X-ray structure and function studies of key enzymes for biomass conversion: GH6 cellobiohydrolases and GH61 lytic polysaccharide monooxygenases (LPMO)

Miao Wu

Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2013

Acta Universitatis agriculturae Sueciae
2013:23

Cover: The crystal structure of *Phanerochaete chrysosporium* GH61D (with the bound copper atom, depicted as a sphere in brown)
(Design: Miao Wu)

ISSN 1652-6880
ISBN 978-91-576-7785-3

Print: SLU Service/Repro, Uppsala 2013

X-ray structure and function studies of key enzymes for biomass conversion: GH6 cellobiohydrolases and GH61 lytic polysaccharide monooxygenases (LPMO)

Abstract

The need for large enzyme quantities due to the difficult hydrolysis of recalcitrant polysaccharides is still a major barrier to economical biomass conversion for biofuel production. To discover or develop new efficient wood degrading enzymes and add into enzyme cocktails are essential for optimizing enzymatic conversion of biomass.

Cellulases in glycoside hydrolase family 6 (GH6) play a key role and are obvious targets for enzyme discovery and engineering. However, convenient substrates for high throughput screening have not been available. **In paper I**, improved fluorogenic substrates for GH6 were rationally designed, synthesized and evaluated. Hydrolysis rates increased by 10–150 times with *Hypocrea jecorina* Cel6A. Enzyme-substrate structures showed that the modifications led to relief of the exo-anomeric effect and a better position of the glycosidic bond for protonation. **In paper II**, the first crystal structure of a bacterial GH6 cellobiohydrolase, *Thermobifida fusca* Cel6B, reveals that the enzymes has a much longer substrate binding tunnel than its fungal GH6 counterparts and that the tunnel exit is closed by a loop that needs to be displaced to allow cellobiose product release for processive action by the enzyme.

Recently, lytic polysaccharide monooxygenases (LPMOs) were discovered as a new class of enzyme for cleavage of recalcitrant polysaccharides with a novel oxidative mechanism. **In paper III and IV**, one fungal LPMO, *Phanerochaete chrysosporium* GH61D, was recombinantly expressed and shown to be metal-dependent, to cleave glycosidic bonds in wood lignocellulose and to oxidize preferentially at carbon C1. The *Pch*GH61D crystal structure is the first structure of an LPMO from basidiomycetes. Its active center shows a type II copper binding configuration, which is common to other LPMOs. Structure comparison and molecular dynamic simulations indicate three loops and a series of aromatic and polar residues near the binding surface that may influence substrate recognition and binding.

Keywords: cellobiohydrolase, GH6, GH61, *Hypocrea jecorina*, LPMO, lytic polysaccharide monooxygenase, methylumbelliferyl-β-cellobioside, *Phanerochaete chrysosporium*, *Thermobifida fusca*, X-ray crystallography

Author's address: Miao Wu, Department of Molecular Biology, SLU,
Box 7016, SE-75007 Uppsala, Sweden
E-mail: miao.wu@slu.se

To Wei

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I **Wu M***, Nerinckx W*, Piens K, Ishida T, Hansson H, Sandgren M, Ståhlberg J. (2013). Rational design, synthesis, evaluation and enzyme-substrate structures of improved fluorogenic substrates for family 6 glycoside hydrolase. *FEBS J* 280*(1),* 184-98.

II Sandgren M, **Wu M**, Karkehabadi S, Mitchinson C, Kelemen BR, Larenas EA, Ståhlberg J, Hansson H. (2013). The structure of a bacterial cellobiohydrolase: the catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. *J Mol Biol* 425(3), 622-635.

III Westereng B, Ishida T, Vaaje-Kolstad G, **Wu M**, Eijsink VGH, Igarashi K, Samejima M, Ståhlberg J, Horn SJ, Sandgren M. (2011). The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PloS One* 6(11):e27807

IV **Wu M**, Beckham GT, Larsson AM, Ishida T, Kim S, Payne CM, Himmel ME, Crowley MF, Horn SJ, Westereng B, Igarashi K, Samejima M, Ståhlberg J, Eijsink VGH, Sandgren M. (2013). Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus *Phanerochaete chrysosporium*. *J Biol Chem* 288*, doi: 10.1074/jbc.M113.459396*

* First authorship shared

Other publications:

Kuang M, Wang S, **Wu M**, Ning G, Yao G, Li L. (2010). Expression of IFNalpha-inducible genes and modulation of HLA-DR and thyroid stimulating hormone receptors in Graves' disease. *Mol Cell Endocrinol* 319(1-2):23-9.

Du B, Han H, Wang Z, Kuang L, Wang L, Yu L, **Wu M**, Zhou Z, Qian M. (2010). Targeted drug delivery to hepatocarcinoma in vivo by phage-displayed specific binding peptide. *Mol Cancer Res* 8(2):135-44.

Gudmundsson M*, **Wu M***, Ishida T, Momeni MH, Vaaje-Kolstad G, Lundberg D, Royant A, Ståhlberg J, Sandgren M. (2013). The structures of a reduced and oxidized copper active site of a CBM33 lytic polysaccharide monooxygenase. (Manuscript)

Wu M*, Bu LT*, Hansson H, Vuong TV, Wilson DB, Sandgren M, Beckham GT and Ståhlberg J. (2013). Structure and mechanism studies of bacterial GH6 cellobiohydrolase using catalytic mutants from *Thermobifida fusca* Cel6B. (Manuscript)

Abbreviations

AA	Auxiliary Activities
BGL	β-glucosidase
CBH	Cellobiohydrolase
CBM	Carbohydrate-binding module
CD	Catalytic domain
CDH	Cellobiose dehydrogenase
ClF3MUF-G2	6-chloro-4-trifluoromethylumbelliferyl-β-cellobioside
ClMUF-G2	6-chloro-4-methylumbelliferyl-β-cellobioside
ClPhUF-G2	6-chloro-4-phenylumbelliferyl-β-cellobioside
CMC	Carboxymethylcellulose
E3	*Thermobifida fusca* Cel6B
Efa/E. faecalis	*Enterococcus faecalis*
EG	Endoglucanase
GH	Glycoside hydrolase
Hj/H. jecorina	*Hyporcrea jecorina*
LPMO	Lytic polysaccharide monooxygenase
MUF-G2	4-methyl-umbelliferyl-β-cellobioside
Ncr/N. crassa	*Neurospora crassa*
Pch/P. chrysosporium	*Phanerochaete chrysosporium*
PDB	Protein Data Bank
PEG	Polyethylene glycol
PhUF-G2	4-phenylumbelliferyl-β-cellobioside
r.m.s	root mean square
RMSF	root mean square fluctuations
Sma/S. marcescens	*Serratia marcescens*
Tau/T. aurantiacus	*Thermoascus aurantiacus*
Tf/T. fusca	*Thermobifida fusca*

1 Background

1.1 Enzymatic conversion of lignocellulosic biomass

Due to depletion of fossil resources and the threat of global climate changes, one of the largest challenges that humanity is facing is the transition towards sustainable future energy supply for the growing global population. In particular, the transportation sector is still highly dependent on fossil fuels and it is necessary to develop sustainable alternative liquid fuels for transportation, so called biofuel. In the past decades, increased amount of biofuels have been produced from sugar and starch crops (e.g. sugarcanes, beets and corn) or vegetable oil in the form of bioethanol, biodiesel, or biogas. However, this has also led to an increased competition with the global food supply (Naik *et al.*, 2010). Therefore, attention is instead turning towards utilization of lignocellulosic biomass (such as agricultural residues, forestry wastes) as a potential sustainable carbon resource of soluble sugars for biofuel fermentation and in a longer perspective also replacement of other petroleum based products. However, the plant cell wall has evolved by nature to resist degradation and the well known 'biomass recalcitrance' make it more difficult to degrade with enzymes than starch materials (Himmel, 2007). Intensive research is going on around the world to improve the efficiency and lower the cost of this process.

Lignocellulosic biomass consists mainly of plant cell walls (Figure 1), which is a composite material of roughly 70-80% polysaccharides and a complex three-dimensional aromatic polymer named lignin. Cellulose, as major component of this material, is a homopolymer of glucose units linked by β-1,4-glycosidic bonds to straight chains that may be up to 15,000 glucose units long. The structure of cellulose is governed by the construction of the cellulose synthase complexes in the cell wall membrane. In higher plants, ~36 glucose chains are synthesized in parallel and aggregate into highly crystalline insoluble microfibrils that are irregularly interrupted by somewhat amorphous

regions. Several microfibrils can pack into even larger bundles. The cellulose bundles are surrounded and entangled by hemicellulose. Hemicellulose is a group of linear or branched homo- or heteropolymers of various polysaccharides, including β-glucan, xylan, xyloglucan, arabinoxylan, mannan, galactomannan, etc.. The cellulose and hemicellulose composite material is subsequently embedded with lignin, which is a three dimensional heterogeneous aromatic polymer of syringyl and guaiacyl hydroxyphenyl units. It is this complex construction of the plant cell wall with high crystallinity and dense packing that make the plant biomass highly recalcitrant and difficult to degrade (Yang *et al.*, 2011).

In nature, plant-degrading microbes, bacteria and fungi, use a multitude of extracellular enzymes to break down the polysaccharides to soluble sugars. Cellulose is the major component of the plant material, and cellulase enzymes generally dominate both in number of genes and in the amount of protein by mass. Cellulases and other carbohydrate active enzymes are divided into structurally and evolutionarily related families in the CAZy database (http://www.cazy.org/). Most of the plant degrading enzymes belong to glycoside hydrolase (GH) families because they cleave glycosidic bonds by hydrolysis. To date, 131 different GH families have been identified. Many GHs have a bimodular organization with two independently folded modules: a large catalytic domain (CD) and a smaller non-catalytic carbohydrate-binding module (CBM). These two modules are linked by a flexible linker. CDs perform the main function of the enzymes where the polysaccharide is cleaved. The CBM of a cellulase typically binds to the surface of insoluble cellulose and acts like an anchor that enhances the hydrolytic activity of the enzyme (Reinikainen *et al.*, 1992; Stahlberg *et al.*, 1988). In addition to glycoside hydrolysis, a novel oxidative mechanism for polysaccharide cleavage has been recently discovered for GH61 and CBM33 enzymes. A new enzyme class 'lytic polysaccharide monooxygenase' (LPMO) was proposed for these enzymes (Horn *et al.*, 2012) and just recently they have been classified into two new CAZy families as Auxiliary Activities families 9 and 10 (AA9, AA10) for GH61s and CBM33s, respectively.

In industry applications, the filamentous fungus *Hypocrea jecorina* is the major source for commercial enzyme cocktails (e.g. Cellic® product (Ctec/Htec) of Novozymes; Accellerase®Trio™ of DuPont, etc.). *Hypocrea jecorina* is an ascomycete soft rot fungus, and its genome has been sequenced and transcriptome analysis identified about 35 genes coding for biomass degrading enzymes when it grows in cellulose containing medium (Foreman *et al.*, 2003). Industrial strains can secret more than 100 grams enzyme per liter (Seiboth *et al.*, 2011). The major components are two cellobiohydrolases,

*Hj*Cel7A (hydrolyze processively from reducing towards non-reducing end of cellulose chain) and *Hj*Cel6A (hydrolyze processively from non-reducing towards reducing end of cellulose chain). Several endoglucanases are included, such as *Hj*Cel7B, *Hj*Cel5A, *Hj*Cel12A, and there are also two LPMOs present in the *H. jecorina* secrectome, *Hj*GH61A and *Hj*GH61B (Foreman *et al.*, 2003). Different enzyme cocktails are blended with different amount of free enzymes and can also be adjusted or optimized based on various biomass components for obtaining high catalysis efficiency.

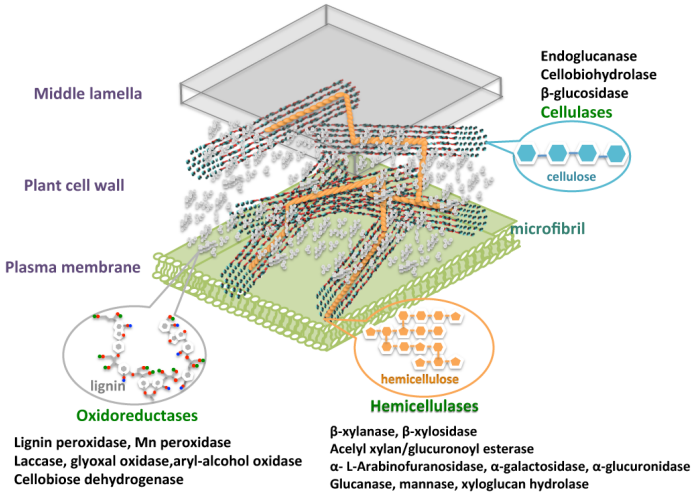


Figure 1. Plant cell wall construction and its degrading enzymes.

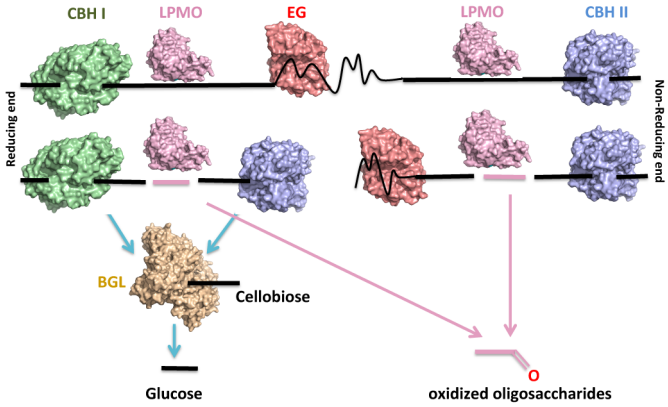


Figure 2. Schematic illustration of synergistic cellulose degradation

The different cellulolytic enzymes including endoglucanase (EG), cellobiohydrolase (CBH), β-glucosidase (BGL) and lytic polysaccharide monooxygenase (LPMO) cooperate synergistically in the degradation of cellulose (Figure 2). Endoglucanases attack amorphous regions of the cellulose, and LPMOs cleave internal glycosidic bonds, thereby creating new chain ends where the cellobiohydrolases can start hydrolysis and proceed processively into crystalline regions of the cellulose. The two *H. jecorina* CBHs (*Hj* Cel7A and *Hj* Cel6A) act processively in opposite direction along cellulose chains producing cellobiose as the main product. Finally, β-glucosidases carry out degradation of cellobiose or other soluble cellodextrins to yield glucose that can be used for fermentation. β-glucosidases play an important role in the efficiency of enzymatic conversion of biomass, because cellulase enzymes and particularly cellobiohydrolases are sensitive to product inhibition by accumulated amounts of cellobiose product. *H. jecorina* typically secrets low levels and therefore needs extra addition of β-glucosidases (Sipos *et al.*, 2010). The lytic polysaccharide monooxygenase (LPMO, including GH61 and CBM33 enzymes) can act directly on crystalline cellulose with an oxidative mechanism and produce soluble oxidized oligosaccharides as products.

1.2 Lytic Polysaccharide Monooxygenases (LPMOs)

Traditionally, the classical endo-exo synergistic model of cellulose degradation comprised endo-acting hydrolases that cut randomly in the amorphous regions of the cellulose and processive exo-acting enzymes that degrade the cellulose chains from the ends, proceeding into the crystalline regions. However, the compact structure of cellulose material makes it difficult for enzymes to get access to the surface. Already since the 1950s it has been suggested that there might be some additional factors that can help to loosen up the cellulose structure and expose cellulose chains to cellulases for further degradation (Reese, 1956). Actually, in *H. jecorina*, it was discovered that two GH61 genes are up-regulated under cellulase inducing conditions. Although these two GH61 proteins showed very weak endo-activities in cellulose degradation, they have a dramatic enhancing effect when added to a cellulase mixture (Merino & Cherry, 2007). The first structure of a GH61 protein, *H. jecorina* Cel61B, showed that the protein has a flat surface with a conserved metal binding site (Karkehabadi *et al.*, 2008). However, the substrate specificity or mechanism of action was not known at that time. The first clues came from the publication of the mechanism of chitin binding protein CBP21 from *Serratia marcescens*, a bacterial protein belonging to family CBM33 (new CAZy family AA10). CBP21 was demonstrated to cleave chitin with a novel metal-dependent

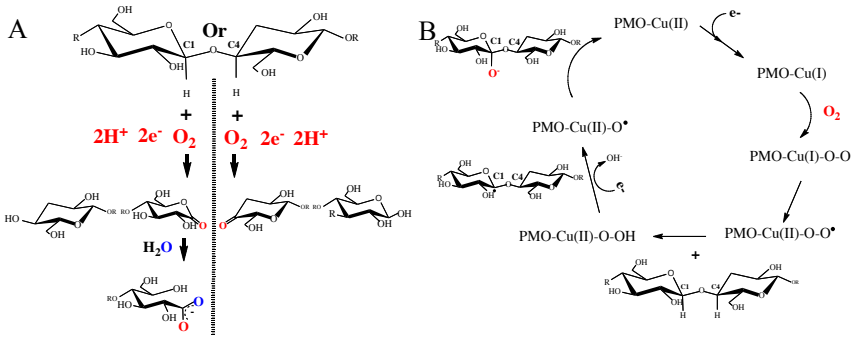
oxidative mechanism that requires both a reducing agent and molecular oxygen (Vaaje-Kolstad *et al.*, 2010). Only one of the oxygen atoms is incorporated into the oxidized product, thus the name 'lytic polysaccharide monooxygenase' was suggested. Based on structure homology and other similarities, it was proposed that GH61 proteins might have similar activity on cellulose as CBP21 has on chitin. This has later been confirmed in several studies (Wu *et al.*, 2013; Phillips *et al.*, 2011; Quinlan *et al.*, 2011; Westereng *et al.*, 2011) and also have one CBM33 (*Streptomyces coelicolor* CelS2) been shown to exhibit LPMO activity on cellulose (Forsberg *et al.*, 2011).

So far only a few LPMOs have been expressed and characterized. CBM33s from *Thermobifida fusca* have been identified to increase substrate accessibility on chitin similarly as CBP21 and potentiate hydrolytic enzymes both on chitin and cellulose substrates (Vaaje-Kolstad *et al.*, 2010; Moser *et al.*, 2008; Vaaje-Kolstad *et al.*, 2005). *S. coelicolor* CelS2, belonging to CBM33, shows cleavage of cellulose and produce aldonic acids (Forsberg *et al.*, 2011). Oxidative cleavage of cellulose has also been shown for GH61s (new CAZy family AA9) from the fungi *Thermoascus aurantiacus* (*Tau*GH61A), *Phanerochaete chrysosporium* (*Pch*GH61D) and *Neurospora crassa* (*Ncr*PMO-2, *Ncr*PMO-3) (Phillips *et al.*, 2011). These studies have proven that LPMOs are divalent metal ion dependent and that copper ion binds strongest and gives the highest activity. The enzymes require a reducing agent for catalysis. So far ascorbic acid, glutathione, gallic acid and the enzyme cellobiose dehydrogenase have been shown to act as electron donors (Langston *et al.*, 2012; Langston *et al.*, 2011; Phillips *et al.*, 2011). The requirement for molecular oxygen has also been demonstrated.

Details about the molecular mechanism of LPMOs and binding behavior on polysaccharides are yet to be elucidated. It has been shown that oxidation may occur either at the C1 position to yield saccharides with aldonic acid at the reducing end, or at the C4 position, which gives 4-keto sugars at the non-reducing end of saccharide products. Oxidation at C6 has also been suggested (Quinlan *et al.*, 2011). For instance, *Pch*GH61D and *Enterococcus faecalis Efa*CBM33 appears to oxidize only at the C1 carbon (Forsberg *et al.*, 2011; Westereng *et al.*, 2011); *Ncr*PMO-2 prefers to cleave at the C4 position and *Ncr*PMO-3 gives both C1 and C4 oxidized products (Beeson *et al.*, 2012; Bey *et al.*, 2012; Li *et al.*, 2012).

A catalytic mechanism was proposed by (Beeson *et al.*, 2012) that the hydrogen atom on either the C1 or the C4 carbon is abstracted and substituted by oxygen. The glycosidic bond is destabilized and broken and yields either a ketone at C4 position or a C1 lactone that is subsequently hydrolyzed spontaneously to a C1 aldonic acid (Scheme 1A). Aachmann et.al claim that in

the reaction, molecular oxygen is activated by the reduced copper ion, and one of the oxygen atoms oxidizes the C1 carbon and the other one is involved in the hydrolysis of the lactone to produce the aldonic acid end product (Aachmann *et al.*, 2012). Further (Li *et al.*, 2012) proposed a scheme for electron transfer during the catalytic cycle (Scheme 1 B). In addition, from a structural view, they pointed out a series of conserved residues on the *Ncr*PMO-3 surface which might provide an electron transfer pathway in the enzyme, from the copper ion at the reaction center to a potential binding site for cellobiose dehydrogenase, where the CDH haem domain was docked in place (Li *et al.*, 2012).



Scheme 1. Proposed reaction pathway of oxidative cleavage of cellulose by LPMOs (A) Proposed by Beeson, et. al., 2011, as described in the text. (B) Proposed by Li, et. al., 2012. Copper II at the metal binding site receives an electron from the reducing agent and is reduced to copper I. An O_2 molecule binds to the copper ion and takes one electron to form a PMO-Cu(II)-superoxo complex. The copper superoxo abstracts a hydrogen atom from the C1 or C4 position of the cellulose, generating a copper hydroperoxo intermediate and a substrate radical; A second electron from the reducing agent and uptake of a proton then cleave the O-O bond within the copper-peroxo, releasing water and generating a copper oxo radical. The PMO-Cu(II)-O$^{\bullet}$ couples with the substrate radical to hydroxylate the polysaccharide. The hydroxylated product is not stable and the glycosidic bond of the glucan chain will break with the elimination of the adjacent carbohydrate moiety.

Currently, there are five fungal LPMO structures in PDB: *Hypocrea jecorina* GH61B (PDBcode 2VTC)(Karkehabadi *et al.*, 2008), *Thielavia terrestris* GH61E (PDBcode 3EII, 3EJA)(Harris *et al.*, 2010), *Thermoaucus aurantiacus* GH61A (PDBcode 2YET, 3ZUD)(Quinlan *et al.*, 2011) and *Neurospora crassa* PMO-2 and PMO-3 (PDB code 4EIR, 4EIS) (Li *et al.*, 2012). There are 4 bacterial CBM33s structures in PDB: *Serratia marcescens* CBP21 (PDBcode 2LHS, 2BEM, 2BEN)(Vaaje-Kolstad *et al.*, 2005), *Vibrio cholerae* CBM33 (PDBcode 2XWX)(Wong *et al.*, 2012), *Enterococcus*

faecalis CBM33 (PDBcode 4A02)(Vaaje-Kolstad *et al.*, 2012); PDB code 4ALC, 4ALE, 4ALO, 4ALR, 4ALS, 4ALT) and *Burkholderia pseudomallei* CBM33 (PDBcode 3UAM). All these LPMO structures show similar active sites, including two conserved histidine residues, that can accommodate a metal ion. The metal ion and aromatic residues are exposed to solvent on a flat surface of the enzyme that provides a hydrophobic surface for binding to insoluble polysaccharides. The LPMOs thus seem to be well suited to cleave glycosidic bonds without the need to first decrystallize polymer chains.

GH61s are abundant in wood degrading fungi, as many as 44 GH61 encoding genes are found in *Chaetomium globosum* and 17 GH61 genes have been identified in *Phanerochaete chrysosporium* (Martinez *et al.*, 2004). Most of the CBM33s are from bacteria, but members have also been found in viruses, one archae, a slime mold and one fungus. Commercial enzyme cocktails for biofuel applications from the fungus *H. jecorina* naturally contains only two GH61s. It is predicted that substantial enhancement of the activity may be achieved by further addition of LPMOs to enzyme cocktails. Intensive research is conducted to deepen our understanding of the diversity among the large repertoire of LPMOs present in Nature and how they can be exploited in the future.

1.3 Glycoside hydrolase family 6

Enzymes of glycoside hydrolase family 6 are common in both cellulolytic fungi and bacteria and play key role in biomass degradation in Nature. GH6 contains only cellulase enzymes, both endoglucanases and cellobiohydrolases. The GH6 cellobiohydrolases act from the nonreducing towards the reducing end of the cellulose polymer (the opposite direction compare to GH7 CBHs) and produce cellobiose units as the main product.

The first structure of a GH6 enzyme, the catalytic domain of *H. jecorina* cellobiohydrolase Cel6A (PDBcode 1CBH), was solved in 1990 in Uppsala, Sweden (Rouvinen *et al.*, 1990). The structure revealed a distorted seven-stranded α/β barrel and the active site located inside a substrate binding tunnel. The tunnel is lined with conserved tryptophan residues that serve as hydrophobic sugar binding platforms in subsites -2, +1, +2 and +4 (W135, W367, W269, W272, respectively; Figure 3A). The substrate binding path extends along 6 to 7 glucosyl units with the subsites numbered -3, -2, -1, +1,+2,+3 and +4 from the non-reducing end to the reducing end of a cellulose chain. The cleavage takes place between subsites -1 and +1.

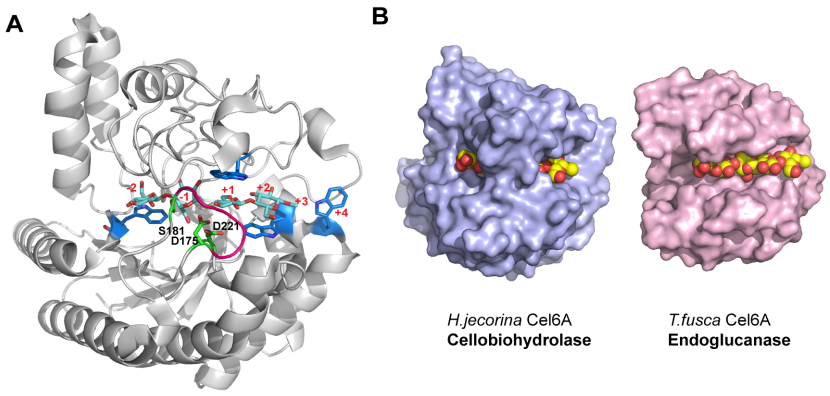


Figure 3. Glycoside hydrolase family 6 structures. (A) Cellobiohydrolase Cel6A from *H. jecorina* (PDB code 1QJW). Sugar-binding tryptophans are shown in and 6 glucosyl subsites, -2 to +4 are indicated. The catalytic acid D221, the proposed catalytic base D175, and S181 are shown in green. The mobile active center loop is highlighted in pink. (B) Surface models of GH6 enzymes. In a cellobiohydrolase (left) the cellulose chain is enclosed in a tunnel, while it is bound in a cleft in an endoglucanase (right).

The first GH6 endoglucanase structure was the catalytic domain of Cel6A (E2) from the thermophilic bacterium *Thermobifida fusca* (PDBcode 1TML) (Spezio *et al.*, 1993). In contrast to the GH6 cellobiohydrolases, the active sites of GH6 endoglucanases are located in a more open cleft rather than a tunnel, which indicate their higher tendency to cut internal bonds in the cellulose polymer (Figure 3B). Other GH6 X-ray structures available at PDB were: three more CBHs, *Humicola insolens* Cel6A (PDBcode 1BVW)(Varrot *et al.*, 1999), *Coprinopsis cinerea* Cel6A (PDBcode 3VOH)(Tamura *et al.*, 2012; Liu *et al.*, 2010) and Coprinopsis *cinerea* Cel6C (PDBcode 3A64)(Liu *et al.*, 2010); two more EGs, *Humicola insolens* Cel6B (PDBcode 1DYS)(Davies *et al.*, 2000) and *Mycobacterium tuberculosum* Cel6A (PDBcode 1UP0)(Varrot *et al.*, 2005). All these GH6 catalytic domains share the same structural fold.

GH6 enzymes hydrolyze β-1,4-glycosidic bonds with an inverting mechanism (Knowles *et al.*, 1988). The anomeric carbon is directly attacked by a water molecule in a single step nucleophilic substitution reaction (SN_1) that cleaves the β-glycosidic bond upon creation of a new reducing end with alpha-anomeric configuration.

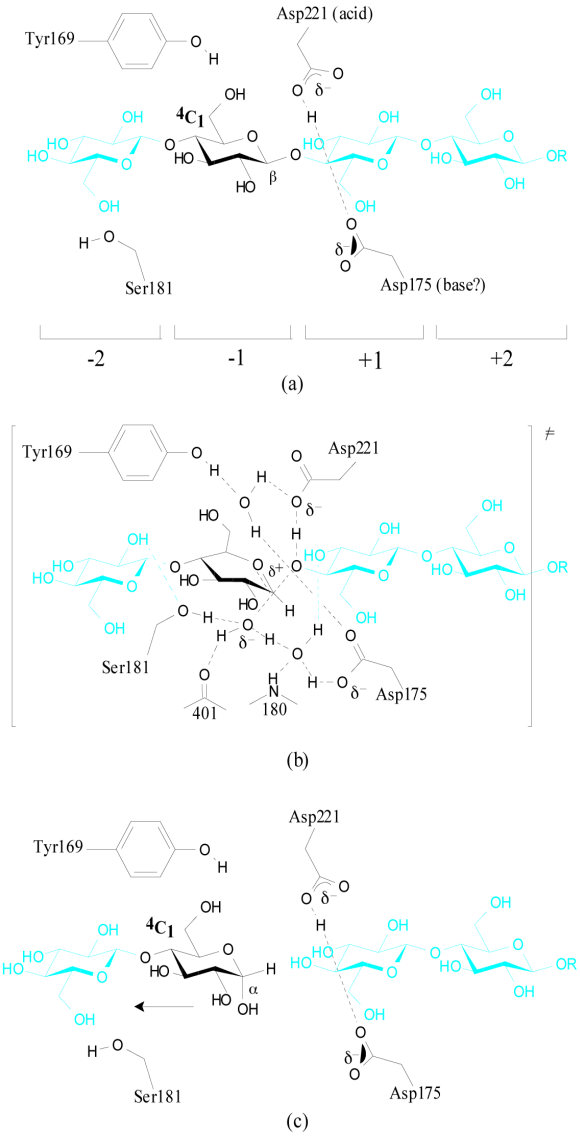


Figure 4. Grotthus inverting mechanism of GH6 enzyme from *H. jecorina* Cel6A as described in the text (Drawn by Jerry Ståhlberg)

Structure studies and mutational experiments of GH6 enzymes provide a detailed picture of the catalytic mechanism, illustrated in Figure 4 (Varrot *et al.*, 2002; Varrot *et al.*, 1999; Zou *et al.*, 1999). The conserved aspartic acid

residue D221 in *H. jecorina* Cel6A acts as the catalytic acid that protonates the scissile glycosidic bond between subsites -1 and +1. A water molecule is held in position by the hydroxyl group of S181 and the backbone carbonyl of D401 for nucleophilic attack on the anomeric carbon of the -1 D-glucose unit. This water molecule is not in direct contact with any carboxylate residue that may act as catalytic base. Instead, it is hydrogen bonded to a second water molecule, which in turn binds to the side chain of an aspartate residue, D175 in *H. jecorina*. Upon attack on the anomeric carbon, the nucleophilic water molecule gives a proton to the second water molecule and the proton is then taken up by D175, which thus acts indirectly as the catalytic base. The attacking water will end up as a new hydroxyl group on the other side of the anomeric carbon, thus inverting the anomeric configuration. The term 'Grotthuss' mechanism is used for this kind of proton transfer through a hydrogen bond network of water molecules (Koivula *et al.*, 2002).

Dual conformations have been observed for both the catalytic residues D175 and D221. In most of the structures they form a hydrogen bond to each other, and this hydrogen bond must be broken and both side chains must be in their 'active' conformation to catalyze the reaction. Actually, in *H. jecorina* the active conformer of D221 has only been observed in one structure so far (molecule B of 1HGW). Furthermore, an active center loop is flexible and needs to swing in to a closed conformation to isolate the two water molecules near the anomeric carbon. The hydroxyl group of S181 moves around 8Å upon the conformation change of the active center loop (Figure 4) (Koivula *et al.*, 2002; Zou *et al.*, 1999).

1.4 *Anti/Syn* protonation and the exo-anomeric effect

The concept of semi-lateral glycosidic oxygen protonation was introduced by (Heightman & Vasella, 1999) and was originally described for beta-equatorial glycoside hydrolases, but appears to be equally applicable to enzymes acting on alpha-axial glycosidic bonds (Nerinckx *et al.*, 2005). The catalytic acid is either positioned on the same side of the glycosidic bond as the ring oxygen of the -1 glycoside unit, then the enzyme is a *syn* protonator, or it is located on the opposite side, in which case it is an anti protonator (Figure 5). Members of the same GH family are always *syn* or always *anti* protonators and this specificity appears to be preserved within clans of families.

Whether an enzyme is *syn* or *anti* protonator has implications for the orientation of the glycosidic bond, due to the so-called 'exo-anomeric effect'. When the orbital of one of the lone electron pairs of the glycosidic oxygen is pointing in the opposite direction as the bond between the anomeric carbon and

the ring oxygen (C1'-O5'), there is hyperconjugative overlap of the C1'-O5' antibonding orbital with the orbital of the antiperiplanar lone pair, thereby creating partial double bond character of the glycosidic bond and stabilization by 4-5 kcal/mol (Johnson *et al.*, 2009; Cramer *et al.*, 1997). *Anti* protonators overcome this exo-anomeric effect by directly protonating the overlapping antiperiplanar lone pair orbital of the glycosidic oxygen, which removes the stabilizing effect (Figure 5, center). However, *syn* protonators need another strategy and instead the glycosidic bond must rotate away from its lowest energy orientation (synclinal ϕ angle) so that hyperconjugative overlap is no longer possible (Figure 5, right).

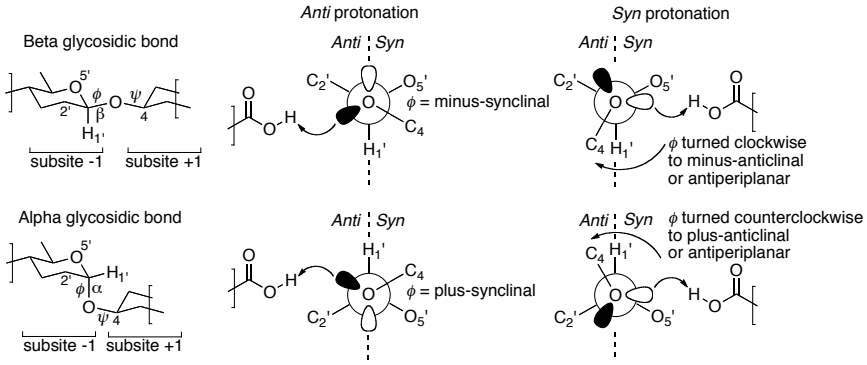


Figure 5. Newman projections, with glycosidic oxygen as proximal atom and the anomeric carbon as tistal atom, showing *anti* (center) versus *syn* (right) semi-later protonation in beta-equatorial (top) and alpha-axial (bottom) glycoside hydrolases. The indicated ϕ is the dihedral angle for O5'-C1'-O4-C4.

The enzymes of GH6 are *syn* protonators. Based on the considerations about the *syn* protonation and the exo-anomeric effect, GH6 substrates need to bind to the enzyme with the dihedral angle ϕ of the scissile anomeric bond in the minus-anticlinal position. The energetic burden (4-5 kcal.mol^{-1}) for removal of the exo-anomeric effect has to be compensated by favorable enzyme-substrate interactions.

1.5 Aim and outline of this thesis

The aim of this thesis was to study novel enzymes for plant biomass degradation including *Thermobifida fusca* Cel6B and *Phanerochaete chrysosporium* GH61D and to develop fluorogenic substrates for GH6 enzyme studies. The specific objectives of the study were

- To rationally design and synthesize fluorogenic substrates for GH6 enzymes, and evaluate these substrates and their applications. (Paper I)

- To solve the structure of *Thermobifida fusca* Cel6B and compare with other known GH6 structures. (Paper II)

- To characterize *Phanerochaete chrysosporium* GH61D, solve its structure, and investigate its interaction with substrates and cofactors. (Paper III and IV)

2 Current investigation

2.1 Fluorogenic substrates for glycoside hydrolase family 6 (paper I)

2.1.1 Introduction

Whether the development of improved enzyme cocktails is done through screening for new enzymes or engineering of already known enzymes, a useful and convenient enzyme assay is an indispensable part of research.

Aryl saccharides have been used for many years as chromogenic substrates for many different glycoside hydrolase enzymes. For example, chromogenic and fluorogenic cellobiosides and lactosides have been very useful substrates in the research on GH family 7 cellulases (e.g. cellobiohydrolase Cel7A and endoglucanase Cel7B of *H. jecorina*). They are applied in expression assays, purification monitoring, high-throughput screening, pH and temperature dependency profiling, enzyme kinetics and inhibition studies. However, similar substrates for such applications have not been available for *H. jecorina* Cel6A and other GH6 enzymes.

4-methylumbelliferyl (MUF), is a common fluorophore in artificial fluorogenic substrates for several kinds of enzymes. Already in the 1980s Claeyssens and coworkers synthesized MUF-cellooligosaccharides and used for analysis of cellulase enzymes (vanTilbeurgh *et al.*, 1982). The substrates were readily hydrolyzed by GH7 enzymes, but they did not work well as substrates for GH6 enzymes because of poor hydrolysis of the heteroglycosidic bond and preferential cleavage at homoglycosidic bonds in the substrate.

Since structures are available of GH6 enzymes with oligosaccharides bound at the active site, the placement of a MUF substrate might give ideas on why it is not hydrolyzed by GH6 enzymes. Therefore, in this paper we applied automated docking of MUF-G2 into the active site of *Hj*Cel6A in order to obtain information on how a umbelliferyl group may be oriented at +1 subsite

and how the aglycone can be modified towards better positioning and more efficient hydrolysis. Four modified umbelliferyl β-cellobiosides were designed, synthesized and tested as substrates for GH6 and some other glycoside hydrolases and enzyme-substrate complex structures have been determined by X-ray crystallography.

2.1.2 Design of fluorogenic substrates for GH6 enzymes

Automated docking of MUF-G2 into *Hj*Cel6A structure 1QJW gave, among several alternative poses, one productive pose with the cellobiosyl unit occupying subsites -2 and -1, the MUF group in +1 subsite, and with the glycosidic oxygen in position for protonation by the catalytic acid. Two positions were observed for substitution towards possibly better recognized variants that can be easily synthesized: (1) the 4-methyl group of the umbelliferone points towards the +2 subsite which can be filled by introduction of a 4-phenyl group. (2) the C6-H6 bond of the umbelliferone moiety points towards an empty local zone in subsite +1. In the original 1QJW structure, that space is occupied by the C6 hydroxymethyl group of the +1 glucoside of the bound oligosaccharide ($(Glc)_2$-S-$(Glc)_2$-O-methyl). Introduction of a chlorine at C6 of the umbelliferone would occupy this zone. This is an electron-withdrawing group that will increase the leaving capacity of the aglycone and because it is close to the anomeric oxygen it may also decrease the exo-anomeric effect. Incorporation of a 4-trifluoromethyl group can further increase the leaving capacity of the aglycone. The set of modified substrates designed and synthesized were ClMUF-G2, ClF3MUF-G2, PhUF-G2, and ClPhUF-G2 (Figure 6).

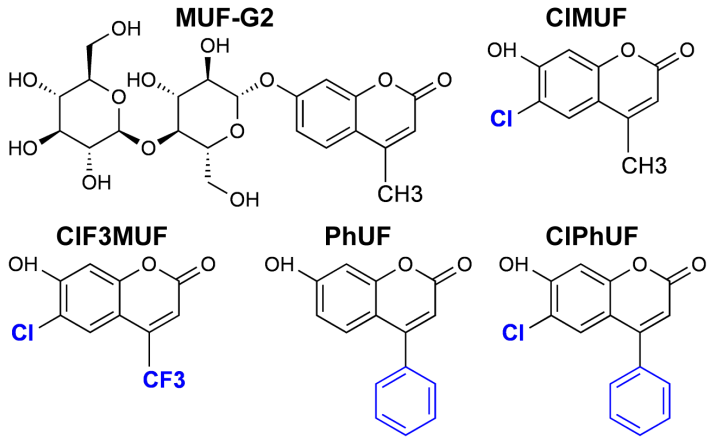


Figure 6. Design of the fluorogenic cellobioside subtrates

Table 1. *Fluorescence intensity of the umbelliferone variants and initial hydrolysis rate of fluorogenic substrates*

Fluorogenic substrate	Fluorophore intensity		Hydrolysis rate of 0.5mM fluorogenic β-cellobiosides at 37°C[a] (min^{-1})[b]				
	pH 5.0	pH 10.2	*Hj* Cel6A	*Tf* Cel6A	*Hj* Cel7B	*Hj* Cel5A	*Hj* Cel45A
MUF-G2	7	460	0.001	0.006	46	0.0005	0.0001
ClMUF-G2	26	327	0.078	0.192	204	0.003	0.0004
ClF3MUF-G2[c]	20	82	0.176	-	250	-	-
PhUF-G2	3	5	0.024	0.018	25.8	0.018	0.008
ClPhUF-G2	0.2	2	0.012	0.012	10.2	0.018	0.006

[a]pH 5.0 with *H. jecorina* Cel6A and Cel7B; pH 5.5 with *T. fusca* Cel6A. [b]Mol fluorophore released per mol enzyme per minute. [c]The concentration of ClF3MUF-G2 was limited to ~0.1 mM due to low solubility.

2.1.3 Fluorescence properties and activities on fluorogenic substrates

ClMUF, PhUF and ClPhUF show similar fluorescence spectra as MUF with excitation (Ex) and emission (Em) maxima at 360 and 450 nm, respectively, while ClF3MUF displays a shift to longer wavelengths (Ex 385 nm, Em 495 nm). All fluorophores showed higher fluorescence intensity at pH 10.2 than at pH 5.0. ClMUF and ClF3MUF gave 4 and 3 times higher intensity at pH 5.0, but lower intensity at pH 10.2, than MUF. Phenyl substituted PhUF and ClPhUF gave weaker fluorescence (Table 1).

The activities of different cellulases against synthesized fluorogenic substrates could be followed by incubation in a microtiterplate while monitoring the fluorescence under UV light. The substrate themselves did not show any visible fluorescence (Figure 7). Both *Hj*Cel6A and *Tf*Cel6A showed clear activity against ClF3MUF-G2 and ClMUF-G2. Weaker fluorescence were seen for ClPhUF-G2 and PhUF-G2 probably due to the low intensity of the free fluorophores. *Hj*Cel7A, Cel7B and *Tf*Cel9B can hydrolyze all the substrates and showed very strong fluorescence, whereas no activity was detected for *Tf*Cel9A.

Initial hydrolysis rate assays were set up by using fluorescence reader with temperature control, and results are shown in Table 1. The GH6 enzymes did actually hydrolyze MUF-G2, but with very low rate. The modified umberlliferyl substrates were hydrolyzed much faster by GH6 enzymes. ClF3MUF-G2 gave the highest rate increase compared to MUF-G2 (150 times with *Hj*Cel6A) followed by ClMUF-G2. However, the ClF3MUF-G2 substrate was not soluble above 0.1 mM concentration. Therefore ClMUF-G2 was chosen as substrate for enzyme kinetics measurements and kinetic parameters

could be determined for both *Hj*Cel6A and *Tf*Cel6A. Interestingly, *Hj*Cel5A and Cel45A also showed up to 30 and 80 times higher hydrolysis rate, respectively, with the modified fluorogenic cellobioside substrates.

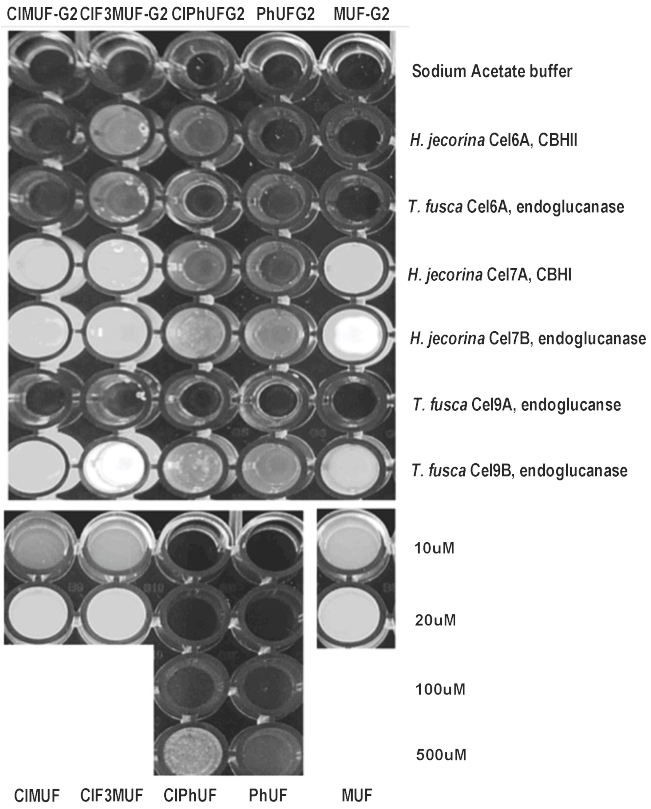


Figure 7. Fluorescence on a UV table of the umbelliferone variants, and corresponding β-cellobiosides after 1 hour incubation at 40 °C with enzymes *H. jecorina* Cel6A, 2 µM; Cel7A, 1 µM; Cel7B, 0.1 µM (pH 5.0); *T. fusca* Cel6A, 2 µM; Cel9A, 1 µM; Cel9B, 0.1 µM (pH 5.5).

2.1.4 Structure study of enzyme complexes with fluorogenic substrates

In order to avoid hydrolysis of the substrate, the catalytic deficient mutant *Hj*Cel6A D221A was used for structure studies. The protein was crystallized using previously established conditions (McPherson, 1982). Crystals were soaked with 1 mM of the substrates for 1.5 hour prior to freezing in liquid nitrogen. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble, France). Three enzyme-complex structures (*Hj*Cel6A D221A with MUF-G2, ClMUF-G2 or ClPhUF-G2,

respectively) were solved by molecular replacement. X-ray crystallography data are summarized in Table 2.

Table 2. *Summary of statistics from X-ray crystallography data of HjCel6A D221A complex structures*

Ligand complex	MUF-G2	ClMUF-G2	ClPhUF-G2
PDB code	4AX7	4AU0	4AX6
Space group	*P*1	*P*1	*P*1
Molecules / asymmetric unit	4 (A,B,C,D)	2 (A,B)	2 (A,B)
Resolution range (Å)[a]	30.0-1.7(1.8-1.7)	32.0-1.7(1.8-1.7)	35.7-2.2(2.3-2.2)
Rwork/Rfree (%)	18.9/22.1	19.4/23.7	20.4/25.5
Avg overall B-factor (Å2) [b]	17.3	15.3	13.2
Avg ligand B-factor (Å2) [b]	19.5	17.5	13.8
Binding sites at substrate binding tunnel of *Hj*Cel6A	+1,+2,+3 (A, C) -2,-1,+1 (B, D)	+1,+2,+3 (A, B)	-2,-1,+1(A, B)

[a] Information in parentheses refers to the highest resolution shell.

[b] Calculated using MOLEMAN2.

The complex with MUF-G2 explains why it is poorly hydrolyzed by GH6 enzymes. Firstly, MUF-G2 binds preferentially in subsite +1 to +3 which is out-of-register with the catalytic center (Table 2). Secondly, when bound at the correct -2 to +1 subsites, MUF-G2 adopts a nonproductive conformation. The D-glucoside moiety at subsite -1 is in the unproductive ground state $^4C_1$ chair conformation. The anomeric carbon, or rather its anti-orbital to the glycosidic bond, is poorly accessible for nucleophilic attack. Furthermore, the heteroglycosidic bond displays a dihedral angle ϕ_2 (-88°) where it is stabilized by the exo-anomeric effect. Finally, the heteroglycosidic oxygen is rather distant (3.9Å) from, and is not pointing a free electron pair straight at the catalytic acid.

The synthesized ClPhUF-G2 binds only at the catalytic center, from -2 to +2 subsite, with the 4-phenyl group occupying subsite +2. The umbelliferyl group turns 180° to 'upside-down' orientation compared to the MUF group in MUF-G2 structure. This complex shows all the requirements for the substrate to be in a catalytically productive pose. The D-glucoside moiety at subsite -1 has adopted the pre-TS ^2S_O skew conformation (Figure 8C). In addition, the distance between the glycosidic oxygen and the catalytic acid is shorter (3.0Å) (Figure 8A) and a free electron pair points towards the catalytic acid. Furthermore, the heteroglycosidic bond has turned to a dihedral angle (ϕ_2=-

105°) where exo-anomeric stabilization is relieved (Figure 8A). All these changes make ClPhUF-G2 hydrolysable from a structural point of view.

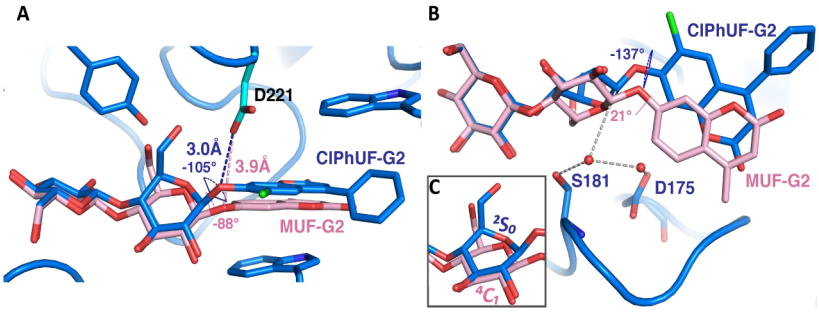


Figure 8. Comparison of MUF-G2 (pink) and ClPhUF-G2 (blue) binding at the catalytic center of *H. jecorina* Cel6A D221A. (A) The glycosidic oxygen of ClPhUF-G2 would be closer to the catalytic acid D221, and its dihedral angle ϕ_2 is in minus-anticlinal position, which relieves exo-anomeric stabilization. (B) Rotated view showing the proposed catalytic base D175 and the two water molecules (red spheres) involved in the Grotthus mechanism. Note that the umbelliferyl units are flipped "upside-down" as reflected by the respective ψ_2 dihedral angles. (C) The D-glucoside unit in subsite –1 adopts a $^4C_1$ chair with MUF-G2 and a ^2S_O skew with ClPhUF-G2. The ligand from chain B of the MUF-G2 complex, and D221 from chain B of pdb entry 1HGW, were superimposed with chain A of the ClPhUF-G2 structure. The active conformation of the catalytic acid D221 was taken from chain B of pdb entry 1HGW.

The ClMUF-G2 complex showed the substrate occupying subsites +1 to +3 which is not a catalytically productive binding position. Obviously the substrate must also be able to bind productively in subsites -2 to +1 and it was actually hydrolyzed faster than ClPhUF-G2 despite its preference for nonproductive binding. It thus appears that introduction of the chlorine substitution is beneficial by forcing the umbelliferyl group to bind in a favorable orientation for hydrolysis, and introduction of the phenyl group can help to shift binding preference towards the productive position.

By comparing the glycosidic bond dihedral angles ϕ_2 and ψ_2 at the –1/+1 subsite junction in crystal structures of GH6 enzymes and the distance between the glycosidic oxygen/sulphur and the catalytic acid (Table 3), we note that with a glucoside unit at subsite +1, the ϕ_2 angle is rotated around –20° further away from the minus-synclinal position compared to the ClPhUF-G2 complex. Perhaps that is necessary to obtain further relief of exo-anomeric stabilization. Therefore, there may be a potential to improve the hydrolysability by optimizing the fit of the aglycone at subsite +1 such that the φ_2 angle will be

restrained within the range observed when D-glucoside is attached at subsite +1. Furthermore, a twist of the dihedral angle ψ_2 around 20° anticlockwise seems to shorten the distance between the glycosidic oxygen and the catalytic acid.

Table 3. *Glycosidic bond dihedral angles ϕ_2 and ψ_2 at the –1/+1 subsite junction in crystal structures of GH6 enzymes, and distance between the glycosidic oxygen/sulphur and the catalytic acid. The MUF-G2 and ClPhUF-G2 complexes are published in Paper I.*

Crystal structure	**+1 group**	**ϕ_2**[a]	**ψ_2**[a]	**Distance (Å)**[b]
MUF-G2, B-mol	MUF	-88°	+21°	3.7
MUF-G2, D-mol	MUF	-90°	+21°	3.9
MUF-G2, A-mol	Glc	-123°	-155°	2.5
MUF-G2, C-mol	Glc	-122°	-150°	2.8
ClPhUF-G2, A-mol	ClPhUF	-105°	-137°	3.0
ClPhUF-G2, B-mol	ClPhUF	-103°	-133°	2.9
1QK2, A-mol[c]	4S-Glc	-125°	-148°	2.5
1QK2, B-mol[c]	4S-Glc	-125°	-149°	2.4
1GZ1[d]	4S-Glc	-132°	-158°	2.6
2BOF[e]	Glc	-140°	-161°	3.3

[a]Dihedral angles ϕ_2 (O5'-C1'-O7$_{umb}$-C7$_{umb}$) or (O5'-C1'-O/S4$_{Glc}$-C4$_{Glc}$) and ψ_2 (C1'-O7$_{umb}$-C7$_{umb}$-C6$_{umb}$) or (C1'-O/S4$_{Glc}$-C4$_{Glc}$-C5$_{Glc}$). [b]Glycosidic O or S distance to the "active" rotamer of D221 from B chain of PDB-entry 1HGW (Koivula *et al.*, 2002). [c]*H. jecorina* Cel6A wt/Glc$_2$-S-Glc$_2$ complex (Zou *et al.*, 1999). [d]*H. insolens* Cel6A D416A/Glc$_2$-S-Glc$_2$ complex (Varrot *et al.*, 2002). [e]*T. fusca* Cel6A Y73S/cellotetraose complex (Larsson *et al.*, 2005).

2.2 *Thermobifida fusca* Cel6B structure and mechanism (paper II)

2.2.1 Introduction

Thermobifida fusca, an aerobic filamentous bacterium, has received attention as a major plant cell wall degrader in heated organic materials such as compost heaps (Wilson, 2004). It secrets a simple set of cellulases including

*Tf*Cel5A(E5), *Tf*Cel6A(E2), *Tf*Cel6B(E3), *Tf*Cel9A(E4), *Tf*Cel9B(E1), *Tf*Cel48A(E6), E7(*Tf*CBM33) and E8(*Tf*CBM33) (Irwin *et al.*, 1993). Two of them belong to the GH6 family. *Tf*Cel6A is an endoglucanase, which has been characterized and whose structure is available in the PDB (PDBcode 1TML) (Spezio *et al.*, 1993). *Tf*Cel6B (EC 3.2.1.91) is the other GH6 family enzyme, classified as cellobiohydrolase with similar properties as *H. jecorina* Cel6A. It hydrolyzes cellulose chains preferentially from the non-reducing end. It has weak activity on polysaccharide substrates on its own, but is a very important component in the *T. fusca* enzyme system since maximum synergistic activity is only achieved in the presence of *Tf*Cel6B (Irwin *et al.*, 1993). In addition, it is attractive for industrial application that *Tf*Cel6B shows higher thermostability and broader pH optimum than the similar cellobiohydrolase Cel6A from *H. jecorina* (Vuong & Wilson, 2009b). David B. Wilson and coworkers have characterized *Tf*Cel6B in terms of substrate specificity, cellulose binding, processivity and synergism, and have introduced several mutations to study its catalytic mechanism (Vuong & Wilson, 2009b; Jung *et al.*, 2003). To date, there are eight GH6 enzyme structures in PDB, including five fungal cellobiohydrolases (Tamura *et al.*, 2012; Thompson *et al.*, 2012; Liu *et al.*, 2010; Varrot *et al.*, 1999; Rouvinen *et al.*, 1990) and three endoglucanases from both fungi and bacteria (Varrot *et al.*, 2005; Davies *et al.*, 2000; Spezio *et al.*, 1993).

In this research, we presented the first crystallographic X-ray structure of a bacterial GH6 cellobiohydrolase, the catalytic domain of *Thermobifida fusca* Cel6B (formerly E3), determined both in the absence of any ligand and in the presence of the product cellobiose.

2.2.2 Crystal structure of *Thermobifida fusca* Cel6B

The pure intact *T. fusca* Cel6B enzyme, including both cellulose binding module and catalytic domain, was produced at DuPont™ Industrial Biosciences. To remove the cellulose binding module from the intact enzyme for crystallization experiments, *T. fusca* Cel6B was digested with papain (100:1 w/w) overnight at room temperature and the catalytic domain was isolated by ion-exchange chromatography.

Crystallization of *Tf*Cel6B catalytic domain was carried out using the hanging drop method with/without cellobiose. Micro-seeding technique was necessary to obtain good quality crystals within several weeks. The crystals without ligand were obtained in 20% polyethylene glycol (PEG) 6000, 0.1M sodium citrate (pH 4.0) and 1 M lithium chloride and the complex *Tf*Cel6B crystal was obtained in 20% polyethylene glycol (PEG) 6000, 0.1 M sodium acetate (pH 4.0), 50 mM calcium chloride, and 10 mM cellobiose. The X-ray

diffraction data sets of *Tf*Cel6B in apo and complex form were collected at ESRF and Maxlab, respectively.

T. fusca Cel6B crystal structures without/with ligand were solved at 1.5 Å and 2.2 Å resolutions at the space group $P2_1$ but different cell dimensions (b= 93.3 Å and 106.7 Å), with R_{work}/R_{free} value of 0.198/0.236 and 0.191/0.228, respectively. Both of the structure models were modeled from the first to the last residue of catalytic module, and showed similar overall structure. However, seven residues are not visible in the electron density of the *Tf*Cel6B apo structure, and are not included in the final model.

The overall fold of *Tf*Cel6B catalytic domain is similar with the CD of other GH6 enzymes, but with a narrower and significantly longer substrate binding tunnel than in the others. In addition, the tunnel is further closed in the ligand bound form of *Tf*Cel6B structure. Although GH6 enzymes show low sequence identity, the substrate binding sites and catalytic residues are highly conserved (with 25 important residues being identical). *Tf*Cel6B catalytic domain sequence contains the largest number of residues (420 residues) among structurally known GH6s from both fungi and bacteria as shown in the structure-based sequence alignment (Figure 9). The bulk of additional residues in *Tf*Cel6B is located in six discrete sections of the sequence. These six sections, denoted I-VI from the N-terminus (red in Figure 10), might be significant in substrate binding, recognition and processivity.

Section I (residues 185-197) constitutes a 13-residue long insertion into an alpha helix near the exit of the substrate binding tunnel. In the ligand complexed *Tf*Cel6B structure this loop bends in towards the product binding sites, while it is heading off in a different direction in the apo structure where most of the loop is not visible (residues 188-194). Thus it appears that the section I loop can change conformation and close to effectively block the tunnel exit or open like a lid for release of product. In the closed form, the loop occupies the -3 subsite, leaving no room for a sugar unit at this subsite. T191 at the tip of section I loop makes hydrogen bonds interactions with conserved residues involved in sugar binding (Figure 10B).

Section II (residue 284-289) is a loop that bends over subsites +3 and +4, and makes interactions with another extended loop, section VI (residue 517-530), from the other side of the substrate binding tunnel. They extend the substrate binding tunnel by approximately 10 Å.

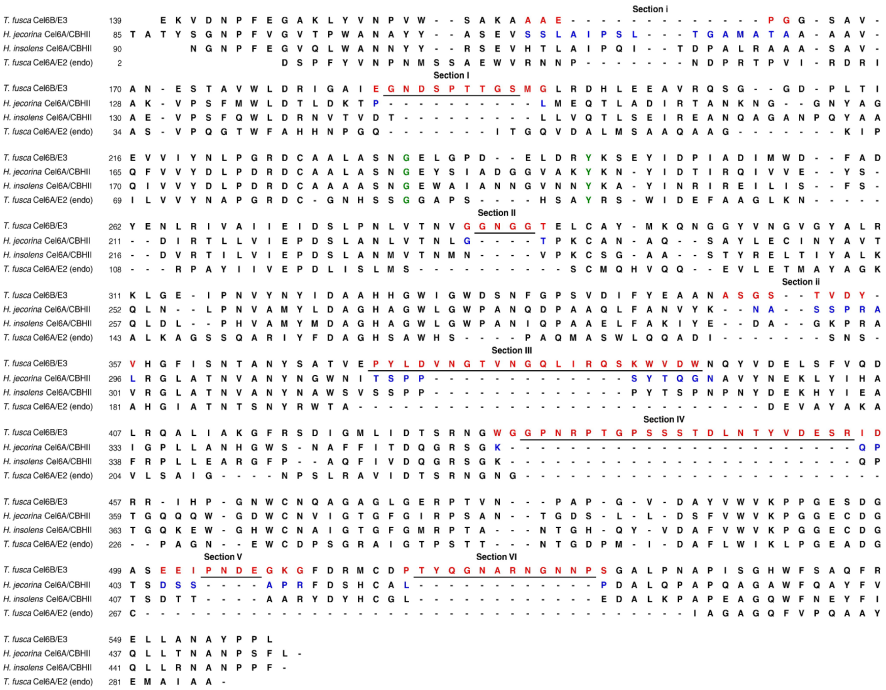


Figure 9. Structure-based sequence alignment of the catalytic domains of *T. fusca* Cel6B, the cellobiohydrolases Cel6A of *Hypocrea jecorina* and Cel6A of *Humicola insolens* and the endoglucanase Cel6A from *Thermobifida fusca*. Regions with notable variation compared to the three known GH6 cellobiohydrolases are colored red for *T. fusca* Cel6B and blue for *H. jecorina* Cel6A. Two other sections in *T. fusca* Cel6B have fewer residues than in the fungal counterparts and are indicated i and ii. The residue numbering is given in the beginning of each row.

Section III (residue 373-394) and IV (residue 431-456) are folded together to form an extra substructure near the substrate tunnel entrance with unknown function (Figure 10C). The last residue of section III, W394 is suitably positioned to form a putative +6 subsite, and the preceeding residue, D393, can form a hydrogen bond to OH2 of the glucose unit at subsite +4. Sequence alignment reveals that this motif in section III, a tryptophan preceded by an aspartate or a glutamate is common among bacterial cellobiohydrolases of GH6.

Section V (residue 501-510) is a horseshoe-shaped loop that forms the roof of the substrate binding tunnel above the -1 and -2 subsites (product binding site). The side chain of K509 (N^{ζ} atom) hydrogen bonds to the OH3 of the glucose residue at subsite -2 (Figure 10D). In addition, in the ligand-bound structure, the section V loop is stabilized by a salt bridge between E507 at the

tip of the loop and H541, which in turn makes van der Waals interaction with P190 at the tip of the section I loop.

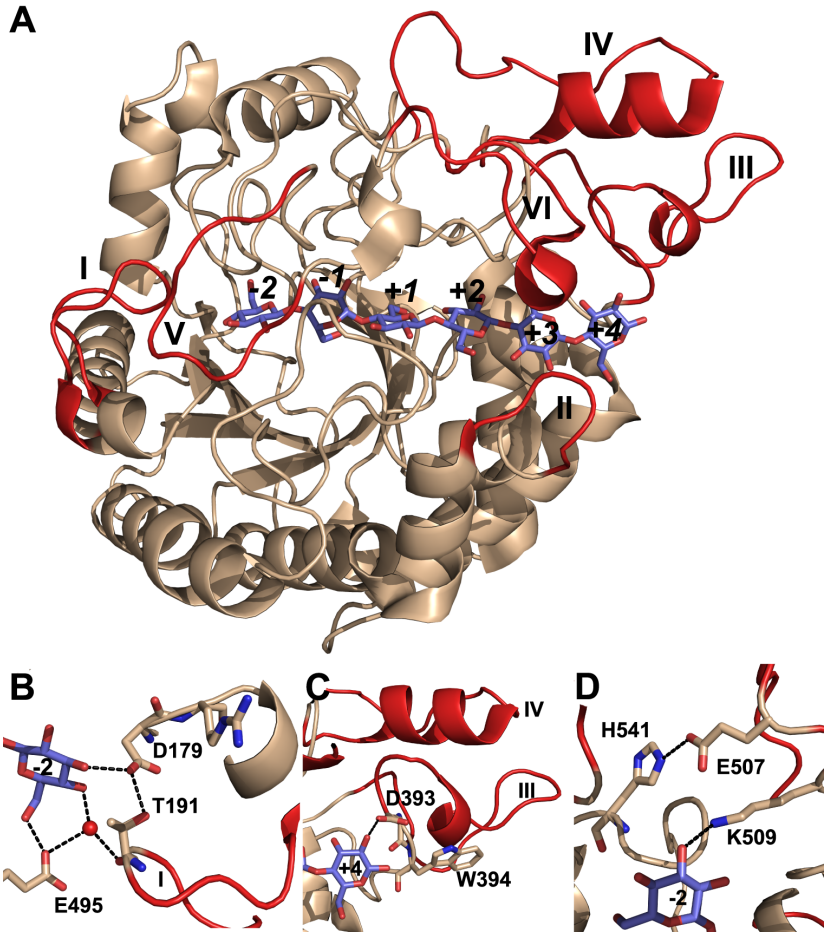


Figure 10. Crystal structure of *Tf*Cel6B with sugar bound. (A) Secondary structure representation of *Tf*Cel6B ligand bound structure. The longer sections (I-VI) compared to *H. jecorina* Cel6A are highlighted in red. (B) Close up view of section I loop. T191 interacts with D179 which in turn hydrogen bonds to OH3 of the -2 glucose unit. The carboxyl group of E495 and the carbonyl of T191 stabilize the -2 glucose by hydrogen bonded network to OH6 and OH4. (C) Section III and IV of *Tf*Cel6B with residues D393 and W394. In this ligand bound *Tf*Cel6B structure D393 moved closer, to make a hydrogen bond with OH2 on the +4 glucose unit. (D) K509 and E507 in section V loop. K509, located at the tip of section V hydrogen bonds to OH3 of the -2 glucoside and E507 forms a salt bridge to H541.

2.2.3 The active site of *T. fusca* Cel6B

As defined previously, the substrate binding tunnel is numbered from the non-reducing end to the reducing end of sugar, -2, -1, +1, +2, +3, +4, with the glycosidic bond to be cleaved between sites -1 and +1. Subsites -1 and -2 are also called 'product sites', since they harbour the cellobiose moiety that is cleaved off from the non-reducing end of a cellulose chain during processive action. In the ligand bound *Tf*Cel6B structure model, contiguous electron density for glucose residues was observed in the substrate binding tunnel from subsite -2 to +4. Guided by the ligand Fo-Fc difference map, cellohexaose and cellotetraose were modeled into molecule A and B of the asymmetric unit, respectively.

Vuong and Wilson have confirmed that D274 is the catalytic acid (Vuong & Wilson, 2009a). In all the *Tf*Cel6B structures, D274 takes on an inactive conformation, not pointing towards the glycosidic oxygen, and making a hydrogen bond to D226. This hydrogen bond must be broken because D274 can not protonate the glycosidic oxygen and D226 can not accept a proton as long as they are sharing a proton with each other. D226 adopts different conformations in the apo and ligand structures, respectively. In the ligand structure, D226 flips towards the catalytic center and points to the +1 subsite. The active center loop in *Tf*Cel6B (residue 229-233) has an open conformation in the apo structure and a closed conformation in the ligand bound structure, which is consistent with observed conformation changes in *H. jecorina* Cel6A (Zou *et al.*, 1999).

As proposed by the GH6 Grotthus mechanism, in *Tf*Cel6B, D274 act as catalytic acid to perform protonation of the glycosidic oxygen between the glucose units. A water molecule is suitably positioned to make a nucleophilic attack on the anomeric carbon, held in place by S232 and the backbone carbonyl oxygen of D497. The water donates a proton to a neighbor water molecule, which in turn gives a proton to D226. Thus the D226 residue acts indirectly as the catalytic base through a hydrogen bonded proton transfer network (Figure 11).

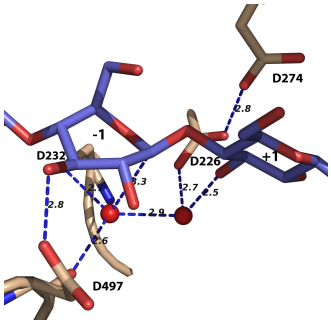


Figure 11. The active site in *Tf*Cel6B ligand structure (PDBcode: 4AVN). Blue dashed lines indicate possible hydrogen bonds with distances given in Angstroms. Red spheres indicate water molecules involved in the catalytic mechanism.

2.2.4 Bacterial *T. fusca* cellulases have higher structural diversity

An alignment of 160 GH6 enzyme sequences from both bacteria and fungi showed that the presence of the extra sections are unique to a subgroup of bacterial sequences, which presumably all are processive CBHs. But not all six sections are present in all bacterial enzymes. Section III-VI exist in cellobiohydrolases from a wide range of species while section II only appears in some species from phyla/genera *Actinomycetales*. Section I (tunnel exit loop) mainly exists in *Actinomycetales* and some *Gammaproteobacteria*.

When superimposing *Tf*Cel6B and other known GH6 structures, it is obvious that *Tf*Cel6B has a longer and more closed substrate binding tunnel. On the other hand, the GH6 endoglucanase *Tf*Cel6A from *T. fusca* has a shorter and more open substrate binding cleft than its fungal counterparts. Thus it suggests that family GH6 displays a larger structure diversity within bacteria than fungi. (Figure 12)

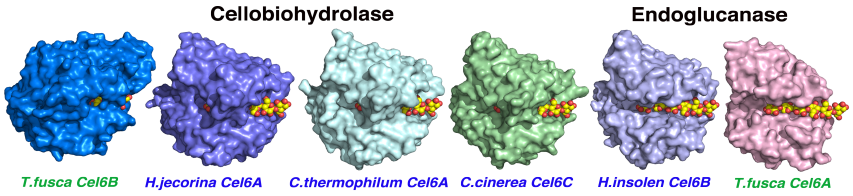


Figure 12. Surface model comparison of GH6 enzymes with cellohexaose in the active site shown as spacefilling sphere model. Names in green color are bacterial enzymes (*T. fusca*); Names in blue color are fungal. The first four GH6s are cellobiohydrolases. The last two are endoglucanases.

2.3 *Phanerochaete chrysosporium* GH61D function and structure (paper III, paper IV)

2.3.1 Introduction

Previously, five LPMO structures had been published from five different fungal GH61s: *H. jecorina* GH61B (PDB code: 2VTC) (Karkehabadi *et al.*, 2008), *Thielavia terrestris* GH61E (3EII, 3EJA) (Harris *et al.*, 2010), *Thermoascus aurantiacus* GH61A (2YET, 3ZUD) (Quinlan *et al.*, 2011), and *Neurospora crassa* PMO-2 and PMO-3 (4EIR and 4EIS, respectively) (Li *et al.*, 2012). All these structures and most of the previous progress on GH61 mechanism elucidation are with enzymes from ascomycete fungi. However, in nature most of the wood decomposition is done by basidiomycete fungi (Fernandez-Fueyo *et al.*, 2012; Eastwood *et al.*, 2011), which are broadly divided in brown rot and white rot fungi. Multiple GH61 genes have been found in genomes of both types (Fernandez-Fueyo *et al.*, 2012; Eastwood *et al.*, 2011; Martinez *et al.*, 2004), and the number of GH61 genes is generally larger in basidiomycetes than in ascomycetes, and it appears to be larger in white rot than in brown rot fungi (Eastwood *et al.*, 2011). It is thus of significant interest to study LPMO enzymes from basidiomycete fungi. *Phanerochaete chrysosporium* is one of the most extensively studied of white rot fungi, and its genome was the first basidomycete to be sequenced. Up to seventeen putative GH61 encoding genes (*Pch*GH61s) were initially identified (Martinez *et al.*, 2004).

The white rot fungus *P. chrysosporium* is claimed to be able to degrade all components of wood including the lignin portion (Martinez *et al.*, 2004). It secrets different extracellular glycoside hydrolases, carbohydrate esterases, and various oxidative enzymes that cooperate in wood depolymerization (Broda *et al.*, 1994). One redox enzyme of particular interest is cellobiose dehydrogenase (CDH), a 89 kDa extracellular bimodular flavocytochrome, which consists of a N-terminal haem domain of 190 residues, and a C-terminal flavin domain of 539 residues connected by a 25 residue linker peptide (Hallberg *et al.*, 2000). The crystal structures of both domains have been determined by Hallberg and Divne and coworkers (Hallberg *et al.*, 2002; Hallberg *et al.*, 2000). CDH oxidizes cellobiose and longer cellooligosaccharides by catalyzing a two electron oxidation at the anomeric C1 position to yield the corresponding 1,5-lactone, which is then spontaneously hydrolyzed to the open aldonic acid form (Henriksson *et al.*, 2000). A large number of compounds have been found to act as electron acceptors for CDH, both one-electron and two-electron

acceptors (Henriksson *et al.*, 2000). Recent studies have brought new light by demonstrating that CDH enzymes may act as electron donor to LPMOs. For example, *Thermoascus aurantiacus* GH61 (*Tau*GH61A) together with *Humicola insolens* CDH, two LPMOs and CDH from *Neurospora crassa*, as well as GH61 and CDH from *Thielavia terrestris*, were shown to synergistically degrade cellulose yielding oxidized oligosaccharides as product (Langston *et al.*, 2011; Phillips *et al.*, 2011). So far there is no report on CDH-LPMO interaction using *Phanerochaete chrysosporium* enzymes.

So far there is no report on CDH-LPMO interaction with *P. chrysosporium* enzymes, but peptide mapping and sequencing by mass spectrometry of culture filtrates have revealed the presence of both CDH and GH61 proteins in the secretome of *Phanerochaete chrysosporium* when grown on cellulose-containing media (Hori *et al.*, 2011). In this section, *P. chrysosporium* GH61D (Protein ID: Phchr1|4691) was identified, expressed and characterized. The crystal structure of *Pch*GH61D was determined and it is the first LPMO structure from a basidiomycete fungus. An initial study on the interaction between *Pch*GH61D and *Pch*CDH was also performed.

2.3.2 *Pch*GH61D expression and characterization

In the research group of M. Samejima and K. Igarashi at University of Tokyo, that we collaborate with, the *Pch*GH61D protein (ID4691 in the *P. chrysosporium* v2.0 genome database) was identified as present in much larger quantity when xylan was added to the cellulose-containing medium. The gene was cloned and inserted in *P. pastoris* for heterologous recombinant expression by T. Ishida who brought the expression strain to Uppsala for further studies.

Since the N-terminal amino group of GH61 proteins is involved in metal binding at the active site, it is essential that the signal peptide is correctly cleaved off. However, heterogeneous signal peptide cleavage is a well known problem with the *P. pastoris* expression system. The problem was initially solved by introduction of an enterokinase cleavage site between the alpha-factor signal peptide and the codon for His1 of the mature protein, and treating the secreted protein with enterokinase enzyme. To express protein for structure determination, a new construct was used with human albumin signal peptide that appeared to be correctly cleaved off without the need for external entrokinase treatment. After purification, the *Pch*GH61D was enzymatically deglycosylated using endoglycosidase H (Endo H). Figure 13 shows the SDS-PAGE gel of expressed and purified *Pch*GH61D enzyme.

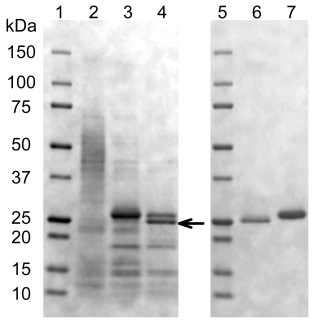


Figure 13. SDS-PAGE analysis of recombinantly expressed *Pch*GH61D. Lanes: 1 & 5, marker (Precision Plus Protein™ Dual Color Standards; BioRad); lane 2, culture medium of the *P. pastoris* strain; lane 3, EndoH treated culture medium; lane 4, EndoH treated culture medium after enterokinase treatment. Lane 6 & 7, protein bands with MW of 25 and 27 kDa, respectively. The protein band used for N-terminal sequencing is indicated by an arrow.

Pure and correctly processed *Pch*GH61D was provided to the research group of Professor Vincent Eijsink, University of Life Sciences (UMB), Norway, for activity measurements. The results proved that *Pch*GH61D is indeed an LPMO, i.e. a metal dependent enzyme that is able to cleave cellulose by an oxidative mechanism. It was shown to oxidize primarily at the C1 anomeric carbon of glucose units, no C4 oxidized sugars could be detected (Paper III). Of different tested metal ions, Cu^{2+} gave the highest activity followed by Mn^{2+}. The activity was negligible with Co^{2+}, Zn^{2+}, Fe^{3+}, Ca^{2+}, Mg^{2+} or without metal bound. *Pch*GH61D showed synergistic effect on cellulose degradation together with a commercial *H. jecorina* enzyme cocktail (Celluclast). Furthermore, we have shown that *Pch*GH61D also is able to cleave glycosidic bonds in wood lignocellulose (steam exploded spruce) and produces C1 oxidized cello-oligosaccharides (Figure 14; Paper IV).

2.3.3 Crystal structure of *Pch*GH61D

In order to make sure that all enzyme molecules have copper bound at the active site, the *Pch*GH61D enzyme was incubated with 1 mM $CuSO_4$ and free $CuSO_4$ was removed prior to crystallization set up. Crystals were obtained already in the first crystallization screening (JCSG+ Suite sparse matrix screen, Qiagen) but only in one crystallization condition (2.1 M DL-malic acid, pH 7.0). Single crystals from there were directly used for X-ray diffraction data collection and structure determination. *Pch*GH61D was crystallized in space group *C*2 with unit-cell parameters of a=149.3 Å, b=37.5 Å, c=79.8 Å and with a β angle of 117.4°. There are 2 molecules in the asymmetric unit. The structure determination was straightforward by molecular replacement using a homology model of *Pch*GH61D that I had previously made for Paper III. The final structure, at 1.75 Å resolution, exhibits crystallographic R and R_{free} values

of 18.6% and 22.3%. All the 217 amino acids of both protein molecules could be included in the final structure model.

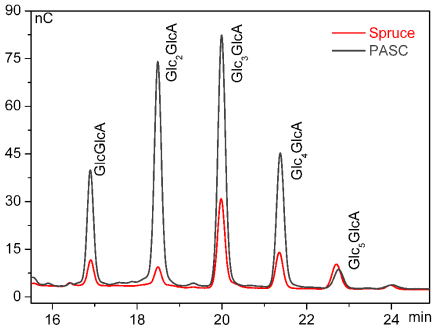


Figure 14. High-Performance Anion-Exchange Chromatography (HPAEC) chromatogram of soluble aldonic acids (DP 2-6) obtained upon incubation of 0.1% (w/v) PASC or 0.5% (w/v) steam exploded spruce with *Pch*GH61D (34 µg/mL) in 25 mM sodium acetate buffer, pH 5.3 including 1.5 mM ascorbate at 50 °C for 20 hours.

The structure of *Pch*GH61D is the first LPMO structure from a basidiomycete fungus. The overall fold is a β-sandwich consisting of two β-sheets which pack onto each other to form the core of the protein. The catalytic center of the enzyme with copper bound is positioned on a flat surface on one side of the β-sandwich. Three extended loops (LC, LS, L2 loops, shown in Figure 15A) are involved in shaping the flat potential substrate binding surface. The L2 loop region (residues 17-57) includes two short η-helices. The LC and LS loops appear to be flexible, especially in some regions (Figure 15 pink) the B factor values are much higher than the average for the protein.

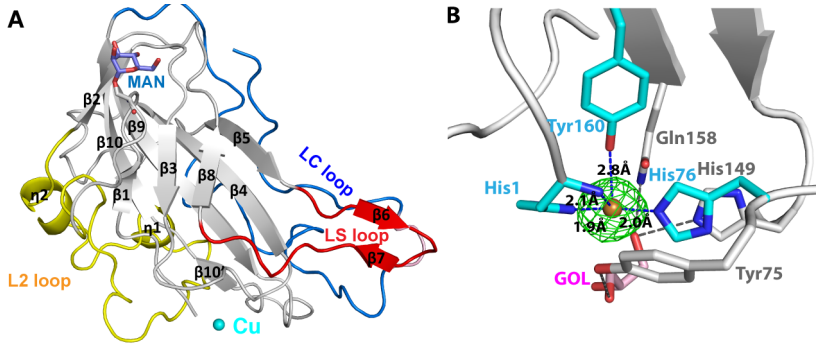


*Figure 15. Pch*GH61D crystal structure (PDB code 4B5Q). (A) Cartoon representation of the *Pch*GH61D structure with the bound copper atom, depicted as a sphere in cyan. The glycosylated residue Ser11 and the attached *O*-linked mannose residue are shown in stick representation in grey and slate blue, respectively. (B) A close up view of the *Pch*GH61D in the vicinity of the copper binding site. The green F_o-F_c map around the copper atom is contoured at 0.41 e/Å^3 (3σ). Cyan colored residues are coordinated to the copper atom. A glycerol molecule was modeled underneath the copper atom at the active site, coloured in pink (denoted GOL).

There are two potential N-glycosylation sites (Asn173 and Asn203) in the sequence (Figure 16, green frame), but no remaining GlcNAc residues are visible in the electron density map at either site. However, one O-linked mannose residue was evident at Ser11 of the A molecule. In Figure 15 B selected residues are shown in the vicinity of copper binding site that might play important roles in catalysis and substrate binding.

In the structure data, clear density of a copper ion was seen at the metal binding site of the catalytic center in both A and B molecules. Copper was modeled with full occupancy and fit the density with B factor values of 15 $Å^2$ (average B factor for the protein is 16.7 $Å^2$). The copper binding site is a type II copper center (Figure 15B), which generally exhibits a hexacoordination. In this type II copper center geometry, a square planar coordination is created by N or N/O atoms (Klinman, 1996; Mccracken *et al.*, 1987). In the *Pch*GH61D structure, square coordination is provided by the main-chain amide group (2.1 Å), Nδ (1.9 Å) of His1, and Nε of His76 (2.0 Å), whereas there is no ligand at the fourth coordination position. In *Hje*GH61B (PDBcode: 2VTC) the corresponding position is occupied by a water molecule. The hydroxyl group of Tyr160 (2.8 Å) occupies one of the axial positions while the other axial position is empty in the hexacoordination geometry. Protein atoms occupy four coordination positions leaving two coordination sites for ligand binding, which are however blocked in the *Pch*GH61D structure by the bound glycerol molecule.

2.3.4 Structure and sequence comparison of *Pch*GH61D

All known LPMO structures were superimposed with the *Pch*GH61D structure in MacPymol program (Figure 17A). The three loop regions (L2, LC and LS loop) that comprise the putative substrate-binding surfaces vary among different LPMOs. Generally, LC and LS loop regions are more extended in fungal LPMOs than bacterial LPMOs. The L2 loop varies within fungal LPMOs. *Pch*GH61D, *Tte*GH61E, *Ncr*PMO-2 have shorter L2 loops compared with *Hje*GH61B, *Tau*GH61A, and *Ncr*PMO-3. However, within GH61 enzymes, there are similarities in the exposure of aromatic residues on the putative substrate binding surface (Figure 17B/C). Our computational study indeed indicates that the three loops and aromatic residues on the binding surface may play important roles in substrate binding.

Nine catalytic domain sequences of the *Pch*GH61s were incorporated into the sequence alignment of known GH61 structures using MAFFT program to

obtain an overall view of the structure-based alignment of the *Pch*GH61s (Figure 16). The *P. chrysosporium* GH61s exhibit significant sequence variability. There is no major length variation in the LS or LC loop regions, but three *Pch*GH61s (Phchr1|129325, 121193, and 10320) have extended L2 loops similar to *Hje*GH61B, *Tau*GH61A, and *Ncr*PMO-3. Residues corresponding to Tyr75 of the *Pch*GH61D vary considerably within *Pch*GH61s, with aspartate, proline, glycine, alanine, or asparagine at this position.

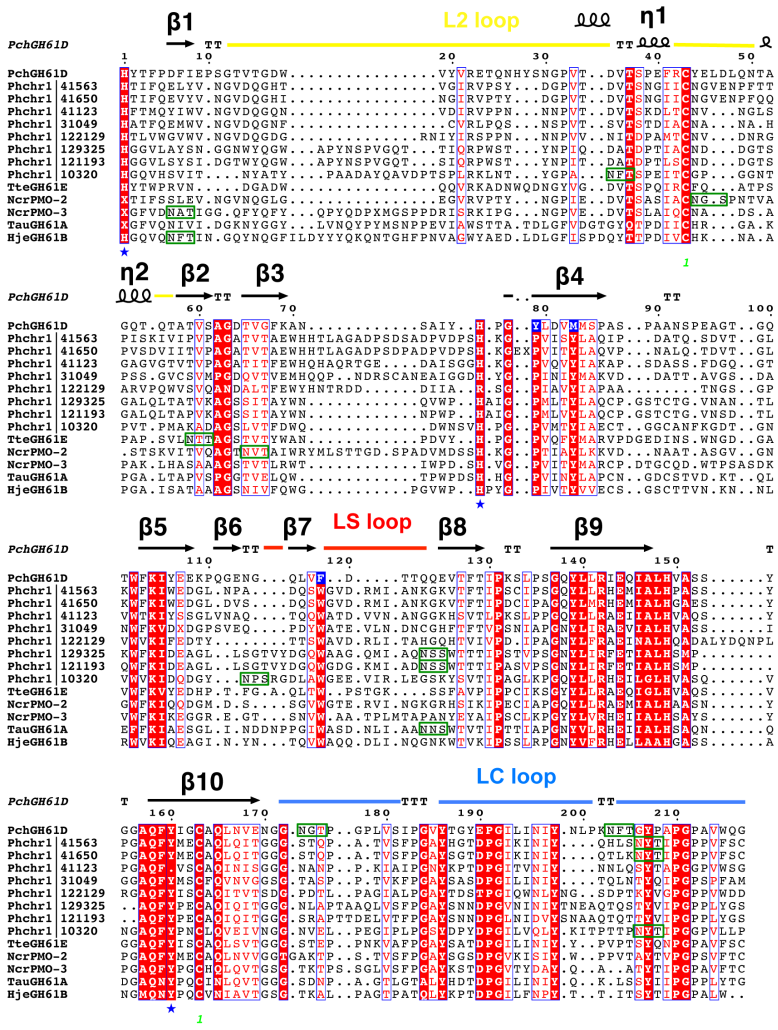


Figure 16. Alignment of the catalytic domains of *P. chrysosporium* GH61s and GH61 enzymes with known structures (*Tte*GH61E/3EJA; *Ncr*PMO-2/4EIR; *Ncr*PMO-3/4EIS; *Tau*GH61A/3ZUD; *Hje*GH61B/2VTC). Residue numbering and secondary structure assignments shown above the sequences refer to *Pch*GH61D. Residues in white and red denote residues with a global similarity score greater than 0.7. Residues shown with a red background are highly conserved (>85%), whereas residues with a blue background are residues at highly conserved positions that are unique for *Pch*GH61D. The three active site residues are marked with blue stars. The green '1' label denotes the disulfide. The L2, LS and LC loop regions are labeled with yellow, red and blue lines, respectively. Green frames indicate potential N-glycosylation sites.

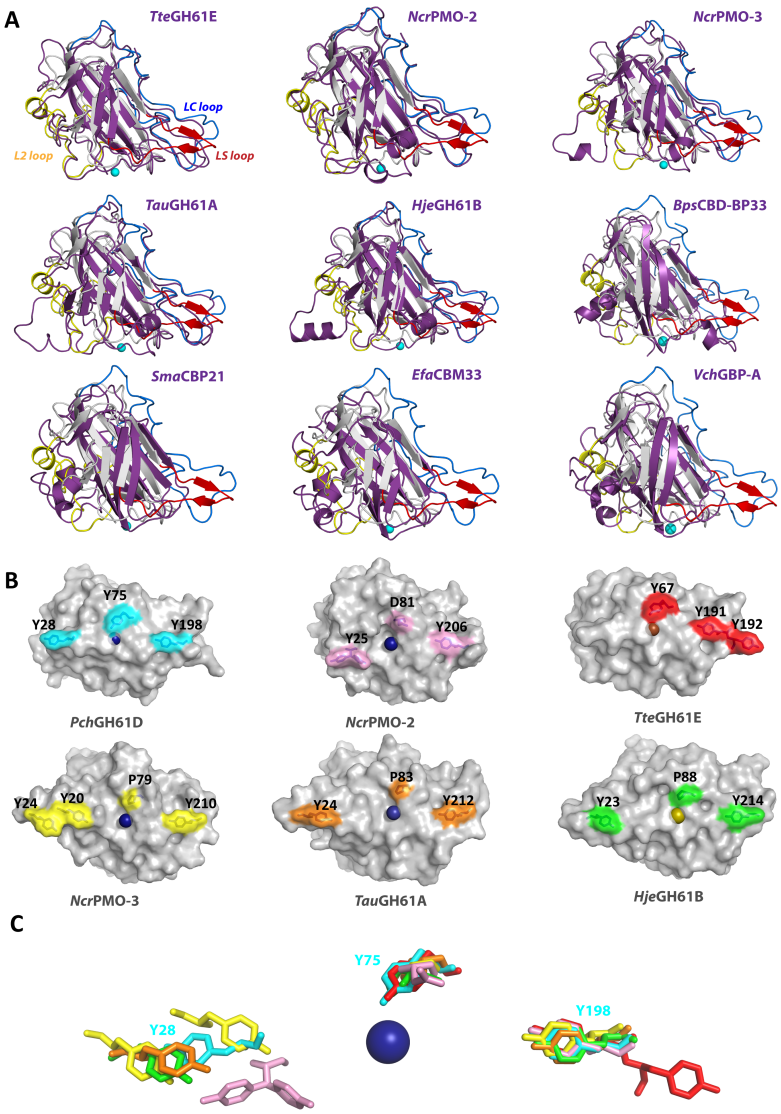


Figure 17. Structural comparison of LPMOs. **(A)** Superimposed structures of *Pch*GH61D, in gray, with other LPMOs, in purple. The first six panels are superpositions with GH61s and the last three are bacterial CBM33s (*Sma*CBP21, *Efa*CBM33, *Vch*GBP-A). (B) Aromatic residues highlighted in contrasting colors on the molecular surface of GH61s. (C) Superposition of the aromatic residues shown in stick model with the same colors as in panel B.

2.3.5 *Pch*GH61D-cellulose interaction study

Molecular dynamic simulations (MD) were conducted by Gregg Beckham and co-workers at the National Renewable Energy Laboratory (NREL), Golden, CO, USA, using our *Pch*GH61D crystal structure as starting model. The enzyme was placed on the hydrophobic face of cellulose Iβ in analogy with the proposed binding and orientation of CBM1 on crystalline cellulose. MD simulation was conducted for 100 ns to examine the enzyme-substrate interactions, and identify potential important residues for binding, or any conformational changes that occur upon interaction with cellulose.

Figure 18 shows the root mean square fluctuations (RMSF) per residue, and the B-factors of the *Pch*GH61D crystal structure for comparison. RMSF results indicate that the largest fluctuations occur at the LC (residues 170-217) and LS (residues 109-124) loops. The B-factors also suggest that the LC and LS loops are the most flexible, although they may not be strictly comparable due to the differences between chemical environments of the crystal and simulated *Pch*GH61D enzymes. Figure 19 shows the position of the copper atom at the active site. The distance from copper to the hydrogen atom on the C1 carbon fluctuates near 5.0 Å during the MD simulation. We do not know yet the exact oxygen binding position on the copper ion, but it seems that it will be close enough to attack the C1 carbon.

All the known GH61 structures show a distribution of aromatic or polar residues on the flat surface (Figure 17 B,C) similar to Type A CBMs, which have been reported to bind to the hydrophobic face of cellulose I (Boraston *et al.*, 2004; Lehtio *et al.*, 2003). In our simulation experiment, three tyrosine residues (Tyr28, Tyr75, Tyr198) were specially examined (Figure 20). Tyr28, and Tyr198 align over the same chain, and their positions are quite stable during the simulation. Tyr75 is bound to the adjacent cellulose chain on the edge of the crystal upon docking, and it retains this conformation over 100 ns. The active site position also remains stable and directly above a glycosidic bond in the middle chain, near the proposed site of attack on the cellulose.

2.3.6 *Pch*GH61D-CDH interaction study (Unpublished)

In order to investigate whether the CDH and GH61D enzymes from *P. chryososporium* do interact with each other, preliminary binding experiments were tried using the surface plasmon resonance technique on a Biacore X-100 system (GE healthcare). *Pch*GH61D was immobilized on a CM5 sensor chip and various concentrations of purified CDH (17.8-90 uM) was injected. The recorded sensorgrams are shown in Figure 21 (left). The largest response units were plotted against the various concentration of *Pch*CDH to show the binding affinity (shown in Figure 21 (right)). The results indicate that there is indeed

specific binding between *Pch*CDH and *Pch*GH61D. The K_D value for CDH binding to the immobilized *Pch*GH61D was estimated at 80 uM by the BiaEvaluation software (GE healthcare). However this value is not the true dissociation constant since the binding reactions did not reach equilibrium. Furthermore, the dissociation of CDH was very slow and several conditions were tried for regeneration. It is uncertain if the immobilized *Pch*GH61D was unaffected by the harsh conditions that were required for complete removal of CDH (e.g. Glycine buffer, pH 10.2).

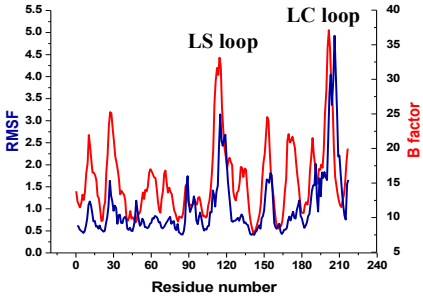


Figure 18. RMSF (Å) of *Pch*GH61D over the 250 ns MD simulation (blue) and B-factors (Å^2) of the *Pch*GH61D from the crystal structure(red).

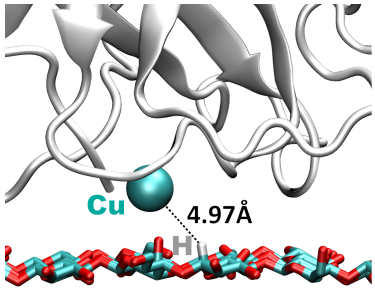


Figure 19. Simulation of *Pch*GH61D on the hydrophobic surface of cellulose. The copper (shown as cyan sphere) fluctuates at approximately 5 Å from the hydrogen atom on the C1 carbon during the MD simulation.

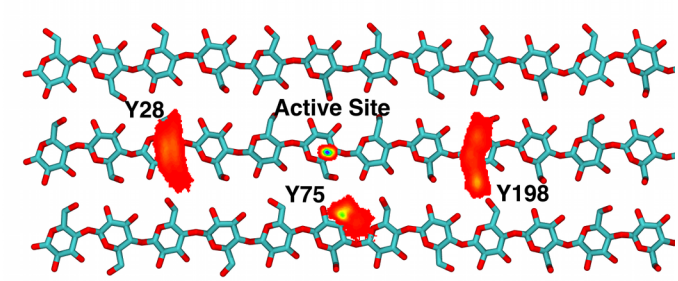


Figure 20. Histogram of the tyrosine residues (Tyr28, Tyr75, and Tyr198) and the *Pch*GH61D active site position on the cellulose surface. The color code denotes the position on a 0.1 Å x 0.1 Å grid on the cellulose surface from red at lower density to blue at the highest density

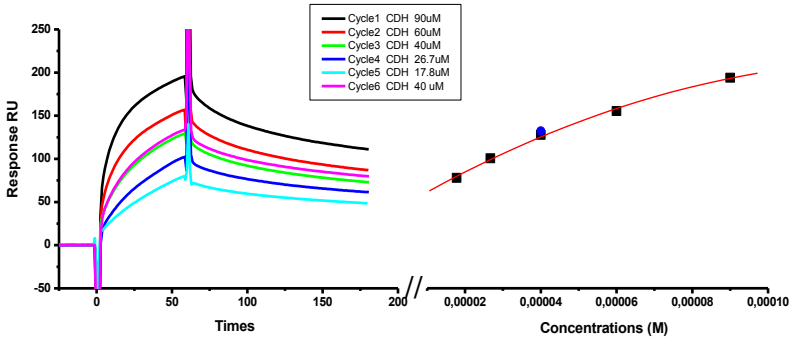


*Figure 21. Pch*CDH binds to *Pch*GH61D. (Left:) Sensorgram from BIAcore binding assay. Contact time: 60s; dissociation time: 120s. The SPR (surface plasmon resonance) signal (expressed in resonance units) is already subtracted from the response obtained over a control surface, and normalized to a baseline of 0 resonance units. (Right:) Binding affinity profile of *Pch*CDH and *Pch*GH61D. Note that the binding did not reach equilibrium, but was measured as the change in RU at the highest point in each sensorgram. Thus, the apparent Kd value estimated by the BioEvaluation software is not the true equilibrium constant for dissociation.

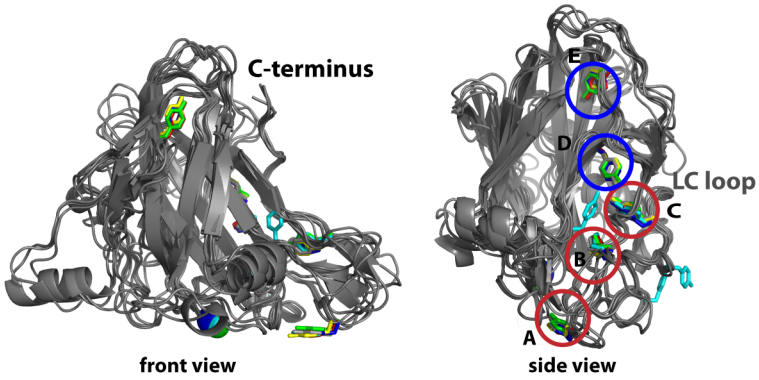


Figure 22. Superposition of GH61 structures showing tyrosine residues that might be involved in electron transfer between the active site copper ion and a putative binding site for CDH near the C-terminus. Cluster A, B and C (red circles) are conserved in *Pch*GH61D, but not D and E (blue circles).

The interaction between *Pch*CDH and *Pch*GH61D indicates that CDH may indeed act as electron donor to LPMOs also in this organism. If CDH binds to the proposed site near the C terminus on the GH61 molecule, and the copper atom is embedded at the enzyme-cellulose interface, there is a need for an electron transfer pathway within the protein. Electron transfer in proteins is facilitated by a hydrophobic environment (Tan *et al.*, 2004). When superimposing the structurally known GH61s, five clusters of aromatic

residues (tyrosines) were found to be conserved in the LC loop region near the place where CDH was docked by (Li *et al.*, 2012), as shown in Figure 22. The A, B and C clusters are conserved in *Pch*GH61D but not D and E. Instead, there is another tyrosine residue (colored cyan in Figure X) located near the C cluster. Intuitively, the aromatic residues seem to form a hydrophobic stair that may act as a path for electron transfer between a bound CDH molecule and the copper atom at the active site of the GH61 enzyme.

3 Conclusions and future perspectives

3.1 Fluorogenic substrates for GH6 enzymes

In paper I, we have figured out that MUF-G2 is poorly hydrolyzed by GH6 enzymes because it adopts a nonproductive pose at the catalytic center (subsite -2 to +1. Rationally designed modifications of the MUF group successfully restrict the conformational freedom in the confined space of the active site as the substrate dockings predicted. The 6-chlorine substitution is beneficial by forcing the umbelliferyl group to turn 'upside down' to a favorable orientation for hydrolysis, and the introduction of the 4-phenyl group helps to shift binding preference towards the productive position (Figure 6). However, the rate of fluorophore release from the modified substrates is still low with GH6 enzymes. Furthermore the fluorescence intensity of the PhUF and ClPhUF became weaker due to the insertion of phenyl group. Thus, there is a potential for further improvement of fluorogenic substrates for GH6 enzymes. By comparing the ligand structures of GH6, especially at -1 subsite, we speculate that hydrolysis will be faster if the ϕ_2 angle of the scissile bond is rotated further away from the minus-synclinal position than observed in the ClPhuF-G2 complex. Nevertheless, our improved substrates can be useful already today for activity assays, screening and enzyme kinetics as demonstrated in this study, and also in other potential applications. For example, ClF3MUF-G2 that was readily hydrolyzed but showed limited solubility may be useful for zymogram staining where low solubility can be an advantage.

3.2 *Thermobifida fusca* Cel6B

In paper II, we have determined the structure of the catalytic domain of the bacterial cellobiohydrolase Cel6B from *T. fusca*. Insertions in six loop sections give a longer substrate binding tunnel and one of the loops is blocking the exit

from the tunnel, which suggest higher processivity and lower endolytic activity than with fungal GH6 cellobiohydrolases. Perhaps *Tf*Cel6B is even a 'true' exo-cellulase. One striking observation is that in the ligand complex, the cellobiose moiety in the product sites, -2/-1, is completely enclosed, raising the intriguing question how product is expelled from the enzyme during processive action. Two routes can be envisioned, either by opening of the exit loop, or by opening of the active center loop. The latter loop must in any case open subsequently to allow the end of the remaining cellulose chain to proceed into the product sites for another cycle of processive action. Molecular dynamic simulations may prove useful here to explore possible influence of the extended loops on substrate binding and product expulsion etc. It is also desirable to complement these structural studies with biochemical measurements of processivity, endo activity and other important enzyme properties and compare with other GH6 cellobiohydrolases.

3.3 *Phanerochaete chrysosporium* GH61D

Since the discovery in 2010 of the novel oxidative mechanism of LPMOs for decomposition of recalcitrant polysaccharides (Vaaje-Kolstad, G, 2010), the research on LPMOs has boosted in the last couple of years. Our research here has so far resulted in two publications, both on *Pch*GH61D, a major LPMO from the wood degrading white-rot fungus *Phanerochaete chrysosporium*. **In paper III,** we have cloned and overexpressed *Pch*GH61D in *Pichia pastoris*, and obtained correctly processed, pure enzyme. We showed that *Pch*GH61D is a copper-dependent LPMO with activity on Avicel, filter paper and phosphoric acid swollen cellulose, which oxidizes at the C1 carbon. **In paper IV**, we solved the structure of *Pch*GH61D, the first LPMO structure from a basidiomycete fungus. Structure comparison and molecular dynamic simulations suggest that three loops and a series of aromatic and polar residues near the binding surface are important for substrate recognition and binding. In order to understand why *Pch*GH61D is specific for C1 oxidation, further studies are needed on the catalytic mechanism and on LPMOs with different oxidation preferences (C1, C4 or possible C6). Going forward in the burgeoning field of LPMO biochemistry, it is likely that a combination of structural, biophysical, and computational studies, such as that presented here, will be necessary to fully understand the diversity of LPMO structures as well as its functional consequences.

References

Aachmann, F.L., Sorlie, M., Skjak-Braek, G., Eijsink, V.G.H. & Vaaje-Kolstad, G. (2012). NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proc. Nat. Acad. Sci. U.S.A.* 109(46), 18779-18784.

Beeson, W.T., Phillips, C.M., Cate, J.H.D. & Marletta, M.A. (2012). Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J. Am. Chem. Soc.* 134(2), 890-892.

Bey, M., Zhou, S., Poidevin, L., Henrissat, B., Coutinho, P.M., Berrin, J.G. & Sigoillot, J.C. (2012). Comparison of two lytic polysaccharide monooxygenases (GH61) from Podospora anserina reveals differences upon cello-oligosaccharides oxidation. *Appl. Environ. Microbiol.* 79 (2): 488-96.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769-781.

Broda, P., Birch, P., Brooks, P., Copapatino, J.L., Sinnott, M.L., Tempelaars, C., Wang, Q., Wyatt, A. & Sims, P. (1994). *Phanerochaete chrysosporium* and its natural substrate. *FEMS Microbiol. Rev.* 13(2-3), 189-196.

Cramer, C.J., Truhlar, D.G. & French, A.D. (1997). Exo-anomeric effects on energies and geometries of different conformations of glucose and related systems in the gas phase and aqueous solution. *Carbohydr. Res.* 298, 1-14.

Davies, G.J., Brzozowski, A.M., Dauter, M., Varrot, A. & Schulein, M. (2000). Structure and function of Humicola insolens family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 angstrom resolution. *Biochem. J.* 348, 201-207.

Eastwood, D.C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., Asiegbu, F.O., Baker, S.E., Barry, K., Bendiksby, M., Blumentritt, M., Coutinho, P.M., Cullen, D., de Vries, R.P., Gathman, A., Goodell, B., Henrissat, B., Ihrmark, K., Kauserud, H., Kohler, A., LaButti, K., Lapidus, A., Lavin, J.L., Lee, Y.H., Lindquist, E., Lilly, W., Lucas, S., Morin, E., Murat, C., Oguiza, J.A., Park, J., Pisabarro, A.G., Riley, R., Rosling, A., Salamov, A., Schmidt, O., Schmutz, J., Skrede, I., Stenlid, J., Wiebenga, A., Xie, X.F., Kues, U., Hibbett, D.S., Hoffmeister, D., Hogberg, N., Martin, F., Grigoriev, I.V. & Watkinson, S.C. (2011). The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333(6043), 762-765.

Fernandez-Fueyo, E., Ruiz-Duenas, F.J., Ferreira, P., Floudas, D., Hibbett, D.S., Canessa, P., Larrondo, L.F., James, T.Y., Seelenfreund, D., Lobos, S., Polanco, R., Tello, M., Honda, Y., Watanabe, T., San, R.J., Kubicek, C.P., Schmoll, M., Gaskell, J., Hammel, K.E., St John, F.J., Vanden Wymelenberg, A., Sabat, G., BonDurant, S.S., Syed, K., Yadav, J.S., Doddapaneni, H., Subramanian, V., Lavin, J.L., Oguiza, J.A., Perez, G., Pisabarro, A.G., Ramirez, L., Santoyo, F., Master, E., Coutinho, P.M.,

Henrissat, B., Lombard, V., Magnuson, J.K., Kues, U., Hori, C., Igarashi, K., Samejima, M., Held, B.W., Barry, K.W., LaButti, K.M., Lapidus, A., Lindquist, E.A., Lucas, S.M., Riley, R., Salamov, A.A., Hoffmeister, D., Schwenk, D., Hadar, Y., Yarden, O., de Vries, R.P., Wiebenga, A., Stenlid, J., Eastwood, D., Grigoriev, I.V., Berka, R.M., Blanchette, R.A., Kersten, P., Martinez, A.T., Vicuna, R. & Cullen, D. (2012). Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium* provide insight into selective ligninolysis. *Proc. Nat. Acad. Sci. U.S.A.* 109(14), 5458-5463.

Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J.M., Yao, J. & Ward, M. (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* 278(34), 31988-31997.

Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunaes, A.C., Stenstrom, Y., MacKenzie, A., Sorlie, M., Horn, S.J. & Eijsink, V.G.H. (2011). Cleavage of cellulose by a CBM33 protein. *Protein Sci.* 20(9), 1479-1483.

Hallberg, B.M., Bergfors, T., Backbro, K., Pettersson, G., Henriksson, G. & Divne, C. (2000). A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Struct. Fold. Des.* 8(1), 79-88.

Hallberg, B.M., Henriksson, G., Pettersson, G. & Divne, C. (2002). Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J. Mol. Biol.* 315(3), 421-434.

Harris, P.V., Welner, D., McFarland, K.C., Re, E., Navarro Poulsen, J.C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. & Lo Leggio, L. (2010). Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49, 3305-3316.

Heightman, T.D. & Vasella, A.T. (1999). Recent Insights into Inhibition, structure, and mechanism of configuration-retaining glycosidases. *Angew. Chem. Int. Ed. Engl.* 38, 751–770.

Henriksson, G., Johansson, G. & Pettersson, G. (2000). A critical review of cellobiose dehydrogenases. *J. Biotechnol.* 78(2), 93-113.

Himmel, M.E. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production (vol 315, pg 804, 2007). *Science* 316(5827), 982-982.

Hori, C., Igarashi, K., Katayama, A. & Samejima, M. (2011). Effects of xylan and starch on secretome of the basidiomycete *Phanerochaete chrysosporium* grown on cellulose. *FEMS Microbiol. Lett.* 321(1), 14-23.

Horn, S.J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V.G.H. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels.* 5, 45.

Irwin, D.C., Spezio, M., Walker, L.P. & Wilson, D.B. (1993). Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.* 42(8), 1002-13.

Johnson, G.P., Petersen, L., French, A.D. & Reilly, P.J. (2009). Twisting of glycosidic bonds by hydrolases. *Carbohydr. Res.* 344(16), 2157-2166.

Jung, H., Wilson, D.B. & Walker, L.P. (2003). Binding and reversibility of *Thermobifida fusca* Cel5A, Cel6B, and Cel48A and their respective catalytic domains to bacterial microcrystalline cellulose. *Biotechnol. Bioeng.* 84(2), 151-9.

Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C. & Sandgren, M. (2008). The structure of a glycoside hydrolase family 61 member, Cel61B from the *Hypocrea jecorina*. *J. Mol. Biol* 383, 144-154.

Klinman, J.P. (1996). Mechanisms whereby mononuclear copper proteins functionalize organic substrates. *Chem. Rev.* 96(7), 2541-2561.

Knowles, J.K.C., Lentovaara, P., Murray, M. & Sinnott, M.L. (1988). Stereochemical course of the action of the cellobioside hydrolase I and II of *Trichoderma reesei*. *J. Chem. Soc., Chem. Commun.* (21), 1401-1402.

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., Stahlberg, J. & Jones, T.A. (2002). The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. *Eur. J. Biochem.* 124(34), 10015-10024.

Langston, J.A., Brown, K., Xu, F., Borch, K., Garner, A. & Sweeney, M.D. (2012). Cloning, expression, and characterization of a cellobiose dehydrogenase from *Thielavia terrestris* induced under cellulose growth conditions. *Biochim. Biophys. Acta* 1824(6), 802-812.

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(19), 7007-7015.

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(39), 12915-12921.

Lehtio, 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. Nat. Acad. Sci. U.S.A.* 100(2), 484-489.

Li, X., Beeson, W.T., Phillips, C.M., Marletta, M.A. & Cate, J.H.D. (2012). Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* 20(6), 1051-1061.

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, *Cc*Cel6C, a cellulase constitutively produced by *Coprinopsis cinerea*. *FEBS J.* 277(6), 1532-1542.

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(6), 695-700.

Mccracken, J., Peisach, J. & Dooley, D.M. (1987). Cu(II) coordination chemistry of amine oxidases. Pulsed EPR studies of histidine imidazole, water, and exogenous ligand coordination. *J. Am. Chem. Soc.* 109(13), 4064-4072.

McPherson, A.J. (1982). *Preparation and analysis of protein crystals*. New York: John Wiley and Sons.

Merino, S.T. & Cherry, J. (2007). Progress and challenges in enzyme development for Biomass utilization. *Biofuels* 108, 95-120.

Moser, F., Irwin, D., Chen, S.L. & Wilson, D.B. (2008). Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng.* 100(6), 1066-1077.

Naik, S.N., Goud, V.V., Rout, P.K. & Dalai, A.K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renew Sust Energ Rev*.14(2), 578-597.

Nerinckx, W., Desmet, T., Piens, K. & Claeyssens, M. (2005). An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy. *FEBS Lett.* 579(2), 302-312.

Phillips, C.M., Beeson, W.T., Cate, J.H. & Marletta, M.A. (2011). Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol.* 6(12), 1399-1406.

Quinlan, R.J., Sweeney, M.D., Lo Leggio, L., Otten, H., Poulsen, J.-C.N., Johansen, K.S., Krogh, K.B.R.M., Jorgensen, C.I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C.P., Dupree, P., Xu, F., Davies, G.J. & Walton, P.H. (2011). Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc.Natl.Acad.Sci.* 108(37), 15079–15084.

Reese, E.T. (1956). Enzymatic hydrolysis of cellulose. *Appl. Microbiol.* 4(1), 39-45.

Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T.A., Knowles, J.K.C. & Teeri, T.T. (1992). Investigation of the function of mutated cellulose-binding domains *reesei* cellobiohydrolase I. *Proteins.* 14(4), 475-482.

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(4967), 380-386.

Seiboth, B., Ivanova, C. & Seidl-Seiboth, V. (2011). Trichoderma reesei: A Fungal Enzyme Producer for Cellulosic Biofuels, Biofuel Production-Recent Developments and Prospects. In: Bernardes, M.A.D.S. (Ed.) *Biofuel*

Production-Recent Developments and Prospects In Tech. ISBN 978-953-307-478-8.

Sipos, B., Benko, Z., Dienes, D., Reczey, K., Viikari, L. & Siika-aho, M. (2010). Characterisation of specific activities and hydrolytic properties of cell-wall-degrading enzymes produced by *Trichoderma reesei* Rut C30 on different carbon sources. *Appl. Biochem. Biotechnol.* 161(1-8), 347-64.

Spezio, M., Wilson, D.B. & Karplus, P.A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* 32(38), 9906-9916.

Stahlberg, J., Johansson, G. & Pettersson, G. (1988). A binding-site-deficient, catalytically active, core protein of endoglucanase-II from the culture filtrate of *Trichoderma Reesei*. *Eur. J. Biochem.* 173(1), 179-183.

Tamura, M., Miyazaki, T., Tanaka, Y., Yoshida, M., Nishikawa, A. & Tonozuka, T. (2012). Comparison of the structural changes in two cellobiohydrolases, *Cc*Cel6A and *Cc*Cel6C, from *Coprinopsis cinerea* - a tweezer-like motion in the structure of *Cc*Cel6C. *FEBS J.* 279(10), 1871-1882.

Tan, M.L., Balabin, I. & Onuchic, J.N. (2004). Dynamics of electron transfer pathways in cytochrome c oxidase. *Biophys. J.* 86(3), 1813-1819.

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(Pt 8), 875-82.

Vaaje-Kolstad, G., Bohle, L.A., Gaseidnes, S., Dalhus, B., Bjoras, M., Mathiesen, G. & Eijsink, V.G.H. (2012). Characterization of the chitinolytic machinery of *Enterococcus faecalis* V583 and high-resolution structure of its oxidative CBM33 enzyme. *J. Mol. Biol.* 416(2), 239-254.

Vaaje-Kolstad, G., Houston, D.R., Riemen, A.H.K., Eijsink, V.G.H. & van Aalten, D.M.F. (2005). Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J. Biol. Chem.* 280(12), 11313-11319.

Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z., Zhai, H., Sorlie, M. & Eijsink, V.G. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330(6001), 219-22.

vanTilbeurgh, H., Claeyssens, M. & Bruyne, C.K.d. (1982). The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. *FEBS Lett.* 149(1), 152-156.

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-beta-cellobioside, at 1.9 Å. *Acta Crystallogr D Biol Crystallogr* 58, 2201-2204.

Varrot, A., Hastrup, S., Schülein, M. & Davies, G. (1999). Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 A resolution. *Biochem. J.* 15(337), 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(21), 20181-20184.

Vuong, T.V. & Wilson, D.B. (2009a). The absence of an identifiable single catalytic base residue in *Thermobifida fusca* exocellulase Cel6B. *FEBS. J.* 276(14), 3837-45.

Vuong, T.V. & Wilson, D.B. (2009b). Processivity, synergism, and substrate specificity of *Thermobifida fusca* Cel6B. *Appl. Environ. Microbiol.*75(21), 6655-61.

Westereng, B., Ishida, T., Vaaje-Kolstad, G., Wu, M., Eijsink, V.G.H., Igarashi, K., Samejima, M., Stahlberg, J., Horn, S.J. & Sandgren, M. (2011). The putative endoglucanase *Pc*GH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative ezyme that cleaves cellulose. *PLoS ONE* 6(11).

Wilson, D.B. (2004). Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chemical Record* 4(2), 72-82.

Wong, E., Vaaje-Kolstad, G., Ghosh, A., Hurtado-Guerrero, R., Konarev, P.V., Ibrahim, A.F.M., Svergun, D.I., Eijsink, V.G.H., Chatterjee, N.S. & van Aalten, D.M.F. (2012). The *Vibrio cholerae* colonization factor GbpA possesses a modular structure that governs binding to different host surfaces. *PLoS Pathog.* 8(1), e1002373.

Wu, M., Beckham, G.T., Larsson, A.M., Ishida, T., Kim, S., Payne, C.M., Himmel, M.E., Crowley, M.F., Horn, S.J., Westereng, B., Igarashi, K., Samejima, M., Stahlberg, J., Eijsink, V.G. & Sandgren, M. (2013). Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus *Phanerochaete chrysosporium*. *J. Biol. Chem.* doi:10.1074/jbc.M113.459396

Yang, B., Dai, Z., Ding, S.-Y. & Wyman, C.E. (2011). Enzymatic hydrolysis of cellulosic biomass. *Future Science* 2(4), 421-450.

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

Acknowledgements

This work was financially supported by the research programme *MicroDrive* at the Swedish University of Agricultural Sciences, Uppsala, Sweden.

I would like to express my sincere gratitude to all the people who have contributed to this thesis and all those around me during the past four years at the joint 'X-ray' at Uppsala University and Swedish University of Agricultural Sciences:

Jerry Ståhlberg, my main supervisor, for accepting me as a Ph.D. student, trusting and supporting me all the time. Your expertise and optimism have given me such an inspiring research environment. Your broad knowledge on carbohydrate degrading enzymes is invaluable for my learning. Your lovely singing and dancing are absolutely the bonus of my Ph.D..

Mats Sandgren, my co-supervisor, my tutor of crystallography and ski, for bringing me to many different symposiums and conferences, introducing me to the worldwide good scientists, which is invaluable for my career network built up and scientific insights cultivation.

Henrik Hansson, for your great contribution on *T. fusca* Cel6B project and the helps on HPLC assay. It is a little bit magic that the experiments we did together all go very well.

Anna Larsson, for sharing your sequence and structure analysis experiences with me, and helpful discussions and suggestions on both GH6 and GH61 works.

Mikael Gudmundsson, for always being there to talk and help me on different subjects: chemistry, computer technique support, crystallography tips, Swedish, culture, travel, food, life and numerous gossips. The memories on Guinness, the Niagara falls, the Times Square, all have you there. If you have to go to Canada before my dissertation party, do not forget my gift before you leave.

Takuya Ishida, for teaching me *Pichia* expression, sharing your lab tips with me, answering my questions any time even you went back to Japan. Your hard work and patience to science make me impressive.
Majid Momeni Haddad, for holding my hands to run out of the building during the earthquake in Japan, sharing the different feelings about work, life, or future career.
Anna Brisova, it is always good to have the other girl in the office to balance the gender and brought us lovely talks and jokes.
Wim Nerinckx, Gregg T. Beckham, Vincent Eijsink, Bjorge Westereng, Svein Jarle Horn, Colin Mitchinson, Kathleen Piens, Kiyohiko Igarashi, Masahiro Samejima, for being the best collaborators, i learned a lot during the project discussion and paper writing with you. Your passions and ambitions on scientific work encourage me to contribute to science in the future.
Erling and Christer, for helping me fix many computer and technique problems of programs; **Nils** and **Saeid,** teach me how to collect data and solve structure; **Elleonor**, for helping me fill in complicated travel expense form and get money back; **Stefan and Yafei**, for being my half-time examiners and very glad to talk to you all the time, not only about science; **Margareta**, for helping me order all what i need for experiments and courses quickly; **Tex**, for sharing crystallization knowledge in the lab and hotel room at Dublin, enjoy the interesting talk about Chinese, Japanese or vegetarian food.

Many other colleagues made my Ph.D. life much more enjoyable, i am very appreciated to have you around me, many thanks to **Anna J, Cha San, Christofer, Henrik, Karin, Sanjeewani, Roy, Maria, Nina, Torleif, Christofer, Evalena, Inger, Wimal, Michiel, Sherry, Alwyn, Lars, Fredrik, Avinash, Annette, Agata, Gelareh, Lotta, Henrik, Ana, Anton, Gunilla, Dirk, …** and our big Chinese team **Yang, Li, Xiaodi, Wangshu, Xiaohu, Lu, Jie, Shiying, Zhen**. There are also many other friends around lunch table, entertaining party, family party, BBQ party, you guys bring me a wonderful life in Sweden, never feel alone.

Finally, i want to thank my parents for being healthy and happy in China, which allowed me focus on my Ph.D. study without worries (谢谢父母们的健康和快乐，使我能够专注于博士论文的研究) ; my husband, **Wei**, your love and trust are the most precious things i have, let's continue our happy life and bring our coming baby a world like heaven.

Miao @ Uppsala