Smith, G.T., Sweredoski, M.J., and Hess, S. (2013). Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers. J Proteome Res *12*, 3071-3086.

Kim, M.S., Kandasamy, K., and Chaerkady, R. (2010). Assessment of resolution parameters for CID-based shotgun proteomic experiments on the LTQ-orbitrap mass spectrometer. J Am Soc Mass Spectrom *21*, 1606-1611.

Lemeer, S., Hahne, H., Pachl, F., and Kuster, B. (2012). Software tools for MS-based quantitative proteomics: A brief overview. In quantitative methods in proteomics, K. Marcus, ed. (Totowa, NJ: Humana Press), pp. 489-499.

Meiring, H.D., and der Heeft, V.E. (2002). Nanoscale LC–MS (n): technical design and applications to peptide and protein analysis. J Sep Sci *25*, 557-568.

Morisaka, H., Matsui, K., Tatsukami, Y., and Kuroda, K. (2012). Profile of native cellulosomal proteins of *Clostridium cellulovorans* adapted to various carbon sources. AMB Express *2*, 37.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. Mol Sys Biol *7*, 548-548.

Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R.A. (2013). Rapid and deep human proteome analysis by single-dimension shotgun proteomics. Mol Cel Proteomics *12*, 3330-3338.

Schmidt, A., Karas, M., and Dülcks, T. (2003). Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? J Am Soc Mass Spectrom *14*, 492-500.

Shishkova, E., Hebert, A.S., and Coon, J.J. (2016). Now, more than ever, proteomics needs better chromatography. Cell Systems *3*, 321-324.

Silva, J.C., Gorenstein, M.V., Li, G.Z., and Vissers, J.P.C. (2006). Absolute quantification of proteins by LCMS$^{(E)}$ a virtue of parallel MS acquisition. Mol Cel Proteomics *5*, 144-156.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533-538.

Xu, P., Duong, D.M., and Peng, J. (2009). Systematical optimization of reverse-phase chromatography for shotgun proteomics. J Proteome Res *8*, 3944-3950.

Yamana, R., Iwasaki, M., Wakabayashi, M., Nakagawa, M., Yamanaka, S., and Ishihama, Y. (2013). Rapid and deep profiling of human induced pluripotent stem cell proteome by one-shot nanoLC–MS/MS analysis with meter-scale monolithic silica columns. J Proteome Res *12*, 214-221.

Chapter II

Elucidation of the recognition mechanisms for hemicellulose and pectin in *Clostridium cellulovorans* using intracellular quantitative proteome analysis

C. cellulovorans is attractive bacteria for degrading and metabolizing plant cell wall polysaccharides. This bacteria can degrade and metabolize all types of plant cell wall polysaccharides (cellulose, hemicellulose and pectin). Furthermore, this bacteria changes the expression pattern of carbohydrate-degrading enzymes depending on each plant cell wall polysaccharides. However the mechanisms for sensing plant cell wall polysaccharides and changing expression patterns of carbohydrate-degrading enzymes are not known in *C. cellulovorans*.

In chapter I, I optimized LC-ESI-MS/MS parameters for maximizing the number of identified proteins and average peak area of MS1. Proteome analysis with optimized LC-ESI-MS/MS enables to identify more proteins. In this chapter, I analyzed the mechanism how *Clostridium cellulovorans* changes expression patterns of carbohydrate-degrading enzymes depending on plant cell wall polysaccharides using intracellular quantitative proteome analysis.

Materials and methods

Cell culture and media

C. cellulovorans 743B (ATCC35296) was grown anaerobically as previously described (Sleat et al., 1984), differing only in the carbon source, which was replaced by 0.3% (w/v) glucose (Nacalai Tesque, Kyoto, Japan), 0.3% (w/v) xylan from beechwood (Sigma, MO, USA), 0.3% (w/v) pectin from apple (Sigma), or 0.3% (w/v) locust bean gum from *Caretonia silliqua* seeds (Sigma).

Estimating the growth of the anaerobic bacteria

The growth curves of *C. cellulovorans* on each medium were determined by bacterial protein estimation, as previously described (Raman et al., 2009) with small modifications. Cells from 1 mL of cell culture were collected by centrifugation (13,000 × *g*, 4°C, 10 min). The cell pellets were washed with 1 mL of phosphate-buffered saline (pH 7.4; PBS) and incubated with 800 μL of sodium deoxycholate (2%) for 20 min at 37°C. Two hundreds microliters of trichloroacetic acid (24%) was added to the suspension, which was then centrifuged (13,000 × *g*, 4°C, 10 min). One hundred microliters of resolubilization solution (5% SDS, 2 N NaOH) was added to the suspension and vigorously mixed. The protein concentration was measured using a protein assay bicinchoninate kit (Nacalai Tesque), with bovine serum albumin used as a standard.

Sample preparation for quantitative proteome analysis

C. cellulovorans was grown in 50 mL cultures to late-logarithmic phase (36 h). Cells were concentrated by centrifugation (6,000 × *g*, 4°C, 10 min), and the supernatant was discarded. Cell pellets were collected by centrifugation (13,000 × *g*, 4°C, 10 min) and washed with 500 μL PBS, and centrifuged again. Cells were resuspended with 500 μL of lysis buffer (2% (w/v) 3-(3-cholamidopropyl)dimethylammonio-1-propanesulfonate, 10 mM dithiothreitol, 1% *(*v/v*)* protease inhibitor cocktail for bacterial cell lysis (Sigma), 7 M urea, and 2 M thiourea in 50 mM tris-HCl (Nacalai Tesque)). The cells were disrupted with a Bioruptor UCD-250T (Cosmo Bio, Tokyo, Japan) at 250 W, 15 s on-and-off cycles for 10 min, on ice. The solution was centrifuged (13,000 × *g*, 4°C, 20 min), and the supernatant was collected and subjected to ultrafiltration with an Amicon Ultra-0.5 Centrifugal Filter Unit (10 kDa, Millipore, MA, USA) and buffer-exchanged with 200 mM triethyl ammonium bicarbonate (TEAB; Sigma). Proteins were reduced by adding tris (2-carboxyethyl) phosphine to 10 mM from a 200 mM stock, and the reaction was allowed to proceed for 60 min at 55°C. Following the incubation, 5 μL of iodoacetamide (375 mM) was added, and the reaction continued for 30 min at room temperature in the dark. Proteins were precipitated by adding 1 mL of ice-cold acetone and incubating the solution overnight at −20°C. The precipitated proteins were

resuspended with 100 µL of 200 mM TEAB, 2 µg of sequencing grade modified trypsin (Promega, WI, USA) was added, and incubated overnight at 37°C. The four samples for proteome analysis (glucose, xylan, pectin, and galactomannan) were labeled using a tandem mass tag (TMT) 6-plex labeling kit (Thermo Fisher Scientific, MA, USA) with reporters at *m/z* = 126, 127, 129, and 130, respectively, in 41 µL acetonitrile. After 60 min at room temperature, 8 µL of 5% (w/v) hydroxylamine was added to each tube and mixed for 15 min. As an internal standard for quantification, a mixture of tryptic fragments from all substrates was combined with TMT-131 (reporter at *m/z* = 131). Aliquots were pooled and evaporated under vacuum, then dissolved in 60 µL of formic acid (0.1%) and subjected to LC-MS/MS analysis.

LC-MS/MS analysis

Proteome analysis was performed using an LC (Ultimate 3000®; Thermo Fisher Scientific)-MS/MS (LTQ Orbitrap Velos Mass Spectrometer®; Thermo Fisher Scientific) system equipped with a long monolithic silica capillary column. Tryptic digests were separated by reverse-phase chromatography using a monolithic silica capillary column (500 cm long, 0.1 mm ID) (Morisaka et al., 2012), at a flow rate of 500 nL/min. A gradient was achieved by changing the ratio of the two eluents: eluent A, 0.1% (v/v) formic acid; eluent B, 80% acetonitrile containing 0.1% (v/v) formic acid. The gradient started with 5% B, increased to 45% B for 600 min, further increased to 95% B to wash the column for 140 min, returned to the initial condition, and was held for re-equilibration of the column. The separated analytes were detected using a mass spectrometer with a full scan range of 350–1,500 *m/z* (resolution 60,000), followed by 10 data-dependent higher-energy c-trap dissociation (HCD) MS/MS scans acquired for TMT reporter ions by using 40% normalized collision energy in HCD with 0.1 ms activation time in quantitative proteome analysis and with a full scan range of 350–1,500 *m/z* (resolution 60,000), followed by 10 data-dependent collision-induced dissociation (CID) MS/MS scans in a qualitative proteome analysis. An electrospray ionization (ESI) voltage was set at 2.3 kV. Triplicate analyses were performed for each sample in three independent experiments, and the collected data were reviewed for protein identification and

quantification in a quantitative proteome analysis. Single analyses were performed for each sample of the three independent experiments in a qualitative proteome analysis. The collected data were reviewed for protein identification and quantification. Blank runs were inserted between runs of different samples.

Data analysis

Data analysis was performed using Proteome Discoverer 1.4 (Thermo Fisher Scientific). Protein identification was performed using the Mascot algorithm against the *C. cellulovorans* protein database (4,254 sequences) from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) with a precursor mass tolerance of 20 ppm and a fragment ion mass tolerance of 50 mmu in quantitative proteome analysis, and mass tolerance of 2.0 Da and a fragment ion mass tolerance of 0.8 Da in qualitative proteome analysis. Carbamidomethylation of cysteine was set as a fixed modification for quantitative proteome analysis, and carbamidomethylation of cysteine and a TMT 6-plex at the N-terminus and lysine were set as fixed modifications for qualitative proteome analysis. Protein quantification was performed using the Reporter Ions Quantifier with the TMT 6-plex method. The data were then filtered with a cut-off criteria of q-value ≤ 0.05, corresponding to a 5% false discovery rate (FDR) on a spectral level. Proteins with no missing values in three replicates were accepted as identified proteins for the protein quantification analysis. In the qualitative proteome analysis, proteins with scores > 10 were accepted as identified proteins. Global median normalization was performed to normalize the quantity of each tryptic digest injected into the mass spectrometer. The heat map was generated using Cluster 3.0 (de Hoon et al., 2004), which can carry out hierarchical cluster analysis (HCA). Euclidean distance was used to measure the similarities of the protein profile patterns within the clustering analysis. To visualize the clustering results from Cluster 3.0, Java TreeView software was used (Saldanha, 2004). The pathway map was constructed from Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) (Kanehisa and Goto, 2000), and identified proteins were adapted to this map.

Results and discussion

To identify substrate recognition systems of *C. cellulovorans*, I carried out a quantitative proteome analysis of cells grown anaerobically on media with each carbon sources. Using a quantitative proteome approach based on isobaric tagging, I obtained protein profiles from cells grown on each carbon source. The workflow of the intracellular proteome analysis is illustrated in Figure 1.

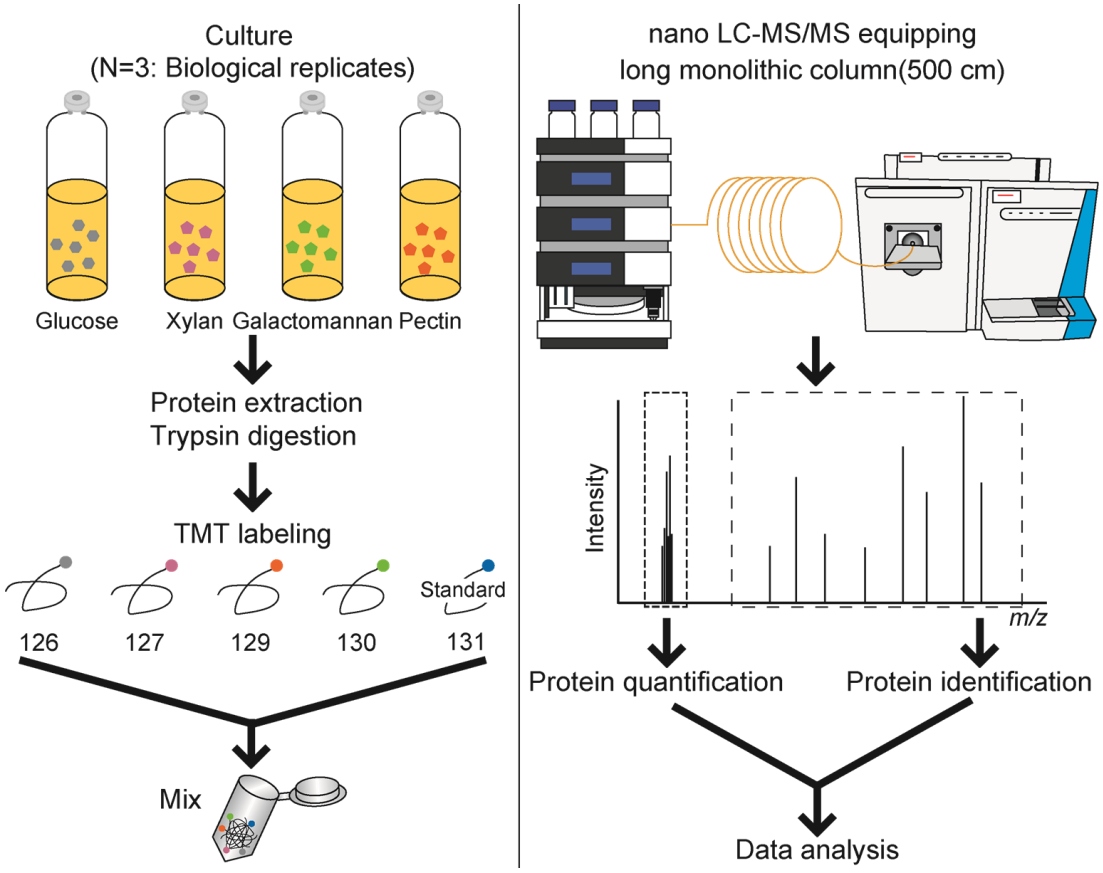


Figure 1. Experimental procedure of *C. cellulovorans* intracellular quantitative proteome analysis

Proteins prepared from cell lysates of *C. cellulovorans* grown in the presence of glucose, xylan, galactomannan, and pectin were individually reductive-alkylated and digested with trypsin. Tryptic fragments were labeled with tandem mass tags (TMTs). The labeled peptides were mixed and injected into the LC-MS/MS system with a long monolithic silica capillary column for mass measurement and data collected were used for protein quantification.

Growth confirmation

To determine the growth of *C. cellulovorans* on each of the substrates, I conducted bacterial protein estimation (Figure 2). As has previously been shown, *C. cellulovorans* can grow on xylan, pectin, or galactomannan as the sole carbon source (Sleat et al., 1984), although other cellulosome-producing Clostridia, such as *C. thermocellum* and *C. cellulolyticum*, cannot grow on pectin and galactomannan (Petitdemange et al., 1984; Prawitwong et al., 2013). Growth on glucose, xylan, and pectin was slower than that on galactomannan, but cells were collected from all cultures at similar growth phases. From the growth analysis, I selected a culture time of 36 h, as cells were in the late-logarithmic phase, which was appropriate for proteome analysis in terms of growth phase and protein concentration.

Qualitative and quantitative proteome analysis

Nano LC-MS/MS equipped with a long monolithic silica capillary using a tandem mass tag (TMT) 6–plex isobaric tag column was employed for intracellular quantitative proteome analysis. *C. cellulovorans* has 4,254 protein-encoding genes in its genome (Tamaru et al., 2010). For protein identification, I constructed a protein database built from the genome of *C. cellulovorans*. In total, I could identify 734 proteins from all samples within our cutoff criteria. To correct for variations in the amount of TMT-labeled peptides infused into the mass spectrometer, I normalized all quantitative data to the median value of that analysis. For all data analysis, I used these normalized relative quantification values. I checked for reproducibility of quantitative protein profiles among three biological replicates for each substrate by HCA (Figure 3). Each array was clustered individually based on each biological replicate. Our results indicate that each sample was reproducibly quantified, and analytical and biological replication of results was ensured. Furthermore, I separately carried out a qualitative proteome analysis to compare the results of quantitative proteome analysis.

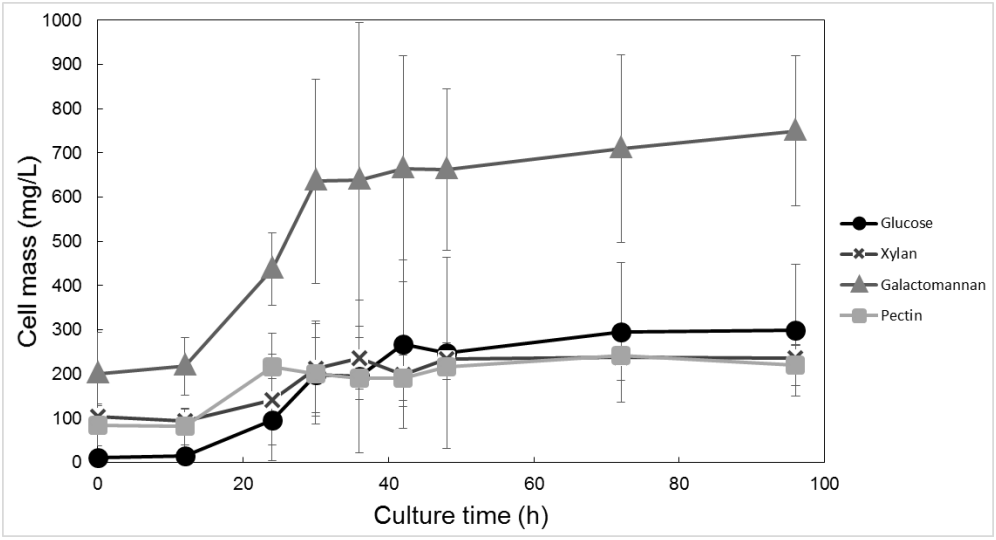


Figure 2. Confirmation of growth of *C. cellulovorans* cultured with four different substrates

Growth of *C. cellulovorans* was measured by estimation of protein in cell lysates to determine an appropriate culture time (Raman et al., 2009). At 36 h, *C. cellulovorans* appears to be in the late-logarithmic phase in all four substrates (glucose, xylan, galactomannan, pectin).

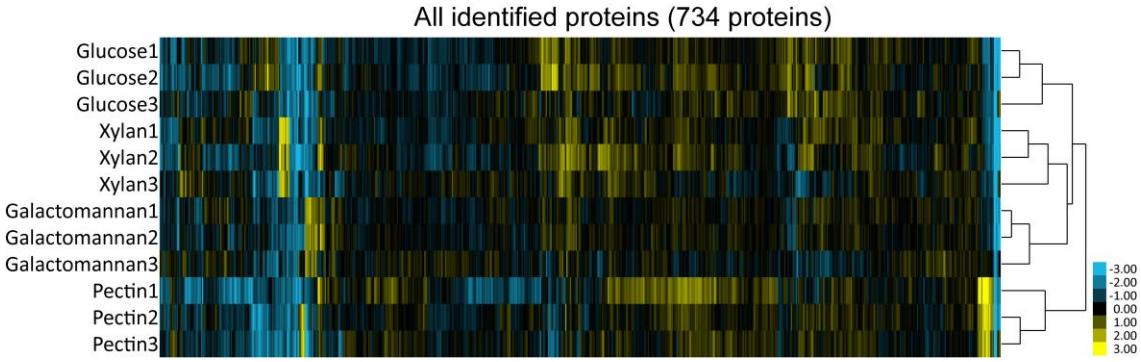


Figure 3 Hierarchical clustering analysis represented proteome profiles of *C. cellulovorans* with four different substrates.

The quantitative proteome data from three biological replicates of each substrate were used for the hierarchical clustering analysis. Each array was clustered in response to substrates, and each biological replicate were grouped together. Color bar indicates changes in protein abundance. Increased and decreased protein levels are shown in yellow and blue, respectively.

Substrate-specific proteins

To discover substrate-specific proteins (those that showed a significant change between different growth substrates), I carried out an empirical Bayes moderated *t* test (Smyth, 2004). *P*-values were adjusted with the Benjamini-Hochberg method to avoid the problem of multiple testing. The thresholds that I adopted were FDR-adjusted *p*-value of < 0.05 and fold-change of protein ratio > 2.0 compared to glucose. Proteins for which the levels significantly changed were defined by comparison between glucose and each polysaccharide (xylan, galactomannan, and pectin) at these thresholds. All substrate-specific proteins detected are shown in Table 1. Using KEGG analysis and cluster analysis based on genome analysis, I focused mainly on metabolism-related and substrate recognition-related proteins for further analysis.

Profiles of metabolism-related proteins

First, I focused on profiles of metabolism-related proteins, such as enzymes involved in substrate degradation and metabolism, and other characteristic metabolic pathway. *C. cellulovorans* is known to change production of carbohydrate-related enzymes secreted into media from exoproteome analyses (Esaka et al., 2015; Matsui et al., 2013) and alternation of production of different metabolic pathways depending on which substrates are available in culture is also predicted.

Table 1 Substrate-specific proteins

Xylan specific proteins		vs Glucose		Pectin specific proteins		vs Glucose	
Locus	Description	log2-fold change	FDR adjusted *p* value	Locus	Description	log2-fold change	FDR adjusted *p* value
Clocel_0589	alpha-L-fucosidase	2.67	2.07E-04	Clocel_0048	transcriptional regulator, AbrB family	1.91	2.50E-03
Clocel_0590	xylose isomerase	3.35	1.43E-04	Clocel_0322	TatD family hydrolase	1.74	8.66E-03
Clocel_0591	transaldolase	3.52	1.46E-06	Clocel_0513	extracellular solute-binding protein	2.12	2.50E-03
Clocel_0592	xylulokinase	2.85	7.29E-05	Clocel_0519	glycogen/starch/alpha-glucan phosphorylase	1.88	1.37E-02
Clocel_1085	dinitrogenase iron-molybdenum cofactor biosynthesis protein	2.32	4.24E-03	Clocel_1243	extracellular solute-binding protein	2.53	9.20E-05
Clocel_1151	methyl-accepting chemotaxis sensory transducer	1.96	1.70E-02	Clocel_1892	acetate kinase	1.31	6.43E-03
Clocel_1430	glycoside hydrolase family protein	1.90	9.48E-03	Clocel_2210	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	1.08	1.47E-02
Clocel_2573	hypothetical protein Clocel_2573	1.60	3.95E-03	Clocel_2214	ATP:corrinoid adenosyltransferase BtuR/CobO/CobP	1.37	4.25E-02
Clocel_2592	two component transcriptional regulator, AraC family	3.72	1.49E-04	Clocel_2222	precorrin-3B C(17)-methyltransferase	1.70	6.63E-03
Clocel_2595	xylan 1,4-beta-xylosidase	4.70	4.53E-06	Clocel_2225	cobalamin (vitamin B12) biosynthesis CbiX protein	1.95	4.32E-03
Clocel_2596	sugar ABC transporter periplasmic protein	4.62	1.06E-03	Clocel_2227	precorrin-4 C(11)-methyltransferase	2.11	7.92E-03
Clocel_2597	inner-membrane translocator	4.01	7.36E-05	Clocel_2250	altronate dehydratase	4.07	1.02E-05
Clocel_2598	ABC transporter	4.34	1.89E-05	Clocel_2251	mannitol dehydrogenase domain-containing protein	4.50	4.39E-06
Clocel_2881	PTS system lactose/cellobiose-specific transporter subunit IIB	1.49	1.70E-02	Clocel_2253	Crp family transcriptional regulator	2.37	2.50E-03
Clocel_2940	putative phosphate transport regulator	1.44	2.45E-02	Clocel_2254	glycoside hydrolase family protein	4.36	4.39E-06
Clocel_3175	PhoH family protein	1.64	3.84E-02	Clocel_2255	major facilitator superfamily protein	4.33	4.02E-05
Clocel_3761	ATP:guanido phosphotransferase	1.31	1.37E-02	Clocel_2256	glycosyl hydrolase family protein	4.19	1.49E-05
Clocel_3762	UvrB/UvrC protein	1.42	3.45E-02	Clocel_2259	PfkB domain-containing protein	3.75	1.49E-06
Clocel_4277	aldo/keto reductase	1.08	4.66E-02	Clocel_2262	short-chain dehydrogenase/reductase SDR	3.21	2.11E-02
Galactomannan specific proteins				Clocel_2263	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase	3.15	4.92E-03
Clocel_0034	glycoside hydrolase family protein	1.73	1.04E-02	Clocel_2403	glucosamine/fructose-6-phosphate aminotransferase	1.45	4.57E-03
Clocel_0391	glycosyltransferase	1.83	1.96E-02	Clocel_2737	small GTP-binding protein	1.10	2.98E-02
Clocel_0684	thiazole biosynthesis protein ThiH	1.57	4.69E-03	Clocel_3380	LPXTG-motif cell wall anchor domain-containing protein	4.03	1.27E-04
Clocel_2225	cobalamin biosynthesis CbiX protein	1.39	4.52E-02	Clocel_3909	quorum-sensing autoinducer 2 (AI-2), LuxS	1.19	2.34E-02
Clocel_2259	PfkB domain-containing protein	1.31	1.04E-02	Clocel_4088	galactose-1-phosphate uridylyltransferase	1.62	1.73E-02
Clocel_2697	sialate O-acetylesterase	1.69	1.69E-02	Clocel_4277	aldo/keto reductase	1.31	8.08E-03
Clocel_2800	Alpha-galactosidase	2.99	5.49E-05				
Clocel_2962	inosine-5\'-monophosphate dehydrogenase	1.31	1.96E-02				
Clocel_3175	PhoH family protein	1.98	1.14E-02				
Clocel_3194	mannose-6-phosphate isomerase	3.77	6.86E-05				
Clocel_3196	glycosidase-like protein	2.65	5.29E-03				
Clocel_3198	N-acylglucosamine 2-epimerase	3.19	2.10E-03				
Clocel_3200	binding-protein-dependent transport system inner membrane protein	4.76	8.61E-04				
Clocel_3201	extracellular solute-binding protein	4.23	5.49E-05				
Clocel_3205	glycoside hydrolase family 2	2.55	1.51E-03				
Clocel_3657	xylan 1,4-beta-xylosidase	1.79	4.26E-02				
Clocel_3857	ABC transporter	2.15	1.03E-03				
Clocel_4053	LPXTG-motif cell wall anchor domain-containing protein	2.28	3.79E-02				
Clocel_4087	aldose 1-epimerase	1.45	2.68E-02				
Clocel_4088	galactose-1-phosphate uridylyltransferase	3.38	1.83E-04				
Clocel_4089	UDP-glucose 4-epimerase	2.09	1.82E-03				
Clocel_4277	aldo/keto reductase	1.08	4.26E-02				

Degradation and metabolism of each substrate

I constructed a substrate degradation pathway from KEGG pathway maps, and presented the fold change of each protein (Figure 4). For xylan degradation- and metabolism- related proteins (Figure 4A), the levels of three proteins (Clocel_0590, 0592, and 2595) were significantly elevated in the presence of xylan (Table 1), but production of Clocel_2900 (Endo-1, 4-beta xylanase) was not specifically elevated.

For galactomannan degradation- and metabolism-related proteins (Figure 4B), the levels of 9 proteins (Clocel_2259, 2800, 3194, 3196, 3198, 3205, 4087, 4088, and 4089) were significantly elevated in the presence of galactomannan (Table 1). By contrast, the levels of mannanase (Clocel_1134 and 4119) were not significantly elevated.

For pectin degradation- and metabolism-related proteins (Figure 4C), levels of 8 proteins (Clocel_2250, 2251, 2254, 2256, 2259, 2262, 2263, and 3380) were significantly elevated in the presence of pectin (Table 1). For the degradation and metabolism of pectin, both the "hydrolase/isomerase pathway" and the "lyase/5-dehydro-4-deoxy-gluconate pathway" were found (Richard and Hilditch, 2009). Clocel_2254 is a member of the Glycoside Hydrolases (GH) 28 family (Cantarel et al., 2009), which is known to have endogalacturonase and exogalacturonase activities. Additionally, Clocel_3380 is a member of the Polysaccharide lyase (PL) 9 family. BLAST analysis indicates that the protein has exogalacturonate lyase activity.

I first focused on metabolic pathways of each substrate in *C. cellulovorans* by KEGG analysis. For xylan metabolism, it is interesting to note that intracellular levels of Clocel_2595 (xylosidase) increased, but levels of Clocel_2900 (xylanase) were not enhanced. Although Clocel_2595 (xylosidase) and some xylanases were categorized as xylan-specific in "exoproteome" analysis (Matsui et al., 2013), *C. cellulovorans* cannot grow in medium containing xylose as the sole carbon source (Sleat et al., 1984). Combining our findings with the inability of *C. cellulovorans* to grow in xylose medium, I predict that *C. cellulovorans* metabolizes xylooligosaccharides intracellularly. I hypothesize the following pathway for xylan metabolism: *C. cellulovorans* degrades xylan to xylooligosaccharides extracellularly and transports them into the cytoplasm. Next, xylooligosaccharides are degraded to xylose in the cell, and xylose is metabolized and used by the pentose phosphate pathway. Indeed, I observed increased levels of Clocel_0591 (transaldolase), which is related to the pentose phosphate pathway, in xylan medium.

For galactomannnan metabolism, I focused on Clocel_3196 (glycosidase-like protein), which is a member of the GH130 family and is annotated as beta-1,4-mannooligosaccharide/beta-1,4-mannosyl-*N*-acetylglucosamine phosphorylase

(Kawahara et al., 2012). Similarly annotated genes are part of mannan metabolic pathways in other bacteria (Senoura et al., 2011), and I observed increased levels of proteins in this pathway (Clocel_3198, 3205) in the presence of galactomannan. However, Clocel_3197, which is known to play an important role in the mannan metabolic pathway (Esaka et al., 2015), was only identified in our qualitative proteome analysis. This protein has been also detected with "exoproteome" analysis (Esaka et al., 2015). It is assumed that this protein acts both intracellularly and extracellularly through an unknown mechanism. In addition, I identified increased levels of alpha-galactosidase (Clocel_2800). In a previous study, this protein was identified in an "exoproteome" analysis, and it is predicted to degrade galactose side chains both intracellularly and extracellularly. Specifically, the extracellular collaboration between mannanase and alpha galactosidase rapidly degrades galactomannan. Next, galactose side chains that were not degraded extracellularly are degraded by "intracellular" alpha galactosidase, and the resulting mannooligosaccharides and galactose are further metabolized.

I predicted a pectin metabolism pathway in *C. cellulovorans* by combining the results of the present study and the previously reported "exoproteome" analysis (Matsui et al., 2013). Extracellularly, pectin is degraded by pectin lyase (Clocel_0873, 1172, and 3380) to oligogalacturonate with both saturated and unsaturated ends. In parallel, pectin is also de-esterified to pectate (Clocel_0211 and 3114), resulting in simpler extracellular degradation pathways. The resulting metabolites are transported into the cell and metabolized with the hydrolase/isomerase and lyase/5-dehydro-4-deoxy-gluconate pathways. In *Erwinia chrysathemi*, which is a well-studied pathogenic bacterium, pectin is degraded using similar systems; i.e. pectin is degraded to oligogalacturonides outside of the cell. Next, oligogalacturonides are transported into cells and metabolized in the hydrolase/isomerase and lyase/5-dehydro-4-deoxy-gluconate pathways (Hugouvieux-Cotte-Pattat et al., 1996). While known for *E. chrysathemi*, this system has not been reported in other cellulosome-producing Clostridia. *C. cellulovorans* commonly degrades all sugar-substrates (polysaccharides) to oligosaccharides, not to monosaccharide, outside of the cell. Degradation of those oligosaccharides to monosaccharides and further metabolism take place intracellularly. Similar degradation systems have not been reported

in Clostridia although *C. cellulolyticum* can metabolize xylose as a sole carbon source. Furthermore, in the presence of pectin, I found increased levels of an oligogalacturonides (esterified-oligogalacturonates) transporter (Clocel_2255). In the current analysis, I did not identify pectinesterase "intracellularly," although it had previously been identified in an "exoproteome" analysis (Matsui et al., 2013). Together, these results suggest that oligogalacturonates are transported into *C. cellulovorans*, and degraded to monogalacturonates and further metabolized intracellularly. In addition to the degradation and metabolism of each substrate, levels of 4 proteins related to the synthesis of vitamin B12 were significantly elevated in cells cultured on pectin (Clocel_2214, 2222, 2225, and 2227). I was unable to find any pectin-specific proteins reported to have a requirement for vitamin B12 (Sañudo-Wilhelmy et al., 2014). For the degradation and metabolism of pectin, *C. cellulovorans* uses pectin lyase (Clocel_2259) (EC 4.2.1.7), which showed elevated levels in our analysis. I speculate that this enzyme may require vitamin B12.

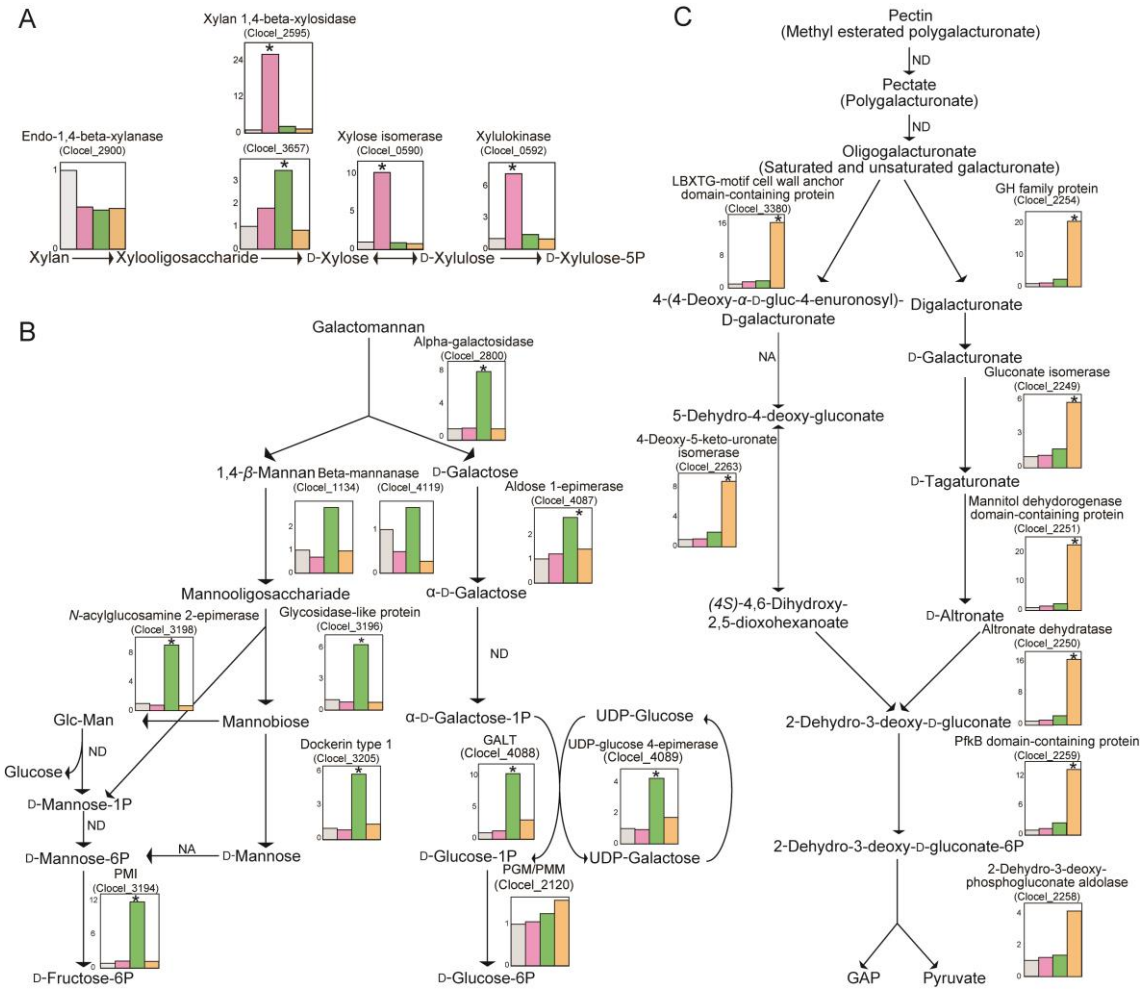


Figure 4. Degradation and metabolism pathways for each substrate

Each substrate degradation and metabolism pathway is shown: (A) xylan, (B) galactomannan, (C) pectin. For each protein, the fold change compared to glucose is shown. Asterisk (*) denotes significantly elevated proteins. Gray bar shows fold change in glucose-grown cells (=1); pink bar shows fold change in xylan-grown cells; green bar shows fold change of galactomannan-grown cells; orange bar shows fold change in pectin-grown cells. ND : protein not detected; NA : genes commonly assigned to the pathway, but not annotated; PMI: Phosphomannose isomerase; GALT: Galactose-1-phosphate uridyltransferase; PGM/PMM: Phosphoglucomutase/ phosphomannomutase alpha/beta/alpha domain I; 1P: 1-phosphate; 6P: 6-phosphate.

Genome and cluster analyses

To identify candidates related to substrate recognition systems, I performed

genomic analysis based on *C. cellulolyticum* TCS-related gene clusters (Xu et al., 2013). For the thresholds of TCS-regulated cluster identification, I applied two criteria. First, a cluster must contain more than two components of a TCS pathway: an integral membrane histidine kinase (sensor histidine kinase), a transcriptional regulator (response regulator), and an extracellular solute-binding protein (sugar binding protein). Second, degradation-, metabolism-, or transport-related gene loci must be located within 2 of the TCS genes identified above. Using this threshold, I identified 14 candidate clusters related to TCS.

Combining the information of our genome analysis and the substrate-specific proteins, I identified xylan- and galactomannan-specific clusters that include TCS-related genes. I identified clusters corresponding to two substrates; the xylan-specific cluster included Clocel_2592, 2595, 2596, 2597, and 2598 (Table 1), and Clocel_2593 and Clocel_2594. The galactomannan-specific cluster included Clocel_3194, 3196, 3198, 3200, 3201, and 3205 (Table 1), and Clocel_3195, 3197, 3199, 3202, 3203, 3204 (Figure 5). Each cluster includes three components of a TCS: a transcriptional regulator AraC, an integral membrane histidine kinase, and an extracellular solute-binding protein. I suggest that these genes are common components of a hemicellulose recognition system.

From pectin, I found increased levels of proteins Clocel_2250, 2251, 2253, 2254, 2255, 2256, 2259, 2262, and 2263 (Table 1) and Clocel_2249, 2257, 2258, and 2260. I suggest that this is a large cluster for the degradation and metabolism of pectin. For the degradation and metabolism of pectin, hydrolase/isomerase and lyase/5-dehydro-4-deoxy-gluconate pathways are known. The pectin-specific cluster which I identified contains genes related to both of these pathways, as well as for transporting pectin.

I also identified a xylose-metabolic cluster containing proteins Clocel_0589, 0590, 0591, and 0592 (Table 1) that had increased levels specifically in the presence of xylan. A galactose-metabolic cluster containing proteins Clocel_4087, 4088, and 4089 (Table 1) and Clocel_4090 had elevated levels specifically in the presence of galactomannan. The xylose-metabolic cluster contained two xylose metabolism-related proteins, Clocel_0590 and 0592 (Table 1; Figure 4A), aldose epimerase (Clocel_0591; Table 1), which is related to the pentose phosphate pathway, and alpha-fucosidase (Clocel_0589; Table 1). In the galactose-metabolic cluster, galactose metabolism-related genes are present (Figure 4B)

and increased levels of proteins Clocel_4087, 4088, and 4089 (Table 1) and Clocel_4090 were found.

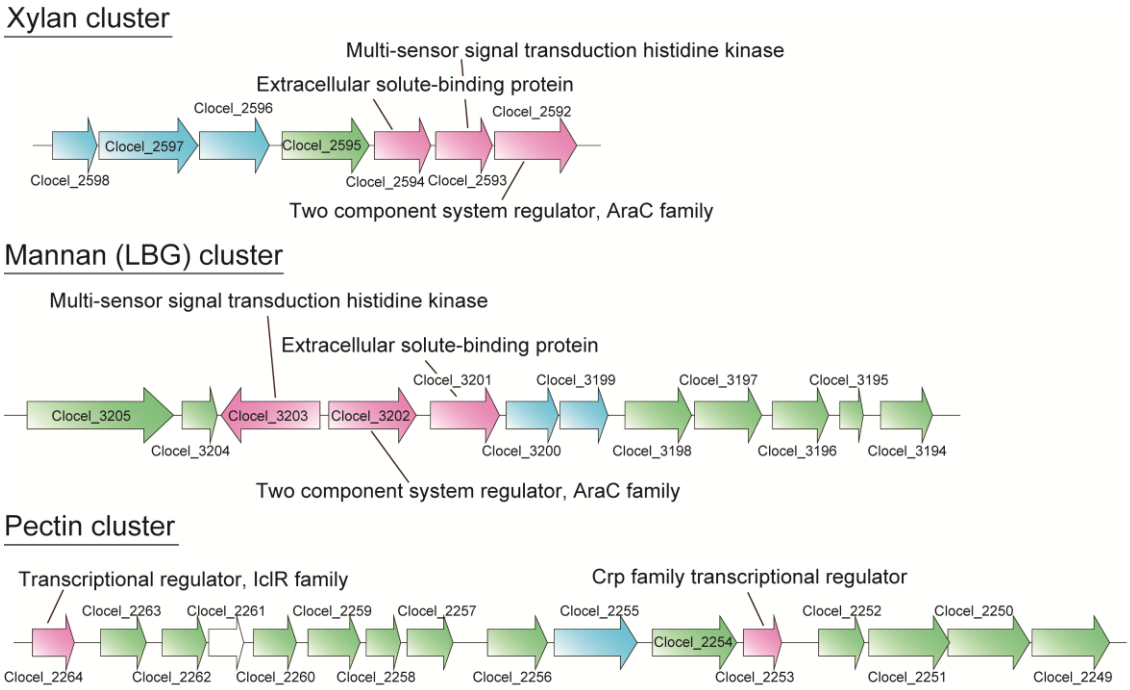


Figure 5. Candidates for gene clusters related to substrate recognition

I carried out genome analysis and identified candidates for gene clusters involved in TCS. Combining these results and substrate specific proteins, I found substrate-specific gene clusters related to recognition of each substrate. Green arrows indicate metabolism-related proteins, pink arrows indicate signal transduction-related proteins, and blue arrows indicate transport-related proteins. White arrows indicate pseudogenes.

Within the pectin clusters, a Crp family transcriptional regulator (Clocel_2253) was identified upstream of the hydrolase/isomerase pathway, and levels of this regulator were significantly increased. Crp family transcriptional regulators are known to regulate gluconate-related genes in *Escherichia coli* (Peekhaus and Conway, 1998), and I propose that this protein similarly regulates the hydrolase/isomerase pathway in *C. cellulovorans*. Upstream of *Clocel_2263*, an IclR family transcriptional regulator encoded genes (*Clocel_2264*) was annotated (Figure 5). In our quantitative proteome analysis, I did not identify this protein, but in the qualitative proteome analysis, this protein was identified

and displayed higher peptide spectrum matches in pectin than in xylan and galactomannan. The KdgR protein regulates the degradation of pectin in *Erwinia chrysanthemi*, and is in the IclR family transcriptional regulator group (Molina-Henares et al., 2006). BLAST analysis indicates that Clocel_2264 has similarity with the KdgR protein (e-value = 5E-48). Based on these findings, I proposed that IclR family transcriptional regulators control the expression of the lyase/5-dehydro-4-deoxy-gluconate pathway.

Based on the metabolism-related protein profiles and cluster analysis performed in the present study, I propose molecular models for substrate recognition of *C. cellulovorans* in hemicellulose (xylan and galactomannan) and pectin (Figure 6). For hemicellulose, polysaccharides (xylan and galactomannan) are first degraded to oligosaccharides (xylooligosaccharides and mannooligosaccharides) in the extracellular environment, and these oligosaccharides are recognized by extracellular solute binding proteins (Clocel_2594 ; Clocel_3201 in Table 1). Oligosaccharide-binding proteins then transduce signals to integral membrane histidine kinases (Clocel_2593 and 3203). Next, intracellular transcriptional regulators (Clocel_2592 in Table 1; Clocel_3202) are phosphorylated by the activation of integral membrane histidine kinases. Finally, genes related to the degradation and metabolism of hemicellulose are upregulated, allowing *C. cellulovorans* to metabolize substrates more efficiently (Figure 6A).

For pectin, two regulators (Clocel_2253 in Table 1; Clocel_2264) work in substrate recognition. I predict that oligosaccharides derived from pectin or other pectin metabolites bind to these regulators allowing oligosaccharides to be transported into the cell (Figure 6B). This is a similar system to the pectin-recognition system in *E. chrysathemi* (Molina-Henares et al., 2006), but *E. chrysathemi* does not have two regulators operating hydrolase/isomerase and lyase/5-dehydro-4-deoxy-gluconate pathways.

The *C. cellulovorans* recognition system for xylan is very similar to that of *C. cellulolyticum*, but based on previous research (Xu et al., 2013), the composition of clusters is clearly different. In the *C. cellulolyticum* cluster, metabolism-related enzymes were not included. I therefore predict that these genes may not be regulated by TCS (Xu et al., 2013). However, the *C. cellulovorans* cluster includes metabolism- and transport-

related genes. Thus, I speculate that the recognition system for xylan in *C. cellulovorans* regulates both the production of degradation- and transport-related enzymes and metabolism-related enzymes.

Substrate recognition systems in *C. cellulovorans* are one of the survival strategies. In other bacteria, monosaccharides such as glucose and xylose are mainly transported, but this bacterium recognizes oligosaccharides of hemicellulose and pectin, and can optimize degradation-, metabolism-, and transport-related proteins' profiles according to the substrate faster than other bacteria. This feature maybe suitable for competing with other anaerobic bacteria.

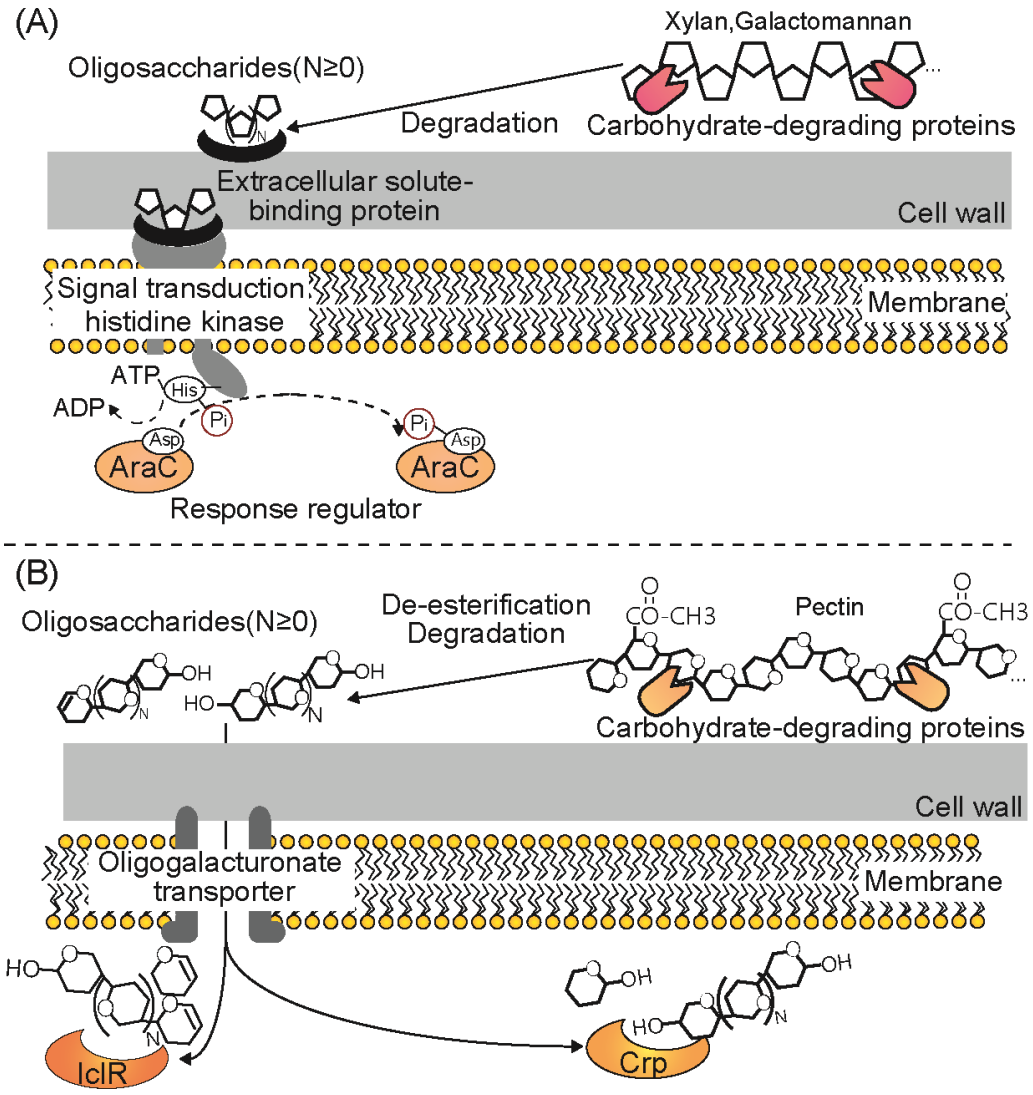


Figure 6. Proposed *C. cellulovorans* substrate recognition systems of hemicellulose and pectin

(A)For hemicellulose, polysaccharides are degraded to derived oligosaccharides outside of the cell, and extracellular solute-binding proteins bind these substrates. Solute-binding proteins induce signal-to-signal transduction integral membrane histidine kinases, and the activated kinases phosphorylate transcriptional regulator AraC, and target genes (shown in Figure 5) are upregulated. (B) For pectin, polysaccharides are de-esterified by pectinesterase and degraded by pectin lyase. Next, derived oligosaccharides are transported into cell, and these or other metabolites bind to a transcriptional regulator. Then, genes belonging to the target cluster (shown in Figure 5) are upregulated.

Summary

In conclusion, I performed “intracellular” quantitative proteome analysis of *C. cellulovorans* with glucose, xylan, galactomannan, and pectin and identified 734 proteins with LC-MS/MS equipped long monolithic silica capillary columns. I focused on substrate-recognition-, degradation-, and metabolism- related proteins, and carried out cluster analysis based on genome analysis together with the intracellular proteome analysis and the previous “exoproteome” analysis (Esaka et al., 2015; Matsui et al., 2013). Based on our results, I propose a mechanism for the recognition of hemicellulose and pectin in *C. cellulovorans*. For hemicellulose recognition, *C. cellulovorans* uses TCS and regulates all substrate-related genes with a single regulator. In the pectin recognition system, *C. cellulovorans* regulates hydrolase/isomerase and lyase/5-dehydro-4-deoxy-gluconate pathways with two regulators, Clocel_2253 and Clocel_2264, respectively. Judging from the presence of an oligogalacturonates transporter in the pectin regulatory clusters, the regulation of pectin metabolism-related genes appears to be controlled by the oligogalacturonates derived from pectin, or other metabolic products. This system has not been described for other cellulosome-producing Clostridia. Due to the multiplicity of regulator proteins involving substrate recognition, I suggest that *C. cellulovorans* has more distinct promoter regions that respond to each substrate. This information should be helpful for the elucidation of the environmental sensing strategy of *C. cellulovorans*.

References

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res *37*, D233-D238.

de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

Esaka, K., Aburaya, S., Morisaka, H., Kuroda, K., and Ueda, M. (2015). Exoproteome analysis of *Clostridium cellulovorans* in natural soft-biomass degradation. AMB Express *5*, 1-8.

Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W., and Reverchon, S. (1996). Regulation of pectinolysis in *Erwinia chrysanthemi*. Annu Rev Microbiol *50*, 213-257.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kawahara, R., Saburi, W., Odaka, R., Taguchi, H., Ito, S., Mori, H., and Matsui, H. (2012). Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, β-1,4-mannooligosaccharide phosphorylase. J Biol Chem *287*, 42389-42399.

Matsui, K., Bae, J., Esaka, K., Morisaka, H., Kuroda, K., and Ueda, M. (2013). Exoproteome profiles of *Clostridium cellulovorans* grown on various carbon sources. Appl Environ Microbiol *79*, 6576-6584.

Molina-Henares, A.J., Krell, T., Guazzaroni, M.E., Segura, A., and Ramos, J.L. (2006). Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. FEMS Microbiol Rev *30*, 157-186.

Morisaka, H., Matsui, K., Tatsukami, Y., Kuroda, K., Miyake, H., Tamaru, Y., and Ueda, M. (2012). Profile of native cellulosomal proteins of *Clostridium cellulovorans* adapted to various carbon sources. AMB Express *2*, 1-5.

Peekhaus, N., and Conway, T. (1998). Positive and negative transcriptional regulation of the *Escherichia coli* gluconate regulon gene gntTby GntR and the cyclic AMP (cAMP)-

cAMP receptor protein complex. J Bacteriol *180*, 1777-1785.

Petitdemange, E., Caillet, F., Giallo, J., and Gaudin, C. (1984). *Clostridium cellulolyticum* sp. nov., a cellulolytic, mesophilic: species from decayed grass. Int J Syst Bacteriol *34*, 155-159.

Prawitwong, P., Waeonukul, R., Tachaapaikoon, C., Pason, P., Ratanakhanokchai, K., Deng, L., Sermsathanaswadi, J., Septiningrum, K., Mori, Y., and Kosugi, A. (2013). Direct glucose production from lignocellulose using *Clostridium thermocellum* cultures supplemented with a thermostable *β*-glucosidase. Biotechnol Biofuels *6*, 184.

Raman, B., Pan, C., Hurst, G.B., Rodriguez Jr, M., McKeown, C.K., Lankford, P.K., Samatova, N.F., and Mielenz, J.R. (2009). Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PloS ONE *4*, e5271.

Richard, P., and Hilditch, S. (2009). D-Galacturonic acid catabolism in microorganisms and its biotechnological relevance. Appl Microbiol Biotechnol *82*, 597-604.

Sañudo-Wilhelmy, S.A., Gómez-Consarnau, L., Suffridge, C., and Webb, E.A. (2014). The role of B vitamins in marine biogeochemistry. Annu Rev Mar Sci *6*, 339-367.

Saldanha, A.J. (2004). Java Treeview—extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., and Wasaki, J. (2011). New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. Biochem Biophys Res Commun *408*, 701-706.

Sleat, R., Mah, R.A., and Robinson, R. (1984). Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. Appl Environ Microbiol *48*, 88-93.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*.

Tamaru, Y., Miyake, H., Kuroda, K., Nakanishi, A., Kawade, Y., Yamamoto, K., Uemura, M., Fujita, Y., Doi, R.H., and Ueda, M. (2010). Genome sequence of the cellulosome-producing mesophilic organism *Clostridium cellulovorans* 743B. J Bacteriol *192*, 901-

902.

Xu, C., Huang, R., Teng, L., Wang, D., Hemme, C.L., Borovok, I., He, Q., Lamed, R., Bayer, E.A., and Zhou, J. (2013). Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. Biotechnol Biofuels *6*, 73.

Chapter III

Temporal proteome dynamics of *Clostridium cellulovorans* cultured with major plant cell wall polysaccharides

In chapter II, I performed comparative proteome analysis to elucidate hemicellulose and pectin recognition mechanisms of *C. cellulovorans* with optimized proteome analysis. As a result, I could suggest plant cell wall polysaccharides recognition mechanisms in *C. cellulovorans*. On the other hand, this analysis could reveal temporal proteome dynamics because I performed end-point analysis. Temporal proteome dynamics of *C. cellulovorans* will be beneficial for the further understanding of its degradation strategy and improving production of biofuels.

In this chapter, I cultured *C. cellulovorans* with five plant cell wall polysaccharides (glucose, cellulose, xylan, galactomannan and pectin) and performed temporal proteome analysis at five time points to understand its metabolism and degradation strategy.

Materials and methods

Bacterial strain and culture medium

C. cellulovorans 743B (ATCC35269) was purchased from American type culture collection. The culture medium was prepared as previously described (Sleat et al., 1984), except for the carbon sources; 0.3% glucose (Nacalai Tesque, Kyoto, Japan), 0.3% Sigmacell Cellulose Type 20 (Sigma, MO, USA), 0.3% xylan from Beechwood (SERVA Electrophoresis GmbH, Heidelberg, Germany), 0.3% locust bean gum from *Ceratonia siliqua* seeds (Sigma) used as galactomannan, and 0.3% pectin from apple (Sigma). *C. cellulovorans* was grown at 37 °C, pH 7.5, in 1-L glass anaerobic fermenters (three biological replicates). During the culture process, the culture media were fractionated and the ATP concentration was measured to determine growth curves (Inamori et al., 2016). For proteome analysis, 40 mL of culture media was collected at each predetermined TP.

The collected media were centrifuged at 6,000 × *g*, at 4 °C for 10 min, and the precipitated cells and supernatants were separated. The collected cells were washed three times with PBS (137 mM NaCl, 8.1 mM Na_2HPO_4, 2.68 mM KCl, 1.47 mM KH_2PO_4, pH 7.4; NIPPON GENE, Tokyo, Japan). The washed cells and culture supernatants were flash-frozen and stored at -80 °C until use.

Preparation of cellular proteins for proteome analysis

The preparation of cellular proteins was conducted according to a previous report (Masuda et al., 2008), with slight modifications. In brief, the frozen cells were dissolved in 200 µL of lysis buffer (12 mM sodium deoxycholate; Wako, Osaka, Japan and 12 mM sodium lauryl sulfate (Nacalai Tesque) in 50 mM Tris-HCl, pH 7.5), and disrupted by sonication at 250 W for 10 min (Bioruptor UCD-250T; Cosmo Bio, Tokyo, Japan). The solution was centrifuged at 13,000 × *g* at 4 °C for 20 min, and the supernatant was collected. The proteins were precipitated using methanol–chloroform precipitation, as described previously (Wessel and Flügge, 1984). The precipitated solution was dissolved in 200 µL of reaction buffer (12 mM sodium deoxycholate, 12 mM sodium lauryl sulfate, and 50 mM triethylammonium bicarbonate [TEAB; Sigma], pH 8.5), and reduced by adding dithiothreitol (Nacalai Tesque) to 50 mM using 1 M stock solution of dithiothreitol, and the reductive reaction proceeded for 30 min at 37 °C. Following the incubation, iodoacetamide (Wako) was added until the solution was at a concentration of 50 mM using a 500 mM stock solution of iodoacetamide, and the alkylation reaction was continued for 30 min at room temperature in the dark. Two micrograms of Lysyl Endopeptidase, Mass Spectrometry Grade (Wako) was added and the solution was incubated at 37 °C for 4 h. Two micrograms of Sequencing Grade Modified Trypsin (Promega, Madison, WI, USA) and 1,000 µL of TEAB were added and the solution was incubated overnight at 37 °C. To remove the surfactant, 400 µL of ethyl acetate (Wako) was added to 400 µL of the digested solution. Eight microliters of trifluoroacetic acid (Wako) was added for acidification, and the mixture was shaken for 1 min. The shaken mixture was centrifuged at 16,000 × *g* for 2 min to separate the aqueous and organic

phases. The aqueous phase was collected and desalted using Monospin C18 (GL science, Osaka, Japan) according to the manufacturer's protocol.

Preparation of the secreted proteins for proteome analysis

The frozen supernatant was dissolved on ice and was subjected to ultrafiltration using an Amicon Ultra-15 centrifugal filter (Merck Millipore, MA, USA) to concentrate the secreted proteins as previously described (Esaka et al., 2015; Matsui et al., 2013). The concentrated proteins were precipitated using methanol–chloroform precipitation. Reductive alkylation, digestion to the peptides, and cleaning up of the peptides are described in the above section.

Tandem mass tag labeling

The desalted samples were dissolved in 50 µL of TEAB. The dissolved peptide solution was labeled using a TMT 6-plex labeling kit (Thermo Fisher Scientific, MA, USA) according to the manufacturer's protocol with a slight modification. The TMT-labeling reagents were dissolved in 41 µL of acetonitrile and mixed with 10 µL of TMT-labeling reagents to 100 µL. Each digested peptide solution was mixed with 10 µL of TMT-labeling reagents as described in Figure 1. The tryptic digests of the lag, early log, middle log, late log, and stationary phases were labeled with TMT-126, -127, -128, -129, and -130, respectively. In addition, an equal volume of all samples was mixed, and the mixture was labeled with TMT-131 as an internal standard. After the labeling reaction, held at room temperature for 60 min, the reactants were quenched by the addition of 8 µL of 5% hydroxylamine and then lyophilized after mixing with 30 µL of each reactant. The dried samples were resolved in 30 µL of 0.1% formic acid.

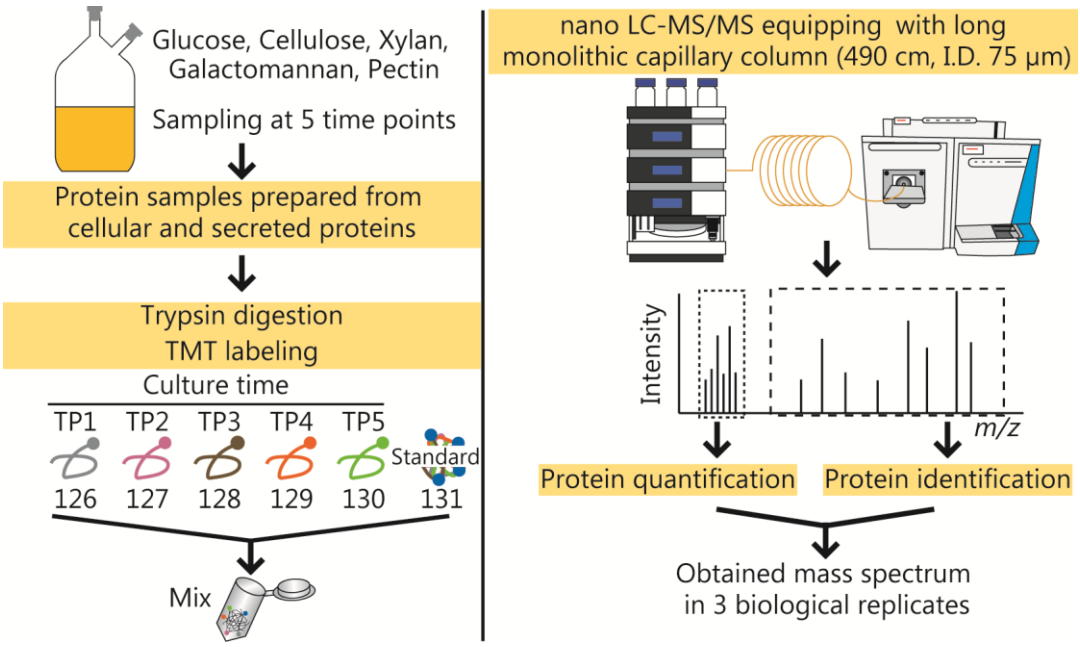


Figure 1 Experimental procedure of temporal quantitative proteome analysis

C. cellulovorans was cultured with glucose, cellulose, xylan, galactomannan, or pectin. Proteins were prepared from cell lysates at various time points. Tryptic fragments were labeled with tandem mass tag (TMT). The labeled peptides were mixed and analyzed with the nano liquid chromatography (LC) -mass spectrometry (MS/MS) system and the obtained mass spectra were used for protein identification and quantification. TP indicates time points at which the proteome samples were collected.

LC-MS/MS analysis

The proteome analysis was conducted according to a previous chapter, with some slight modifications. Five microliters of the labeled samples were injected into the LC-MS/MS system. Proteome analysis was performed using an LC (Ultimate 3000; Thermo Fisher Scientific)-MS/MS (LTQ Orbitrap Velos Mass Spectrometer; Thermo Fisher Scientific) system equipped with a long monolithic silica capillary column (Morisaka et al., 2012) (490-cm long, 75-μm internal diameter) at a flow rate of 280 nL/min. A gradient was achieved by changing the ratio of the two eluents; eluent A, 0.1% (v/v) formic acid; eluent B, 80% acetonitrile containing 0.1% (v/v) formic acid. The gradient started with 5% B, increased to 45% for 750 min, further increased to 95% B to wash the column for

140 min, returned to the initial condition, and was held for re-equilibration of the column. The separated analytes were detected using a mass spectrometer with a full scan range of 350–1,500 *m/z* (resolution, 60,000), followed by 10 data-dependent higher-energy c-trap dissociation MS/MS scans acquired for TMT-reporter ions by using 35% normalized collision energy. The temperature of the ion transfer tube was set to 280 °C and the dynamic exclusion was 180 s. An electrospray ionization voltage was set at 2.3 kV.

Data analysis

Data analysis was performed using Proteome Discoverer 2.1 (Thermo Fisher Scientific). Protein identification was performed using MASCOT (Matrix Science, London, UK) against the *C. cellulovorans* protein database from NCBI, with a precursor mass tolerance of 20 ppm and a fragment ion mass tolerance of 50 mmu. Carbamidomethylation of cysteine and TMT 6-plex of N-term and lysine were set as fixed modifications, and oxidation of methionine was set as a dynamic modification. The data were then filtered with the cutoff criteria of ≤0.01 (*q*-value), corresponding to a 1% false discovery rate on a spectrum level. Proteins with missing values only in the lag phase or with no missing values in the three triplicates were accepted as identified proteins. The global median normalization was performed to normalize the amount of tryptic digests injected into the mass spectrometer in cellular protein. For k-means clustering of secreted proteins, all quantitative values were divided with $\log_{10}$ (Total ATP concentration of each TP) to normalize the amount of proteins in each TP. For calculation of fold change relative to glucose, I used permutation-based false discover rate calculation in Perseus (Tyanova et al., 2016) and the differences were plotted as fold changes. *k*-means clustering was conducted using R software. Functional annotation was performed using the KEGG BRITE annotation tool (Kanehisa et al., 2011; Mao et al., 2005). The coefficient correlation of each quantitative value was calculated using Pearson's correlation coefficient, and the protein co-production network was visualized by Cytoscape 3.5.0 (Shannon et al., 2003).

Results and discussion

Growth curve analysis

The experimental workflow is described in Figure 1. First, I measured the growth curves of *C. cellulovorans* cultured with each carbon source (Figure 2) as performed previously (Inamori et al., 2016; Miyake et al., 2016). *C. cellulovorans* grew on all carbon sources and I found that *C. cellulovorans* degraded and metabolized galactomannan and produced ATP more efficiently on this carbon source than on others. In contrast, *C. cellulovorans* degraded cellulose very slowly. Next, I determined five sampling points as follows: lag phase (time point 1 [TP1]), early log phase (TP2), middle log phase (TP3), late log phase (TP4), and stationary phase (TP5) (Table 1). At each TP, I collected the culture media for proteome analysis.

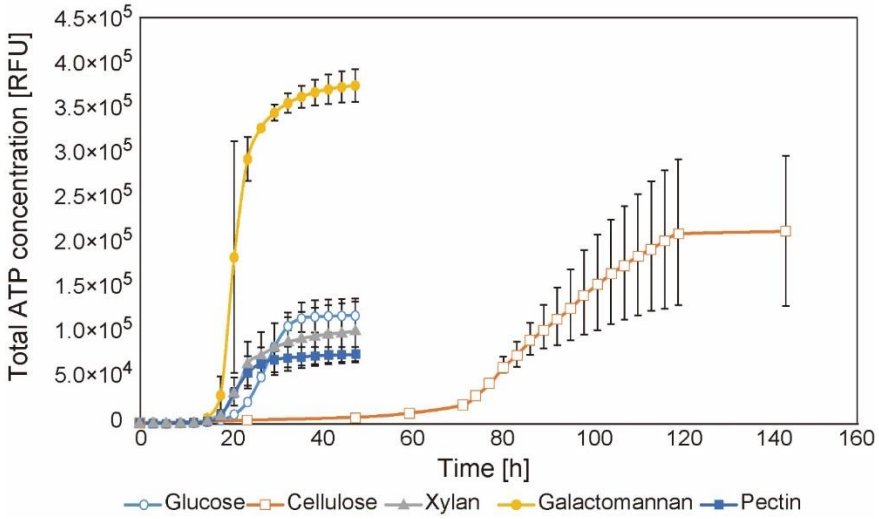


Figure 2 Confirmation of growth curves of *C. cellulovorans* cultured with five different carbon sources

The growth curves of *C. cellulovorans* were measured by accumulated ATP concentration with five different carbon sources, glucose (open circle), cellulose (open square), xylan (closed triangle), galactomannan (closed circle), and pectin (closed square). Error bars indicate mean ± standard deviations (n = 3).

Table 1 Culture time of each growth phase

	Time point 1	Time point 2	Time point 3	Time point 4	Time point 5
Growth phase	Lag	Early log	Middle log	Late log	Stationary
Glucose	13.5 h	21 h	24 h	30 h	39 h
Cellulose	24 h	60 h	75 h	81 h	144 h
Xylan	12 h	15 h	21 h	24 h	36 h
Galactomannan	12 h	18 h	19.5 h	21 h	27 h
Pectin	12 h	16.5 h	18 h	21 h	30 h

Proteome analysis

Proteome samples from cellular and secreted proteins were prepared from three biological replicates in each carbon source. I performed quantitative proteome analysis using TMT labeling and capillary monolith nano liquid chromatography (LC) -tandem mass spectrometry (MS/MS). From this analysis, I identified a total of 1,895 cellular proteins. Among them, 865 cellular proteins were common to all carbon sources. In addition, I identified a total of 879 secreted proteins, 361 of which were common to all carbon sources (Figure 3). Our proteome analysis covered approximately 50% of the genomic products of *C. cellulovorans*; this proteome coverage is the highest in *C. cellulovorans* proteome researches (Esaka et al., 2015; Matsui et al., 2013).

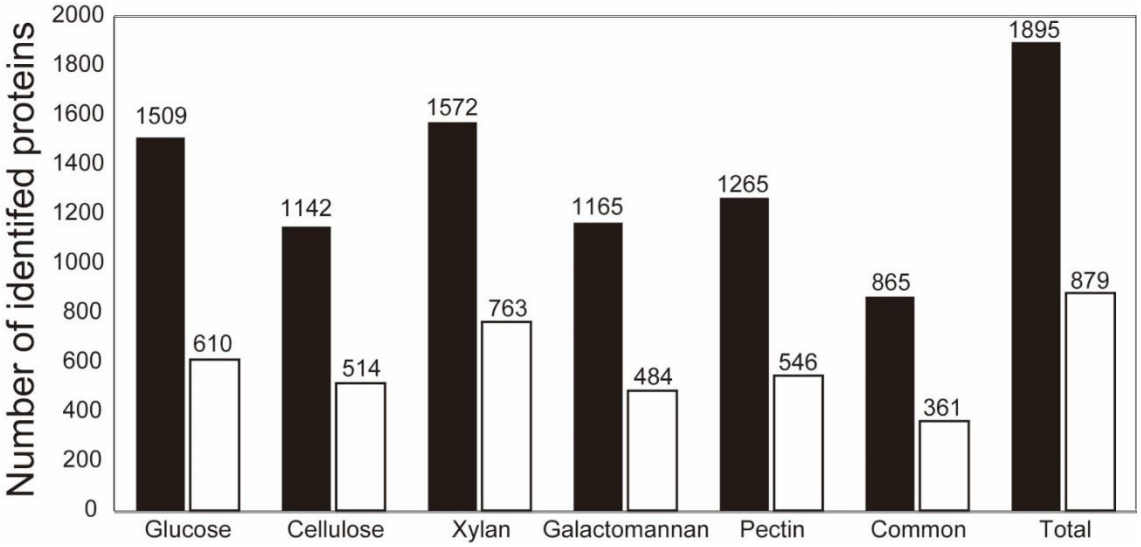


Figure 3 Number of identified cellular and secreted proteins

Number of identified proteins (cellular, closed bar; secreted, open bar). 'Common' indicates proteins identified in all carbon sources, and 'Total' indicates proteins identified in any carbon source.

Uniquely identified proteins of each carbon source

I assumed that the uniquely identified proteins in each carbon source (Table 2) are important for assimilation of the substrates. Among them, I found uniquely-identified carbohydrate-degrading enzymes and cellulosomal proteins among the cellular and/or secreted proteins.

Table 2 Number of uniquely identified proteins in each carbon source

	Cellular protein	Secreted protein
Glucose	62	31
Cellulose	50	13
Xylan	176	138
Galactomannan	19	12
Pectin	32	13

Enzymes possibly related to carbohydrate degradation among the uniquely-identified proteins are shown in Table 3. Among the proteins uniquely identified in the cellulose sample, I identified two cellulosomal proteins, PeeB and Hyp3, and five non-cellulosomal proteins. The activities of these proteins are unknown and further research is required.

As uniquely-identified proteins in the xylan sample, I identified five non-cellulosomal proteins. All of them have not been annotated as xylan degradation-related proteins, and these proteins belong to the GH4, GH13, and PL9 families, isomerase, and hypothetical proteins. Among them, two proteins (belonging to the GH4 and PL9 families) are annotated as pectin-degrading enzymes in the carbohydrate-active enzymes (CAZy) database (Lombard et al., 2013). The main component of xylan from hardwood is *O*-acetyl-4-*O*-methylglucuronoxylan (Ebringerová et al., 2000), and the side chain cannot be broken with xylanase and xylosidase. Therefore, some pectin-degrading enzymes might be required to degrade the side chain of *O*-acetyl-4-*O*-methylglucuronoxylan. To our knowledge, it has not been reported that the PL9 and GH4 family proteins are required for degradation of xylan or *O*-acetyl-4-*O*-methylglucuroxylan (Lombard et al., 2013; Polizeli et al., 2005). This implies that *C. cellulovorans* changes the protein profile of carbohydrate-degrading proteins depending not only on the main chain but also on the side chain of the polysaccharides.

Table 3 Enzymes possibly related to carbohydrate degradation among the uniquely identified proteins

Cellulose

Accession	Description	Protein name	CAZy family (Lombard et al., 2013)
gi302576230	**Dockerin type 1**	**PeeB**	**PL11**
gi302578484	**Dockerin type 1**	**Hyp3**	**-**
gi302577108	Pullulanase, type I	pullulanase	CBM48,GH13
gi302576119	Glycosyl transferase family 2	-	GT2
gi302578575	Glycogen/starch synthase, ADP-glucose type	-	GT5
gi302577027	Glycogen/starch/alpha-glucan phosphorylase	-	GT35
gi302576735	Ig domain protein group 2 domain protein	-	CBM32

Xylan

Accession	Description	Protein name	CAZy family
gi302576819	Glycoside hydrolase family 4	Alpha-galactosidases	GH4
gi302577083	Carbohydrate-binding CenC domain protein	-	CBM4
gi302579889	Protein of unknown function DUF1565	Pectate	PL9
gi302579926	Alpha amylase catalytic region	Alpha-amylase	GH13

Galactomannan

Accession	Description	Protein name	CAZy family
gi302577082	Beta-galactosidase	-	GH1
gi302577679	Mannan endo-1,4-beta-mannosidase	-	CBM35,GH26
gi302578777	Protein of unknown function DUF291	ManGH2B	GH5,GH5,CBM23
gi302579006	Hypothetical protein Clocel_3338	-	GH113
gi302579773	Mannan endo-1,4-beta-mannosidase	Beta-mannanase	GH26,CBM59

Pectin

Accession	Description	Protein name	CAZy family
gi302576984	Alpha-N-arabinofuranosidase	-	GH43
gi302578311	Glycosyl hydrolase family 88	-	GH88
gi302578753	Protein of unknown function DUF1680	-	GH127
gi302578333	Pectinesterase	exopolygalacturonate	CE8,PL9

*The bold proteins are cellulosomal proteins.

***k*-means clustering analysis**

To elucidate the temporal proteome dynamics in each carbon source, I performed non-hierarchical *k*-means clustering analysis. In *k*-means clustering analysis, I grouped the cellular and secreted proteins in each carbon source into four clusters. As a result, in the cellular proteins, I found that these clusters in each carbon source seemed to show the following five trends: most increasing (cluster A), slightly increasing (cluster B), stable (cluster C), slightly decreasing (cluster D), and decreasing (cluster E) (Figure 4a). In the secreted proteins, almost all protein clusters showed an increasing trend, but there were a few decreasing protein clusters in cellulose, pectin, and galactomannan (Figure 4b).

I performed Kyoto encyclopedia of genes and genomes (KEGG) orthology analysis to categorize the functional classes of each protein cluster with cellular proteins (Kanehisa et al., 2011; Mao et al., 2005). In almost all clusters of each carbon source, the top 3 ranked protein category classes were classified into carbohydrate metabolism, amino acid metabolism, translation, or membrane transport proteins. In contrast, in the cluster A of pectin and xylan, proteins for the metabolism of cofactors and vitamins were ranked as top 1, and in cluster E of xylan, proteins for energy metabolism were ranked as top 1. In a previous chapter, *C. cellulovorans* was found to increase the production of proteins related to cobalamin metabolism when pectin is used as a carbon source. In addition to the previous chapter, our data also suggests that xylan and pectin increased cobalamin-related proteins. Our research cannot reveal the role of cobalamin in xylan and pectin degradation or metabolism. The increase of cobalamin-related proteins might be a secondary effect of the degradation or metabolism of xylan and pectin. I need further investigation into these mechanisms.

From *k*-means clustering of cellular proteins, I suggest that cellulosomal proteins on the cell surface might not change during the degradation of polysaccharides. Most cellulosomal proteins were included in cluster B, C, or D in cellulose, xylan, and pectin, and the expression profiles of these proteins did not change dramatically (Figure 4a). On the other hand, in glucose, about half of the cellulosomal proteins were included in cluster A, and in galactomannan almost all cellulosomal proteins were included in cluster E (Figure 4a).

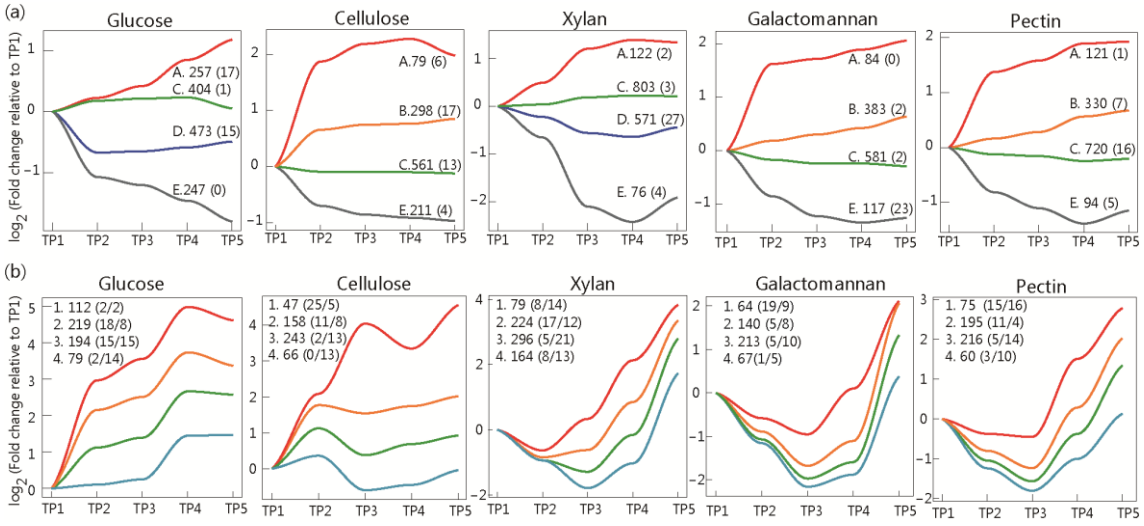


Figure 4 *k*-means clustered proteome changes in cellular and secreted proteins

(a) The cellular protein profiles were grouped by using *k*-means clustering. I grouped the cellular proteins with four clusters, and the protein clusters was divided into five classes as follows: most increasing (cluster A; red), slightly increasing (cluster B; orange), stable (cluster C; green), slightly decreasing (cluster D; blue), and decreasing (cluster E; gray) clusters. The number shown in cluster bars indicates the protein belonging to its cluster, and the number shown in parentheses indicates the number of identified cellulosomal proteins in each cluster. (b) The secreted protein profiles were grouped with *k*-means clustering. I grouped the cellular proteins with into four clusters, and the protein clusters were divided into four classes (1st [red], 2nd [orange], 3rd [green], and 4th [blue] clusters). The numbers on the graph indicate the number of proteins belonging to each cluster, and the number shown in parentheses indicates the number of identified cellulosomal (before slash) and non-cellulosomal (after slash) proteins.

In galactomannan-cultured *C. cellulovorans*, released cell-free cellulosome might have an important role in cell wall degradation. In the secreted proteins of galactomannan-cultured *C. cellulovorans*, the most increasing cluster included 19 cellulosomal proteins (Figure 4b), and in the cellular proteins of galactomannan-cultured *C. cellulovorans*, almost all cellulosomal proteins showed decreasing trend (Figure 4a). In other *Clostridium* species, it is reported that cell-free cellulosome has degradation activity (Mohand-Oussaid et al., 1999; Xu et al., 2016). Cell-free cellulosome is a cellulosome

which does not bond to cell surfaces (Xu et al., 2016). In *C. cellulovorans*, cell-free cellulosome has not been reported. However, in previous secreted proteome analysis, the cellulosomal proteins are more abundant than non-cellulosomal proteins (Matsui et al., 2013), and a higher xylanolytic activity was reported in the cell-free fraction than in the cell-attached fraction of *C. cellulovorans* (Kosugi et al., 2001). Furthermore, in our analysis, the most increasing cluster of cellulose, galactomannan, and pectin included more cellulosomal proteins than other clusters (Figure 4b). I postulated that the released cell-free cellulosomal proteins in *C. cellulovorans* degrade plant cell wall polysaccharides.

Of the secreted proteins, 25 in the cluster with the most increasing trend of cellulose were cellulosomal proteins, and five proteins in the cluster with the most increasing trend were non-cellulosomal proteins (Figure 4b). I postulated that cellulosomal proteins were mainly used but that non-cellulosomal proteins were changed secondarily for the degradation of cellulose. In a previous report, *C. cellulovorans* was reported to increase non-cellulosomal proteins dramatically when cultured with phosphoric-acid-swollen cellulose comparing to cellobiose (Matsui et al., 2013). In contrast to cellulose, other polysaccharides-grown *C. cellulovorans* showed different proportions of cellulosomal and non-cellulosomal proteins at the most increasing protein cluster among secreted proteins (Figure 4b). In glucose, two cellulosomal and two non-cellulosomal proteins were identified in the cluster with the most increasing trend in the secreted proteins. Coordination between cellulosomal proteins and between cellulosomal and non-cellulosomal proteins is important for the degradation of carbon sources (Matsui et al., 2013; Tamaru et al., 2011). In cellulose, clusters 1 and 2 contained 36 cellulosomal proteins and 13 non-cellulosomal proteins. However, in other polysaccharides, cluster 1 and 2 contained 24–26 cellulosomal proteins and 17–26 non-cellulosomal proteins (Figure 4b). More cellulosomal proteins were found in cellulose than in other polysaccharides, and more non-cellulosomal proteins were found in other polysaccharides than in cellulose. Our data implies that cellulosomal proteins might play important roles in the degradation of cellulose, whereas non-cellulosomal proteins may not. On the other hand, both cellulosomal and non-cellulosomal proteins might play an important role in the degradation of hemicellulose. In the clusters with the most increasing

trends among the secreted proteins in xylan, pectin, and galactomannan, each carbohydrate degradation protein demonstrated degradation of each carbon source.

Pathway analysis of cellular proteins in central metabolism

To confirm the response of *C. cellulovorans* to each carbon source, I performed pathway analysis. I plotted the fold changes in glucose at TP1 for each carbon source (Figure 5–10). During glycolysis, the production levels of almost all proteins of all polysaccharides did not change compared to glucose and TP1, but phosphofructokinase was increased about 2–4 fold in cellulose, galactomannan, and pectin between TP1, TP4, and TP5 (Figure 5). Phosphofructokinase controls glycolysis, and in *C. thermocellum* cultured in cellobiose, phosphofructokinase increases about 2-fold in the stationary phase compared to the exponential phase (Rydzak et al., 2012). Our analysis confirmed this profile when *C. cellulovorans* was cultured in all carbon sources. In the pentose phosphate pathway, one transketolase (gi302577000) increased more than 2-fold at TP3, TP4 and TP5 compared to TP1 in pectin, and transaldolase (gi302576999) also increased approximately 2-fold in pectin. In contrast, in the time course analysis of xylan, one transaldolase (gi302576999) decreased 2-fold, but another transaldolase (gi302576350) increased 4-fold (Figure 6). These expression levels might reflect enzyme specificity or function for these two isozymes. In the TCA cycle, the acetate-metabolizing pathway was activated in pectin compared to glucose. Pectin has some acetyl groups (Mohnen, 2008), and its acetyl group might be released with acetyl esterase; therefore, enzymes in the acetate-metabolizing pathway might increase (Figure 7). Central metabolism enzymes were not changed dramatically, although different major plant cell wall polysaccharides were utilized as carbon sources.

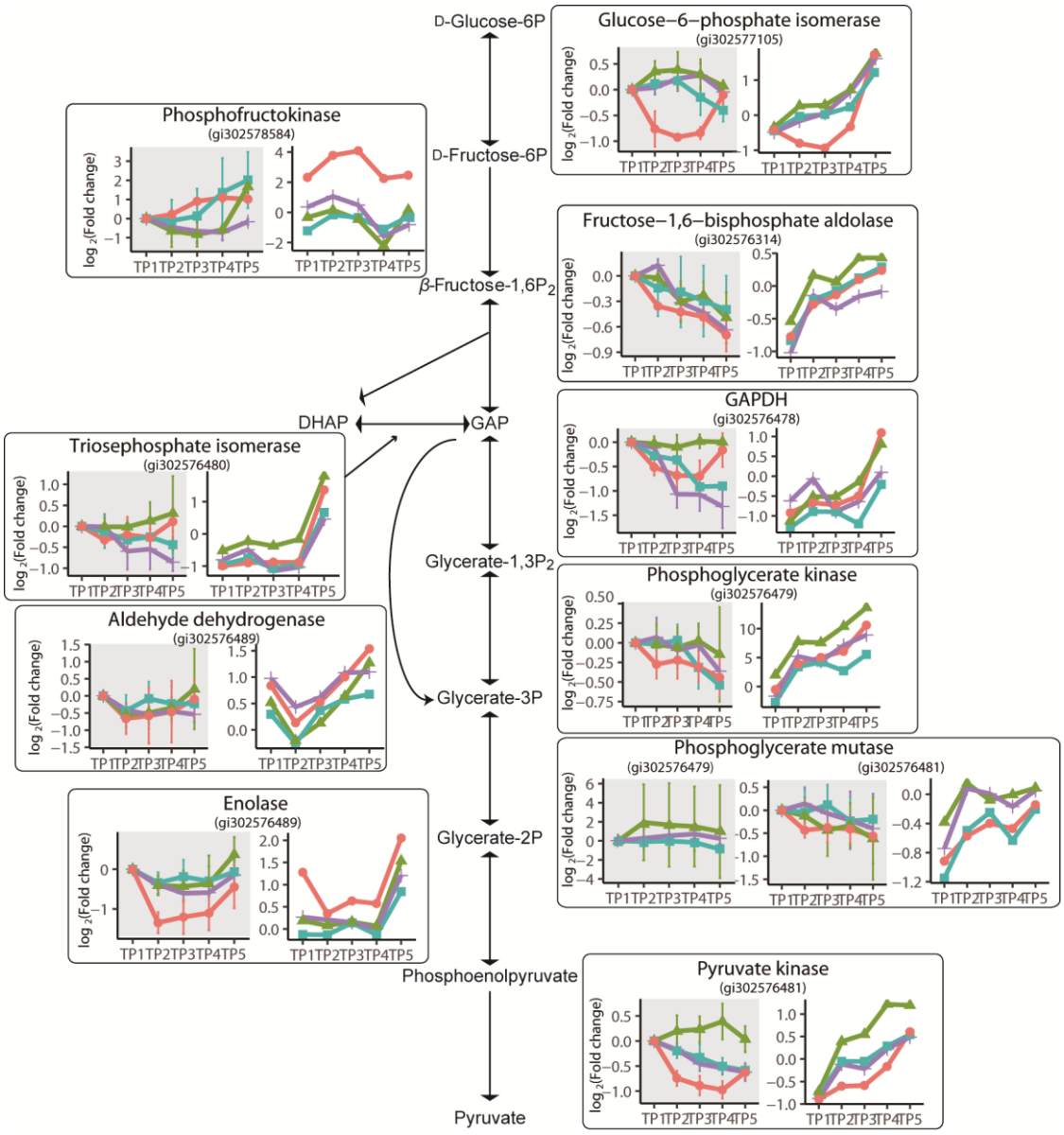


Figure 5 Pathway analysis of glycolysis

Glycolysis pathway was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

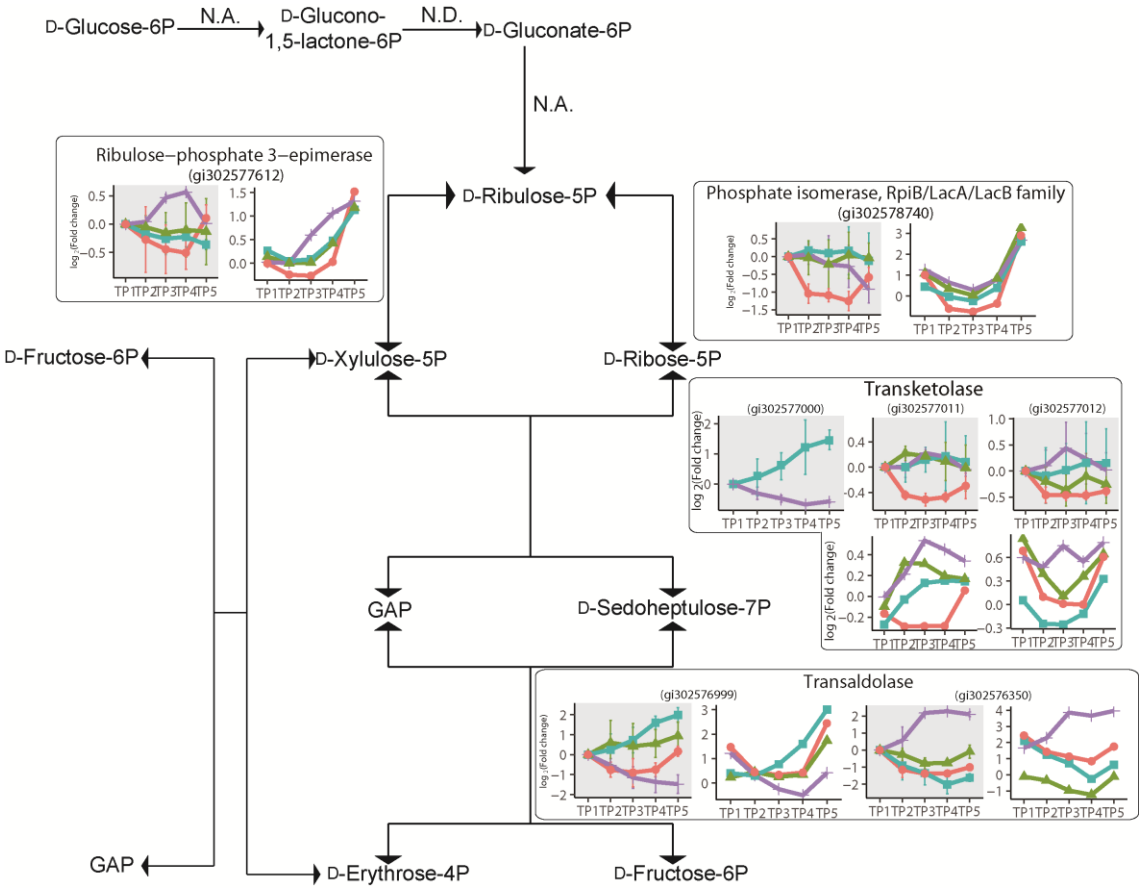


Figure 6 Pathway analysis of pentose phosphate pathway

Pentose phosphate pathway was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

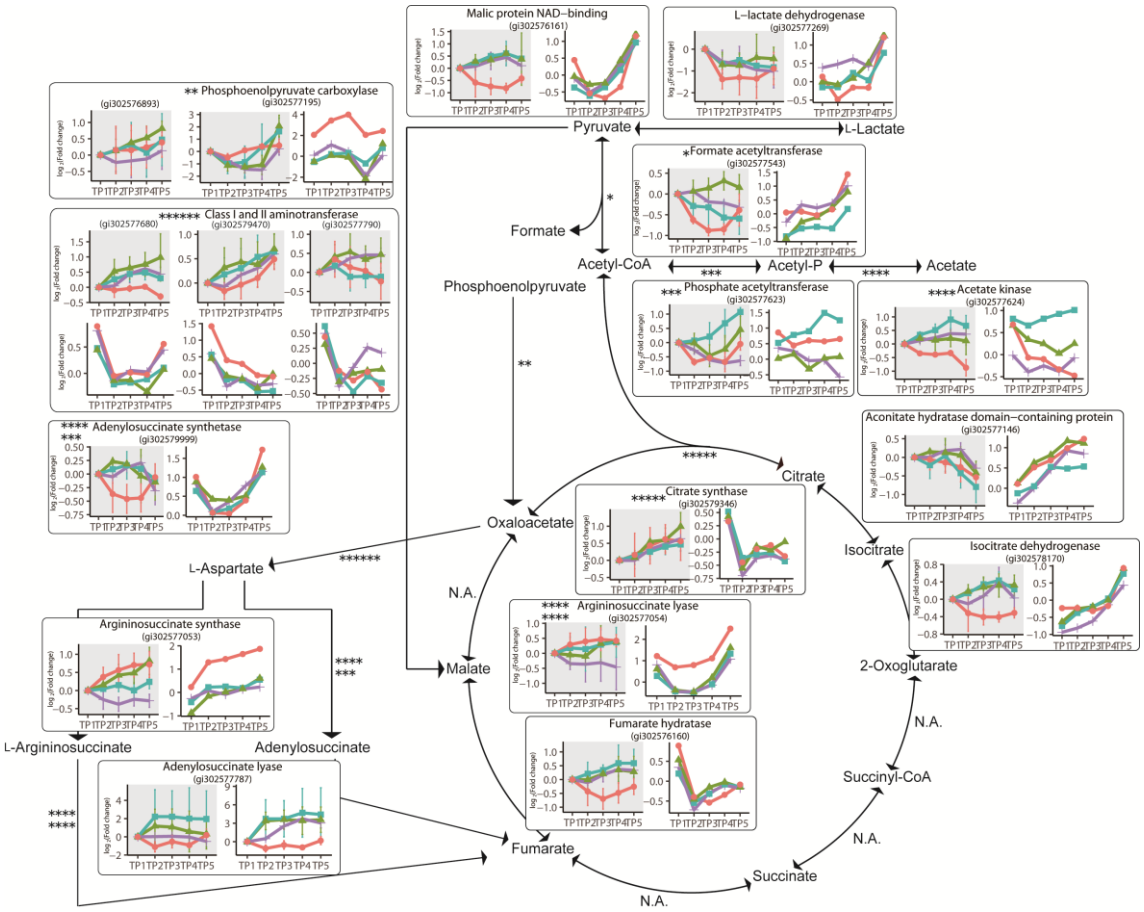


Figure 7 Pathway analysis of tricarboxylic acid cycle

Tricarboxylic acid cycle was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

Pathway analysis of cellular proteins in major plant cell wall polysaccharides metabolism

In the metabolic pathway of xylan, metabolic enzymes of xylan were increased compared to glucose when xylan is used as a carbon source. In other carbon sources, almost all metabolic enzymes of each carbon source were also increased (Figure 8–10).

In the galactomannan degradation pathway, galactokinase and galactose-1-phosphate uridylyltransferase were up-produced in pectin-cultured *C. cellulovorans*

(Figure 9 b). In pectin, rhamnogalacturonan-1, which has a galactose as a constituent sugar residue, was identified. I identified a rhamnogalacturonan-1-degrading enzyme in our secreted proteome analysis of galactomannan-cultured *C. cellulovorans*. Therefore, I speculate that galactose derived from rhamnogalacturonan-1 is metabolized in this pathway.

In the pectin-degrading and -metabolizing pathway, all proteins (except for pectinesterase and pectate lyase) were not up-produced in the cellular proteins compared to glucose (Figure 10). However, pectinesterase and pectate lyase were identified as secreted proteins in pectin. Pectin (polysaccharides of galacturonate) might be degraded to oligogalacturonides (oligosaccharides of galacturonate) by the secreted proteins.

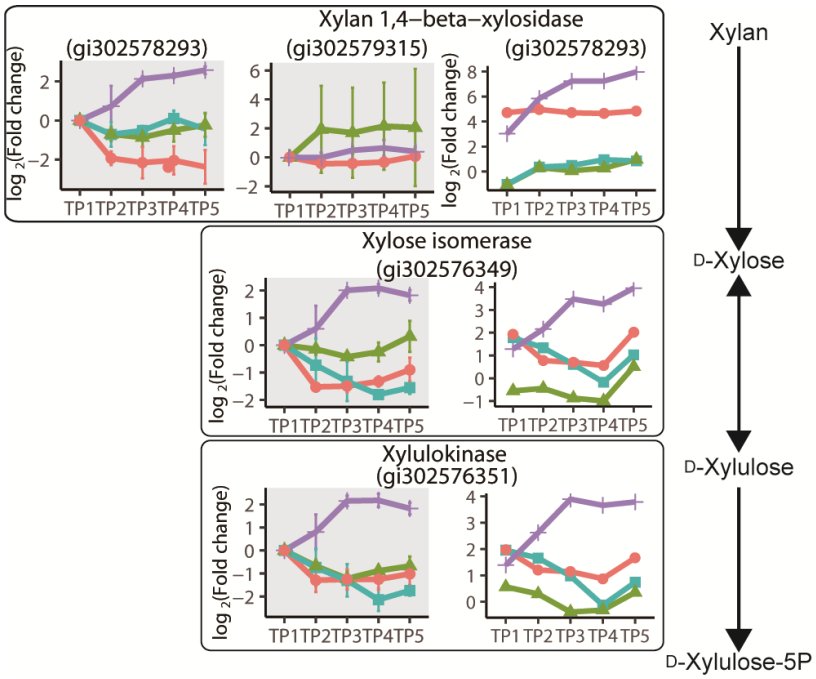


Figure 8 Pathway analysis of xylan degradation and metabolism

Pathway of xylan degradation and metabolism was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

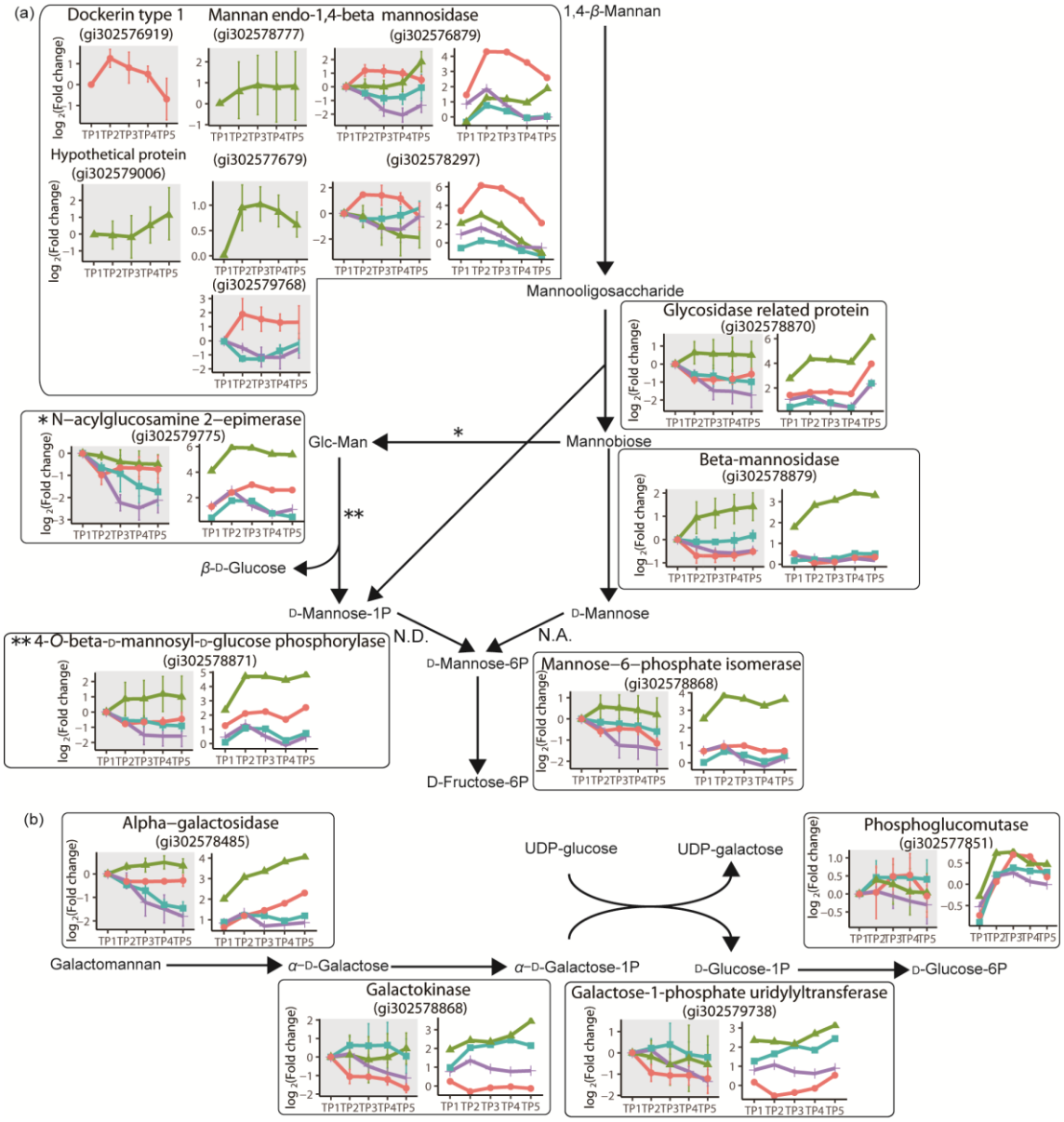


Figure 9 Pathway analysis of galactomannan degradation and metabolism

(a) Pathway of mannan degradation and metabolism (b) Galactose degradation and metabolism pathway

Each pathway was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

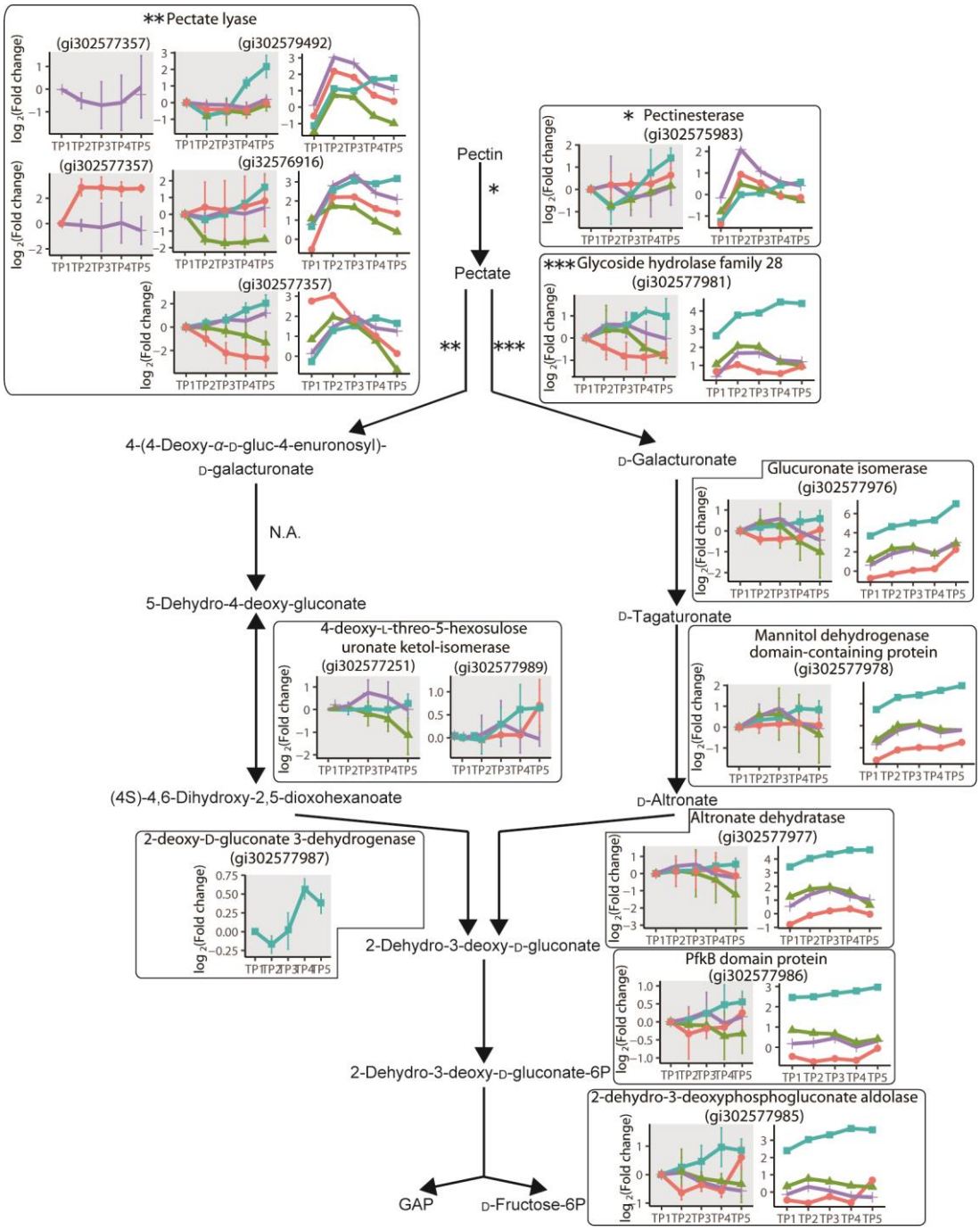
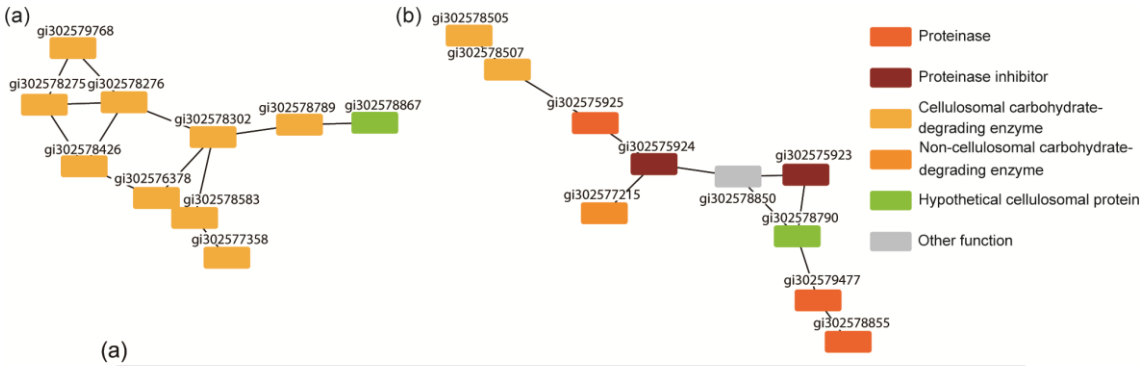


Figure 10 Pathway analysis of pectin degradation and metabolism

Pathway of pectin degradation and metabolism was constructed from KEGG. White box indicates the fold change of each protein relative to glucose at the same time point, and gray box indicates temporal profile of each protein relative to TP1 in each carbon sources. Each line indicates fold change or temporal profile (Cellulose: red, xylan: purple, green: galactomannan, blue: pectin).

Co-production analysis of the secreted proteins

I carried out a co-production analysis (Matsuda et al., 2017) against commonly identified secreted proteins to determine co-production patterns in an unbiased manner. Proteins with associated functions are often produced simultaneously (Lundquist et al., 2012). Therefore, proteins which have high correlation coefficients imply a functional or physical link (Lundquist et al., 2012). I calculated the Pearson's correlation coefficient using R and the protein co-production networks were visualized with Cytoscape 3.5.0 (Shannon et al., 2003). In these protein clusters, there was one big cluster, two medium-size clusters, and dozens of mini clusters. I focused on the medium-size clusters, which include proteins with similar functions (Figure 11). In the two medium-size clusters, cellulosomal proteins (cluster (a)) or proteinases/proteinase inhibitors (cluster (b)) were enriched, respectively. The abundance of proteins in these two clusters was coordinately regulated in the secreted proteins. The cellulosomal proteins in cluster (a) showed an increasing trend from *k*-means clustering analysis (Figure 11). Furthermore, in a previous report, the expression of almost all proteins in cluster (a) did not change regardless of the carbon source (Matsui et al., 2013). Moreover, proteins in this cluster have activity against cellulose (Arai et al., 2006), xylan (Han et al., 2004), and galactomannan (Jeon et al., 2011). These proteins are identified in all carbon sourcews and have activity for multiple carbon sources. I speculate that these cellulosomal proteins might be core enzymes in the degradation of cellulose and hemicellulose among the secreted proteins.



(a)

Proteins belonging to cellulosomal proteins cluster			
Accession	Description	Protein name	CAZy family
gi302576378	Glycoside hydrolase family 5	EngGH5F	GH5
gi302577358	Glycoside hydrolase family 9	EngGH9Y (EngY)	CBM30,GH9
gi302578275	Dockerin type 1	ManGH26C	CBM35,GH26
gi302578276	Glycoside hydrolase family 9	EngGH9B	GH9,CBM3
gi302578302	Dockerin type 1	ManGH26B (ManB)	CBM35,GH26
gi302578426	Glycoside hydrolase family 9	EngGH9Z (EngZ)	GH9,CBM3
gi302578583	Glycoside hydrolase family 10	XynGH10B (XynB)	CBM22,GH10
gi302578789	Glycoside hydrolase family 5	EngGH5D	GH5
gi302578867	Dockerin type 1	Hyp1	-
gi302579768	Dockerin type 1	ManGH26A	CBM35,GH26

(b)

Proteins belonging to proteinase/proteinase inhibitors cluster			
Accession	Description	Protein name	CAZy family
gi302575923	Proteinase inhibitor I42, chagasin	CpiC	-
gi302575924	Proteinase inhibitor I42, chagasin	CpiA	-
gi302575925	Ig domain protein group 2 domain protein	CSPB	-
gi302577215	Glycoside hydrolase family 9	-	CBM4,GH9
gi302578505	Glycoside hydrolase family 9	EngGH9K (EngK)	CBM4,GH9
gi302578507	Glycoside hydrolase family 48	ExgGH48S (ExgS)	GH48
gi302578790	Hypothetical protein Clocel_3112	Hyp8	-
gi302578850	Ig domain protein group 2 domain protein	-	-
gi302578855	M6 family metalloprotease domain protein	-	-
gi302579477	Peptidase M15B and M15C DD-carboxypeptidase VanY/endolysin	-	-

Figure 11 Co-production analysis of the secreted proteins

Two medium-sized clusters and proteins belonging to each cluster. Nodes and edges indicate proteins and their correlations. Edges indicate correlation coefficients larger than 0.80 or smaller than −0.80. (a) The medium-sized cluster comprised cellulosomal carbohydrate-degrading enzymes. (b) The medium-sized cluster comprised proteinases, proteinase inhibitors, and some cellulosomal proteins.

In cluster (b), I identified three proteinases, two cellulosomal proteinase inhibitors, two cellulosomal carbohydrate enzymes, one non-cellulosomal carbohydrate enzyme, and one cellulosomal hypothetical protein. All proteinase inhibitors identified in the secreted proteins belonged to this protein cluster. In the cellular proteins, the correlation coefficients of proteinases and proteinase inhibitors were between 0.000 and 0.702 and in the secreted proteins were higher than 0.80. In *C. cellulolyticum*, the expression of cellulosomal protease inhibitors is important for the activity of cellulase (Xu et al., 2014). In *C. cellulovorans*¸ cellulosomal proteinase inhibitors inhibit plant cysteine proteinases (Meguro et al., 2011). The proteinases co-expressed in this analysis were metalloproteases and a serine protease (Figure 11). Plants secrete anti-bacterial peptides and proteins as a defense against pathogens (Broekaert et al., 1997; Thomma et al., 2002). I suggest that *C. cellulovorans* secretes proteases and proteinase inhibitors in a coordinated way to protect its carbohydrate enzymes from plant proteinases. Using this protease protection system, *C. cellulovorans* can degrade polysaccharides in plant cell walls without losing carbohydrate-degrading enzyme activity. I speculate that *C. cellulovorans* might degrade anti-bacterial peptides and proteins with secreted proteases.

Summary

In order to clarify the temporal cellular and secreted proteome profiles, I performed temporal proteome analysis with five different carbon sources. Using an optimized LC-MS/MS system, equipped with a long monolithic silica capillary column, I identified 1,895 and 879 proteins in total as cellular and secreted proteins, respectively. *k*-means clustering and co-production analyses suggested new evidence for degradation mechanisms of polysaccharides and protease protection systems. I identified carbohydrate-degrading enzymes from each carbon source from uniquely identified proteins in each. Some of the identified proteins have not been previously reported as being involved in the degradation of carbon sources. *k*-means clustering in the cellular and secreted proteins suggested that cellulosomal proteins at the cell surface might be released as secreted proteins during degradation of plant cell wall polysaccharides (Figure 12a). Co-production analysis implied that protease and protease inhibitors are produced simultaneously (Figure 12b). These proteins might protect plant proteases and anti-microbial proteins/peptides. Our data are useful for the further understanding of *C. cellulovorans*' strategies and mechanisms of degradation and assimilation of major plant cell wall polysaccharides.

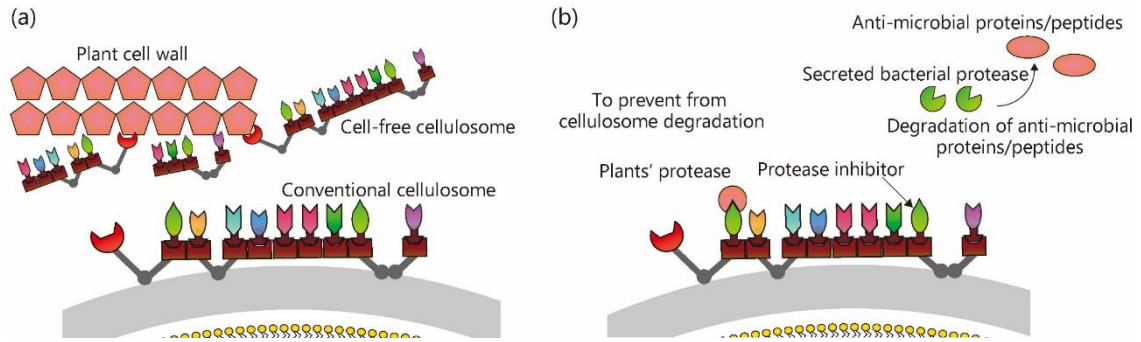


Figure 12 Proposal models of *C. cellulovorans* obtained in this study

(a) Plant cell wall degrading model of cell-free cellulosome.

Conventional cellulosome is anchored to the cell surface, but some studies reveal that cell-free cellulosome has an essential role in other cellulosome-producing *Clostridium* species. Cell-free cellulosome is a cellulosome which does not bind to the cell surface. Our analysis indicates that *C. cellulovorans* secretes cellulosome to the supernatant. Cell-free cellulosome might have a role as a secreted protein.

(b) Co-production of proteases and protease inhibitors prevents cellulosome from degradation.

Our analysis indicates that proteases and protease inhibitor are co-produced in all carbon sources. *C. cellulovorans*' protease inhibitors inhibit plant proteases. Furthermore, our analysis indicates that some proteases are co-produced. I speculate that the secreted bacterial proteases might have the same role as protease inhibitors and degrade anti-microbial proteins/peptides to protect from plant attack.

References

Arai, T., Kosugi, A., Chan, H., Koukiekolo, R., Yukawa, H., Inui, M., and Doi, R.H. (2006). Properties of cellulosomal family 9 cellulases from *Clostridium cellulovorans*. Appl Microbiol Biotechnol *71*, 654-660.

Broekaert, W.F., Cammue, B.P., De Bolle, M.F., Thevissen, K., De Samblanx, G.W., Osborn, R.W., and Nielson, K. (1997). Antimicrobial peptides from plants. CRC Crit Rev Plant Sci *16*, 297-323.

Ebringerová, A., Alföldi, J., Hromádková, Z., Pavlov, G., and Harding, S. (2000). Water-soluble *p*-carboxybenzylated beechwood 4-*O*-methylglucuronoxylan: structural features and properties. Carbohyd Polym *42*, 123-131.

Esaka, K., Aburaya, S., Morisaka, H., Kuroda, K., and Ueda, M. (2015). Exoproteome analysis of *Clostridium cellulovorans* in natural soft-biomass degradation. AMB Express *5*.

Han, S.O., Yukawa, H., Inui, M., and Doi, R.H. (2004). Isolation and expression of the xynB gene and its product, *XynB*, a consistent component of the *Clostridium cellulovorans* cellulosome. J Bacteriol *186*, 8347-8355.

Inamori, T., Aburaya, S., Morisaka, H., Kuroda, K., and Ueda, M. (2016). Characteristic strategy of assimilation of various saccharides by *Clostridium cellulovorans*. AMB Express *6*, 64.

Jeon, S.D., Yu, K.O., Kim, S.W., and Han, S.O. (2011). A celluloytic complex from *Clostridium cellulovorans* consisting of mannanase B and endoglucanase E has synergistic effects on galactomannan degradation. Appl Microbiol Biotechnol *90*, 565-572.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2011). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res *40*, D109-D114.

Kosugi, A., Murashima, K., and Doi, R.H. (2001). Characterization of xylanolytic enzymes in *Clostridium cellulovorans*: expression of xylanase activity dependent on growth substrates. J Bacteriol *183*, 7037-7043.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res *42*, D490-D495.

Lundquist, P., Poliakov, A., Bhuiyan, N.H., Zybailov, B., Sun, Q., and van Wijk, K.J. (2012). The functional network of the *Arabidopsis thaliana* plastoglobule proteome based on quantitative proteomics and genome-wide co-expression analysis. Plant Physiol, pp. 111.193144.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics *21*, 3787-3793.

Masuda, T., Tomita, M., and Ishihama, Y. (2008). Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J Proteome Res *7*, 731-740.

Matsuda, F., Kinoshita, S., Nishino, S., Tomita, A., and Shimizu, H. (2017). Targeted proteome analysis of single-gene deletion strains of *Saccharomyces cerevisiae* lacking enzymes in the central carbon metabolism. PloS ONE *12*, e0172742.

Matsui, K., Bae, J., Esaka, K., Morisaka, H., Kuroda, K., and Ueda, M. (2013). Exoproteome profiles of *Clostridium cellulovorans* grown on various carbon sources. Appl Environ Microbiol *79*, 6576-6584.

Meguro, H., Morisaka, H., Kuroda, K., Miyake, H., Tamaru, Y., and Ueda, M. (2011). Putative role of cellulosomal protease inhibitors in *Clostridium cellulovorans* by their gene expression and measurement of the activities. J Bacteriol, *193*, 5527-5530.

Miyake, H., Maeda, Y., Ishikawa, T., and Tanaka, A. (2016). Calorimetric studies of the growth of anaerobic microbes. J Biosci Bioeng *122*, 364-369.

Mohand-Oussaid, O., Payot, S., Guedon, E., Gelhaye, E., Youyou, A., and Petitdemange, H. (1999). The extracellular xylan degradative system in *Clostridium cellulolyticum* cultivated on xylan: evidence for cell-free cellulosome production. J Bacteriol *181*, 4035-4040.

Mohnen, D. (2008). Pectin structure and biosynthesis. Curr Opin Plant Biol *11*, 266-277.

Morisaka, H., Matsui, K., Tatsukami, Y., Kuroda, K., Miyake, H., Tamaru, Y., and Ueda, M. (2012). Profile of native cellulosomal proteins of *Clostridium cellulovorans* adapted

to various carbon sources. AMB Express *2*, 37.

Polizeli, M., Rizzatti, A., Monti, R., Terenzi, H., Jorge, J.A., and Amorim, D. (2005). Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol *67*, 577-591.

Rydzak, T., McQueen, P.D., Krokhin, O.V., Spicer, V., Ezzati, P., Dwivedi, R.C., Shamshurin, D., Levin, D.B., Wilkins, J.A., and Sparling, R. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiol *12*, 214.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Sleat, R., Mah, R.A., and Robinson, R. (1984). Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. Appl Environ Microbiol *48*, 88-93.

Tamaru, Y., Miyake, H., Kuroda, K., Nakanishi, A., Matsushima, C., Doi, R.H., and Ueda, M. (2011). Comparison of the mesophilic cellulosome‐producing *Clostridium cellulovorans* genome with other cellulosome‐related clostridial genomes. Microb Biotechnol *4*, 64-73.

Thomma, B.P., Cammue, B.P., and Thevissen, K. (2002). Plant defensins. Planta *216*, 193-202.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods *13*, 731.

Wessel, D., and Flügge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem *138*, 141-143.

Xu, Q., Resch, M.G., Podkaminer, K., Yang, S., Baker, J.O., Donohoe, B.S., Wilson, C., Klingeman, D.M., Olson, D.G., and Decker, S.R. (2016). Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Sci Adv *2*, e1501254.

Xu, T., Li, Y., He, Z., and Zhou, J. (2014). Dockerin‐containing protease inhibitor

protects key cellulosomal cellulases from proteolysis in *Clostridium cellulolyticum*. Mol Microbiol *91*, 694-705.

Conclusions

The present proteome study was carried out to understand the molecular mechanisms behind the changing protein profile with sensing of carbohydrates in the culture medium, and the temporal dynamics of the cellular and secreted proteins of *C. cellulovorans* depending on culture polysaccharides.

In Chapter I, I conducted definitive screening for optimization of the LC-ESI-MS/MS parameters to increase the number of proteins identified and the peak intensity of MS1. As a result of the optimization, I confirmed that eight parameters were significantly influential in increasing the number of proteins identified. Furthermore, I found two-factor interaction between the LC parameter and the MS parameter in increasing the number of proteins identified. This increase in sensitivity makes it possible to perform more comprehensive proteome analyses in *C. cellulovorans*.

In Chapter II, using comparative proteome analysis, I elucidated the molecular mechanism behind the changing proteome profile when sensing hemicellulose and pectin in the culture medium. On the basis of the proteome and genome analyses, I have suggested the molecular mechanism that changes the proteome profile depending on hemicellulose and pectin. For xylan and galactomannan, *C. cellulovorans* regulates hemicellulose-degrading and -metabolizing proteins by a two-component system. For pectin, *C. cellulovorans* regulates pectin-degrading and -metabolizing proteins by two transcriptional regulators.

In Chapter III, I revealed the temporal dynamics of cellular and secreted proteins by temporal proteome analyses with cellulose, hemicellulose, and pectin. From the temporal analysis, I suggested that secreted “cell-free cellulosomes” might be involved in degrading plant cell-wall polysaccharides. Furthermore, I found co-production of the proteinase and proteinase inhibitor in the secreted proteins. I suggested that these proteins are used to protect the cellulosome and bacterium from plant proteases and antibacterial proteins. My study revealed that *C. cellulovorans* showed different characteristic responses to plant cell-wall polysaccharides and subsequent molecular mechanisms for changing its proteome profile.

Acknowledgements

This thesis was submitted by the author to Kyoto University for the Doctor degree of Agriculture. The studies presented here have been carried out under the direction of Professor Mitsuyoshi Ueda in the Laboratory of Biomacromolecular Chemistry, Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, during 2013-2019.

I would like to express greatest appreciation to Mitsuyoshi Ueda who offered insightful comments, meaningful discussion, and continuous support throughout the course of my study. I also would like to thank associate professor Kouichi Kuroda who always gave me detailed and critical comments to my study. I am grateful to assistant professor Wataru Aoki for his valuable comments, which improved my study. I am deeply grateful to the late Hironobu Morisaka for his guidance, experimental suggestion, and skilled technical advice. I express my appreciation to the secretary Fukuko Suzuki for her support of my laboratory life. I greatly thank to Dr. Hiroyoshi Minakuchi for his technical support.

I am also grateful to all my colleagues in the laboratory for wide-ranging discussion, encouragements, and warm supports. I would like to thank financial supports from the research fellow ship of the Japan Society for the Promotion of Science for Young Scientists.

Shunsuke ABURAYA

Laboratory of Biomacromolecular Chemistry
Division of Applied Life Sciences
Graduate School of Agriculture
Kyoto University

Publications

Chapter I

Shunsuke Aburaya, Wataru Aoki, Hiroyoshi Minakuchi, Mitsuyoshi Ueda, Definitive screening design enables optimization of LC-ESI-MS/MS parameters in proteomics, *Bioscience, Biotechnology, and Biochemistry*, 81, 2237-2243 (2017)

Chapter II

Shunsuke Aburaya, Kohei Esaka, Hironobu Morisaka, Kouichi Kuroda, Mitsuyoshi Ueda, Elucidation of the recognition mechanisms for hemicellulose and pectin in *Clostridium cellulovorans* using intracellular quantitative proteome analysis. *AMB Express* 5, 29 (2015)

Chapter III

Shunsuke Aburaya, Wataru Aoki, Kouichi Kuroda, Hiroyoshi Minakuchi, Mitsuyoshi Ueda, Temporal proteome dynamics of *Clostridium cellulovorans* cultured with major plant cell wall polysaccharides, submitted

Other publications

Kohei Esaka, Shunsuke Aburaya, Hironobu Morisaka, Kouichi Kuroda, Mitsuyoshi Ueda, Exoproteome analysis of *Clostridium cellulovorans* in natural soft-biomass degradation, *AMB Express*, 5, 2 (2015)

Takako Inamori, Shunsuke Aburaya, Hironobu Morisaka, Kouichi Kuroda, Mitsuyoshi Ueda, Characteristic strategy of assimilation of various saccharides by *Clostridium cellulovorans*, *AMB Express*, 6, 64 (2016)

Masatoshi Murai, Hiroyuki Inaoka, Takahiro Masuya, Shunsuke Aburaya, Wataru Aoki, Hideto Miyoshi, Specific methylation of Asp160 (49 kDa subunit) located inside the quinone binding cavity of bovine mitochondrial complex I, *Biochemistry*,

55, 3189-3197 (2016)

Sakihito Kitajima, Kenji Miura, Wataru Aoki, Katsuyuki T Yamato, Toki Taira, Ryuta Murakami, Shunsuke Aburaya, Transcriptome and proteome analyses provide insight into laticifer's defense of *Euphorbia tirucalli* against pests, *Plant Physiology and Biochemistry*, 108, 434-446 (2016)

Kenji Okuda, Masatoshi Murai, Shunsuke Aburaya, Wataru Aoki, Hideto Miyoshi, Reduction of synthetic ubiquinone QT catalyzed by bovine mitochondrial complex I is decoupled from proton translocation, *Biochemistry*, 55, 470-481 (2016)

Toshiyuki Takagi, Hironobu Morisaka, Shunsuke Aburaya, Yohei Tatsukami, Kouichi Kuroda, Mitsuyoshi Ueda, Putative alginate assimilation process of the marine bacterium *Saccharophagus degradans* 2-40 based on quantitative proteomic analysis, *Marine Biotechnology*, 18, 15-23 (2016)

Seiji Shibasaki, Miki Karasaki, Shunsuke Aburaya, Hironobu Morisaka, Yumiko Takeda, Wataru Aoki, Sachie Kitano, Masayasu Kitano, Mitsuyoshi Ueda, Hajime Sano, Tsuyoshi Iwasaki, A comparative proteomics study of a synovial cell line stimulated with TNF-α, *FEBS Open Bio*, 6, 418-424 (2016)

Jihye Chung, Shunsuke Aburaya, Wataru Aoki, Mitsuyoshi Ueda, Molecular changes in appearance of a cancer cell among normal HEK293T cells, *Journal of Bioscience and Bioengineering*, 123, 281-286 (2017)

Takahiro Masuya, Masatoshi Murai, Takeshi Ito, Shunsuke Aburaya, Wataru Aoki, Hideto Miyoshi, Pinpoint chemical modification of the quinone-Access channel of mitochondrial complex I via a two-step conjugation reaction, *Biochemistry*, 56, 4279-4287 (2017)

Yohei Tatsukami, Toshiyuki Takagi, Shunsuke Aburaya, Natsuko Miura, Wataru Aoki, Mitsuyoshi Ueda, Metabolite profiling of the fermentation process of "*yamahai-ginjo-shikomi*" Japanese sake, *PLOS ONE*, 13, e0190040 (2018)

Keisuke Motone, Toshiyuki Takagi, Shunsuke Aburaya, Wataru Aoki, Natsuko Miura, Hiroyoshi Minakuchi, Haruko Takeyama, Yukio Nagasaki, Chuya Shinzato, Mitsuyoshi Ueda, Protection of coral larvae from thermally induced oxidative stress by redox nanoparticles, *Marine Biotechnology*, 20, 542-548 (2018)

Ayaki Aoyama, Masatoshi Murai, Naoya Ichimaru, Shunsuke Aburaya, Wataru Aoki, Hideto Miyoshi, Epoxycyclohexenedione-type compounds make up a new class of inhibitors of the bovine mitochondrial ADP/ATP carrier, *Biochemistry*, 57, 1031-1044 (2018)

Sakihito Kitajima, Wataru Aoki, Daisuke Shibata, Daisuke Nakajima, Nozomu Sakurai, Kazufumi Yazaki, Ryosuke Munakata, Toki Taira, Masaru Kobayashi, Shunsuke Aburaya, Eric Hyrmeya Savadogo, Susumu Hibino, Haruna Yano, Comparative multi-omics analysis reveals diverse latex-based defense strategies against pests among latex-producing organs of the fig tree (*Ficus carica*), *Planta*, 247, 1423-1438 (2018)

Takahiro Tsuji, Hiroaki Ozasa, Wataru Aoki, Shunsuke Aburaya, Tomoko Funazo, Koh Furugaki, Yasushi Yoshimura, Hitomi Ajimizu, Ryoko Okutani, Yuto Yasuda, Takashi Nomizo, Kiyoshi Uemasu, Koichi Hasegawa, Hironori Yoshida, Yoshitaka Yagi, Hiroki Nagai, Yuichi Sakamori, Mitsuyoshi Ueda, Toyohiro Hirai, Young Hak Kim, Alectinib resistance in ALK-rearranged lung cancer by dual salvage signaling in a clinically paired resistance model, *Molecular Cancer Research* in press

Yasuhiro Kato, JeongHoon Park, Hyota Takamatsu, Hachirou Konaka, Wataru Aoki, Shunsuke Aburaya, Mitsuyoshi Ueda, Masayuki Nishide, Shohei Koyama,

Yoshitomo Hayama, Yuhei Kinehara, Toru Hirano, Yoshihito Shima, Masashi Narazaki, Atsushi Kumanogoh, Apoptosis-derived membrane vesicles drive the cGAS–STING pathway and enhance type I IFN production in systemic lupus erythematosus, *Annals of the Rheumatic Diseases*, 77, 1507-1515 (2018)

Kojiro Takanashi, Yukimi Nakagawa, Shunsuke Aburaya, Kenta Kaminade, Wataru Aoki, Yuka Saida-Munakata, Akifumi Sugiayama, Mitsuyoshi Ueda, Kazufumi Yazaki, Comparative proteomic analysis of *Lithospermum erythrorhizon* reveals regulation of a variety of metabolic enzymes leading to comprehensive understanding of the shikonin biosynthetic pathway, *Plant and Cell Physiology* in press

Natsuko Miura, Keisuke Motone, Toshiyuki Takagi, Shunsuke Aburaya, Sho Watanabe, Wataru Aoki, Mitsuyoshi Ueda, Ruegeria sp. strains isolated from the reef-building coral *Galaxea fascicularis* inhibit growth of the temperature-dependent pathogen *Vibrio coralliilyticus*, *Marine Biotechnology* in press