博士論文

Structural and functional analysis of the GH family 6 cellobiohydrolase from the basidiomycete *Phanerochaete chrysosporium*

（担子菌 *Phanerochaete chrysosporium* 由来
GH ファミリー6 に属するセロビオヒドロラーゼの構造と機能解析）

立岡美夏子

Contents

List of publications

1. Tachioka, M., Sugimoto, N., Nakamura, A., Sunagawa, N., Ishida, T., Uchiyama, T., Igarashi, K., Samejima, M., 2016. Development of simple random mutagenesis protocol for the protein expression system in *Pichia pastoris*. *Biotechnol. Biofuels.* 9, 1–10.

2. Tachioka, M., Nakamura, A., Ishida, T., Igarashi, K., Samejima, M., 2017. Crystal structure of a family 6 cellobiohydrolase from the basidiomycete *Phanerochaete chrysosporium*. *Acta Crystallogr. Sect. F Struct. Biol. Commun*. 73, 398–403.

3. Tachioka, M.*, Nakamura, A.*, Ishida, T., Takahashi, S., Yan, B., Tanaka, H., Furubayashi, N., Inaka, K., Igarashi, K., Samejima, M., 2017. 水素原子の可視化を目指したセルラーゼの大型結晶作製 (Production of large-volume cellulase crystals for visualization of hydrogen atoms). *Int. J. Microgravity Sci. Appl.* 34, 340108-1–6.

* First author ship shared.

Abbreviations

CAZy	Carbohydrate active enzymes
CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CD	Catalytic domain
Cc	*Coprinopsis cinerea*
Cf	*Cellulomonas fimi*
CMC	Carboxy methyl cellulose
EG	Endoglucanase
GH	Glycoside hydrolase
Glc_n	Glucosyl unit at subsite ‘n’
Hi	*Humicola insolens*
MDA	Multiple displacement amplification
PASC	Phosphoric acid-swollen cellulose
Pc	*Phanerochaete chrysosporium*
PDB	Protein data bank
*p*NPG3	*p*-nitrophenyl β-D-cellotrioside
P. pastoris	*Pichia pastoris*
RCA	Rolling circle amplification
Tf	*Thermobifida fusca*
Tr	*Trichoderma reesei*

1. Introduction

1.1. Cellulose

Cellulose is biosynthesized by many living organisms including plant, algae, bacteria, fungi and animals (Pérez & Samain, 2010). Cellulose is the most abundant organic compound on earth and it is estimated that several tons of cellulose is fixed annually in the form of plant cell wall (Coughlan *et al.*, 1985; Hon, 1994; Percival Zhang *et al.*, 2006).

Cellulose is a simple, long polymer composed of β-1,4-linked D-glucopyranose units (Fig. 1-1A), where the directionality of the chain is identified by the reducing and non-reducing ends. As the molecules are rotated ~180˚ relative to their neighbors, a cellobiosyl unit is the smallest repeating unit of the linear polymer. Its degree of polymerization (DP) is estimated to range from hundreds to several tens of thousands depending on the source organisms (Hon, 1994; Pérez & Samain, 2010).

Owing to the ability to form hydrogen bonding and hydrophobic interactions with each other, cellulose chains pack into tight crystalline complex called 'microfibril'. It contains both highly crystalline and less ordered amorphous region, and shapes and crystallinity of microfibril are also diverse among source organisms (Hon, 1994; Pérez & Samain, 2010; Moon *et al.*, 2011; Payne *et al.*, 2015). Most of the naturally occurring crystalline form is a complex of two allomorph I_α and I_β, in which chains are organized in a parallel fashion. In both cellulose I_α and I_β, these chains are making up flat sheets stabilized by 2-dimensional intra-layer hydrogen bonds, and the sheets are stacked by van der Waals interactions and C–H···O hydrogen bonds

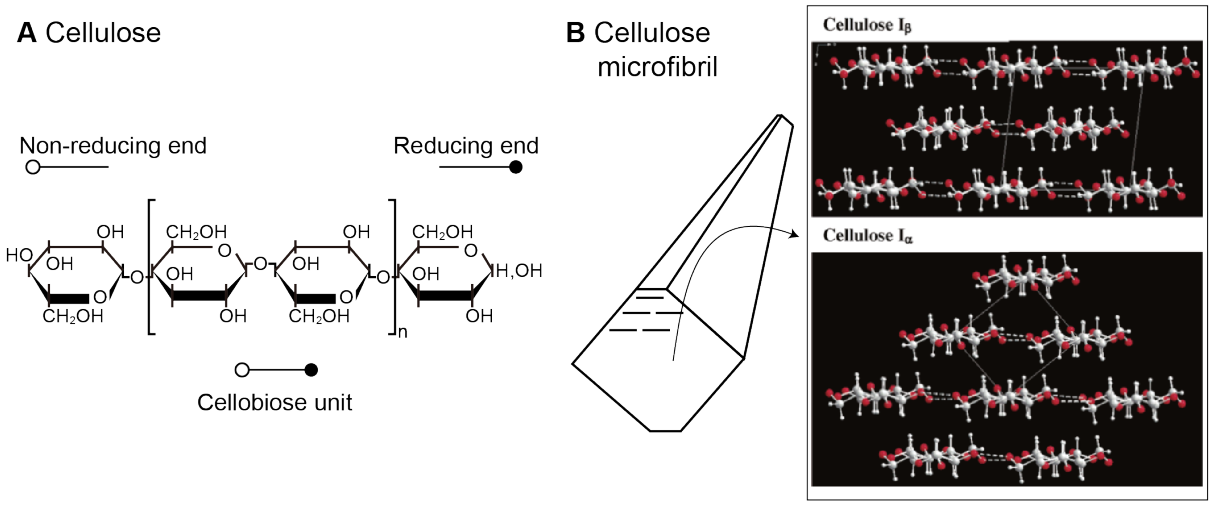

Figure 1-1. (A) Chemical structure of a single cellulose chain. (B) Microfibril and crystal structures of cellulose I_β and I_α. Pictures of the crystal structures are from Wada *et al.*, 2004 .

(Nishiyama *et al.*, 2002, 2003; Nishiyama, 2009) (Fig. 1-1B). In the case of plant cell wall, the crystallinity is generally 50–90% (Hon, 1994), and the 36-chain diamond shape or much smaller chain organization has been proposed as the model of microfibril (Payne *et al.*, 2015). The microphibril is further assembled with other microfiblils and making a complex with hemicellulose and lignin (Bidlack *et al.*, 1992; Carpita & Gibeaut, 1993). These crystalline structure and complexed interactions with other polymers makes the plant cell wall resistant to biological and chemical degradation.

1.2. Cellulases

As cellulose is accounting for large fraction of the fixed carbon, the biodegradation of cellulose is ecologically important and plays key roles in the global carbon cycle (Lynd *et al.*, 2002). Cellulose-degrading organisms have been found in relatively selected but diverse taxonomic groups (Himmel *et al.*, 2010; Cragg *et al.*, 2015). They have been evolved to produce an array of extracellular proteins to access cellulose and to convert it into soluble oligosaccharides. In many cellulose degradation systems, glycoside hydrolases (GHs) are primary enzymes. Efficient degraders have a variety of genes of enzymes in their genome, and tailors the enzyme cocktail to fit surrounding environments.

Cellulase is a generic term used to refer to enzymes that hydrolyze β-1,4-glycosidic bond of cellulose, or those that are active on cellulose. Currently, proteins which are active on carbohydrates are discussed according to the classification system in the carbohydrate-active enzymes (CAZy) database (Lombard *et al.*, 2014). In the CAZy system, proteins are classified on the basis of amino-acid sequence similarities. Therefore it reflects structural features of proteins in contrast to the conventional nomenclatures which are related to substrate specificities (Davies & Sinnott, 2008; Cantarel *et al.*, 2009). The database is launched in the late 1980s (Henrissat *et al.*, 1989, 1995) and continuously updated to cover glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and carbohydrate binding modules (CBMs) in addition to GHs.

GH families and catalytic mechanisms

GH is the biggest class in the CAZy database consisting of more than 140 families at this moment. As the hydrolysis occurs at the anomeric position, GHs can be divided into two groups: ‘retaining’ or ‘inverting’, in which anomeric configurations of products are retained or inverted, respectively. This catalytic pattern is mostly conserved within GH families whereas some GH families (e.g. families 23, 97) has been reported to include both retaining and inverting enzymes. Approximately half of the families are characterized (or predicted) as retaining and ~30% of the families as inverting (Fig. 1-2). Cellulases are found in several GH

families: 5, 7, 12, 44 and 51 as retaining enzymes, and 6, 8, 9, 45, 48, 74 and 124 as inverting enzymes.

The fundamental mechanism of enzymatic glycosidic bond hydrolysis is proposed by Koshland in 1953 (Koshland, 1953), which has been a conceptual basis for long year research history of GHs. Retaining mechanism occurs in two-step reaction, where a glycosyl-enzyme complex is formed as an intermediate (Fig. 1-3A). In the first step, the glycosidic bond is cleaved by proton donation from an acid/base residue and simultaneous attack of a nucleophile residue at C1 atom. Then the intermediate is cleaved by nucleophilic attack of a water with an aid of acid/base residue. On the other hand, inverting mechanism is a single displacement mechanism by a direct attack of a water molecule (Fig. 1-3B). An acid residue protonates the glycosidic bond oxygen and a base residue increases the nucleophilicity of the attacking water.

Aspartic acid (Asp) or glutamic acid (Glu) residues are the most frequent catalytic residues in GHs. Generally, the acid/base and nucleophile residues are located closer together (4.5–5.5 Å) in the structures of retaining GHs, and the acid and base residues are located apart (9–10.5 Å) in those of inverting GHs (McCarter & Withers, 1994; Davies & Henrissat, 1995; Zechel & Withers, 2000). In addition to the pair of acidic residues, the distortion of the sugar is also known as an important and general feature of GH mechanism (Davies *et al.*, 2003, 2012; Nerinckx *et al.*, 2003; Heather B Mayes *et al.*, 2014). There are also exceptions of catalytic residues as reviewed in several papers (Vuong & Wilson, 2010; Fushinobu *et al.*, 2013; Jongkees & Withers, 2013). Especially, identification of catalytic residues of inverting GHs are less secure, and the distances between two carboxyl residues are updated to a much larger range of 6–12 Å in the recent book by Sinnott (Sinnott, 2013).

1	2	3	4	**5**	**6**	**7**	**8**	**9**	10	11	**12**	13	14	15	16	17	18	19	20
	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
	42	43	**44**	**45**	46	47	**48**	49	50	**51**	52	53	54	55	56	57	58	59	
AA9	62	63	64	65	66	67	68	PL16	70	71	72	73	**74**	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	**124**	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
141	142	143	144	145															

Figure 1-2. GH families. The anomeric configurations of products are retained or inverted in families colored in yellow or green, respectively. Empty cells are deleted families.GH family 61 was moved to auxiliary activities (AA) family 61 and GH family 69 was moved to polysaccharide lyases (PL) family 16. GH family 23 and 97 includes both retaining and inverting enzymes. GH families that include cellulases are shown in bold.

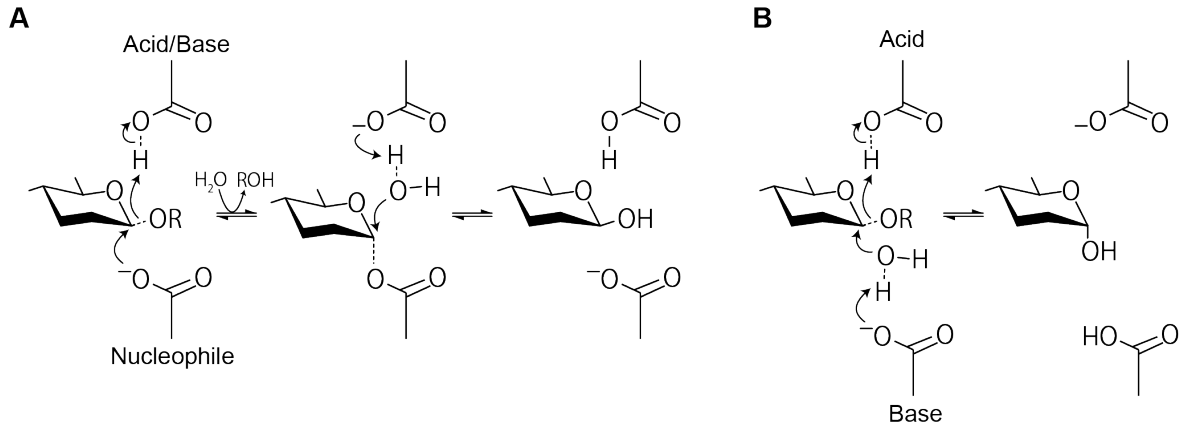

Figure 1-3. Retaining (A) and inverting (B) mechanism

Functions and structures of cellulases

Since the model of enzymatic cellulose degradation was first documented in 1950 (Reese *et al.*, 1950), there are extensive investigations on the mechanism of cellulase action. Reese *et al*. revealed that only limited numbers of microorganisms had abilities to produce a complete set of enzymes for degrading native cellulose, though enzymes capable of hydrolyzing soluble cellulose derivatives were much more widespread in microorganisms. Detailed properties of the components that were able to attack native cellulose (termed as "C_1" in Reese's paper) was not clear at that time. Later on, great efforts put into isolation of pure enzymes, and cellobiohydrolases (CBHs) were identified as C_1 enzymes in soft-rot fungi *Trichoderma* species (Wood & McCrae, 1972; Berghem & Pettersson, 1973). With that background, cellulases have been historically classified into either endoglucanase (EG) or CBH, according to their catalytic efficiencies in crystalline cellulose degradation. EGs are known to hydrolyze carboxymethyl cellulose (CMC) and cause decreased DP of cellulose, thus considered to hydrolyze glycosidic bonds randomly in the middle of cellulose chains. On the other hand, CBHs are exo-acting cellulases that have little ability to attack CMC and produce mainly cellobiose from cellulosic substrates. CBHs are necessary for crystalline cellulose degradation, and it is contrast that endoglucanases (EGs) alone cannot degrade crystalline cellulose.

Cellulase activities are found in the listed GH families in Table1-1, though some families involve mainly in degradations of other carbohydrates rather than cellulose. Aerobic filamentous fungi, such as soft-rot *T. reesei* (*Tr*), *Humicola insolens* (*Hi*) and white-rot *Phanerochaete chrysosporium* (*Pc*), share similar sets of cellulases (Uzcategui *et al.*, 1991; Schülein, 1997; Teeri, 1997). They generally produce two types of CBHs, GH family 6 and 7, and several EGs such as GH family 5, 6, 7, 12, 45 (Payne *et al.*, 2015). Most of the cellulases present modular structures consisting of a catalytic domain (CD) and binding modules. Several

aerobic bacteria, such as the thermophilic bacteria *Thermobifica fusca* (*Tf*) and actinobacteria *Cellulomonas fimi* (*Cf*), employ enzyme systems composed of GH family 5, 6, 9, and 48 (Wilson, 2004; Christopherson *et al.*, 2013). Anaerobic microorganisms such as the thermophilic bacteria *Clostridium thermocellum* employ multiple enzyme complexes termed 'cellulosome' (Johnson *et al.*, 1982; Lamed *et al.*, 1983), where GH family 5, 9 and 48 are found dominantly (Doi & Kosugi, 2004). Though the strategies of the cellulose degradation differ among microorganisms, the fungal system is the main focus of this thesis. Fungi are responsible for majority of biomass degradation (Lynd *et al.*, 2002), therefore understanding the fungal cellulases is unquestionably important.

The molecular-level explanation to the different modes of action of CBHs and EGs was brought from the atomic resolution structures of each domain that started to appear from the beginning of 1990s. The CDs of *Tr*CBH2 (Cel6A) (Rouvinen *et al.*, 1990), EG (*Tf*Cel6A, previously named as E2) (Spezio *et al.*, 1993) and *Tr*CBH1 (Cel7A) (Divne *et al.*, 1994) were determined by X-ray crystallography and CBM1 from *Tr*Cel7A was determined by NMR (Kraulis *et al.*, 1989). Davies and Henrissat surveyed available structures in 1995 and reviewed that the modes of action are closely related to overall architectures (topologies) (Davies & Henrissat, 1995). Similar examples of fungal GH structures are shown in Fig. 1-4, together with the oxidative enzymes that are involved in cellulose degradation.

Table 1-1. Currently known GH families with cellulase activities

Retaing cellulases

family	nuc.*	acid/base*	number**	major activity***
5	Glu	Glu	10313	EG (EC 3.2.1.4); endo-β-1,4-mannanase (EC 3.2.1.78)
7	Glu	Glu	5106	EG (EC 3.2.1.4); reducing end-acting CBH (EC 3.2.1.176)
12	Glu	Glu	762	EG (EC 3.2.1.4); xyloglucan endo-hydrolase (EC 3.2.1.151)
44	Glu	Glu	143	xyloglucanase (EC 3.2.1.151)
51	Glu	Glu	2376	α-L-arabinofuranosidase (EC 3.2.1.55)

Inverting cellulases

family	acid*	base*	number**	major activity***
6	Asp	Asp, Grotthuss	1136	EG (EC 3.2.1.4); CBH (EC 3.2.1.91)
8	Glu	(a) Asp (b) Glu ?Grotthuss	3039	EG (EC 3.2.1.4); chitosanase (EC 3.2.1.132); licheninase (EC 3.2.1.73); endo-β-1,4-xylanase (EC 3.2.1.8)
9	Glu	Asp	2555	EG (EC 3.2.1.4)
45	Asp	(a) Asp	375	EG(EC 3.2.1.4); distantly related to plant expansins
48	Glu	unknown	990	reducing end-acting CBH (EC 3.2.1.176); EG (EC 3.2.1.4); chitinase (EC 3.2.1.14)
74	Asp	Asp	384	xyloglucanase (EC 3.2.1.151)
124	Glu	?Grotthuss	6	EG (EC 3.2.1.4)

***Catalytic nucleophile, acid/base, acid and base residues. Alphabets in parenthesis indicate subfamily groups in GH family. 'Grotthus' indicates proton transfering mechanisms.
**Number registered in CAZy database.
***According to CAZyPedia (http://www.cazypedia.org/)

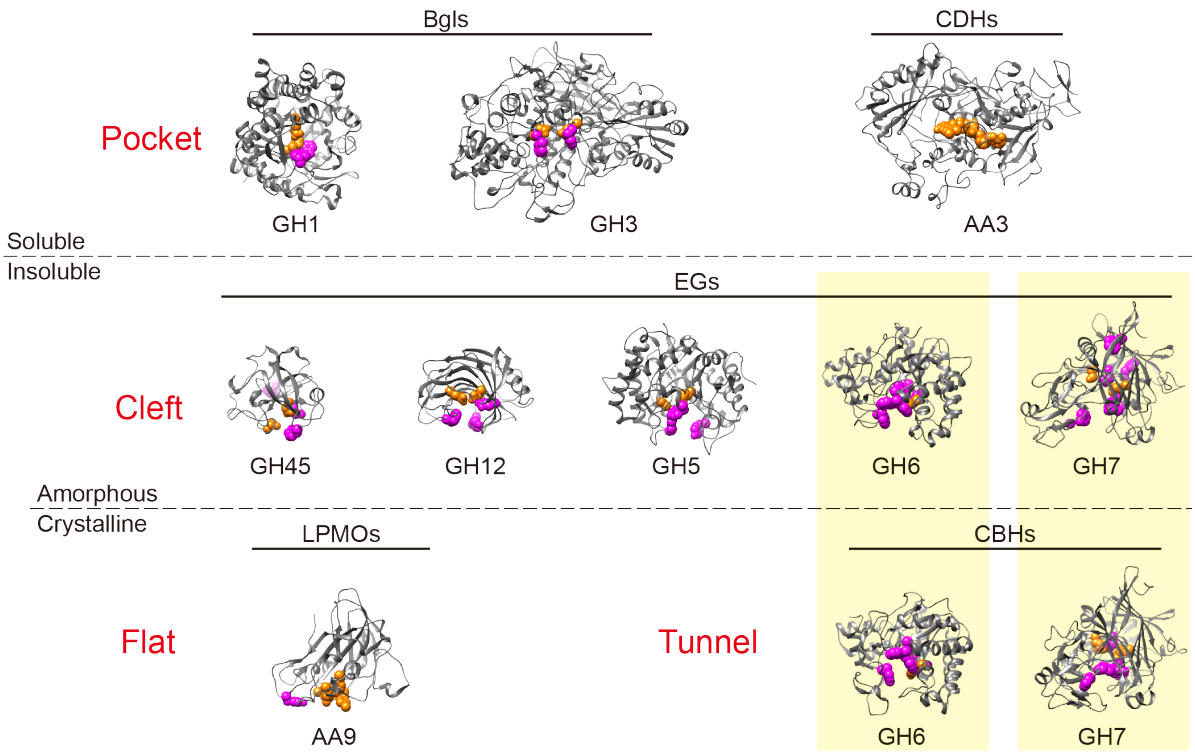

Figure 1-4. Examples of functional domains of fungal cellulolytic enzymes, with CAZy family numbers, their modes of action, and rough categorizations by active site topologies. Aromatic residues of substrate-binding sites are shown in magenta, and catalytic residues and co-factors are shown in orange. Abbreviations: Bgl, β-glucosidase, CDH, cellobiose dehydrogenase; EG, endoglucanase; CBM, carbohydrate binding module; LPMO, lytic polysaccharide mono-oxigenases; CBH, cellobiohydrolase; AA, auxiliary activities. Models were taken from protein data bank (PDB) entry 2E40, 4IIC, 4QI7, 3X2M, 1NLR, 3AYS, 1DYS, 1PJJ, 2YET, 1QK2 and 4C4C, from top left to bottom right.

The active sites of EGs are found as clefts (Fig. 1-4, middle), which is suitable for accommodating a single polymer chain in the active site and hydrolyze it randomly. In the case of CBHs, the active sites are covered by additional loops which make tunnel-like shapes (Fig. 1-4, bottom). This structure is not suitable for acquiring a chain into the active site relative to the cleft-type EGs, thus facilitates exo-type initiations of catalysis. At the same time, once CBHs acquire a cellulose chain in its active site, it is advantageous in keeping a cellulose chain to perform continuous catalysis. The pocket-type active sites are found in enzymes acting on soluble smaller substrates like β-glucosidases (Bgls) and cellobiose dehydrogenases (CDHs) (Fig. 1-4, upper). Flat-type binding sites (Fig. 1-4, bottom) found in Lytic polysaccharide mono-oxygenases (LPMOs) are optimal for recognizing the surface of crystalline cellulose. LPMOs are enzymes that cleave glycosidic bonds at crystalline surface by oxidative reaction with reducing agents and a copper in the active site (Vaaje-Kolstad *et al.*, 2010; Langston *et al.*, 2011). LPMOs have attracted attentions in recent years for their boosting effects in cellulose degradation and reviewed in several papers (Fushinobu, 2014; Vaaje-Kolstad *et al.*, 2017).

The cleft or tunnel type active sites that accommodate a single cellulose chain are common structures in cellulases. It is contrast that the deep-pocket exo-enzymes are found in glucoamyses and β-amylases, possibly because of the large availability of chain ends in their substrates (Davies & Henrissat, 1995). The CBHs are only enzymes that degrade crystalline cellulose into soluble oligosaccharides, and the tunnel structure is considered to be an essential feature. It is also contrast that chitinases utilize deep-cleft type active sites to degrade crystalline chitin (Vaaje-Kolstad *et al.*, 2013).

Main features of CBHs

The existence of two different types of CBHs, CBH1 and CBH2, in fungal cellulase systems was designated with the isolation of CBH2 from *T. reesei* (Fägerstam & Pettersson, 1979, 1980). The *Tr*CBH1 and *Tr*CBH2 genes were cloned in 1980s (Teeri *et al.*, 1983, 1987), and their catalytic domains (CDs) are currently classified into GH family 7 and 6, respectively. The CBHs are produced most dominantly among biomass degrading enzymes in *T. reesei*, as 40–70% and 10–20% of its total extracellular proteins have been reported to be CBH1 (Cel7) and CBH2 (Cel6), respectively (Nidetzky & Claeyssens, 1994; Nidetzky *et al.*, 1994; Teeri, 1997; Rosgaard *et al.*, 2007). The gene-knockout study also proved that deletions of *Tr*CBHs cause the most drastic effect on its cellulose degradation ability (Suominen *et al.*, 1993). It is widely considered that the hydrolysis of crystalline cellulose by CBHs is a rate-limiting step in cellulase degradation.

The Cel7 and Cel6 have been known as exo-enzymes that attack cellulose from opposite ends, reducing- and non-reducing ends, respectively (Teeri, 1997). The Cel7s mainly produce β-cellobiose by retaining mechanism, and Cel6s mainly produce α-cellobiose by inverting mechanism (Knowles *et al.*, 1988; Claeyssens *et al.*, 1990). They share similar two-domain modular structures, where CDs are connected by linker peptide to a cellulose binding domain (CBM1). The small angle X-ray scattering revealed that the intact *Tr*CBHs are 'tadpole'-like structures (Abuja *et al.*, 1988a, 1988b). The active sites of CBHs are asymmetric: Cel7s have binding subsites of -7 to +2 and that of Cel6s have binding subsites of -2 to +4 (Fig. 1-5).

The processivity, generally defined as the ability of performing multiple rounds of catalysis before dissociating from its substrates, is closely related to enclosed subsite structures (Breyer & Matthews, 2001). Cel7s are known as essentially processive enzymes. Microscopic studies visualized that Cel7s induce thinning of crystalline cellulose (Imai *et al.*, 1998; Boisset *et al.*, 2000), and the unidirectional movements of *Tr*Cel7A on the cellulose surface was observed by high-speed atomic force microscopy (AFM) (Igarashi *et al.*, 2009, 2011). It was also shown that the high processivities of Cel7s are associated with the lengths of the loop regions (Nakamura *et al.*, 2014). On the other hand, Cel6s induce sharpening of

crystalline tips, indicating that Cel6s show lower processivity and more localized attack (Chanzy & Henrissat, 1985; Boisset *et al.*, 2000). The processive actions of Cel6s are also questioned because Cel6s did not move significantly in the AFM observation (Igarashi *et al.*, 2011). At the same time, several researches demonstrated that endo-mode initiations are common events for Cel7s and Cel6s, indicating that the tunnel-enclosing loops have some flexibility and do not restrict CBHs for exo-initiation only (Ståhlberg *et al.*, 1993; Amano & Shiroishi, 1996; Armand *et al.*, 1997; Boisset *et al.*, 2000; Kurasin & Väljamäe, 2011).

The role of the CBM1 is implicated in the adsorption to hydrophobic surface of cellulose (Lehtiö *et al.*, 2003), and the elimination of CBM1 from CBHs causes significant losses of activity toward crystalline cellulose (Van Tilbeurgh *et al.*, 1986; Tomme *et al.*, 1988; Ståhlberg *et al.*, 1993). As well as the CDs, CBM1s and linkers of Cel6s and Cel7s are slightly different from each other. The aromatic residues of Cel7s are three tyrosine residues, but one of these residues is replaced by a tryptophan residue in Cel6s that result in increased affinity (Linder *et al.*, 1995; Carrard & Linder, 1999; Takashima *et al.*, 2007). A single molecule imaging technique revealed different morphological preferences of CBM1s, where that of *Tr*Cel7A had higher affinity to ordered crystalline regions than that of *Tr*Cel6A (Fox *et al.*, 2013). Binding specificities to reducing end was also proposed for *Tr*Cel6A-CBM1 (Guo & Catchmark, 2013).

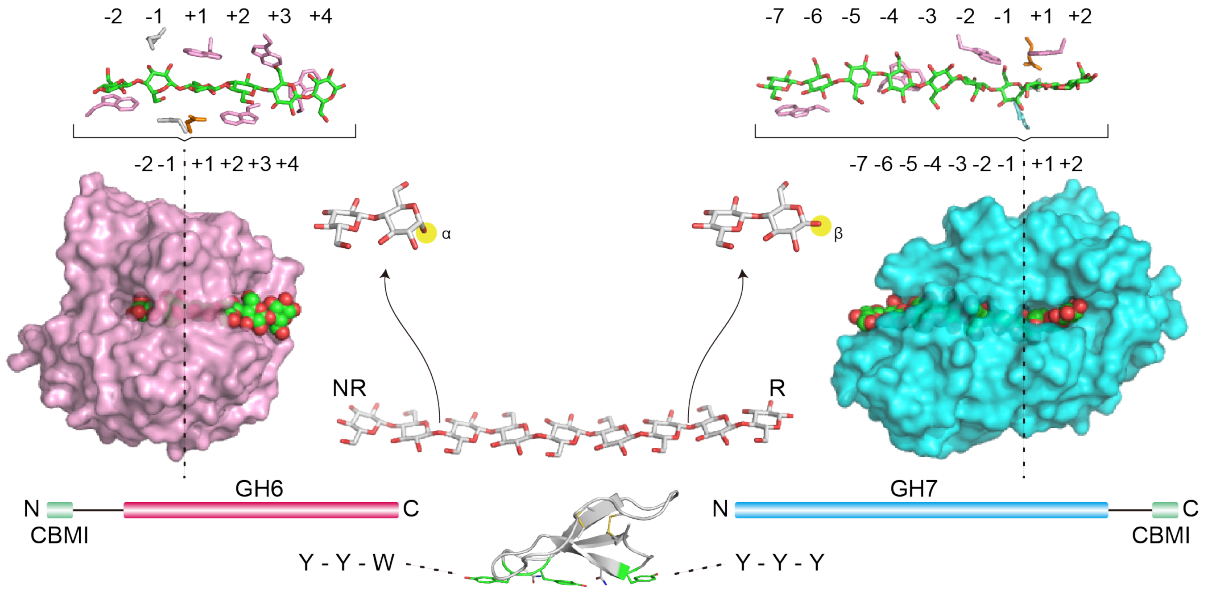

Figure 1-5. GH family 6 and 7 CBHs. Cel6s attack cellulose from non-reducing (NR) end and produce α-cellobiose by inverting mechanism. Cel7s attack cellulose from reducing (R) end and produce β-cellobiose by retaining mechanism. Cel6s and Cel7s have CBM1 in N- and C-terminus, respectively. Cel7-CBM1 has three tyrosine at bottom (PDB entry: 1CBH), on the other hand, one of the three tyrosine is replaced by tryptophan in Cel6-CBM1.

Linker regions are generally longer (~45 peptides) in Cel6 than in Cel7 (~30 peptides) (Sammond *et al.*, 2012) and the glycosylation of the linker also play important role in binding affinity (Payne *et al.*, 2013b).

It is interesting that Cel7s and Cel6s differ on several points such as catalytic mechanisms, asymmetrical structures of active sites, degree of processivity and endo-type initiations, and properties of CBMs and linkers. It has been long known from 1980 that two CBHs exhibit the "exo-exo synergy" (Fägerstam & Pettersson, 1980). These differences of two CBHs are likely developed during evolution to enable cooperative degradation of crystalline cellulose. Rate-determining steps in actions of CBHs are often discussed among researchers, and one consensus is that non-productive binding is a factor in the case of Cel7s (Igarashi *et al.*, 2011; Kurasin & Väljamäe, 2011; Payne *et al.*, 2013a). Heterogeneous crystalline surface does not allow Cel7s to show their full processivity and Cel7s are stacked on cellulose surface. The earlier theory of synergy explained the roles of non-processive enzymes as increasing attacking points of CBHs, but it is now pointed out that they also help the dissociation of CBHs (Jalak *et al.*, 2012). Cel6s can play such roles in crystalline cellulose degradation, but their modes of action are still unclear. In order to get the full picture of crystalline cellulose degradation, further investigations are necessary, especially for GH family 6 CBHs.

1.3. GH family 6

There are currently ~1,000 proteins listed in GH family 6 in the CAZy database. This family is mainly composed of cellulases, as the hydrolysis activities of β-1,4 glycosidic bond in cellulosic substrates and β-glucans are reported. Diverse sets of EGs and CBHs from bacteria and fungi are extensively investigated, and their three-dimensional structures are also available. As shown in Fig 1-6, the diversity is remarkable in the loop regions near the active sites. Both fungal and bacterial EGs have open active sites with shortened loops (Spezio *et al.*, 1993; Davies *et al.*, 2000). Fungal CBHs have tunnel-like active site covered by two loops (Rouvinen *et al.*, 1990), which are labeled as N- and C- terminal loop. Bacterial CBHs have additional and much more extended loops (Sandgren *et al.*, 2013; M. Wu *et al.*, 2013a). The domain architectures are different between fungal and bacterial cellulases and binding domains are found in either N- or C- termini (Mertz *et al.*, 2005). GH family 6-like protein are found in tunicate cellulose synthase with a lack of conserved catalytic residues suggesting that it does not maintain original cellulase activity (Nakashima *et al.*, 2004). In an exceptional case, GH family 6-like protein with xylanase activity was reported (Kim *et al.*, 2012).

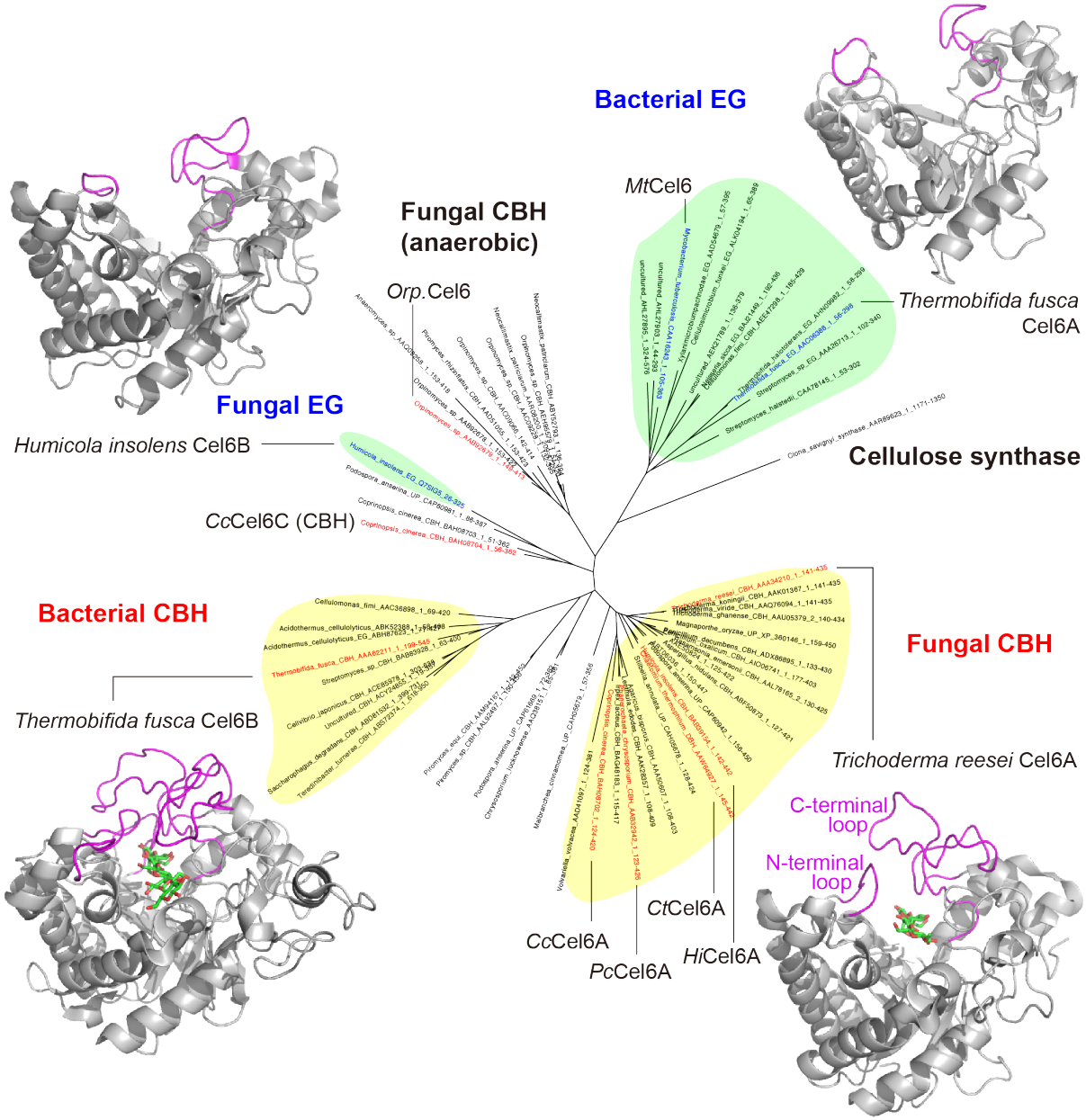

Figure 1-6. Phylogenetic tree of GH family 6 domains that are listed as 'characterized' in the CAZy database. The phylogenetic analysis was performed as follows: sequences of characterized GH family 6 enzymes were obtained by *CAZytool* by Alexander H. Viborg (http://research.ahv.dk/); the sequences were annotated by *Domosaics* (Moore *et al.*, 2014) with *Pfam* libraries and the sequences encoding GH family 6 domains were extracted; the alignment was then performed by *MAFFT* with E-INS-i method (Katoh *et al.*, 2017). The enzymes whose three-dimensional structures are determined are colored in red (CBH) and blue (EG). The models were taken from *H. insolens* Cel6B (PDB entry 1DYS), *T. fusca* Cel6A (1TML), *T. fusca* Cel6B (4B4F), and *T. reesei* Cel6A (1QK2).

Catalytic mechanism of GH family 6 cellulases

The core active sites numbered +2 to -2 are required for catalysis (Van Tilbeurgh *et al.*, 1985, 1989; Claeyssens *et al.*, 1989) and conserved in GH family 6 cellulases, as structures of *Tr*Cel6A and *Hi*Cel6A are shown in Fig. 1-7B and C. The subsites -2, +1 and +2 consist of tryptophan residues and accommodate glycosil units in normal chair conformations. The glycosil unit at the subsite -1 was frequently observed as distorted conformations between $^{2}S_{O}$ skew-boat and $^{2,5}B$ boat in crystal structures (Zou *et al.*, 1999; Varrot *et al.*, 2003b). This distortion is derived by the steric hindrance of the tyrosine residue (Tyr169 of *Tr*Cel6A) (Koivula *et al.*, 1996; Larsson *et al.*, 2005). The N- and C-terminal loops have been observed in several conformations in GH family 6 CBHs, and two major conformations (Fig. 1-7A) are designated as “open” and “closed” conformations for convenience.

The GH family 6 enzymes perform hydrolysis with inversion of anomeric configuration, which was determined experimentally by NMR studies of *Tr*Cel6A (Knowles *et al.*, 1988; Claeyssens *et al.*, 1990). Generally, inverting mechanism is considered as a single displacement mechanism, where proton donation and nucleophilic attack of water occur simultaneously (p. 6, Fig. 1-3B). In this case, the catalytic acid promotes leaving group departure and the symmetrically located catalytic base activates a nucleophilic water molecule. However, the catalytic base is not identified in GH family 6 enzymes. Currently, a Grotthuss-type proton-transferring mechanism is a leading hypothesis of the mechanism of GH family 6, which was first proposed by Koivula *et al*. in*Tr*Cel6A (Koivula *et al.*, 2002). In this mechanism, Asp221 is initially protonated and make a hydrogen bond with neighboring Asp175 (Fig. 1-7D, left). When a substrate comes into the active site, Asp221 gives up the hydrogen bond with Asp175 and protonates the glycosidic bond, and the nucleophilic water is coordinated by Ser181 on the N-terminal loop and the main chain carbonyl oxygen of Asp405 (Fig. 1-7D, middle). Serine and carbonyl group cannot accept proton from the nucleophilic water, thus the proton is transferred through an additional water and accepted by Asp175 (Fig. 1-7D, right). This proposal is on the basis of crystal structures, where the residues Asp216, Asp175 and Ser176 are observed in several conformations. Although definitive confirmations of the Grotthuss mechanism are still lacking, this hypothesis is widely accepted as the mechanism of GH family 6 (Payne *et al.*, 2015). The more detailed history and experimental basis are discussed below.

The X-ray crystal structure of *Tr*Cel6A was solved nearly three decades ago in 1990 (Rouvinen *et al.*, 1990), which was the first structure determination of cellulase. Rouvinen *et al.* noted that two aspartic acid residues, Asp221 and Asp175, are located close to the scissile bond. Catalytic acid is considered to be Asp221 and was examined by careful biochemical studies in both fungal and bacterial cellulases as summarized in Table 1-2 (Rouvinen *et al.*,

1990; Damude *et al.*, 1995; Wolfgang & Wilson, 1999; Vuong & Wilson, 2009, see original papers for other mutants such as Glu and Asn). The mutants of catalytic acid residues show almost no activities on cellulosic substrates, but substantial activities remain on substrates with good leaving groups such as 2,4-dinitrophenol-β-D-cellobioside (DNPC). The roles of Asp175 and Asp263 (and corresponding residues) are proposed that they locate close to catalytic acid

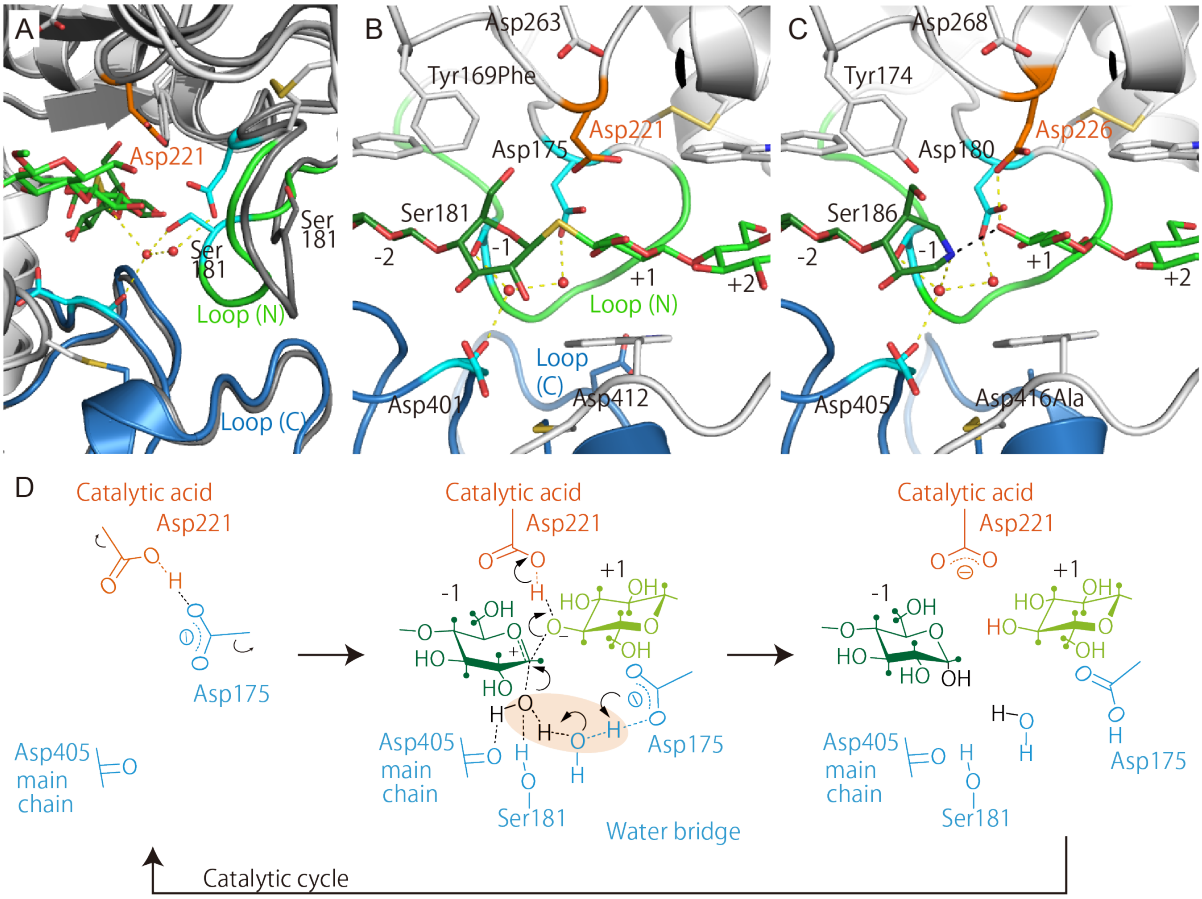

Figure 1-7. Snapshots of active site crystal structures and proposed Grotthuss mechanism. (A) Open (PDB entry 1CB2) and closed (1QJW) structure of *Tr*Cel6A. (B) Thio-oligosaccharide complex structure of *Tr*Cel6A (1QJW). (C) Cellobio-derived isofagomine complex structure of *Hi*Cel6A (1OCN). (D) Proposed catalytic mechanism with grotthuss proton transfer through a water wire (residue numbers are corresponding to *Tr*Cel6A).

Table 1-2. Biochemical studies of acid mutants

Enzyme			Mutation	Substrate	Activity (%)	Substrate (better LG)	Activity (%)	Ref
Fungal CBH	*Trichoderma reesei*	Cel6A	D221A		ND			Rouvinen, J (1990)
Bacterial EG	*Cellulomonas fimi*	Cel6A	D252A	CMC	0.0004	DNPC	80	Damude, HG (1995)
				PASC	ND			
Bacterial EG	*Thermobifida fusca*	Cel6A	D117A	CMC	0.03	DNPC	110	Wolfgan,g DE (1999)
				PASC	0.02			
				FP	9			
Bacterial CBH	*Thermobifida fusca*	Cel6B	D274A	CMC	18	DNPC	660	Vuong, TV (2009)
				PAC	1.5			
				PASC	2.7			
				BMCC	1.1			

Table 1-3. Biochemical studies of base/binding mutants

Enzyme			Mutation	Substrate	Activity (%)	Substrate (binding)	Fold	Ref
Bacterial EG	*Cellulomonas fimi*	Cel6A	D392A	DNPC	>0.005			Damude, HG (1995)
				CMC	0.003			
				PASC	0.003			
Bacterial EG	*Thermobifida fusca*	Cel6A	D265A	CMC	2.5	MU(Glc2)	No affinity	Wolfgang, DE (1999)
				PASC	2	MU(Glc3)	>0.001	
				FP	26	Glc3	ND	
Fungal CBH	*Humicola insolens*	Cel6A	D405N/A		1–0.3			Varrot, A (2002)
Fungal EG	*Humicola insolens*	Cel6B			ND			Varrot, A (2002)
Bacterial CBH	*Thermobifida fusca*	Cel6B	D497A	DNPC	2	DNPC	0.0003	Vuong, TV (2009)
				CMC	21			
				PAC	86			
				PASC	56			
				BMCC	13			

Table 1-4. Biochemical studies of base/pK_a mutants

Enzyme			Mutation	Substrate	Activity (%)	Substrate (binding)	Fold	Ref
Fungal CBH	*Trichoderma reesei*	Cel6A	D175A		20			Rouvinen, J (1990)
Fungal CBH	*Coprinopsis cinerea*	Cel6A,C	D164A,D102A	PASC	ND			Tamura, M (2012)
Bacterial EG	*Cellulomonas fimi*	Cel6A	D216A	CMC	0.8			Damude, HG (1995)
				PASC	0.08			
				DNPC	5			
Bacterial EG	*Thermobifida fusca*	Cel6A	D79A	CMC	1.3	MU(Glc2)	1.3	Wolfgang, DE (1999)
				PASC	0.7	MU(Glc3)	1.5	
				FP	42	Glc3	1.7	
Bacterial CBH	*Thermobifida fusca*	Cel6B	D226A	DNPC	32	DNPC	0.1	Vuong, TV (2009)
				CMC	110			
				PAC	20			
				PASC	8			
				BMCC	11			

Table 1-5. Biochemical studies of mutants of residues near cleavage points

Enzyme			Mutation	Substrate	Activity (%)	Proposed role in ref	Ref
Fungal CBH	*Trichoderma reesei*	Cel6A	Y169F	Glc3/Glc4	25	Ring distortion	Koivula, A (1996)
Bacterial EG	*Thermobifida fusca*	Cel6A	Y73F	CMC	8	Ring distortion	Lasson, AM (2005)
				PASC	6		
Bacterial CBH	*Thermobifida fusca*	Cel6B	Y220A	DNPC	26	Ring distortion	Vuong, TV (2009)
				CMC	ND		
				PAC	ND		
				PASC	0.9		
				BMCC	2		
Bacterial CBH	*Thermobifida fusca*	Cel6B	S232A	CMC	44	Water network	Vuong, TV (2009)
				PAC	123		
				PASC	79		
				BMCC	58		
Bacterial CBH	*Thermobifida fusca*	Cel6B	D226A/S232A	CMC	14	Double mutant of bases (asp-ser)	Vuong, TV (2009)
				PAC	1		
				PASC	3		
				BMCC	3		
Bacterial EG	*Thermobifida fusca*	Cel6A	D79N/D265N	CMC	0.23	Double mutant of bases (asp-asp)	Wolfgang, DE (1999)
				PASC	0.14		
				FP	27		
Bacterial EG	*Cellulomonas fimi*	Cel6A	D287A	DNPC	3	pKa modulation of acid	Damude, HG (1995)
				CMC	0.006		
				PASC	0.008		
Bacterial EG	*Thermobifida fusca*	Cel6A	D156A	DNPC	45	pKa modulation of acid	Wolfgang, DE (1999)
				CMC	2.2		
				PASC	1		
				FP	67		
Fungal CBH	*Humicola insolens*	Cel6A	D416A		10	Rescue of the base mutant	Varrot, A (2002)

Abbreviations in Tables 1-2 to 1-5: BMCC, bacterial micro crystalline cellulose; CMC, carboxy methyl cellulose; DNPC, 2,4-dinitrophenol-β-D-cellobioside; FP, filter paper; LG, leaving group; MU, 4-methylumbelliferyl; PAC, phosphoric acid treated cotton; PASC, phosphoric acid swollen cellulose.

and modulate its pK_a (Rouvinen *et al.*, 1990; Damude *et al.*, 1995; Wolfgang & Wilson, 1999). Exceptionally, the position of Asp79 (*Tr*Asp175) in EG *Tf*Cel6A is different from those in other GH family 6 structures. The Asp79 is ~11 Å away from the active site (Spezio *et al.*, 1993), but it also participates in the pK_a modulation of catalytic acid (Wolfgang & Wilson, 1999).

On the other hand, a candidate for catalytic base differs between fungal CBHs and bacterial EGs. Damude *et al*. performed precise kinetic studies of *Cf*Cel6A (EG) and proposed Asp392 (corresponding to Asp401 in *Tr*Cel6A) as a catalytic base (Damude *et al.*, 1995). Actually, this residue positions at opposite side of catalytic acid and its mutant is reported to cause a drastic loss of activity in EGs *Cf*Cel6A (Damude *et al.*, 1995) and *Hi*Cel6B (Varrot *et al.*, 2002) (Table 1-3). The pK_a analysis and typical Hehre resynthesis-hydrolysis kinetics in hydrolysis of α-fluoride substrates are also supportive of the existence of classical catalytic base residue in *Cf*Cel6A (Damude *et al.*, 1996). However, this proposal is suspected by several researchers. The Asp401 does not locate close to the cleavage point even though the distance between catalytic acid and Asp401 is 9.5 Å that is typical for inverting enzyme (Varrot *et al.*, 1999a). The Asp401 makes a salt bridge with a neighboring arginine, therefore probably ionized, but unlikely to participate in catalysis (Koivula *et al.*, 1996). The crystal structures with ligands at subsite -1 show that Asp401 interacts with O3 of the ligands. Several studies suggests that Asp401 is involved in substrate binding and the loss of activity came from the loss of binding ability (Wolfgang & Wilson, 1999; Vuong & Wilson, 2009) (Table 1-3). Tyrosine or arginine residues were also candidates of other possible catalytic residues in EG *Tf*Cel6A (Taylor *et al.*, 1995; André *et al.*, 2003).

The proposal of Grotthuss mechanism in 2002 provided a new understanding of the catalytic mechanism of GH family 6 that was alternative to the classical Koshland mechanism. The existence of catalytic base was investigated in *Tr*Cel6A by kinetic studies (Konstantinidis *et al.*, 1993; Becker *et al.*, 2000), and it was concluded that *Tr*Cel6A lacks a catalytic base residue in its typical sense (Koivula *et al.*, 2002). Observations of ordered water molecules near the active site led to the proposal of Grotthuss mechanism with MD calculations for confirmation of the nucleophilic water. Koivula *et al*. emphasized the role of Asp170 to stabilize a uniquely electron-deficient transition state of *Tr*Cel6A, and also mentioned the possibility of Asp175 to accept proton from the water chain as an indirect catalytic base. The crystal structures of fungal CBHs supported the existence of water network (Varrot *et al.*, 2003b; Thompson *et al.*, 2012) (Fig. 1-7C), and mutagenesis studies also suggested that the bacterial CBH *Tf*Cel6B employs this mechanism (Vuong & Wilson, 2009). Recent computational simulations also confirm that Grotthuss proton transfer occurs with and without the proton accepting residue (Mayes *et al.*, 2016). However, the water network, locations of proton accepting residue, and residues corresponding to Ser181 are not conserved in EGs (Larsson

et al., 2005; Varrot *et al.*, 2005). Cockburn *et al*. also reported additional data that are supportive again for *Cf*Asp392 (*Tr*Asp401) as catalytic base (Cockburn *et al.*, 2010).

A plenty of mutagenesis data is available in GH family 6 and only a portion of them are listed in Table 1-2 to 1-5. Catalytically important residues are well elucidated, but there are still many questions about their individual roles. Catalytic mechanisms of glycosidic bond hydrolysis are generally conserved within the same GH family because of conserved arrangements of catalytic residues, but the Grotthuss mechanism itself does not seem to be a common feature of GH family 6. Experimental proofs of the Grotthuss mechanism would be difficult and, to the contrary, an unexpectedly small or no solvent kinetic isotope effects were observed in hydrolysis of cellotriose by *Tr*Cel6A (Teleman *et al.*, 1995; Koivula *et al.*, 2002).

Fungal GH family 6 CBHs

As described in the previous sections, fungal GH family 6 CBHs (referred to as Cel6s), play a key role in crystalline cellulose degradation. Therefore, it has been questioned how Cel6s interact with cellulose and degrade it by either processive/non-processive or endo/exo modes of action. In contrast to perfectly processive GH7 CBHs, there remains much unanswered questions on the mode of action of Cel6s.

Experiments with soluble cello-oligosaccharides revealed that *Tr*Cel6A processively hydrolyzes cello-oligosaccharides, because no cellotetraose was released as an intermediate (Nidetzky *et al.*, 1994; Harjunpää *et al.*, 1996). As shown in the Fig. 1-5 (p. 10), several tryptophan residues construct the active sites numbered -2 to +4. These hydrophobic platforms allow cellulose to slide through the active site, as various intermediates observed in crystal structures (Varrot *et al.*, 2003a). The tryptophan at the subsite -2 is required for activity and has the highest affinity to glucose (Ruohonen *et al.*, 1993; Varrot *et al.*, 2003a; Payne *et al.*, 2011). On the other hand, the tryptophan at the subsite +4 is proved to be necessary only for crystalline cellulose degradation (Koivula *et al.*, 1998). Actually, the subsite +4 is not conserved in GH family 6 EGs (Spezio *et al.*, 1993; Varrot *et al.*, 2005). Additional tryptophan constructing the subsite +6 is found in bacterial CBHs such as *Tf*Cel6B (Sandgren *et al.*, 2013) or in some fungal CBHs such as Cel6C from the coprophilous fungi *Coprinopsis cinerea* (*Cc*) (Liu *et al.*, 2010). It was also shown in *Tr*Cel7A that the aromatic residues at the entrance (subsite +7) is important to degrade crystalline cellulose (Nakamura *et al.*, 2013).

The role of the C-terminal loop (Fig. 1-6, p. 12) was investigated in the *Cf*CBH, which demonstrated that the deletion of C-terminal loop residues enhance EG activity (Meinke *et al.*, 1995). At the same time, activity toward 2,4-dinitrophenol-β-D-cellobioside (DNPC) was decreased 1/5th and cellotetraose was accumulated during cellulose degradation, indicating that deletion of loop caused multiple effects. There are no other deletion experiments, but the

properties of EG *Hi*Cel6A offer an insight as well. The *Hi*Cel6A has high similarity with CBHs, including the tryptophan at the subsite +4, but lacks the C-terminal loop and does not degrade crystalline cellulose efficiently (Davies *et al.*, 2000). The coupled opening of the N-terminal loop is observed in the crystal structure (Fig. 1-6, top left, p. 12), though there is still possibility of conformational change of loops with ligand binding.

The endo-mode initiation is often proposed as an important feature of Cel6s' activities. Ståhlberg *et al*. revealed that *Tr*Cel6A creats new chain ends in amorphous cellulose and filter paper (Ståhlberg *et al.*, 1993). Boisset *et al*. observed fragmentations of bacterial cellulose ribbons and proposed that *Hi*Cel6A is an endo-processive enzyme (Boisset *et al.*, 2000). Crystal structures support this endo-processive character, as the active site tunnels of Cel6s are shorter than those of Cel7s (Fig. 1-5, p.10). In addition, unique surface loops and catalytic residues were observed in several conformations to show open and closed active site structures (Fig. 1-7, p. 14) (Varrot *et al.*, 1999a, 1999b; Zou *et al.*, 1999). Binding studies demonstrated that glucose binding enhances binding of oligosaccharides, indicating that an induced conformational change occurrs in subsites (Van Tilbeurgh *et al.*, 1989).

However, there are still large gaps in understanding the modes of action in actual degradation of crystalline cellulose. The Cel6s was currently observed at single molecule level by high-speed AFM (Igarashi *et al.*, 2011) and by total internal reflection fluorescence microscopy (TIRFM) (Nakamura *et al.*, 2016). Many *Tr*Cel6A molecules were found on the crystalline cellulose surface in the AFM experiment, but processive movements like *Tr*Cel7A were not observed for *Tr*Cel6A. Since starting points of *Tr*Cel7A movements seemed to be increased by addition of *Tr*Cel6A, endo-type attack of crystalline cellulose by *Tr*Cel6A was suggested (Igarashi *et al.*, 2011). On the other hand, Nakamura *et al.* observed two types of movements of *Tr*Cel6A on crystalline cellulose surface. The fast movements of *Tr*Cel6A (30–40 nm/s) suggested a diffusional searching of attacking points on cellulose surface. The rates of slower movements (~9 nm/s) were similar to those of Cel7s observed by high-speed AFM (7–15 nm/s) (Nakamura *et al.*, 2014), and the estimated processivity was as large as ~70 (Nakamura *et al.*, 2016).

It is a general agreement that Cel6 itself has ability to processively perform reactions, as shown in experiments with soluble cello-oligosaccharides. The asymmetrical active site structure indicates that the directionality of its motion is from non-reducing end toward the reducing end. However, its mode of action on crystalline surface is still under discussions, and may be highly dependent on cellulosic substrates. Interestingly, the laboratory hydrolysis experiments of cellulose often report that the synergistic performance of Cel7 and Cel6 is maximized in approximately one-to-one or in higher proportion of Cel6s (Tomme *et al.*, 1988; Rosgaard *et al.*, 2007; Lantz *et al.*, 2010), while many fungi produce Cel7s more than Cel6s.

Although processivity has been thought as a key factor for CBHs to be efficient in crystalline cellulose degradation, it does not seem to be a necessary strategy in the case of Cel6s.

Cel6A from the basidiomycete *Phanerochaete chrysosporium*

Phanerochaete chrysosporium (*Pc*) is an extensively studied white-rot basidiomycete, because of its ability to degrade all major components of plant cell walls including cellulose, hemicellulose and lignin. The genome sequence was published in 2004, as a first genome representing basidiomycete (Martinez *et al.*, 2004). Its cellulolytic enzymes were studied by Eriksson and Petterson (Eriksson & Pettersson, 1975) and one GH family 6 cellobiohydrolase (referred to as *Pc*Cel6A) was purified and characterized in 1991 (Uzcategui *et al.*, 1991). The cloning of *Pc*Cel6A gene was done in 1994 and revealed that *Pc*Cel6A also consists of a catalytic domain (CD) and an N-terminal carbohydrate-binding module (CBM) connected by a linker region, with the 54 % sequence similarity to *Tr*Cel6A (Tempelaars *et al.*, 1994).

Characterizations of cellulases have been majorly performed with native cellulose materials, the allomorph cellulose I. On the other hand, Igarashi *et al.* demonstrated that *Pc*Cel6A is quite efficient in degradation of crystalline cellulose III_I that was obtained by ammonia treatment of *Cladophora* cellulose (Igarashi *et al.*, 2012). Cellulose III_I is an allomorph that has inter-sheet hydrogen bonding patterns but less inter-sheet stacking interactions, which differs from native cellulose I_α and I_β (Wada *et al.*, 2004, Fig. 1-8A). Several studies demonstrated that cellulose III_I is more susceptible to enzymatic hydrolysis, and pointed out the importance of conversion of crystalline allomorph for better saccharification of cellulosic biomass (Igarashi *et al.*, 2007, 2012; Beckham *et al.*, 2011; Chundawat *et al.*, 2011; Gao *et al.*, 2013). In the case of Cel7, the enhancement of activity in cellulose III_I was proposed as the higher accessible surface area (available lanes for processive movement) because no large differences were found in binding and dissociation rate constants and velocity of processive movement (Shibafuji *et al.*, 2014). In degradation of *Cladophora* cellulose III_I, the hydrolysis

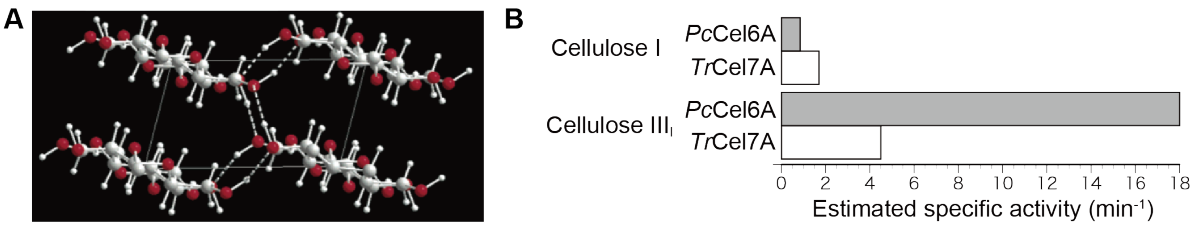

Figure 1-8. (A) Structure of crystalline cellulose III_I. The picture was taken from Wada *et al.*, 2004**. (B) The activities of *Pc*Cel6A and *Tr*Cel7A toward cellulose I and III_I** (Igarashi *et al.*, 2012)**. Igarashi *et al* proposed surface density (ρ) that is calculated by dividing adsorbed enzyme by maximum amount of that. The figure shows estimated specific activity at ρ→0.**

velocity of *Pc*Cel6A was estimated ~4 times higher than that of *Tr*Cel7A (Igarashi *et al.*, 2007, 2012, Fig. 1-8B). Igarashi *et al*. speculated that the enhancement of hydrolysis for *Pc*Cel6A is due to the susceptibility of cellulose III_I to its endo-type attack. However, the mechanism of endo-type attack of crystalline cellulose is still unclear. Typical EGs are not capable of degrade cellulose III_I, thus the mechanism of Cel6s is of great interest. The rate-limiting point of crystalline cellulose by Cel6s is probably different from that by Cel7s.

According to the studies on Cel7s and GH family 18 chitinases, the efficiency toward crystalline substrates, processivity, and ligand binding affinity are closely related factors. High processivity is achieved by strong substrate binding at enzyme subsites, which results in slower speed of enzyme movements and lower dissociation rates (Horn *et al.*, 2006; Payne *et al.*, 2013a; Nakamura *et al.*, 2014). Therefore, processivity becomes a cost in degradation of susceptible substrates, and it is one of the promising strategy to convert substrates into an easily degradable form and use enzymes with low-processivity and low-adsorption property to achieve rapid hydrolysis. Cel6s would play the expected role for more efficient cellulose degradation, and protein engineering of Cel6s would lead to major advances in cellulosic biomass utilization.

1.4. Aims

The *Pc*Cel6A has been investigated in this study from the following aspects: the catalytic mechanism of hydrolysis of β-1,4-glycosidic bond in the cellulose molecular chain, and the modes of action toward crystalline cellulose. Although the three-dimensional structures and plenty of biochemical data have been available for GH family 6 cellulases, these molecular mechanisms are not yet fully understood and there are still challenges for structure-based protein engineering.

GH family 6 cellulases have several conserved aspartic acid residues in their active sites, whose roles are still under discussions, and the questions whether or why GH family 6 utilizes the unique Grotthuss mechanism are still unanswered. One interesting feature is the conformational changes of catalytic residues and tunnel-enclosing loops. Tunnel-enclosing loops are known to be generally important for CBH activities, but those of GH family 6 are not restricted in just covering the active sites for processivity. Therefore, in chapter 2 and 3, Cel6A from *Phanerochaete chrysosporium* (*Pc*Cel6A) was studied by X-ray crystallography and its catalytic mechanism was discussed. At first, the X-ray crystal structure of *Pc*Cel6A was solved and flexibility of the loops was investigated (chapter 2). In order to further elucidate the roles of catalytic resides, X-ray crystal structures of *Pc*Cel6A was determined at sub-atomic resolution (chapter 3). The presented result is the first insight of protonation states of aspartic acid residues in GH family 6, because protonation states of residues cannot be determined by

X-ray crystallography in typical data quality. Based on these new information, I propose that the conformational change of N-terminal loop is necessarily associated with the catalytic proton transfer in *Pc*Cel6A.

As is the case with catalytic hydrolysis mechanism, understanding of the structure-function relationships responsible for crystalline cellulose degradation is not easy. Non-theoretical approach, such as random mutagenesis, is a powerful tool for protein engineering. By mimicking the natural evolution process, designing of enzymes for desired properties are possible. However, one of the technical difficulties of CBH engineering is that activities toward soluble substrates does not correlate with activities toward crystalline cellulose. Therefore, the screening of CBHs should be performed in appropriate substrates, but the direct screening by insoluble substrate requires substantial extracellular protein expressions. Therefore, in chapter 4, a new random mutagenesis technique was developed aiming to engineer the properties of *Pc*Cel6A. In this method, random mutagenesis can be performed by using the methylotrophic yeast *Pichia pastoris*, which is excellent host for fungal extracellular enzymes. By direct screening with crystalline cellulose III_I, the *Pc*Cel6A mutants with altered properties were found and their crystalline cellulose degradation activities were discussed.

2. Analysis of the overall structure of *Pc*Cel6A

(Paper 2)

Although biochemical studies and protein engineering works of *Pc*Cel6A have been conducted by several research groups (Heinzelman *et al.*, 2009; Igarashi *et al.*, 2012; Ito *et al.*, 2013), the three-dimensional structure has not been reported yet. Here, the X-ray crystal structure of CD of *Pc*Cel6A was determined in its apo-form and in the complex with a ligand, cellobiose.

2.1. Enzyme production, crystallization and structure determination

The CD of *Pc*Cel6A, residues 82–439, was produced in *P. pastoris* using 5-liter jar-fermenter. The concentration of crude protein reached ~2 g/L and the culture supernatant gave two apparent bands on SDS-PAGE gel, which arised from differential glycosylation. The protein solution was purified by hydrophobic chromatography (Phenyl-Toyopearl 650S) and anion exchange chromatography (DEAE-Toyopearl 650S) as a single band with a molecular weight of 37 kDa. Finally, the protein solution was dialyzed into 5 mM Tris-HCl buffer, pH 7.5, containing 100 mM NaCl. Crystallization was performed by the sitting-drop vapor diffusion method with the reservoir solution composed of 20% (w/v) polyethylene glycol 3350, 200 mM calcium acetate, 50 mM acetate buffer, pH 5.0, and 10% (w/v) 2-methyl-2,4-pentandiol. To introduce ligands, a crystal was incubated with *p*-nitrophenyl β-D-cellotrioside (*p*NPG3) for 10 hours prior to data collection. The structure was solved by molecular replacement using a model structure generated by *Phyre2* server (Kelley & Sternberg, 2009).

The structure of the CD of native *Pc*Cel6A (hereafter *Pc*Cel6A-apo) and the structure from *p*NPG3-soaked crystal which contained one α-cellobiose molecule (therefore referred to as *Pc*Cel6A-G2), were determined at 1.2 Å and 2.1 Å resolution, respectively. Both belonged to space group $P2_12_12_1$ with one molecule in the asymmetric unit, and the *Pc*Cel6A-apo and *Pc*Cel6A-G2 were refined to R_{work}/R_{free} values of 13.8/15.9% and 17.7/23.5%, respectively. Summary of data-collection and refinement statistics are shown in Table 2-1.

2.2. Overall structure

The overall structure of CD of *Pc*Cel6A consists of a distorted seven-stranded β/α_8-barrel, where a tunnel-like active site is enclosed by pair of loops, N- and C-terminal loops, as other fungal CBHs (Fig. 2-1). The subsite -2 to +4, all catalytically important residues and two

disulfide bridges were conserved in *Pc*Cel6A. Structural similarity search by the *Dali* server (Holm & Rosenström, 2010) revealed that the greatest similarity of the overall structure of *Pc*Cel6A to previously determined structures of Cel6A from *Coprinopsis cinerea* (*Cc*Cel6A, Tamura *et al.*, 2012): the sequence similarity of *Pc*Cel6A and *Cc*Cel6A was 65% and the root mean square deviation (r.m.s.d.) for 357 C-α atoms of *Pc*Cel6A-apo to *Cc*Cel6A-apo (PDB entry 3VOG) was 0.7 Å (Z score = 60.8), and that of *Pc*Cel6A-G2 to *Cc*Cel6A-*p*NPG3 (PDB entry 3VOI) was 0.9 Å (Z score = 61.3). The CD of *Pc*Cel6A has one potential *N*-glycosylation site, but no electron density of sugars was visible in the structures, indicating that the protein used to prepare the crystals are non-glycosylated.

Table 2-1. Summary of X-ray data-collection and refinement statistics

	***Pc*Cel6A-apo**	***Pc*Cel6A-G2**
Data collection		
Resolution (Å)[a]	50.00–1.20 (1.22–1.20)	50.00–2.10 (2.14–2.10)
Completeness (%)[a]	99.9 (99.2)	99.5 (99.0)
Redundancy (%)[a]	3.1 (2.8)	6.8 (6.8)
Average I/σ (I)[a]	43.7 (5.9)	15.5 (3.1)
R_{sym} (%)[a]	5.5 (28.8)	7.2 (48.4)
Mosaicity range (˚)	0.26–0.32	0.79–1.27
Refinement		
R_{work}/R_{free} (%)	13.8/15.9	17.7/23.5
PDB entry code	5XCY	5XCZ

The data sets were processed using the *HKL2000* (Otwinowski & Minor, 1997) and molecular replacement was done by *MOLREP* (Winn et al., 2011) with a model structure obtained by *Phyre2* (Kelley & Sternberg, 2009). Refinement were performed *Phenix* (Adams et al., 2010) and by *Coot* (Emsley et al., 2010). [a] Values in parentheses are for the highest resolution shell.

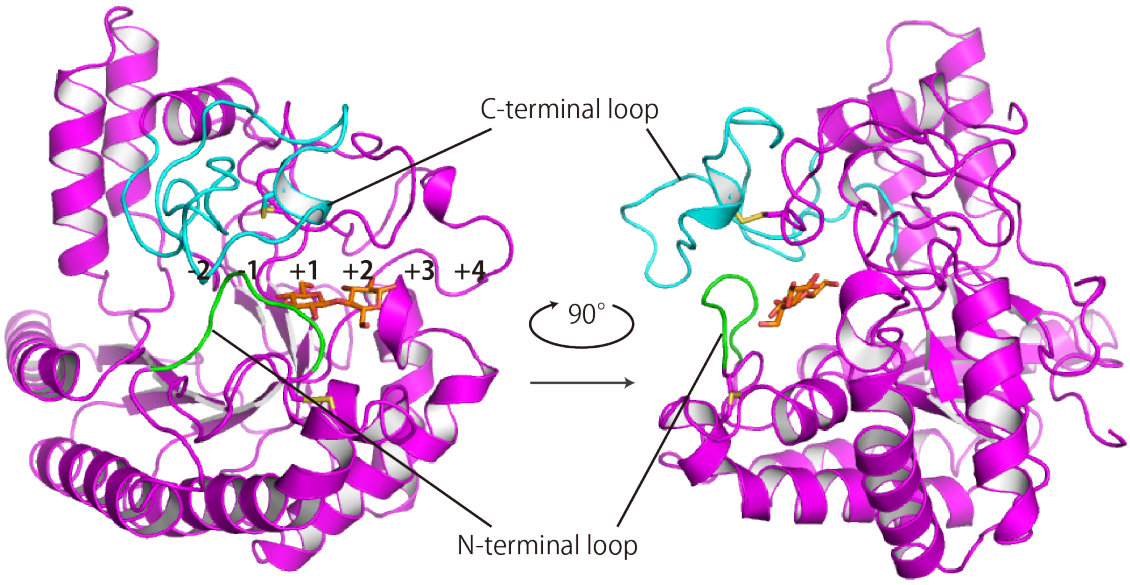

Figure 2-1. Overall structure of *Pc*Cel6A with cellobiose bound at subsites +1 and +2. The N-terminal and C-terminal loops covering the catalytic center are colored in green and cyan, respectively.

2.3. Ligand binding site

When the *Pc*Cel6A crystal was incubated with *p*NPG3 under the crystallization conditions at pH 5, α-cellobiose is observed in subsites +1 and +2 as a product of the hydrolysis of *p*NPG3 (Fig. 2-2). The observation of α-cellobiose in the subsites is plausible because the GH family 6 enzymes perform hydrolysis of β-1,4-glycosidic bonds with inversion of anomeric configuration, and are known to cleave chromophoric cello-oligosaccharides to produce cellobiose units (Claeyssens *et al.*, 1989). The cleavage pattern of *p*NPG3 in this structure is different from the *Cc*Cel6A structure in which non-hydrolyzed *p*NPG3 bound at subsites +1 to +4 (PDB: 3VOI, Tamura *et al.*, 2012), and from the *Cc*Cel6C structure in which two *p*NPG2 molecules bind at subsites -3 to -1 and +1 to +3 in *Cc*Cel6C (PDB: 3ABX, Liu *et al.*, 2010).

Other small molecules were not found with a clear electron density map, but unmodeled electron density blob was found at the product subsite -1 of *Pc*Cel6A-G2. In GH family 6 enzymes, the subsite -1, where the sugar is distorted from a stable chair conformation for catalysis, is often occupied by molecules other than waters and sugars, such as cations and low-molecular weight compounds. In *Pc*Cel6A-G2, tris(hydroxymethyl)aminomethane (tris) molecule with multiple conformations appears to be the most favorable model and found to make hydrogen bonds with O4 of glucose at subsite +1, OD1 of Y164, NZ of K388, carbonyl O, OD1 of D394, and several water molecules. No water molecule corresponding to catalytic water was found in the structure, probably because they were excluded from catalytic site by the putative tris molecule.

2.4. Loop flexibility

The existence of tunnel-enclosing loops is a distinct feature of GH family 6 CBHs, and the flexibility and mobility of the loops must be important for their catalytic activity. The N- and C-terminal loops were observed as “open” conformations in the *Pc*Cel6A-apo and a “closed” conformation in the *Pc*Cel6A-G2. In the *Pc*Cel6A-apo structure, the both loops modeled as

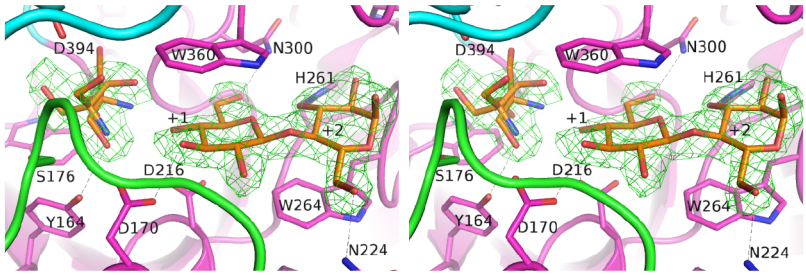

Figure 2-2. Close-up view of subsites +1 and +2 of *Pc*Cel6A with cellobiose as a ligand. The $|F_o|-|F_c|$ map was calculated without ligand atoms and contoured at the 3 σ level. Tris molecule at subsite -1 was modeled in double conformations.

double conformations with similar occupancies ranging from 0.46 to 0.54 (Fig. 2-3A). In contrast, these loops were observed as a single conformation in the *Pc*Cel6A-G2. The two conformations of *Pc*Cel6A-apo and the *Pc*Cel6A-G2 structure were superposed with each other (Fig. 2-3), in order to further examine the flexibility and mobility of the loops.

The *B*-factors represent the significance of vibrations in crystal structures and are a good indicator of flexibility of residues. Thus, relative *B*-factor was calculated by dividing the average *B*-factor of each residue by that of the whole protein (11.6 $Å^2$ for the *Pc*Cel6A-apo and 20.0 $Å^2$ for the *Pc*Cel6A-G2) and were visualized by means of a blue-red color scale in Fig. 2-3. The residues of the N- and C-terminal loops have consistently high *B*-factors in *Pc*Cel6A-apo, which indicates the flexibility of both loops in the open structure. The r.m.s.d values of two conformations, which were visualized in proportion to the thickness of worm representations, revealed the different characteristics of two loops. The most dynamic conformational change occurred in the N-terminal loop concomitantly with ligand binding. On the other hand, the r.m.s.d. values between the *Pc*Cel6A-apo and *Pc*Cel6A-G2 were similar to that between the two conformations of the *Pc*Cel6A-apo.

Therefore, both loops showed clear flexibility, but their mobility and response to ligand binding were different. The flexibility of both loops in open structure is speculated to be a machinery to allow easier acquisition and sliding of cellulose chains, which should be important for efficient initiation of catalysis. At the same time, *B*-factor of both loops was similar to the average value of whole protein region in closed conformation. As these loops cover the subsite

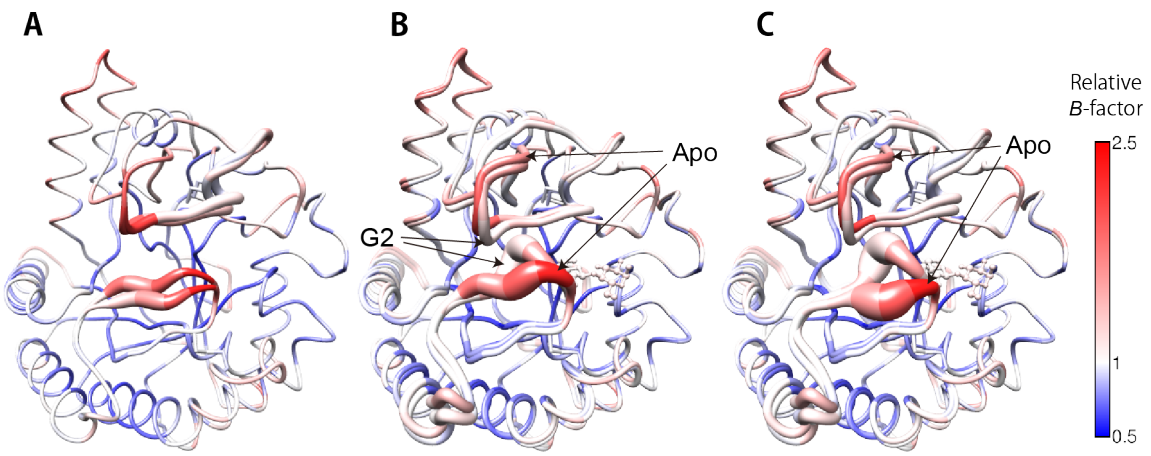

Figure 2-3. Superposition of the apo- and cellobiose-bound structures of *Pc*Cel6A. (A) The apo structure was observed in multiple conformations: “relatively open” and “most open” conformations are shown. (B) Superposition of the relatively open apo-structure and the cellobiose-bound structure. (C) Superposition of the most open apo-structure and the cellobiose-bound structure. The width of the ribbon in each figure represents the r.m.s.d. value between the superposed conformations. The relative *B*-factors were calculated for each apo- and cellobiose-bound structure.

-1, the protein should not be too flexible in order to secure the positioning of catalytic residues and to distort substrates at the transition state. The C-terminal loop is placed relatively apart from the ligand binding sites with no direct interaction between the ligand and residues in the flexible loop region (Fig. 2-1), and the position of C-terminal loop was not drastically changed regardless of the ligand binding. Its stable positioning may be related to the efficiency of crystalline cellulose degradation as the C-terminal loop probably interacts the crystalline cellulose surface and plays a role to anchor *Pc*Cel6A on it and to keep a single cellulose chain detached. Interestingly, the rigidity of loop seems to be an evolutionary acquired feature of GH family 6 CBHs. The *Cc*Cel6C, which is phylogenetically distinct from CBHs (Fig. 1-6, p. 12), is capable to open and close the C-terminal loop. *Cc*Cel6C is constantly produced enzymes, has higher activity toward CMC and lack the CBM (Liu *et al.*, 2009, 2010; Tamura *et al.*, 2012), indicating that it does not specialized to degrade crystalline cellulose. In *Pc*Cel6A, the disadvantage in the endo-initiation may be canceled out by the advantages in crystalline cellulose degradation. In contrast, the N-terminal loop of *Pc*Cel6A drastically moves into the catalytic site in response to the ligand binding as well as other GH family 6 CBHs. It is speculated that the large movement of N-terminal loop is closely related to the chemical hydrolysis reaction, and further investigations are required to elucidate its precise role and mechanism.

3. Analysis of the active site residues of *Pc*Cel6A

(Paper 3 and supplement)

Although many chemical reactions catalyzed by GHs involve proton transfer, protonation states of residues are not known in many enzymes including GH family 6. This is because protein structures are usually determined by X-ray crystallography and the typical data quality (resolution) does not provide information about hydrogen atoms (Afonine *et al.*, 2010). However, protonation states and hydrogen bonding networks in proteins are important to understand precise catalytic mechanism of enzymes. In this study, protonation states of the aspartic acid residues of *Pc*Cel6A were determined by subatomic resolution X-ray crystal structures. The role of each residue and catalytic mechanism were discussed.

3.1. Quality of crystals grown in space

Preparation of crystals with high resolution and low mosaicity is indispensable to the accurate structure determination. In order to obtain high-quality crystals, microgravity experiments were performed at protein crystallization research facility of “Kibo”. In microgravity conditions, the convection in crystallization solution is reduced by several orders of magnitude, which results in the stable depletion zone around crystals and low incorporation of impurities (Lorber, 2002). The crystallization conditions were optimized from that described in the section 2 to be suitable for the counter diffusion method. The CD of native *Pc*Cel6A was crystallized at its optimum pH range (pH 4–6, determined by using fluorogenic substrates developed by Wu *et al*. (M. Wu *et al.*, 2013b), Supplementary Fig. S1): at pH 5.25 in the substrate-free condition (hereafter *Pc*Cel6A-apo) and at pH 4.5 with cellotriose (*Pc*Cel6A-G3) as shown in Table 3-1. The microgravity and ground control experiment were conducted with the crystallization tool specialized for JAXA-PCG experiments (JCB-SGT, Fig. 3-1A, Takahashi *et al.*, 2013; Tanaka *et al.*, 2004). Aiming at large crystal growth, the capillary with an inner diameter of 2 mm (SLC, Fig. 3-1B) was used for *Pc*Cel6A-G3.

The sizes of crystals obtained in microgravity experiments were bigger relative to those obtained in ground experiments (Fig. 3-2). The maximum length of the *Pc*Cel6A-apo crystal and *Pc*Cel6A-G3 crystal were ~1 mm and ~0.6 mm, both of which were obtained in microgravity. The highest resolution of *Pc*Cel6A-apo and *Pc*Cel6A-G3 was 0.8 Å and 0.85 Å, respectively. The highest resolution was obtained from both microgravity and ground experiments for *Pc*Cel6A-apo, but that was obtained only from microgravity in the case of

*Pc*Cel6A-G3. The total data quality including resolutions, mosaicities and *R* values of microgravity-grown crystals was always superior to that of ground-grown crystals.

Therefore, in the case of *Pc*Cel6A, the microgravity condition was effective in both data quality and the sizes of crystals, as reported in several papers (Tanaka *et al.*, 2011; Takahashi *et al.*, 2013; Ng *et al.*, 2015). The availability of large, high-quality crystals is quite important for accurate structure determination, because the X-ray radiation damage is more significant at high resolution than in medium and low resolutions (Takeda *et al.*, 2010). Radiation damages result in increase of atom *B*-factors, reduction of disulfide bridges, decarboxylation of carboxyl residues, and lowered occupancies of water molecules, all of which lead erroneous interpretations of enzyme mechanisms (Holton, 2009). The increasing risk of X-ray radiation damage can be avoided by positional changes of exposed regions during data collection using a homogenous crystal. Takeda *et al*. demonstrated that the X-ray dose for high resolution data collection should be in the order of 10^5–10^6, which is 1/100 of the previously proposed limit for medium resolution.

Table 3-1. Crystallization conditions

	***Pc*Cel6A-apo**	***Pc*Cel6A-G3**
Precipitant	50 mM Acetate buffer (pH 5.3), 30% PEG3350, 10% MPD, 100 mM Calcium acetate	50 mM Acetate buffer (pH 4.5), 40% PEG3350, 200 mM NaCl, 5 mM Cellotriose
Protein solution	50 mg/ml Protein, 100 mM NaCl, 5 mM Tris-HCl, 5% PEG3350	40 mg/ml Protein, 10 mM Cellotriose 90 mM NaCl, 4 mM Tris-HCl, 4% PEG3350,
JAXA-PCG No.	#2-2	#2-4, #2-5
Crystallization tool	JCB (dia. 0.5 mm)	JCB (dia. 0.5 mm), SLC (dia. 2 mm)

[a] Microseeding was used in all conditions. Abbreviations: PEG, polyethylene glycol; MPD, 2-methyl-2,4-pentandiol.

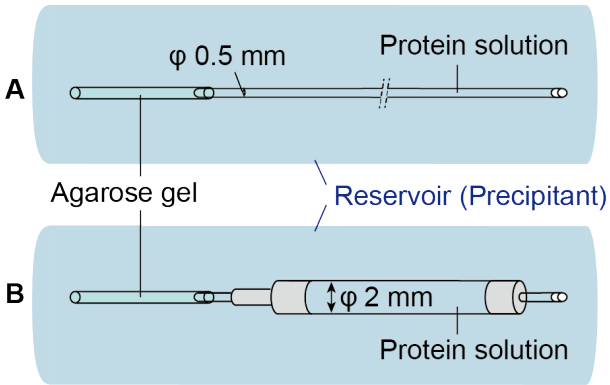

Figure 3-1. Counter diffusion system for usual JAXA-PCG experiments, JCB-SGT (A) and large-volume crystal growth, SLC (B)

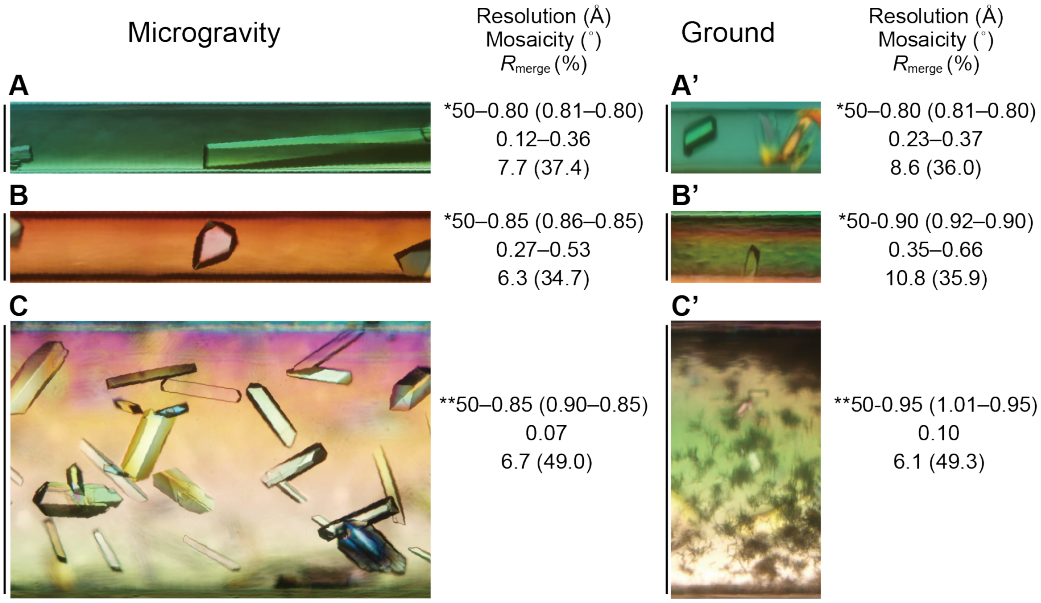

Figure 3-2. Crystals grown in microgravity (left) and ground (right) experiments. (A, A') *Pc*Cel6A-apo (Capillary diameter: 0.5 mm, JAXA-PCG#2-2) (B, B') *Pc*Cel6A cocrystallized with cellotriose (Capillary diameter: 0.5 mm, JAXA-PCG#2-4) (C, C') *Pc*Cel6A cocrystallized with cellotriose (Capillary diameter: 2 mm, JAXA-PCG#2-5). *X-ray data sets were processed by *HKL2000* (Otwinowski & Minor, 1997). **X-ray data sets were processed by *XDS* (Kabsch, 2010).

3.2. Structures in open and closed forms

Structural analysis of *Pc*Cel6A-apo and *Pc*Cel6A-C3 was performed by using X-ray data sets that were collected from the crystals of Fig. 3-2A and Fig. 3-2C, respectively. In order to reduce the radiation damage, X-ray data sets were collected by changing positions in single crystals. Maximum doses were estimated by *RADDOSE* program (Zeldin et al., 2013) to be 7 x 10^5 Gy and 6 x 10^5 Gy for *Pc*Cel6A-apo and *Pc*Cel6A-C3, respectively. The *Pc*Cel6A-apo and *Pc*Cel6A-C3 was refined to R_{work}/R_{free} values of 11.1/12.4%. and 9.9/11.4%, respectively. Summary of the data collection and refinement statistics were shown in Table 3-2 (details were shown in Supplementary Table S1).

The overall structures of *Pc*Cel6A-apo and *Pc*Cel6A-G3 were the "open" and "closed" conformations, respectively (Fig 3-3A). In *Pc*Cel6A-apo, three 2-methyl-2,4-pentandiol (MPD) molecules and three acetate molecules, which were contained in the crystallization solutions at 10% and 50 mM, respectively, were found in the active site tunnel and near the surface of protein (Supplementary Fig. S2). For *Pc*Cel6A-G3, a β-glucose was found at subsite -2 and a cellobiose was found at subsites +1 to +2 (Fig 3-3B to D). The cellobiose molecule was modeled as double conformations of α- and β- form with occupancy of 0.58 and 0.42, respectively (Fig. 3-3D). The cellotriose is generally hydrolyzed slowly by GH family 6 cellulases, and these ligands were considered as products. The hydrolysis may occur when

subsites -2 to +1 were occupied, therefore the products were considered to be transformed to the high affinity sites after hydrolysis. The subsite -1 was empty, but a sodium ion (Na^+) was found coordinated with water molecules and the O4 atom of glucose unit of subsite +1 (Glc_{+1}-O4) (Fig. 3-3E). The binding of Na^+ is likely an artifact because sodium chloride was contained in the crystallization condition at 200 mM. Similar case was also reported in Cel6A (CBH) from *Chaetomium thermophilum*, where a lithium ion occupied at subsite -1 (Thompson *et al.*, 2012). The overall structures of *Pc*Cel6A-apo and *Pc*Cel6A-G3 was similar to the published *Pc*Cel6A structures in open- (PDB entry 5CXY) and closed-form (PDB entry 5CXZ) with r.m.s.d values of C-α atoms of 0.16 and 0.18, respectively. In contrast to the previous open-form structure (5CXY), the only narrower conformation of N-terminal loop was found in *Pc*Cel6A-apo (Supplementary Fig. S3).

3.3. Bond distance analysis of carboxyl groups

In the subatomic-resolution structures, each of atoms could be identified separately in contrast to the structures that were previously determined at medium resolutions (1.2 Å and 2.1 Å). The hydrogen atoms with low vibrations such as hydrogens of C-H bonds were visible in residual $|F_o|-|F_c|$ map (Fig. 3-3C to E). However, low intensity of hydrogens in X-ray is not reliable enough to discuss the protonation states of residues. Therefore, bond length analysis was developed as a promising method to determine the protonation states of residues by taking an advantage of high accuracy of heavy atom positions (Ahmed *et al.*, 2007; Fisher *et al.*, 2012). Carboxylic acid residues have different bond length in its neutral form and

Table 3-2. Summary of X-ray data-collection and refinement statistics

	***Pc*Cel6A-apo**	***Pc*Cel6A-G3**
Data collection		
Resolution (Å)[a]	50.00–0.80 (0.82–0.80)	50.00–0.85 (0.90–0.85)
Completeness (%)[a]	97.4 (93.0)	98.7 (92.4)
Redundancy (%)[a]	4.3 (3.4)	6.5 (4.5)
Average I/σ (I)[a]	8.2 (1.7)	15.2 (1.7)
R_{sym} (%)[a]	9.6 (92.8)	7.4 (83.1)
Mosaicity range (˚)	0.07–0.27	0.05–0.17
Refinement		
R_{work}/R_{free} (%)	11.1/12.4	9.9/11.4

The data sets were processed and scaled using *XDS* (Kabsch, 2010). The initial refinement with isotropic parameters was performed using *REFMAC* in the *CCP4* suite (Winn *et al.*, 2011). Further refinements were performed with the *SHELXL* (Sheldrick, 2015). Manual model rebuilding were performed using *Coot* (Emsley *et al.*, 2010). [a] Values in parentheses are for the highest resolution shell.

ionized form, whose values in Cambridge structural database (CSD) is 1.31 Å for O-H bond, 1.21 Å for C=O bond, and 1.26 Å for C-O$^-$ bond (Fig. 3-4A). In this study, unrestrained refinement was conducted for Asp and Glu residues of *Pc*Cel6A according to the method described by Fisher *et al*. (Fisher *et al.*, 2012).

Hydrogen bond of acid (Asp216) and putative base (Asp176) in apo enzyme

In *Pc*Cel6A-apo structure, the catalytic acid Asp216 was observed in the inactive conformation by flipping away from the tunnel center (Fig. 3-3B, dark grey). Asp216 was in a hydrogen bonding distance of 2.54 Å with Asp170, the pK_a modulator and indirect proton acceptor residue. This hydrogen bond must be broken prior to catalysis and is likely playing some roles because the proximity of these catalytic aspartates is conserved in 9 of 11 known

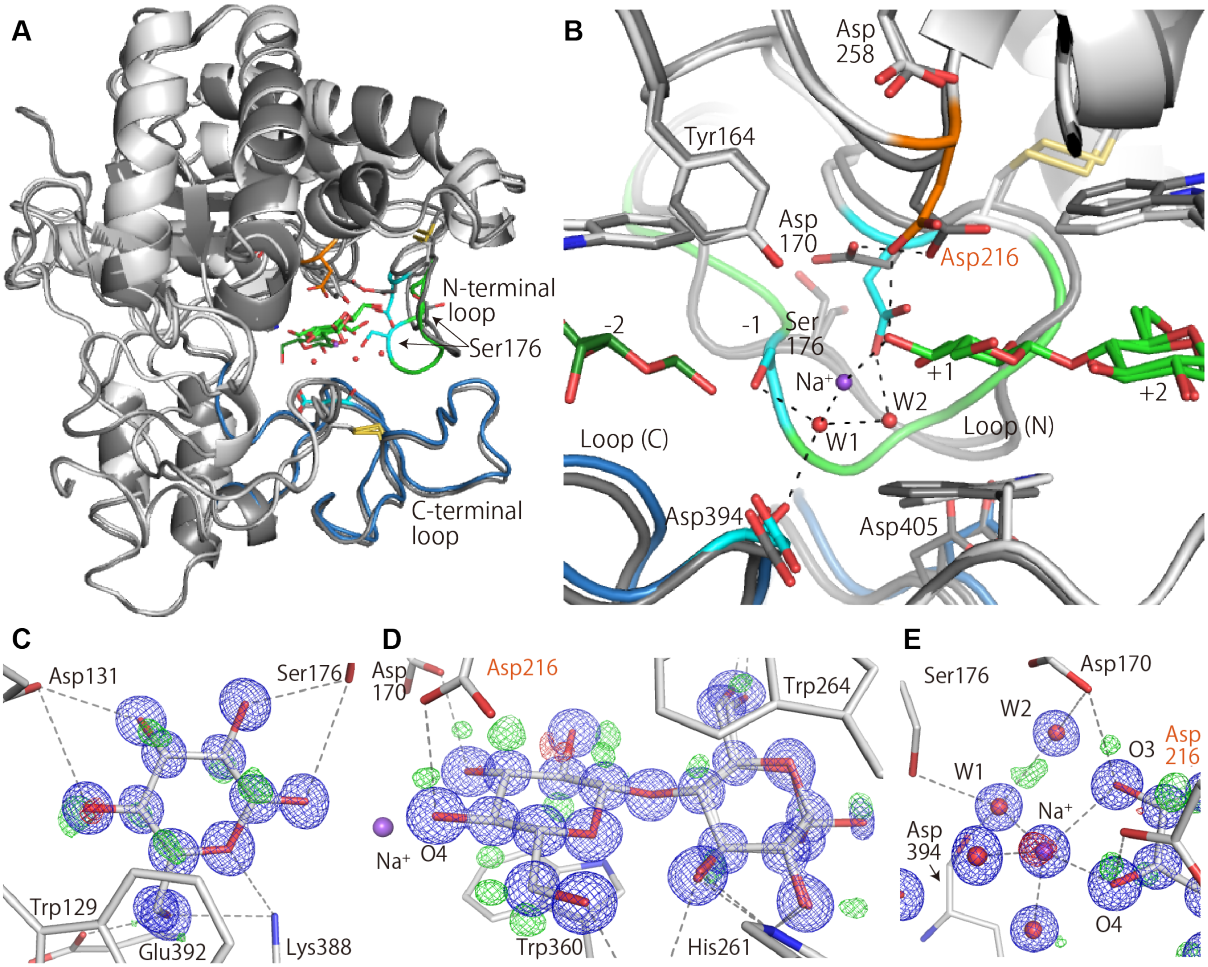

Figure 3-3. Overall (A) and active site (B) structures of the apo and ligand-complex. *Pc*Cel6A-apo was colored in dark gray. *Pc*Cel6A-G3 was colored in light gray, orange (catalytic acid, Asp216), cyan (Asp170, Ser176, Asp394) and green (ligand). The putative catalytic water and chain water are indicated as W1 and W2, respectively. **The glucose (C), cellobiose (D) and Na$^+$ (E) found in *Pc*Cel6A-G3.** $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps were countered at 2 and 3 σ levels, respectively. The distances to the Na$^+$ ion was 2.34 Å for Glc$_{+1}$-O4, 2.59 Å for Glc$_{+1}$-O3, 2.31 Å for W1, and 2.31 and 2.33 Å for the other two waters. These lengths are typical for Na$^+$ ion interactions (Harding, 2002; Rupp, 2009; Heather B. Mayes *et al.*, 2014).

structures in GH family 6 (listed in Supplementary Table S2). In order to understand the properties of this catalytic aspartates, the bond distances of pairs of carboxylic residues were compared in Fig. 3-4B. In addition to the Asp216-Asp170 pair, there are three carboxylic acid pairs in *Pc*Cel6A, Asp165-Glu179, Glu392-Glu101 and Asp412-Asp359 (Supplementary Fig. S4). These residues locate close to the N- or C-terminal loops, and engineering of corresponding residues in *Tr*Cel6A proved that these residue are responsible for local stabilizations especially at acidic pH (Wohlfahrt *et al.*, 2003; Wohlfahrt, 2005).

The most clearly protonated aspartic acid residue was Asp165 with OD1 bond length of 1.21 Å and 1.32 Å, which were in 2.57 Å distance from ionized Glu179 (Fig. 3-4C). In contrast, the carboxylic pairs of Glu392-Asp101 and Asp412-Asp359 (Fig. 3-4D, E), had shorter hydrogen bonding distance of 2.52 Å and 2.43 Å, and the difference in bond length of these residues were similar to each other. It was indicated that they have similar pK_a values and therefore are sharing protons with each other. In the case of Asp216-Asp179 pair (Fig. 3-4B), bond lengths were determined with estimated standard deviation (ESD) of ~0.007 Å. Asymmetric OD1 lengths of 1.24 Å and 1.30 Å indicated that the residue is protonated. In contrast, those of Asp170 was rather symmetric, 1.27 Å and 1.25 Å, indicating that the residue is negatively charged. Therefore, this result means that the proton is possessed by Asp216 when they are making hydrogen bond in *Pc*Cel6A-apo, which is feasible for Asp216 to work as catalytic acid.

Protonation state changes and loop conformations

In *Pc*Cel6A, the aspartic acid residues, Asp216 (catalytic acid), Asp170, Asp258, Asp394 and Asp405 are located in the catalytic site (Fig. 3-3B), and their determined bond lengths were summarized in Table. 3-3. Among these aspartic acid residues, Asp216 was

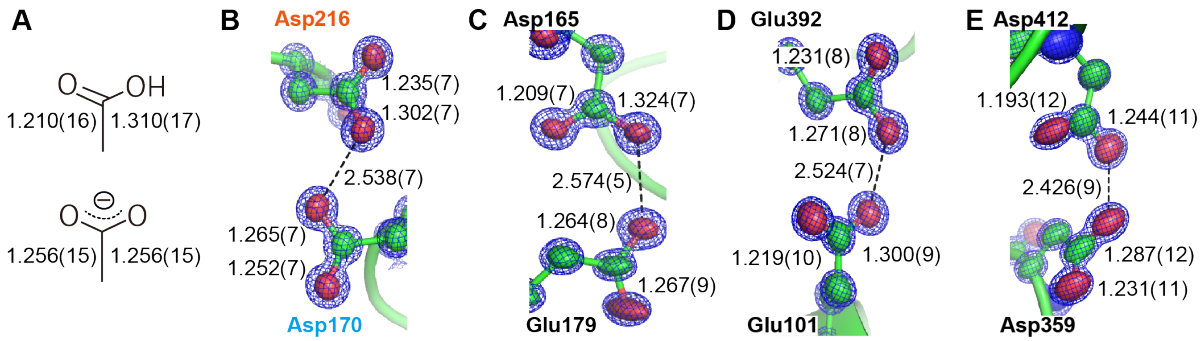

Figure 3-4. Hydrogen bonding distance and bond lengths of asp and glu residues. (A) Dictionary values. (B–E) Asp and Glu pairs in *Pc*Cel6A-apo. Those of *Pc*Cel6A-G3 was similar except for the Asp216-Asp170 pair. $2|F_o|-|F_c|$ map was countered at 2.5 σ level. Estimated standard deviation (ESD) values are shown in parentheses. The anisotropy of atomic displace parameter were shown in ball representations.

reliably identified as protonated in *Pc*Cel6A-apo, with a good match to dictionary values as discussed above. On the other hand, Asp216 was in the deprotonated form with bond lengths of 1.27 Å and 1.26 Å in *Pc*Cel6A-apo (Table. 3-3). The Asp258 was an only candidate for the protonated residue in *Pc*Cel6A-G3 with bond lengths of 1.23 and 1.28 Å. There are variations in dictionary values (Ahmed *et al.*, 2007), and Fisher *et al*. proposed a criterion where significance levels greater than 3 can be inferred for protonated residues (Fisher *et al.*, 2012). Therefore, Asp258 is potentially protonated with this criterion, and it is also reasonable to consider that Asp258 represents a partially protonated residue.

In this protonation state analysis, the catalytic acid Asp216 was distinguished from other residues as this residue was found in two clearly different protonation states. The Asp216 took different conformations between two states, which were shown in Figure 3-5A and B with surrounding hydrogen bonding interactions. In *Pc*Cel6A-G3, Asp216 was in the active conformation pointing toward Glc_{+1}-O4 with a distance of 2.73 Å (Fig. 3-5B). Asp216 was making no direct interaction with the protein in *Pc*Cel6A-G3, but still keeping an interaction with Asp170 by a water-mediated hydrogen bond (Supplementary Fig. S5B). At the same time, another water-mediated interaction was newly created between Asp216 and Asp258 in *Pc*Cel6A-G3. In reported mutagenesis studies, mutations of residues corresponding to Asp170 and Asp258 lowered activities especially at basic pH (Damude *et al.*, 1995; Wolfgang & Wilson, 1999; Koivula *et al.*, 2002). As suggested by these previous reports, Asp170 is contributing to the protonation of Asp216. On the other hand, Asp258 is likely contributing to the deprotonation of Asp216 because Asp258 is protonated rather than Asp216 and interacts with Asp216 only in its deprotonated state.

The Asp170 also shows two different conformations and therefore playing an especially important role in the micro-environmental change around the catalytic acid. In addition, the location of Asp170 is just in the middle of the Glc_{+1} and N-terminal loop, indicating that the

Table 3-3. C-O bond lengths of Asp residues in the catalytic site.

	***Pc*Cel6A-apo**			***Pc*Cel6A-G3**		
	Length (e.s.d.*) (Å)		significance**	Length (e.s.d.*) (Å)		significance**
	OD1	OD2		OD1	OD2	
Asp170	1.265 (0.007)	1.252 (0.007)	1.3	1.277 (0.007)	1.258 (0.007)	2.0
Asp216	1.235 (0.007)	1.302 (0.007)	6.4	1.266 (0.007)	1.255 (0.007)	1.1
Asp258	1.245 (0.006)	1.264 (0.006)	2.3	1.231 (0.007)	1.279 (0.007)	4.8
Asp394	1.243 (0.009)	1.215 (0.009)	2.2	1.271 (0.008)	1.240 (0.009)	2.6
Asp405	Multiple conformations			1.259 (0.009)	1.252 (0.010)	0.5

*e.s.d.: estimated standard diviation

**significance level: $|l_{C\text{-}OD1}-l_{C\text{-}OD2}|/(\sigma_{C\text{-}OD1}^2+\sigma_{C\text{-}OD2}^2)^{1/2}$, according to Fisher *et al.*, 2012.

conformational change of Asp170 is closely related to the ligand binding and conformational change of the N-terminal loop. As shown in Fig. 3-5C, the mobile region of N-terminal loop is composed of Ala and Glu residues with both ends fixed by the disulfide bridge, Cys171-Cys230, and the pair of Asp165-Glu179 (Supplementary Fig. S5). The N-terminal loops itself does not interact with ligand except for Ser164, whose OG atom was in a distance of 2.66 Å from Glc_{-2}-O2. In contrast, it is likely that Asp170 responds the ligand binding to induce the conformational change of the N-terminal loop. In *Pc*Cel6A-G3, Asp170 moved apart from Asp216, Tyr164 and Arg169, and started to interact with Glc_{+1}-O3 and the main chain nitrogen atom of Ser176 on the N-terminal loop (Supplementary Fig. S5).

Identification of the potential catalytic base is of great interest, but it was also difficult to find a candidate by this protonation state analysis. The aspartic acid, Asp170, Asp394, and Asp405, locate near from the α-face of the sugar (Fig. 3-3B), but did not show significant changes in protonation states (Table 3-3). The water network proposed in Grotthuss mechanism (Koivula *et al.*, 2002) was also found in *Pc*Cel6A-G3 as shown in Fig. 3-5D. The superposition of the *Hi*Cel6A structure (1OCN, p. 14, Fig. 1-7C, Varrot, Macdonald *et al.*, 2003) revealed that the position of the Na^+ ion was nearly identical to that of the nitrogen atom of cellobio-derived isofagomine inhibitor. Putative catalytic water (W1) and networking water (W2) were similarly coordinated by Ser176, Asp394 and Asp170, though slight positional shifts of Glc_{+1} was found probably owing to the shorter interactions of the Na^+ ion. The hydrogen bonding distance of W1 and W2 was in a usual distance of 2.72 Å. The bond lengths of Asp170 were 1.28 Å in OD2 and 1.26 Å in OD1, which showed slightly shorter bond length for OD2 (Table 3-3). This indicates that Asp170 pointing the less electron rich oxygen (OD2) to networking water (W2), which is less preferable to accept hydrogen. Therefore, it was not probable that Asp170 activates the networking water in its deprotonated state, though there are still possibilities of different microenvironments with the presence of sugar at subsite -1. At least in *Pc*Cel6A-G3, Asp170 was seemed to be stable in its deprotonated form by being acceptors of many hydrogen bonds.

For the purpose of examination of the actual catalytic residues for hydrolysis reaction and of the feature for the crystalline cellulose degradation in in GH family 6 CBHs, the known structures of GH family 6 has been investigated in Supplementary Table S2. According to the phylogenetic analysis (Fig. 1-6, p. 12), the structures were classified as fungal CBHs (group A), bacterial CBHs (group B), bacterial EGs (group D), and *Cc*Cel6C, *Hi*Cel6B, and Cel6 from *Orpinomyces* sp. are grouped together as group C. The cellulases in group C lack at least one property that is important for CBH activity, such as CBM, C-terminal loop and subsite +4 tryptophan. In these reported structures, the N-terminal loop, catalytic acid residue and base/pK_a residue (corresponding to Asp170 of *Pc*Cel6A) are observed in different

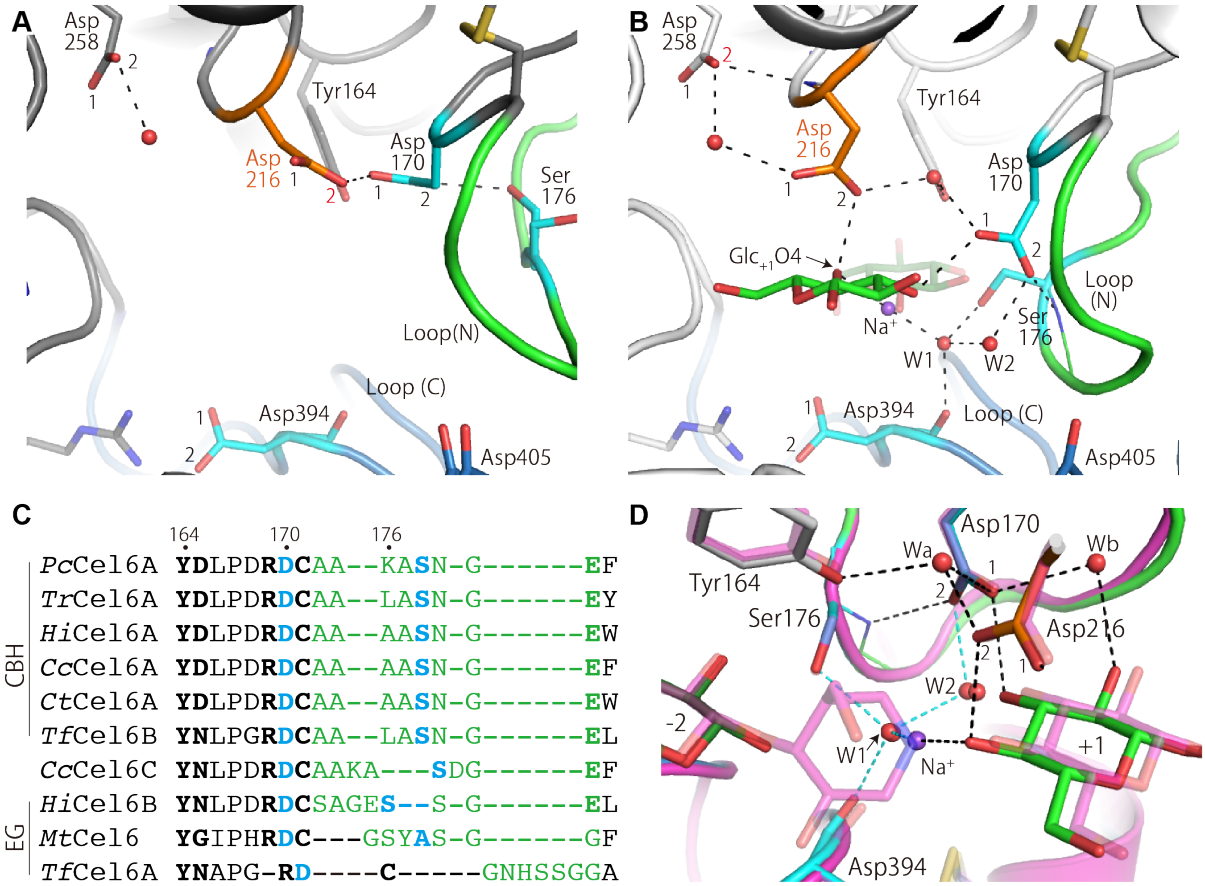

Figure 3-5. Hydrogen bonding patterns around Asp216 and Asp170 in *Pc*Cel6A-apo (A) and *Pc*Cel6A-G3 (B). The small numbers near the aspartic acid residues indicate the OD1 or OD2 in Table. 3-3. **(C) Sequences of the N-terminal loop.** The alignment was generated by UCSF *Chimera* (Pettersen *et al.*, 2004) by superposition of structures (5XCZ, 1QJW, 1OCB, 3VOH, 4A05, 3ABX, 4B4F 1DYS, 1UP3, 2BOD, up to bottom) under the residue-residue distance cutoff 10 Å. **(D) Water network in *Pc*Cel6A-G3.** Isofagomine complex of *Hi*Cel6A (PDB entry 1OCN) were shown in magenta with transparent stick representations with superposition to *Pc*Cel6A-G3.

conformations. In accordance with the postulation that Asp170 is the critical residue for loop enclosure event, the open/closed states of N-terminal loops of CBHs and the conformations of base/pK_a residues had a clear correlation (Supplementary Table S2). On the other hand, the positions of base/pK_a residues are not conserved in *Tf*CelA and Cel6 from *Orpinomyces* sp. as close to the cleavage point. Therefore, the base/pK_a residue is considered to be responsible for CBH activity by linking substrate binding, proton transfer and loop enclosure events. It is the reason why its locations are highly conserved in GH family 6 CBHs and have great contribution to the activity.

The absence of catalytic base and the alternative Grotthuss mechanism are still open questions. It is still a mystery whether the inverting enzymes require a classical catalytic base residue, as the GH family 6 is not an exceptional case of the absence of the catalytic base

(Table 1-1, p.7). In this study, the properties of catalytic acid were discussed as the ability to take both protonated and deprotonated states. To enable this two different protonation states, *Pc*Cel6A employ an elegant mechanism involving the conformational change of catalytic acid. Asp216 locates in between the two aspartic acid residues, Asp170 and Asp258, and the change of hydrogen bonding pattern with them enable Asp216 to accelerate the facile proton transfer to the glycosidic oxygen. One interesting thing is that deprotonated catalytic acid seems to make environment to coordinate positively charged Na^+ ion in the subsite -1 (Fig. 3-3). The positively charged atoms are frequently reported in the crystal structures of GH family 6, such as Li^+ and nitrogen atom of isofagomine (Varrot *et al.*, 2003b; Thompson *et al.*, 2012). Teleman *et al*. also suggested by biochemical studies that ligand complex was stabilized in presence of 40 mM NaCl (Teleman *et al.*, 1995). They considered shielding effects of side chains as sugars have no charges in themselves. In the proposal of Koivula *et al*., they suggested that a uniquely electron deficient transition state was stabilized by base/pK_a residue, Asp175 in *Tr*Cel6A (Koivula *et al.*, 2002). It is likely that stabilization of the positively charged species in the active site is one of the most important features for GH family 6 enzymes, as Warshel emphasizes that the electrostatic effect is the origin of catalytic power of enzymes (Warshel *et al.*, 2006). The mechanism of resetting the protonation states of residues for another catalysis must be also elucidated in the case of inverting GH family 6. Grotthuss mechanism answers it as the proton can be put back to the catalytic acid from the proton accepting residue, Asp170 of *Pc*Cel6A, by the direct hydrogen bond formation. For other possibilities, it is also presumable that pK_a modulating residue, Asp258 of *Pc*Cel6A, gives proton to catalytic acid thorough the water-mediated hydrogen bond. The Asp258 is unique in that this residue accept both protonated and deprotonated states.

4. Development of a random mutagenesis protocol for structure-function studies on *Pc*Cel6A

(*Paper 1*)

Random mutagenesis is a powerful tool for protein engineering and structure-function relationship studies. Error-prone PCR and DNA shuffling techniques have been used successfully for directed evolution of enzymes (Bloom *et al.*, 2005). However, the limited expression hosts remain a key issue, especially for fungal cellulases. In this study, a simple random mutagenesis method for the *P. pastoris* expression system was developed, and was successfully applied to the structure-function relationship study of *Pc*Cel6A.

4.1. Random mutagenesis protocol for *P. pastoris* expression system

The methylotrophic yeast *P. pastoris* is an excellent heterologous expression host, whose productivity reaches at the gram per liter of culture level (Macauley-Patrick *et al.*, 2005). *P. pastoris* is able to secrete heterologous proteins directly into the culture medium, which is especially advantageous for screening of mutant libraries toward insoluble cellulosic substrates. However, *P. pastoris* has rarely been used in random mutagenesis (Pourmir & Johannes, 2012) because transformation of *P. pastoris* requires a large amount (μg) of plasmids (Wu & Letchworth, 2004). Laborious steps are required in usual techniques (Fig. 4-1, left), therefore, a novel two-step method was developed with combinational isothermal reactions of error-prone rolling circle amplification (RCA) and multiple displacement activity (MDA) (Fig. 4-1, right).

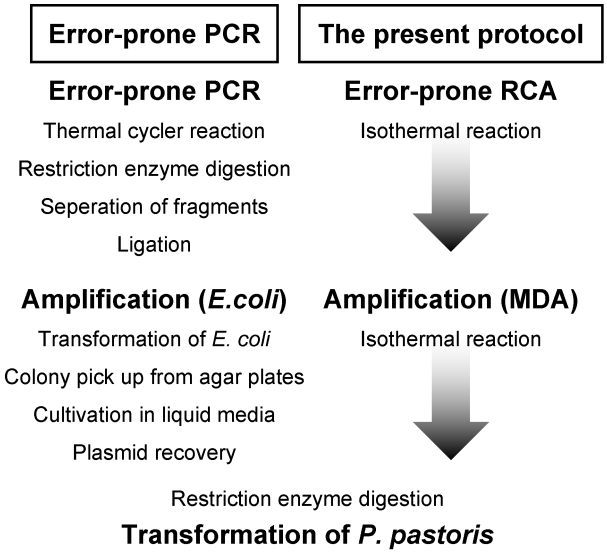

Figure 4-1. Comparison of random mutagenesis methods for the *P. pastoris* expression system.

Error-prone RCA and MDA

As the scheme shown in Fig. 4-2, both of the error-prone RCA and MDA are amplification reactions by Phi29 DNA polymerase. The first step is introduction of random mutations and the second step is amplification of mutated DNA. Phi29 DNA polymerase has a strong strand displacement activity, and amplifies circular DNAs isothermally to yield linear DNAs composed of tandem repeats of the circular DNA plasmids (Dean *et al.*, 2001; Nelson *et al.*, 2002). In addition, Phi29 DNA polymerase can be also utilized for amplification of long linear DNA such as genomic DNA, which is generally known as MDA (Dean *et al.*, 2002).

The replication fidelity of Phi29 DNA polymerase is reduced by the Mn^{2+} ions because of the replacement of Mg^{2+} ions required for activity (Esteban *et al.*, 1993). According to the error-prone RCA protocol that was developed by Fujii *et al.* (Fujii *et al.*, 2004, 2006), various amounts of circular pGAPZα vectors including *cel6A* gene (50–500 pg) were amplified in the presence of 0–2 mM $MnCl_2$ in the first step. The amplified DNA products were more than 10 kbp when directly analyzed by agarose gel electrophoresis, and were digested with restriction enzyme (BlnI) to linear fragments of 4.4 kbp which is identical to the sum of vector and insert. The largest amount of amplified plasmids was estimated as ~1 μg in a total volume of 10 μl, however, the yield of the products decreased with increase of Mn^{2+} concentration and the obtained DNA in the presence of 2 mM Mn^{2+} were less than 10 ng/μl.

In order to obtain larger (μg) amounts, Phi29 DNA polymerase was used again for amplification of error-prone RCA products. With optimization of the reaction mixture, linear error-prone RCA products could be used as a template DNA for Phi29 DNA polymerase amplification. Almost the same level of amplification was obtained in the absence of Mn^{2+}: approximately 5 μg in a total volume of 50 μl for all samples.

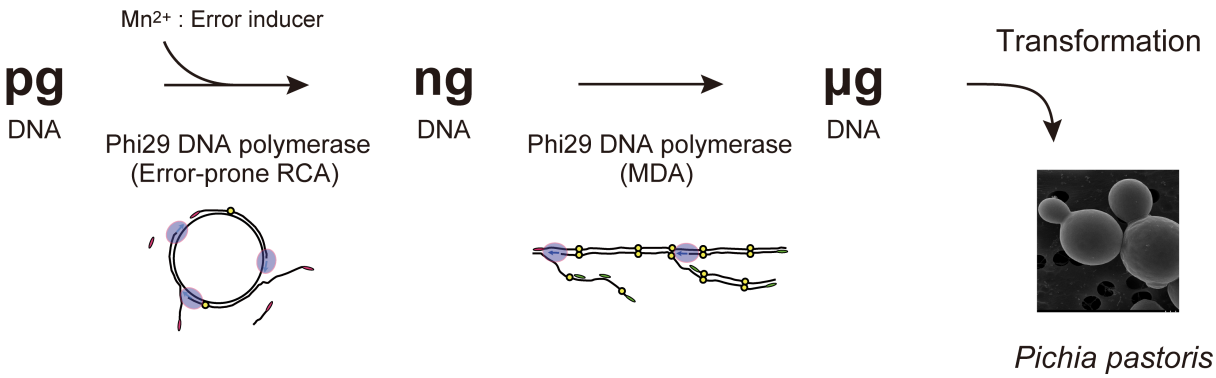

Figure 4-2. Schematic representation of the random mutagenesis method. The circular protein expression vector is amplified repeatedly by strand displacement reaction of Phi29 DNA polymerase. Mutations are introduced in the first step by adding Mn^{2+} to reduce the fidelity of the polymerase (this is known as error-prone RCA). Subsequent amplification with Phi29 DNA polymerase (MDA) provides μg amounts of mutated DNA, sufficient for transformation into *P. pastoris* for enzyme production.

Mutation frequencies

The mutation frequencies were analysed by large-scale sequencing of 12 samples of error-prone RCA-MDA (Table 4-1). The number of errors increased with increasing concentration of Mn^{2+} in accordance with findings by Fujii *et al.* (Fujii *et al.*, 2004). At high template concentration (500 pg), the mutation frequency was high (2.07 kb^{-1}) even in the absence of Mn^{2+} in the reaction mixture. The maximum mutation frequency (2.60 kb^{-1}) was obtained with 2 mM Mn^{2+} and 100 pg template. Mutation frequencies were consistently in the range of 2–3 base substitutions per kb in the presence of 2 mM manganese ions, and this error rate is considered to be favorable for accumulation of adaptive mutations (Arnold *et al.*, 2001).

The types of substitution, inherent bias, and distribution of mutations in relation to GC content of the template sequence are described in detail in *paper 1.* The mutations were also found in the whole vector regions including promoter regions and selection marker gene, which potentially affect the transformation efficiency and protein productivity.

4.2. Properties of *Pc*Cel6A mutants

Enzyme production and activity screening

Approximately 5 µg of the linearized products obtained under two conditions (2 mM manganese and 100, 250 pg template) was used for electroporation of *P. pastoris*. The number of colonies was ~100 per plate, which is slightly fewer than the plates of wild-type colonies obtained by the usual plasmids preparation protocol (100–300 colonies per plate).

In order to handle multiple samples, enzyme production was performed in 96 deep-well plates containing 1 mL liquid media, and gene expression was driven by the constitutive GAP

Table 4-1. Mutation frequency.

Template (pg)	Mn^{2+} (mM)	Total bases* Minimum	Total bases* Maximum	Averaged mutation frequency (kb^{-1})
50	0	452854	697971	0.85 ± 0.56
	1	768351	1250127	1.55 ± 1.40
	2	394019	782970	1.98 ± 1.76
100	0	459038	721539	0.86 ± 0.55
	1	112234	169952	2.15 ± 2.08
	2	414024	782275	2.60 ± 4.63
250	0	141208	223967	1.32 ± 1.38
	1	272842	419288	1.98 ± 1.92
	2	220309	433775	2.03 ± 1.84
500	0	186120	294496	2.07 ± 1.95
	1	213880	463802	1.99 ± 1.76
	2	184115	406056	2.11 ± 1.92

* The total number of bases used for calculation of mutation frequency at each reference base of *cel6A* (1320 bp).

promoter. Ten μl aliquots of yeast culture supernatants were used for cellulose-degrading activity measurements, which were performed at pH 5.0 for 2 hours at 40°C in duplicate by using 96 well plates. Each well contained β-glucosidase (Novozymes 188) to convert hydrolysis products to glucose, and the amount of released sugars was quantified by a glucose oxidase method with Wako Glucose CII Test kit.

The activity of the crude enzymes to phosphoric acid-swollen cellulose (PASC), that were prepared from Avicel cellulose powder, were shown in Fig. 4-3. The 96 colonies obtained by the usual plasmid preparation method was used as a control, and their dispersed activities indicate that the differing amounts of enzymes in the culture resulting from the different productivity of the enzymes (Fig. 4-3C). It is probably because multiple gene integration events occur with detectable frequency and greatly enhance expression levels of target proteins (Clare *et al.*, 1991; Cregg, 2007; Zheng *et al.*, 2014). The numbers of transformants whose activities were less than 10% of that of wild-type *Pc*Cel6A (the median activity of 96 control transformants was 0.37 mM glucose in 2-hour incubation) were 40% and 37% under the error-prone RCA-MDA conditions with 100 pg and 250 pg template, respectively (Fig. 4-3A, B). In contrast, the corresponding number for wild-type *Pc*Cel6A was 2% (2 of 96 colonies, Fig. 4-3C). It is highly probable that the high levels of transformants with markedly lowered activity from the error-prone RCA-MDA plates are mainly due to the introduction of mutations.

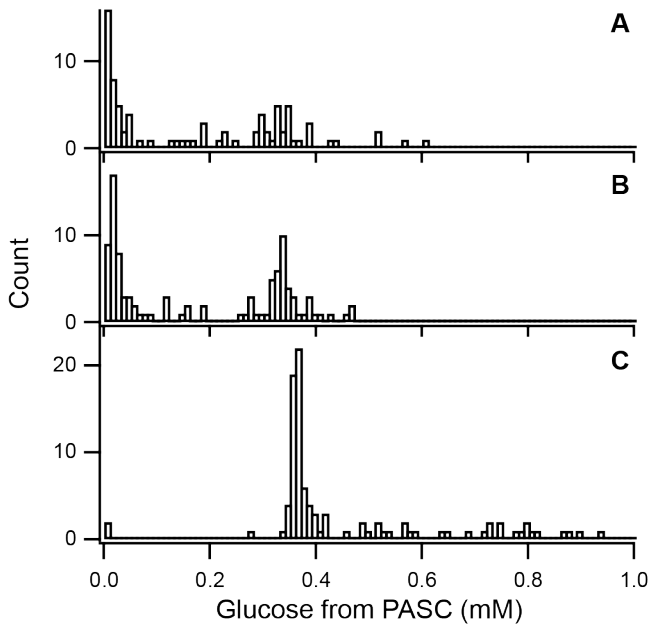

Figure 4-3. Histogram of amorphous cellulose (PASC)-degrading activities of transformants. (A) The activities of 87 transformants obtained by error-prone RCA-MDA under the conditions of 2 mM Mn^{2+} and 100 pg template. (B) The activities of 100 transformants obtained by error-prone RCA-MDA under the conditions of 2 mM $Mn^{2+}$and 250 pg template. (C) The activities of 96 transformants obtained by the usual plasmids preparation protocol (wild-type control).

Degradation of amorphous- and crystalline-cellulose

In order to find functionally informative mutants by direct screening on crystalline cellulose, hydrolysis experiments were performed with crystalline cellulose III_I that was prepared from green algae *Cladophora* sp. (Igarashi *et al.*, 2007). Fig. 4-4A shows the amorphous cellulose (PASC)-degrading versus crystalline cellulose (cellulose III_I)-degrading activities of the 87 transformants that were obtained by error-prone RCA-MDA. The data points in blue shows the wilde-type Cel6A, whose DNA sequencing revealed no mutation or no change in amino acid sequence. Mutants that had at least one mutation in the *cel6A* gene are shown in red. The transformants contained relatively large portions of wild-type enzymes, probably because these samples were randomly selected from the transformants whose activity was more than 10% of the average of wild-type controls. The protein concentration of culture supernatants varied from 0.01 to 0.2 mg/ml, and the results of SDS-PAGE analysis of some mutants were shown in Fig. 4-4B.

Some mutants revealed clearly different characters even in this small library. The most prominent examples were mutants #13 and #15, which had lowered degrading activity toward crystalline cellulose III_I while retaining activity toward amorphous cellulose (Fig. 4-4A, numbered in black). The DNA sequencing of #15 revealed a single mutation, W267C. On the other hand, the DNA sequencing of #13 revealed several mutations: C25Y, A105D, G346D, as well as two mutations that would not cause any amino acid substitution. The DNA sequencing traces of #13 also showed overlaps with the wild-type sequence, probably due to multiple-copy gene integration.

A search for advantageous mutations is generally difficult than that for deleterious mutations. At least, other mutants showed amorphous/crystalline cellulose-degrading activity ratio similar to that of the wild type. On the other hand, SDS-PAGE analysis revealed a candidate of mutant with increased specific activity, mutant #64. The mutant #64 had substantial activity to both amorphous and crystalline cellulose (Fig. 4-4A, numbered in black), but the apparent protein amount was not large (Fig. 4-4B). The DNA sequencing of #64 revealed a single mutation of G421A.

Structural aspects

The location of mutations found in mutant #13, #15, #64 was shown in Figure 4-5. The mutation of W267C was found in mutant #15 to be critical for degradation of cellulose III_I. This tryptophan composes of subsite +4, at the entrance of the active site tunnel of *Pc*Cel6A (Fig. 4-5A). There are a preceding research on corresponding residue of *Tr*Cel6A (W272), where the residue was not necessary for hydrolysis of oligosaccharides and amorphous cellulose but requisite for crystalline cellulose degradation (Koivula *et al.*, 1998). Therefore, the tryptophan

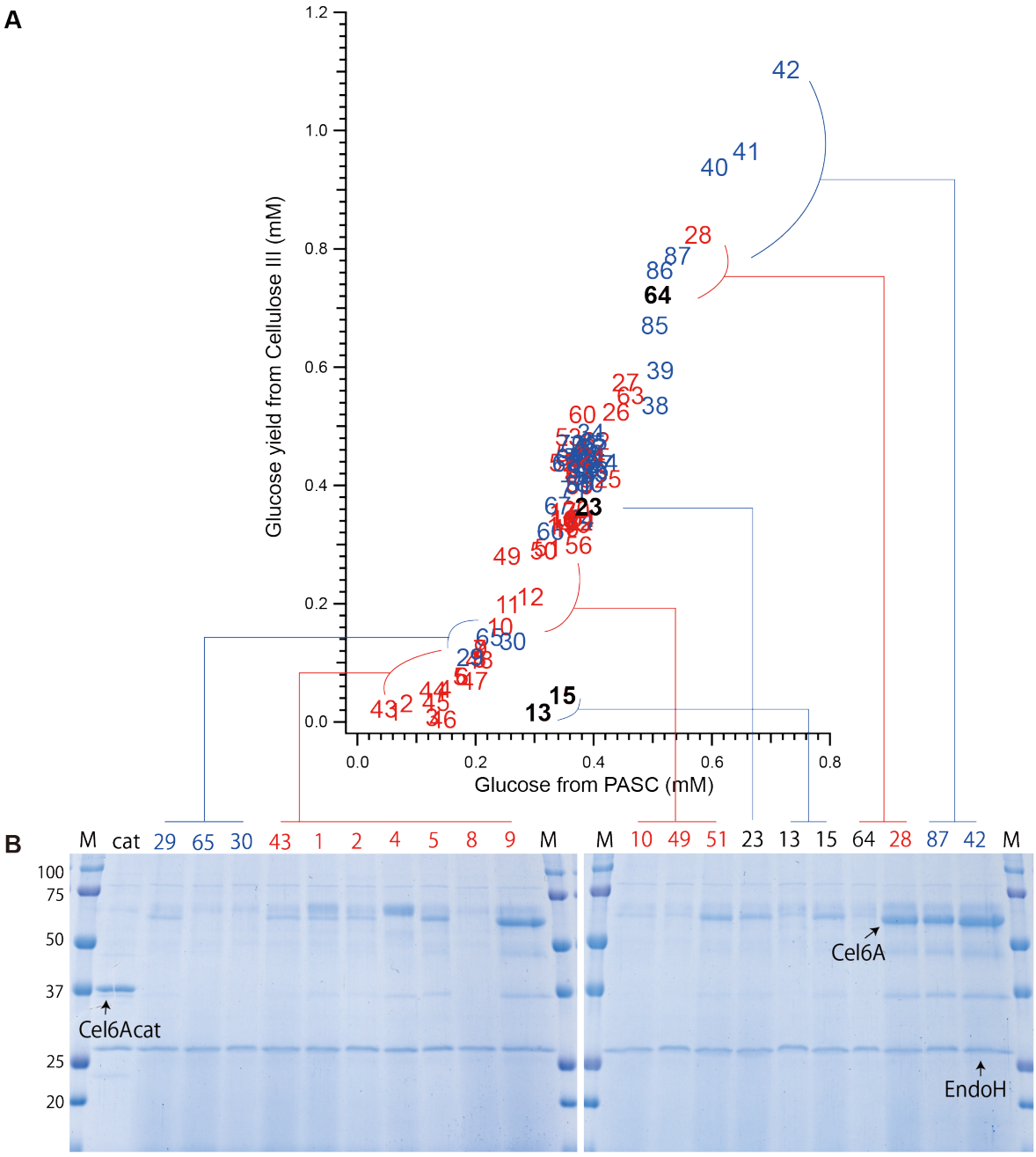

Figure 4-4. (A) Plot of amorphous cellulose (PASC)-degrading activity versus crystalline cellulose (cellulose III_I)-degrading activity of *Pc*Cel6A mutants. Forty-two transformants were selected from error-prone RCA-MDA plates under the conditions of 2 mM Mn^{2+} and 100 pg template (numbered 1–42), and 45 transformants from error-prone RCA-MDA plates under the conditions of 2 mM Mn^{2+} and 250 pg template (numbered 43–87), and their activities were measured. The transformants with no mutation or no change in amino acid sequence are indicated in blue. Mutants with at least one mutation in the *cel6A* gene are shown in red (or black). **(B) SDS-PAGE of 25 µl culture supernatants after the digestion with endoglycosidase H (EndoH) and α-mannosidase.**

at subsite +4 probably has a specialized role for recognition and recruiting single cellulose chain from crystalline surface. The interpretation of mutant #13 was uncertain because of multiple mutations, but the most influential mutation might be C25Y. The C25 is expected to form a disulfide bridge with C8 in the CBM of *Pc*Cel6A (Fig. 4-5B), thus this mutation likely result in reduced affinity on cellulose crystalline surface. The importance of CBM in crystalline cellulose degradation was already reported in several studies (Tomme *et al.*, 1988; Ståhlberg *et al.*, 1993). The possible candidate for the advantageous mutation for specific activity, G412D found in mutant #64, was located at the product binding subsite -2 (Fig. 4-5C). The Cα atom of G412 is only 3.5 Å apart from Glc_{-2}-O3, and there are two carboxyl groups involving in recognition of hydroxyl groups of the Glc_{-2} (E392, D131), and a flexible arginine on the C-terminal loop (R403). Therefore, it would be possible that the additional asp residue had an effect on specific activity.

Screening of mutants was performed only for activity in this study, but the thermostability is also of great interest for protein engineering of *Pc*Cel6A. At least two research groups performed this kind of work previously by combinatorial approaches of sequence analysis and screening, and it is worthy pointing out the mutations and their location in the structure.

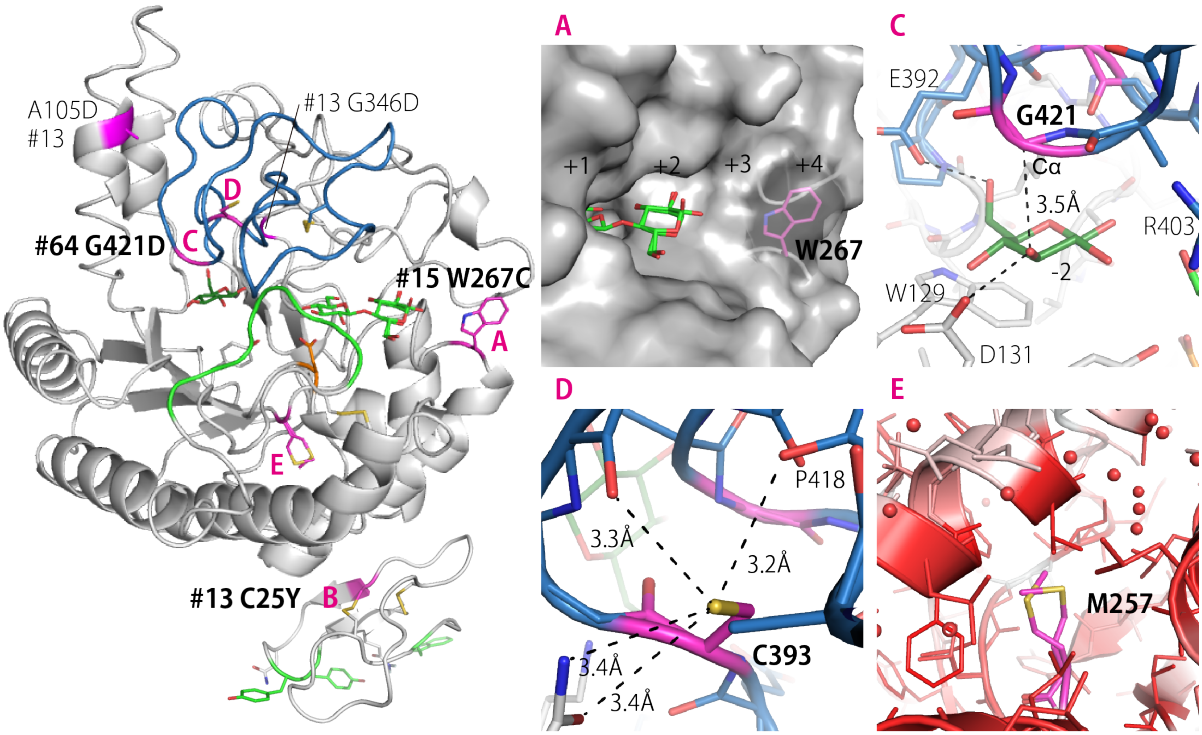

Figure 4-5. Residues that was focused by mutagenesis studies on *Pc*Cel6A. The locations of altered amino acids in this study were indicated with the mutant numbers (#).The structure of CBM was predicted by using *Phyre2* (Kelley & Sternberg, 2009). (A) W267 (#15) (B) C25 (#13) (C) G421 (#64) (D) C393 discussed in Heinzelman *et al.*, 2009 (E) M257 discussed in Ito *et al.*, 2013.

Heinzelman *et al.* successefully engineered *Pc*Cel6A to improve its thermostability by remarkable 10˚C by a single mutation of C393S (Heinzelman *et al.*, 2009; I. Wu *et al.*, 2013). This free cysteine makes only weak hydrogen bonds (Fig. 4-5D), which is consistent with their discussion. Ito *et al*. also improved the thermostability of *Pc*Cel6A by cumulative 16 mutations (Ito *et al.*, 2013). The largest stabilization of 1.2˚C was achieved by a mutation of M257I, and the authors deduced that exchange to more hydrophobic residues stabilize the structure in its buried protein environment. In accordance with their proposal, the inner sides of α-helixes were composed of hydrophobic side chains and no water was found near M257 (Fig. 4-5E). In addition, M257 was observed in double conformations in the crystal structure and therefore probably making negative contributions to thermostability. Mutant M257I was also found in the mutant library of this study (mutant #11, M257I and A103T).

As discussed above, screening of mutant libraries brings different and valuable insights on properties of non-catalytic residues of *Pc*Cel6A. Especially, further investigations of the G421D mutant (mutant #64) will be interesting. The sequence conservation of G421 was unexpectedly as high as 98% in 62 sequences of 'characterized' GH family 6 enzymes in CAZy database. Therefore, it is probable that this mutation could be obtained only by the fully randomized method. It was surprising that G421D had the amorphous/crystalline cellulose-degrading activity ratio similar to wild-type. In many cases, the higher activity to amorphous cellulose can be achieved by the mutation to allow easier diffusion of cellulose chain, but the higher activity to crystalline cellulose might be difficult to generalize. Purification and determination of the accurate specific activity is necessary for further discussions.

5. Conclusion

GH family 6 CBHs are playing an important role for degradation of crystalline cellulose in many cellulolytic microorganisms. The focus of this thesis was to elucidate the structure-function relationships of the GH family 6 CBH from *P. chrysosporium* (*Pc*Cel6A). Hydrolysis of the stable glycosidic bond itself is a challenging reaction, therefore GHs must be evolved at first to facilitate the chemical reaction by structural arrangements of catalytic residues. In addition, cellulases face a particularly difficult challenge of extracting degradable chains from crystalline surface, and CBHs are only enzymes that are able to do it efficiently. Especially, *Pc*Cel6A is known to be efficient in degradation of crystalline cellulose III_I although its mechanism is not well elucidated.

In chapter 2, the overall structure of the CD of *Pc*Cel6A was determined in both apo and substrate-complexed forms. The CD of *Pc*Cel6A has a tunnel-like structure enclosed by two surface loops, as commonly found in fungal GH family 6 CBHs. Analysis of flexibility and mobility of the N- and C-terminal tunnel-enclosing loops revealed differences in their characteristics. The C-terminal loop is an immobile loop regardless of the substrate binding, on the other hand, the N-terminal loop makes a dynamic movement in association with cellobiose binding at subsite +1 and +2.

For further investigation of the catalytic mechanism of *Pc*Cel6A, protonation states of the catalytic aspartic acid residues were examined in chapter 3. Optimizations of the crystallization conditions and trials of crystallization experiment under microgravity enabled to obtain high-quality crystals that diffract to atomic resolutions. Bond distance analysis of aspartic acid residues gave direct evidence that the Asp216 works as the catalytic acid to protonate glycosidic oxygen. The conformational changes and reconstructions of hydrogen bonding network with Asp170 and Asp258, which have opposite natures in terms of protonation states, enable the catalytic acid to show two distinctly different protonation states. The Asp170 was prominent in anchoring many hydrogen bonds including the catalytic acid, substrate, and N-terminal loop, indicating that the Asp170 plays unquestionably important roles in making the micro-environment of the catalytic acid and inducing the associated loop enclosure events.

In chapter 4, a new random mutagenesis protocol was developed to obtain *Pc*Cel6A mutants in a quick and efficient way. In order to utilize *P. pastoris* as the protein expression host, the unique amplification system of phi29 DNA polymerase was repeatedly used. The mutation frequencies of the obtained DNA products were 2–3 mutation/kb, which was suitable for protein engineering studies. Two mutants with altered activities toward crystalline cellulose III_I were found in the relatively small library, supporting the idea that the developed method

was useful for structure-function relationship studies. In addition, the mutation of Gly421Asp at subsite -2 was found to potentially improve the specific activity of the *Pc*Cel6A.

On the basis of the results from this study, the protonation of glycosidic bond by catalytic acid seems to drive the hydrolysis reaction of *Pc*Cel6A. The mechanism of promoting nucleophilic attack of the catalytic water will be a subject for future investigation, but it is speculated that the presence of strong acid can compensate for the lack of catalytic base. The ligand binding, proton transfer, and movement of the N-terminal loop is linked by the highly-conserved residue of GH family 6 CBHs, Asp170 of *Pc*Cel6A. The GH family 6 CBHs are potentially evolved to be efficient toward crystalline substrates by the finely tuned active site around the catalytic acid and tunnel loop regions. With the interactions of the tunnel enclosing loops and crystalline cellulose surface, it is hypothesized that the movement of loops work to loosen the crystalline packing and help to make proximal region to be more accessible. In the mutagenesis experiment, several mutations of non-catalytic residues were found to be affect the specific activity of *Pc*Cel6A and introducing a mutation in the product binding subsite was suggested to be advantageous for higher catalytic power. This indicates that the rate-limiting step of the function of *Pc*Cel6A is not related to the simple hydrolysis of the glycosidic bond but to the other steps where interactions between enzyme and cellulose are involved. Although there are still large gaps between the structure-based understanding and design of CBHs for improved function, the structural information and non-theoretical random mutagenesis approaches in this study provided a future path for engineering of GH family 6 CBHs.

References

Abuja, P.M., Pilz, I., Claeyssens, M., Tomme, P., 1988a. Domain structure of cellobiohydrolase II as studied by small angle X-ray scattering: Close resemblance to cellobiohydrolase I. Biochem. Biophys. Res. Commun. 156, 180–185.

Abuja, P.M., Schmuck, M., Pilz, I., Tomme, P., Claeyssens, M., Esterbauer, H., 1988b. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*. Eur. Biophys. J. 15, 339–342.

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. Biol. Crystallogr. 66, 213–221.

Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., Adams, P.D., 2010. Joint X-ray and neutron refinement with phenix.refine. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 1153–1163.

Ahmed, H.U., Blakeley, M.P., Cianci, M., Cruickshank, D.W.J., Hubbard, J.A., Helliwell, J.R., 2007. The determination of protonation states in proteins. Acta Crystallogr. Sect. D Biol. Crystallogr. 63, 906–922.

Amano, Y., Shiroishi, M., 1996. substrate specificities of four exo-type cellulases produced by *Aspergillus niger*, *Trichoderma reesei*, and *Irpex lacteus* on (1→ 3),(1→ 4)-β-D-glucans and xyloglucan. J. Biochem. 120, 1123–1129.

André, G., Kanchanawong, P., Palma, R., Cho, H., Deng, X., Irwin, D., Himmel, M.E., Wilson, D.B., Brady, J.W., 2003. Computational and experimental studies of the catalytic mechanism of *Thermobifida fusca* cellulase Cel6A (E2). Protein Eng. 16, 125–134.

Armand, S., Drouillard, S., Schulein, M., Henrissat, B., Driguez, H., 1997. A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. J. Biol. Chem. 272, 2709–2713.

Arnold, F.H., Wintrode, P.L., Miyazaki, K., Gershenson, A., 2001. How enzymes adapt: lessons from directed evolution. Trends Biochem. Sci. 26, 100–106.

Becker, D., Johnson, K.S., Koivula, A., Schülein, M., Sinnott, M.L., 2000. Hydrolyses of alpha- and beta-cellobiosyl fluorides by Cel6A (cellobiohydrolase II) of *Trichoderma reesei* and *Humicola insolens*. Biochem. J. 345 Pt 2, 315–9.

Beckham, G.T., Matthews, J.F., Peters, B., Bomble, Y.J., Himmel, M.E., Crowley, M.F., 2011. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. J. Phys. Chem. B 115, 4118–27.

Berghem, L., Pettersson, L., 1973. The mechanism of enzymatic cellulose degradation. Eur. J. Biochem. 37, 21–30.

Bidlack, J., Malone, M., Benson, R., 1992. Molecular structure and component integration of secondary cell walls in plants. Proc. Okla. Acad. Sci 56, 51–56.

Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D., Arnold, F.H., 2005. Evolving strategies for enzyme engineering. Curr. Opin. Struct. Biol. 15, 447–452.

Boisset, C., Fraschini, C., Schülein, M., Henrissat, B., Chanzy, H., 2000. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. Appl. Environ. Microbiol. 66, 1444–52.

Breyer, W.A., Matthews, B.W., 2001. A structural basis for processivity. Protein Sci. 10, 1699–1711.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, D233-8.

Carpita, N.C., Gibeaut, D.M., 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1–30.

Carrard, G., Linder, M., 1999. Widely different off rates of two closely related cellulose-binding domains from *Trichoderma reesei*. Eur. J. Biochem. 262, 637–643.

Chanzy, H., Henrissat, B., 1985. Undirectional degradation of *valonia* cellulose microcrystals subjected to cellulase action. FEBS Lett. 184, 285–288.

Christopherson, M.R., Suen, G., Bramhacharya, S., Jewell, K.A., Aylward, F.O., Mead, D., Brumm, P.J., 2013. The genome sequences of *Cellulomonas fimi* and "*Cellvibrio gilvus*" reveal the cellulolytic strategies of two facultative anaerobes, transfer of "*Cellvibrio gilvus*" to the genus *Cellulomonas*, and proposal of Cellulomonas gilvus s. PLoS One 8, e53954.

Chundawat, S.P.S., Bellesia, G., Uppugundla, N., da Costa Sousa, L., Gao, D., Cheh, A.M., Agarwal, U.P., Bianchetti, C.M., Phillips, G.N., Langan, P., Balan, V., Gnanakaran, S., Dale, B.E., 2011. Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. J. Am. Chem. Soc. 133, 11163–74.

Claeyssens, M., Tomme, P., Brewer, C., Hehre, E., 1990. Stereochemical course of hydrolysis and hydration reactions catalysed by cellobiohydrolases I and II from *Trichoderma reesei*. FEBS Lett. 263, 89–92.

Claeyssens, M., Van Tilbeurgh, H., Tomme, P., Wood, T.M., McRae, S.I., 1989. Fungal cellulase systems. Comparison of the specificities of the cellobiohydrolases isolated from *Penicillium pinophilum* and *Trichoderma reesei*. Biochem. J. 261, 819–25.

Clare, J., Rayment, F., Ballantine, S., Sreekrishna, K., Romanos, M., 1991. High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Biotechnology 9, 455–460.

Cockburn, D.W., Vandenende, C., Clarke, A.J., 2010. Modulating the pH-activity profile of cellulase by substitution: replacing the general base catalyst aspartate with cysteinesulfinate in cellulase A from *Cellulomonas fimi*. Biochemistry 49, 2042–50.

Coughlan, M., Wood, T., Montenecourt, B., 1985. Cellulases : Production , Properties and Applications. Biochem. Soc. Trans 13, 405–406.

Cragg, S.M., Beckham, G.T., Bruce, N.C., Bugg, T.D.H., Distel, D.L., Dupree, P., Etxabe, A.G., Goodell, B.S., Jellison, J., McGeehan, J.E., McQueen-Mason, S.J., Schnorr, K., Walton, P.H., Watts, J.E.M., Zimmer, M., 2015. Lignocellulose degradation mechanisms across the Tree of Life. Curr. Opin. Chem. Biol. 29, 108–119.

Cregg, J.M., 2007. DNA-Mediated Transformation, in: Cregg, J.M. (Ed.), *Pichia* Protocols. Humana Press, pp. 27–42.

Damude, H.G., Ferro, V., Withers, S.G., Warren, R. a, 1996. Substrate specificity of endoglucanase A from *Cellulomonas fimi*: fundamental differences between endoglucanases and exoglucanases from family 6. Biochem. J. 315, 467–72.

Damude, H.G., Withers, S.G., Kilburn, D.G., Miller, R.C., Warren, R. a, 1995. Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. Biochemistry 34, 2220–4.

Davies, G.J., Brzozowski, a M., Dauter, M., Varrot, A., Schülein, M., 2000. Structure and function of *Humicola insolens* family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 A resolution. Biochem. J. 348 Pt 1, 201–7.

Davies, G.J., Ducros, V.M., Varrot, A., Zechel, D.L., 2003. Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. Biochem Soc Trans 31, 523–527.

Davies, G.J., Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–9.

Davies, G.J., Planas, A., Rovira, C., 2012. Conformational analyses of the reaction coordinate of glycosidases. Acc. Chem. Res. 45, 308–316.

Davies, G.J., Sinnott, M.L., 2008. Sorting the diverse: the sequence-based classifications of carbohydrate-active enzymes. Biochem. J. 30, 26–32.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, a F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., Lasken, R.S., 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. U. S. A. 99, 5261–5266.

Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification 1095–1099.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T., Jones, T.A., 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science 265, 524–8.

Doi, R.H., Kosugi, A., 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat. Rev. Microbiol. 2, 541–551.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of *Coot*. Acta Crystallogr. D. Biol. Crystallogr. 66, 486–501.

Eriksson, K.E., Pettersson, B., 1975. Extracellular enzyme system utilized by the fungus *Sporotrichum pulverulentum* (*Chrysosporium lignorum*) for the breakdown of cellulose. 3. Purification and physico-chemical characterization of an exo-1,4-beta-glucanase. Eur. J. Biochem. 51, 213–8.

Esteban, J. a., Salas, M., Blanco, L., 1993. Fidelity of phi29 DNA Polymerase. J. Biol. Chem. 268, 2719–2726.

Fägerstam, L.G., Pettersson, L.G., 1980. The 1.4-β-glucan cellobiohydrolases of *Trichoderma reesei* QM 9414. FEBS Lett. 119, 97–100.

Fägerstam, L.G., Pettersson, L.G., 1979. The cellulolytic complex of *Trichoderma reesei* QM 9414. An immunochemical approach. FEBS Lett. 98, 363–367.

Fisher, S.J., Blakeley, M.P., Cianci, M., McSweeney, S., Helliwell, J.R., 2012. Protonation-state determination in proteins using high-resolution X-ray crystallography: Effects of resolution and completeness. Acta Crystallogr. Sect. D Biol. Crystallogr. 68, 800–809.

Fox, J.M., Jess, P., Jambusaria, R.B., Moo, G.M., Liphardt, J., Clark, D.S., Blanch, H.W., 2013. A single-molecule analysis reveals morphological targets for cellulase synergy. Nat. Chem. Biol. 9, 356–61.

Fujii, R., Kitaoka, M., Hayashi, K., 2006. Error-prone rolling circle amplification: the simplest random mutagenesis protocol. Nat. Protoc. 1, 2493–2497.

Fujii, R., Kitaoka, M., Hayashi, K., 2004. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res. 32, e145.

Fushinobu, S., 2014. Metalloproteins: A new face for biomass breakdown. Nat. Chem. Biol. 10, 88–9.

Fushinobu, S., Alves, V.D., Coutinho, P.M., 2013. Multiple rewards from a treasure trove of novel glycoside hydrolase and polysaccharide lyase structures: new folds, mechanistic details, and evolutionary relationships. Curr. Opin. Struct. Biol. 23, 652–9.

Gao, D., Chundawat, S.P.S., Sethi, A., Balan, V., Gnanakaran, S., Dale, B.E., 2013. Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. Proc. Natl. Acad. Sci. 110, 10922–10927.

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

Harding, M.M., 2002. Metal-ligand geometry relevant to proteins and in proteins: Sodium and potassium. Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 872–874.

Harjunpää, V., Teleman, A., Koivula, A., Ruohonen, L., Teeri, T.T., Teleman, O., Drakenberg, T., 1996. Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*. Association and rate constants derived from an analysis of progress curves. Eur. J. Biochem. 240, 584–91.

Heinzelman, P., Snow, C.D., Smith, M.A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., Arnold, F.H., 2009. SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. J. Biol. Chem. 284, 26229–26233.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. U. S. A. 92, 7090–7094.

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., Mornon, J.P., 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene 81, 83–95.

Himmel, M.E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R., Bayer, E.A., 2010. Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels 1, 323–341.

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, 545–549.

Holton, J.M., 2009. A beginner's guide to radiation damage. J. Synchrotron Radiat. 16, 133–142.

Hon, D.N.-S., 1994. Cellulose: a random walk along its historical path. Cellulose 1, 1–25.

Horn, S.J., Sikorski, P., Cederkvist, J.B., Vaaje-Kolstad, G., Sørlie, M., Synstad, B., Vriend, G., Vårum, K.M., Eijsink, V.G.H., 2006. Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. Proc. Natl. Acad. Sci. U. S. A. 103, 18089–94.

Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttilä, M., Samejima, M., 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. J. Biol. Chem. 284, 36186–36190.

Igarashi, K., Maruyama, M., Nakamura, A., Ishida, T., Wada, M., Samejima, M., 2012. Degradation of crystalline celluloses by *Phanerochaete chrysosporium* cellobiohydrolase II (Cel6A) heterologously expressed in methylotrophic yeast *Pichia pastoris*. J. Appl. Glycosci. 59, 105–110.

Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T., Samejima, M., 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. Science 333, 1279–82.

Igarashi, K., Wada, M., Samejima, M., 2007. Activation of crystalline cellulose to cellulose III(I) results in efficient hydrolysis by cellobiohydrolase. FEBS J. 274, 1785–92.

Imai, T., Boisset, C., Samejima, M., Igarashi, K., Sugiyama, J., 1998. Unidirectional processive action of cellobiohydrolase Cel7A on *Valonia* cellulose microcrystals. FEBS Lett. 432, 113–116.

Ito, Y., Ikeuchi, A., Imamura, C., 2013. Advanced evolutionary molecular engineering to produce thermostable cellulase by using a small but efficient library. Protein Eng. Des. Sel. 26, 73–79.

Jalak, J., Kurašin, M., Teugjas, H., Väljamäe, P., 2012. Endo-exo synergism in cellulose hydrolysis revisited. J. Biol. Chem. 287, 28802–15.

Johnson, E.A., Sakajoh, M., Halliwell, G., 1982. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Appl. Environ. Microbiol. 43, 1125–1132.

Jongkees, S., Withers, S., 2013. Unusual Enzymatic Glycoside Cleavage Mechanisms. Acc. Chem. Res. 47, 226–235.

Kabsch, W., 2010. XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 1–7.

Kelley, L.A., Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363–371.

Kim, D.Y., Ham, S.-J., Kim, H.J., Kim, J., Lee, M.-H., Cho, H.-Y., Shin, D.-H., Rhee, Y.H., Son, K.-H., Park, H.-Y., 2012. Novel modular endo-β-1,4-xylanase with transglycosylation activity from *Cellulosimicrobium* sp. strain HY-13 that is homologous to inverting GH family 6 enzymes. Bioresour. Technol. 107, 25–32.

Knowles, J.K.C., Lentovaara, P., Murray, M., Sinnott, M.L., 1988. Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. J. Chem. Soc. Chem. Commun. 1401.

Koivula, A., Kinnari, T., Harjunpää, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T.A., Teeri, T.T., 1998. Tryptophan 272 : an essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. FEBS Lett. 429, 341–346.

Koivula, A., Reinikainen, T., Ruohonen, L., Valkeajärvi, A., Claeyssens, M., Teleman, O., Kleywegt, G.J., Szardenings, M., Rouvinen, J., Jones, T.A., Teeri, T.T., 1996. The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. Protein Eng. 9, 691–9.

Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T.T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M.L., Zou, J.-Y., Kleywegt, G.J., Szardenings, M., Ståhlberg, J., Jones, T.A., 2002. The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. J. Am. Chem. Soc. 124, 10015–24.

Konstantinidis, a K., Marsden, I., Sinnott, M.L., 1993. Hydrolyses of alpha- and beta-cellobiosyl fluorides by cellobiohydrolases of *Trichoderma reesei*. Biochem. J. 291 (Pt 3, 883–8.

Koshland, D., 1953. Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. 28, 416–436.

Kraulis, J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J., Gronenborn, A.M., 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28, 7241–57.

Kurasin, M., Väljamäe, P., 2011. Processivity of cellobiohydrolases is limited by the substrate. J. Biol. Chem. 286, 169–77.

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

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

Lantz, S.E., Goedegebuur, F., Hommes, R., Kaper, T., Kelemen, B.R., Mitchinson, C., Wallace, L., Ståhlberg, J., Larenas, E. a, 2010. *Hypocrea jecorina* CEL6A protein engineering. Biotechnol. Biofuels 3, 20.

Larsson, A.M., Bergfors, T., Dultz, E., Irwin, D.C., Roos, A., Driguez, H., Wilson, D.B., Jones, T.A., 2005. Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: the role of tyrosine Y73 in substrate ring distortion. Biochemistry 44, 12915–22.

Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., Teeri, T.T., 2003. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc. Natl. Acad. Sci. U. S. A. 100, 484–9.

Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T.T., Pettersson, G., 1995. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. FEBS Lett. 372, 96–98.

Liu, Y., Igarashi, K., Kaneko, S., Tonozuka, T., Samejima, M., Fukuda, K., Yoshida, M., 2009. Characterization of glycoside hydrolase family 6 enzymes from *Coprinopsis cinerea*. Biosci. Biotechnol. Biochem. 73, 1432–1434.

Liu, Y., Yoshida, M., Kurakata, Y., Miyazaki, T., Igarashi, K., Samejima, M., Fukuda, K., Nishikawa, A., Tonozuka, T., 2010. Crystal structure of a glycoside hydrolase family 6 enzyme, CcCel6C, a cellulase constitutively produced by *Coprinopsis cinerea*. FEBS J. 277, 1532–42.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, 490–495.

Lorber, B., 2002. The crystallization of biological macromolecules under microgravity: A way to more accurate three-dimensional structures? Biochim. Biophys. Acta - Proteins Proteomics 1599, 1–8.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S., 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. 66, 506–577.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., Harvey, L.M., 2005. Heterologous protein production using the *Pichia pastoris* expression system. Yeast 22, 249–270.

Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D.S., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., Coutinho, P.M., Henrissat, B., Berka, R., Cullen, D., Rokhsar, D., 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. Nat. Biotechnol. 22, 695–700.

Mayes, H.B., Broadbelt, L.J., Beckham, G.T., 2014. How sugars pucker: Electronic structure calculations map the kinetic landscape of fivebiologically paramount monosaccharides and their implications for enzymatic catalysis. J. Am. Chem. Soc. 136, 1008–1022.

Mayes, H.B., Knott, B.C., Crowley, M.F., Broadbelt, L.J., Ståhlberg, J., Beckham, G.T., 2016. Who's on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolases. Chem. Sci. 7, 5955–5968.

Mayes, H.B., Tian, J., Nolte, M.W., Shanks, B.H., Beckham, G.T., Gnanakaran, S., Broadbelt, L.J., 2014. Sodium ion interactions with aqueous glucose: Insights from quantum mechanics, molecular dynamics, and experiment. J. Phys. Chem. B 118, 1990–2000.

McCarter, J.D., Withers, S.G., 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struct. Biol. 4, 885–92.

Meinke, A., Damude, H.G., Tomme, P., Kwan, E., Kilburn, D.G., Miller, R.C., Warren, R.A.J., Gilkes, N.R., 1995. Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. J. Biol. Chem. 270, 4383–6.

Mertz, B., Kuczenski, R.S., Larsen, R.T., Hill, A.D., Reilly, P.J., 2005. Phylogenetic analysis of family 6 glycoside hydrolases. Biopolymers 79, 197–206.

Moon, R.J., Martini, A., Nairn, J., Simonsen, J., Youngblood, J., 2011. Cellulose nanomaterials review: structure, properties and nanocomposites. Chem. Soc. Rev. 40, 3941–94.

Moore, A.D., Heldy, A., Terrapon, N., Weiner, J., Bornberg-Bauer, E., 2014. DoMosaics: Software for domain arrangement visualization and domain-centric analysis of proteins. Bioinformatics 30, 282–283.

Nakamura, A., Tasaki, T., Ishiwata, D., Yamamoto, M., Okuni, Y., Visootsat, A., Maximilien, M., Noji, H., Uchiyama, T., Samejima, M., Igarashi, K., Iino, R., 2016. Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *Trichoderma reesei* Cel6A and its domains on crystalline cellulose. J. Biol. Chem. 291, 22404–22413.

Nakamura, A., Tsukada, T., Auer, S., Furuta, T., Wada, M., Koivula, A., Igarashi, K., Samejima, M., 2013. The tryptophan residue at the active site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important for initiation of degradation of crystalline cellulose. J. Biol. Chem. 288, 13503–10.

Nakamura, A., Watanabe, H., Ishida, T., Uchihashi, T., Wada, M., Ando, T., Igarashi, K., Samejima, M., 2014. Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. J. Am. Chem. Soc. 136, 4584–92.

Nakashima, K., Yamada, L., Satou, Y., Azuma, J.-I., Satoh, N., 2004. The evolutionary origin of animal cellulose synthase. Dev. Genes Evol. 214, 81–8.

Nelson, J., Cai, Y., Giesler, T., 2002. TempliPhi, Phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. Biotechniques 32, S44–S47.

Nerinckx, W., Desmet, T., Claeyssens, M., 2003. A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. FEBS Lett. 538, 1–7.

Ng, J.D., Baird, J.K., Coates, L., Garcia-Ruiz, J.M., Hodge, T.A., Huang, S., 2015. Large-volume protein crystal growth for neutron macromolecular crystallography. Acta Crystallogr. Sect. F Struct. Biol. Commun. 71, 358–370.

Nidetzky, B., Claeyssens, M., 1994. Specific quantification of trichoderma reesei cellulases in reconstituted mixtures and its application to cellulase-cellulose binding studies. Biotechnol. Bioeng. 44, 961–6.

Nidetzky, B., Zachariae, W., Gercken, G., Hayn, M., Steiner, W., 1994. Hydrolysis of cellooligosaccharides by *Trichoderma reesei* cellobiohydrolases: Experimental data and kinetic modeling. Enzyme Microb. Technol. 16, 43–52.

Nishiyama, Y., 2009. Structure and properties of the cellulose microfibril. J. Wood Sci. 55, 241–249.

Nishiyama, Y., Langan, P., Chanzy, H., 2002. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J. Am. Chem. Soc. 124, 9074–9082.

Nishiyama, Y., Sugiyama, J., Chanzy, H., Langan, P., 2003. Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron X-ray and neutron fiber diffraction. J. Am. Chem. Soc. 125, 14300–14306.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode, in: Methods in Enzymology. pp. 307–326.

Payne, C.M., Bomble, Y.J., Taylor, C.B., McCabe, C., Himmel, M.E., Crowley, M.F., Beckham, G.T., 2011. Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. J. Biol. Chem. 286, 41028–41035.

Payne, C.M., Jiang, W., Shirts, M.R., Himmel, M.E., Crowley, M.F., Beckham, G.T., 2013a. Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. J. Am. Chem. Soc. 135, 18831–9.

Payne, C.M., Knott, B.C., Mayes, H.B., Hansson, H., Himmel, M.E., Sandgren, M., Ståhlberg, J., Beckham, G.T., 2015. Fungal Cellulases. Chem. Rev. 115, 1308–1448.

Payne, C.M., Resch, M.G., Chen, L., Crowley, M.F., Himmel, M.E., Taylor, L.E., Sandgren, M., Ståhlberg, J., Stals, I., Tan, Z., Beckham, G.T., 2013b. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. Proc. Natl. Acad. Sci. U. S. A. 110, 14646–14651.

Percival Zhang, Y.-H., Himmel, M.E., Mielenz, J.R., 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol. Adv. 24, 452–81.

Pérez, S., Samain, D., 2010. Structure and engineering of celluloses. Adv. Carbohydr. Chem. Biochem. 64, 25–116.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Pourmir, A., Johannes, T., 2012. Directed evolution: selection of the host organism. Comput. Struct. Biotechnol. J. 2, e201209012.

Reese, E.T., Siu, R.G.H., Levinson, H.S., 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. J. Bacteriol. 59, 485–97.

Rosgaard, L., Pedersen, S., Langston, J., Akerhielm, D., Cherry, J.R., Meyer, A.S., 2007. Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated barley straw substrates. Biotechnol. Prog. 23, 1270–1276.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., Jones, T.A., 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science 249, 380–386.

Ruohonen, L., Koivula, A., Reinikainen, T., A, V., TelemanA., Claeyssens, M., Szardenings, M., Jones, T.A., Teeri, T.T., 1993. Active site of *T. reesei* cellobiohydrolase II., in: Suominen, P., Reinikainen, T. (Eds.), Proceedings of the Second TRICEL Symposium on *Trichoderma Reesei* Cellulases and Other Hydrolases. pp. 87–96.

Rupp, B., 2009. Biomolecular crystallography: Principles, practice, and application to structural biology, Garland Science.

Sammond, D.W., Payne, C.M., Brunecky, R., Himmel, M.E., Crowley, M.F., Beckham, G.T., 2012. Cellulase linkers are optimized based on domain type and function: insights from sequence analysis, biophysical measurements, and molecular simulation. PLoS One 7, e48615.

Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B.R., Larenas, E.A., 2013. The structure of a bacterial cellobiohydrolase : The catalytic core of the *Thermobifida fusca* family GH6. J. Mol. Biol. 425, 622–635.

Schülein, M., 1997. Enzymatic properties of cellulases from *Humicola insolens*. J. Biotechnol. 57, 71–81.

Sheldrick, G.M., 2015. Crystal structure refinement with SHELXL. Acta Crystallogr. Sect. C, Struct. Chem. 71, 3–8.

Shibafuji, Y., Nakamura, A., Uchihashi, T., Sugimoto, N., Fukuda, S., Watanabe, H., Samejima, M., Ando, T., Noji, H., Koivula, A., Igarashi, K., Iino, R., 2014. Single-molecule imaging analysis of elementary reaction steps of *Trichoderma reesei* cellobiohydrolase I (Cel7A) hydrolyzing crystalline cellulose Iα and IIII. J. Biol. Chem. 289, 14056–65.

Sinnott, M., 2013. Carbohydrate chemistry and biochemistry: Structure and mechanism: Edition 2, 2nd ed.

Spezio, M., Wilson, D.B., Karplus, P.A., 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry 32, 9906–16.

Ståhlberg, J., Johansson, G., Pettersson, G., 1993. *Trichoderma reesei* has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. Biochim. Biophys. Acta 1157, 107–13.

Suominen, P.L., Mäntylä, A.L., Karhunen, T., Hakola, S., Nevalainen, H., 1993. High frequency one-step gene replacement in *Trichoderma reesei*. II. Effects of deletions of individual cellulase genes. Mol. Gen. Genet. 241, 523–30.

Takahashi, S., Ohta, K., Furubayashi, N., Yan, B., Koga, M., Wada, Y., Yamada, M., Inaka, K., Tanaka, H., Miyoshi, H., Kobayashi, T., Kamigaichi, S., 2013. JAXA protein crystallization in space: ongoing improvements for growing high-quality crystals. J. Synchrotron Radiat. 20, 968–73.

Takashima, S., Ohno, M., Hidaka, M., Nakamura, A., Masaki, H., Uozumi, T., 2007. Correlation between cellulose binding and activity of cellulose-binding domain mutants of *Humicola grisea* cellobiohydrolase 1. FEBS Lett. 581, 5891–5896.

Takeda, K., Kusumoto, K., Hirano, Y., Miki, K., 2010. Detailed assessment of X-ray induced structural perturbation in a crystalline state protein. J. Struct. Biol. 169, 135–144.

Tamura, M., Miyazaki, T., Tanaka, Y., Yoshida, M., Nishikawa, A., Tonozuka, T., 2012. Comparison of the structural changes in two cellobiohydrolases, CcCel6A and CcCel6C, from *Coprinopsis cinerea*--a tweezer-like motion in the structure of CcCel6C. FEBS J. 279, 1871–82.

Tanaka, H., Inaka, K., Sugiyama, S., Takahashi, S., Sano, S., Sato, M., Yoshitomi, S., 2004. A simplified counter diffusion method combined with a 1-D simulation program for optimizing crystallization conditions. J. Synchrotron Radiat. 11, 45–48.

Tanaka, H., Tsurumura, T., Aritake, K., Furubayashi, N., Takahashi, S., Yamanaka, M., Hirota, E., Sano, S., Sato, M., Kobayashi, T., Tanaka, T., Inaka, K., Urade, Y., 2011. Improvement in the quality of hematopoietic prostaglandin D synthase crystals in a microgravity environment. J. Synchrotron Radiat. 18, 88–91.

Taylor, J., Teo, B., Wilson, D., 1995. Conformational modering of substrate binding to endocellulase E2 from *Thermomonospora fusca*. Protein Eng. 8, 1145–1152.

Teeri, T., 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotechnol. 15, 160–167.

Teeri, T., Lehtovaara, P., Kauppinen, S., 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene 1, 43–52.

Teeri, T., Salovuori, I., Knowles, J., 1983. The molecular cloning of the major cellulase gene from *Trichoderma reesei*. Nat. Biotechnol. 1, 696–699.

Teleman, A., Koivula, A., Reinikainen, T., Valkeajärvi, A., Teeri, T.T., Drakenberg, T., Teleman, O., 1995. Progress-curve analysis shows that glucose inhibits the cellotriose hydrolysis catalysed by cellobiohydrolase II from *Trichoderma reesei*. Eur. J. Biochem. 231, 250–8.

Tempelaars, C.A.M., Birch, P.R.J., Sims, P.F.G., Broda, P., 1994. Isolation, characterization, and analysis of the expression of the cbhII gene of *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. 60, 4387–4393.

Thompson, A.J., Heu, T., Shaghasi, T., Benyamino, R., Jones, A., Friis, E.P., Wilson, K.S., Davies, G.J., 2012. Structure of the catalytic core module of the *Chaetomium thermophilum* family GH6 cellobiohydrolase Cel6A. Acta Crystallogr. D. Biol. Crystallogr. 68, 875–82.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., Claeyssens, M., 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. 170, 575–81.

Uzcategui, E., Ruiz, A., Montesino, R., Johansson, G., Pettersson, G., 1991. The 1,4-beta-D-glucan cellobiohydrolases from *Phanerochaete chrysosporium.* I. A system of synergistically acting enzymes homologous to *Trichoderma reesei*. J. Biotechnol. 19, 271–85.

Vaaje-Kolstad, G., Forsberg, Z., Loose, J.S., Bissaro, B., Eijsink, V.G., 2017. Structural diversity of lytic polysaccharide monooxygenases. Curr. Opin. Struct. Biol. 44, 67–76.

Vaaje-Kolstad, G., Horn, S.J., Sørlie, M., Eijsink, V.G.H., 2013. The chitinolytic machinery of *Serratia marcescens* - a model system for enzymatic degradation of recalcitrant polysaccharides. FEBS J. 280, 3028–3049.

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

Van Tilbeurgh, H., Loontiens, F.G., Engelborgs, Y., Claeyssens, M., 1989. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Binding of small ligands to the 1,4-beta-glucan cellobiohydrolase II and influence of glucose on their affinity. Eur. J. Biochem. 184, 553–9.

Van Tilbeurgh, H., Pettersson, G., Bhikabhai, R., De Boeck, H., Claeyssens, M., 1985. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Reaction specificity and thermodynamics of interactions of small substrates and ligands with the 1,4-beta-glucan cellobiohydrolase II. Eur. J. Biochem. 148, 329–34.

Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., Pettersson, G., 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*: Separation of functional domains. FEBS Lett. 204, 223–227.

Varrot, A., Frandsen, T.P., Driguez, H., Davies, G.J., 2002. Structure of the *Humicola insolens cellobiohydrolase* Cel6A D416A mutant in complex with a non-hydrolysable substrate

analogue, methyl cellobiosyl-4-thio-β-cellobioside, at 1.9 Å. Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 2201–2204.

Varrot, A., Frandsen, T.P., Von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schülein, M., Davies, G.J., 2003a. Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens*. Structure 11, 855–864.

Varrot, A., Hastrup, S., Schulein, M., Davies, G., 1999a. Crystal structure of the catalytic domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution. Biochem. J 304, 297–304.

Varrot, A., Leydier, S., Pell, G., Macdonald, J.M., Stick, R. V, Henrissat, B., Gilbert, H.J., Davies, G.J., 2005. *Mycobacterium tuberculosis* strains possess functional cellulases. J. Biol. Chem. 280, 20181–4.

Varrot, A., Macdonald, J., Stick, R. V, Pell, G., Gilbert, H.J., Davies, G.J., 2003b. Distortion of a cellobio-derived isofagomine highlights the potential conformational itinerary of inverting beta-glucosidases. Chem. Commun. (Camb). 946–7.

Varrot, A., Schülein, M., Davies, G., 1999b. Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. Biochemistry 38, 8884–8891.

Vuong, T. V, Wilson, D.B., 2010. Glycoside hydrolases: catalytic base/nucleophile diversity. Biotechnol. Bioeng. 107, 195–205.

Vuong, T. V, Wilson, D.B., 2009. The absence of an identifiable single catalytic base residue in *Thermobifida fusca* exocellulase Cel6B. FEBS J. 276, 3837–45.

Wada, M., Chanzy, H., Nishiyama, Y., Langan, P., 2004. Cellulose III I crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. Macromolecules 37, 8548–8555.

Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H., Olsson, M.H.M., 2006. Electrostatic basis for enzyme catalysis. Chem. Rev. 106, 3210–3235.

Wilson, D.B., 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. Chem. Rec. 4, 72–82.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E. a, Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. D. Biol. Crystallogr. 67, 235–42.

Wohlfahrt, G., 2005. Analysis of pH-dependent elements in proteins: Geometry and properties of pairs of hydrogen-bonded carboxylic acid side-chains. Proteins Struct. Funct. Genet. 58, 396–406.

Wohlfahrt, G., Pellikka, T., Boer, H., Teeri, T.T., Koivula, A., 2003. Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. Biochemistry 42, 10095–103.

Wolfgang, D.E., Wilson, D.B., 1999. Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2. Biochemistry 38, 9746–9751.

Wood, T.M., McCrae, S.I., 1972. The purification and properties of the C1 component of *Trichoderma koningii* cellulase. Biochem. J. 128, 1183–92.

Wu, I., Heel, T., Arnold, F.H., 2013. Role of cysteine residues in thermal inactivation of fungal Cel6A cellobiohydrolases. Biochim. Biophys. Acta 1834, 1539–44.

Wu, M., Bu, L., Vuong, T. V, Wilson, D.B., Crowley, M.F., Sandgren, M., Ståhlberg, J., Beckham, G.T., Hansson, H., 2013a. Loop motions important to product expulsion in the

Thermobifida fusca glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. J. Biol. Chem. 288, 33107–17.

Wu, M., Nerinckx, W., Piens, K., Ishida, T., Hansson, H., Sandgren, M., Ståhlberg, J., 2013b. Rational design, synthesis, evaluation and enzyme-substrate structures of improved fluorogenic substrates for family 6 glycoside hydrolases. FEBS J. 280, 184–98.

Wu, S., Letchworth, G.J., 2004. High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. Biotechniques 36, 152–154.

Zechel, D.L., Withers, S.G., 2000. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. Acc. Chem. Res. 33, 11–18.

Zeldin, O.B., Gerstel, M., Garman, E.F., 2013. RADDOSE-3D: Time- and space-resolved modelling of dose in macromolecular crystallography. J. Appl. Crystallogr. 46, 1225–1230.

Zheng, J., Guo, N., Lin, F. lai, Wu, L. shuang, Zhou, H. bo, 2014. Screening of multi-copy mannanase recombinants of *Pichia pastoris* based on colony size. World J. Microbiol. Biotechnol. 30, 579–584.

Zou, J.Y., Kleywegt, G.J., Ståhlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T.T., Jones, T.A., 1999. Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. Structure 7, 1035–1045.

Acknowledgements

本論文は、森林化学研究室で学部4年生から取り組んできた研究をまとめたものです。

研究を進めるにあたり、いつも指針を示し導いてくださった教授の**鮫島正浩**先生に心より感謝申し上げます。私は迷走することがしばしばありましたが、その度にデータのとらえ方や研究の進め方など議論を重ねてくださり、深く正しく考えられるようにご指導をいただきました。研究室で過ごした6年間は、私にとって非常に充実して貴重な時間でした。未熟な私に多くの成長の機会をくださり、また真摯に向き合ってくださったこと重ねて深く感謝申し上げます。

いつも温かく私を鍛えて下さった准教授の**五十嵐圭日子**先生に深く感謝申し上げます。異なった視点でものごとを考える面白さを教えてくださり、消化しきれないほどの発見がある議論がとても刺激的でした。時には遅い時間まで辛抱強く議論をしてくださり、成長させてくださったこと誠に感謝致します。今後も大きな財産として、自立した研究者になれるように頑張ります。

培養、精製から結晶化、構造解析、活性測定に至るまで多くの実験をご指導くださった**中村彰彦**博士に厚く感謝申し上げます。特に結晶化は条件検討まで完了した状態で研究を始めさせてくださり、高分解能X線や中性子構造解析に挑戦することができたことも全て中村博士のご指導のおかげです。**石田卓也**博士には、構造解析データの測定、解析をはじめ、投稿論文の執筆、留学なども含めて数々のご指導をいただき、常に研究生活全般で大変お世話になりました。結晶構造解析を一から教えていただき、反応機構についても丁寧に教えてくださったこと深く感謝申し上げます。卒論生として研究を始める頃から実験の進め方を教えてくださった**杉本直久**博士に心より感謝申し上げます。実験が下手な私に粘り強く遺伝子実験をご指導してくださり、またたくさん議論をしてくださって、私は研究の楽しさを知り研究を進めてくることができました。5年間に渡って大学院での研究生活の全てで大変お世話になりました**砂川直輝**博士に厚く感謝申し上げます。酵母の扱いや活性試験をはじめ多くのご指導いただき、また困ったことがある度に相談に乗っていただき、いつも実験方法や研究の方向性に関する的確なアドバイスをいただきました。**内山拓**博士には、スクリーニング実験や活性測定方法などをご指導いただき、厚く御礼申し上げます。特に、ランダム変異導入の論文の執筆の際は、再投稿に向けて詳細に検討を重ねてくださり、アクセプトをもらうことができました。学会等でも気にかけてくださり、いつも鋭い指摘をしてくださいました**徳安健**准教授に厚く感謝申し上げます。ゼミ等でアドバイスをくださりまた研究室を支えて下さっている**寺田珠実**助教に心より感謝申し上げます。いつも優しくお声掛け頂き大変嬉しかったです。事務手続き等をはじめ多くの仕事をお引き受けくださいました**古久保美**

樹研究員、**中山陽子**研究員に深く感謝申し上げます。これまで森林化学研究室でお世話になりました先輩、同期、後輩の皆様に深く感謝致します。

JAXA-PCGプロジェクトの結晶化実験をご指導いただきました株式会社コンフォーカルサイエンスの**田仲広明**様、**高橋幸子**様、**厳斌**様、株式会社丸和栄養食品の**伊中浩治**様、**古林直樹**様、宇宙航空研究開発機構の**太田和敬**様に心より感謝申し上げます。

高分解能X線回折データの測定と解析をご指導いただきました京都大学大学院の**竹田一旗**准教授、**三木邦夫**教授に深く感謝申し上げます。SHELXでの解析方法をご指導くださいました**長谷川修之**様をはじめ同研究室の皆様に厚く御礼申し上げます。

中性子結晶構造解析のデータ測定・解析をご指導いただきました**日下勝弘**准教授、**山田太郎**准教授、**矢野直峰**助教、J-PARC職員の皆様に誠に感謝申し上げます。

X線回折データの測定をさせていただきました高エネルギー加速器研究機構、Spring8の皆様に厚く感謝申し上げます。

I would like to express my appreciation to Drs. **Jerry Ståhlberg**, **Mats Sandgren**, **Anna Borisova**, **Miao Wu**, **Bing Liu** and lab members in Swedish University of Agricultural Sciences at Uppsala.

I am deeply grateful to Drs. **Anu Koivula**, **Merja Penttilä**, **Jenni Rahikainen**, **Sanni Voutilainen**, **Ann Westerholm-Parvinen** and lab members in VTT technical research center in Finland.

東京大学大学院、酵素学研究室の**伏信進矢**教授、高分子材料学研究室の**岩田忠久**教授、木材化学研究室の**横山朝哉**准教授には副査として本論文の審査をしていただき、貴重なご助言をいただいて議論を深めることができました。心より感謝申し上げます。

また、本研究はJSPS特別研究員奨励費（15J10657）、NEDOセルロース系エタノール革新的生産システム開発事業(09000833-0)、JSPS基盤研究B（24380089）、革新的研究開発推進プログラムImPACT、JAXA-JEM利用高品質タンパク質結晶生成実験、JAXAオープンラボ公募共同研究、文部科学省宇宙科学研究拠点形成プログラム、新学術領域研究（2411400, 24114008）等の助成をいただきました。

応援してくれた友人、また学会等で関わって下さった皆様に厚く感謝申し上げます。そしていつもそばで支えてくれた佐野峻平さんに心より感謝いたします。最後に、自分のやりたいことを応援してこれまで育ててくれた父、母、弟の一央、祖父母に感謝を表し、謝辞とさせて頂きます。

2018 年 1 月

立岡 美夏子