

Structural Insights into the  
Catalytic Mechanism, Protein Dynamics,  
Inhibition and Thermostability of GH7  
Cellobiohydrolases

Majid Haddad Momeni

*Faculty of Natural Resources and Agricultural Sciences  
Department of Chemistry and Biotechnology  
Uppsala*

Doctoral Thesis  
Swedish University of Agricultural Sciences  
Uppsala 2014

Acta Universitatis agriculturae Sueciae

2014:32

Cover: The crystal structure of the Michaelis complex of *Hypocrea jecorina*  
Cel7A with bound cellononaose (*Hje*Cel7A; PDB code 4C4C)

By: Brandon C. Knott, 2013

ISSN 1652-6880

ISBN (print version) 978-91-576-8012-9

ISBN (electronic version) 978-91-576-8013-6

© 2014 Majid Haddad Momeni, Uppsala

Print: SLU Service/Repro, Uppsala 2014

# Structural Insights into the Catalytic Mechanism, Protein Dynamics, Inhibition and Thermostability of GH7 Cellobiohydrolases

## Abstract

Glycoside hydrolase family 7 cellobiohydrolases (GH7 CBH) are typically the most abundant enzymes of cellulolytic fungi and play a key role in biomass recycling in Nature, as well as in biofuel production from plant biomass. This thesis examines molecular properties of this biologically and industrially important class of enzymes.

**Paper I** shows that *HirCel7A* is the most abundant protein of the serious forest pathogen *Heterobasidion irregulare*. The *HirCel7A* exhibits intermediate dynamical and structural properties between CBHs with the most closed and most open tunnels known in GH7. The results point to tunnel-enclosing loops as important for carbohydrate processivity and association-dissociation on cellulose. **Paper II** presents the first Michaelis complex, with cellononaose spanning 42 Å of the active site, and the first glycosyl-enzyme intermediate trapped in a GH7 CBH. QM/MM calculations determine optimal reaction coordinates, and rate constants at 11 s<sup>-1</sup> for Step1 and 5300 s<sup>-1</sup> for Step2, showing that the glycosylation step is rate-limiting. A product-assisted mechanism is revealed for the deglycosylation step, indicating that expulsion of the cellobiose product is not required prior to hydrolysis of the intermediate.

**In Paper III**, *HgrCel7A* from *Humicola grisea* var. *thermoidea* showed 10 °C higher T<sub>m</sub> and 75% higher yield in a biomass performance assay at 65 °C than the canonical *HjeCel7A* of *Hypocrea jecorina*. The crystal structure of *HgrCel7A* indicates higher flexibility in tunnel-defining loops, and structural features potentially related to thermostability and enhanced activity, including a putative conformational switch in an active-center loop not reported previously in GH7. **In Paper IV**, structures of *HjeCel7A* in complex with xylooligosaccharides of DP 3-5 show predominant binding in the beginning of the tunnel and partial occupancy for a second binding mode near the catalytic centre. Birchwood xylan displayed ~100-fold stronger inhibition based on mass, suggesting that it may penetrate further into the tunnel and occupy a longer stretch of the active site.

*Keywords:* cellulose, cellulase, glycoside hydrolase, cellobiohydrolase, inhibition, molecular dynamics, retaining mechanism, xylobiose, xylooligosaccharides

*Author's address:* Majid Haddad Momeni, SLU, Department of Chemistry and Biotechnology, P.O. Box 7015, SE-750 07 Uppsala, Sweden

*E-mail:* majid.haddad@slu.se

## Dedication

*To my beloved parents and siblings...*

# Contents

<b>List of Publications</b>	<b>7</b>
<b>Abbreviations</b>	<b>9</b>
<b>1 Introduction</b>	<b>11</b>
1.1 Ligoncellulosic biomass	11
1.2 Biodegradation of lignocellulose	13
1.3 Main features of GH7 cellobiohydrolases (CBHs)	15
1.4 Known family 7 CBHs	16
1.5 Catalytic mechanisms of GHs	17
1.6 Inhibition of cellulases	19
1.7 Thermostable GH7 CBHs	19
1.8 Molecular dynamic (MD) approaches	20
<b>2 Present investigations</b>	
<b>2.1 Structural and biochemical studies of the Glycoside Hydrolase Family 7 CBH of the Tree-killing fungus <i>Heterobasidion irregulare</i> (paper I)</b>	
2.1.1 Enzyme preparation	22
2.1.2 Structural solution and refinement	22
2.1.3 Quality of final structure models	22
2.1.4 Overall structural comparison of <i>HirCel7A</i> with related enzymes	23
2.1.5 Highlights of computational characterization	27
<b>2.2 The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies (paper II)</b>	
2.2.1 Crystallisation and structural determination	29
2.2.2 Michaelis complex structure	30
2.2.3 Glycosyl Enzyme Intermediate structure (GEI)	30
2.2.4 Computational highlights	31
<b>2.3 Expression, crystal structure and cellulase activity of the thermostable Cellobiohydrolase Cel7A from the fungus <i>Humicola grisea</i> var. <i>thermoida</i> (paper III)</b>	
2.3.1 Expression of GH7 CBHs	35
2.3.2 Biochemical characterization	36
2.3.3 Crystallization and Structural determination of <i>H. grisea</i> Cel7A	36

2.3.4	Overall structure of <i>HgrCel7A</i> and comparison with other CBHs	37
2.3.5	Potential thermostabilizing interactions in <i>HgrCel7A</i>	40
<b>2.4</b>	<b>Structural insights into the inhibition of Cellobiohydrolase Cel7A by xylooligosaccharides (paper IV)</b>	
2.4.1	Experimental procedures	41
2.4.2	Overview of the crystallographic structures	41
2.4.3	Tunnel-entrance binding sites	42
2.4.4	Second binding mode observed in X4 and X5 complexes	43
2.4.5	Inhibition studies	45
<b>3</b>	<b>Conclusions and future perspectives</b>	47
	<b>References</b>	<b>51</b>
	<b>Acknowledgements</b>	<b>57</b>

## List of Publications

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

- I **Momeni MH\***, Payne CM\*, Hansson H, Mikkelsen NE, Svedberg J, Engström Å, Sandgren M, Beckham GT and Ståhlberg J. (2013). Structural, Biochemical, and Computational Characterization of the Glycoside Hydrolase Family 7 cellobiohydrolase of the Tree-killing Fungus *Heterobasidion irregulare*. *J Biol Chem* **288**, 5861-5872.
- II Knott BC, **Momeni MH**, Crowley MF, Mackenzie LF, Götz AW, Sandgren M, Withers SG, Ståhlberg J and Beckham GT (2013). The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J Am Chem Soc* **136**, 321-329.
- III **Momeni MH**, Goedegebuur F, Hansson H, Karkehabadi S, Askarieh G, Mitchinson, C, Larenas E, Ståhlberg J, Sandgren M. Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus *Humicola grisea* var. *thermoidea*. (*Submitted to Acta cryst D*)
- IV **Momeni MH**, Ubhayasekera W, Hansson H, Sandgren M., Stahlberg J. Structural insights into the inhibition of Cellobiohydrolase Cel7A by xylooligosaccharides. (*Manuscript*)

Papers I-II are reproduced with the permission of the publishers.

\* First authorship shared

## Other publications

- V Dererie DY, Trobro S, **Momeni MH**, Hansson H, Blomqvist J, Passoth V, Schnurer A, Sandgren M, Ståhlberg J (2010). Improved bio-energy yields via sequential ethanol fermentation and biogas digestion of steam exploded oat straw. *Bioresource Technol* **102**, 4449-55.
- VI Blomqvist J, South E, Tiukova I, **Momeni MH**, Hansson H, Ståhlberg J, Horn JS, Schnurer J, Passoth V (2011). Fermentation of lignocellulosic hydrolysate by the alternative industrial ethanol yeast *Dekkera bruxellensis*. *Lett Appl Microbiol* **53**, 73-8.
- VII Tiukova I, de Barros Pita W, Sundell D, **Momeni MH**, Horn SJ, Ståhlberg J, de Morais MA Jr, Passoth V (2014). Adaptation of *Dekkera bruxellensis* to lignocellulose substrate. *Biotechnol Appl Bioc* **61**, 51-57.
- VIII Gudmundsson M, Kim S, Wu M, Ishida T, **Momeni MH**, Vaaje-Kolstad G, Lundberg D, Royant A, Ståhlberg J, Eijsink VGH, Beckham GT, and Sandgren M. Activation of a lytic polysaccharide monooxygenase active site: Structural and electronic snapshots during the transition from a Cu(II) to Cu(I) metal center by X-ray photo-reduction. (*Submitted to J Biol Chem*)

## Abbreviations

AA	Auxiliary Activities
AFEX	Ammonia fiber expansion
AS	Aimless Shooting
Bgl	$\beta$ -glucosidase
CBH	Cellobiohydrolase
CBM	Carbohydrate-Binding Module
CD	Catalytic domain
DNP-2F-G3	2-deoxy-2-fluoro- $\beta$ -cellotrioside
EG	Endoglucanase
GEI	Glycosyl Enzyme Intermediate
GH	Glycoside Hydrolase
<i>Hir/H.irregulare</i>	<i>Heterobasidion irregulare</i>
<i>Hje/H.jecorina</i>	<i>Hypocrea jecorina</i>
<i>Hgr/H. grisea</i>	<i>Humicola grisea</i>
MD	Molecular dynamics
MM	Molecular mechanics
PASC	Phosphoric acid swollen cellulose
<i>Pch/P.chrysosporium</i>	<i>Phanerochaete chrysosporium</i>
PDB	Protein Data Bank
PMF	Potential mean force
<i>pNP-Lac</i>	para-nitrophenyl Lactosidase
QM	Quantum mechanical
RMSF	Root mean square fluctuations
XOS	xylo-oligosaccharide



# 1 Introduction

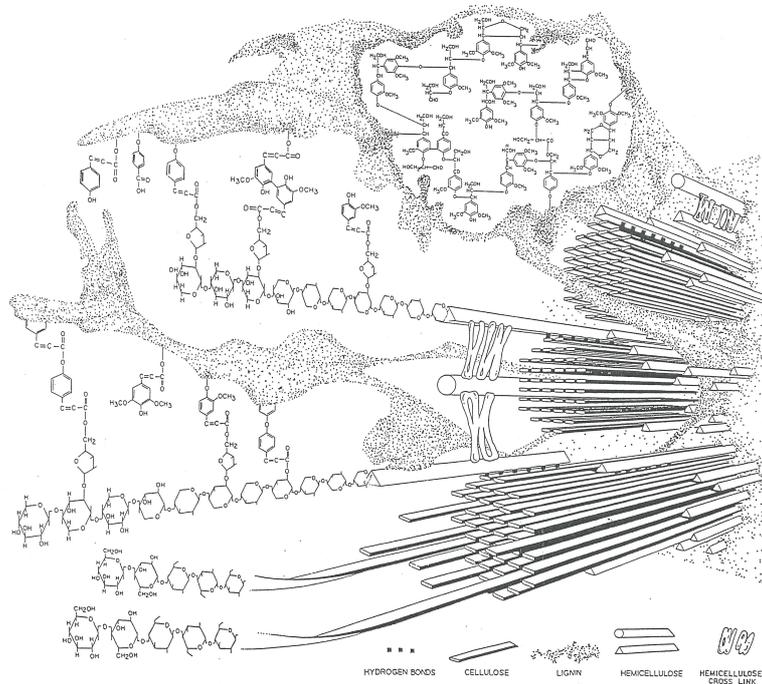
## 1.1 Lignocellulosic biomass

Lignocellulose is the collective name for the three major constitutive components of plant material, known as cellulose, hemicellulose and lignin. Cellulose builds up roughly 45-50% of the dry weight of plant cell walls. It is a glucan homopolymer that consists of unbranched linear chains of  $\beta$ -1,4 linked glucose residues that are typically several thousand residues in length. The chains are bound together with hydrogen bonds and van der Waals interactions into highly regular crystalline microfibrils, which provide tensile strength to lignocellulosic materials. Microfibrils in turn are joined together into larger bundles called fibrils. Although cellulose is highly regular and crystalline by nature, there are forms and regions that are less well-ordered, usually referred to as amorphous cellulose (Beguin P. & Aubert JP, 1994).

In addition to cellulose, plant biomass contains a variety of other polysaccharides collectively named hemicellulose that typically constitutes about 25-30% of total plant cell wall materials. Hemicellulose polysaccharides vary in composition, type of linkages, branching and substitutions (Scheller HV & Ulvskov P, 2010). Major monosaccharides are D-xylose, D-mannose, D-galactose, D-glucose, L-arabinose, 4-O-methyl-glucuronic, D-galacturonic and D-glucuronic acids. They are usually built on a linear  $\beta$ -1,4 linked backbone, in some cases with occasional  $\beta$ -1-3-glycosidic bonds. The major component of hardwood hemicellulose is glucuronoxylan, whereas glucomannan is a dominant building block in softwood. In contrast to cellulose, hemicelluloses have lower degree of polymerization (DP), are usually branched and are heterogeneous in both composition and structure, and they are more susceptible to hydrolytic degradation (Jeffries, 1994).

The hemicellulose polysaccharides are secreted alongside with cellulose synthesis and they surround and interconnect the cellulose fibrils. At later stages of cell wall formation, lignin monomers (phenylpropane units) are secreted into the

cell wall where they condensate through radical polymerization into a complex three-dimensional polymer called lignin, which fills up space between the polysaccharide components (Figure 1). It gives structural support, impermeability and resistance against oxidative stress and/or microbial attack to the lignocellulosic materials.



*Figure 1.* Secondary cell-wall structure. Cellulose microfibrils and hemicellulose chains are surrounded by lignin. Specific linkages and components of non-core lignin are indicated for a generalized grass secondary cell-wall.

Consequently, two main obstacles hinder the hydrolysis of lignocellulosic materials, the crystalline structure of cellulose (tightly bound in microfibrils) and the tightly bound and resistant lignin surrounding it (Aziz & Sarknen, 1989).

The biological degradation of lignocellulosic materials has acquired substantial attention by microbiologists and biotechnologists for several decades. Due to the heterogeneity and recalcitrant nature of these materials, the industrial process for conversion of cellulosic materials requires mechanical size reduction combined with harsh physical, and/or chemical pretreatment prior to enzymatic degradation, e.g. ammonia fiber expansion (AFEX) and dilute sulfuric acid pretreatments. These techniques usually result in partial hydrolysis of hemicellulose, liberation of

pentoses (5-carbon sugars), redistribution of lignin and decreased crystallinity of cellulose fibrils (Fan *et al.*, 1982; Sun & Cheng, 2002).

## 1.2 Biodegradation of lignocellulose

In nature, lignocellulose is recycled mainly by microorganisms, fungi and bacteria, by the use of extracellular enzymes. A multitude of enzymes with different specificities are required to degrade the various components of the material. The cellulose and hemicellulose polysaccharides are mainly degraded by different glycoside hydrolases (GH), whereas oxidative–ligninolytic enzymes are responsible for lignin depolymerization. Cellulases and hemicellulases are widespread among diverse groups of organisms, whereas efficient ligninolytic systems are found only among higher fungi (brown rot and white rot basidiomycetes).

Cellulose is the main component of plant biomass and cellulase enzymes are often found as both the largest number of GH genes and the majority of proteins by mass in the secretomes of biomass degrading organisms. Cellulases are enzymes that are able to hydrolyze the  $\beta$ -1,4 glycosidic bonds in insoluble cellulose. In common with other enzymes acting on insoluble polysaccharides, cellulases often exhibit multi-modular architecture where the catalytic domain (CD) is connected to one or several carbohydrate binding modules (CBM) by flexible linkers and sometimes further CDs with other specificities (Shoseyov *et al.*, 2006). Almost all modular cellulases of fungi, such as *H. jecorina* Cel7A, consist of a single CD, a highly glycosylated 30–50 residue linker and a small family 1 CBM (CBM1) of ~35 residues. Apparently, the function of the CBM is to hold the enzyme on the cellulose surface to facilitate for the CD to sequester cellulose chains into the active site (Din *et al.*, 1991).

Several cellulases from different GH families cooperate synergistically in the degradation of biomass, due to the heterogeneity of the lignocellulosic material and the presence of both crystalline and amorphous cellulose regions. The enzymes can be broadly divided into two main groups, cellobiohydrolases (CBH) and endoglucanases (EG), based on their architecture and mode of action. EGs generally have an open cellulose-binding cleft. They cleave preferentially bonds internally in cellulose chains and are most active on amorphous cellulose and are less efficient on crystalline regions (Barr *et al.*, 1996). On the other hand, in CBHs the cellulose-binding path is surrounded by longer loops that cover the cleft so that a tunnel is formed through the catalytic domain. Therefore, CBHs act preferentially from either the reducing or the non-reducing ends of cellulose chains in a processive manner, i.e. they cleave off several cellobiose units before release of the cellulose chain. The soluble products, mainly cellobiose and small amounts of longer oligosaccharides are then hydrolyzed to glucose by  $\beta$ -glucosidases (Bgl).

Traditionally, the cooperative effect has been attributed to the classical ‘endo-exo synergism’ model, which postulates that EGs increase the number of available end groups for initiation by CBHs. However, more recently a new class of enzymes has been discovered called lytic polysaccharide monooxygenases (LPMOs) that cleave glycosidic linkages in polysaccharides using a copper-mediated, oxidative mechanism (Vaaje-Kolstad, *et al.*, 2010; Hemsworth *et al.*, 2014). Fungal LPMOs were previously classified in family GH61 and bacterial LPMOs in CBM33, however they have recently been moved to the new Auxiliary Activities families AA9 and AA10, respectively, in the Carbohydrate Active enZYmes (CAZy) database. These enzymes are thought to act directly on surfaces of crystalline polysaccharides. Unlike GHs they are able to make chain breaks in crystalline regions without the need to detach single cellulose chains from the insoluble substrate. Thus, LPMOs represent a new enzyme mechanism for the decomposition of recalcitrant polysaccharides, which can act synergistically with traditional hydrolytic enzymes (Harris, *et al.*, 2010; Merino & Cherry, 2007).

To date, over 115 GH families have been discovered and defined in the CAZy database. The classification is based on sequence similarities (Henrissat & Davies, 1997; Henrissat & Bairoch, 1996; Henrissat, 1991). Several families contain members from both bacteria and fungi, and enzymes with different activities and substrate specificities.

Filamentous fungi play a central role in the global recycling of lignocellulose and account for the majority of plant biomass decomposition on land. Cellulose utilization is distributed throughout the entire fungal kingdom, from primitive Chytridiomycetes to the advanced Basidiomycetes and thus represents a large genetic diversity (Lynd *et al.*, 2002).

The ascomycete fungus *Hypocrea jecorina* (anamorph to *Trichoderma reesei*) is the predominant source of cellulase enzymes for industrial use (Geddes *et al.*, 2011), and has long served as a model organism for research on mechanism, specificity and synergistic cooperation of cellulose-degrading enzymes. The genome of *H. jecorina* has been sequenced, and transcriptional analysis identified about 35 upregulated genes coding for known or predicted biomass active proteins, including two CBHs and several EGs and BglS (Foreman *et al.*, 2003).

Major components in the *H. jecorina* secretome are two processive CBHs, Cel7A (40–50%) and Cel6A (~20%), which act preferentially from opposite ends and in opposite directions along cellulose chains, Cel7A from the reducing end and Cel6A from the nonreducing end. They also utilize different reaction mechanisms. Cel7A hydrolyses glycosidic bonds with net retention of the  $\beta$ -anomeric configuration, whereas Cel6A is an inverting enzyme. The third most abundant enzyme is the endoglucanase Cel7B, which is homologous to Cel7A and similar in

structure, but with an open cellulose binding cleft instead of a tunnel due to deletions in several loops along the active site (Kleywegt *et al.*, 1997).

Natural biomass materials are not well suited for fundamental research due to their inherent complexity described above. Therefore, more well-defined model substrates are commonly applied in cellulase research. Avicel and phosphoric acid swollen cellulose, PASC, are among the most widely used crystalline and amorphous substrates, respectively (Zhang *et al.*, 2004, 2006). Avicel is prepared by partial acid hydrolysis of wood cellulose. It is mainly crystalline and exhibits a relatively low DP around 200. PASC is usually prepared from Avicel by partial dissolution in concentrated phosphoric acid followed by re-precipitation in water (Schulein, 1997). The DP is maintained while the physical properties are significantly moderated. The accessible surface area is increased, whereas the crystallinity is drastically reduced (Zhang *et al.*, 2006).

### 1.3 Main features of GH7 cellobiohydrolases

In several cellulolytic fungi, the major proteins produced under cellulase-inducing conditions are GH7 CBHs. The *H. jecorina* Cel7A (*Hje*Cel7A) constitutes approximately half of total secreted protein, and gene knock out studies have shown that it is a rate-limiting factor in cellulose degradation (Ilmen *et al.*, 1997), pointing at a key role for GH7 enzymes in biomass degrading fungi. Due to their large biological importance and utilization in industrial applications, GH7 cellobiohydrolases are main targets of several structural, functional and engineering studies.

The most extensively studied GH7 CBH is *Hje*Cel7A and it has served as a model for processive CBHs. The first GH7 structure was obtained with this enzyme after removal of the linker-CBM portion using papain (Divne *et al.*, 1994). Structural studies of cellodextrin binding using catalytically deficient mutants has allowed the complete mapping of cellulose binding along the entire active site (Divne *et al.*, 1998).

The structure of the catalytic domain of *Hje*Cel7A exhibits a  $\beta$ -jelly roll fold with two largely antiparallel  $\beta$ -sheets packing face to face to form a curved  $\beta$ -sandwich. The main structural feature is a 50 Å long cellulose-binding tunnel formed by long loops extending from the edges of the  $\beta$ -sandwich. The tunnel contains 11 glucosyl-binding subsites, numbered -7 to +4 from the non-reducing end to the reducing end of the cellulose chain, and with the point of cleavage between subsites -1 and +1 (Divne *et al.*, 1998). The indole side chains of four tryptophan residues are exposed at subsites -7, -4, -2 and +1 to serve as hydrophobic sugar-binding platforms as often seen in carbohydrate active enzymes.

These tryptophans as well as most other protein-substrate interactions present in *HjeCel7A* are highly conserved among the known GH7 CBHs.

#### 1.4 Known family 7 CBHs

To date, there are three dimensional structures available in the PDB of ten GH7 enzymes; seven CBHs: *Hypocrea jecorina* Cel7A (*HjeCel7A*) (Divne 1994, Divne, Ståhlberg *et al.* 1998), *Phanerochaete chrysosporium* Cel7D (*PchCel7D*) (Munoz *et al.*, 2001), *Melanocarpus albomyces* Cel7B (*MalCel7B*) (Parkkinen, *et al.*, 2008), *Rasamsonia emersonii* Cel7A (*RemCel7A*) (Grassick *et al.*, 2004), *Trichoderma harzianum* Cel7A (Textor *et al.*, 2012) and *Heterobasidion irregulare* Cel7A (*HirCel7A*; Momeni *et al.*, 2013) and *Limnoria quadripunctata* Cel7B (*LquCel7B*; Kern *et al.*, 2013); and three EGs: *H. jecorina* Cel7B (Kleywegt *et al.*, 1997), *Humicola insolens* Cel7B (Davies *et al.*, 1997) and *Fusarium oxysporum* Cel7B (Sulzenbacher *et al.*, 1996). They are all very similar along the  $\beta$ -sandwich core and around the active center with a conserved catalytic triad and tryptophan platforms at subsites -2 and +1. However, as shown in Figure 2, the CBHs reveal more closed tunnels, whereas the EGs exhibit open clefts, due to the variation in length and sequence of loops along the cellulose binding sites (von Ossowski *et al.*, 2003).

Among the GH7 CBHs, most of the protein-sugar interactions observed in *HjeCel7A* are well conserved. However, differences in length and sequence of loops along the cellulose binding path vary the accessibility of the active site and will affect the dynamics of loop movements, which in turn will influence processivity, product inhibition, probability of endo-initiation and release of nonproductively bound enzyme. *HjeCel7A* reveals the most closed tunnel, whereas *PchCel7D* exhibits the most open active site of known GH7 CBHs (Figure 2).

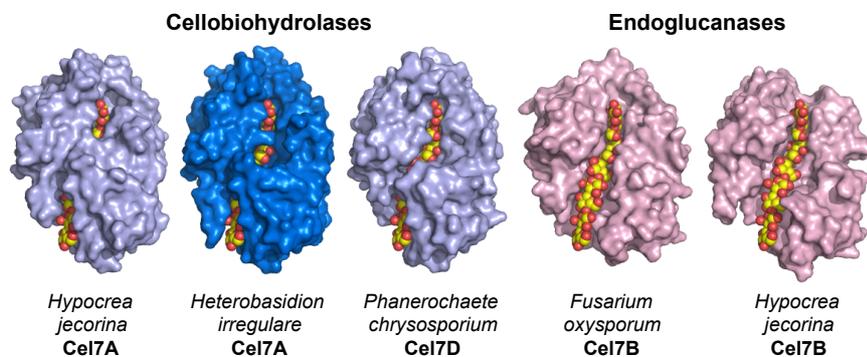


Figure 2. Comparison of the molecular surface between *Heterobasidion irregulare* Cel7A (*HirCel7A*), *P. chrysosporium* Cel7D (*PchCel7D*), *H. jecorina* Cel7A (*HjeCel7A*) and *H. jecorina* Cel7B and *Fusarium oxysporum* Cel7B (Endos).

## 1.5 Catalytic mechanisms of glycoside hydrolases

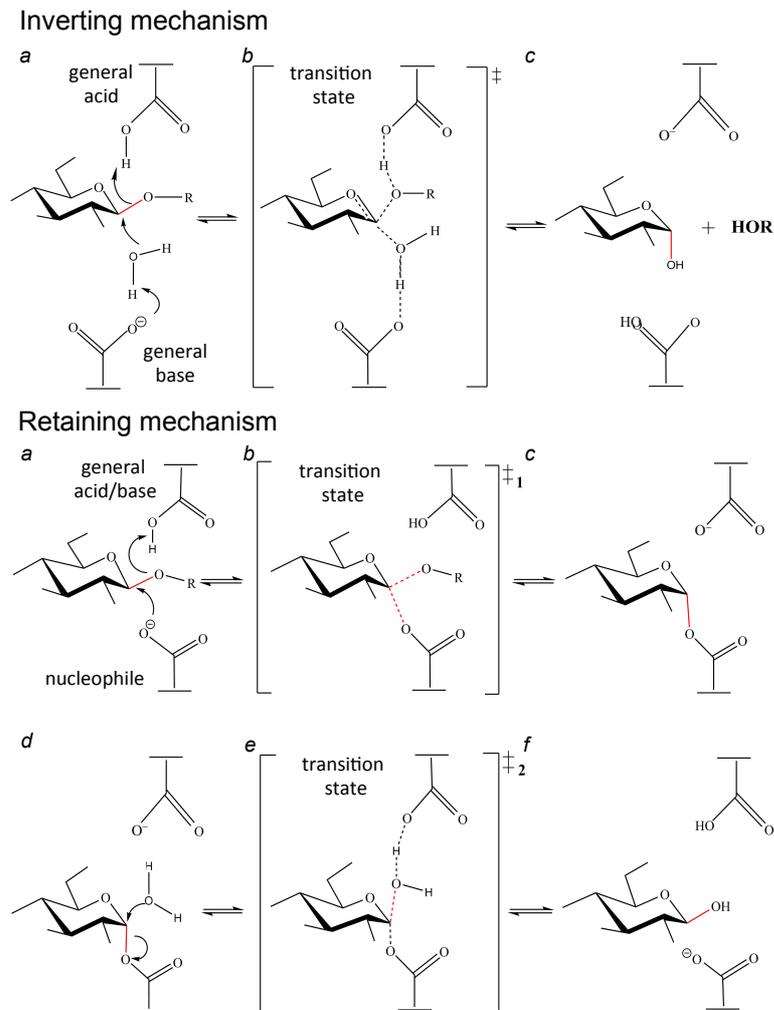
GH enzymes employ either one or the other of two distinct mechanisms, which were postulated by Koshland already in 1953. They differ in the stereochemical outcome of the anomeric configuration and are commonly known as ‘Retaining’ and ‘Inverting’ mechanisms, respectively.

The Inverting mechanism is achieved by a one step, single displacement reaction over a single transition state. Two suitably positioned carboxylic acid residues act as general acid and general base, respectively. The acid protonates the glycosidic oxygen at the same time as the base abstracts a proton from and activates a water molecule held in place for nucleophilic attack on the anomeric carbon of the glycosidic bond to be cleaved (Figure 3, top panel). This results in the inversion of the anomeric configuration, hence the name of the mechanism.

The Retaining mechanism refers to a two-step, double displacement mechanism over two transition states, using an acid/base and a nucleophile residue, again generally aspartate and/or glutamate residues. During the first step, the glycosylation step, proton transfer is occurring from the acid/base residue to the glycosidic oxygen, similarly to the inverting mechanism. However, the other catalytic residue, the nucleophile, is positioned in direct proximity to, and simultaneously making a nucleophilic attack at the anomeric carbon of the glycosyl unit in subsite -1. The glycosidic bond is broken and replaced by a glycosyl-ester bond to the nucleophile to form a glycosyl-enzyme intermediate (GEI).

In the second step of the reaction, known as the deglycosylation step, a water molecule is bound near the anomeric carbon and is deprotonated by the acid/base residue, which now acts as catalytic base. The activated water molecule makes a nucleophilic attack at the anomeric carbon, which breaks the GEI bond and simultaneously restores the acid/base and nucleophile residues (Figure 3, bottom panel).

As far as known to date, all enzymes in the same GH family utilize the same catalytic mechanism. GH7 cellulases employ the retaining mechanism. For these and other GH enzymes acting on  $\beta$ -1,4 glycosidic bonds, it is necessary to distort the -1 glucosyl unit from the ground state  ${}^4C_1$  chair conformation in the ES-complex. Otherwise the anti-orbital to the glycosidic bond on the anomeric carbon will not be accessible to the nucleophile for attack.



*Figure 3.* The two major mechanisms of enzymatic glycosidic hydrolysis (Koshland *et al.*, 1953). In the *Inverting mechanism*, the protonation of the glycosidic oxygen and the release of glycosyl units are accomplished in a single step. The *Retaining mechanism* is initiated through protonation of the glycosidic oxygen and a nucleophilic attack on the anomeric carbon leading to bond cleavage and glycosylation of the nucleophile. The resulting glycosyl-enzyme intermediate is hydrolyzed by a water molecule, which donates a proton to the acid/base residue to complete the catalytic action.

The catalytic action of GH7 is accomplished by a triad of catalytic amino acids positioned near to each other on the same  $\beta$ -strand in the highly conserved EXDXXE motif (also found in GH16). The catalytic function of these amino acids was implied after the discovery of the first GH7 structure, that of *HjeCel7A* (Divne *et al.*, 1994). Glu212 and Glu217 were proposed to function as the nucleophile and acid/base residues, respectively. This was subsequently confirmed by three isosteric

single mutations to produce variants E212Q, D214N and E217Q (Ståhlberg *et al.*, 1996). The activity was essentially abolished in the E212Q and E217Q mutants and also drastically reduced (~80-fold) in D214N showing that Asp214 substantially promotes the catalytic action, probably due to hydrogen bonding with Glu212.

## 1.6 Inhibition of cellulases

Lignocellulosic biomass is designed by nature to resist microbial degradation and is substantially more recalcitrant to enzymatic degradation than e.g. starch-based polysaccharides due to its complex structure containing hemicellulose and cellulose surrounded by lignin. (Zhang *et al.*, 2004). As mentioned above, physicochemical pretreatment is essential for disruption of the complex structure, to facilitate the liberation of hemicellulose and/or lignin components and consequently increase the accessibility of the cellulose polymer for enzymatic hydrolysis (Modenbach & Nokes, 2012). Different alkaline, acidic and/or hydrothermal pretreatments are among the most applied methods worldwide.

It has been recognized that pretreatments often liberate and/or generate compounds that are inhibitory to enzymes, which seriously reduces the efficiency of subsequent saccharification processes. Xylan and xylooligosaccharides (XOS) have been experimentally confirmed to be potent inhibitors of both GH6 and GH7 CBHs and have been proposed to play a significant role for this inhibition (Baumann *et al.*, 2011).

## 1.7 Thermostable GH7 CBHs

The economy of an industrial scale enzymatic conversion of biomass to fermentable sugars would benefit from an improved thermostability of the enzyme mixture (Viikari *et al.*, 2007), since the lifetime of the cellulases are expected to increase with thermostability. Thus, thermostable cellulases are good candidates to be used for industrial biomass conversion processes since higher thermal stability could lead to shorter hydrolysis time and higher specific activity at elevated temperatures. Significant effort has therefore been invested in the search for thermostable enzymes, including GH7 CBHs with enhanced thermostability compared to *HjeCel7A* ( $T_m \sim 62$  °C).

The thermophilic ascomycete fungus *Melanocarpus albomyces* (anamorph *Myriococcum albomyces*) secretes both xylanases and cellulases with significant thermal activity and stability including the structurally determined *MalCel7B* (Voutilainen *et al.*, 2007). *Ramsonia emersonii*, isolated from composting biomass, is another thermophilic aerobic fungus that is able to produce a mixture of thermostable CBHs from both GH6 and GH7 (*RemCel7A*).

Further thermostable GH7 CBHs have been isolated from the thermophilic fungi *Thermoascus aurantiacus*, *Acremonium thermophilum* and *Chaetomium thermophilum*. They exhibit temperature optima at ~70 °C, ~60 °C and ~65 °C, respectively (Voutilainen *et al.*, 2008). Structures of these enzymes have not been reported to date, but structures of *MalCel7B* and *RemCel7A* are available.

## 1.8 Molecular dynamic (MD) approaches

X-ray crystallography is of paramount importance for protein structure studies since it is the only technique available today for experimental determination of atom coordinates within larger macromolecules. However, crystal structures provide rather static pictures and limited information about the flexibility of proteins in solution. Computational approaches offer complementary insights into protein dynamics, and interactions with ligand, substrate and solvent.

The computational chemistry group at National Renewable Energy Laboratory (*NREL*), USA, has done pioneering work recently to set up MD simulation methods relevant for biomass processing (Beckham *et al.*, 2011a), in which both the dynamics of enzyme and the insoluble cellulose substrate can be examined. For example, MD simulation studies cast new light on the role of the cellulose-binding module (CBM) and the linker in *HjeCel7A* (Payne *et al.*, 2013); have revealed that enzyme binding to the cellulose surface affects loop dynamics and ligand binding in *HjeCel7B* (Lin *et al.*, 2011); and have enabled the calculation of the energy required for decrystallisation of cellulose chains from the surface of crystalline cellulose (Beckham *et al.*, 2011b). Furthermore, computational work on the energetics of cellobiose binding to the product sites has suggested that processive and nonprocessive GH7 enzymes from *H. jecorina* vary significantly, which has implications for product inhibition (Bu *et al.*, 2013).

## 2 Present investigations

In this thesis, I have exclusively investigated structural, biochemical and mechanistic features of selected family 7 cellobiohydrolases (GH7 CBHs) secreted by different fungi, *H. irregulare*, *H. jecorina* (mesophilic) and *H. grisea* (thermophilic fungus). Our structural snapshots were combined with molecular simulation studies to examine aspects of protein dynamics, catalytic mechanism, inhibition and thermostability within this biologically and industrially important class of enzymes.

### 2.1 Structural and biochemical studies of the Glycoside Hydrolase Family 7 CBH of the Tree-killing fungus *Heterobasidion irregulare*, HirCel7A. (Paper I)

Root-rot fungi of the *Heterobasidion annosum sensu lato* complex are among the most serious forest pathogens and are responsible for considerable economical loss in temperate forests. The genome of the species *Heterobasidion irregulare*, which infects pine, has been sequenced and contains over 280 carbohydrate-active enzymes.

In this study, we have isolated the most abundant protein in the secretome of *H. irregulare* when grown on woody biomass, the GH7 CBH HirCel7A. The enzyme consists of a GH7 catalytic domain of 440 residues without any linker-CBM. The HirCel7A enzyme was subjected to crystallization experiments, followed by structure determination of the Apo form and soaked with ligand (SX<sub>5</sub>) at 1.9 Å and 1.7 Å resolution and  $R_{free}$  of 25.5% and 17.7%, respectively.

The overall structure confirms that the HirCel7A is a cellobiohydrolase rather than an endoglucanase, with a cellulose-binding tunnel that is more closed than PchCel7D and more open than HjeCel7A. In addition, molecular simulations of *H. irregulare*, *H. jecorina*, and *P. chrysosporium* were conducted and revealed

differences in enzyme-ligand interactions, ligand solvation and flexibility of loops surrounding the substrate binding sites.

### **2.1.1 Enzyme preparation**

The *H. irregulare* strain TC 32-1 (Olson *et al.*, 2012) was cultivated at ambient temperature using minimal medium (pH 5.0) and microcrystalline cellulose, milled spruce heartwood, or Aspen sawdust as carbon sources. Previous to this study, major proteins in the culture filtrate from a culture on spruce were identified by peptide mapping. The most abundant protein was identified as *HirCel7A* (Wang, 2009). In the present investigation, the *HirCel7A* was purified from a culture on aspen sawdust by anion exchange and gel filtration chromatography. The purified protein was concentrated to 9.0 mg/ml at pH 5.0.

### **2.1.2 Structure solution and Refinement**

Crystals appeared already in the initial screening using the JCSG+ screen. After optimization of the crystallization conditions, crystals of sufficient quality for data collection were obtained. The *HirCel7A* structure was determined by molecular replacement with Phaser (McCoy *et al.*, 2007) using the structure of *PchCel7D* as a search model (PDB code 1GPI). For structure refinement, a homology model was built by threading the *HirCel7A* sequence onto the *PchCel7A* structure using SwissPDBviewer (Guex & Peitsch, 1997). Further model building and refinement was done using REFMAC 5 and COOT and other programs of the CCP4 suite. Swiss PDBviewer was used for structural alignment of the structure models. The atomic coordinates and structure factors have been deposited at the Protein Data Bank with PDB codes 2YG1 and 2XSP for the apo and SX<sub>5</sub> soaked structures, respectively.

### **2.1.3 Quality of final structure models**

Structures were obtained of *HirCel7A* crystallized without ligand (Apo) and in presence of 0.8 mM  $\beta$ -1,4-D-thioxylopentaoside (SX<sub>5</sub>). The structure of SX<sub>5</sub>, contains only one molecule in the asymmetric unit and all 440 amino acid residues show well-defined electron density. However, the Apo structure exhibits two molecules in the asymmetric units, both of which display areas with relatively poor densities associated with elevated temperature factors. In chain A, all 440 residues could be modeled, but in chain B three loops showed insufficient density for model building and the following amino acid residues in those loops are excluded from chain B in the final model: 45-47 (3 residues), 52-55 (4 residues), and 197-203 (7 residues).

The SX<sub>5</sub> structure shows electron density for one xylose residue at site +1 and also one molecule of HEPES at binding site -2. The xylose residue binds very similarly to the +1 glucosyl unit in the structures of *Hje*Cel7A E212Q mutant in complex with cellobiose and cellotetraose (PDB codes, 3CEL and 5 CEL, respectively). However, the xylose residue here is in the form of the alpha anomer and is obviously lacking the 6-hydroxymethyl group present in glucose.

#### 2.1.4 Overall structural comparison of *Hir*Cel7A with related enzymes

As expected from high sequence similarities (Figure 4), the overall structure of *Hir*Cel7A (Figure 5) is similar to the catalytic modules of *Pch*Cel7D (69% identity) and *Hje*Cel7A (56% identity) with root mean square distances (RMSD) of 0.6 and 0.8 Å for 421 and 407 matching Ca atoms, respectively. Major differences are observed at four regions along the substrate binding path: at the tunnel entrance at subsites -7/-6 (Figure 6A, loop A1), loop contacts around subsite -4 (Figure 6B, loop A3 and B2), near the catalytic center (Figure 6C, loop A3 and B3) and adjacent to the product-binding sites (Figure 6D), which are discussed in the following sections.

Hir_7A_2xsp	1	XQVGTQTAEN	HPKLTVSQCS	AGGSCITTESR	SVVLDSNWRW	LHTTSGTTNC	YTGNTW	DASL	CPDPV	TCAQN
Hje_7A_8cel	1	XSACTLQSET	HPPLTWQKCS	SGGCTCQQTG	SVVIDANWRW	THATNSSTNC	YDGNLW	SSTL	CPDNET	CAKN
Pch_7D_1z3v	1	XQAGNTAEN	HPQLQSQQCT	TSGGCKPLST	KVVLDSNWRW	VHSTSGYTNC	YTGNEW	DTSL	CPDGK	TCAAN
Hir_7A_2xsp	71	CALDGDADYSG	TYGISTSGNA	LTLKFV	YST	NIGSRVY	LMSADDTNYE	IFKLNQEFA	FDVDM	SNLPC
Hje_7A_8cel	71	CCLDGAAYAS	TYGVTTSGNS	LSIDFV	QSA	-QK	NVNGARLY	LMA	SDTTYQ	EFTLLGNEFS
Pch_7D_1z3v	71	CALDGDADYSG	TYGITSTGTA	LTLKFV	G	-	NVNGSRVY	LMA	DDTHYQ	LLKLLNQEFT
Hir_7A_2xsp	141	GLNGALYFVE	MDADGGLSRF	PNNKAGSKYG	TGYCDTQCPQ	DIKFINGEAN	ILGWTP	PSSSD	SN	ANT
Hje_7A_8cel	139	GLNGALYFVS	MDADGGVSKY	PTNTAGAKYG	TGYCDSQCPR	DLKFINGEAN	VEGWE	PSSNN	ANT	GT
Pch_7D_1z3v	136	GLNGALYLSA	MDADGGMSKY	PGNKAGAKYG	TGYCDSQCPK	DIKFINGEAN	VGNWT	E	-	G
Hir_7A_2xsp	211	CCNEMDVWEA	NINSAAVTPH	VCNVQGQTRC	SGTQC	GD	-	G	DERYD	GICDK
Hje_7A_8cel	209	CCSEMDIWEA	NSISEALTPH	PCTTVGQEIC	EGDGC	GGTYS	DNRYG	GGTCDP	DGCDWNP	YRLL
Pch_7D_1z3v	204	CCSEMDIWEA	NNDAAAFTPH	PCTTTGQTRC	SGDDC	A	-	-	-	RNT
Hir_7A_2xsp	279	--TVNTNSKF	TVVTQFLTSD	NTTTGTLHEI	RRLVYQNGKV	IANSKTN	IAG	MSQFDS	ITDD	FCNAQKTA
Hje_7A_8cel	279	SFTLDTTKKL	TVVTQFETS-	-	-	-	-	GAI	NRYYVQNGVT	FQQPNAEL
Pch_7D_1z3v	268	--TVDTSKPF	TVVTQFLTND	NTSTGTLSEI	RRIYIQNGKV	IQNSVAN	IPG	VDPVNS	ITDN	FCAQKTA
Hir_7A_2xsp	347	DTNS	FENLGG	LNVMGQAFDK	GVVLMVMSVWD	DHEAN	MLWLD	SDYP	TTSSAS	TPGV
Hje_7A_8cel	340	GSS	FSDKGG	LTQFKKATSG	GMVLMVMSLWD	DYYAN	MLWLD	STYP	TNETSS	TPGAV
Pch_7D_1z3v	336	DTNW	FAQKGG	LKQMGEAALGN	GMVLLALS1WD	DHAAN	MLWLD	SDYP	TDKDP	SPGVA
Hir_7A_2xsp	417	SQN	PNSVVF	SNIKIGPIGS	TYTA					
Hje_7A_8cel	409	SQS	PNAKVTF	SNIKFGPIGS	TFSGTS					
Pch_7D_1z3v	406	SQV	PNSQVVF	SNIKFGDIGS	TGNPSG					

Figure 4. Structure-based sequence alignment of the catalytic domains of *H. irregularis* Cel7A, *H. jecorina* Cel7A and *P. chrysosporium* Cel7D. Loop regions with notable variation are shown in boxes and labeled as in Figure 5 where they are highlighted in red. The amino acid numbering is given in the beginning of each row.

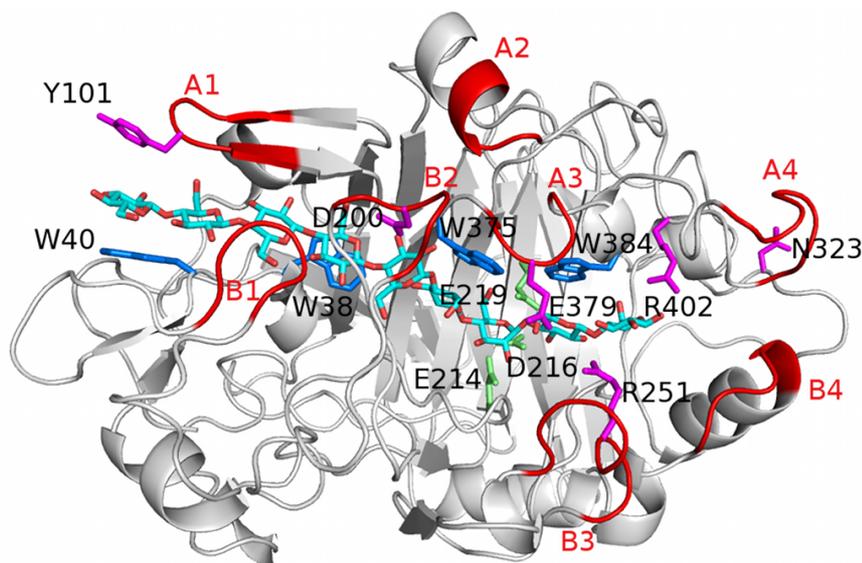
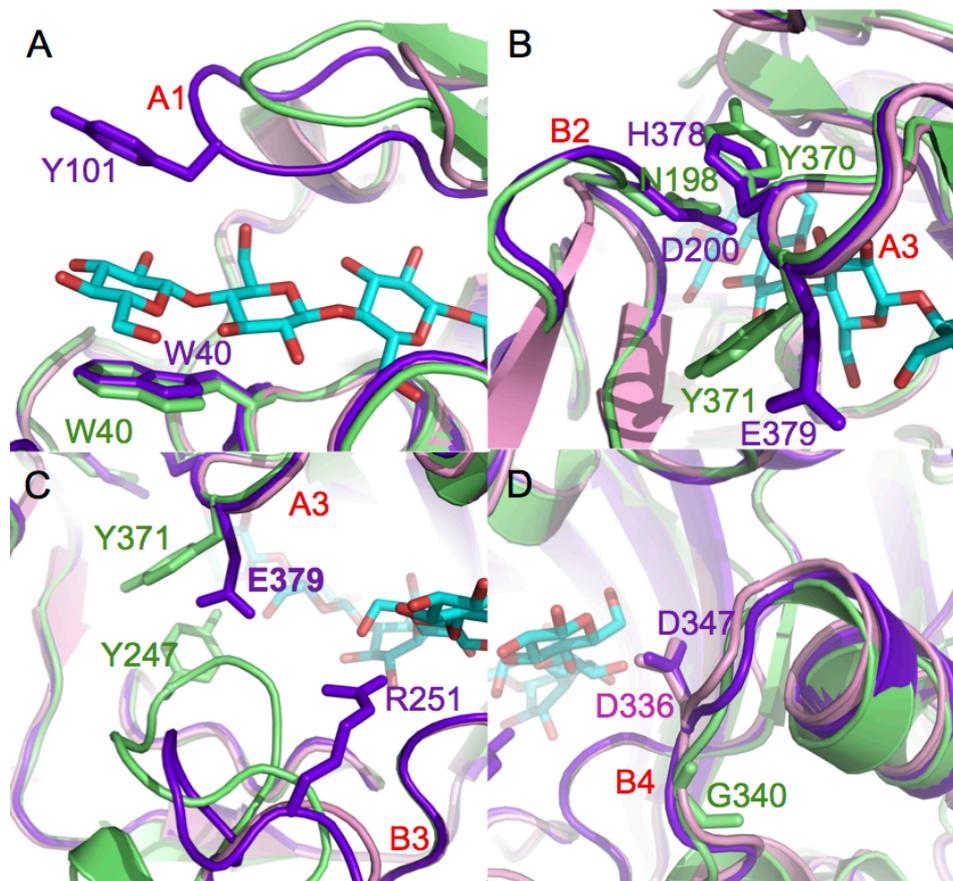


Figure 5. Overall structure of the *HirCel7A* enzyme (SX<sub>5</sub> structure) with a docked ligand from the *H. jecorina* 8CEL structure (Divne *et al.*, 1998). The ligand is shown in cyan, and loops and residues of interest are labeled.

#### *Comparison of the loop above subsite -7*

Loop A1 above the tunnel entrance differs in length and sequence between the three enzymes (Figure 5). *HirCel7A* has the longest loop (aa 98-103), containing a tyrosine residue (Tyr101) at the tip, which is stacking with the cellulose chain at subsite -7, opposite to the Trp40 glucosyl-binding platform. The structure of *MalCel7B* also reveals a tyrosine residue at the tip of the entrance loop, and it has been proposed to be involved in acquisition of a cellulose chain into the tunnel (Parkkinen *et al.*, 2008). The corresponding loop in *HjeCel7A* and *PchCel7D* is shorter by one and four residues, respectively, and consequently these two enzymes are lacking a tyrosine here. Furthermore, the *HirCel7A* structures show two distinct conformations of loop A1, indicating certain flexibility in this region. In *HjeCel7A*, the loop mobility may be more restricted due to presence of a salt-link that anchors the base of the loop (Lys102 and Glu408 in *HjeCel7A*). A corresponding salt-link is not present in *HirCel7A* and *PchCel7D*.



**Figure 6.** **A)** Superposition of loop A1 at the tunnel entrance over subsite -7 from *HirCel7A* (violet), *HjeCel7A* (green), and *PchCel7D* (pink). The *HirCel7A* loop A1 contains a tyrosine residue (Tyr101), not present in *HjeCel7A* or *PchCel7D*. Loop A1 in *PchCel7D* is significantly truncated. **B)** Superposition of loop A3 and loop B2 over the -4 subsite. The three enzymes exhibit different loop-loop contacts and sequence diversity. **C)** Superposition of loops A3 and B3 near the catalytic center subsite -1 and the product sites +1/+2. *HjeCel7A* exhibits a longer loop B3, which forms stable contacts to loop A3 across the catalytic center. The *HirCel7A* loop B3 is two residues shorter and interacts with water with Glu379 on loop A3. *PchCel7D* exhibits the most exposed catalytic center because of a six-residue deletion in loop B3. **D)** Superposition of loop B4 shows the aspartate residue that may interact with the reducing end of the product at subsite -2, which is present in *HirCel7A* (Asp347, D347) and *PchCel7D* (Asp336, D336) but is deleted in *HjeCel7A*.

#### *Comparison of the loops near subsite -4*

The second loop of interest is loop B2 that consists of 13-15 amino acid residues in GH7 CBHs. In the *HirCel7A* structure, Ser199 and Asp200 at the tip of loop B2 interact with His378 on the opposing loop A3 from the other side of the  $\beta$ -sandwich, which effectively encloses the cellulose chain in a tunnel around the -4

subsite (Figure 6B). *HjeCel7A* exhibits the same length for loop B2, where Asn197 and Asn198 at the tip interact with Tyr370 on loop A3. Interestingly, in MD simulations, loop B2 displayed considerably larger fluctuations in *HirCel7A* than in *HjeCel7A*. Higher flexibility is also reflected by the crystal structures. In the *HirCel7A* SX<sub>5</sub> structure, the loop adopts a similar conformation as in *HjeCel7A*. However, in chain B of the Apo structure the loop B2 appears to be opened, since the tip of the loop was disordered and could not be modeled. The reason for this difference in mobility is probably due to the Tyr to His substitution in loop A3 where the larger Tyr residue sidechain enables stronger interaction with loop B2. Actually, loop B2 is recognized as the most flexible region within the entire *HirCel7A* structure, as it exhibited the highest RMSF values (root mean square fluctuations) in the MD simulations.

#### *Comparison of catalytic center loops*

Loop B3, also called the ‘exo-loop’ shows considerable variation among GH7 CBHs. The loops of *HirCel7A* and *PchCel7D* are two and six residues shorter respectively than in *HjeCel7A*. In the latter enzyme, Tyr247 at the tip of loop B3 interacts with Tyr371 on the loop A3 across the tunnel, which covers the catalytic center at subsite -1. The shorter loop B3 of *HirCel7A* adopts a different conformation, which leaves the active center open and exposed to solvent.

For loop A3, the Tyr371 in *HjeCel7A* is replaced by Glu379 in *HirCel7A*, which stretches over the active site towards loop B3, but is not in direct contact (Figure 6C). The side chain of Glu379 appears to be flexible as it displays different rotamers in the three *HirCel7A* structure models. There is also space for Glu379 to bend down towards the cellulose chain. In the MD simulations, it is occasionally within contact distance to OH6 of the D-glucose unit at subsite +1 (data not shown), pointing at a putative role in substrate binding and/or mediation of cellulose sliding over the catalytic center. Glutamate at this position is rare within GH7 and in addition to *HirCel7A* it is only found in few basidiomycete homologs (e.g. *Puccinia graminis* and *Pleurotus ostreatus*).

#### *Product binding site*

Loop A4 (392-403) adjacent to the product binding sites is the second most flexible region in *HirCel7A*, and is more flexible than in the other two enzymes (Figure 5), which may affect product binding and expulsion, although the loop does not interact directly with substrate. This is reflected by ~2.5 Å backbone shifts in loop A4 between the *HirCel7A* crystal structures.

Beyond the reducing end of the cellulose chain, the end of the active site cleft is formed by loop B4. *HjeCel7A* and other GH7 CBHs of *Hypocrea/Trichoderma* species have a one-residue deletion in loop B4. Most other GH7 CBHs have an

additional aspartate residue here (347 in *HirCel7A*; 336 in *PchCel7D*) that points towards the reducing end of a bound cellulose chain, which has been proposed to aid in product expulsion by encouraging cellobiose to tilt away from the catalytic center (Ubhayasekera *et al.*, 2005).

### 2.1.5 Highlights of computational characterization

The MD simulation of *HirCel7A* with the bound cellononaose ligand displays an apparent conformational change in loop A1 compared to the Apo crystal structure, which leads to the binding of Tyr101 to the glucose unit at subsite -7. A cluster representation of ~50 structures, uniformly spaced during the MD simulations, is shown in Figure 7, with color-coding from low (blue) to high (red) RMSF. The three regions associated with the largest fluctuations during the course of simulations, are loop B2 and loop A4 in *HirCel7A* and loop A2 in *PchCel7D*. The ligand fluctuations as a function of binding site from -7 to +2 show significant differences at subsite -7 and -6 between the enzymes, whereas the other subsites are quite similar.

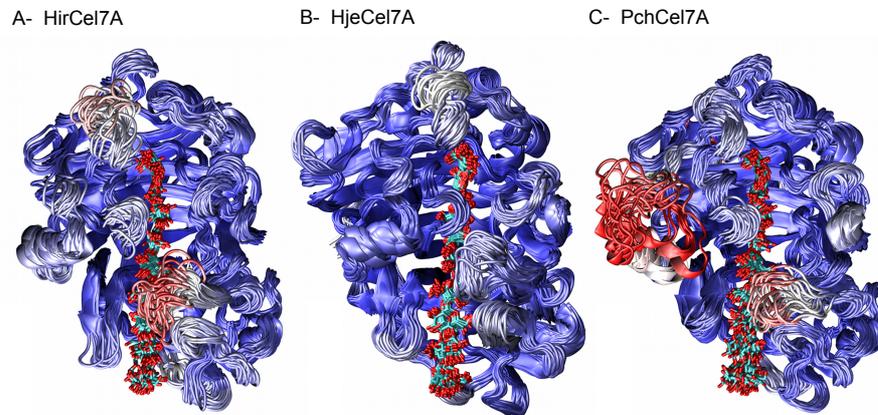


Figure 7. A-C, cluster representations of A) *HirCel7A*, B) *HjeCel7A*, C) *PchCel7A* calculated within a 250-ns trajectory at 5-ns intervals. The enzymes are colored by RMSF, where red indicates the largest fluctuations, and blue indicates the lowest fluctuations.

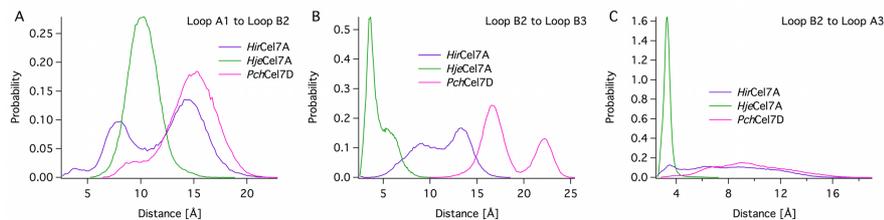


Figure 8. Histograms of the distances of the center of mass from Figure S5 of **A)** Loop A1 to B2, **B)** Loop B2 to B3, and **C)** Loop B2 to A3 for *HirCel7A*, *HjeCel7A*, and *PchCel7D*. These loops correspond to those hypothesized to participate in endo/exo initiation events.

CBHs have often been referred to as exo-cellulases. However, endolytic cleavage has been experimentally shown for GH7 CBHs, demonstrating that tunnel-enclosing loops may open occasionally to allow the enzyme to grab an internal part of a cellulose chain (Ståhlberg *et al.*, 1993; Kurasin & Våljamäe, 2011). Histograms of the minimum distances between opposing loops in the three enzymes (Figure 8) indicate that tunnel opening occurs more frequently in both *HirCel7A* and *PchCel7D* than in *HjeCel7A*.

Overall, the computational and structural studies show that the accessibility of the active site of *HirCel7A* is intermediate between the more closed *HjeCel7A* and the more open *PchCel7D*, suggesting that *HirCel7A* will exhibit intermediate properties e.g. in terms of degree of processivity and endo initiation.

## 2.2 The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies (Paper II).

Capturing a GH structure with its natural substrate productively bound is exceptionally rare. In none of the previous structures of GH7 CBHs with an occupied -1 binding site is the -1 glucose residue productively bound. For example, in the complex of *HjeCel7A* E212Q with two cellotetraose molecule occupying -7 to -4 and -2 to +2 (PDB code 5CEL), the -1 glucosyl unit adopts a ground state  ${}^4C_1$  chair conformation with its anomeric carbon at 7 Å distance to the nucleophile equivalent (Gln212), obviously too far away for nucleophilic attack. However, one structure of a GH7 EG with a thio-oligosaccharide bound with the -1 glucosyl in near boat conformation could be used to build a model of *HjeCel7A* with a productively bound cellulose chain (PDB code 8CEL; Divne *et al.*, 1998). As mentioned above the retaining mechanism of GH7s involves the formation of a covalent glycosyl-enzyme intermediate (GEI). Structures have been reported of a trapped GEI in several other retaining GH families, but not previously in GH7.

This study presents two new structures of *Hje*Cel7A, namely a Michaelis complex with a nine-Glc long cellodextrin spanning the entire tunnel from subsite -7 to +2, and a GEI with a hexasaccharide in subsites -6 to -1 covalently bound to the catalytic nucleophile, and a cellobiose molecule in the product sites +1/+2. These structures reveal details of the ‘static’ reaction coordinate. They were used as starting structures for computational analysis (transition path sampling, TPS) to determine the ‘dynamic’ reaction coordinate, combined with quantum mechanical (QM) and molecular mechanics (MM) calculations to calculate free energy profiles and activation energy barriers for the catalytic cycle of the enzyme.

### 2.2.1 Crystallization and structural determination

The Michaelis complex structure was obtained by co-crystallisation of *Hje*Cel7A acid/base mutant E217Q with an insoluble cellooligosaccharide preparation with Glc<sub>10</sub> as major component (provided by Prof. Jurgen Puls, University of Hamburg; Figure 9A). The GEI was trapped in a structure by incubating *Hje*Cel7A E217Q with the mechanism-based suicide inhibitor 2,4-dinitrophenyl 2-deoxy-2-fluoro- $\beta$ -cellotrioside (DNP-2F-G3; Figure 9B) prior to crystallization setup.

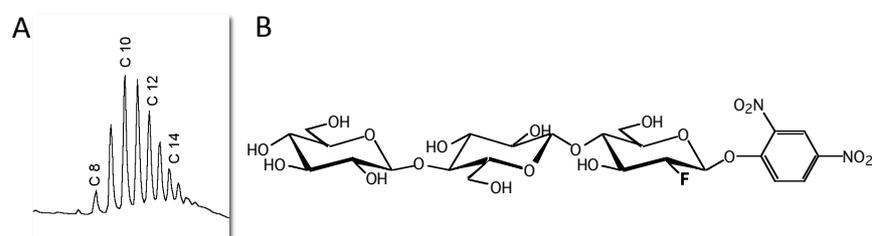


Figure 9. **A)** Insoluble cellooligosaccharides with Glc<sub>10</sub> as the major component. **B)** The mechanism-based inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro- $\beta$ -cellotrioside (**DNP-2F-G3**).

The Michaelis and GEI structures were refined at 1.45 and 1.32 Å resolution and  $R_{work}/R_{free}$  of 0.171/0.196 and 0.168/0.187, respectively, and they have been deposited with PDB accession codes 4C4C and 4C4D. The overall structures are shown in Figure 10.

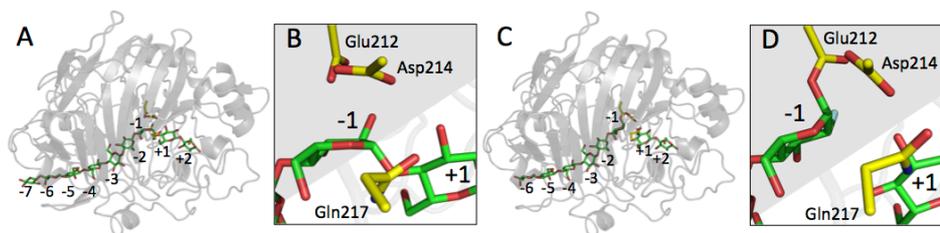


Figure 10. **A)** Overall structure of the Michaelis complex. **B)** Active site of the Michaelis complex suggests a  ${}^4E$  envelope conformation of the glucopyranoside unit in subsite -1. **C)** Overall structure of the Glycosyl-Enzyme Intermediate (GEI). **D)** Schematic view of the active site of the GEI structure. Detailed comparison with the Michaelis complex in fig 10B reveals a dihedral rotation of the Glu212 sidechain upon formation of the covalent bond with the anomeric carbon.

### 2.2.2 Michaelis complex structure

The Michaelis complex structure is the first of a GH7 CBH, and shows a fully intact cellonanoase chain spanning subsites -7 to +2. Two apparent twists of the cellulose chain are identified within the 50 Å tunnel at subsites -4/-3 and -3/-2 that prime the cellulose for catalytic hydrolysis. The celooligomer of the theoretical model (8CEL) is very similar to the Michaelis complex in this study (Figure 10 A, B).

The glucose residue in site -1 adopts a  ${}^4E$  envelope conformation with its anomeric carbon 3.4 Å from the nearest carboxylate oxygen of the catalytic nucleophile Glu212 and its C2 hydroxyl at hydrogen bonding distance from the other carboxylate oxygen of Glu212. The mutated acid/base residue Gln217 displays double conformations where the dominant conformation of the side chain is hydrogen bonded to the -1/+1 glycosidic oxygen and an alternate form turns towards the +1 glycosyl residue (Figure 11 a).

### 2.2.3 Glycosyl-enzyme intermediate structure (GEI)

The second structure is the GEI of the *HjeCel7A* mutant E217Q with a hexasaccharide in sites -6 to -1 covalently linked to the catalytic nucleophile Glu212, and a cellobiose molecule in +1/+2. The density of the -1 unit is weaker than for the other sugars, thus it seems that the crystal contains two populations of structures, a major form with a covalently linked 2-fluoro-glucoside moiety in site -1, and a minor form wherein the -1 site is empty. The -1 glucosyl is a  ${}^4C_1$  chair, which is  $\alpha$ -linked to the nucleophile Glu212. The two populations are reflected by dual conformations of all three residues of the catalytic triad (Figure 11 b). The cellobiose product has turned slightly outwards with respect to the Michaelis complex, but it is still in the ‘unprimed GEI’ mode, as explained below.

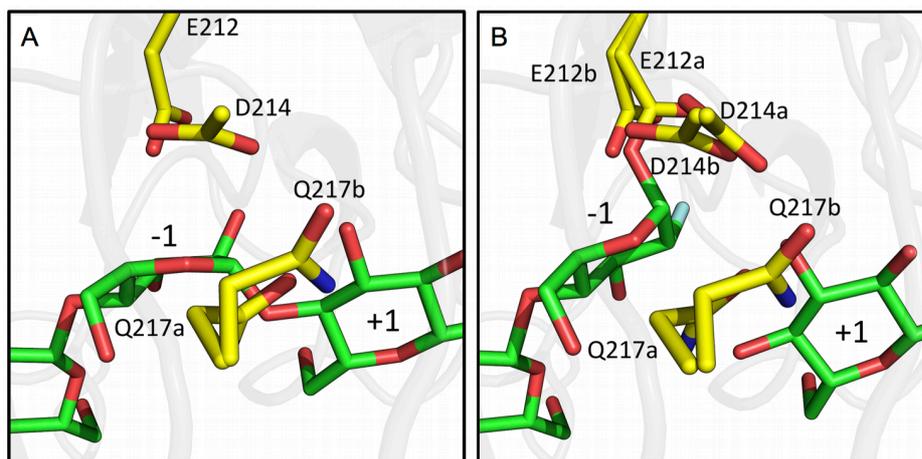


Figure 11. Alternate conformations of the catalytic residues. **A)** Q217 exhibits dual conformations in the structure of the Michaelis complex. **B)** All three catalytic residues display double conformations in the GEI structure.

### 2.2.4 Computational highlights

The Michaelis and GEI structures are essentially identical except for a 30° rotation of the carboxyl group of the nucleophile Glu212, a small shift of the sugar unit in subsite +1 and significant changes of the -1 sugar ring.

To fully elucidate the ‘dynamic’ reaction coordinate (RC) for *HjeCel7A*, QM/MM molecular dynamics simulations of cellononaose hydrolysis were performed. Using the Aimless Shooting (AS) method of Transition Path Sampling (TPS), 21,000 trajectories were harvested for Step 1 (see the Supporting Information file of Knott *et al.*, 2014 for full description). Snapshots from a candidate trajectory for the glycosylation step (Step 1) are illustrated in Figure 12 A-C for the reactant, transition state and product, respectively. Figures 12 D and E show a schematic representation of the process, and the computed free energy profile along the RC of Step 1, respectively. The computational procedure for the deglycosylation step (Step 2) closely follows that of Step 1 and the results are illustrated similarly in Figure 13.

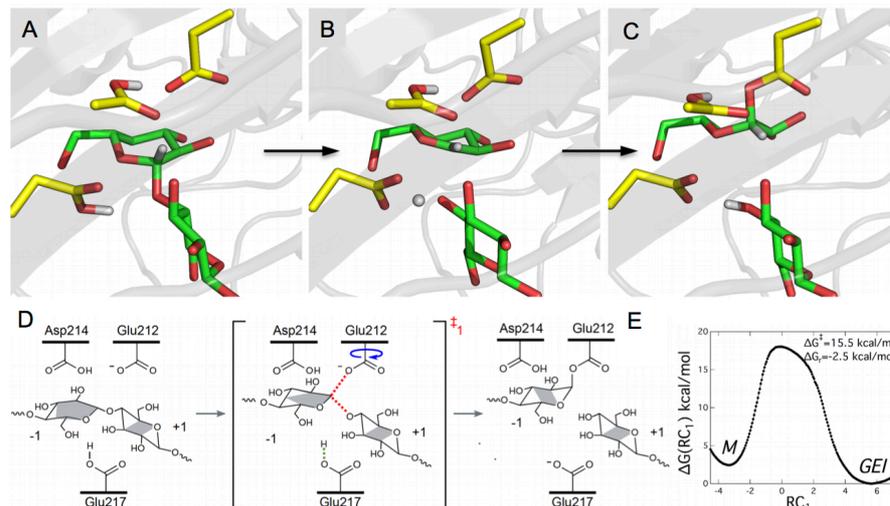
After Step 1, there is not sufficient space for a water molecule to attack the anomeric carbon of the -1 glucopyranose ring to complete the catalytic cycle. In the literature, reaction schemes of the retaining mechanism usually depict that the product dissociates from the enzyme after the glycosylation step to allow access of water. However, Ubhayasekera *et al.*, (2005) recognized the presence of two distinct binding modes of the cellobiose product in GH7 crystal structures, one of which leaves room for a nucleophilic water molecule, and proposed that product release might not happen until after the deglycosylation step. The two binding

modes for the cellobiose product may correspond to an ‘unprimed GEI’ and a ‘primed GEI’ mode, respectively (Figure 14 A, B). The product in the GEI structure and at the end of the Step 1 simulation binds in the unprimed GEI mode observed in some previous crystal structures. In the primed GEI mode, there is sufficient space for a water molecule to reside between the enzyme-bound substrate and the product. Thus, it is hypothesized that between Step 1 and Step 2 of the catalytic cycle, the cellobiose product translates towards the tunnel outlet, creating space for a water molecule to move into the active site near Glu217 and serve as the Step 2 nucleophile.

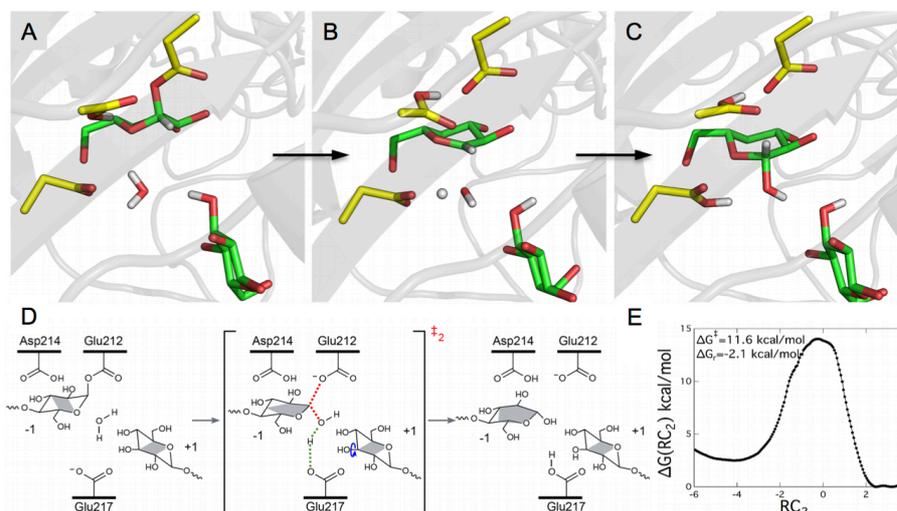
To connect the Step 1 product to the Step 2 reactant, MD simulation was performed where the catalytic water is pulled into the active site and the cellobiose product translates away from the active site towards the binding tunnel exit. In the unprimed mode, the C3 hydroxyl of the +1 glucosyl ring forms a hydrogen bond to the catalytic water. As the water molecule moves towards the -1 anomeric carbon, this hydrogen bond is broken resulting in a rotation of the C3 hydroxyl. This dynamical transition is a crucial reaction coordinate component for deglycosylation. Via umbrella sampling, the free energy (potential of mean force, PMF) between the unprimed GEI and primed GEI is calculated along the water distance coordinate, revealing a 2.0 kcal/mol barrier, and essentially equal stability in the unprimed GEI and primed GEI modes (Figure 13C).

Enabled by an accurate RC, free energy and reaction rate calculations can be employed in a kinetically meaningful manner. For step 1, the free energy barrier height is 15.5 kcal/mol and the difference between reactants and products is -2.5 kcal/mol downhill. The predicted reaction rate constant is  $10.8 \text{ s}^{-1}$ , which is in good agreement with the rate of processive hydrolysis by *HjeCel7A* on crystalline cellulose estimated at  $7.1 \pm 3.9 \text{ s}^{-1}$  using high-speed atomic force microscopy (Igarashi *et al.*, 2011). The step 2 free energy profile reveals an 11.6 kcal/mol barrier for deglycosylation. The products are stabilized by -2.1 kcal/mol relative to reactants suggesting an overall reaction free energy of -4.2 kcal/mol. The predicted reaction rate constant for Step 2 is  $5.29 \times 10^3 \text{ s}^{-1}$ .

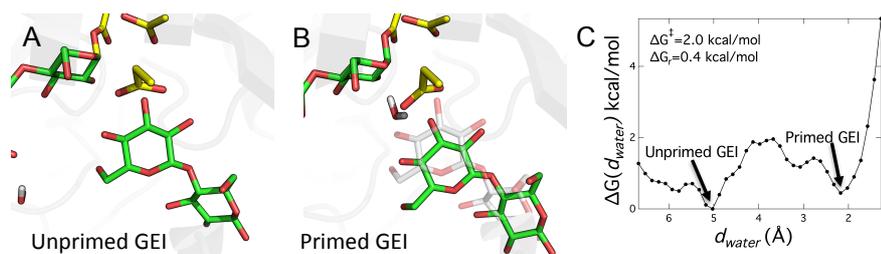
The results reveal a product-assisted mechanism for the deglycosylation step as well as the first comparison of the rate constants for both steps of the catalytic cycle. Step 1 is slower than Step 2 by nearly three orders of magnitude; supporting the hypothesis that glycosylation is the rate-limiting step on natural substrates in these enzymes.



**Figure 12.** **A)** Snapshot of the reactant conformation from a representative AS trajectory for the glycosylation step. **B)** Snapshot of the transition state. **C)** The product of glycosylation reaction indicates the glycosyl-enzyme intermediate along -1 sugar in the  $C_1$  conformation. **D)** Schematic illustration of the overall glycosylation reaction by the collective variables recognized by LM color-coded at the transition state. **E)** Reaction free energy and barrier for glycosylation step, M denotes the Michaelis complex.



**Figure 13.** **A)** Snapshot of the reactant conformation from a representative AS trajectory. **B)** A representative of the transition state. Note the distorted conformation of the -1 sugar, as the nucleophilic water molecule is ripped apart. **C)** A snapshot of the product illustrating that glycosyl enzyme bond has been broken, and the catalytic amino acids are restored. **D)** Schematic presentation of the overall deglycosylation step with collective variables suggested by LM as color-coded at the transition state. **E)** Reaction free energy and barrier for deglycosylation step.



*Figure 14.* Cellobiose product transition between the unprimed to primed GEI along approaching the the nucleophilic water. **A)** Snapshot of the *Hje*Cel7A GEI obtained by AS. The nucleophile water in the vicinity of the active site is also shown. **B)** Snapshot of the ‘primed GEI’ from an umbrella sampling simulation. The cellobiose product shifts downward in to the binding tunnel exit in concert with this water movement. **C)** The free energy as a function of the distance from the nucleophilic water to the active site displays only a small barrier between these two states.

## 2.3 Expression, crystal structure and cellulase activity of the thermostable Cellobiohydrolase Cel7A from the fungus *Humicola grisea* var. *thermoidea* (paper III)

The economy of industrial scale enzymatic conversion of biomass to fermentable sugars would benefit from an improved thermostability of the enzyme mixtures used (Viikari *et al.*, 2007), since the lifetimes of the cellulases are expected to increase with thermostability. Thus, thermostable cellulases are good candidates to be used for industrial biomass conversion since higher thermal stability could lead to higher specific activity at elevated temperatures and shorter hydrolysis time. As mentioned above, *HjeCel7A* is the major enzyme and the key rate-limiting component in commercial enzyme cocktails for biomass conversion. Although this enzyme is fairly thermostable with a  $T_m$  of 62.5 °C, it is less thermo-tolerant than other major enzymes of the enzyme cocktail. Therefore, it is of interest to search for GH7 homologues with higher thermostability and equal or superior hydrolytic properties.

However, enzymes of GH7 have been notoriously difficult to express recombinantly. Numerous trials in bacterial hosts have failed and expression in yeast has often resulted in very low yields and/or only partially active enzyme. Most of the recombinant GH7s reported to date come from a limited number of expression systems developed in a few filamentous fungi. This may be related to the fact that while GH7 CBHs are ubiquitous in cellulolytic fungi, they have only been found in a few other eukaryotes and do not appear to exist at all in the Bacteria and the Archaea kingdoms.

The thermophilic fungus *Humicola grisea* var. *thermoidea* has shown to be an efficient producer of a hydrolytic system of thermostable enzymes including cellulase (Ferreira-Filho *et al.*, 1996), trehalase (Zimmermann *et al.*, 1990), glucoamylase (Campos & Felix, 1995) and xylanase activities (Da Silva *et al.*, 1994). Among cellulases, the fungus has been shown to produce different CBHs and EGs with pronounced activity at elevated temperatures (Takashima *et al.*, 1996). The only three-dimensional structure available previously from *H. grisea* is that of EG Cel12A (Sandgren *et al.*, 2003).

### 2.3.1 Expression of GH7 CBHs

In this study, a host/vector system was developed for heterologous expression in the filamentous fungus *Aspergillus niger* var. *awamori* AP4. Gene specific primers were then used against genomic DNA isolated from a diverse set of fungi, to amplify GH7 CBH-encoding genes for expression in this system. Ten cloned Cel7 genes, including *Hypocrea jecorina* Cel7A (*HjeCel7A*) as reference, were successfully expressed (paper III, Table 1). The homologs share 56-97% protein sequence identity with *HjeCel7A*. The genes were expressed under the control of a

constitutive promoter in order to minimize the background of host proteins and potential interference from other carbohydrases. Consequently, the Cel7 enzymes from shake flask cultivations could be purified to apparent homogeneity in a single hydrophobic interaction chromatography step.

### 2.3.2 Biochemical characterization

#### *Protein melting temperature ( $T_m$ )*

Thermostability was examined by monitoring the thermal denaturation of the *A. niger*-expressed enzymes using CD spectroscopy in order to estimate protein-melting temperatures ( $T_m$ ). The most stable of the homologues, *Humicola grisea* var. *thermoidea* Cel7A, *Hgr*Cel7A, exhibits 10 °C higher melting temperature ( $T_m$  72.5 °C) than *H. jecorina* Cel7A.  $T_m$  values for the expressed Cel7 homologues are presented in Paper III, Table 1.

#### *Enzymatic activity on PASC and pretreated corn stover*

Comparison of the activity of *Hgr*Cel7A and *Hje*Cel7A when acting alone on phosphoric acid swollen cellulose (PASC) reveals a much higher hydrolytic rate for *Hgr*Cel7A at both high (65 °C; ~4.8-fold higher initial rate) and moderate temperature (38 °C; ~3.3-fold higher) as shown in Paper III, Figure 2.

The *Hgr*Cel7A was also assayed for cellulosic conversion performance on an industrially relevant biomass material, thermochemically pretreated corn stover (PCS), at elevated temperature (65 °C for 24 h). The *Hje*Cel7A was included for comparison as well as three less thermostable homologs, to range from the lowest to the highest  $T_m$  value. This assay combines the Cel7 sample to be tested at about 1:1 mass ratio with proteins from the growth of a *H. jecorina cbh1* delete strain, *i.e.* lacking native Cel7A due to disruption of the *cbh1* gene. The performance on PCS at 65 °C correlates with the  $T_m$  values of the Cel7 enzymes, and the highest cellulose conversion was indeed obtained with *Hgr*Cel7A (Paper III, Figure 3). A 75% higher yield of soluble sugar clearly demonstrates that *Hgr*Cel7A performs better than *Hje*Cel7A at high temperature.

### 2.3.3 Crystallization and structure determination of *H. grisea* Cel7A

In order to produce large amount of protein for structural studies, the *Hgr*Cel7A was also expressed under control of the *cbh1* (Cel7A) promoter in an engineered *H. jecorina* strain that is devoid of production of the four major native cellulases, Cel5A, Cel6A, Cel7A and Cel7B. The C-terminal linker-CBM1 was proteolytically removed from full-length *Hgr*Cel7A using papain, and the isolated catalytic domain (CD) was subjected to crystallization experiments. Crystals were obtained of space group  $P2_12_12_1$  with two molecules in the asymmetric unit. The structure of *Hgr*Cel7A was refined at 1.8 Å resolution and final  $R_{work}/R_{free}$  of 0.167/0.210.

### 2.3.4 Overall structure of *HgrCel7A* and comparison with other CBHs

The structure of the *HgrCel7A* CD is very similar to other GH7 CBHs along the  $\beta$ -sandwich core, as reflected by RMSD values of 1.0-1.2 Å over matching Ca positions between *HgrCel7A* (chain A) and *HjeCel7A*, *PchCel7D*, *MalCel7B* and *RemCel7A*. Nearly all amino acids identified by Divne *et al.* (1998) as important for cellulose binding are conserved at similar positions. As with *HirCel7A* in Paper I, major differences presumably related to function occur at four regions along the cellulose-binding path of the enzyme: the tunnel entrance at subsite -7/-6 (loop A1; Figure 15B), loop contacts around subsite -4 (loop B2; Figure 15C), near the catalytic center (loop B3; Figure 15D), and adjacent to the product binding subsites (Figure 15E).

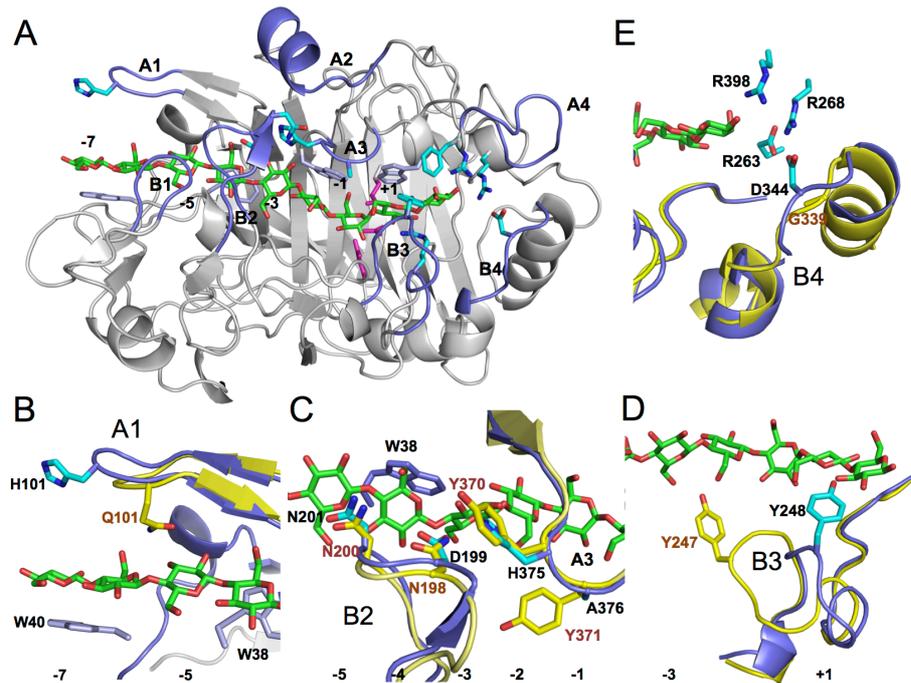
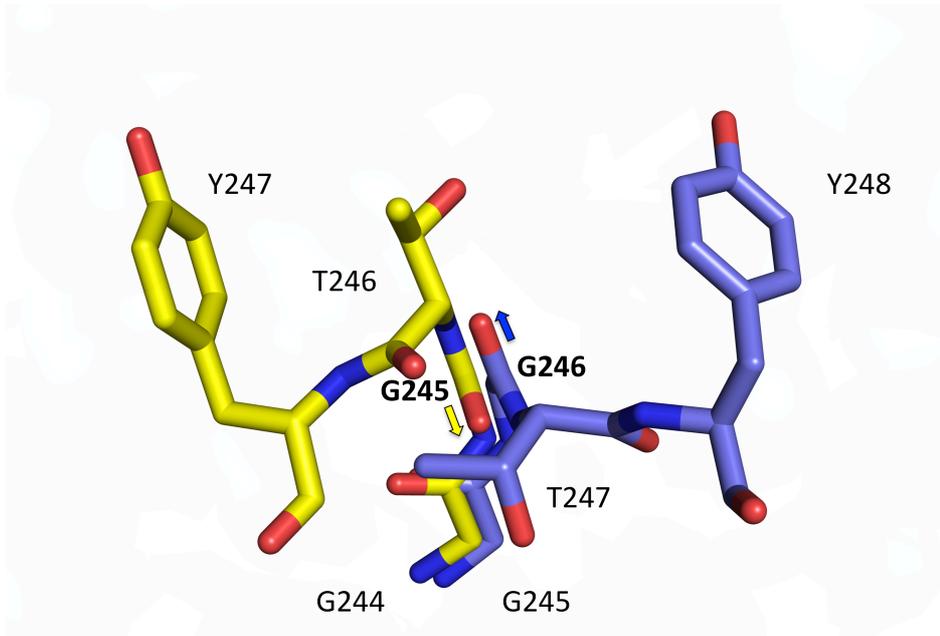


Figure 15. **A)** Structure of *HgrCel7A* (A chain) with a cellulose chain from the *HjeCel7A* 8CEL structure superimposed (Divne *et al.*, 1998). The ligand is highlighted in green and residues of interest are labeled. **B)** Superposition of loop A1 at subsite -7 of *HgrCel7A* (blue) and *HjeCel7A* (yellow). In *HgrCel7A*, His101 is positioned at the tip of loop A1, which is one residue longer than the corresponding loop in *HjeCel7A*. **C)** Superposition of loops A3 and B2 over subsite -4, where His375 and Ala376 are present in *HgrCel7A* instead of Tyr370 and Tyr371, respectively. **D)** Loop B3 of *HgrCel7A*, adopts a new conformation with Tyr248 at the tip pointing into subsite +2. **E)** Loop B4 has a one-residue insertion, of Asp 344, relative to *HjeCel7A*.

#### *Tunnel entrance at subsite -7/-6*

Loop A1 at the entrance to the tunnel shows variations both in terms of length and sequence among GH7 CBHs. In *HgrCel7A*, His101 at the tip of this loop is located near the -7 glucosyl unit and may guide a cellulose chain into the tunnel, in analogy with the role proposed for tyrosine present at the corresponding position in *MalCel7B*, *LquCel7B* and *HirCel7A* (Voutilainen *et al.*, 2008; Kern *et al.*, 2013; Momeni *et al.*, 2013). Flexibility in loop A1 has been observed among GH7 CBH structures. The conformation of the loop in *HgrCel7A* when in solution may differ somewhat from the crystal structure since it seems to be affected by the crystal packing. In both chain A and B, the A1 loop sticks into the tunnel and occupies the -7 subsite of the other protein molecule in the asymmetric unit.



*Figure 16.* Superposition of exo-loop in *HgrCel7A* (chain A) and *HjeCel7A* (4C4C) along aa 245-248. Gly246 in *HgrCel7A* is turned almost 180° about the psi angle compared to Gly245 in *HjeCel7A* as shown via arrows.

#### *Loop contacts near the -4 subsite and the catalytic centre*

Loop B2 folds over the  $\beta$ -sandwich core to build up the roof of the tunnel (Figure 15C). The loop appears to exhibit considerable flexibility in *HgrCel7A*, as indicated by the crystal structure. In chain A, loop B2 displays the highest B values for main chain atoms, and in chain B, the loop is open and disordered. Residues 193-200 had to be excluded from chain B of the refined model due to insufficient electron

density. The B3 loop of *HgrCel7A* also exhibits somewhat elevated B-factors, although considerably less than loop B2 (Figure 15D). Loop B3 adopts a new conformation where Tyr248 points into subsite +2 of the active site, which has not been seen previously in any GH7 structure. In other GH7 CBH structures with the same length of loop B3, the tip of the loop points towards the catalytic center (Figure 16). In *HjeCel7A*, Tyr247 at the tip interacts with both the substrate in the -2 subsite and with Tyr371 on loop A3 across the binding tunnel.

Loops B2 and B3 of *HgrCel7A* exhibit the same length, similar sequence and similar surroundings as in *HjeCel7A*. This suggests that the reasons for difference in behavior may not reside within the loops themselves. Rather we believe that the dynamics of these loops are primarily governed by their interaction opportunities across the active site. In particular two residues at the tip of the opposing loop A3 appear to play an important role here. In *HjeCel7A*, tyrosines 370 and 371 of loop A3 interact with the tips of loops B2 (Asn198) and B3 (Tyr247), respectively (Figure 15C).

Loops B2 and B3 of *HgrCel7A* exhibit the same length, similar sequence and similar surroundings as in *HjeCel7A*. This suggests that the reasons for difference in behavior may not reside within the loops themselves. Rather we believe that the dynamics of these loops are primarily governed by their interaction opportunities across the active site. In particular two residues at the tip of the opposing loop A3 appear to play an important role here. In *HjeCel7A*, tyrosines 370 and 371 of loop A3 interact with the tips of loops B2 (Asn198) and B3 (Tyr247), respectively (Figure 15C).

The corresponding residues in *HgrCel7A* are His375 and Ala376. His375 is in contact with loop B2 (Asp199) in chain A, but not in chain B where loop B2 appears to be open. A histidine is also found in the same position in *HirCel7A*, where MD simulations show larger fluctuations in loop B2 and more frequent tunnel opening relative to *HjeCel7A*, primarily because of a stable hydrogen bond to Tyr370 in the latter enzyme (Momeni *et al.*, 2013). MD simulations of *Trichoderma harzianum* Cel7A (*ThaCel7A*) and *HjeCel7A* also point to the importance of loop A3 for the mobility of loop B3 (Textor *et al.*, 2013). These enzymes share over 80% sequence identity but Tyr371 in loop A3 of *HjeCel7A* is replaced by an alanine in *ThaCel7A* (as in *HgrCel7A*). In *HjeCel7A* the loops remain in contact throughout the MD simulation, whereas in *ThaCel7A*, loop B3 shows larger fluctuations and is frequently opened and the active site is exposed.

It is noteworthy that in the *HgrCel7A* structure, Tyr248 at the tip of loop B3 partially obstructs the +2 subsite. At least the Tyr248 side chain needs to retract some 1-2 Å from subsite +2 during enzyme action on cellulose. We modeled the B3 loop of *HgrCel7A* onto that of *HjeCel7A*, i.e. in the '-1 position' and it seems to fit well into the *HgrCel7A* structure without any steric hindrance, which make us

believe that loop B3 is flexible and can switch between these conformations in *HgrCel7A*. Furthermore, superposition of the structures indicates that loop B3 may be able to adopt the '+2 position' in other Cel7 homologs as well, including *HjeCel7A*, *MalCel7B* and *RemCel7A*. The shift is accomplished by rotation about the psi angle of Gly246, by almost 180°, relative to Gly245 in *HjeCel7A*. This conserved glycine residue seems to act as a hinge that makes the peptide chain proceed in the opposite direction (Figure 16).

Our *HgrCel7A* structure points to a new alternate conformation of loop B3 and a putative conformational switch within homologous GH7 CBHs. However, further studies are needed to investigate how often such conformational changes may occur in different enzymes and to elucidate possible connections with enzyme action.

#### *Product-binding region*

The product-binding region is well conserved relative to other GH7 CBHs apart from the conformational change of loop B3 discussed above, and the substitution of Phe386 in loop A4 in vicinity of subsite +2, where a tyrosine residue is present in other structures and the majority of GH7 CBH sequences. Loop B4 defines the end of the active site and exhibits similar sequence and structure as in other GH7s. The side chain of Asp344 in loop B4 points towards and can hydrogen bond to the reducing end of the cellulose chain at subsite +2. An aspartate is conserved here in most GH7 CBH sequences, but is missing in *Hypocrea/Trichoderma* species due to a one-residue deletion in loop B4 (Figure 15E).

#### **2.3.5 Potential thermostabilizing interactions in *HgrCel7A***

The structure of *HgrCel7A* indicates considerable mobility of surface loops and yet the enzyme is ~10 °C more thermostable than *HjeCel7A*. The *HgrCel7A* thus seems to be able to avoid that loop movements propagate into the core of the protein structure towards irreversible protein unfolding. Upon closer examination of the base of certain loops, i.e. the regions where they connect to the secondary structure framework, some potentially stabilizing interactions were recognized.

At the base of loop B1, Gln43 and Ile60, make a hydrophobic interaction interface that may have a stabilizing effect primarily on the 43-48 region, which appears to be rather loosely connected at the surface of the protein near the tunnel entrance.

The long and remarkably mobile loop B2 is anchored by a salt bridge between Glu191 and Arg206 at the N- and C-termini of the loop. The glutamine is conserved in most of the structures, but an arginine at this position is unique to *HgrCel7A*. The Arg206 is further involved in a salt bridge with Asp240 at the base of loop B3, crosslinking these regions. Thus, it may have a crucial stabilizing role in *HgrCel7A*.

The side chains of Val434 and Leu437 (glycine and serine in *HjeCel7A*) at the C-terminus of the catalytic domain form a hydrophobic cluster together with Val290, Phe307 and Ile314. This indicates that the linker peptide is more firmly anchored and that the native, full-length *HgrCel7A* may tolerate larger dynamics of the linker-CBM tail without propagation of unfolding into the core of the catalytic domain.

## 2.4 Structural insights into the inhibition of Cellobiohydrolase Cel7A by xylo-oligosaccharides (paper IV)

Enzymatic decomposition of lignocellulose in industrial scale is a complex process involving enzyme cocktails with multiple components, heterogeneous insoluble substrates and various soluble products and other compounds, and the composition changes as the reaction proceeds. It has been recognized that enzymes are sensitive to inhibition by various components, present in the biomass material, released by thermochemical pretreatment and/or during the enzymatic saccharification, and by the accumulation of hydrolysis products (Jonsson *et al.*, 2013). For example, xylan and xylooligosaccharides (XOS) liberated during the pretreatment have been demonstrated to have strong inhibitory effects on GH7 CBHs (Holtzapfle *et al.*, 1990; Zhang & Viikari, 2012; Baumann *et al.*, 2011; Kont *et al.*, 2013). To elucidate the mechanism behind this inhibition at a molecular level, we have determined structures of *HjeCel7A* in complex with xylooligosaccharides.

### 2.4.1 Experimental procedures

Crystallization was performed using wild type and catalytically deficient variants E212Q and E217Q of *HjeCel7A*, without addition of ligand. Grown crystals were thereafter soaked with 5-10 mM of xylose (X1), xylobiose (X2), xylotriose (X3), xylotetraose (X4) and xylopentaose (X5) prior to x-ray diffraction data collection.

In addition, inhibition experiments were performed by measuring the activity of intact *HjeCel7A* on *p*-nitrophenyl lactoside (*p*NP-lac; 0.1 to 5.0 mM) as substrate in the presence and absence of XOSs. The yellow color of the liberated *p*-nitrophenol was measured at 400 nm in a microtitre plate reader. Michaelis-Menten kinetic parameters ( $K_m$ ,  $V_{max}$ ) and mixed-type inhibition constants  $K_{ic}$  and  $K_{iu}$  were derived by non-linear regression using the software Simfit.

### 2.4.2 Overview of the crystallographic structures

No xylose was seen in the structures from crystals soaked with X1, but rather surprisingly, lactose was found in the product binding sites +1/+2, probably carried over from the fungal cultivation. Several structures were obtained from crystals with X2, both co-crystallized and soaked, but in none of them xylobiose could be

identified in the electron density maps. Instead cellobiose was bound in the +1/+2 product sites. We could trace the origin of this cellobiose to the commercial X2 material, where the concentration of cellobiose was determined to be approximately 150  $\mu\text{M}$  in a 20 mM X2 stock solution (data not shown). When X3 was added for soaking, many crystals did crack within minutes and most crystals tested for x-ray diffraction exhibited severe disorder and could not be used for structure determination. However, by trying several crystals after short soaking times, structures could be obtained with X3 bound in the beginning of the tunnel of the enzyme, with both the E212Q and E217Q variants. Crystals soaked with X4 and X5 did not show a similar tendency to crack. They appeared to maintain their integrity even after overnight soaking and structure complexes with both of these ligands could be obtained.

In total, six *HjeCel7A*-XOS complex structures, refined at 1.3-1.9 Å resolution, are reported in this study. The overall protein structures are essentially identical with each other as shown by 0.09-0.14 Å pairwise RMS deviations of C $\alpha$  carbons. In both the E212Q<sub>X3</sub> and E217Q<sub>X3</sub> structures, X3 binds in the beginning of the tunnel, approximately spanning subsites -6 to -3. In the E217Q<sub>X4</sub> structure, X4 overlaps with the position of X3 and with the fourth xylopyranose unit closer to the tunnel entrance, approximately at the -6 subsite. There is also a low occupancy binding of a xylo-oligosaccharide at the -2/-1 subsites in E217Q<sub>X4</sub>, a saccharide that appears to be open at the reducing end (Figure 17). With X5, the electron density maps show four xylopyranosyl units binding similarly as X4, whereas the density is too weak for modeling a fifth Xyl unit. In the WT<sub>X5</sub> structure, there is additional density for two Xyl units in subsites -2/-1, but in contrast to the E217Q<sub>X4</sub> structure, the -1 xylosyl unit exhibits a closed pyranoside conformation.

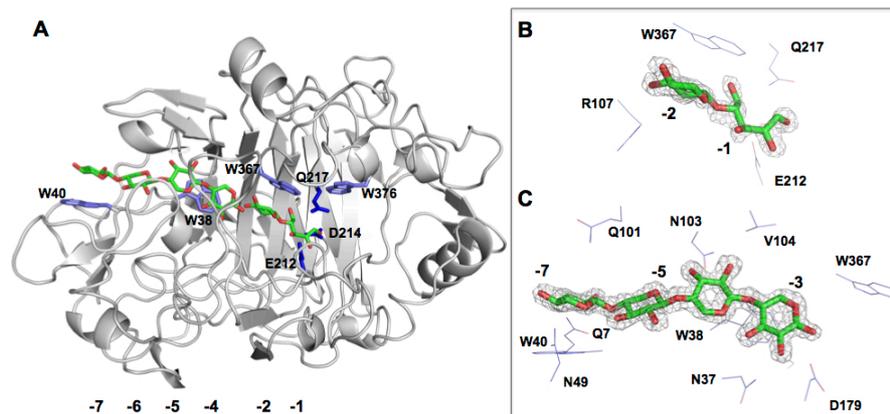
All modeled XOSs are well supported by electron density and the binding is essentially identical for overlapping xylopyranose units. Furthermore, the xylo-oligomers bind with the same directionality and overall orientation as has previously been reported for cellodextrins (e.g. 5CEL; 6CEL; 7CEL; Divne *et al.*, 1998), but out of register at subsites -6 to -3.

### 2.4.3 Tunnel-entrance binding sites

Clear electron density is observed for a bound XOS at the beginning of the tunnel in all the six structure complexes. The electron densities show three (E212Q<sub>X3</sub>, E217Q<sub>X3</sub>) or four (E217Q<sub>X4</sub>, E212Q<sub>X5</sub>, E217Q<sub>X5</sub>, WT<sub>X5</sub>) xylopyranosyl residues bound in the same manner, indicating a preferred binding mode of xylopyranosyl sugars. The first and innermost of the xylose units is located about halfway between subsites -4 and -3 for glucose units of cellodextrins. The side chains of Asp179 and Asn37 hydrogen bond with OH2 and OH3 of the sugar. The second xylopyranosyl unit, between subsites -5 and -4, makes hydrogen bonds to the other side of the

tunnel. The OH2 and OH3 hydroxyls bind to the backbone amide and carbonyl groups of Val104, and to the side chain carbonyl of Asn103. A third xylopyranoside is located between subsites -6 and -5. It has no direct polar contacts, only van-der-Waals and solvent-bridged contacts with the enzyme. The E217Q<sub>X4</sub>, E212Q<sub>X5</sub> and E217Q<sub>X5</sub> structure complexes contain one additional xylopyranosyl residue near Trp40. The sugar ring is nearest to and stacks with the pyrrole ring of Trp40, in contrast to the cellodextrin binding to *Hje*Cel7A (pdb ID 4C4C), where the glucopyranosyl ring bound at subsite -7 shows ring stacking mainly with the benzene ring of Trp40.

We observe the same binding pattern for XOS in the different structure complexes, with a registry shift where each xylopyranose residue is positioned 2.8-3.1 Å further into the tunnel compared to cellodextrin binding (Figure 18).



**Figure 17.** **A)** Overall structure of E217Q<sub>X4</sub> bound with X4 and X2 moiety. The backbone of the catalytic domain shown in light grey, tryptophan binding platform and catalytic residues are in purple and dark blue, respectively. X4 bound at the entrance of tunnel and xylobiose at -1/-2 sites. X4 is in red- phosphoric green. **B)** Strong positive electron density (in grey) in the Fo-Fc difference map for xyp-xls in binding sites -2/-1 and **C)** X4 in sites -7/-3.

#### 2.4.4 Second binding mode observed in X4 and X5 complexes

In addition to the XOS molecule observed at the beginning of the tunnel, partial occupancy of two consecutive xylose residues in subsites -1 and -2 could be seen in the E217Q<sub>X4</sub> and WT<sub>X5</sub> complexes. This suggests an alternate binding mode for XOS in the vicinity of the catalytic center. In both cases only two xylose residues could be clearly seen of this second binding mode. Interestingly, the xylose unit at subsite -1 differs in conformation between the two structures. In E217Q<sub>X4</sub>, the electron density at subsite -1 did not fit a pyranose/hemiacetal representation of xylose, but rather an open aldehyde form, whereas the -1 xylosyl in E217Q<sub>X5</sub>, exhibits a regular ring-closed pyranoside conformation (Figure 19).

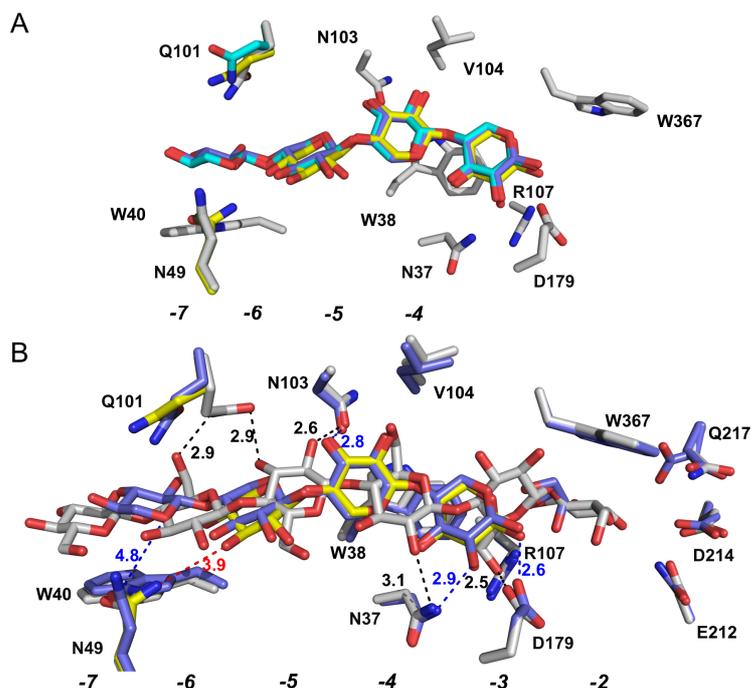


Figure 18. **A)** Comparison of X3, X4 and X5 in E217Q and E212Q structures colored in yellow, cyan and purple, respectively. **B)** Comparison of X3, X5 and G6 in E217Q<sub>X3</sub>, E212Q<sub>X5</sub> and 7CEL structures. Cellose chain is in grey, X3 in yellow and X5 in blue.

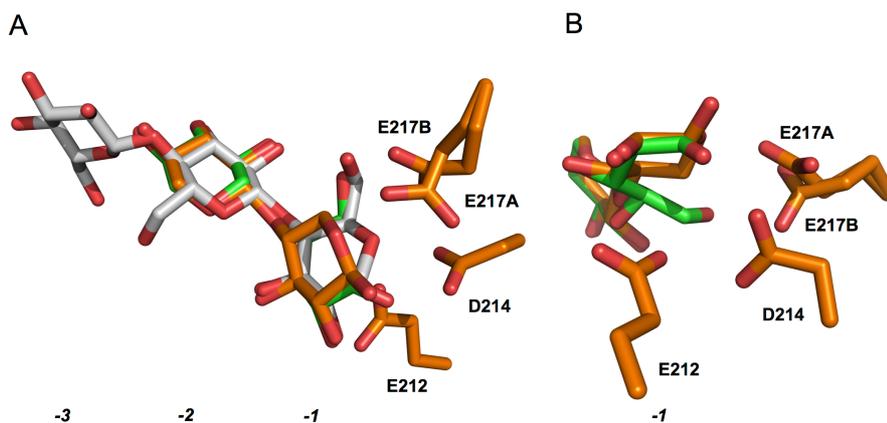


Figure 19. **A)** Superposition of subsites -3 to -1 in WT<sub>X5</sub> (orange), E217Q<sub>X4</sub> (green) and 4C4C (grey) showing that the binding of XOS is in register with the binding of cellononaose. **B)** The xylopyranosyl of subsite -1 have a regular <sup>4</sup>C<sub>1</sub> chair conformation in WT<sub>X5</sub> (orange) while in E217Q<sub>X4</sub> (green) the xylose unit is in an open aldehyde form.

#### 2.4.5 Inhibition studies

Activity of intact full-length *HjeCel7A* enzyme on *pNP-Lac* as substrate was measured in the presence of different XOS. Birchwood xylan showed the strongest inhibitory effect followed by X5, which was approximately twice as inhibitory as X3 (Figure 20). The presence of xylose had a negligible effect on *HjeCel7A* hydrolytic activity. These findings agree well with the results by Baumann *et al.*, (2011). The fitted results show that both the  $V_{max}$  and the Michaelis Menten constant  $K_m$  are affected, indicating mixed type inhibition and thus both the competitive and uncompetitive inhibition constants,  $K_{ic}$  and  $K_{iu}$ , need to be derived (Paper IV, Table 2). The inhibition experiments indicated stronger inhibition by X2 than X3 and X5. However, as mentioned above, the commercial X2 material used was found to be contaminated with small amounts of cellobiose, which is probably responsible for most of the inhibition effect in this case.

It is noteworthy that the Birchwood xylan displays ~100-fold stronger inhibition based on mass than X3 and X5. This suggests that longer XOS probably extend further into the tunnel and occupies a longer stretch of the active site, than the predominant binding at the beginning of the tunnel observed in the complex structures with X3, X4 and X5.

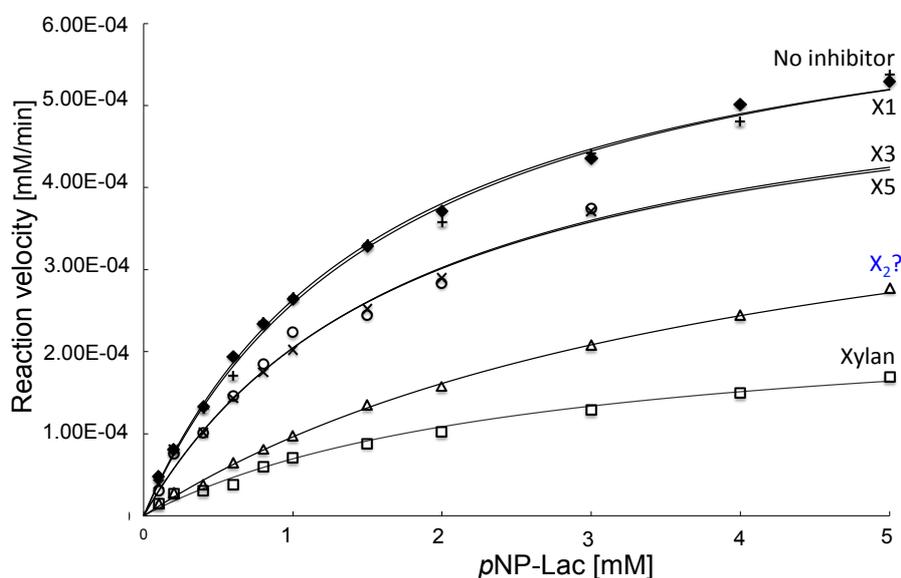


Figure 20. Inhibition of *HjeCel7A* by different xylooligosaccharides and xylan (Birchwood). Absorbance at 405 nm was measured after 60 min incubation of *pNP-Lac* at 37 °C, pH 5.0 as described in Experimental procedures. ◆, No inhibitor; +, 50 mM X1; Δ, 25 mM X2; ×, 10mM X3; o, 5mM X5 and □, 0.31 g/L xylan (Birchwood).



## 3 Conclusions and future perspectives

### 3.1 Structural, biochemical, and computational characterization of the *HirCel7A*

In paper I, we have isolated, characterized and determined the structure of *HirCel7A*, the most abundant enzyme in the secretome of the tree-killing fungus *Heterobasidion irregulare*. This fungus is an amazingly efficient pathogen due to its dual ability to infect living trees and to digest wood, via necrotrophic and saprotrophic life styles, respectively.

In a previous transcript profiling study, the *Cel7A* gene was not recognized among up-regulated genes during growth on wood (Olson et al., 2012). However, according to our results, *HirCel7A* is clearly the major protein of the fungus when cultivated on woody materials. This suggests that further studies are needed, e.g. by proteomics, to obtain a clear picture of which enzymes are actually expressed and utilized by this organism under different conditions.

The structure of *HirCel7A* shows a cellulose-binding tunnel that is more closed than *P. chrysosporium Cel7D* and more open than *H. jecorina Cel7A*, and interesting differences in loop regions along the cellulose-binding path. Structural comparisons and MD simulations reveal significant differences in loop dynamics, ligand fluctuations and interactions in the -7 to -4 binding sites, and that a tyrosine residue at the tunnel entrance of *HirCel7A* may serve as an additional ligand-binding site. Overall, the results indicate intermediate enzyme properties for *HirCel7A*, which is likely to have implications for the degree of processivity, endo-initiation, and substrate dissociation.

In light of these results, it would be interesting to conduct complementary studies to correlate our findings with activity measurements and performance assays on biologically and industrially relevant lignocellulose materials in comparison with other GH7 CBHs.

### 3.2 Mechanism of Cellulose Hydrolysis by a Two-step, Retaining CBH

In paper II, we present the first Michaelis complex, with a full cellobiose ligand, and the first glycosyl-enzyme intermediate of *HjeCel7A*. These structures reveal details of the ‘static’ reaction coordinate and enabled us to perform TPS (transition path sampling) to determine the ‘dynamic’ reaction coordinate for the catalytic cycle and to perform free energy and reaction rate calculations.

The discovery of a product-assisted mechanism for the deglycosylation step suggests that the cellobiose product does not need to be expelled until after deglycosylation, which also has implications for product inhibition. Furthermore, our study provides the first comparison of the rate constants for both steps of the catalytic cycle. Step 1 is slower than Step 2 by nearly three orders of magnitude; supporting the assumption that glycosylation is the rate-limiting step of the hydrolytic mechanism on natural substrates in these enzymes.

The computed reaction rate constant of  $11 \text{ s}^{-1}$  is in good agreement with the rate of processive hydrolysis by *HjeCel7A* on a crystalline cellulose surface estimated by AFM at  $7 (\pm 4) \text{ s}^{-1}$  (Igarashi *et al.*, 2008). However, reported turnover rates of cellulose degradation by *HjeCel7A* are much lower, and other steps in the process have been suggested to be rate-limiting, for example the advancement of the cellulose chain within the tunnel after each hydrolytic event. Our results provide an excellent starting point for follow-up studies of the reaction coordinate and energy calculations for cellulose chain translocation in order to describe the complete hydrolytic-processive cycle of this biologically and industrially important class of enzymes. Such work has already been initiated.

### 3.3 Expression, crystal structure and cellulase activity of the thermostable *HgrCel7A*

In paper III, we report the discovery, characterization and 3D structure of *HgrCel7A* with considerable higher thermostability ( $10 \text{ }^\circ\text{C}$  higher  $T_m$ ) and cellulase activity compared to the canonical *HjeCel7A*. The enzyme gave  $\sim 75\%$  higher yield of soluble sugar from biomass in a performance assay at elevated temperature. The structure of the *HgrCel7A* catalytic domain indicates considerable mobility of active-site defining loops, which may relate to higher cellulase activity. Some potentially stabilizing interactions were recognized at the base of certain loops, which may prevent loop movements to propagate into the core of the protein structure that could otherwise lead to irreversible protein unfolding. Furthermore, the structure reveals a putative conformational switch in an active-center loop that has not been reported previously in any GH7 structure.

Further studies can be envisioned along two parallel routes. The significance of the structural features that we have recognized in *HgrCel7A*, and their influence on cellulase activity and thermostability, could be investigated by mutational studies and protein engineering. It would also be interesting to perform computational analyses on this enzyme in order to compare e.g. loop dynamics, ligand solvation and enzyme-ligand interactions with other GH7 CBHs. An important aspect here would be to investigate if and how often the conformational switch of the B3 loop may occur in different enzymes, and to elucidate possible connections with enzyme action.

### 3.4 Structural insights into inhibition of GH7 by XOS

In paper IV, structures of *H. jecorina* Cel7A in complex with X3, X4 and X5 reveal a predominant binding mode for XOS at the beginning of the substrate-binding tunnel of the enzyme, which is out of register with cellobiose binding. We also observed partial occupancy of a second binding mode in subsites -2/-1 near the catalytic center. This second binding mode is in register with cellobiose binding and overlaps with the other binding mode. Thus, the same XOS molecule cannot employ both binding modes simultaneously.

Interestingly, several structures were obtained after soaking crystals with X2, but consistently showed the binding of cellobiose in the product sites +1/+2. Therefore, we analyzed the commercial X2 preparation used, and found that it was contaminated with cellobiose. The cellobiose contamination is sufficient to account for the observed electron density in the structures, and is probably also responsible for most of the inhibition effect. This may explain why X2 appeared to be a stronger inhibitor than X3 and X5, in inhibition experiments by us as well as by others. In connection with the presented results, further crystallization attempts and inhibition experiments will be conducted with X2 that has been enzymatically treated to remove the cellobiose. These experiments will hopefully lead us to proper examination of the binding of X2 to *HjeCel7A*.

Birchwood xylan displayed ~100-fold stronger inhibition based on mass than X3 and X5. This suggests that longer XOS probably extend further into the tunnel and occupies a longer stretch of the active site than X3, X4 and X5. In order to examine this structurally, crystals will be exposed to xylan with the aim of obtaining structures with longer XOS bound.



## References

- Aziz, S., & Sarkanen, K. (1989). Organosolv Pulping - a Review. *Tappi Journal* 72(3), 169-175.
- Barr B.K., Hsieh Y-L, Ganem B., Wilson, D.B. (1996). Identification of two functionally distinct classes of exocellulases. *Biochemistry* 35(2), 586-592.
- Baumann, M.J., Borch, K. & Westh, P. (2011). Xylan oligosaccharides and cellobiohydrolase I (TrCel7A) interaction and effect on activity. *Biotechnology for Biofuels* 4(1), 45.
- Beckham, G.T., Bomble, Y.J., Bayer, E.A., Himmel, M.E. & Crowley, M.F. (2011a). Applications of computational science for understanding enzymatic deconstruction of cellulose. *Current Opinion in Biotechnology* 22(2), 231-238.
- Beckham, G.T., Matthews, J.F., Peters, B., Bomble, Y.J., Himmel, M.E. & Crowley, M.F. (2011b). Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *The Journal of Physical Chemistry. B* 115(14), 4118-4127.
- Beguín, P. & Aubert, J.P. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews* 13(1), 25-58.
- Bu, L. Crowley, M.F., Himmel, M.E. & Beckham, G.T. (2013). Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. *Journal of Biological Chemistry* 288(17), 12175-12186.
- Campos, L. & Felix, C.R. (1995). Purification and Characterization of a Glucoamylase from *Humicola grisea*. *Applied and Environmental Microbiology* 61(6), 2436-2438.
- Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K. & Westh, P. (2012). Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A. *Journal of Biological Chemistry* 287(22), 18451-18458.
- Da Silva, R., Yin, D.K. & Park, Y.K. (1994). Application of thermostable xylanases from *Humicola* sp. for pulp improvement. *Journal of Fermentation and Bioengineering* 77, 109-111.
- Davies, G.J., Ducros, V., Lewis, R.J., Borchert, T.V. & Schulein, M. (1997). Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. *Journal of Biotechnology* 57(1-3), 91-100.
- Din, N., Gilkes, N.R., Tekant, B., Miller, R.C., Warren, A.J. & Kilburn, D.G. (1991). Non-Hydrolytic Disruption of Cellulose Fibers by the Binding Domain of a Bacterial Cellulase. *Bio-Technology* 9(11), 1096-1099.

- Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T. & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265(5171), 524-528.
- Divne, C., Stahlberg, J., Teeri, T.T. & Jones, T.A. (1998). High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *Journal of Molecular Biology* 275(2), 309-325.
- Fan, L., Lee, Y. & Gharpuray, M. (1982). The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Advance Biochemistry Engineering Biotechnology* 23, 158-183.
- Ferreira Filho, E.X. (1996). Purification and characterization of a beta-glucosidase from solid-state cultures of *Humicola grisea* var. *thermoidea*. *Canadian Journal of Microbiology* 42(1), 1-5.
- 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. *et al.* (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry* 278(34), 31988-31997.
- Fox, J.M., Levine, S.E., Clark, D.S. & Blanch, H.W. (2012). Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry-Us* 51(1), 442-452.
- Geddes, C.C., Nieves, I.U. & Ingram, L.O. (2011). Advances in ethanol production. *Current Opinion in Biotechnology* 22(3), 312-319.
- Grassick, A., Murray, P.G., Thompson, R., Collins, C.M., Byrnes, L., Birrane, G., Higgins, T.M. & Tuohy, M.G. (2004). Three-dimensional structure of a thermostable native cellobiohydrolase, CBHIB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *European Journal of Biochemistry* 271(22), 4495-4506.
- Guex, N. & Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18(15), 2714-2723.
- Harris, P.V., Welner, D., McFarland, K.C., Re, E., Navarro Poulsen, J.C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S. *et al.* (2010). Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry-Us* 49(15), 3305-3316.
- Hemsworth, G.R., Henrissat, B., Davies, G.J. & Walton, P.H. (2014). Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature Chemical Biology* 10(2), 122-126.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 280( Pt 2), 309-316.
- Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*. 316 ( Pt 2), 695-696.
- Henrissat, B. & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion Structural Biology* 7(5), 637-644.
- Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M., Ando, T. & Samejima, M. (2011). Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* 333(6047), 1279-1282.

- Ilmen, M., Saloheimo, A., Onnela, M.L. & Penttila, M.E. (1997). Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Applied and Environmental Microbiology* 63(4), 1298-1306.
- Jeffries, T.W. & Kurtzman, C.P. (1994). Strain Selection, Taxonomy, and Genetics of Close-Fermenting Yeasts. *Enzyme and Microbial Technology* 16(11), 922-932.
- Jonsson, L.J., Alriksson, B. & Nilvebrant, N.O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnology for Biofuels* 6(1), 16.
- Kern, M., McGeehan, J.E., Streeter, S.D., Martin, R.N., Besser, K., Elias, L., Eborall, W., Malyon, G.P., Payne, C.M., Himmel, M.E., *et al.* (2013). Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceeding of National Academy of Science of the United States of America* 110(25), 10189-10194.
- Kim, S., Stahlberg, J., Sandgren, M., Paton, R.S. & Beckham, G.T. (2014). Quantum mechanical calculations suggest that lytic polysaccharide monoxygenases use a copper-oxy, oxygen-rebound mechanism. *Proceeding of National Academy of Science of the United States of America* 111(1), 149-154.
- Kleywegt, G.J., Zou, J.Y., Divne, C., Davies, G.J., Sinning, I., Stahlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T.T. & Jones, T.A. (1997). The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 angstrom resolution, and a comparison with related enzymes. *Journal of Molecular Biology* 272(3), 383-397.
- Knott, B.C., Haddad Momeni, M., Crowley, M.F., Mackenzie, L.F., Gotz, A.W., Sandgren, M., Withers, S.G., Stahlberg, J. & Beckham, G.T. (2014). The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *Journal of the American Chemical Society* 136(1), 321-329.
- Kont, R., Kurasin, M., Teugjas, H. & Valjamae, P. (2013). Strong cellulase inhibitors from the hydrothermal pretreatment of wheat straw. *Biotechnology for Biofuels* 6(1), 135.
- Koshland, D.E. (1953). Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews* 28(4), 416-436.
- Kurasin, M. & Valjamae, P. (2011). Processivity of cellobiohydrolases is limited by the substrate. *Journal of Biological Chemistry* 286(1), 169-177.
- Lee, Y.H. & Fan, L.T. (1982). Kinetic studies of enzymatic hydrolysis of insoluble cellulose: analysis of the initial rates. *Biotechnology and Bioengineering* 24(11), 2383-2406.
- Lin, Y., Silvestre-Ryan, J., Himmel, M.E., Crowley, M.F., Beckham, G.T. & Chu, J.W. (2011). Protein allostery at the solid-liquid interface: endoglucanase attachment to cellulose affects glucan clenching in the binding cleft. *Journal of the American Chemical Society* 133(41), 16617-16624.
- Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* 66(3), 506-577.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* 40(Pt 4), 658-674.
- Merino, S.T., & Cherry, J. (2007). Progress and challenges in enzyme development for biomass utilization. *Advances in Biochemistry and Engineering/Biotechnology* 108, 95-120.

- Modenbach, A.A. & Nokes, S.E. (2012). The use of high-solids loadings in biomass pretreatment—a review. *Biotechnology and Bioengineering* 109(6), 1430-1442.
- Momeni, M.H., Payne, C.M., Hansson, H., Mikkelsen, N.E., Svedberg, J., Engstrom, A., Sandgren, M., Beckham, G.T. & Stahlberg, J. (2013). Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *Journal Biological Chemistry* 288(8), 5861-5872.
- Munoz, I.G., Ubhayasekera, W., Henriksson, H., Szabo, I., Pettersson, G., Johansson, G., Mowbray, S.L. & Stahlberg, J. (2001). Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. *Journal of Molecular Biology* 314(5), 1097-1111.
- Olson, A., Aerts, A., Asiegbu, F., Belbahri, L., Bouzid, O., Broberg, A., Canback, B., Coutinho, P.M., Cullen, D., Dalman, K. *et al.* (2012). Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New phytologist*. 194(4), 1001-1013.
- Parkkinen, T., Koivula, A., Vehmaanpera, J. & Rouvinen, J. (2008). Crystal structures of *Melanocarpus albomyces* cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. *Protein Science* 17(8), 1383-1394.
- Payne, C.M., Resch, M.G., Chen, L., Crowley, M.F., Himmel, M.E., Taylor, L.E., 2nd, Sandgren, M., Stahlberg, J., Stals, I., Tan, Z. *et al.* (2013). Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceeding of National Academy of Science of the United States of America* 110(36), 14646-14651.
- Sandgren, M., Gualfetti, P.J., Paech, C., Paech, S., Shaw, A., Gross, L.S., Saldajeno, M., Berglund, G.I., Jones, T.A. & Mitchinson, C. (2003). The *Humicola grisea* Cel12A enzyme structure at 1.2 Å resolution and the impact of its free cysteine residues on thermal stability. *Protein Science* 12(12), 2782-2793.
- Scheller, H.V. & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology* 61, 263-289.
- Shoseyov, O., Shani, Z. & Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews* 70(2), 283-295.
- Schulein, M. (1997). Enzymatic properties of cellulases from *Humicola insolens*. *Journal of Biotechnology* 57, 71-81.
- Stahlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T.T. & Jones, T.A. (1996). Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *Journal of Molecular Biology* 264(2), 337-349.
- Stahlberg, J., Johansson, G. & Pettersson, G. (1993). *Trichoderma reesei* has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. *Biochimica et Biophysica Acta* 1157(1), 107-113.
- Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M. & Davies, G.J. (1996). Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry-US* 35(48), 15280-15287.
- Sun, Y. & Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* 83(1), 1-11.

- Takashima, S., Nakamura, A., Hidaka, M., Masaki, H. & Uozumi, T. (1996). Cloning, sequencing, and expression of the cellulase genes of *Hemicola grisea* var. *thermoidea*. *Journal of Biotechnology* 50(2-3), 137-147.
- Textor, L.C., Colussi, F., Silveira, R.L., Serpa, V., de Mello, B.L., Muniz, J.R., Squina, F.M., Pereira, N., Jr., Skaf, M.S. & Polikarpov, I. (2013). Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from *Trichoderma harzianum*: deciphering the structural features of cellobiohydrolase catalytic activity. *FEBS Journal* 280(1), 56-69.
- Ubhayasekera, W., Munoz, I.G., Vasella, A., Stahlberg, J. & Mowbray, S.L. (2005). Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. *FEBS Journal* 272(8), 1952-1964.
- Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z.L., Zhai, H., Sorlie, M. & Eijsink, V.G.H. (2010). An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* 330(6001), 219-222.
- Viikari, L., Alapuranen, M., Puranen, T., Vehmaanpera, J. & Siika-Aho, M. (2007). Thermostable enzymes in lignocellulose hydrolysis. *Advances in Biochemical Engineering/Biotechnology* 108, 121-145.
- von Ossowski, I., Stahlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., *et al.* (2003). Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. *Journal of Molecular Biology* 333(4), 817-829.
- Voutilainen, S.P., Boer, H., Linder, M.B., Puranen, T., Rouvinen, J., Vehmaanpera, J., & Koivula, A. (2007). Heterologous expression of *Melanocarpus albomyces* cellobiohydrolase Cel7B, and random mutagenesis to improve its thermostability. *Enzyme and Microbial Technology* 41(3), 234-243.
- Voutilainen, S.P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., Hooman, S., Viikari, L., Vehmaanpera, J. & Koivula, A. (2008). Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnology and Bioengineering* 101(3), 515-528.
- Wang, X (2009). Identification of proteins expressed by the root-rot fungus *Heterobasidion annosum* when growing on lignocellulose. Second cycle, A2E, Master of Science thesis. Swedish University of Agricultural Sciences, Uppsala, Sweden. <http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-8-268>.
- Zhang, Y.H., Himmel, M.E. & Mielenz, J.R. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnology Advances* 24, 452-481.
- Zhang, Y.H. & Lynd, L.R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and Bioengineering* 88(7), 797-824.
- Zimmermann, A.L., Terenzi, H.F. & Jorge, J.A. (1990). Purification and properties of an extracellular conidial trehalase from *Hemicola grisea* var. *thermoidea*. *Biochimica et Biophysica Acta* 1036(1), 41-46.



## Acknowledgements

Completing a PhD study, is truly a marathon competition, and I would not have been able to fulfill this task without support and assistance of several people during my doctoral studies.

First and foremost, I would like to express my special appreciation and gratefulness to my main supervisor **Jerry Ståhlberg** for his guidance, understanding and most importantly his friendship, during my master thesis and Ph.D education. His mentorship was excellent in supporting me with a well round of experience and advice. I also want to express my gratitude to my co-supervisors; **Henrik Hansson** for his substantial support and fruitful discussions, both in implementing analytical experiments and helping me on several other tasks; and **Mats Sandgren** for his excellent work to maintain and reinforce the integrity of our group and expand the scientific network through collaboration, workshops and conferences, from which I had great benefit already during my PhD education.

I want to thank several graduate and undergraduate colleagues, who I have been working with in our group: **Saeid Karkehabadi & Nils Mikkelsen**, who have been an inspiration on how to make things 'perfect', **Miao Wu** and **Mikael Gudmundsson** for being great PhD companions and office mates, and **Anna Borisova** for your fruitful discussions and countless encouragements.

Furthermore, I also want to acknowledge with much appreciation our excellent collaborators Drs. **Gregg Beckham, Brandon Knott** and **Christina Payne**, at NREL, who have conducted MD simulation on two projects, Dr. **Edmundo Larenas** and Dr. **Frits Goedegebuur**, at Dupont Industrial Biosciences, and Professor **Masahiro Samejima** and Drs. **Kiyohiko Igarashi, Takuya Ishida**, at University of Tokyo, for their countless support.

I would like to thank my opponent, Dr. **Anu Koivula**, and committee members Professor **Sherry Mowbray**, Professor **Gunnar Johansson**, Dr. **Jinyu Zou** and the chairman Professor **Torleif Hård**, for contributions at my thesis defense.

I also thank dozens of people, who have helped and taught me immensely at our x-ray group: **Alwyn Jones, Terese Bergfors, Stefan Knight, Wimal, Margaretha, Evalena, Christopher L, Glareh, Roy, Mahfuz, Frida, Umair, Li, Lotta, Anatoly, Jonas, Anton, Torsten, Åsa, Lars L, Avinash, Agata, Yafei, Wangshu, Dee, Inger, Maria, Yang, Benjamin, Ulla, Henrik B, Erling, Nina.**

Thanks to my fantastic colleagues at the Department of Microbiology and Department of Plant Pathology, **Anna Schnurer, Volkmar Passoth, Johanna B, Åke Olson, Jevgenija T** and so forth.

I would also like to thank my dear friends: **Aboozar, Sara H, Masoud, Saeid A, Katja, Simon, Marcus, Anna K, George** and **Mona-Mercedes V** for their encouragements and of course all countless friends who made shared fantastic memories with me in Sweden.

*Last, but not least, I offer my regards and blessings to my fantastic parents, and lovely brother and sister, who I love so much, and who always supports me the strongest in all my pursuits.*

**May 9<sup>th</sup> 2014.**