Structure and Functional Studies of Plant Cell Wall Degrading Enzymes

Glycoside Hydrolase Family 3 β-Glucosidases and Lytic Polysaccharide Monooxygenases

Mikael Gudmundsson

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
2014:83

Cover: *That's no moon! It's a β-glucosidase! (The ReCel3A dimer)*

ISSN 1652-6880
ISBN (print version) 978-91-576-8114-0
ISBN (electronic version) 978-91-576-8115-7

Print: SLU Service/Repro, Uppsala 2014

Structure and Function of Plant Cell Wall Degrading Enzymes Glycoside Hydrolase Family Three β-Glucosidases and Lytic Polysaccharide Monooxygenases

Abstract

Presently, plant biomass is considered as one of the major future renewable sources for the production of second-generation biofuels. While the first generation biofuels essentially are based on starch and sucrose rich feed stocks and which production may compete with food production, the second-generation biofuels may be based on lignocellulose as feedstock, which is less problematic from an ethical point of view. The degradation of carbohydrates in plant biomass to fermentable sugars requires the concerted action of several diverse classes of carbohydrate active enzymes (CAZymes) for a total and efficient conversion of the plant biomass. Through a carefully balanced synergism mechanistically different CAZymes are able to degrade the stable and recalcitrant polymers in the plant cell walls, such as cellulose, to soluble and fermentable monosaccharides. It is crucial to study the properties and function of these enzymes if we want to strive for a sustainable production of chemicals and biofuels, as they serve as a reservoir of environmentally friendly molecular tools. The main focus of the research work presented in this thesis is biochemical and structure-function characterizations of two classes of CAZymes: fungal glycoside hydrolase family 3 (GH3) β-glucosidases, and bacterial lytic polysaccharide monooxygenases, often referred to as LPMOs. GH3 β-glucosidases catalyse the conversion of disaccharides, produced by other CAZymes e.g. cellulases, to glucose. *H. jecorina* Cel3A, *R. emersonii* Cel3A and *N. crassa* NcGH3-3 are three industrially relevant fungal GH3 β-glucosidases for which the structures have been determined using X-ray crystallographic methods. The *H. jecorina* Cel3A, *R. emersonii* Cel3A enzymes has also been characterized biochemically. The LPMOs act in the very initial stage of plant cell wall degradation and cleave glycosidic bonds in crystalline polysaccharides via an oxidative mechanism, which facilitates access to new chain ends for other CAZymes. To elucidate the structural and biochemical properties of LPMOs with bacterial origin, the structure of an AA10 LPMO the LPMO10A from *Enterococcus faecalis* was determined using X-ray crystallography. Furthermore, structural changes of the active site metal configuration by so-called X-ray induced photoreduction, were determined. During this reduction process, which mimics the active enzyme, the bound active site copper atom is reduced from Cu(I) to Cu(II), which causes changes in the ligation configuration.

Keywords: β-glucosidase, LPMO, X-ray crystallography, carbohydrate active enzymes

Author's address: Mikael Gudmundsson, Swedish University of Agricultural Sciences, Department of Chemistry and Biotechnology, P.O. Box 7015, 750 07 Uppsala, Sweden
E-mail: mikael.gudmundsson@slu.se

Dedication

Till Johanna

Nullum magnum ingenium sine mixtura dementiae
Seneca

List of Publications

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

I Karkehabadi S, Helmich KE, Kaper T, Hansson H, Mikkelsen NE, **Gudmundsson M**, Piens K, Fujdala M, Banerjee G, Scott-Craig JS, Walton JD, Phillips GN and Sandgren M. (2014). Biochemical Characterization and Crystal Structures of a Fungal Family 3 β-Glucosidase, Cel3A from *Hypocrea jecorina*. *The Journal of Biological Chemistry.* **289**(45), 31624-31637.

II **Gudmundsson M**, Larsson AM, Hansson H, Stals I, Larenas E, Kaper T and Sandgren M. (2014). Crystal Structure of Glycoside Hydrolase Family 3 β-Glucosidase Cel3A from the Moderately Thermophilic Fungus *Rasamsonia emersonii*. (Submitted manuscript)

III **Gudmundsson M**, Karkehabadi S, Mikkelsen NE, Kaper T, and Sandgren M. Crystal structure of a Glycoside Hydrolase Family 3 β-Glucosidase from *Neurospora crassa*. (Manuscript)

IV **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. (2014). Structural and Electronic Snapshots during the Transition from a Cu(II) to Cu(I) Metal Center of a Lytic Polysaccharide Monooxygenase by X-ray Photoreduction. *The Journal of Biological Chemistry.* **289**(27), 18782–18792.

* First authorship shared

The contribution of Mikael Gudmundsson to the papers included in this thesis was as follows:

I Planning the work together with the other co-authors of the paper and reviewing relevant literature within the scope of study. Taking part in data interpretation, structure refinements and writing the paper together with the co-authors. Conclusions made together with all co-authors of the paper.

II Planning the work together with the other co-authors of the paper and reviewing the literature. Conducting all lab work for the experiments included in the paper, including refining and depositing the structures included in the study. Taking an active role in interpretation and writing the paper together with the co-authors. Overall conclusions made together with the co-authors.

III Planning the work together with the co-authors of the paper and reviewing the literature. Refining the structure, data interpretation and coordinating the work writing the paper together with the co-authors. Conclusions made together with all co-authors of the paper.

IV Planning the work together with the co-authors and reviewing relevant literature within the scope of the study. Performed lab work and data interpretation together with co-authors. Taking part in data interpretation, structure refinements and writing the paper together with the co-authors. Overall conclusions were made together with co-authors.

Abbreviations

AA	Auxiliary activities
aa	Amino acid
*Aa*BGLI	*Aspergillus aculeatus* β-glucosidase I
BGL	β-glucosidase
BXL	β-xylosidase/ β-D-xylopyranosidase
CAZy	Carbohydrate-active enzymes Database
CAZymes	Carbohydrate-active enzymes
CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CNPG	2-chloro-4-nitrophenyl-β-D-glucopyranoside
DNA	Deoxyribonucleic acid
DP	Degree of polymerization
EC	Enzyme classification
EG	Endoglucanase, endo-β-1,4-glucanase
GH	Glycoside hydrolase
GlcNAc	*N*-acetyl-D-glucosamine
NagZ	β-D-*N*-acetylhexosaminidase
GT	Glycosyltransferase
*Hj*Cel3A	*Hypocrea jecorina* Cel3A
*Hv*ExoI	*Hordeum vulgare* β-glucan exohydrolase I
LPMO	Lytic Polysaccharide Monooxygenase
*Nc*Cel3A	*Neurospora crassa* Cel3A
PASC	Phosphoric acid swollen cellulose
PCS	Pretreated corn stover
PL	Polysaccharide lyase
*p*NP-G	*p*-nitrophenyl-β-D-glucopyranoside
*Re*Cel3A	*Rasamsonia emersonii* Cel3A
RMSD	Root mean square deviation

RNA	Ribonucleic acid
*Tn*Bgl3B	*Thermotoga neapolitana* β-glucosidase 3B

1 Introduction

In the spring of 2009 as a freshly baked confident chemistry graduate I was asked what I had been taught about fungi (basidiomycete fungi to be exact) and fungal enzymes. My glazed look and silence probably betrayed my ignorance, as I was soon handled a book detailing the life and behaviour *Heterobasidion annosum*, the hot fungi du jour in the Sandgren-Ståhlberg lab. Thus I started my journey amongst my filamentous friends and their fascinating enzymes by which they eek out a living on this planet.

Fungi are responsible for degrading the majority of all plant matter on land. As their tools for this momentous task they have an intricate arsenal of enzymes developed over the ages in an ancient evolutionary race with the plants. We can see traces of this evolutionary back and forth by for example digging into the ground and seeing that there is no coal deposits younger than 300 million years. Coal was formed when the non-degraded lignin in trees and plant matter were accumulated and buried, but roughly 300 million years ago a new breed of fungi appeared that could degrade lignin and thus radically slowed the formation of coal (Floudas *et al.*, 2012). These had new tools, new enzymes, that could, by utilizing radicals, break up the tough and previously unattainable lignin for consumption by the fungi.

The ability of fungi to enzymatically degrade plant biomass has in the last decades garnered large interest from industrial actors as the need for renewable fuels has been realized. Currently the most widely produced and used biofuel is ethanol. But expansion of ethanol production, from the so called 1st generation biofuels, produced from sugar and starch crops, has raised concerns about the competition with food crops and natural resources such as water and arable land (FAO., 2011). One solution to this conflict of interest is to utilize non-edible lignocellulosic biomass. Such resources can be obtained from plants cultivated on non-arable land (e.g. switchgrass), from agricultural waste (e.g.

corn stover, wheat straw, sugarcane bagasse) or residues from the forest industry.

Among the many technical hurdles for producing economically viable lignocellulosic ethanol is the composition and production of efficient enzyme cocktails that can liberate the recalcitrant carbohydrates in plant cell walls at high conversion rates. A starting point for such enzyme cocktails has been use the whole cellulase mixture produced by the ascomycete fungi *Hypocrea jecorina*. This fungus secretes a complete set of cellulose degrading enzymes. There has during the past decades been a constant development of *H. jecorina* strains that gradually have been capable of produce higher titres of (hemi)cellulose degrading enzymes (Seidl *et al.*, 2009). The effort of improving the efficacy of the biocatalysts performing the degradation has mainly followed two approaches: (*i*) the mining of genomes for diversity and novel enzyme paradigms and (*ii*) rational knowledge based protein engineering. The latter approach requires intimate knowledge of mechanisms of action of these enzymes in regards to their substrates at a molecular level.

1.1 Aim of the thesis

The overall aim of this thesis was to study the structure and function relationship among fungal glycoside hydrolase family 3 (GH3) β-glucosidases from industrially relevant host organisms. To achieve this, the structures of three fungal GH3 β-glucosidases were determined using X-ray crystallographic methods: Cel3A from *Hypocrea jecorina* (*Hj*Cel3A; Paper I), Cel3A from the thermophilic fungus *Rasamsonia emersonii* (*Re*Cel3A: Paper II); and GH3-3 from *Neurospora crassa* (*Nc*GH3-3; Paper III). In addition to structure determination, some of the biochemical properties for two of these enzymes (*Hj*Cel3A and *Re*Cel3A) where characterized and the data analysed in the light of the available structural information (Paper I and II).

A second research focus within the thesis work was to structurally and biochemically characterize the bacterial enzyme, LPMO10A from *Enterococcus faecalis* (*Ef*LPMO33A), from a newly discovered and partially characterized class of enzymes called lytic polysaccharide monooxygenases (LPMOs), using dose resolved X-ray crystallography.

It is crucial to study the properties and function of glycoside hydrolase in general, and specifically in these in this fungal GH3 β-glucosidases and LPMOs, if we want to strive for efficient and sustainable production of chemicals and biofuels using plant biomass as feedstock, as enzymes from biomass degrading microorganisms serve as a reservoir of environmentally friendly molecular tools.

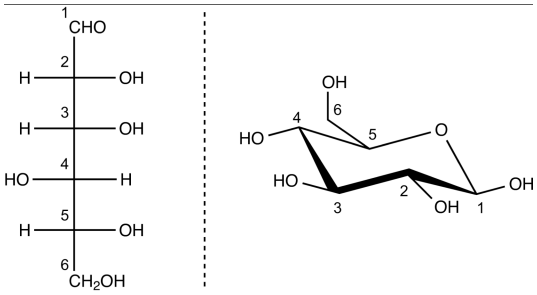
This thesis also contains a summary of the history and current situation of the field of microbial biomass degradation, with the focus towards the role of β-glucosidases and their importance in biotechnological applications. Also discussed briefly is the recently discovered enzymes LPMO.

I have used protein X-ray crystallography, spectroscopic and biochemical methods to investigate the enzymes in focus of this thesis, as well as collaborated with groups utilizing quantum chemical computational methods to further understand the enzymatic mechanism. The crystal structures included in this thesis will be thoroughly analysed in the rest of the thesis, and the stage for future characterization work will be set.

2 Background

2.1 Carbohydrates

Carbohydrates are one of the most ubiquitous groups of biomolecules found in nature. They are involved in most major cellular functions, from fuel for metabolism to signalling and intra-cellular communication. The definition of carbohydrates is that they are polyhydroxyaldehydes, polyhydroxyketones and simple derivatives thereof or compounds hydrolysable into such units. Monosaccharides are a chain of hydroxylated hydrocarbons with either a terminal aldehyde or internal ketone group that cannot be hydrolysable into a simple form. The coupling of monosaccharides via glycosidic linkages gives rise to oligosaccharides (which will be referred to as glycans in this thesis unless otherwise specified) that can be either branched or linear (Varki & Sharon, 2009) or glycosides.



ß-D-glucose

Figure 1. Open chain and ring forms of β-D-glucose. The open form and the ring form exist in equilibrium in water where the ring form is highly favoured.

Monosaccharides or glycans can be joined together to form polysaccharides like chitin or cellulose, which in turn can be organized into higher hierarchical

structures and ultimately form complexes that make up the major parts of the plant cell walls, the fungal cell walls, the exoskeleton of arthropods, the protein-sugar co-polymer peptidoglycan and lipopolysaccharide in bacteria and the glycoprotein cell-envelope of some archaea (Albers & Meyer, 2011; Cummings & Doering, 2009; Esko *et al.*, 2009; Etzler & Mohnen, 2009; Tiemeyer *et al.*, 2009). Animals also produce polysaccharides like the very complex mucin glycoprotein that line mucus membranes and in large varieties of glycoproteins (Brockhausen *et al.*, 2009; Helenius & Aebi, 2001).

2.2 Lignocellulosic biomass

Human civilisation has, since the early days of our species, been dependent on plant biomass as a source of energy for a variety of purposes. It was not displaced until the advent of crude oil that gave rise to inexpensive liquid fuel that led to industrialization and subsequently to our current high standard of living for a relatively big fraction of the human population on our planet. Today the world's societies are largely affected by increasing environmental and political concerns regarding the utilization of the slowly depleting fossil resources. A depletion of our fossil based resources will make biomass the only sustainable source of organic carbon that can be utilized to produce replacements for fuels and chemicals that currently are now based on fossil resources such as crude oil, natural gas and coal.

Plant biomass is produced from CO_2 and H_2O with O_2 as a by-product and sunlight as an energy source (Hill, 2012). The main components produced by plants are hexose (C_6) and pentose (C_5) carbohydrates, which in turn constitute the major polymeric components of cellulose and hemicelluloses. There is also a third major plant component formed, lignin. Lignin is composed of a heterogeneous and highly cross-linked polymer of substituted phenols that in combination with cellulose and hemicellulose confer rigidity to the plant.

Cellulose (*Figure 2*) is a linear polymer composed of glucose units linked by β-1,4 glycosidic bonds (Klemm *et al.*, 2005), and constitutes roughly 40-50 % of the dry weight of plant biomass (Pettersen, 1984). Each glucose unit in a cellulose polymer is rotated by 180° relative to its two neighbours, thus creating a minimal repeating unit of two glucose units in cellulose. Two glucose units attached to one and another by a β-1,4 glycosidic bond is called cellobiose. The chain length, or degree of polymerization (DP), of cellulose polymers in plant cell walls varies from 100 to 10000 glucose units in length (Klemm *et al.*, 2005). Cellulose has a strong tendency to aggregate with other cellulose chains through intra- and intermolecular hydrogen bonding, into partially crystalline microfibrils. These microfibrils can adopt a multitude of structures and

morphologies with varying degrees of crystallinity (Klemm *et al.*, 2005; Béguin & Aubert, 1994).

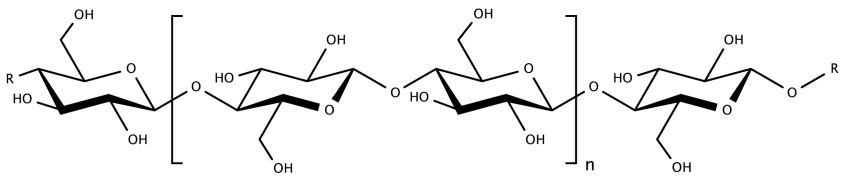


Figure 2. Molecular structure of cellulose, composed of n units of cellobiose.

Hemicellulose is a heterogeneous group of plant polysaccharides that unlike cellulose, have lower DP, in the range of 50 – 300 glycan units, and these are often decorated with various types of glycan's attached to the main chain, and different types of hemicelluloses are essentially amorphous. The amounts of hemicellulose found in plant cell walls are typically 25 - 30 % (Pettersen, 1984). The variability of carbohydrates, linkages, branching and substitutions among different types of hemicellulose vary greatly. The linear main chain components are generally β(1-4) linked glucose, xylose or mannose, with β(1,3)/(1,4)-glucans appearing more rarely and mainly in grasses (Scheller & Ulvskov, 2010). In softwood galactoglucomannans and arabinoglucuronoxylan are the main hemicelluloses, whereas hardwood contains mostly glucuronoxylans. Functionally what the hemicelluloses contribute is to strengthen the plant cell wall by tethering to the cellulose fibrils, which they cover and interconnect.

The second most abundant biopolymer, after cellulose, in plant cell walls, as well on the planet, is lignin (Lu, 2014). Lignin, or lignins, is a complex mix of aromatic heteropolymers derived from a set of hydroxycinnamyl-alcohol monomers (which in turn are derived ultimately from phenylalanine). Lignin is deposited in the nascent plant cell walls after cellulose and hemicelluloses have been synthesized and deposited. The main functions attributed to lignin are to provide stiffness and strength to the cell wall and acting as glue in and especially between the cells. In addition lignin also waterproofs the cells and the first appearance of lignin is thought to be intimately linked to the development of vascular water transport systems in plants (Boerjan *et al.*, 2003; Whetten & Sederoff, 1995). Lignin and lignification is also important in plant pathogen defence, where the lignin itself by being rigid and hydrophobic provides protection from mechanical pressure and water-soluble cell-wall degrading enzymes. Lignins are also accumulated by the plants at positions of external microbial attack where they exercise a more offensive antimicrobial role

(Bhuiyan *et al.*, 2009; Nicholson & Hammerschmidt, 1992; Vance *et al.*, 1980).

2.3 Microbial biomass degradation

The organisms that are the predominant degraders of lignocellulosic material in nature are fungi, especially ascomycetes and basidiomycetes (Kirk & Farrell, 1987). The ability to efficiently degrade lignocellulose is thought to be related to the ability of growing mycelia that can transport needed nutrients, such as nitrogen an iron, to the otherwise nutrient poor carbon source lignocellulose. Degradation of the plant cell wall components occurs outside of the fungus, either in association with the fungal cell wall or extracellularly by secreted proteins and other molecules. The fungi are usually described as having two enzyme systems for degrading lignocellulose: a hydrolytic system of enzymes that hydrolyses and modifies polysaccharides; and an oxidative system that breaks down lignin and opens phenyl rings. As will be described later in this thesis the advent of a new class of plant degrading enzymes called lytic polysaccharide monooxygenases has drastically complicate this picture. Part of this thesis is dedicated to studies on this new class of oxidative plant degrading enzymes, see section 1.10 and paper IV.

2.4 Fungal cellulose biodegradation

The synergistic action of fungal cellulases have been known for a long time (Wood & McCrae, 1978), and the archetypical cellulase system has been that of the ascomycete *Hypocrea jecorina* (teleomorph of *Trichoderma reesei*). *H. jecorina* was initially isolated at Bougainville Island (Solomon Islands) in the pacific during World War II when the rapid disintegration of tents and uniforms prompted an inquiry, and quarantine, by researchers. The native strain, QM6a, is the ancestor of all enzyme producing *H. jecorina* strains used in the industry worldwide for large-scale production of enzymes. Besides being the main producer of industrial cellulases and hemicellulases with reported enzyme yields of more than 100 g/l, *H. jecorina* has also been a model organism within academia for studies of cellulase and hemicellulose regulation (Kubicek *et al.*, 2009; Martinez *et al.*, 2008; Merino & Cherry, 2007; Aro *et al.*, 2005; Cherry & Fidantsef, 2003).

Traditionally the cellulolytic system of *H. jecorina* has been divided into three classes of enzymes: endoglucanases (endo-β-1,4-glucanases, EG), cellobiohydrolases (β-1,4-glucanases, CHB), and β-glucosidases. Endo-acting cellulases cleave cellulose chains internally and create new chain ends for the pro-

cessive cellobiohydrolases to attach to and thereafter processively degrade upon release of cellobiose. Cellobiose is then further cleaved to glucose by enzymes called β-glucosidases. Characterization of enzymes from this last class of enzymes is one of the main focus areas within this thesis, and the topic of paper I to III.

EGs, represented in *H. jecorina* by Cel7B, Cel5A, Cel12A and Cel45A, are characterized by having an open substrate binding cleft (Rouvinen *et al.*, 1990), which allow these enzymes to bind exposed cellulose chains and thus cleave in the middle of the cellulose polymer. These enzymes show highest activity on amorphous cellulose rather than crystalline cellulose (Barr *et al.*, 1996). CBHs, while having a similar catalytic domain as homologous EGs they have extended loops protruding over the active site cleft forming a tunnel through the catalytic domain (Divne *et al.*, 1994). This tunnel facilitates the CBHs ability to processively cleave off cellobiose units from the cellulose chain while still staying bound to the polymer. *H. jecorina* has two CBHs: Cel7A and Cel6A. Both enzymes are composed of a catalytic domain and an associated non-catalytic carbohydrate-binding module (CBM). Both *Hj*Cel7A and *Hj*Cel6A are processive enzymes with *Hj*Cel7A acting from the reducing end of the cellulose chain and *Hj*Cel6A acting from the non-reducing end. CBHs, and to lesser extent EGs, are sensitive to inhibition by the cellobiose released during the degradation of cellulose (Gruno *et al.*, 2004). This product inhibition of the CBHs by its own catalytic end product cellobiose is relieved by further degradation of the disaccharide by β-glucosidases to the monosaccharide glucose.

H. jecorina has a set of seven β-glucosidases: Cel1A, Cel1B, Cel3A, Cel3B, Cel3C, Cel3D and Cel3E. The major role for the β-glucosidases is thought to act as a cellobiase, which hydrolyses the released cellobiose and to some extent longer cellodextrins to glucose. An alternative and/or additional role for these enzymes is to act as “transglycosidases“ that convert disaccharides and monosaccharides to more potent cellulase gene inducers such as sophorose (Foreman *et al.*, 2003; Suto & Tomita, 2001).

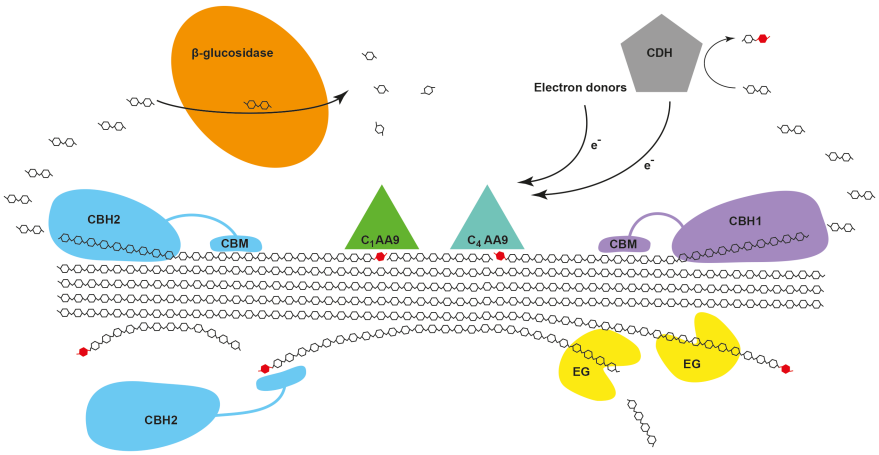


Figure 3. The new paradigm of cellulose degradation (Horn *et al.*, 2012). Exo acting cellobiohydrolases processively degrade cellulose from the reducing ends (CBH1) and the non-reducing end (CBH2). Endo acting endoglucanases (EG) hydrolyse exposed amorphous cellulose chains. Cellobiose is further cleaved by β-glucosidases to for glucose or is oxidized by cellobiose dehydrogenase (CDH). Electrons from CDH or from other electron donors are used by the lytic polysaccharide monooxygenases that via oxidation of the C_1 or C_4 carbon cleaves the glycosidic linkages and thus decrystallises the cellulose.

2.5 Carbohydrate active enzymes

Carbohydrates, glycans and polysaccharides can be exceedingly if not infinitely complex. Microorganisms and their carbohydrate active enzymes have been evolving over the ages to adapt to the increasing complexity offered by the carbohydrates. The enzymes acting on carbohydrates utilize a relatively small number of protein structure folds and domains with minor adaptations for specific substrates. Since 1991 carbohydrate active enzymes have been organized into families based on sequence, structure and catalytic activities in the Carbohydrate active enzymes database (CAZy, www.cazy.org). The CAZy families are defined by sequence similarity to a biochemically characterized founding member (Henrissat, 1991), which fortunately revealed a correlation between sequence and substrate specificity. Thus since every substrate specificity has a common ancestor which can be traced in contemporary proteins, CAZymes (Cantarel *et al.*, 2009) exhibit the opposite of the problems posed by carbohydrates, that are chemically similar and difficult to separate and distinguish. Distinguishing the exact substrate preference of highly specific enzymes such as DNAses, RNAses or proteases from sequence alone is nigh impossible whereas the glycobiology of an organism can be deduced relatively quickly after assigning CAZy families (Cantarel *et al.*, 2012). The CAZy database currently contains over 350 protein families and several hundred thousands of sequences and

is continuously updated as new proteins are functionally characterized and structurally determined. Since 2013 CAZy is organized into six main classes, five enzyme classes and the associated proteins carbohydrate binding modules (CBM) (Lombard *et al.*, 2014; Levasseur *et al.*, 2013). These six main classes of proteins in the CAZy database are currently organized as follows:

- **Glycoside hydrolases (GH)**
  Glycoside hydrolases (EC 3.2.1.x) are enzymes that catalyse the hydrolysis of glycosidic bonds of glycans, releasing a carbohydrate hemiacetal and the free aglycon, as well as the reverse transglycosylation reaction. There are as of this writing over 130 GH families classified in CAZy. GHs will be discussed in further detail below.

- **Glycosyltransferases (GT)**
  Glycosyltransferases (EC 2.4.x.y) are responsible for the synthesis of disaccharides, glycans and polysaccharides by catalysing the formation of glycosidic linkages by transferring an activated sugar moiety to a nucleophilic acceptor, usually an alcohol (Lairson *et al.*, 2008).

- **Polysaccharide lyases (PL)**
  Polysaccharide lyases (EC 4.2.2.x) cleave uronic acid-containing polysaccharides using a β-elimination mechanism creating an unsaturated uronic acid at the non-reducing end. The PLs are a complimentary class of enzymes to the GHs that can degrade glycans and are ubiquitous in nature. Their abilities to degrade alginates, pectins and heparins have garnered them quite some interest from food and medical industries (Lombard *et al.*, 2010; Yip & Withers, 2004; Sutherland, 1995).

- **Carbohydrate esterases (CE)**
  Man glycans, especially in hemicelluloses, are substituted with acyl moieties. Carbohydrate esterases catalyses the de-*O* or *N*-aceteylation of glycans. The CE class involves a wide range of reaction mechanisms, most common is a Ser-His-Asp(Glu) triad, analogous to that in lipases and serine proteases, but His-Asp dyad and Zn^{2+} catalyzed deacetylation occurs in other families (Biely, 2012; Dodd & Cann, 2009).

- **Auxiliary activities (AA)**
  After it was discovered that the glycoside hydrolase family 61 (GH61) and carbohydrate binding module family 33 (CBM33) were in fact lytic polysaccharide monooxygenases, of which there will be further discussion about

below, it was decided to establish a new class of CAZymes named auxiliary activities. The AA family is composed of the aforementioned LPMOs, GH61 and CBM33 renamed to AA9 and AA10 respectively, and lignin degrading enzymes and few other carbohydrate oxidizing families (Levasseur *et al.*, 2013).

- **Carbohydrate binding domains (CBM)**
  Carbohydrate binding domains are a non-catalytic class of proteins with the capability to bind to crystalline and soluble carbohydrates. These proteins are most often found within the protein sequence of larger CAZymes as auxiliary modules (Boraston *et al.*, 2004). There are cases of isolated CBM proteins (Barral *et al.*, 2005) and previously the CBM33 family was thought to be one such family before its LPMO activity was determined (Vaaje-Kolstad *et al.*, 2010).

2.6 Glycoside hydrolases

The glycosidic bonds, which the various types of glycoside hydrolases hydrolyse, are primarily found among polysaccharides that make up a large portion of the carbon in the biosphere. The stability of these polymers is a feature that makes them highly suitable for energy and as structural elements. Cellulose is estimated to have a half-life of ~4.7 billion years at room temperature. Thus it is not surprising that GHs, which degrade such molecules, are one of natures most potent catalysts with rate enhancements up to 10^{17} times that of the uncatalysed rate (Wolfenden *et al.*, 1998).

As previously stated GHs are a ubiquitous and diverse group of enzymes found in all types of life. They have for more than 20 years been organized into families based on sequence and structure conservation (Henrissat, 1991), as listed in the CAZy database. Where enzymes from the same enzyme family share an overall similarity mode of action, structure and usually also catalytic mechanism. There are two main classes of glycoside hydrolase catalytic mechanisms; retaining and inverting glycoside hydrolases, depicted in *Figure 4* (Koshland, 1953). Both mechanism classes utilize two key carboxyl groups that stabilize an oxocarbenium-ion transition state. In the retaining enzymes the two catalytic carboxyl groups are situated around 5.5 Å apart from one and another, which facilitates the formation of a covalent glycosyl-enzyme intermediate (*Figure 4a*) (Lovering *et al.*, 2005; Sinnott & Souchard, 1973), whereas in the inverting enzymes the distance between the two catalytic carboxyl groups is usually larger and varies considerably, the increased distance allows the insertion of a water molecule that can act as an external nucleophile and

attack the substrate (*Figure 4b*.) (Zechel & Withers, 2000; McCarter & Withers, 1994).

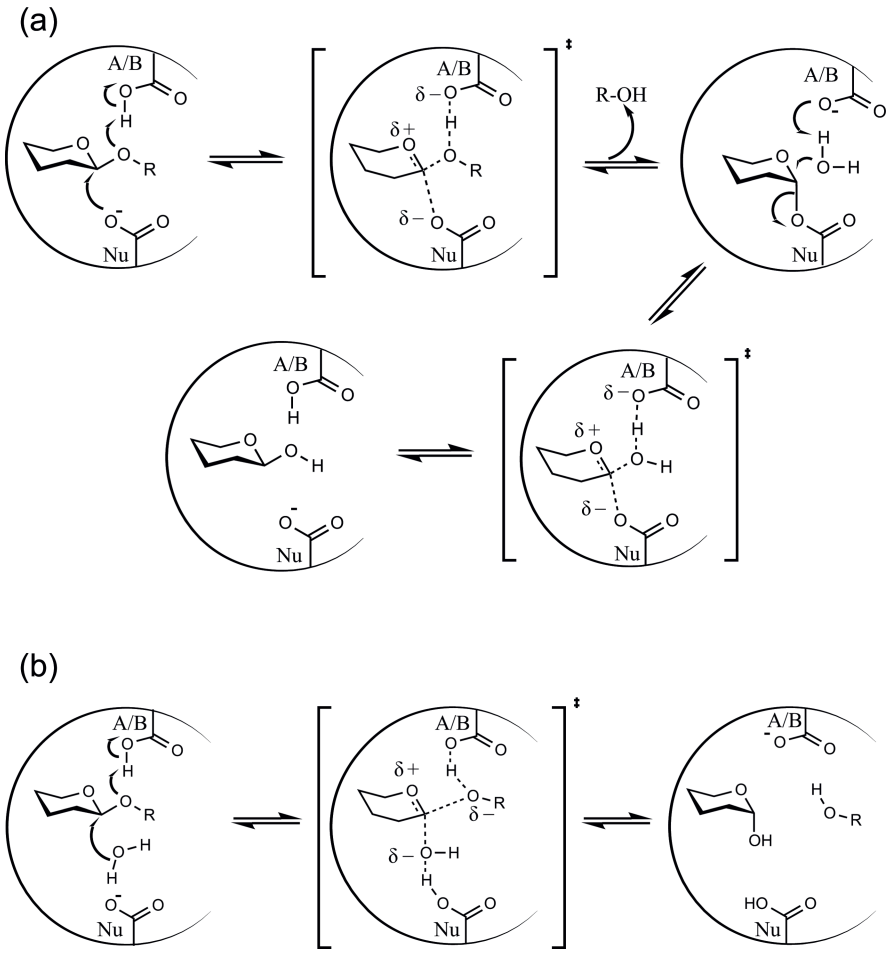


Figure 4. Proposed mechanism of retaining glycosidases **(a)**. Proposed mechanism for inverting glycosidases **(b)**. A/B refers to the acid/base residue; Nu refers to the nucleophile residue.

In 1997, Davies *et. al.* proposed a system of nomenclature for the numbering of sugar-binding subsites in glycoside hydrolase active sites (Davies et al., 1997). In sugar chemistry carbohydrate oligomers are by convention represented by drawing the non-reducing end on the left-hand side and the reducing end on the right-hand side, and thus numbering them from left to right. Glycoside hydrolases bind and hydrolyse, at minimum, a single carbohydrate attached via a glycosidic bond to another carbohydrate or moiety. The generally accepted numbering scheme for the subsites in GH enzymes starts from the point of

cleavage and labels the binding sites from -1 to –n (where n is an integer), towards the non-reducing end of the sugar polymer, and from +1 to +n towards the reducing end. The cleavage thus occurs between subsites –1 and +1 sites.

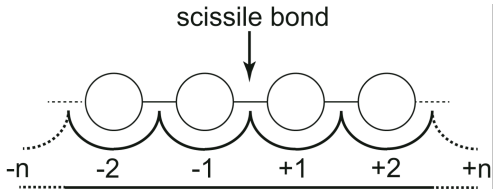


Figure 5. Schematic of glycosides hydrolase binding site nomenclature. Binding sub-sites are labelled from –n from the reducing end to +n at the reducing end. Cleavage occurs between –1 and +1 binding sites, the scissile bond.

2.7 Glycoside hydrolase family 3

Glycoside hydrolase family 3 (GH3) is one of the largest enzyme families in the CAZy database (Lombard *et al.*, 2014) and as of this writing contains over 6800 gene sequences. These sequences are widely distributed in bacteria, fungi plants and have also been found in some archaea. They carry out a range of biological functions concerning biomass degradation, cell wall remodelling and pathogen defence. The activities found in GH3 include β-glucosidase (BGL) β-D-xylopyranosidase (BXL), *N*-acetyl-β-D-glucosaminidase (s), α-L-arabinofuranosidase, exo 1,3-β-glucanase and 1,4-β-glucanase as well as several activities concerning the hydrolysis of glycoconjugates containing at least a single β–linked glucopyranoside (Harvey *et al.*, 2000).

A common feature amongst GH3 enzymes is their broad substrate specificity in respect to single monosaccharide residues, the position of linkage and chain lengths present in the substrates. ExoP from *Hordeum vulgare* (barley), classified as a β-D-glucan glucohydrolase, has been shown to be able to remove a single non-reducing end glycosyl residue from a wide range of β-glucans, oligosaccharides, Aryl β-D-glucopyranosides, cyanogenic β-D-glucosides and some β-D-oligoxyloglucosides (Hrmova & Fincher, 1997). Some GH3 enzymes are characterized as bifunctional like ARA-I from barley, which has both α-L-arabinofuranosidase and β-D-xylopyranosidase activity (Lee *et al.*, 2003) and Nag3 from *Cellulomonas fimi* characterized as a *N*-acetyl-β-D-glucosaminidase/β-glucosidase (Mayer *et al.*, 2006).

On class of GH3 enzymes that does not adhere to the family's propensity for promiscuity are the GH3 *N*-acetyl-β-D-glucosaminidases (NagZs). The, predominantly prokaryotic, GH3 NagZ are selective for *N*-acetyl-β-D-glucosamine (GlcNAc) (with a notable exception of the aforementioned *Cellu-*

lomonas fimi Nag3 (Mayer *et al.*, 2006)) and have been found involved in cell wall recycling of gram-negative bacteria (Litzinger *et al.*, 2010; Cheng *et al.*, 2000). They are also notable for having members composed solely of an A-domain (*Figure 7*). In fact, this have been found to be case for most NagZs from Gram-negative bacteria (Bacik *et al.*, 2012) The enzyme NagZ from *Bacillus subtilis* (*Bs*NagZ) have been shown to utilize a catalytic mechanism that, while retaining the GH3 aspartate nucleophile, utilizes as the acid/base a histidine-aspartate dyad located on a flexible loop on the A-domain, not on the B-domain (which is present in the case of *Bs*NagZ) (Litzinger *et al.*, 2010).

GH3 is a rather large enzyme family and there have been some efforts to identify and classify sub family clusters within GH3. In 2000 Harvey *et. al.* presented an analysis of the 99 GH3 genes annotated in CAZY at the time, and identified six major GH3 clusters. However, the analysis may have been hampered because available activity data was scarce (Harvey *et al.*, 2000). A later study from 2003 by Cournoyer and Faure analysed about 500 GH3 gene sequences that they organized into sub-families, clusters and sub-clusters as well as highlighting species and reported activities (*Figure 6*; (Cournoyer & Faure, 2003)). The GH3 sub-families AB, AB' and AB'' represent different domain arrangements. Enzymes in the largest GH3 sub-family, AB, have an *N*-terminal triosephosphate isomerase (TIM) barrel domain (A), coloured in blue in *Figure 7a-f*, containing the catalytic aspartate nucleophile, followed by a β-sandwich domain (B), coloured in yellow in *Figure 7b-f*, containing the catalytic glutamate acid/base. B' refers to a truncated form of domain B (Faure *et al.*, 1999) and B'' refers to the type of β-sandwich domain found in the NagZs.

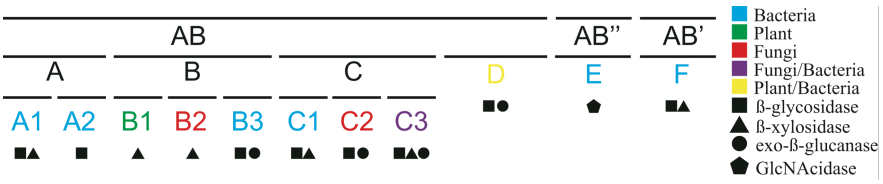


Figure 6. Phylogenetic organization of the GH3 as organized by Faure *et. al.* GH3 sub-families (AB, AB' and AB''), clusters (A, B, C, D, E, and F), and sub-clusters (A1, A2, B1, B2, B3, C1, C2, C3) are shown, species representation in each sub cluster designated by colour: bacteria are indicated in blue, from fungi in red, plants in green, both fungi and bacteria in purple and plants and bacteria in yellow. The reported activities (EC) are shown by black shapes: β-glucosidases and laminaribiase in squares, β-xylosidase in triangles, exo-β-glucanase in circles, and *N*-acetylglucosaminidase in pentagons.

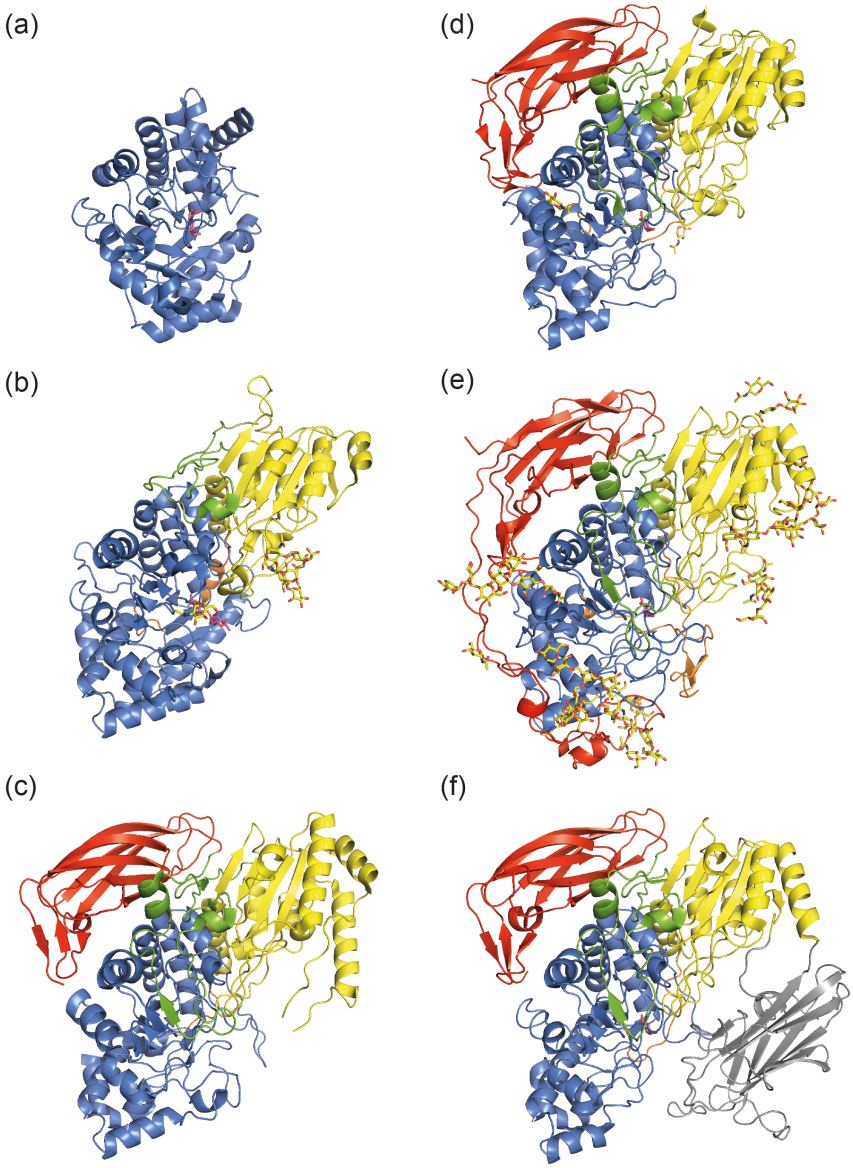


Figure 7. Ribbon representation of **(a)** *Vc*NagZ (PDB ID: 1Y65), **(b)** *Hv*ExoI (PDB ID: 1IEX), **(c)** *Tn*Bgl3B (PDB ID: 2X41), **(d)** *Hj*Cel3A (PDB ID: 3ZYZ), **(e)** *Re*Cel3A (PDB ID: 4D0J) and **(f)** *Km*BglI (PDB ID: 3AC0). Domain A coloured in blue, domain B coloured in yellow, FnIII-domain coloured in red, linker 1 coloured in orange, linker 2 coloured in green, PA14-domain coloured in grey. Active site bound glycans shown in magenta sticks, *N*-glycosylations are shown in yellow sticks.

2.8 Structure and function in GH3

The first published crystal structure of a GH3 enzyme was that of barley (*Hordeum vulgare*) the structure of the barley β-glucan exohydrolase isoenzyme ExoI (*Hv*ExoI, *Figure 7b*.) (Varghese *et al.*, 1999). *Hv*ExoI consists of 603 amino acids (aa) organized into two domains, one *N*-terminal $(\alpha/\beta)_8$ TIM-barrel domain followed by a $(\alpha/\beta)_6$ β-sandwich domain. The substrate-binding pocket is formed between the two domains of the enzyme. The first domain provides the catalytic nucleophile (Asp285) and the second domain provides a glutamate (Glu491) as the catalytic acid/base. This assignment was subsequently validated by solving crystal structures of *Hv*ExoI containing non-hydrolysable thio-glycoside Michaelis-complex mimics (Hrmova *et al.*, 2004; Hrmova *et al.*, 2002).

A decade after the first structure of barley *Hv*ExoI, the structure of Bgl3B from the thermophilic bacterium *Thermotoga neapolitana* (*Tn*Bgl3B, *Figure 7c*) (Pozzo *et al.*, 2010) was published. Interestingly, *Tn*Bgl3B has a third domain in addition to the standard GH3 AB two-domain core and falls into subcluster C2 in phylogenetic tree. The third domain is an FnIII-like domain (coloured in red in *Figure 7c-f*) that straddles the A and B domains on the opposite side to the active site cleft. FnIII-domains are found in ~2% of all known proteins, and they are as well found in many GH families. The function of this domain is still not known, but since the FnIII domains observed in GH3 is far from the active site cleft in GH3s, a substrate binding related function seems unlikely. Rather, it is more likely to confer structural or stabilizing properties. In *Tn*Bgl3B and other FnIII containing GH3 structures, the *N*-terminal domain lacks two α-helices of the canonical $(\alpha/\beta)_8$ TIM-barrel. They are replaced with a short loop that forms an antiparallel β-sheet, in practice making the *N*-terminal domain a $\beta\beta(\beta/\alpha)_6$ TIM-barrel like domain.

In 2010 the first four domain containing GH3 structure, β-glucosidase I from the yeast *Kluyveromyces marxianus* (*Km*BglI, *Figure 7f*) was published (Yoshida *et al.*, 2010). *Km*BglI has a 172 residue PA14-domain inserted in the middle of domain B, coloured in grey in *Figure 7f*. This PA14-domain is a β-barrel domain found in a variety of bacterial and eukaryotic glycosidases, glycosyl transferases, adhesins and cell signalling proteins (Rigden *et al.*, 2004). The main function of the PA14-domain is attributed to binding of carbohydrate containing ligands rather than catalysis. PA14-domain containing GH3 proteins are located in sub-clusters C3 and C1, and fungal enzymes like *Km*BglI are exclusively located in sub-cluster C3. The PA14 containing GH3 enzymes exhibit a substrate preference for smaller carbohydrate containing substrates compared to other GH3 β-glucosidases. Zmudka *et. al.* did in 2012 published the structure of the PA14 containing GH3 enzyme DesR from *Streptomyces*

venezuelae. DesR is activating macrolide antibiotics by the hydrolysis of an attached β-linked glucose moiety (Zmudka *et al.*, 2013; Zhao *et al.*, 2003). The PA14-domain in DesR is rotated by 116° compared to *Km*BglI, and it has been hypothesized that the biological function of this type of GH3 proteins is not as oligosaccharide-degrading enzymes but as part in organism defence (Zmudka *et al.*, 2013).

More recently the structure of an industrially relevant fungal β-glucosidase, BGL1 from *Aspergillus aculeatus* (*Aa*BGL1) was published (Suzuki *et al.*, 2013). *Aa*BGL1 is a three-domain GH3 β-glucosidase, belonging to the C2-sub-cluster, and being structurally overall similar to *Tn*Bgl3B. In this thesis the structure of and biochemical characterization of an additional 3 GH family 3 enzymes will be described.

2.9 Lytic polysaccharide monooxygenases

As has been described above the discovery and characterisation of various GHs has revolutionised our understanding of microbial plant cell wall degradation. However polysaccharides do often pack together into tight crystal lattices that the GHs must overcome and act upon. Decrystallization of polysaccharides, especially highly ordered and crystalline structures such as cellulose and chitin, requires thermodynamic work (Payne *et al.*, 2013; Beckham & Crowley, 2011; Beckham *et al.*, 2011; Cho *et al.*, 2011; Payne *et al.*, 2011), and over 60 years ago Reese and co-workers hypothesised that organisms utilize other mechanisms in addition to GHs to depolymerize plant cell walls (Reese *et al.*, 1950). It was not until 2010 when the crucial discovery of a group of enzymes, now called lytic polysaccharide monooxygenases (LPMOs), overturned the classical cellulose depolymerisation paradigm and validated the Reese *et. al.* hypothesis (Vaaje-Kolstad *et al.*, 2010). LPMOs utilized a completely novel mechanism of for cleaving glycosidic, bonds and even before their mode of action had been discovered this group of proteins had been known to act synergistically together with traditional GH enzyme cocktails, but without any knowledge how they did this (Harris *et al.*, 2010; Vaaje-Kolstad *et al.*, 2005a).

In 2005 the small, apparently non-catalytic, chitin-binding protein 21 (CBP21) was isolated from the chitinolytic bacterium *Serratia marcescens,* which seemingly enhanced the activities of the endogenous chitinases ChiA and ChiB from *S. marcescens*. ChiA and ChiB are processive enzymes acting in opposite directions along a chitin polymer. Thus, in the chitinolytic enzyme system, ChiA and ChiB are equivalents to the two types of processive cellobiohydrolases (CBHs) of GH family 6 and GH family 7, respectively, found in fungal cellulose degrading enzyme systems (Vaaje-Kolstad *et al.*, 2005a;

Vaaje-Kolstad *et al.*, 2005b). *Sm*CBP21 was at that time classified as a member of carbohydrate-binding module family 33 (CBM33) in the CAZy classification. This CBM family has now been reclassified into auxiliary activity family 10 (AA10) in the CAZy database. It was speculated that this was a class of CBMs that could bind-to and thereby some how disrupt the chitin polymer and thus facilitate accessibility for the hydrolytic enzymes. Similar results for CBM33/AA10 proteins from *Thermobifida fusca* were also reported in 2007 by David Wilson and co-workers (Moser *et al.*, 2008), but with cellulose activity being enhanced by the *T. fusca* CBM33 protein.

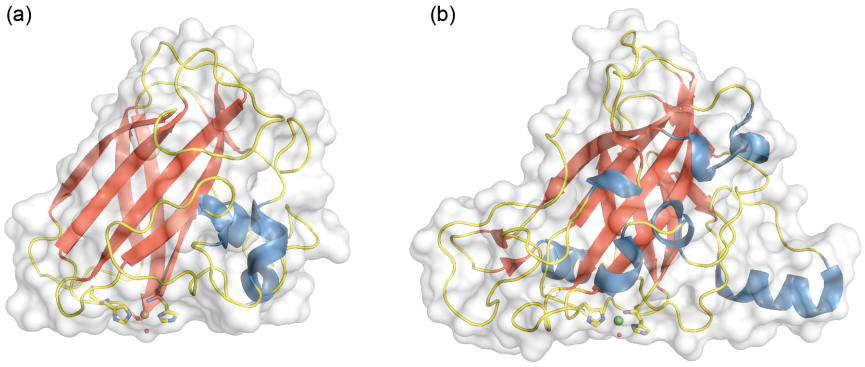


Figure 8. Structures of the chitin active *Ef*CBM33A from *Enterococcus faecalis* (new name *Ef*LPMO10A, PDB ID: 4ALC) **(a)** and the cellulose active *Hj*Cel61B rom *Hypocrea jecorina* (alternative or new names *Hj*LPMO9B and EG7, PDB ID: 2VTC) **(b)**.

Genes encoding for CBM33/AA10 enzymes have mainly been found in bacteria and viruses, but are exceedingly rare in eukaryotes (there are five annotated eukaryotic sequences as of this writing). Surprisingly when the first crystal structure of a fungal protein from the enigmatic glycoside hydrolase family 61 (GH61, now reclassified as AA9), Cel61B from *Hypocrea jecorina* (*Hj*Cel61B), was published in 2008, and later when the structure of GH61E from *Thielavia terrestris* (*Tt*GH61E) was published, there were several striking structural similarities between this structure and to the one of the bacterial enzyme *Sm*CBP21 (*Figure 8*) (Harris *et al.*, 2010; Karkehabadi *et al.*, 2008a). Both *Sm*CBP21 and the two GH61 have a metal binding site consisting of two histidine residues that form a "histidine brace" (Quinlan *et al.*, 2011). One of the histidines is the *N*-terminal residue of the enzyme, and in addition to the δ-amine of the sidechain also the *N*-terminal amine itself takes part of the ligation of the bound metal. In the metal binding site of *Sm*CBP21 a sodium atom was modelled and in the *Hj*Cel61B a nickel atom was modelled. Harris *et. al.* also performed mutations in the metal binding site, removed metals using EDTA and tested different divalent metal ions. They showed that without any of the

histidines the synergism disappeared as it also did when there was no metal present (Harris *et al.*, 2010). In both GH61 structural reports, the lack of conserved and suitably positioned carboxylate pairs for GH action was noted, further indicating the miss-attributed GH classification of the GH61 family

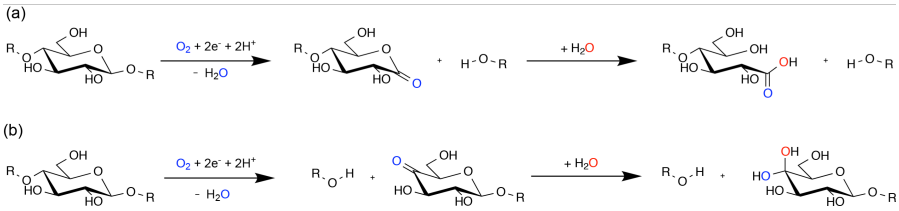


Figure 9. Oxidative cleavage of cellulose by a C1-oxidizing LPMO produces a lactone species that is subsequently hydrolysed to an aldonic acid **(a)** A C4-oxidizing LPMO that produces a 4-keto sugar that via hydrolysis can form a gemdiol **(b)**.

The aforementioned seminal study presented in 2010 by Vaaje-Kolstad *et. al.* showed unambiguously that *Sm*CBP21 is an enzyme that cleaves glycosidic bonds in chitin via an oxidative reaction mechanism (Vaaje-Kolstad *et al.*, 2010). *Sm*CBP21 generated normal non-reducing ends and non-reducing chain ends that are C1-oxidized carbohydrates called aldonic acids (*Figure 9a*). It was also shown that *Sm*CBP21 activity is drastically boosted by the presence of electron donors such as ascorbic acid or other reducing agents, and inhibited by metal-chelators such as EDTA that can chelate divalent metal ions such as copper. Experiments with ^{18}O-labelled O_2 or H_2O confirmed that the reaction involves molecular oxygen and that one of the oxygen atoms is incorporated in the sugar product, while the other O atom goes to water, thus demonstrating that the enzyme is a monooxygenase (*Figure 9*).

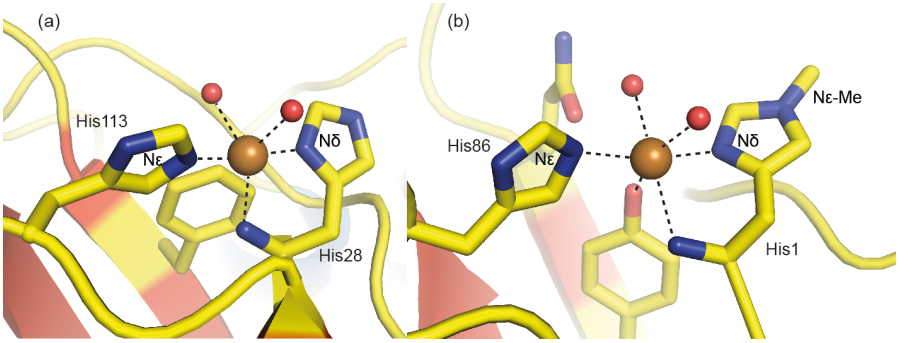


Figure 10. The trigon-bipyrimidal copper binding sites of *Enterococcus faecalis* CBM33A (now *Ef*LPMO10A PDB ID: 4ALC) **(a)** and the octahedral copper binding-site of GH61D from *Phanerochaete chrysosporium* (now *Pc*LPMO9D, PDB ID: 4B5Q) **(b)**. Of note are that the *N*-terminal amine is a ligand of copper atom, and the methylated Nε of His1 in *Pc*LPMO9D.

Following the initial discoveries of the LPMO oxidative catalytic reaction mechanisms of both AA9 and AA10 representatives, several characteristics and features of LPMOs were elucidated. LPMOs utilizes copper (Forsberg *et al.*, 2011; Phillips *et al.*, 2011; Quinlan *et al.*, 2011), coordinated in a mononuclear type(II) copper binding site (Gudmundsson *et al.*, 2014; Hemsworth *et al.*, 2013b), and can oxidize either the C1 or the C4 carbon, generating aldonic acids and 4-keto sugars respectively (*Figure 9*). Structures of the oxidized, Cu(II) and photoreduced , Cu(I) state of bacterial AA10 LPMOs have been presented (Paper IV in this thesis) (Gudmundsson *et al.*, 2014; Hemsworth *et al.*, 2013b). Also Aachmann *et. al.* has shown that CBP21 binds Cu(I) with higher affinity than Cu(II) (Aachmann *et al.*, 2012). In fungal AA9 LPMOs, the ε-amine in the *N*-terminal histidine has been found to be methylated (*Figure 10b*) (Quinlan *et al.*, 2011). However, the function of this modification is unclear. There are examples of fungal AA9 LPMOs that are active without methylation at His1 (Wu *et al.*, 2013a), and no effect of methylation was seen in quantum mechanical calculations of the reaction mechanism (Kim *et al.*, 2014). More recently LPMOs with activity towards soluble oligosaccharides (Isaksen *et al.*, 2014), xyloglucan (Agger *et al.*, 2014) and starch (Vu *et al.*, 2014) have been discovered. Revealing that the LPMO field is not limited to enzymes acting on crystalline cellulose and chitin, but appears to be a more widely used mechanism in regards carbohydrate cleavage.

Langston *et. al.* has shown that a *Thermoascus aurantiacus* LPMO (*Ta*GH61) can utilize cellobiose dehydrogenase (CDH) from *Humicola insolens* (*Hi*CDH) as a reducing agent (Langston *et al.*, 2011). It has been known that CDHs are genetically co-regulated with cellulases, but the true function for CDHs besides oxidizing cellobiose has yet not been elucidated (Henriksson *et al.*, 2000; Henriksson *et al.*, 1995). CDHs are two-domain enzymes consisting of a flavin adenine dinucleotide (FAD) domain that oxidizes cellobiose and a heme domain that is believed to be responsible for transferring the electrons generated by the FAD domain to an electron acceptor. The ability of CDHs to act as electron donors to LPMOs points to a putative role of CDHs in lignocellulose degradation.

3 Experimental

3.1 Protein X-ray crystallography

Macromolecular X-ray crystallography is a technique that allows for the observation of atoms and molecules organized into a crystal lattice. Atoms in a crystal lattice scatter incident X-rays that diffract in many directions. By measuring the angles and intensities of the diffracting beams it is possible to retrieve an image of the electron density of the atoms in the crystal. In the electron density the relative placements of atoms, the bonds connecting them and information of disorder can be modelled.

The first protein structure to be determined was that of myoglobin in 1958 by Sir John Cowdery Kendrew and his research colleagues in the Cavendish Laboratory at Cambridge University, UK (Kendrew *et al.*, 1958). After this initial discovery the number of protein crystal structures have grown immensely until date. The field has been awarded a total 28 Nobel Prizes and the main repository for protein structures, the PDB (www.rcsb.org/pdb), currently contains almost a hundred thousand entries for protein structures determined by X-ray (Berman *et al.*, 2000).

Macromolecular crystals, unlike their small inorganic and organic counterparts, constituted of up to 90% solvent. The interactions holding the crystal lattices together are weak, mainly hydrogen bonds and van der Waals interactions. Dehydration, radiation damage or any other disturbance of the crystal lattice can quickly damage a protein crystal irreparably. For a diffraction experiment to be successful the crystallinity must be kept intact.

The initial and biggest bottleneck for macromolecular structure solution is the growth of good diffraction quality crystals. Although the parameters of crystallization are more and more well understood, it is not possible to predict under which conditions a given protein will crystallize. The standard approach is too coarse-screen a wide range of crystallization conditions; such as buffer

type, pH, small molecule additives, temperature, precipitants (salts, solvents and polymers), ions and temperature, in hope of finding a hit that can be optimized by fine-tuning the initial crystallization condition found.

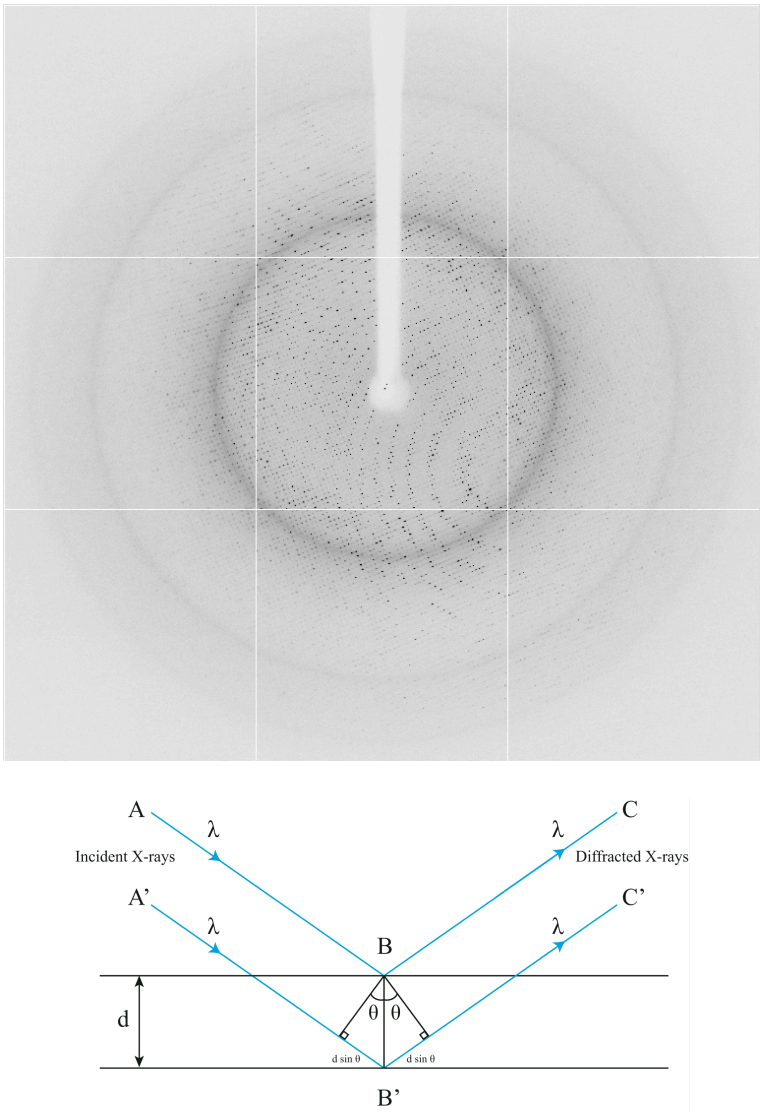


Figure 11. Top: Protein X-ray diffraction pattern of a protein crystal exposed in an X-ray beam. Below: Bragg diffraction. When two beams of identical wavelength and phase travels the paths ABC and A'B'C' the lower beam travels an extra length of 2d sin θ. Constructive interference occurs when that length equals an integer multiple of the wavelength (λ).

Optimized crystals are retrieved from a crystallization experiment via cryoloops and immediately frozen by plunging them into liquid nitrogen, which has a temperature of approximately 77K (−196 °C). The freezing of the crystal prohibits the formation of ice crystals formed in the solution the macromolecular crystals are embedded in. Successful freezing of protein crystals generally requires the presence of a cryoprotectant such as e.g. glycerol or various types of polyethylene glycol. The diffraction experiment is then carried out either at a home X-ray source or at one of the synchrotrons available around the world, e.g. Max-Lab in Lund, Sweden. Synchrotrons provide high flux densities (photons per s per mm^2), highly focused and tuneable incident X-ray beams. When an incident X-ray beam interacts with the electrons around the atoms in a crystal the X-rays are scattered elastically in all directions. These scattered beams will diffract, meaning that they will interfere with each other constructively and destructively resulting in a diffraction pattern, as can be seen in the top part of *Figure 11*. Diffraction peaks from a crystal will occur according to Braggs law (*Figure 11 bottom*) (Bragg, 1913). To be able to calculate the electron density of a molecule from the diffraction pattern collected from a crystal, the position and intensity of the reflections must be indexed and integrated.

When making the physical measurements the phase information of the beam is unfortunately lost, since light detectors, like CCDs, only measure the intensity of light that hits them not the phase of the wave of the beam. This phase problem can be solved by different means, several involving the effects of heavy metals (or atoms significantly heavier than carbon oxygen and nitrogen, the main building blocks of a protein molecule) on the diffraction pattern but the most widely used technique to sole a structure of a macro molecule nowadays is called molecular replacement (MR). In this phasing technique one utilizes an already known structures of homologous structures to the one to be determined as search models.

These processes to handle collected X-ray diffraction data are handled by crystallographic software packages such as e.g. the Collaborative Computational Project Number 4 (CCP4) software package (Winn *et al.*, 2011) or PHENIX (Adams *et al.*, 2010). After an initial structure model is obtained of a macromolecule such as a protein molecule, the structure model is further gradually improved by iterative building cycles using manual protein building software like *Coot* (Emsley *et al.*, 2010), and maximum-likelihood structure refinement software like REFMAC (Murshudov *et al.*, 1997). The refined structure model is thereafter validated to ensure that the structure model do not violate and geometry and other structure quality parameters before being deposited at the protein data bank (PDB) (Berman *et al.*, 2000).

3.1.1 Time resolved X-ray crystallography

To more fully elucidate the intricacies of a biochemical process, it is necessary to study not only the resting state but also initial and end states, as well as intermediates along the reaction co-ordinates. X-ray crystallography gives, in principle, an average structure of the crystallized component thus careful consideration needs to be taken when one wants to study the short lived transient states that an enzyme and its substrate adopt. In the case of redox enzymes, where the catalysed reactions involve several oxidation states and intermediates of their metal co-factors, this is not trivial since the X-rays during the data-collection liberate electrons that readily alter the redox state of active site (Sjögren & Hajdu, 2001; Schlichting *et al.*, 2000; Logan *et al.*, 1996; Chance *et al.*, 1984; Chance *et al.*, 1980). That X-ray radiation induces damage to crystalline macromolecules have been known for a long time (Blake & Phillips, 1962), and the dose limits for a cryo-cooled crystals is commonly taken as the 2×10^7 Gy (1 Gray (Gy) = 1 J kg^{-1}), the dose which reduces the diffraction pattern to half of its original intensity, calculated by Henderson (Henderson, 1990). However the metals in redox enzymes active sites, in part by having evolved to channel electrons to an oxidized active site, are readily reduced by very low absorbed doses of X-rays (e.g. 3 MGy for reduction of 50 % of Mn(II) centres in photosystem II (Yano *et al.*, 2005)). Thus to be able to capture the structure of a redox centre at a certain oxidation state care must be taken to minimize the amount and impact of these side reactions.

In the seminal 2002 Nature paper by Berglund and co-workers they used the approach of spreading the X-ray dose over many protein crystals combined with single crystal microspectrophotometry that could probe the redox state at a specific X-ray dose (Berglund *et al.*, 2002). To be able to retrieve a complete dataset at a given dose of each dataset on each crystal was collected with a certain offset, i.e. the starting angle for crystal one was 0°, 10° for crystal two, and so forth until the whole angular range, the angle range needed to determine the structure, had been covered. Then each frame of images of e.g. 10° from each dataset was merged to create a set of composite datasets each representing the structure at a given dose.

In Paper IV (Section 4.4) of this thesis we performed a similar experiment on a bacterial LPMO. Whereas in the Berglund *et. al.* paper they utilized multiple crystals we performed this on one single crystal. We could achieve this by utilizing the rod-like shape of the crystal used for the experiment and using a very thin X-ray beam that can be achieved at a state of the art synchrotron. By doing so we could distribute the X-ray dose over the length of the crystal by translating the crystal along its axis during the whole duration of the data collection and thereby continuously feeding a new non-exposed part of the crystal

into the X-ray beam. This data collection strategy allowed us to simply repeat the data-collection following the same path through the crystal each time to retrieve successive dataset with an increased X-ray dose for each of the collected data sets.

4 Results and discussion

4.1 Biochemical Characterization and Crystal Structures of a Fungal GH3 β-Glucosidase, Cel3A from *Hypocrea jecorina* (Paper I)

Hypocrea jecorina as have been described is model organism for cellulose degradation and a workhorse for enzyme production. *H. jecorina* secretes three GH3 enzymes, Cel3A, Cel3B and Cel3E (Foreman *et al.*, 2003; Barnett *et al.*, 1991; Chirico & Brown, 1987b; Chirico & Brown, 1987a). *Hj*Cel3A comprises ~1% of the secretome of *H. jecorina* (Gritzali M. & D., 1979), a level that is suboptimal for *in vitro* cellulose degradation. It has been shown that supplementing *H. jecorina* whole cellulase mixture with endogenous or exogenous β-glucosidases significantly boost the degradation efficiency of this (Banerjee *et al.*, 2010; Salvi *et al.*, 2010; Kumar & Wyman, 2009). At the time of this projects initiation the only published eukaryotic GH3 structure was that of the barley 1-3/1-4-β-glucanase, *Hordeum vulgare* ExoI (PDB ID: 1EX1) (Varghese *et al.*, 1999). This lack of structural information then on GH family 3 β-glucosidase triggered further efforts to solve structures of this type of enzymes and was also the rational for further biochemical and structural characterizations of GH3 enzymes from the model organism *Hypocrea jecorina.* In the first paper included in this thesis we present the crystal structures of *Hj*Cel3A expressed in either in *H. jecorina* or *Pichia pastoris,* and a glucose soaked ligand complex structure. We also characterized the *Hj*Cel3A enzyme kinetics on a range of natural and synthetic substrates as well as presenting the increased efficiency of saccharification by increased amounts of HjCel3A in *H. jecorina* whole cellulase.

4.1.1 Production of *Hj*Cel3A from *H. jecorina* and *P. pastoris*

The native *HjCel3A* gene (UniProt Q12715) was cloned into and subsequently overexpressed in a *H. jecorina* strain lacking four genes coding for cellulases (*cbh1, cbh2, egl1, egl2*). Transformants were grown in Vogel's minimal medium with glucose and sophorose as carbon source (Vogel, 1956). The overexpressed protein constituted 80 % of the produced and secreted protein as determined by SDS-PAGE. The protein was subsequently purified using a Superdex 200 16/60 GL column (GE Healthcare) and concentrated. The purity of the enzyme was determined by SDS-PAGE and isoelectric focusing.

The production of *Hj*Cel3A in *P. pastoris* (*Pp-Hj*Cel3A) was performed by cloning and transforming the *Hjcel3A* gene into *P. pastoris* X33 (Invitrogen). Transformants were grown in BMGY medium and expression was performed in BMM medium with addition of methanol as an inducer. *Pp-Hj*Cel3A was concentrated and buffer-exchanged into 25 mM sodium acetate, pH 5.0 for crystallization.

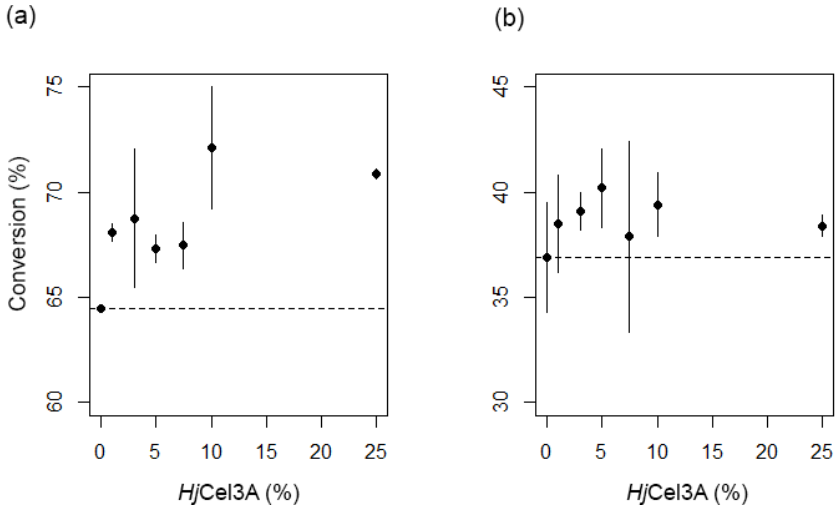


Figure 12. The effect of *Hj*Cel3A produced in *H. jecorina* on saccharification of PASC **(a)** and PCS **(b)** by whole cellulase. The total enzyme loading was kept constant at 20 mg of enzyme/g of cellulose (solid symbols). X axis, weight percentage of *Hj*Cel3A; horizontal lines, conversion of whole cellulase without added *Hj*Cel3A; vertical lines, percentage of *Hj*Cel3A at optimal conversion; error bars, S.D. of quadruplicate assays.

4.1.2 Effects of *Hj*Cel3A on Cellulose Degradation by Cellulase Mixtures

As stated in section 1.5 cellobiose is a major inhibitor of enzymatic cellulose degradation, especially *Hj*Cel7A in case of *H. jecorina* (Gruno *et al.*, 2004). The hydrolysis of cellobiose to glucose by β-glucosidases will synergistically enhance the efficiency of cellulose degradation. A portion of the whole cellu-

lase mixture was replaced with purified *Hj*Cel3A on an equal protein basis, the conversion of both phosphoric acid swollen cellulose (PASC) and pretreated corn stover (PCS) increased (*Figure 12*). Optimal conversion was observed for mixtures that had increased *Hj*Cel3A levels (Fujdala & Larenas, 2010).

4.1.3 Broad substrate specificity of *Hj*Cel3A

*Hj*Cel3A has been known to hydrolyse oligosaccharides from a degree of polymerisation (DP) of 2 to DP6, as well as of various β-linked disaccharides (Korotkova *et al.*, 2009). We determined the kinetic parameters for *Hj*Cel3A in detail for a range of oligosaccharides (cellobiose to cellotetraose), β-1,2 , β-1,3 and β-1,6 linked disaccharides (sophorose, laminaribiose and gentiobiose) and the chromogenic substrate CNPG. The values are summarized in Table 1. The results confirm the broad substrate specificity of *Hj*Cel3A. It is thus capable of hydrolysing different β-linked disaccharides with the highest affinity for the β-1,3-D-linked laminaribiose. Similar results have also been observed for two other well-characterized GH family 3 β-glucosidases *Hv*ExoI and *Tn*Bgl3B (Hrmova *et al.*, 2002; Zverlov *et al.*, 1997). Our experimental data also show that *Hj*Cel3A prefers hydrolysis of slightly longer oligosaccharides than disaccharides (i.e. cellotriose and cellotetraose. For longer substrates, subsites further away from the active site may be important for specificity (Hrmova *et al.*, 1998; Hrmova *et al.*, 1995) The dramatic increase in specificity for the hydrolysis of cellotriose versus cellobiose suggests that the +2 subsite should be important for activity on β-1,4-linked oligosaccharides.

Table 1. *Hj*Cel3A enzyme kinetics

Substrate	K_M	k_{cat}	k_{cat}/K_M
	mM	s^{-1}	$M^{-1}s^{-1}$
Cellobiose	0.35 ± 0.04	16.0 ± 0.48	0.5 x10^5
Cellotriose	0.036 ± 0.006	31 ± 0.8	8.5 x10^5
Cellotetraose	0.036 ± 0.006	24 ± 0.11	8.0 x10^5
Gentiobiose	0.53 ± 0.08	8.0 ± 0.40	0.2 x10^5
Laminaribiose	0.25 ± 0.03	28.0 ± 0.84	1.1 x10^5
Sophorose	0.45 ± 0.03	23.0 ± 0.46	0.5 x10^5
CNPG	0.087 ± 0.01	28.0 ± 1.12	3.2 x10^5

In the structure of *Hv*Exo1 in complex with thio-linked laminaribiose (Hrmova *et al.*, 2002), it was shown that the glucopyranose residue in the +1 subsite bound with the β (apolar) face toward Trp286, whereas in the structure complex with thio-linked cellobiose, the glucopyranose residue bound with the less

hydrophobic α face toward Trp286 (*Figure 13b*). In *Hj*Cel3A, the corresponding tryptophan residue, Trp237, has swung inward to the +1 subsite, whereas Trp37, with the side chain oriented 90° in relation to Trp237, together with Phe260, forms a hydrophobic patch at the approximate location of Trp286 of *Hv*Exo1. On the other side of the +1 subsite, the phenyl ring of Tyr443 is likely to play the same role as the benzene ring of the Trp434 in *Hv*Exo1. The hydrophobic patch may better complement the β face of a predicted bound laminaribiose than the assumed α face that would be the case for cellobiose and sophorose (*Figure 13a*). This may contribute to the preference of *Hj*Cel3A for (1-3)- over (1-4)- and (1-2)- β-D-linked disaccharides seen in our experiments.

Judged by the structure, a putative +2 subsite of *Hj*Cel3A has only two residues, Asp370 and Phe260 that are in position to interact with a glucopyranose. None of these two residues are conserved among GH3 enzymes, although they commonly occur in sequences of fungal GH3 β-glucosidases. Given the importance of the +2 subsite in *Hj*Cel3A, it is surprising that the assumed +2 site cannot show more potential interactions with a glycan residue.

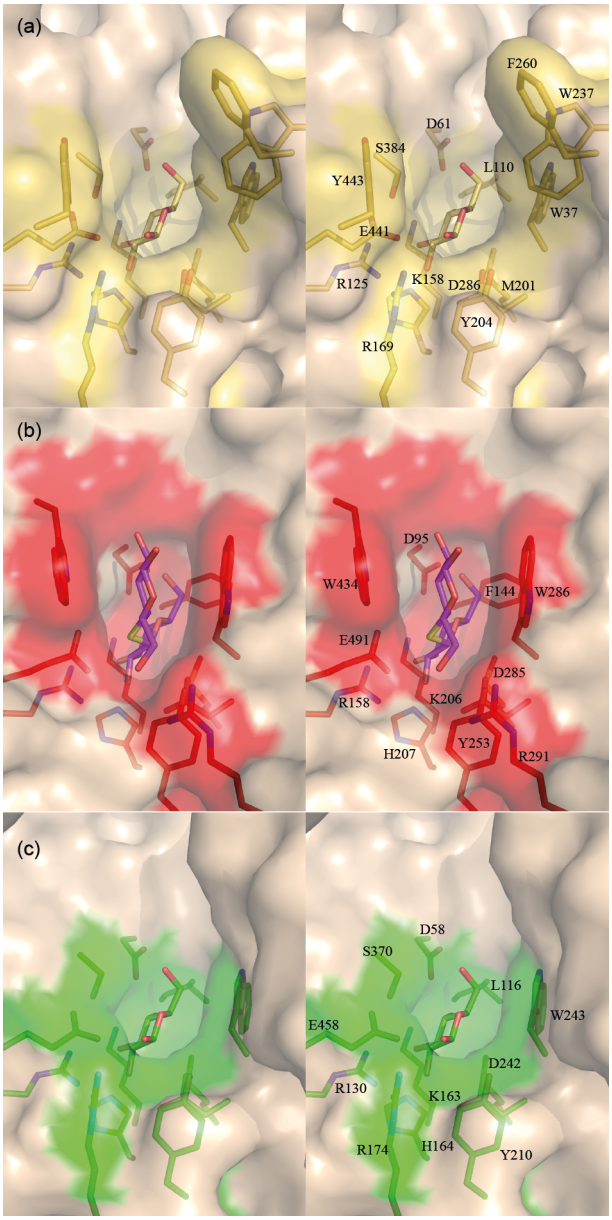


Figure 13. Stereoviews of the -1 and +1 substrate binding subsites with a van der Waals surface representation of *Hypocrea jecorina* Cel3A + glucose, yellow (PDB ID: 3ZYZ) **(a)**, *Hordeum vulgare* Exo1, red, + thio- cellobiose, violet (PDB ID: 1IEX) **(b)**, and *Thermotoga neapolitana* Bgl3B + glucose, green (PDB ID: 2X41) **(c)**.

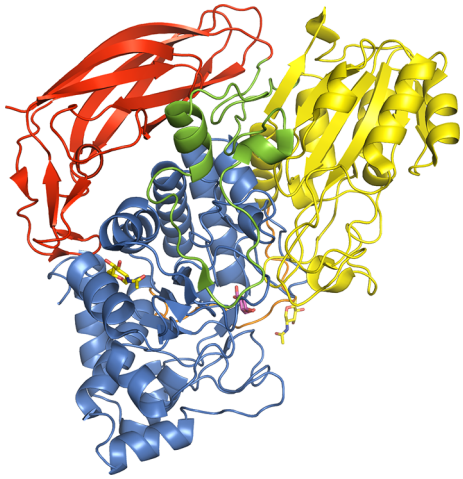


Figure 14. Schematic cartoon representation of the overall structure of *Hj*Cel3A. The three domains are colored blue (A domain), yellow (B domain), and red (FnIII domain). The two domain linker regions are shown in orange (linker 1) and green (linker 2). The glucose bound in the active site and the *N*-glycans are depicted in magenta and yellow sticks respectively.

4.1.4 HjCel3A Crystal Structures

H. jecorina-expressed *Hj*Cel3A crystallized with one molecule in the asymmetric unit in space group $P2_12_12_1$ for both the apo and glucose-complexed forms and the *P. pastoris*-expressed *Hj*Cel3A (*Pp*-*Hj*Cel3A) crystallized with two molecules in the asymmetric unit in space group $C2_1$. The structures were solved to 2.1 Å (*Hj*Cel3A, *Hj*Cel3A + glucose) and 2.5 Å (*Pp*-*Hj*Cel3A + glucose). The crystallographic *R*-factors for the final structure models of the *Hj*Cel3A, *Hj*Cel3A-glucose complex, and *Pp*-*Hj*Cel3A are 17.5, 18.3, and 20.1%, respectively, whereas the *R*-free values are 22.2, 22.8, and 27.0%, respectively.

The overall fold of *Hj*Cel3A bears a strong resemblance to that of *Tn*Bgl3B (Pozzo *et al.*, 2010) and is composed of three distinct domains (*Figure 14*). Secondary-structure matching (SSM) superposition (Krissinel & Henrick, 2004) of the *Hj*Cel3A structure and of the *Tn*Bgl3B gives a root mean square distance (RMSD) of 1.6 Å. The A-domain encompasses residues 7–300. This domain is joined to the B-domain with a 16-residue long linker (residues 301–316). The B-domain, a five-stranded α/β sandwich, comprises residues 317–522 and is followed by a third domain, the FnIII-domain, which is composed of residues 580–714 and has an immunoglobulin type topology. The folds represented by domains A and B together are present in many GH3 β-glucosidases, and the fold was first described for *Hv*Exo1 (Varghese *et al.*, 1999). Whereas the A-domain of *Hv*Exo1 has a canonical TIM barrel fold, with an alternating repeat of eight α-helices and eight parallel β-strands in an α/β barrel, domain A

of *Hj*Cel3A lacks three of the parallel β-strands and the two intervening α-helices. Similarly to what was reported for *Tn*Bgl3B, domain A has instead three short antiparallel β-strands, which, together with five parallel β-strands and six α-helices, form an incomplete or collapsed TIM-barrel like $\beta\beta(\beta/\alpha)_6$ fold.

4.1.5 Active Site Geometry and Substrate Specificity

As in other GH3 β-glucosidases the *Hj*Cel3A active site is situated at the interface between domain A and B. The nucleophile and acid/base active site residues of *Hj*Cel3A (Asp236 and Glu441) have their side chains pointing toward the active site and are present at almost identical positions as the catalytic residues of *Hv*ExoI (Asp285 and Glu491) and TnBgl3B (Asp242 and Glu458) (*Figure 13a-c*). The structures of apo-*Hj*Cel3A and *Hj*Cel3A-glucose are essentially identical except for the presence of a glucose molecule bound at the -1 subsite in the complex structure. Clear density is observed for this glucose accommodated at the -1 subsite. The -1 subsite seems to have a highly conserved com- position among the structures of *Hj*Cel3A, *Tn*Bgl3B, and *Hv*ExoI (Pozzo *et al.*, 2010; Hrmova *et al.*, 2005; Hrmova *et al.*, 2004; Varghese *et al.*, 1999) with a tight network of hydrogen bonding involving residues Asp62, Arg125, Lys158, His159, Tyr204, Asp236, and Glu441 of *Hj*Cel3A.

In comparison to *Tn*Bgl3B, *Hj*Cel3A possesses a much more shallow active site binding cleft. On domain A there are two loops that connect α-helix Asn24–Ser33 and β-strand Gly44– Thr46 and β-strand Leu57–Gln60 and α-helix Pro77–Thr84 that are both shorter in the structure of *Hj*Cel3A than the corresponding loops of the *Tn*Bgl3B structure. Opposite the extended loops of domain A *Tn*Bgl3B also has peculiar loop protruding from an α-helix, (residues 393-410) that substitutes a β-strand of canonical β-sandwich domain, which is partly unmodelled but is situated so that a significant deepening of the active site cleft occurs compared to *Hj*Cel3A.

The active site loop of *Hj*Cel3A (residues 438 – 460) in domain B in which the acid/base amino acid Glu441 is situated is 8 residues longer (Tyr443–Ala450) in *Hj*Cel3A than the corresponding loops in *Hv*ExoI and *Tn*Bgl3B. At each side of this loop there are two glycine residues (Gly442 and 451). Glycines are known to promote structural flexibility (Schwarzinger *et al.*, 2002; Callebaut *et al.*, 1994). These glycines could potentially serve as hinges for opening and closing of this loop and thus play a role in activity and/or substrate recognition of the enzyme in regards to its substrate. In the structure of *Hj*Cel3A, this loop is bent away from the active site, and as a result, the side chain of Tyr443 is brought to a position that has become an integral part of the +1 subsite of the enzyme (*Figure 13a*).

4.1.6 *Hj*Cel3A and *Pp-Hj*Cel3A comparison and *N*-Glycosylation

*Pp-Hj*Cel3A is considerably more glycosylated than *Hj*Cel3A. In SDS- PAGE analysis, the *Pp-Hj*Cel3A sample ran at a significantly higher apparent molecular mass than *Hj*Cel3A. Previous studies on fungal GH3 β-glucosidases shown no difference in activity and stability of heterologously produced forms from *P. pastoris* (Liu *et al.*, 2012; Hong *et al.*, 2007). Limited characterization of the two forms on short cellooligosaccharides show that this applies to *Hj*Cel3A as well.

*Hj*Cel3A was analyzed for glycosylation sites by peptide analysis using MS. For six of the seven *N*-glycosylation sequons that are present in *Hj*Cel3A, the glycosylated form was identified (positions 45, 208, 310, 417, 566, and 613). Although on average three sites were glycosylated due to the presence of multiple glycoforms. For the *Pp-Hj*Cel3A all seven site were found to be glycosylated. The two monomers in the asymmetric unit of *Pp-Hj*Cel3A appear to have an orientation with an axis of pseudosymmetry between the Asn208 *N*-glycosylations, for which continuous electron density supports eight glycans that have been modeled as two NAGs and six mannose residues.

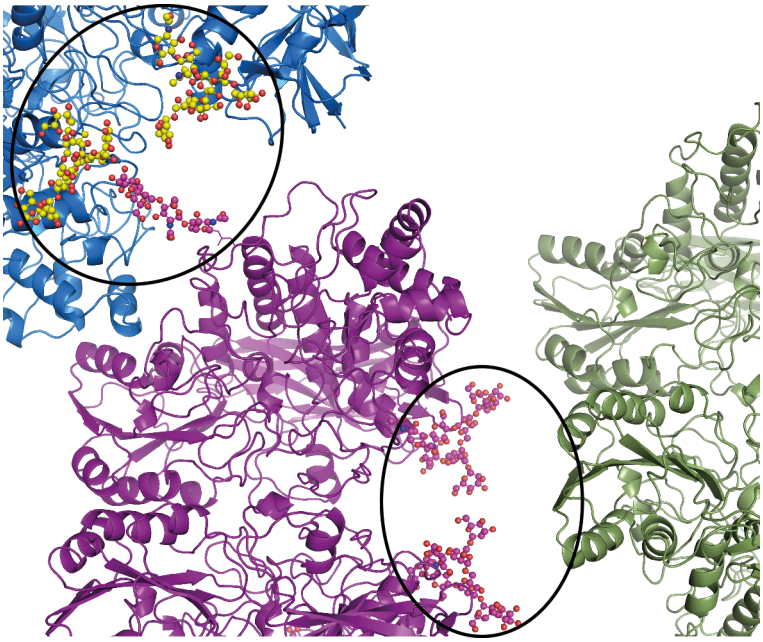


Figure 15. Three asymmetric units of *Pp-Hj*Cel3A, each containing two protein molecules, are shown in blue, purple, and green. The crystal contacts with van der Waals interaction between glycan molecules are indicated by black circles. The intermolecular interactions involve both of the glycosylation sites at Asn208, located at the dimer interface of the non-crystallographic symmetry-related molecules, and at Asn310.

4.1.7 Conclusions

The β-glucosidase *Hj*Cel3A from *H. jecorina* is an enzyme of industrial relevance and, as part of cellulase mixtures, contributes to the efficient production of fermentable sugars from lignocellulosic substrates. Enhanced levels of *Hj*Cel3A in *H. jecorina* cellulase mixtures benefit cellulose conversion. *Hj*Cel3A contributes to cellulose degradation by converting cellobiose to glucose and is especially efficient in converting longer cellooligosaccharides to glucose.

*Hj*Cel3A was produced in *H. jecorina* as well as in *P. pastoris*. Both samples crystallized, and the determined protein structures are essentially identical. The -1 subsite of *Hj*Cel3A is well conserved and displays the same geometry as determined for other GH3 β-glucosidases. The +1 subsite of *Hj*Cel3A appears to be narrower than those of the other known GH3 β-glucosidases for which the three-dimensional structure has been determined. Features were identified that are in agreement with the determined substrate specificity of *Hj*Cel3A.

The *N*-linked glycosylations attached to the enzymes, which are more extensive in the *P. pastoris*-produced sample changes neither the thermal stability nor the activity of the enzyme. It appears to be a coincidence that in both structures, the same two asparagine residues are glycosylated, and in both structures, the glycosylations are involved in crystal contacts and as such contribute to the formation of the crystallographic space group. The protein structures obtained from the two samples are identical; thus, the main effect of the differences in glycosylation appears to be crystallization in either the $P2_12_12_1$ or $C2_1$ space groups.

4.2 Structural and Functional Studies of Glycoside Hydrolase Family 3 β-Glucosidase Cel3A from the Moderately Thermophilic Fungus *Rasamsonia emersonii* (Paper II)

We did in a previous study show (Paper I) that enriching the *H. jecorina* secretome with additional amounts of the endogenous β-glucosidase Cel3A (*Hj*Cel3A) increase the performance of the mix for conversion of cellulose to glucose (Karkehabadi et al., 2014; Barnett et al., 1991). An approach for further improving the enzyme mixtures is to substitute components with homologs from alternative sources. The most widely studies biomass degrading enzymes are from organisms, such as *Trichoderma*, *Aspergillus* and *Penicillium*, that are active in a temperature range of about 40 to 50 °C. Increasing the reaction temperature has a large impact and advantages on enzymatic biomass degradation via increased hydrolytic rates, enhanced mass transfer and decreased

substrate viscosity. The caveat for enzymes being that they quickly become unstable and inactivated at elevated temperatures. To overcome this deficiency of enzymes from mesophilic fungi one can deploy enzymes from thermophilic hosts. Enzymes from thermophilic fungi can often tolerate higher temperatures than their mesophilic counterparts, with some being active up to 70 – 80 °C (Margaritis & Merchant, 1986; Margaritis & Merchant, 1984). It has been reported that cellolytic ability of thermophilic species was several times higher than for mesophiles and examples of their enzymes showing higher hydrolytic capabilities although their extracellular enzyme titres were lower than for traditionally used species (Wojtczak *et al.*, 1987; Tansey, 1971).

We did in our study find that the secreted β-glucosidase Cel3A from the thermophilic fungi *Rasamsonia emersonii* (previously *Talaromyces emersonii*) (*Re*Cel3A) yielded significant performance boost when supplemented in cellulase enzyme mixtures compared to adding same amount of *Hj*Cel3A. We subsequently biochemically characterised *Re*Cel3A and solved the three-dimensional structure of the enzyme to 2.2 Å.

4.2.1 Production of *Re*Cel3A

The *cel3a* gene from *R. emersonii* (Gene bank nr. AAL69548.3) was cloned and transformed into a derivative of *H. jecorina* RL-P37 deleted for four major cellulases (*cel5A*, *cel6A*, *cel7A, and cel7B)* (Foreman *et al.*, 2005). Successful transformants were subsequently grown on Vogel's minimal medium (Vogel, 1956) with a mixture of glucose and sophorose as carbon source. The resulting *H. jecorina* strain did express *Re*Cel3A at levels of more than several grams per litre, constituting more than 50 % of the total secreted protein, as judged by SDS-PAGE. The concentrated culture filtrate was then purified for crystallization on a Superdex 200 16/60 GL column (GE Healthcare) followed by concentration.

4.2.2 Biochemical properties of *Re*Cel3A

Previously we demonstrated that *H. Jecorina* cellulase mixtures with increased levels of native β-glucosidase *Hj*Cel3A have enhanced cellulose degradation activity (Paper I) (Karkehabadi *et al.*, 2014). Here *H. Jecorina* P37 D*bgl1* whole cellulase (lacking *Hj*Cel3A) mixtures supplemented with increasing levels of either *Re*Cel3A or *Hj*Cel3A were compared for the degradation of dilute acid pretreated corn stover (PCS) (*Figure 16*). The mixtures containing *Re*-Cel3A showed up to a 25 % increase of glucose release compared to mixtures with an equal amount of *Hj*Cel3A added.

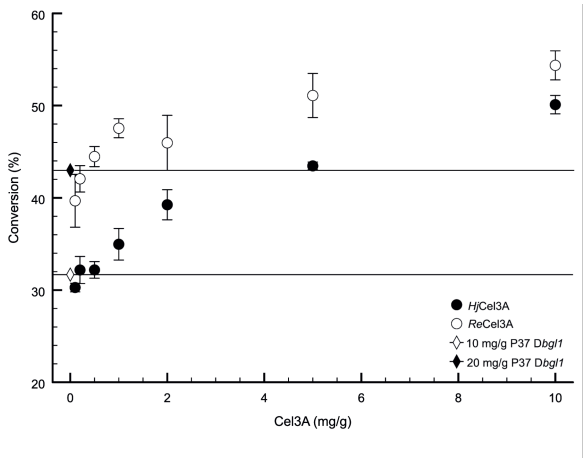


Figure 16. Saccharification of acid pretreated corn stover with 10 mg/g *H. jecorina* strain P37 Δ*bgl1* supplemented with 0.1 – 10 mg/g β-glucosidases. Horizontal lines indicate the conversion levels of 10 and 20 mg/g P37 Δ*bgl1* as indicated.

Table 2. *Kinetic parameters of ReCel3A with chromophoric substrates, and disaccharides and oligosaccharides.*

Substrate	K_M	k_{cat}	k_{cat}/K_M
	mM	s^{-1}	$M^{-1}S^{-1}$
CNPG	0.40	14	3.6×10^4
p-NPG	0.40	5.4	1.4×10^4
CNPX	0.66	0.23	3.5×10^2
cellobiose	0.78	5.5	7.1×10^3
cellotriose	0.39	0.72	1.9×10^3

Table 3. *Melting temperatures of HjCel3A and ReCel3A with and without the presence of glucose.*

β-glucosidase	Tm (°C)
*Hj*Cel3A	77.6
*Hj*Cel3A + 1 mM glucose	79.0
*Re*Cel3A	87.3
*Re*Cel3A + 1 mM glucose	87.3

Initial biochemical characterization of *Re*Cel3A has been performed by Murray *et. al.* (Murray *et al.*, 2004). In that study it was shown that *Re*Cel3A was a rather thermostable GH3 β-glucosidase and retained much of its activity even at higher temperatures. We compared the melting temperatures of *Re*Cel3A and *Hj*Cel3A, shown in Table *3*. The melting temperature of *Re*Cel3A, 87.3 °C, is 10 °C higher than the 77.6 °C of *Hj*Cel3A in absence of glucose. In the

presence of 1 mM glucose *Hj*Cel3A shows a slight increase in melting temperature to 79 °C, whereas that of *Re*Cel3A is unaffected. We subsequently investigated the enzymatic properties of *Re*Cel3A on different soluble glucan substrates, seen in Table *2*. The highest catalytic efficiency of *Re*Cel3A, among the substrates tested was for hydrolysing CNPG, but more interestingly there was a higher K_{cat}/K_m towards cellobiose over cellotriose. Hrmova *et. al.* (Hrmova *et al.*, 1998) has previously shown that the barley β-glucanase ExoI (*Hv*ExoI) has an increased affinity towards longer cellodextrins. This is also the case for *Hj*Cel3A (Karkehabadi *et al.*, 2014), which indicates, combined with their reported broad substrate affinity, that hydrolysing accumulating cellobiose during the degradation of cellulose might not be their primary or only biological function. The superior performance of *Re*Cel3A on PCS, compared to *Hj*Cel3A, could be explained by its apparent preference for cellobiose compared to other types of disaccharides and cellodextrins and potentially also its higher stability compared to *Hj*Cel3A.

4.2.3 *Re*Cel3A crystal structure

*Re*Cel3A was crystallized in the orthorhombic space group $P2_12_12_1$, with refined unit-cell parameters of; a = 137.3 Å, b = 148.6 Å, c = 196.4 Å. The molecular replacement solution using the program *Phaser* (McCoy *et al.*, 2007) gave the best solution with four protein molecules (MW = 90.4 kDa) in the asymmetric unit, with a calculated V_M of 2.77 Da^{-1} (Matthews, 1968), and an estimated solvent content of 56%. Initial phases were obtained using *H. jecorina* Cel3A (PDB ID: 3ZZ1) as search model. The *Re*Cel3A structure, at 2.2 Å resolution, was refined to final R_{work} and R_{free} values of 18.8 and 23.8, respectively. The final *Re*Cel3A structure model, consisting of four non-crystallographic symmetry (NCS) related *ReCel3A* molecules in the asymmetric unit, *Figure 17*, contains in total 3348 amino-acid residues, 1842 water molecules and a total 181 carbohydrate residues. Each protein chain consists of 834 amino acid residues and the first and last visible residue in all four *Re*Cel3A molecules in the crystal structure is D21 and P855, respectively, of the translated deposited *Re*Cel3A DNA sequence (GenBank: AAL69548.3). Residues 1 to 20 of the *Re*Cel3A sequence constitute the signal peptide, as predicted by the SignalP server (Petersen *et al.*, 2011), and is cleaved off prior the secretion of the mature protein.

4.2.4 *Re*Cel3A crystal structure model and overall fold

The crystal structure model of *Re*Cel3A is composed of four NCS related protein molecules. It has been shown that *Re*Cel3A form dimers in solution (Murray *et al.*, 2004), which was confirmed in this study when performing gel

filtration characterization of *Re*Cel3A, and is clearly supported by the crystal structure and in the *Aa*BGL1 structure (Suzuki *et al.*, 2013). The two molecules in the dimer are related by a 180 degrees rotation, *Figure 17*.

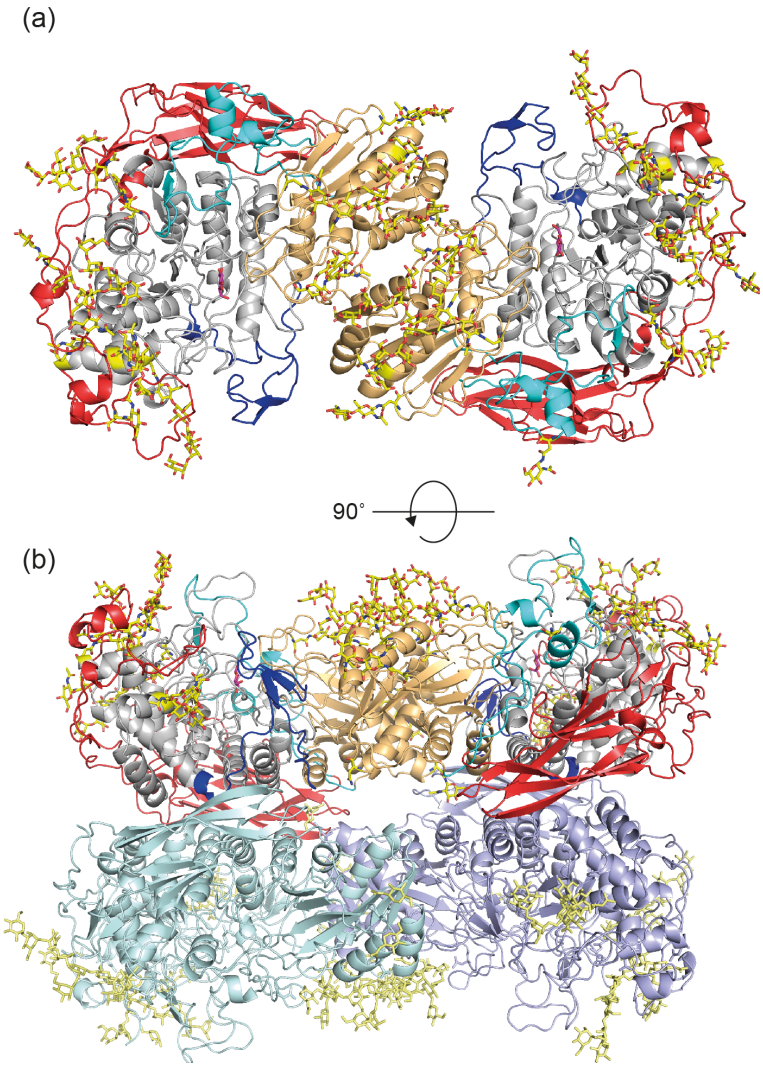


Figure 17. **(a)** Cartoon representation of the *Rasamsonia emersonii* Cel3A dimer. The A-domain is coloured in light grey, the linker-1 in dark blue, the B-domain in gold, the linker-2 in cyan, the C-terminal FnIII-domain in red, the *N*-linked glycosylations in yellow, and the glucose in the -1 subsite in depicted magenta. **(b)** The view is rotated 90° compared to (a), and now shows the tetrameric assembly of *Re*Cel3A in the asymmetric unit. The two protein chains in the second dimer are coloured in teal and purple. *N*-linked glycosylations are depicted in yellow sticks.

The overall structure of *Re*Cel3A is composed of three distinct domains, similar to *Hj*Cel3A, and can be observed in *Figure 17*. The first domain of *Re*Cel3A (residues 21 - 342 and coloured light grey in *Figure 17*) has a TIM barrel-like ββ(β/α)$_6$-barrel fold, which varies in some significant ways compared to the canonical TIM-barrel present in *Hv*ExoI (Yoshida *et al.*, 2010). The most remarkable difference is the anti-parallel direction of the second β-strand of the TIM-barrel, as was also noticed for *Thermotoga neapolitana* β-glucosidase GH3 enzyme *Tn*Bgl3B, solved by Pozzo et al. 2010 (Pozzo *et al.*, 2010). Another difference is the absence of the first and second α-helices in the TIM-barrel, a feature that also was found in the *Kluyveromyces marxianus* GH3 enzyme *Km*BglI solved by Yoshida et al., 2010 (Yoshida *et al.*, 2010). Noteworthy is also the low amount of hydrogen bonds between the β-strands in the *Re*Cel3A ββ(β/α)$_6$-barrel. In the region of the first and the last β-strand of the barrel, *Re*Cel3A has only two residues in each of the secondary structure elements, which according to the Stride definitions (Frishman & Argos, 1995) are not defined as β-strands as in the canonical TIM barrel fold but as isolated β-bridges. Each of the two loops preceding the isolated β-bridges contains one aromatic side chain that together is forming a hydrophobic patch on one side of the assumed glycan binding subsite +1. Furthermore, there are no hydrogen bonds between the fourth and fifth β-strands in this collapsed TIM barrel.

An interesting observation is that out of all currently available three-dimensional structures of GH3 proteins that share the ββ(β/α)6-barrel fold all also have a C-terminal FnIII-domain present in the overall structure organization. This third domain might contribute in stabilizing the fold of the first TIM barrel domain of the structure, and allow the otherwise stable TIM barrel to collapse during evolution and open up for changes around the catalytic centre

The A and B domains of *Re*Cel3A are linked by a 39 residues long linker (residues 343-382). This loop that previously has been observed to affect the catalytic activity of GH3 β-glucosidases (Hong *et al.*, 2006). This loop is also present in *Re*Cel3A, and constitutes part of the linker between domain A and domain B in the structure. This region of the linker is completely absent in the structures of *Hj*Cel3A, *Tn*Bgl3B and *Km*BGLI. This loop interacts with both the extended C-terminal loop of *Re*Cel3A as well as with the large glycosylation attached to Asn319 on domain one in *Re*Cel3A, *Figure 18*. One could speculate that one importance of the extended loop on domain three of *Re*Cel3A, and the proximal glycosylations, is to stabilize the hydrophobic regions of linker 1, which may in itself promote association to and binding with natural substrates. There are no proof for this hypothesis and will have to be investigated further.

The second domain of *Re*Cel3A (residues 383 – 584, coloured gold in *Figure 17*) is an $(\alpha/\beta)_6$ sandwich domain, a domain that is commonly present in GH3 enzymes and when it is present contains the glutamate that acts as general catalytic acid/base (Varghese *et al.*, 1999). The two long loops succeeding the second and forth β-strands (residues 418-457 and 502-525, respectively) in the second domain are heading towards the catalytic centre. The loop proceeding the forth β-strand of the second domain contains a short β-hairpin motif (residues 509-510 and 513-514), the catalytic general acid Glu505, and Tyr507 that is positioned to enable aromatic ring stacking with a glycan unit bound in an assumed substrate binding subsite +1.

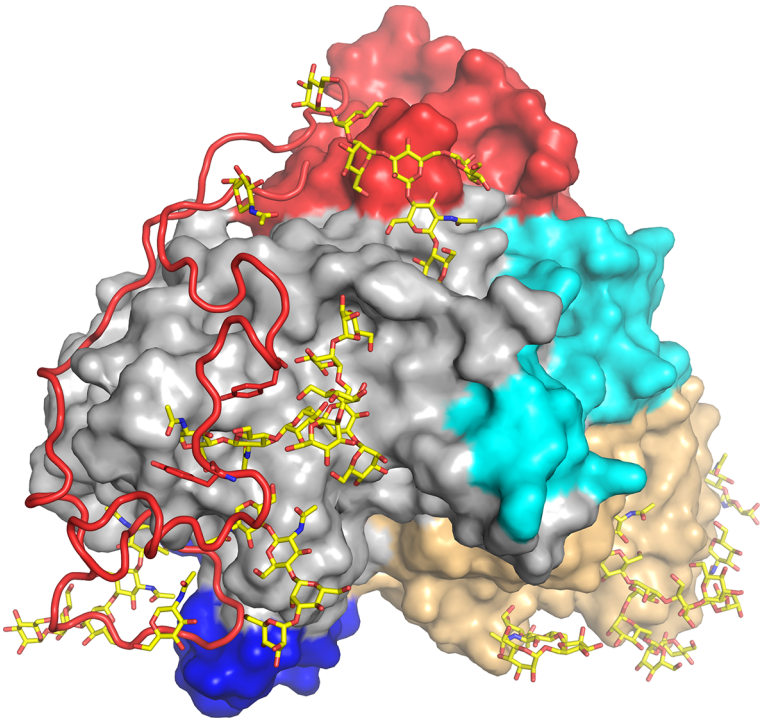


Figure 18. Surface representation of *Rasamsonia emersonii* Cel3A with the corresponding two *N*-glycosylations coloured in yellow. The extended C-terminal loop of *Re*Cel3A (red) is showed as ribbon.

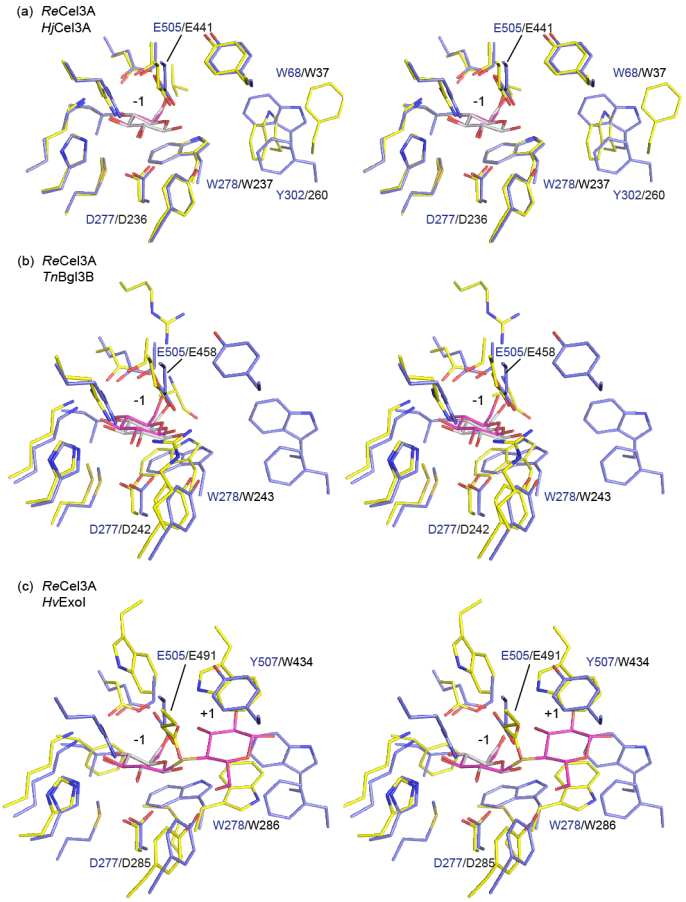


Figure 19. Stereo representations of the catalytic centres of GH3 β-glucosidases in sticks. *Rasamsonia emersonii* Cel3A is shown in yellow and the -1 glucose monomer in grey, Aligned structures are shown in blue and respective ligands from aligned structures are shown in magenta. **(a)** shows *Hypocrea jecorina* Cel3A (PDB ID: 3ZYZ), **(b)** shows *Thermotoga neapolitana* Bgl3B (PDB ID: 2X41) and **(c)** shows *Hordeum vulgare* ExoI (PDB ID: 1IEX).

The third domain of *Re*Cel3A (residues 649 – 855, coloured red in *Figure 17*) consists of seven β-strands forming two β-sheets with an FnIII or immunoglobulin-like fold, which could be classified as an s-type collapsed β-barrel (Bork *et al.*, 1994). Two additional short β-strands of domain three form a small sheet in the *N*-terminal part of the domain. After the first β-strand a 94 residues long loop protrudes (residues 663-756), embracing the *N*-terminal collapsed TIM barrel domain, shown in *Figure 18*. The tip of the loop ends up near the active site of *Re*Cel3A where it interacts with loops that build up the sides of the catalytic centre. This “arm” crosses over the extended *N*-terminus of *Re*Cel3A, and also covers three extensive glycosylation chains (chains ema-

nating from asparagine 61, 312 and 319) of the first domain. The long loop protruding from the third domain of *Re*Cel3A that "embraces" domain one, *Figure 18*, seems to be found only among members of cluster C2 of the Cournoyer & Faure classification. In *Hj*Cel3A, this loop is much shorter and consists of only 21 residues (residues 594-614 of *Hj*Cel3A) compared to 94 residues in *Re*Cel3A. The elongated linkers (residues 343–382, 585–648, and 663-756 that all are pointing toward the active centre), and contribute in deepening the substrate binding cleft in *Re*Cel3A compared to *Hj*Cel3A.

4.2.5 Binding subsites

The extensive amount of hydrogen bonds between the protein and the ligand bound in the catalytic centre of GH3 glucosidases makes the binding in subsite -1 in these enzymes highly specific. All the glucose binding residues in subsite -1 are conserved within the currently available structures of GH3 β-glucosidases, except for the tyrosine binding to the anomeric oxygen, which is substituted by a phenylalanine in ExoP from the marine bacterium *Pseudoalteromonas sp.* BB1 (Nakatani *et al.*, 2012). The two catalytic residues in *Re*Cel3A were identified based on homology to other GH3 structures, to be Asp277 (nucleophile) and Glu505 (acid/base) based on homology to other GH3 structures. In accordance with other GH3 enzymes (Hrmova *et al.*, 2001; Varghese *et al.*, 1999), Asp277 is located on the A-domain and Glu505 on the B-domain. The subsite -1 pocket is positioned at the carboxy side of the barrel-like fold of domain A and is mainly built up by amino acids from domain A of the molecule. Clear density is observed in the *Re*Cel3A structure for a glucose unit in the -1 subsite in all four NCS molecules. No indication of distortion from the relaxed chair conformation can be seen. The only direct hydrogen bond between the glucose unit in subsite -1 and domain two is through the carboxyl group of the general acid/base Glu505 and the anomeric carbon oxygen of the glucose molecule bound in the -1 subsite, with a distance of 2.62 Å. A comparison of the -1 sites and their immediate environments are shown for *Re*Cel3A overlaid with *Hj*Cel3A, *Tn*Bgl3N and *Hv*ExoI in *Figure 19*.

In *Hv*ExoI, the suggested +1 site is lined by two tryptophan residues, Trp286 and Trp434 (*Figure 19c*), which have been referred to as a "coin slot" and are proposed to be the basis of the broad substrate specificity of *Hv*ExoI (Varghese *et al.*, 1999). In *Re*Cel3A, Trp278 aligns in sequence with Trp268 in *Hv*ExoI but has a similar inward shifting towards the -1 subsite as the corresponding tryptophan residues in *Hj*Cel3A and *Tn*Bgl3B (*Figure 19a-b*). The inward shifting of the tryptophan residue breaks the "coin slot" and the rearrangement is a direct consequence of the collapsed TIM barrel in the *Re*Cel3A structure, as described previously. The -1 subsite widens when the 2nd barrel β-

strand is shorter and anti-parallel and, similar to what was reported for *Tn*Bgl3B, the +1 subsite is narrowed on one side by a loop (residues 197 – 212 in *Re*Cel3A) not present in *Hv*Exo1 while present in many β-glucosidases. The sidechains of Arg200 and Gln201 of *Re*Cel3A located on this loop would restrict the allowed conformations of a di- or oligo-saccharide substrate when it is bound with the non-reducing end in the -1 subsite. This restriction would "push" the linked glucopyranose residues towards Trp278 and the +1 subsite. Further away from the active site the same loop is extended in *Re*Cel3A compared to other β-glucosidases with known structure. Only BGL1 from Aspergillus aculeatus (*Aa*BGL1) contains a similar but slightly longer loop (Suzuki *et al.*, 2013). Next to Trp278 in *Re*Cel3A and replacing one side of the "coin slot" described for *Hv*ExoI are the side chains of Phe302 and Trp68 (*Figure 19*). These are pointing in parallel towards the active site and form a hydrophobic knob in the +1 and +2 subsites. This stacking of the phenylalanine and tryptophan residues narrows the +1 subsite close to the active site and was also observed in *Aa*BGL1.

As we showed in paper II *Hj*Cel3A prefers hydrolysis of slightly longer oligosaccharides, *i.e.* cellotriose and cellotetraose to that of cellobiose. Our data for *Re*Cel3A show that this enzyme prefers cellobiose to cellotriose. There is no increase in activity on cellotriose compared to cellobiose, which indicates that the +2 subsite contributes relatively little to the substrate recognition. For *Hj*Cel3A, the activity increased for cellotriose compared to cellobiose, thus indicating the importance of a +2 subsite for *Hj*Cel3A. The enzyme structures show that in the +1 subsite of *Re*Cel3A the plane of Trp68 sidechain has turned almost 90 degrees away from the +1 subsite when compared to the structure of *Hj*Cel3A. This allows for the stacking of Phe302 and Trp68 sidechains, as described above, and puts the phenylalanine residue in the +1 subsite rather than in the +2 subsite as in *Hj*Cel3A. As a consequence the existence of a +2 subsite is less pronounced in *Re*Cel3A than in *Hj*Cel3A, where the Phe260 also is complemented with an asparagine (Asn261 in *Hj*Cel3A) to form a +2 subsite. The lack of a +2 subsite in *Re*Cel3A could explain the activity profile for the enzyme of being a more pronounced cellobiase than *Hj*Cel3A.

4.2.6 *N*-glycosylation

*Re*Cel3A is highly *N*-glycosylated. There are a total 16 potential glycosylation sites in *Re*Cel3A, with the canonical *N*-glycosylation recognition sequence N-X-S/T. The *Re*Cel3A structure model contains a total of 181 attached glycosylation residues. In spite of this relatively generous glycosylation of the *Re*Cel3A molecules, it was possible to crystallize the protein without enzymatic removal of the *N*-glycans prior the crystallization experiments. A large number of car-

bohydrate chains attached to the *Re*Cel3A molecules in the structure model can be observed and modelled, the longest chain being composed of ten carbohydrate residues. We can also see that the glycosylation chains are contributing with crystal contacts to NCS and symmetry related molecules in the structure.

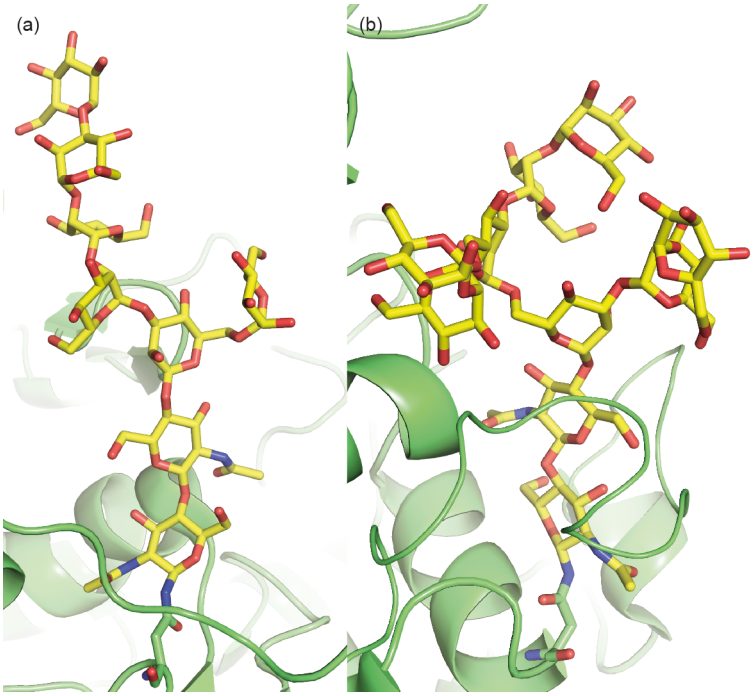


Figure 20. **(a)** $Glc_1Man_5GlcNAc_2$ *N*-glycan bound to Asn249. **(b)** $Man_8GlcNAc_2$ *N*-glycan bound to Asn319 in the *Rasamsonia emersonii* Cel3A structure.

In ReCel3A we can model carbohydrate chains, such as Man7GlcNAc2, known to be prevalent in the Rut-C30 derived strains of H. jecorina (Stals et al., 2004). Wild-type strains of *H. jecorina* show a more normal endoplasmic reticulum (ER) glycosylation trimming, yielding $Man_{5-6}GlcNAc_2$ chains. Such glycans are the result of the trimming of $Glc_3Man_9GlcNAc_2$, which then is transferred to the nascent peptide chain in the ER, by α-glucosidases found in the ER. Further trimming occurs normally by action of α-mannosidases and β-*N*-Acetylglucosaminidases. The Rut-C30 derived strains have an inefficient ER-α-glucosidase, which accounts for the presence of untrimmed monoglycosylated *N*-glycans (Stals *et al.*, 2004). We can clearly observe both the longer incompletely trimmed glycan chains, $Man_8GlcNAc_2$ (*Figure 20b*), and shorter monoglycosylated $Glc_1Man_5GlcNAc_2$ (*Figure 20a*) and $Man_{5-6}GlcNAc_2$ chains as well as single *N*-acetylglucosamine (GlcNAc) residues in the *Re*Cel3A

structure. Modelled carbohydrate glycans are not by themselves evidence of an *N*-glycosylation pattern. It is expected that glycosylation chains that are flexible and not restrained by the protein crystal packing cannot be observed through crystallographic methods. However, single GlcNAc residues are observed in the *Re*Cel3A tetramer that cannot be part of a longer glycan as they pack tightly between protein chains and presumably provide important crystal contacts.

The exact mechanisms of how glycosylations affects the structure and function of cellulases and other proteins are unknown. The classical example for cellulases is that the *O*-glycosylation on flexible linkers impart protease resistance (Langsford *et al.*, 1987). Glycosylation has been shown to increase solubility, reduce aggregation and enhance thermal stability of proteins (Kayser *et al.*, 2011; Wang *et al.*, 2010; Ioannou *et al.*, 1998). . In *Re*Cel3A most of the larger glycans reside on the first domain. The glycan chains at asparagines 438 and 519 are situated on the second domain close to the proposed dimer interface of *Re*Cel3A. The *N*-glycosylation chain at Asp319 shows a remarkable feature of being buried by the extended C-terminal loop. Two conserved aromatic residues; Tyr720 and Tyr727, on this loop provide stacking interactions to the two buried NAG residues, 1201 and 1202. As has been stated previously*Re*Cel3A most likely exists as a dimer in nature. Interestingly, the overall glycosylation pattern for the dimer shows that the active site on each of the monomer seems encircled by glycans, of which some glycans originate from the other monomer, *Figure 17*. Interestingly the opposite face of *Re*Cel3A is seemingly devoid of glycosylations, both modelled glycans and predicted sites, with the notable exception of Asn470 and its single GlcNAc residue that is only observed in two of the NCS molecules of the *Re*Cel3A structure. The GlcNAc moiety interacts with the adjacent symmetry molecule. One function of the extensive glycosylations of *Re*Cel3A, as well of other enzymes from the same GH3 subfamily, could be to protect the active site from aromatic compounds such as e.g. lignin compounds from the degraded biomass.

4.2.7 Conclusions

The GH family 3 β-glucosidase Cel3A from *R. emersonii* is an industrially relevant enzyme and an efficient supplement to whole cellulase mixtures expressed by *H. jecorina* from an application point of view for the production of fermentable sugars from lignocellulosic biomass. *Re*Cel3A appears to have a preference for disaccharides over longer β-1,4 glucans indicating a primary role for this β-glucosidase as a cellobiose in the degradation of cellulosic biomass. The three-domain structure architecture of *Re*Cel3A, the collapsed TIM-barrel, α/β sandwich and FnIII domain, also contains an extended C-terminal

loop and a relatively high number of attached *N*-glycans. The majority of the attached glycans are either covered by the extended loop present in the *Re*-Cel3A structure, or are situated at the dimer interface between two *Re*Cel3A molecules. This might suggest that the glycans are functional in the sense of stabilizing the loop covering the collapsed TIM-barrel domain, and possibly providing binding interactions at the dimeric interface. *Re*Cel3A exhibits a higher thermostability compared to *Hj*Cel3A, and enhances the PCS saccharification compared to *Hj*Cel3A. Which would make *Re*Cel3A a potential candidate for replacing an enzyme like *Hj*Cel3A in commercial enzyme mixtures for conversion of ligno-cellulosic biomass into fermentable sugars. Due to its thermal stability, *Re*Cel3A could be part of an enzyme mixture that is optimized to operate at elevated process temperatures and thereby potentially resulting in overall increased saccharification reaction rates for the process.

4.3 Structural studies of a Glycoside Hydrolase Family 3 β–glucosidase from the Model Fungus *Neurospora crassa* (Paper III)

Organisms such as the filamentous fungi *Neurospora crassa*, which in nature is found decomposing and consuming dead plant material, produces and secretes a full suite of enzymes that in an orchestral fashion are able to completely decrystallize and depolymerize cellulose as well as other plant cell wall polysaccharides (Tian *et al.*, 2009). With the knowledge that *N. crassa* is a plant degrader and as *H. jecorina* is considered a model organism because of its well characterised sexual cycle and genomics we decided to investigate one of its secreted GH3 enzymes. In the genome of *N. crassa* there are at least eight genes encoding for GH3 enzymes. Three of these genes have a signal peptide and are expected to be secreted GH3 enzymes, GH3-1 (Bgl7, NCU03641), GH3-3 (Bgl6, NCU08755) and GH3-4 (Bgl2, NCU04952). These three gene products are upregulated when wild-type *N. crassa* is grown with cellulose as the main carbon source (Wu *et al.*, 2013b). Of these, the GH3-3 and GH3-4 have been experimentally characterized as true β-glucosidases (Bohlin *et al.*, 2010; Tian *et al.*, 2009). The GH3-3 has been identified in the conidia cell walls of *N. crassa* (Maddi *et al.*, 2009) and GH3-4 was identified, using mass spectrometry, in supernatant of *N. crassa* grown on Avicel and *Miscanthus* (Tian *et al.*, 2009). Both these enzymes have homologs in other biomass degrading fungi. *E.g.,* GH3-4 and Cel3A of *Hypocrea jecorina* (*Hj*Cel3A), which is the major secreted β-glucosidase of *H. jecorina*, GH3-3 (*Nc*GH3-3) is homologous to the recently published crystal structure *Aa*BGLI (Suzuki *et al.*, 2013) and *Re*Cel3A that is presented here in Paper II, both have similar struc-

tural features as *Nc*GH3-3 and they exhibit the properties of dedicated cellobiases as well. GH3-1 is homologous to GH3 β-xylosidases.

In the study presented in Paper III we present the crystallization and structure determination of a GH3 β-glucosidase from *Neurospora crassa (Nc*GH3-3) solved to 2.25 Å resolution. These results are discussed in the light of differences and similarities to other GH3 enzymes with known structure. *Nc*GH3-3 is most similar to two recently solved GH3 β-glucosidases crystal structures *Aa*BGL1 (PDB ID: 4IIB) and *Re*Cel3A (Section 4.2, Paper II) (Suzuki *et al.*, 2013). These three GH3 structures all have in common certain major structural features such as; they all have a high number of *N*-glycosylations that can be observed in the structures of these, over 40 modeled glycans per protein molecule, curiously localized only on one face of these proteins. These enzymes also all have an extended C-terminal loop protruding from the C-terminal domain and cover large parts of the first domain and several of its *N*-glycosylations. Linkers connecting the three domains of these enzymes extend much longer towards the active site cleft of the enzyme compared to other fungal GH3 β-glucosidases such as *Hj*Cel3A and *Tn*Bgl3B. Unlike *Hj*Cel3A and *Tn*Bgl3B this class of GH3 β-glucosidases all have been shown to exist as dimers in solution (Suzuki *et al.*, 2013; Murray *et al.*, 2004).

4.3.1 Production of *Nc*GH3-3

The gene encoding *Nc*GH3-3 (GenBank: EAA26868.1) was overexpressed in a *H. jecorina* strain lacking eight genes coding for cellulases (*cbh1*, *cbh2*, *egl1*, *egl2, egl3, egl4, egl5, egl6)* and one coding for a mannanase (*man1*). Transformants of *H. jecorina* were picked from Vogel's minimal medium plates (Vogel, 1956) containing acetamide after 7 days incubation at 37 °C, and grown in liquid cultures containing Vogel's minimal medium with a mixture of glucose and sophorose as carbon sources. The overexpressed protein appeared as the dominant protein in the culture supernatants. Culture filtrate from production of *Nc*GH3-3 in *H. jecorina* was subsequently purified via gel filtration on a Superdex 200 HiLoad 16/60 column (GE Healthcare) followed by concentration for crystallization.

4.3.2 Crystallization of *Nc*GH3-3

The purified and concentrated *Nc*GH3-3 was crystallized in the orthorhombic space group $P2_12_12$ with the unit-cell parameters of a = 142.93 Å, b = 286.84 Å, c = 58.05 Å. The molecular replacement solution of the structure gave a best solution of two NCS related protein molecules (MW = 93.6 kDa) in the asymmetric unit, with a calculated V_m of 3.17 Da^{-1} (Matthews, 1968) and a solvent content of 61 %. The *Nc*GH3-3 structure was refined, at 2.25 Å resolu-

tion, to a final R_{work} and R_{free} values of 17.9 and 21.6%, respectively. The final *Nc*GH3-3 structure is composed of two NCS molecules composed of 842 and 843 protein residues respectively, 875 water molecules and 85 carbohydrate residues. No gaps in the protein chains are found in the electron density of the structure.

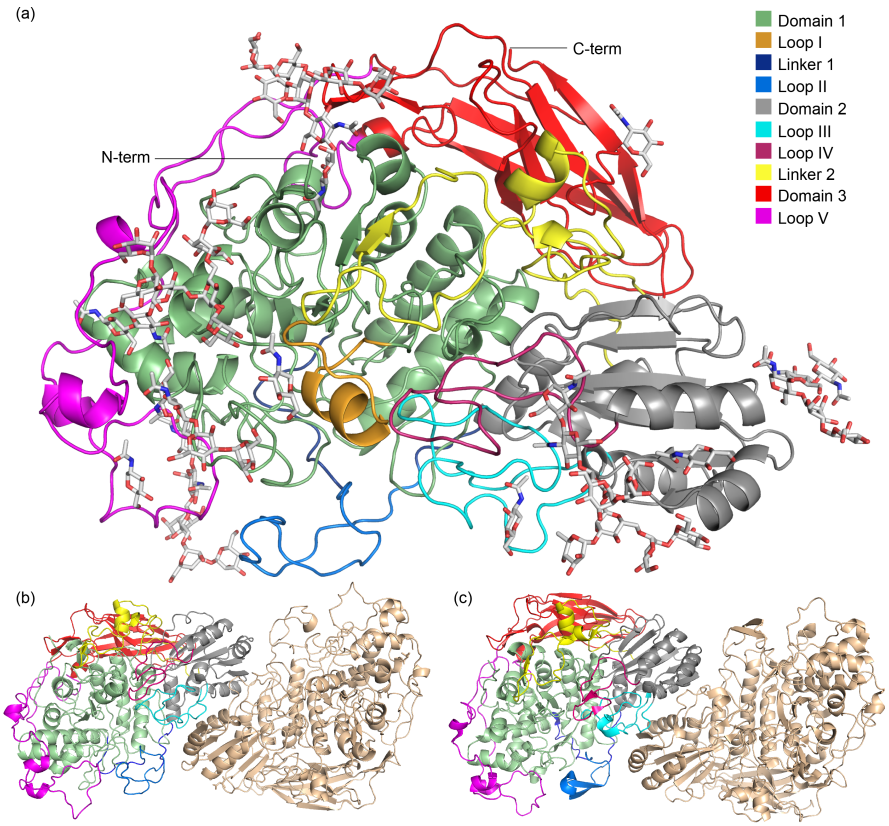


Figure 21. Cartoon representation of the *Neurospora crassa* GH3-3 structure displayed in ribbon **(a)**. The three domains of the protein are coloured in green (A-domain), grey (B-domain) and red (FnIII-domain). Loops and linkers are highlighted in colours respective to the legend in the top right corner. *N*-glycosylations are shown in grey sticks. The quaternary structure of *Nc*Gh3-3 **(b)** and *Rasamsonia emersonii* Cel3A showing the dimer formation found in the two structures **(c)**.

4.3.3 The fold and structure of *Nc*GH3-3

The *Nc*GH3-3 crystal structure model is composed of two non-crystallographically symmetry related protein molecules in the asymmetric unit. *Nc*GH3-3 chain A contains 843 amino acids and the first modelled residue is Ser34 of the deposited *Nc*GH3-3 DNA sequence (GenBank: EAA26868.1), while chain B contains 842 residues and the first modelled residue in the struc-

ture is Leu35. The last modelled residue, in both NCS related protein chains, is Pro875. The predicted signal peptide cleavage site is between residues 18 and 19 according to SignalP 4.1 (Petersen *et al.*, 2011), the non-observed residues at the *N*-terminus of the protein are thought to be non-visible due to high flexibility. The overall structure of *Nc*GH3-3 is as for the other β-glucosidases from the same GH3 subfamily composed of three separate domains, connected by two linkers and has a high degree of *N*-glycosylation. The two *Nc*GH3-3 protein chains forms a dimer similar to what has been observed for *Re*Cel3A and *Aa*BGL1 (Suzuki *et al.*, 2013). There are 11 NXT/S *N*-glycosylation sequence motives in *Nc*GH3-3 that have modelled *N*-glycan moieties in *Nc*GH3-3 structure, ranging in length from a single *N*-acetylglucosamine (GlcNAc) residues to longer $Man_7GlcNAc_2$ chains. The *Nc*GH3-3 structure is highly homologous to two GH family 3 β–glucosidase structures *Re*Cel3A and *Aa*BGL1, with 61% sequence identity to both proteins, and an RMSD of 0.73 Å and 0.81 Å, respectively. All three-domain GH3 enzymes fall into the sub-cluster C2, as specified by Cournoyer et al. (Cournoyer & Faure, 2003) Compared to the other two fungal GH3 structures available, *H. jecorina* Cel3A (*Hj*Cel3A) and *K. marxianus* BglI (*Km*BglI) the sequence identities are 46% and 31% respectively, and the RMSD, value to these two structures, are 1.28 Å and 1.56 Å respectively.

4.3.4 The A-domain of *Nc*GH3-3

The first domain of *Nc*GH3-3 (residues 34 to 340, coloured light grey in *Figure 21*) is composed of a collapsed TIM barrel fold (or $\beta\beta(\beta/\alpha)_6$ fold), described for the first time by Pozzo et. al. in the paper describing the structure of the GH3 β-glucosidase *Thermotoga neapolitana* TnBgl3B (Pozzo *et al.*, 2010) and more recently also by other groups presenting new GH3 β-glucosidase structures (Karkehabadi *et al.*, 2014; Suzuki *et al.*, 2013; Yoshida *et al.*, 2010). Domain 1 of *Nc*GH3-3 contains the catalytic nucleophile Asp276, as well as the majority of residues comprising the substrate binding subsite -1, with the catalytic centre located between subsites -1 and +1. All three-domain containing GH3 structures, where the C-terminal domain is an FNIII-like domain, lack two α-helices, compared to that of the canonical TIM-barrel of the barley GH3 structure *Hv*ExoI (Varghese *et al.*, 1999), *Hv*ExoI (*Figure 22*). Loss of protein overall stability due to the lack of central secondary elements may be mitigated by the introduction of a disulphide bridge between the two strands (Cys69 and Cys85 in *Nc*GH3-3). The missing loops allows for a more wider binding cleft in *Hj*Cel3A, whereas in *Nc*Gh3-3 this space is occupied, and extended, by loops from other domains. The structure of *Nc*GH3-3, as well as *Re*Cel3A and *Aa*BGL1, has a protruding loop (loop I in *Figure 21* and *Figure 22*), which is not present in the three-domain GH3 structures *Hj*Cel3A (*Figure 22b*), *Km*BglI

and *Tn*Bgl3B. This loop extends one side of the active-site-cleft whereas the Loop II in linker 1, (*Figure 22*), extends the opposite side of the cleft. The α-helical part of loop I presents three residues towards the active-site-cleft, Glu200, Asp203 and Tyr204. In the *Re*Cel3A structure (described in Paper II) there is a two amino acid deletion, compared to *Nc*GH3-3 and *Aa*Bgl1, resulting in a loss of α-helical structure that leaves a slightly wider active-site cleft.

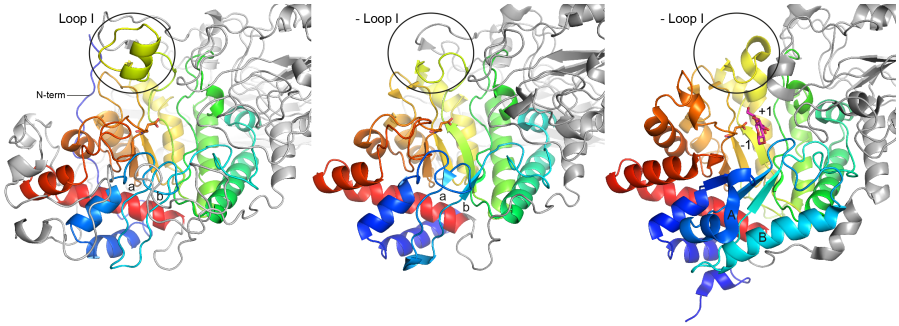


Figure 22. Cartoon representation of the A-domain of *Neurospora crassa* GH3-3 **(a)**, *Hypocrea jecorina* Cel3A **(b)** (PDB ID: 3ZYZ (Karkehabadi *et al.*, 2014)), and *Hordeum vulgare* ExoI **(c)** (PDB ID: 1IEX (Hrmova *et al.*, 2001) respectively, shown in ribbon. Loops a and b in *Nc*GH3-3 and *Hj*Cel3A highlights the deleted helices which are present and marked A and B in *Hv*ExoI. Loop I is highlighted in the circle.

4.3.5 Linker 1 and Loop II

In *Nc*GH3-3 the A-domain is connected to the B-domain by a 42 residues long linker (residues 341-383). In *Hj*Cel3A and *Hv*ExoI this linker is only 18, and 16 residues long, respectively. The insertion of 25 residues (residues 351-376) observed in *Nc*GH3-3, as well as *Re*Cel3A and *Aa*Bgl1, denoted as loop II in *Figure 23*, has previously been described as a hydrophobic linker that activates *T. aurantiacus* BGLII in organic solvents (Hong *et al.*, 2006), and is present with a similar fold also in the *Re*Cel3A and the *Aa*Bgl1 structure (Karkehabadi *et al.*, 2014; Suzuki *et al.*, 2013; Yoshida *et al.*, 2010). Loop II in *Nc*GH3-3 is not present in the other fungal GH3 structures *Hj*Cel3A (*Figure 23d*), and *Km*BGLI. This loop extends the opposite side of the active-site-cleft to that of loop I. In *Nc*GH3-3 loop II has two tryptophan residues and one phenylalanine (Trp355, Trp365 and Phe354), which constitute one side of the putative substrate binding subsites +1 and +2 (discussed in more detail in section 4.3.8). Loop II also contains two tyrosines and a tryptophan (Tyr360, Tyr371 and Trp376), which are positioned outside of the active-site-cleft. Interestingly this part of loop II resembles the flat binding surface of a carbohydrate binding molecule type one (CBM1), which also consists of two tyrosines and one tryptophan (Mattinen *et al.*, 1998). There seems to be a high variability in loop II

among GH family 3 β–glucosidase, The two aromatic residues, Phe352 and Trp355 in *Nc*GH3-3 are the most conserved residues in this loop, and these two residues are also present in the *Re*Cel3A and *Aa*Bgl1 GH3 structures (*Figure 23a, c*). Tyr360 and an aromatic residue at the position of Trp366 are also present in all loop II containing GH3 structures. In the loop II containing GH3 structures there is an *N*-glycan chain, emanating from Asn57 in *Nc*GH3-3, which is wedged in between loop II and the C-terminal loop V (*Figure 25*). This apparent feature of using *N*-glycans as an apparent structural element could be unique to this class of enzymes.

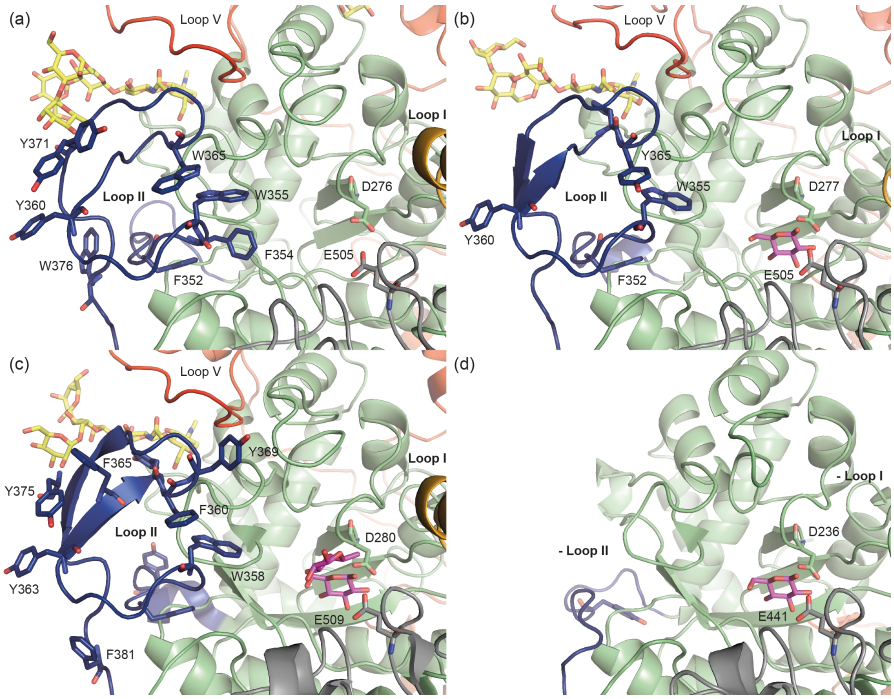


Figure 23. Cartoon ribbon representation of linker 1 and loop II (blue). Panel **(a)** shows *Neurospora crassa* GH3-3, **(b)** *Rasamsonia emersonii* Cel3A, **(c)** *Aspergillus aculeatus* Bgl1 (PDB ID: 4IIH (Suzuki *et al.*, 2013)) and **(d)** *Hypocrea jecorina* Cel3A (PDB ID: 3ZYZ (Karkehabadi *et al.*, 2014)). Domain A is coloured in green, domain B in grey, The C-terminal domain in red, loop I in orange, active site ligands in magenta sticks, active site residues in sticks and *N*-glycosylations in yellow sticks.

4.3.6 Loops in the B-domain and the second linker

The B-domain of the *Nc*GH3-3 structure (residues 383 to 584) has an $(\alpha/\beta)_6$ sandwich fold, which is structurally well preserved among all GH3 enzymes with known structure, containing at least two domains, except in the bacterial *Tn*Bgl3B, which has substituted one edge β-strand, corresponding to strand c in

Figure 24, with an additional α-helix and a flexible loop (Pozzo *et al.*, 2010). Domain B of *Nc*GH3-3 has two loops, III and IV, (residues 421- 455 and 501 – 524 respectively) that constitute one side of the active site cleft (*Figure 24*), in between loops I and II. Loop III of *Nc*GH3-3 is a 34 residues long loop extending between strand b and helix B. The loop folds back over itself and is stabilized by a disulphide bridge (residues Cys430 and Cys435). Residue Ser446 and Asp432 are two conserved residues in loop III, in the *Nc*GH3-3 structure, that are directed towards the active site. Ser446 is especially important as it is positioned pointing directly towards the hexose ring of a substrate bound in the -1 subsite. In *Hv*ExoI this position is occupied by a tryptophan residue (Trp343; *Figure 24c*), which is a very common motif in carbohydrate binding. Trp343 constitutes half of the "coin-slot" binding pocket of *Hv*ExoI (Varghese *et al.*, 1999), which is not present in any other three-domain containing GH3 enzymes. Loop IV is a 23 residue long loop where the catalytic acid/base Glu505 is located (*Figure 24*). At position 507 *Nc*GH3-3 has phenylalanine whereas the other related GH family 3 structures have a tyrosine located at this position. Domain B of the *Nc*GH3-3 structure is followed by the second linker region (residues 585 to 649). This linker is extended compared to the one that is described as a C-terminal extension in the *Hv*ExoI structure (Varghese *et al.*, 1999). This extension covers and stabilizes loop I and IV, which are not present in the barley enzyme *Hv*ExoI, as well as comprising a large part of the interface between the B-domain and the C-terminal domain.

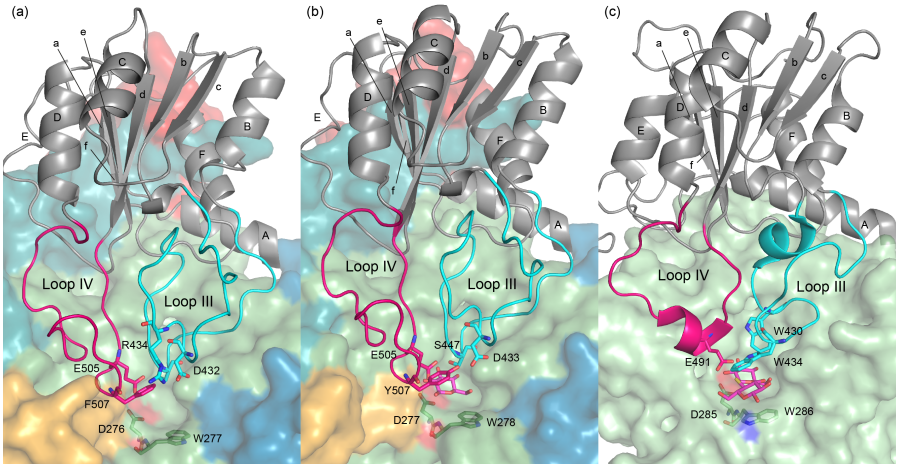


Figure 24. Overview of domain 2 (grey) in ribbon. *Neurospora crassa* GH3-3 **(a)**, *Rasamsonia emersonii* Cel3A **(b)** and *Hordeum vulgare* ExoI **(c)** (PDB ID: 1IEX (Hrmova *et al.*, 2001). Loop III and IV are highlighted in cyan and magenta respectively. Other domains are represented with surfaces coloured according to the scheme presented in *Figure 21*.

4.3.7 C-terminal domain of *Nc*GH3-3

The third domain in *Nc*GH3-3 is an FnIII-like or immunoglobulin s-type domain (residues 650 to 857), first structurally observed in GH3 in *Tn*Bgl3B (Pozzo *et al.*, 2010). The FnIII domain is a β-sandwich composed of two-layers of β-sheets of three and four β-strands, respectively. The extended loop V first observed in structures of *A. niger* Bgl1 and *Re*Cel3A, encompasses the A-domain and interacts with loop II on the opposite side of the molecule from the third domain *Figure 25*. Fascinating is that loop V folds over three large *N*-glycan chains (bound to Asn61, Asn311 and Asn318). Several conserved aromatic residues are π-stacking with GlcNAc residues in loop V (Tyr706, Tyr708, Tyr723 and Phe730). It has recently been proposed that the homologous loop V from *An*Bgl1 is flexible and would allow the FnIII-domain to extend and bind to lignin thus explaining the tad-pole-like structure that was observed by the SAXS experiments carried out with *An*Bgl1 (Lima *et al.*, 2013). We argue though that such domain reorganization as speculated upon by Lima *et al.* seems very unlikely. First, many of the conserved residues form seemingly crucial stacking interactions with *N*-glycans, a fact not accounted for in their model. Secondly, presence of flexibility within a protein crystal commonly connects to bad or even disappearing electron density, *e.g.* as in the *Ps*ExoP structure, the only published GH3 structure with a highly flexible domain, one domain is not visible in the electron density though it was expressed as part of the protein. We thus believe it is unlikely that the loop V has that degree of flexibility in *Nc*GH3-3 and other GH family 3 proteins that has this loop.

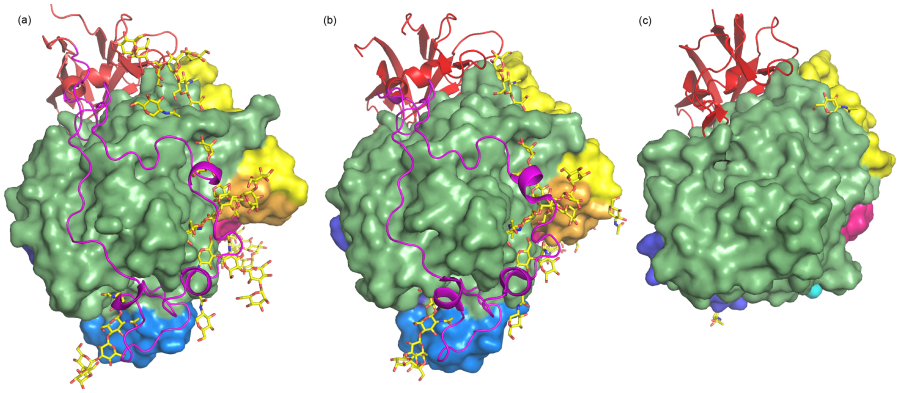


Figure 25. Domain 3 and loop V displayed in ribbon (red and magenta respectively) for *Neurospora crassa* GH3-3 **(a)**, *Rasamsonia emersonii* Cel3A **(b)** and *Hypocrea jecorina* Cel3A **(c)** (PDB ID: 3ZYZ (Karkehabadi *et al.*, 2014)). *N*-glycans attached to the three structures are displayed in yellow sticks, and other domains are represented as surfaces.

4.3.8 Catalytic subsites

The location of the catalytic centre binding subsite -1 of *Nc*GH3-3 is positioned at the carboxy side of domain one. The two catalytic residues of *Nc*GH3-3 were identified based on homology to other GH3 structures with known catalytic residues. The nucleophile of *Nc*GH3-3 is Asp276, and the acid/base is Glu505 (*Figure 23a*). In subsite -1 of *Nc*GH3-3 no distinct density corresponding to a bound glucose was observed, which often is observed for several other GH family 3 structures. There is extra electron density present in the -1 subsite that was difficult to interpret and model something into, but this density may be due to a partially bound buffer molecule, a PEG molecule and/or partial density of a bound glucose molecule.

The putative +1 subsite of *Nc*GH3-3 is very similar to that of *Re*Cel3A and *Aa*BGL1 but different from the suggested +1 subsite in *Hv*ExoI, which is lined by two tryptophan residues (*Figure 24c*), proposed to be the basis of the broad substrate specificity of *Hv*ExoI (Hrmova *et al.*, 2002). Trp277 of *Nc*GH3-3 corresponds to one of these but the sidechain has shifted to become an essential part of -1 subsite rather than +1 subsite, as in *Hv*ExoI. This shift causes a rearrangement of the core residues and contributes to the collapsing of the TIM barrel fold described above. In the collapsed TIM barrel fold, a feature that seems to be shared in many fungal and bacterial β-glucosidases, the 2nd barrel β-strand is shorter and anti-parallel, which makes the -1 subsite wider compared to the active site in GH3 enzymes with a complete TIM barrel fold. Similar to the structures of *Re*Cel3A and *Aa*BGL1, one side of the +1 subsite is formed by a phenylalanine, Phe301, which is stacking with the sidechain of Trp64 only slightly further away from the active site. On the opposite side of the +1 subsite and the active site entrance the aromatic residue Phe507 is found, which structurally seems to correspond to the "coin slot" Trp434 side chain in *Hv*Exo1 structure. The corresponding residues in the structures of *Re*-Cel3A and *Aa*BGL1 have almost the same side chain conformations but are in both these enzymes tyrosines (Tyr507 and Tyr511, respectively). In both the *Re*Cel3A and the *Aa*BGL1 structure, the hydroxyl group of the tyrosine has a hydrogen bond interaction with an aspartate residue (Asp433 and Asp437) in a potential +2 subsite (*Figure 24*). In the *Nc*GH3-3 structure, the corresponding residue (Asp432) is instead interacting with the sidechain of Arg434, which should stabilize the aspartate residue and compensate for the slight increase of hydrophobicity in the +1 binding site. This arginine residue is not present in *Re*Cel3A and *Aa*Bgl1 enzymes. It thus seems as if not only the presence of the aspartate residue but also its flexibility/stability may be important for substrate and/or product interaction in this class of β-glucosidases. Also forming part of the +1 subsite and with potentially important interactions with the substrate are

the sidechains of Arg196 and Gln197, which both are conserved residues in *Re*Cel3A and *Aa*BGL1.

Previously, we have shown that *Re*Cel3A prefers cellobiose to cellotriose while *Hj*Cel3A prefers hydrolysis of slightly longer cello-oligosaccharides, such as cellotriose and cellotetraose to that of cellobiose (Karkehabadi *et al.*, 2014). In analogy with the *Re*Cel3A structure, the plane of the Trp64 side chain in *Nc*GH3-3 enzyme has turned almost 90 degrees in the structure, if compared with the corresponding tryptophan residue in the *Hj*Cel3A structure, and stacks with the sidechain of Phe301. This puts the phenylalanine residue in subsite +1 rather than in a tentative +2 subsite, as is the case in the structure of *Hj*Cel3A. Thus similar to *Re*Cel3A the existence of a +2 subsite is less pronounced in *Nc*GH3-3 than in *Hj*Cel3A and the enzyme may thus also have a substrate specificity similar to the one of *Re*Cel3A.

4.3.9 Conclusion

We have in this study determined the structure of a glycoside hydrolase family 3 β–glucosidase, GH3-3 from *Neurospora crassa,* at 2.2 Å resolution and have shown that this β–glucosidase is similar in structure to two other fungal β-glucosidases *Aspergillus aculeatus* BglI and *Rasamsonia emersonii* Cel3A, with which it shares several structural features that may be unique to this class of GH3 β-glucosidases. Among these features, and most pronounced, are the likely dimer form of the active enzyme and the large and seemingly conserved glycosylations. The structural analysis further showed that *Nc*GH3-3 should have a substrate specificity similar to the previously structurally and biochemically characterized *Re*Cel3A.

4.4 Structural and Electronic Snapshots during the Transition from a Cu(II) to Cu(I) Metal Center of a Lytic Polysaccharide Monooxygenase by X-ray Photoreduction (Paper IV)

The chitin-active LPMO from the Gram-negative chitinolytic bacterium *Serratia marcescens*, CBP21, was the first LPMO to be identified and characterised as such (Vaaje-Kolstad *et al.*, 2010; Vaaje-Kolstad *et al.*, 2005a; Vaaje-Kolstad *et al.*, 2005b). CBP21 catalysis was shown to be dependent on molecular oxygen, an external electron donor, and the presence of a metal ion cofactor (Vaaje-Kolstad *et al.*, 2010), later identified as copper (Aachmann *et al.*, 2012). Copper ions have been identified to activate AA10 (Aachmann *et al.*, 2012; Bohle *et al.*, 2011), AA9 (Phillips *et al.*, 2011; Quinlan *et al.*, 2011; Westereng *et al.*, 2011), and AA11 LPMOs (Hemsworth *et al.*, 2014). At the time of this

study, in addition to CBP21, LPMO activity has only been demonstrated for two other CBM33s so far, a cellulose-active CBM33 from *Streptomyces coelicolor* (CelS2, (Forsberg *et al.*, 2011)) and a chitin-active CBM33 from *Enterococcus faecalis* (*Efa*CBM33A, (Vaaje-Kolstad *et al.*, 2012b)), the latter of which is the subject of this study. *Efa*CBM33A is the only LPMO found in the genome of *E. faecalis* and constitutes, along with a GH family 18 chitinase (*Efa*Chi18A), the chitinolytic machinery of the bacterium. *E. faecalis* is an opportunistic pathogen and both *Efa*CBM33A and *Efa*Chi18A are virulence factors (Vebo *et al.*, 2010; Vebo *et al.*, 2009), suggesting a putative second role for these enzymes beyond biomass depolymerisation.

Until recently, there has been a dearth of structural data for metal binding in AA10 LPMOs compared to fungal AA9 LPMOs. Recently, Hemsworth *et al.* reported the structure of *Bacillus amyloliquefaciens* CBM33 (*Bam*CBM33), with unknown catalytic activity, binding Cu(I) (Hemsworth *et al.*, 2013b). It was shown that *Bam*CBM33 is stabilized in the presence of copper, and that the active site of *Bam*CBM33 with a Cu(I) ion adopts a T-shaped ligation geometry (PDB codes: 2YOX, 2YOY). A Cu(II) form of *Bam*CBM33 was not crystallized, but X-ray Absorption Near Edge Structure (XANES) and Electron paramagnetic resonance (EPR) spectroscopic methods were used to demonstrate that the enzyme was readily photo-reduced during crystallization from the Cu(II) form to a Cu(I) state (Hemsworth *et al.*, 2013b).

In this study, we investigate the active site of *Efa*CBM33 by progressively photo-reducing the catalytic copper from Cu(II) to Cu(I) in the X-ray beam at a synchrotron, using a data collection strategy minimizing the X-ray dose that is deposited in the protein crystal. During photo-reduction, we determine consecutive structural states by collecting X-ray diffraction datasets on the same crystal by controlled X-ray dose per dataset of the crystal used to collect X-ray diffraction data on. By comparing the catalytic copper bound centre of *Efa*CBM33 to known Cu(I) and Cu(II) analogues found in the Cambridge Structural Database (CSD) (Allen, 2002), we ascertain that the obtained structures of the *Efa*CBM33A unambiguously describe varying oxidation states ranging from Cu(II) to Cu(I). Lastly, we conduct quantum mechanical calculations on an active site model of the Cu(II) and Cu(I) forms of *Efa*CBM33, which suggest that the electronic structure of the active site remains quite similar as measured by atomic charges. As initial reduction of the bound copper from Cu(II)) to Cu(I) is likely a requirement for LPMO activity, these results offer a structural and electronic picture of how LPMO active sites are pre-activated for oxygen binding and subsequent catalysis.

4.4.1 Dose resolved diffraction data collection

X-ray diffraction experiments and controlled X-ray induced reduction of the bound copper from Cu(II) to Cu(I) were performed at beamline ID14-EH1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Six diffraction data sets were collected using the same protein crystal. By utilizing a rod-shaped crystal, monitoring the evolution of UV-visible absorption spectra with X-ray dose (McGeehan *et al.*, 2009) and a strategy for helical data collection (Flot *et al.*, 2010), the radiation dose was minimized and data of a minimally photo-reduced state of *Efa*CBM33 could be collected (Adam *et al.*, 2004). A helical data collection strategy consists in defining two points on the crystal along the rotation axis of the goniometer. While the crystal is rotated over a total 97° angular wedge by 1° steps, it is automatically translated along the rotation axis in-between two consecutive rotation steps, thus presenting a fresh part of the crystal to the beam for each diffraction frame. Eventually, the X-ray dose deposited in the sample will approximately be d/w smaller than that deposited with a standard data collection protocol, where d is the horizontal distance between the two points and w is the horizontal width of the beam. Two points on a limited region (320 x 53 x 40) µm, of the crystal were set up as start and end points for data collection with a 50 x 100 µm beam. A different exposure dose per image was used for some datasets, as shown in Table *4*. Collecting subsequent datasets by this method allowed for the analysis of the effects of photo-reduction on the active site copper with minimal systematic errors, as all datasets were collected from multiple and subsequent exposures of the same crystal volume.

4.4.2 Overall structure of *Efa*CBM33A in complex with copper

*Efa*CBM33A with a bound copper atom was crystallized in space group $P2_12_12_1$ with cell dimensions of 43.4 x 48.6 x 68.5 Å, one protein molecule per asymmetric unit, and a calculated V_m (Matthews coefficient) (Matthews, 1968) of 1.97 $Å^3$/Da including all Cα atoms in the structure. We present six structures of *Efa*CBM33A along the process of X-ray induced photoreduction, all refined at 1.5 Å and final R and R_{free} values of 15.6-16.1% and 18.3-19.6%, respectively for the six structures. Data collection and refinement statistics are summarized in Table *4*. In all the structure models, there is clear electron density for all the 169 amino acid residues of the enzyme, approximately 285 water molecules, and one copper atom bound to the protein. Negligible pairwise RMSD values of 0.03-0.04 Å over all protein atoms show that the structures are essentially identical. The primary differences between the six structures are found in the coordination geometry of the copper ion as a function of the X-ray dose. With the exception of the active site, the structure of the copper-bound

*Efa*CBM33A herein is very similar to the previously published apo form without copper (PDB ID: 4A02 (Vaaje-Kolstad *et al.*, 2012a); 0.54 Å RMSD on Cα atoms).

The overall structure and the active site of *Efa*CBM33A, as observed in the structure determined from the dataset obtained after the lowest radiation dose (4ALC), exhibits a trigonal bipyramidal (*tbp*) structure coordinated by two conserved histidine residues (previously named "the histidine brace") and two water molecules (*Figure 26a-c*). In this configuration, the *N*-terminal histidine (His29) forms a bidentate coordination to the copper ion wherein the backbone *N* atom occupies one of the three equatorial coordination positions and the sidechain *N*δ atom occupies one axial position. The Nε atom in His114 occupies the other axial position. The remaining equatorial positions in the Cu(II) structure are occupied by two water molecules. Three additional residues conserved in AA10 LPMOs are shown in *Figure 26c*, Glu64, Ala112, and Phe185. Ala112 is not conserved in AA9 LPMOs, and is thought to play a role in the potential mechanistic differences between AA9 and AA10 LPMOs (Hemsworth *et al.*, 2013a; Hemsworth *et al.*, 2013b). Phe185 is located in a similar position to a conserved tyrosine in AA9 LPMOs, which in the latter case therein imparts an octahedral coordination state around the Cu(II) ion (Wu *et al.*, 2013a; Li *et al.*, 2012; Quinlan *et al.*, 2011; Harris *et al.*, 2010; Karkehabadi *et al.*, 2008b). In AA9 LPMOs, the Glu64 residue is replaced by a conserved glutamine residue. The electron density map for the *Efa*CBM33A structure 4ALC has two spherical electron densities (1.93 and 1.91 e/Å^3, respectively) at 2.21 and 2.19 Å from the copper ion which were modelled and refined as water molecules and are shown as red spheres in *Figure 26c*. For the water molecules bound to copper, there are no other stabilizing interactions with the enzyme. Thus, the positions of the water molecules are primarily dictated by coordination to the copper ion.

The only other known AA10 structure with a copper ion bound reported to date is *Bam*CBM33 from Hemsworth *et al*. (Hemsworth *et al.*, 2013b), wherein all copper ions were photo-reduced to a Cu(I) oxidation state. The *Bam*CBM33 enzyme active site is illustrated in *Figure 26d*. The coordination geometry therein is in a T-shaped (*Tsh*) geometry with no water molecules coordinated to the copper ion. The corresponding protein-copper interactions retain the structure of the histidine brace. The difference in observed geometry between the 4ALC structure and the *Bam*CBM33 structure indicates a difference in copper oxidation state, as described in detail further below.

Table 4. *X-ray data collection and processing, structure refinement, and final model statistics*

PDB code	*4ALC*	*4ALE*	*4ALR*	*4ALQ*	*4ALS*	*4ALT*
Data quality						
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	a = 43.42	a= 43.42	a= 43.42	a= 43.46	a= 43.44	a= 43.40
	b= 48.56	b= 48.57	b= 48.58	b= 48.62	b= 48.61	b= 48.57
	c= 68.45	c= 68.46	c= 68.46	c= 68.51	c= 68.50	c= 68.43
Data-collections statistics						
Wavelength (Å)	0.9334	0.9334	0.9334	0.9334	0.9334	0.9334
Resolution range (Å)	48.56 – 1.49	48.57 – 1.48	39.68 – 1.49	48.68 – 1.48	48.61 – 1.47	39.60 – 1.49
No of unique reflections	23919 (3317)	24395 (3216)	23936 (3324)	24358 (3235)	24138 (3119)	23899 (3307)
Multiplicity	3.4 (3.0)	3.4 (2.9)	3.4 (3.0)	3.3 (2.9)	2.9 (2.3)	3.4 (3.1)
Completeness	98.6 (96.1)	97.8 (90.5)	98.6 (96.1)	97.8 (91.2)	96.6 (87.9)	98.6 (95.9)
R_{merge} [‡]	0.094 (0.477)[†]	0.096 (0.503)	0.094 (0.484)	0.067 (0.237)	0.056 (0.142)	0.100 (0.564)
Exposure time/frame (s)	1	1	1	3	6	1
Total exposure time (s)	97	194	485	776	1358	1455
Dose (Gy)	8.27×10^4	1.65×10^5	2.47×10^5	4.94×10^5	9.86×10^5	1.07×10^6
Refinement						
Resolution range (Å)	39.60 – 1.49	39.61 - 1.48	39.62 - 1.49	39.65 – 1.48	39.64 – 1.47	39.60 – 1.49
Observations	80452 (10080)	81915 (9417)	80300 (10018)	81296 (9301)	69202 (7187)	80875 (10226)
No. of reflections used (working set)	22667	23130	22679	23087	22872	22647
R_{work} (%)	16.1	16.1	16.1	15.6	15.6	16.1
R_{free} (%)	18.8	18.4	18.8	18.3	19.6	18.8
No. of residues	167	167	167	167	167	167
No. of water molecules	289	286	282	281	292	280
No. of copper atoms	1	1	1	1	1	1
RMSD bond lengths (Å)	0.008	0.008	0.008	0.008	0.008	0.008
RMSD angles (°)	1.230	1.229	1.230	1.210	1.221	1.242
Ramachandran [§]						
In Favoured Regions (%)	100	100	100	100	100	100
Outliers (%)	0	0	0	0	0	0

[†] Values for the highest resolution shell are given in parentheses

[‡] $R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ where $I(hkl)$ is the intensity of reflection hkl, Σ_{hkl} is the sum over all reflections and Σ_i is the sum over i measurements of reflection hkl.

[§] Calculated using a strict-boundary Ramachandran definition given by Kleywegt and Jones (Kleywegt & Jones, 1996)

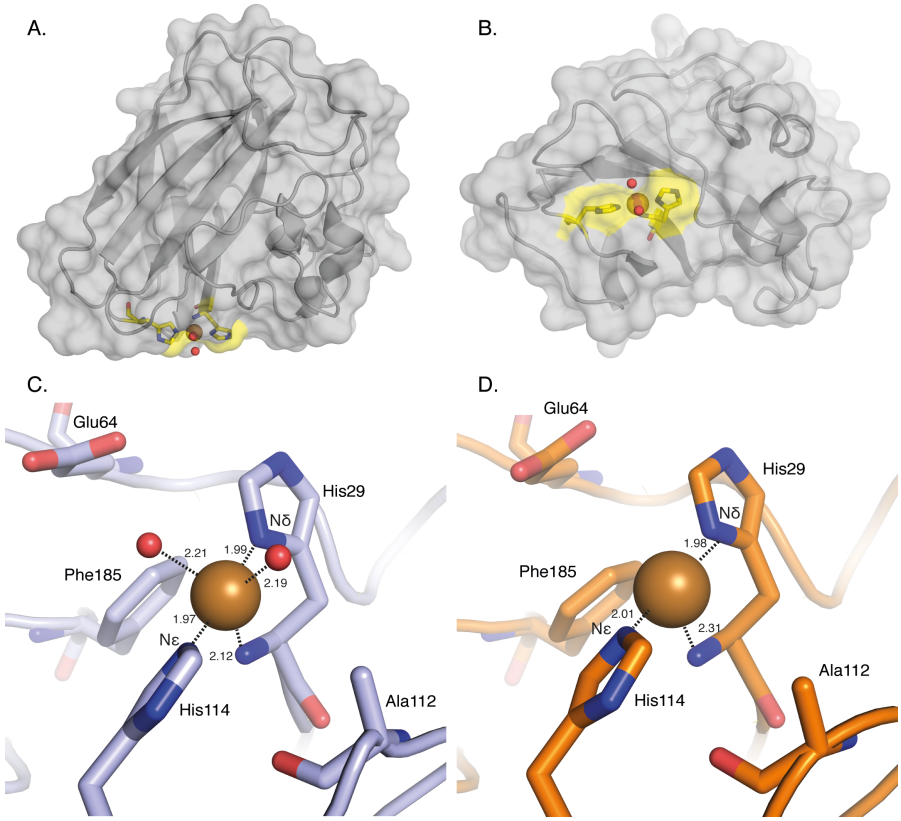


Figure 26. Overall structure of *Enterococcus faecalis* CBM33. (A) Side and (B) bottom view of the crystal structure of Cu(II)-bound *Efa*CBM33A with a cartoon and transparent surface model in grey (PDB code: 4ALC; the structure with the lowest radiation dose). The active center is highlighted in yellow and the two residues making up the histidine brace are shown in stick format. The copper atom is shown as a brown sphere and two water molecules coordinated to the copper are shown as red spheres. (C) The active site in the oxidized [Cu(II)] form of *Efa*CBM33 (PDB code: 4ALC). Note that there are not other stabilizing interactions between the two coordinating water molecules and the enzyme. (D) The active site in *Bam*CBM33 binding Cu(I) (PDB code: 2YOX) (Hemsworth *et al.*, 2013b). Distances to the copper ion are provided in Å

4.4.3 Structural changes induced by X-ray photo-reduction

The structural changes caused by increased X-ray dosage during photo-reduction were limited to the local environment of the copper ion (*Figure 27*). Omit-map analysis of the copper-coordinated water molecules shows a continuous decay of electron density correlated with X-ray exposure, and at approximately 1 MGy accumulated radiation dose (4ALT) both coordinated water molecules are completely lost. The electron density for the water molecule closest to Ala112 is retained slightly longer than the other. The decay of the

electron density for the two water molecules coordinated to the copper represents a change in the fraction of Cu(II) to Cu(I) populations between the six structures, as a result of the accumulated radiation dose of the exposed crystal region used for data collection. The structures obtained at gradually higher doses of X-ray radiation reveal a continuous shift in the copper coordination configuration from *tbp* coordination to *Tsh* geometry in the structures that lack the copper-bound water molecules (*Figure 27*).

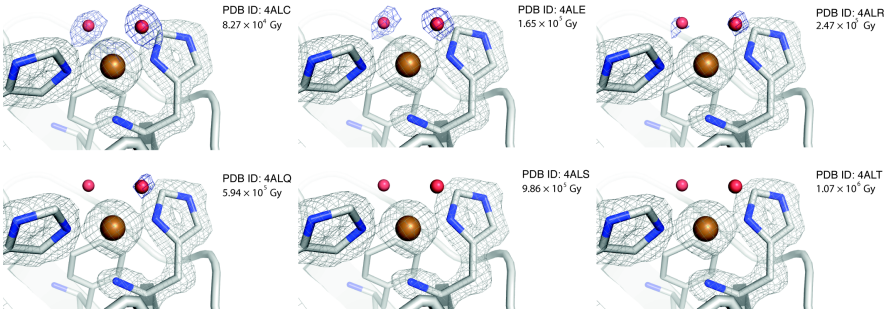


Figure 27. Close up view of the catalytic centers of *Enterococcus faecalis* CBM33A at different levels of X-ray exposure. The blue omit-calculated $2mF_o$-F_c maps (Terwilliger *et al.*, 2008) are contoured at 0.87 e/Å^3, the grey $2F_o$-F_c maps of are contoured at 0.89 e/Å^3. The red oxygen-atom coordinates are taken from the 4ALC structure, and are placed throughout the series for reference and were not used in the calculation of the $2mF_o$-F_c omit maps. The structural images are labeled by PDB ID and the dose of X-ray exposure.

4.4.4 LPMO copper oxidation state determination by analogy to small-molecule copper complexes

The most obvious structural change upon X-ray photo reduction of the copper centre is that the two water molecules coordinating to the copper ion gradually disappear, as shown in *Figure 27*. This demonstrates that the coordination number (CN) for the copper ion drops from five to three; the conformation of the copper site changes from a five-coordinated *tbp* structure to a three-coordinated *Tsh* geometry (*Figure 28*). Gradual disappearance of electron density upon increasing the X-ray dosage was not observed for any other water molecule in the structure, suggesting that the effects seen for the copper-bound waters relates to a change in the copper ion. Among more than 40,000 copper structures in the CSD (Allen, 2002), nearly half fit our initial search criteria ($1 \leq CN \leq 8$; only nitrogen and/or oxygen as coordinating atoms), including 9,727 five-coordinate and 564 three-coordinate structures (data not shown). Limiting the search to only include those with histidine-like coordination surroundings and excluding strained structures left 10 five-coordinate structures, all Cu(II) and 24 three-coordinate structures, all Cu(I).

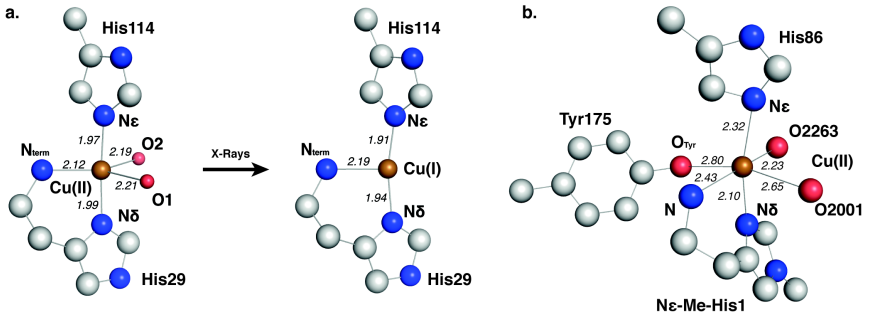


Figure 28. Copper coordination in *Enterococcus faecalis* CBM33A and *Thermoauscus aurantiacus* GH61A. Important residues, atoms, and coordination distances are indicated where appropriate. **(a)** The copper binding site of *Efa*CBM33A (PDB ID: 4ALC) displays a trigonal bipyramidal (*tbp*) coordination of copper and after X-ray exposure adopts a T-shaped (*Tsh*) configuration (PDB ID: 4ALT). **(b)** An octahedral Cu(II) coordination in the GH61A from *T. aurantiacus* (PDB ID: 2YET). In most AA10 LPMOs, including *Efa*CBM33A, the tyrosine residue labeled Tyr175 is replaced by phenylalanine.

The *Efa*CBM33 structures have two axial copper-nitrogen (Cu-N_{ax}, Cu-$N_{ax'}$) bond distances, which both decrease by 0.05 Å going from *tbp* to *Tsh*, while the equatorial copper-nitrogen (Cu-N_{eq}) bond distance becomes 0.075 Å longer. Additionally, the near-linear N_{ax}-Cu-$N_{ax'}$ angle in *tbp*, 176.2°, bends a bit off-axis in *Tsh*, 167.5°, while the N_{eq}-Cu-N_{ax} and N_{eq}-Cu-$N_{ax'}$ angles increase by 5.6° and 3.1°, respectively. The five-coordinated form is almost identical to the Cu(II) structure reported by Casella *et al*. ((Casella *et al.*, 1996); CSD code: ZUBHOT), whereas the three-coordinated counterpart most closely mimics the *Tsh* Cu(I) structure reported by Sorrell *et al*. ((Sorrell *et al.*, 1994); CSD code: PIVNOX). A CSD search for five-coordinated, dihydrate copper structures also coordinated by three nitrogen atoms, returned 7 structures – all Cu(II) – with several *tbp* examples similar to the hydrated version of *Efa*CBM33A seen in 4ALC. Taken together, these observations show that the structural changes observed upon irradiation of *Efa*CBM33A reflect photo-reduction of Cu(II) to Cu(I).

4.4.5 Quantum mechanical calculations of the LPMO active site

The structures presented above enable density functional theory (DFT) calculations to quantify how the electronic structure of the active site changes upon reduction. The active sites of 4ALC, the Cu(II) structure, and 4ALT, the Cu(I) structure, were both examined with the M06-L functional and the 6-31G(d) basis set, by employing an active site model (ASM) representation of the system. Quantum mechanical geometry optimizations were conducted with a range of ASMs. The model consisting of the residues His29, Glu64, Ala112,

His114, Trp176, Ile178, Phe185 was found to yield the smallest RMSD values for a size that was still computationally tractable with a full quantum mechanical treatment of the ASM in both structure. *Figure 29* shows comparisons between the crystal structures and the quantum mechanically optimized ASMs. All computed distances between coordinating nitrogen atoms and the copper differ from the crystallographically observed distances by less than 0.07 Å, which is well within the resolution of the structure.

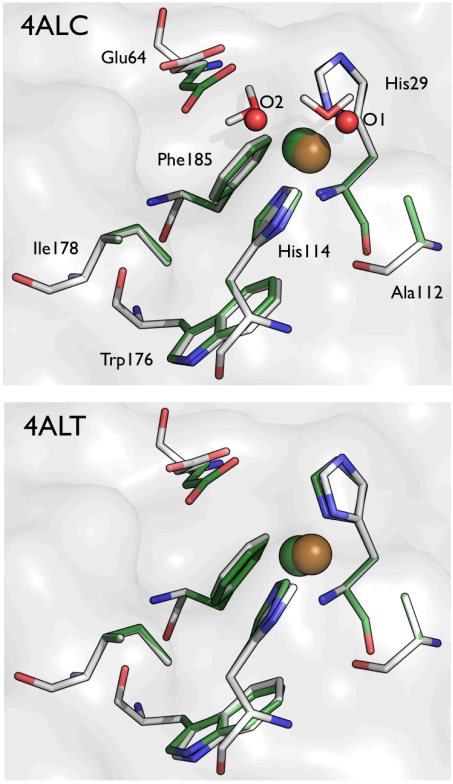


Figure 29. Comparison of the crystallographically determined active sites of 4ALC (top) and 4ALT (bottom) with quantum mechanically optimized active site models. The residues in gray (carbon), blue (nitrogen), and red (oxygen) represent the crystal structures, and the residues shown in green represent the geometry optimized structures from the DFT calculations. The copper is colored gold and green, respectively. The water molecules from the crystal structure are shown as red spheres in 4ALC and the optimized water molecules are shown in stick format.

Subsequent to the geometry optimizations, Natural Population Analysis was conducted to examine the charge distributions for both states. The copper ion charges in the oxidized and reduced states of the active site are +1.48 and +0.99, respectively. These values agree well with the formal oxidation states of

Cu(II) and Cu(I), and also agree remarkably well with the charges found in both the formal Cu(II) and Cu(I) oxidation states of +1.48 and +0.92, respectively, in an AA9 LPMO with a similar ASM approach (Kim *et al.*, 2014). Interestingly, the charge distribution of the coordinating histidine residues does not show a significant change, despite the substantial change in the copper ion oxidation state. This result suggests that the LPMO active site is able to readily accommodate both oxidation states of copper with little overall change in the charge distribution in the enzyme.

4.4.6 Conclusions

The study presented in paper IV presents the second structure of a CBM33 with copper bound and the first structure of a CBM33 with a Cu(II) ion. Using a smart data collection strategy allowing for the structure determination of LPMO structures in both copper oxidation states, we were able to visualize structural and copper-coordination changes associated with reduction of the bound copper from Cu(II) to Cu(I). This experimental methodology is quite generalizable, and can be used to capture the electronic and structural transitions in metalloenzyme reduction at advanced light sources available at some modern state of the art synchrotron beam lines.

X-ray photo-reduction causes clear changes in the active site of *Efa*CBM33, namely the loss of the coordinating water molecules. By correlating the structural data with the CSD, the two forms of *Efa*CBM33A were assigned as a Cu(II) and Cu(I) state with a trigonal bipyramidal and T-shaped geometry, respectively. DFT calculations reveal only minor changes in the atomic charges required for binding to either oxidation state of the copper ion, similar to what was found in a theoretical study for an AA9 LPMO (Kim *et al.*, 2014). The study presented in paper IV provides the first experimental dataset to provide insight in the reductive step that activates the catalytic centre on a LPMO for catalysis.

5 Concluding remarks

5.1.1 Structure and function of fungal GH3 β-glucosidases

Three fungal GH3 β-glucosidases, from *Hypocrea jecorina, Rasamsonia emersonii* and *Neurospora* crassa, were produced. The three-dimensional crystal structure for each of the three enzymes was elucidated. *Hypocrea jecorina* Cel3A and *Rasamsonia emersonii* Cel3A were characterized also biochemically in terms of thermal stability and enzymatic activity towards cellooligosaccharides and synthetic chromogenic substrates. The structure of *Hj*Cel3A revealed that this enzyme is a three-domain protein with a wider active site cleft than homologous enzymes with known structures. Furthermore, *Hj*Cel3A is a rather promiscuous enzyme in respect to the type of β-glycosidic-linkage and substrate length, and it is more efficient on longer oligosaccharides than several other biochemically characterized GH3 β-glucosidases. Supplementing *H. jecorina* whole cellulase with additional amounts of *Hj*Cel3A benefits cellulose conversion most likely primarily by the conversion of cellobiose to glucose.

The β-glucosidases Cel3A from the thermophilic fungi *R. emersonii* was found to increase the saccharification of lignocellulosic biomass to a greater extent than *Hj*Cel3A. The enzyme was also found to be more thermostable than *Hj*Cel3A, which is beneficial for industrial processes that often are performed at highest possible temperatures to increase the conversion rate constancies of these. *Re*Cel3A appears to have a preference for disaccharides rather than longer oligosaccharides. The three-dimensional structure of *Re*Cel3A revealed that this enzyme is a three-domain glycoprotein that exists as a functional dimer in solution. The major differences to homologous GH3 enzymes are a number of loops that deepens the active site cleft of the enzyme as well as a curiously positioned C-terminal loops that covers the *N*-terminal domain and buries extensive *N*-glycan chains.

GH3-3 from *N. crassa* was structurally determined and found to be a dimeric three-domain glycoprotein. *Nc*GH3-3 is structurally highly homologous to

*Re*Cel3A, which allowed us to compare structural features in detail of what is likely a sub-class of fungal GH3 β-glucosidases that mainly hydrolyse disaccharides such as cellobiose.

5.1.2 Single crystal X-ray induced photoreduction snapshots of an AA10 LPMO

Using a carefully designed helical data collection strategy we were able to collect six diffraction datasets of the bacterial AA10 LPMO *Ef*LPMO10A at increasing X-ray radiation doses for the consecutive X-ray datasets. The three dimensional crystal structures based on these six X-ray datasets revealed that the catalytic copper of *Ef*LPMO10A transitions from an oxidized Cu(II) with a octahedral ligation configuration to a reduced Cu(I) with a T-shaped ligation configuration. This structural rearrangement was confirmed via DFT computational techniques.

5.1.3 Future perspectives

In paper I-III we investigated a series of fungal β-glucosidases that are, or can potentially be, key enzymes for industrial degradation of lignocellulosic biomass into fermentable glucans such as glucose. The key feature of these β-glucosidases, used or to be used in biotechnological applications, are their ability to readily and specifically drive the process of depolymerisation of cellulose forward by hydrolysing cellobiose to monomeric glucose. Since the work presented in paper I-III was done we have established a fungal transformation and expression system in our lab. This will allow us to more readily perform point mutations and alterations of these large protein molecules, which will allow for further investigations of the nature of more distal substrate binding sites (+1 and +2) and possible improvements of catalytic parameters such as substrate inhibition.

The continuation of the work with LPMOs will proceed in multiple directions, as the field is still new and rapidly expanding. Questions remain though about the finer details of the catalytic mechanism of these oxidative enzymes, as well as the exact relationship of the fungal LPMOs with CDH. Another interesting fact is that many organisms have a very large set of different LPMO genes and novel substrate preferences are discovered continuously (Agger *et al.*, 2014; Vu *et al.*, 2014). It is not clear that this class of enzymes are restricted to biomass degradation, other roles for enzymes from this class of enzymes will most likely be discovered. With this in mind, it would be interesting to study on what kind of substrates this class of enzymes are active.

References

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

Adam, V., Royant, A., Nivière, V., Molina-Heredia, F.P. & Bourgeois, D. (2004). Structure of superoxide reductase bound to ferrocyanide and active site expansion upon X-ray-induced photo-reduction. *Structure,* 12(9), pp. 1729-1740.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66(Pt 2), pp. 213-21.

Agger, J.W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W.G., Ludwig, R., Horn, S.J., Eijsink, V.G. & Westereng, B. (2014). Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci. U S A,* 111(17), pp. 6287-92.

Albers, S.V. & Meyer, B.H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology,* 9(6), pp. 414-426.

Allen, F.H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta crystallographica. Section B, Structural science,* 58(Pt 3 Pt 1), pp. 380-8.

Aro, N., Pakula, T. & Penttilä, M. (2005). Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiol Rev,* 29(4), pp. 719-39.

Bacik, J.P., Whitworth, G.E., Stubbs, K.A., Vocadlo, D.J. & Mark, B.L. (2012). Active Site Plasticity within the Glycoside Hydrolase NagZ Underlies a Dynamic Mechanism of Substrate Distortion. *Chemistry & Biology,* 19(11), pp. 1471-1482.

Banerjee, G., Car, S., Scott-Craig, J.S., Borrusch, M.S., Aslam, N. & Walton, J.D. (2010). Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. *Biotechnol Bioeng,* 106(5), pp. 707-20.

Barnett, C.C., Berka, R.M. & Fowler, T. (1991). Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Biotechnology (N Y),* 9(6), pp. 562-7.

Barr, B.K., Hsieh, Y.L., Ganem, B. & Wilson, D.B. (1996). Identification of two functionally different classes of exocellulases. *Biochemistry,* 35(2), pp. 586-592.

Barral, P., Suarez, C., Batanero, E., Alfonso, C., Alché, J.D., Rodríguez-Garcia, M.I., Villalba, M., Rivas, G. & Rodriguez, R. (2005). An olive pollen protein with allergenic activity, Ole e 10, defines a novel family of carbohydrate-binding modules and is potentially implicated in pollen germination. *Biochemical Journal,* 390, pp. 77-84.

Beckham, G.T. & Crowley, M.F. (2011). Examination of the α-Chitin Structure and Decrystallization Thermodynamics at the Nanoscale. *J. Phys. Chem. B,* 115(15), pp. 4516-4522.

Beckham, G.T., Matthews, J.F., Peters, B., Bomble, Y.J., Himmel, M.E. & Crowley, M.F. (2011). Molecular-Level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs. *J. Phys. Chem. B,* 115(14), pp. 4118-4127.

Béguin, P. & Aubert, J.P. (1994). The biological degradation of cellulose. *FEMS Microbiol Rev,* 13(1), pp. 25-58.

Berglund, G.I., Carlsson, G.H., Smith, A.T., Szoke, H., Henriksen, A. & Hajdu, J. (2002). The catalytic pathway of horseradish peroxidase at high resolution. *Nature,* 417(6887), pp. 463-8.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic acids research,* 28(1), pp. 235-42.

Bhuiyan, N.H., Selvaraj, G., Wei, Y. & King, J. (2009). Role of lignification in plant defense. *Plant Signal Behav,* 4(2), pp. 158-9.

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnol Adv,* 30(6), pp. 1575-88.

Blake, C.C.F. & Phillips, D.C. (1962). Biological Effects of Ionizing Radiations at the Molecular Level. *International Atomic Energy Agency, Vienna*, pp. 183–191.

Boerjan, W., Ralph, J. & Baucher, M. (2003). Lignin biosynthesis. *Annu Rev Plant Biol,* 54, pp. 519-46.

Bohle, L.A., Mathiesen, G., Vaaje-Kolstad, G. & Eijsink, V.G.H. (2011). An endo-β-*N*-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type *N*-linked glycans. *FEMS Microbiol. Lett.,* 325(2), pp. 123-129.

Bohlin, C., Olsen, S.N., Morant, M.D., Patkar, S., Borch, K. & Westh, P. (2010). A comparative study of activity and apparent inhibition of fungal β-glucosidases. *Biotechnol Bioeng,* 107(6), pp. 943-52.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal,* 382(Pt 3), pp. 769-81.

Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of Molecular Biology,* 242(4), pp. 309-20.

Bragg, W.L. (1913). The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society,* 17, pp. 43–57.

Brockhausen, I., Schachter, H. & Stanley, P. (2009). O-GalNAc Glycans. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

Callebaut, I., Tasso, A., Brasseur, R., Burny, A., Portetelle, D. & Mornon, J.P. (1994). Common prevalence of alanine and glycine in mobile reactive centre loops of serpins and viral fusion peptides: do prions possess a fusion peptide? *J Comput Aided Mol Des,* 8(2), pp. 175-91.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research,* 37, pp. D233-D238.

Cantarel, B.L., Lombard, V. & Henrissat, B. (2012). Complex carbohydrate utilization by the healthy human microbiome. *PloS One,* 7(6), p. e28742.

Casella, L., Carugo, O., Gullotti, M., Doldi, S. & Frassoni, M. (1996). Synthesis, structure, and reactivity of model complexes of copper nitrite reductase. *Inorganic Chemistry,* 35(5), pp. 1101-1113.

Chance, B., Angiolillo, P., Yang, E.K. & Powers, L. (1980). Identification and assay of synchrotron radiation-induced alterations on metalloenzymes and proteins. *FEBS Lett.,* 112, pp. 178-182.

Chance, B., Powers, L., Ching, Y., Poulos, T., Schonbaum, G.R., Yamazaki, I. & Paul, K.G. (1984). X-ray absorption studies of intermediates in peroxidase activity. *Arch Biochem Biophys,* 235(2), pp. 596-611.

Cheng, Q., Li, H., Merdek, K. & Park, J.T. (2000). Molecular characterization of the β-*N*-acetylglucosaminidase of *Escherichia coli* and its role in cell wall recycling. *J Bacteriol,* 182(17), pp. 4836-40.

Cherry, J.R. & Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol,* 14(4), pp. 438-43.

Chirico, W.J. & Brown, R.D., Jr. (1987a). Purification and characterization of a β-glucosidase from *Trichoderma reesei*. *European Journal of Biochemistry,* 165(2), pp. 333-41.

Chirico, W.J. & Brown, R.D., Jr. (1987b). *β*-glucosidase from *Trichoderma reesei*. Substrate-binding region and mode of action on [1-3H]cello-oligosaccharides. *European Journal of Biochemistry,* 165(2), pp. 343-51.

Cho, H.M., Gross, A.S. & Chu, J.W. (2011). Dissecting Force Interactions in Cellulose Deconstruction Reveals the Required Solvent Versatility for Overcoming Biomass Recalcitrance. *Journal of the American Chemical Society,* 133(35), pp. 14033-14041.

Cournoyer, B. & Faure, D. (2003). Radiation and functional specialization of the family-3 glycoside hydrolases. *Journal of Molecular Microbiology and Biotechnology,* 5(3), pp. 190-198.

Cummings, R.D. & Doering, T.L. (2009). Fungi. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

Davies, G.J., Wilson, K.S. & Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal,* 321 (Pt 2), pp. 557-9.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T. & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science,* 265(5171), pp. 524-8.

Dodd, D. & Cann, I.K.O. (2009). Enzymatic deconstruction of xylan for biofuel production. *Global Change Biology Bioenergy,* 1(1), pp. 2-17.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66(Pt 4), pp. 486-501.

Esko, J.D., Doering, T.L. & Raetz, C.R.H. (2009). Eubacteria and Archaea. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

Etzler, M.E. & Mohnen, D. (2009). Viridiplantae. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

FAO. (2011). *The state of the world's land and water resources for food and agriculture : managing systems at risk.* Abingdon: Earthscan.

Faure, D., Desair, J., Keijers, V., Ali Bekri, M., Proost, P., Henrissat, B. & Vanderleyden, J. (1999). Growth of *Azospirillum irakense* KBC1 on the Aryl β-glucoside salicin requires either *salA* or *salB*. *J Bacteriol,* 181(10), pp. 3003-3009.

Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation,* 17(1), pp. 107-18.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martinez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., Aerts, A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P.M., de Vries, R.P., Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Gorecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J.A., Kallen, N., Kersten, P., Kohler, A., Kues, U., Kumar, T.K.A., Kuo, A., LaButti, K., Larrondo, L.F., Lindquist, E., Ling, A., Lombard, V., Lucas, S., Lundell, T., Martin, R., McLaughlin, D.J., Morgenstern, I., Morin, E., Murat, C., Nagy, L.G., Nolan, M., Ohm, R.A., Patyshakuliyeva, A., Rokas, A., Ruiz-Duenas, F.J., Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J.C., John, F.S., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A.,

Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D.C., Martin, F., Cullen, D., Grigoriev, I.V. & Hibbett, D.S. (2012). The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science,* 336(6089), pp. 1715-1719.

Foreman, P., Goedegebuur, F., Van Solingen, P. & Ward, M. (2005). *Novel Trichoderma genes*. PCT/US2004/016881.

Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J. & Ward, M. (2003). Transcriptional Regulation of Biomass-degrading Enzymes in the Filamentous Fungus *Trichoderma reesei*. *The Journal of biological chemistry,* 278(34), pp. 31988-31997.

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

Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins-Structure Function and Genetics,* 23(4), pp. 566-579.

Fujdala, M.K. & Larenas, E.A. (2010). *β-Glucosidase Enhanced Filamentous Fungal Whole Cellulase Compositions and Methods Of Use*.

Gritzali M., J. & D., B.R. (1979). The cellulase system of *Trichoderma*. The relationship between purified extracellular enzymes from induced or cellulose grown cells. *Adv Chem Ser* 181, pp. 237-260.

Gruno, M., Väljamäe, P., Pettersson, G. & Johansson, G. (2004). Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol Bioeng,* 86(5), pp. 503-11.

Gudmundsson, M., Kim, S., Wu, M., Ishida, T., Momeni, M.H., Vaaje-Kolstad, G., Lundberg, D., Royant, A., Ståhlberg, J., Eijsink, V.G., Beckham, G.T. & Sandgren, M. (2014). Structural and Electronic Snapshots during the Transition from a Cu(II) to Cu(I) Metal Center of a Lytic Polysaccharide Monooxygenase by X-ray Photoreduction. *The Journal of biological chemistry,* 289(27), pp. 18782-92.

Harris, P.V., Welner, D., McFarland, K.C., Re, E., Poulsen, J.C.N., Brown, K., Salbo, R., Ding, H.S., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. & Lo Leggio, L. (2010). Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase Family 61: Structure and function of a large, enigmatic family. *Biochemistry,* 49(15), pp. 3305-3316.

Harvey, A.J., Hrmova, M., De Gori, R., Varghese, J.N. & Fincher, G.B. (2000). Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins-Structure Function and Genetics,* 41(2), pp. 257-269.

Helenius, A. & Aebi, M. (2001). Intracellular functions of *N*-linked glycans. *Science,* 291(5512), pp. 2364-9.

Hemsworth, G.R., Davies, G.J. & Walton, P.H. (2013a). Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Current Opinion in Structural Biology,* 23(5), pp. 660-668.

Hemsworth, G.R., Henrissat, B., Davies, G.J. & Walton, P.H. (2014). Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nat Chem Biol,* 10(2), pp. 122-126.

Hemsworth, G.R., Taylor, E.J., Kim, R.Q., Gregory, R.C., Lewis, S.J., Turkenburg, J.P., Parkin, A., Davies, G.J. & Walton, P.H. (2013b). The copper active site of CBM33 polysaccharide oxygenases. *J. Am. Chem. Soc.,* 135(16), pp. 6069-6077.

Henderson, R. (1990). Cryoprotection of Protein Crystals against Radiation-Damage in Electron and X-Ray-Diffraction. *Proceedings of the Royal Society B-Biological Sciences,* 241(1300), pp. 6-8.

Henriksson, G., Ander, P., Pettersson, B. & Pettersson, G. (1995). Cellobiose Dehydrogenase (Cellobiose Oxidase) from *Phanerochaete chrysosporium* as a Wood Degrading Enzyme - Studies on Cellulose, Xylan and Synthetic Lignin. *Applied Microbiology and Biotechnology,* 42(5), pp. 790-796.

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

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal,* 280 (Pt 2), pp. 309-16.

Hill, J.F. (2012). *Chemical research on plant growth : a translation of Theodore de Saussure's recherches chimiques sur la vegetation*. 1st. ed. New York: Springer.

Hong, J., Tamaki, H. & Kumagai, H. (2006). Unusual hydrophobic linker region of β-glucosidase (BGLII) from *Thermoascus aurantiacus* is required for hyper-activation by organic solvents. *Applied Microbiology and Biotechnology,* 73(1), pp. 80-88.

Hong, J., Tamaki, H. & Kumagai, H. (2007). Cloning and functional expression of thermostable β-glucosidase gene from *Thermoascus aurantiacus*. *Appl Microbiol Biotechnol,* 73(6), pp. 1331-9.

Horn, S.J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V.G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels,* 5(1), p. 45.

Hrmova, M., De Gori, R., Smith, B.J., Fairweather, J.K., Driguez, H., Varghese, J.N. & Fincher, G.B. (2002). Structural Basis for Broad Substrate Specificity in Higher Plant β-D-Glucan Glucohydrolases. *Plant Cell,* 14(5), pp. 1033-52.

Hrmova, M., De Gori, R., Smith, B.J., Vasella, A., Varghese, J.N. & Fincher, G.B. (2004). Three-dimensional Structure of the Barley β-D-Glucan Glucohydrolase in Complex with a Transition State Mimic. *The Journal of biological chemistry,* 279(6), pp. 4970-80.

Hrmova, M. & Fincher, G.B. (1997). Barley β-D-glucan exohydrolases. Substrate specificity and kinetic properties. *Carbohydrate Research,* 305(2), pp. 209-221.

Hrmova, M., Garrett, T.P. & Fincher, G.B. (1995). Subsite affinities and disposition of catalytic amino acids in the substrate-binding region of barley 1,3-β-glucanases. Implications in plant-pathogen interactions. *The Journal of biological chemistry,* 270(24), pp. 14556-63.

Hrmova, M., MacGregor, E.A., Biely, P., Stewart, R.J. & Fincher, G.B. (1998). Substrate Binding and Catalytic Mechanism of a Barley β-D-Glucosidase/(1,4)-β-D-Glucan Exohydrolase. *The Journal of biological chemistry,* 273(18), pp. 11134-43.

Hrmova, M., Streltsov, V.A., Smith, B.J., Vasella, A., Varghese, J.N. & Fincher, G.B. (2005). Structural Rationale for Low-Nanomolar Binding of Transition State Mimics to a Family GH3 β-D-Glucan Glucohydrolase from Barley. *Biochemistry,* 44(50), pp. 16529-39.

Hrmova, M., Varghese, J.N., De Gori, R., Smith, B.J., Driguez, H. & Fincher, G.B. (2001). Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant β-D-glucan glucohydrolase. *Structure,* 9(11), pp. 1005-16.

Ioannou, Y.A., Zeidner, K.M., Grace, M.E. & Desnick, R.J. (1998). Human α-galactosidase A: glycosylation site 3 is essential for enzyme solubility. *Biochemical Journal,* 332, pp. 789-797.

Isaksen, T., Westereng, B., Aachmann, F.L., Agger, J.W., Kracher, D., Kittl, R., Ludwig, R., Haltrich, D., Eijsink, V.G.H. & Horn, S.J. (2014). A C4-oxidizing Lytic Polysaccharide Monooxygenase Cleaving Both Cellulose and Cello-oligosaccharides. *J. Biol. Chem.,* 289(5), pp. 2632-2642.

Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C. & Sandgren, M. (2008a). The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from *Hypocrea jecorina*, at 1.6 Å Resolution. *Journal of Molecular Biology,* 383(1), pp. 144-154.

Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C. & Sandgren, M. (2008b). The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from *Hypocrea jecorina*, at 1.6 angstrom Resolution. *Journal of Molecular Biology,* 383(1), pp. 144-154.

Karkehabadi, S., Helmich, K.E., Kaper, T., Hansson, H., Mikkelsen, N.E., Gudmundsson, M., Piens, K., Fujdala, M., Banerjee, G., Scott-Craig, J.S., Walton, J.D., Phillips, G.N., Jr. & Sandgren, M. (2014). Biochemical Characterization and Crystal Structures of a Fungal Family 3 β-Glucosidase, Cel3A from *Hypocrea jecorina*. *The Journal of biological chemistry,* 289(45), pp. 31624-31637.

Kayser, V., Chennamsetty, N., Voynov, V., Forrer, K., Helk, B. & Trout, B.L. (2011). Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnology Journal,* 6(1), pp. 38-44.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. & Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature,* 181, pp. 662–666

Kim, S., Ståhlberg, J., Sandgren, M., Paton, R.S. & Beckham, G.T. (2014). Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci. U S A,* 111(1), pp. 149-154.

Kirk, T.K. & Farrell, R.L. (1987). Enzymatic Combustion - the Microbial-Degradation of Lignin. *Annual Review of Microbiology,* 41, pp. 465-505.

Klemm, D., Heublein, B., Fink, H.P. & Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angew Chem Int Ed Engl,* 44(22), pp. 3358-93.

Kleywegt, G.J. & Jones, T.A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure,* 4(12), pp. 1395-1400.

Korotkova, O.G., Semenova, M.V., Morozova, V.V., Zorov, I.N., Sokolova, L.M., Bubnova, T.M., Okunev, O.N. & Sinitsyn, A.P. (2009). Isolation and Properties of Fungal β-glucosidases. *Biochemistry (Mosc),* 74(5), pp. 569-77.

Koshland, D.E. (1953). Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews of the Cambridge Philosophical Society,* 28(4), pp. 416-436.

Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr,* 60(Pt 12 Pt 1), pp. 2256-68.

Kubicek, C.P., Mikus, M., Schuster, A., Schmoll, M. & Seiboth, B. (2009). Metabolic engineering strategies for the improvement of cellulase production by Hypocrea jecorina. *Biotechnol Biofuels,* 2, p. 19.

Kumar, R. & Wyman, C.E. (2009). Effect of additives on the digestibility of corn stover solids following pretreatment by leading technologies. *Biotechnol Bioeng,* 102(6), pp. 1544-57.

Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem,* 77, pp. 521-55.

Langsford, M.L., Gilkes, N.R., Singh, B., Moser, B., Miller, R.C., Warren, R.A.J. & Kilburn, D.G. (1987). Glycosylation of Bacterial Cellulases Prevents Proteolytic Cleavage between Functional Domains. *Febs Letters,* 225(1-2), pp. 163-167.

Langston, J.A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E. & Sweeney, M.D. (2011). Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microbiol,* 77(19), pp. 7007-15.

Lee, R.C., Hrmova, M., Burton, R.A., Lahnstein, J. & Fincher, G.B. (2003). Bifunctional Family 3 Glycoside Hydrolases from Barley with α-L-Arabinofuranosidase and β-D-Xylosidase activity. Characterization, primary structures, and COOH-terminal processing. *The Journal of biological chemistry,* 278(7), pp. 5377-87.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M. & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels,* 6.

Li, X., Beeson, W.T.t., Phillips, C.M., Marletta, M.A. & Cate, J.H. (2012). Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure,* 20(6), pp. 1051-61.

Lima, M.A., Oliveira-Neto, M., Kadowaki, M.A.S., Rosseto, F.R., Prates, E.T., Squina, F.M., Leme, A.F.P., Skaf, M.S. & Polikarpov, I. (2013). *Aspergillus niger* β-Glucosidase Has a Cellulase-like Tadpole Molecular Shape - Insights into Glycoside Hydrolase family 3 (GH3) β-Glucosidase

Structure and Function. *Journal of Biological Chemistry,* 288(46), pp. 32991-33005.

Litzinger, S., Fischer, S., Polzer, P., Diederichs, K., Welte, W. & Mayer, C. (2010). Structural and kinetic analysis of *Bacillus subtilis N*-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *The Journal of biological chemistry,* 285(46), pp. 35675-84.

Liu, D., Zhang, R., Yang, X., Zhang, Z., Song, S., Miao, Y. & Shen, Q. (2012). Characterization of a thermostable β-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. *Microb Cell Fact,* 11, p. 25.

Logan, D.T., Su, X.D., Aberg, A., Regnstrom, K., Hajdu, J., Eklund, H. & Nordlund, P. (1996). Crystal structure of reduced protein R2 of ribonucleotide reductase: The structural basis for oxygen activation at a dinuclear iron site. *Structure,* 4(9), pp. 1053-1064.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M. & Henrissat, B. (2010). A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal,* 432, pp. 437-444.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M. & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research,* 42(D1), pp. D490-D495.

Lovering, A.L., Lee, S.S., Kim, Y.W., Withers, S.G. & Strynadka, N.C. (2005). Mechanistic and Structural Analysis of a Family 31 α-Glycosidase and its Glycosyl-enzyme Intermediate. *The Journal of biological chemistry,* 280(3), pp. 2105-15.

Lu, F. (2014). *Lignin : structural analysis, applications in biomaterials and ecological significance*. (Biochemistry research trends. New York: Nova Publishers.

Maddi, A., Bowman, S.M. & Free, S.J. (2009). Trifluoromethanesulfonic acid-based proteomic analysis of cell wall and secreted proteins of the ascomycetous fungi *Neurospora crassa* and *Candida albicans*. *Fungal Genet Biol,* 46(10), pp. 768-81.

Margaritis, A. & Merchant, R. (1984). Production and Thermal Stabiity Characteristics of Cellulase and Xylanase Enzymes from *Thielavia terrestris*. In: *Scott, C. D.* (Biotechnology and Bioengineering Symposium, pp. 299-314. Available from: <Go to ISI>://BCI:BCI198528054884.

Margaritis, A. & Merchant, R.F.J. (1986). Thermostable Cellulases from Thermophilic Microorganisms. *Crc Critical Reviews in Biotechnology,* 4(3), pp. 327-367.

Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., Danchin, E.G., Grigoriev, I.V., Harris, P., Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Larrondo, L.F., de Leon, A.L., Magnuson, J.K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A.A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C.L., Yao, J., Barabote, R., Nelson, M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Richardson, P., Rokhsar, D.S., Lucas, S.M., Rubin, E.M., Dunn-Coleman,

N., Ward, M. & Brettin, T.S. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol,* 26(5), pp. 553-60.

Matthews, B.W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology,* 33(2), pp. 491-7.

Mattinen, M.L., Linder, M., Drakenberg, T. & Annila, A. (1998). Solution structure of the cellulose-binding domain of endoglucanase I from *Trichoderma reesei* and its interaction with cello-oligosaccharides. *European Journal of Biochemistry,* 256(2), pp. 279-86.

Mayer, C., Vocadlo, D.J., Mah, M., Rupitz, K., Stoll, D., Warren, R.A. & Withers, S.G. (2006). Characterization of a β-*N*-acetylhexosaminidase and a β-*N*-acetylglucosaminidase/β-glucosidase from *Cellulomonas fimi*. *Febs Journal,* 273(13), pp. 2929-41.

McCarter, J.D. & Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol,* 4(6), pp. 885-92.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr,* 40(Pt 4), pp. 658-674.

McGeehan, J., Ravelli, R.B.G., Murray, J.W., Owen, R.L., Cipriani, F., McSweeney, S., Weik, M. & Garman, E.F. (2009). Colouring cryo-cooled crystals: online microspectrophotometry. *Journal of Synchrotron Radiation,* 16(2), pp. 163-172.

Merino, S. & Cherry, J. (2007). Progress and Challenges in Enzyme Development for Biomass Utilization. In: Olsson, L. (ed. *Biofuels*. (Advances in Biochemical Engineering/Biotechnology, 108) Springer Berlin Heidelberg, pp. 95-120.

Moser, F., Irwin, D., Chen, S. & Wilson, D.B. (2008). Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol Bioeng,* 100(6), pp. 1066-77.

Murray, P., Aro, N., Collins, C., Grassick, A., Penttila, M., Saloheimo, M. & Tuohy, M. (2004). Expression in *Trichoderma reesei* and characterisation of a thermostable family 3 β-glucosidase from the moderately thermophilic fungus *Talaromyces emersonii*. *Protein Expression and Purification,* 38(2), pp. 248-257.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr,* 53(Pt 3), pp. 240-55.

Nakatani, Y., Cutfield, S.M., Cowieson, N.P. & Cutfield, J.F. (2012). Structure and activity of exo-1,3/1,4-β-glucanase from marine bacterium *Pseudoalteromonas sp.* BB1 showing a novel C-terminal domain. *Febs Journal,* 279(3), pp. 464-478.

Nicholson, R.L. & Hammerschmidt, R. (1992). Phenolic-Compounds and Their Role in Disease Resistance. *Annual Review of Phytopathology,* 30, pp. 369-389.

Payne, C.M., Himmel, M.E., Crowley, M.F. & Beckham, G.T. (2011). Decrystallization of Oligosaccharides from the Cellulose Iβ Surface with Molecular Simulation. *J. Phys. Chem. Lett.,* 2(13), pp. 1546-1550.

Payne, C.M., Jiang, W., Shirts, M.R., Himmel, M.E., Crowley, M.F. & Beckham, G.T. (2013). Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J Am Chem Soc,* 135(50), pp. 18831-9.

Petersen, T.N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods,* 8(10), pp. 785-6.

Pettersen, R.C. (1984). The Chemical-Composition of Wood. *Advances in Chemistry Series*(207), pp. 57-126.

Phillips, C.M., Beeson, W.T., Cate, J.H. & Marletta, M.A. (2011). Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by *Neurospora crassa*. *Acs Chemical Biology,* 6(12), pp. 1399-1406.

Pozzo, T., Pasten, J.L., Karlsson, E.N. & Logan, D.T. (2010). Structural and Functional Analyses of β-Glucosidase 3B from *Thermotoga neapolitana*: A Thermostable Three-Domain Representative of Glycoside Hydrolase 3. *Journal of Molecular Biology,* 397(3), pp. 724-739.

Quinlan, R.J., Sweeney, M.D., Lo Leggio, L., Otten, H., Poulsen, J.C., Johansen, K.S., Krogh, K.B., Jorgensen, C.I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C.P., Dupree, P., Xu, F., Davies, G.J. & Walton, P.H. (2011). Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U S A,* 108(37), pp. 15079-84.

Reese, E.T., Siu, R.G. & Levinson, H.S. (1950). The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J Bacteriol,* 59(4), pp. 485-97.

Rigden, D.J., Mello, L.V. & Galperin, M.Y. (2004). The PA14 domain, a conserved all-β domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends in Biochemical Sciences,* 29(7), pp. 335-339.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K. & Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science,* 249(4967), pp. 380-6.

Salvi, D.A., Aita, G.M., Robert, D. & Bazan, V. (2010). Dilute ammonia pretreatment of sorghum and its effectiveness on enzyme hydrolysis and ethanol fermentation. *Appl Biochem Biotechnol,* 161(1-8), pp. 67-74.

Scheller, H.V. & Ulvskov, P. (2010). Hemicelluloses. *Annu Rev Plant Biol,* 61, pp. 263-89.

Schlichting, I., Berendzen, J., Chu, K., Stock, A.M., Maves, S.A., Benson, D.E., Sweet, B.M., Ringe, D., Petsko, G.A. & Sligar, S.G. (2000). The catalytic pathway of cytochrome P450cam at atomic resolution. *Science,* 287(5458), pp. 1615-1622.

Schwarzinger, S., Wright, P.E. & Dyson, H.J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry,* 41(42), pp. 12681-6.

Seidl, V., Seibel, C., Kubicek, C.P. & Schmoll, M. (2009). Sexual development in the industrial workhorse *Trichoderma reesei*. *Proc. Natl. Acad. Sci. U S A,* 106(33), pp. 13909-14.

Sinnott, M.L. & Souchard, I.J. (1973). Mechanism of Action of β-Galactosidase - Effect of Aglycone Nature and α-Deuterium Substitution on Hydrolysis of Aryl Galactosides. *Biochemical Journal,* 133(1), pp. 89-98.

Sjögren, T. & Hajdu, J. (2001). Structure of the bound dioxygen species in the cytochrome oxidase reaction of cytochrome cd1 nitrite reductase. *J. Biol. Chem.,* 276(16), pp. 13072-13076.

Sorrell, T.N., Garrity, M.L., Richards, J.L. & White, P.S. (1994). Synthesis, structural characterization and dioxygen reactivity of imidazole-ligated Cu (I) complexes. *Inorganica chimica acta,* 218(1), pp. 103-108.

Stals, I., Sandra, K., Geysens, S., Contreras, R., Van Beeumen, J. & Claeyssens, M. (2004). Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the *O*- and *N*-glycosylation pattern of Cel7A. *Glycobiology,* 14(8), pp. 713-24.

Sutherland, I.W. (1995). Polysaccharide Lyases. *Fems Microbiology Reviews,* 16(4), pp. 323-347.

Suto, M. & Tomita, F. (2001). Induction and catabolite repression mechanisms of cellulase in fungi. *Journal of Bioscience and Bioengineering,* 92(4), pp. 305-311.

Suzuki, K., Sumitani, J., Nam, Y.W., Nishimaki, T., Tani, S., Wakagi, T., Kawaguchi, T. & Fushinobu, S. (2013). Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from *Aspergillus aculeatus*. *Biochemical Journal,* 452, pp. 211-221.

Tansey, M.R. (1971). Agar-Diffusion Assay of Cellulolytic Ability of Thermophilic Fungi. *Archiv Fur Mikrobiologie,* 77(1), pp. 1-&.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Adams, P.D., Read, R.J., Zwart, P.H. & Hung, L.W. (2008). Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta crystallographica. Section D, Biological crystallography,* 64(Pt 5), pp. 515-24.

Tian, C., Beeson, W.T., Iavarone, A.T., Sun, J., Marletta, M.A., Cate, J.H. & Glass, N.L. (2009). Systems analysis of plant cell wall degradation by the model filamentous fungus *Neurospora crassa*. *Proc. Natl. Acad. Sci. U S A,* 106(52), pp. 22157-62.

Tiemeyer, M., Selleck, S.B. & Esko, J.D. (2009). Arthropoda. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

Vaaje-Kolstad, G., Bohle, L.A., Gaseidnes, S., Dalhus, B., Bjoras, M., Mathiesen, G. & Eijsink, V.G.H. (2012a). Characterization of the Chitinolytic Machinery of Enterococcus faecalis V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme. *Journal of Molecular Biology,* 416(2), pp. 239-254.

Vaaje-Kolstad, G., Bøhle, L.A., Gåseidnes, S., Dalhus, B., Bjørås, M., Mathiesen, G. & Eijsink, V.G.H. (2012b). Characterization of the Chitinolytic Machinery of Enterococcus faecalis V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme. *Journal of Molecular Biology,* 416(2), pp. 239-254.

Vaaje-Kolstad, G., Horn, S.J., van Aalten, D.M.F., Synstad, B. & Eijsink, V.G.H. (2005a). The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem.,* 280(31), pp. 28492-28497.

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

Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z.L., Zhai, H., Sørlie, M. & Eijsink, V.G.H. (2010). An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science,* 330(6001), pp. 219-222.

Vance, C.P., Kirk, T.K. & Sherwood, R.T. (1980). Lignification as a Mechanism of Disease Resistance. *Annual Review of Phytopathology,* 18, pp. 259-288.

Varghese, J.N., Hrmova, M. & Fincher, G.B. (1999). Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure,* 7(2), pp. 179-90.

Varki, A. & Sharon, N. (2009). Historical Background and Overview. In: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. & Etzler, M.E. (eds) *Essentials of Glycobiology*. 2nd. ed. Cold Spring Harbor (NY).

Vebo, H.C., Snipen, L., Nes, I.F. & Brede, D.A. (2009). The transcriptome of the nosocomial pathogen *Enterococcus faecalis* V583 reveals adaptive responses to growth in blood. *PloS One,* 4(11), p. e7660.

Vebo, H.C., Solheim, M., Snipen, L., Nes, I.F. & Brede, D.A. (2010). Comparative genomic analysis of pathogenic and probiotic *Enterococcus faecalis* isolates, and their transcriptional responses to growth in human urine. *PloS One,* 5(8), p. e12489.

Vogel, H.J. (1956). A convenient growth medium for *Neurospora* (medium N). *Microbial Genetics Bullentin,* 13, pp. 42-43.

Vu, V.V., Beeson, W.T., Span, E.A., Farquhar, E.R. & Marletta, M.A. (2014). A family of starch-active polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U S A,* 111(38), pp. 13822-7.

Wang, W., Nema, S. & Teagarden, D. (2010). Protein aggregation-Pathways and influencing factors. *International Journal of Pharmaceutics,* 390(2), pp. 89-99.

Westereng, B., Ishida, T., Vaaje-Kolstad, G., Wu, M., Eijsink, V.G.H., Igarashi, K., Samejima, M., Ståhlberg, J., Horn, S.J. & Sandgren, M. (2011). The Putative Endoglucanase PcGH61D from *Phanerochaete chrysosporium* Is a Metal-Dependent Oxidative Enzyme that Cleaves Cellulose. *PloS One,* 6(11), p. e27807.

Whetten, R. & Sederoff, R. (1995). Lignin Biosynthesis. *Plant Cell,* 7(7), pp. 1001-1013.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read,

R.J., Vagin, A. & Wilson, K.S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67(Pt 4), pp. 235-42.

Wojtczak, G., Breuil, C., Yamada, J. & Saddler, J.N. (1987). A Comparison of the Thermostability of Cellulases from Various Thermophilic Fungi. *Applied Microbiology and Biotechnology,* 27(1), pp. 82-87.

Wolfenden, R., Lu, X.D. & Young, G. (1998). Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society,* 120(27), pp. 6814-6815.

Wood, T.M. & McCrae, S.I. (1978). The cellulase of *Trichoderma koningii*. Purification and properties of some endoglucanase components with special reference to their action on cellulose when acting alone and in synergism with the cellobiohydrolase. *Biochemical Journal,* 171(1), pp. 61-72.

Wu, M., Beckham, G.T., Larsson, A.M., Ishida, T., Kim, S., Payne, C.M., Himmel, M.E., Crowley, M.F., Horn, S.J., Westereng, B., Igarashi, K., Samejima, M., Ståhlberg, J., Eijsink, V.G.H. & Sandgren, M. (2013a). Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus *Phanerochaete chrysosporium*. *J. Biol. Chem.,* 288(18), pp. 12828-12839.

Wu, W., Kasuga, T., Xiong, X., Ma, D. & Fan, Z. (2013b). Location and contribution of individual β-glucosidase from *Neurospora crassa* to total β-glucosidase activity. *Arch Microbiol,* 195(12), pp. 823-9.

Yano, J., Kern, J., Irrgang, K.D., Latimer, M.J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A. & Yachandra, V.K. (2005). X-ray damage to the Mn4Ca complex in single crystals of photosystem II: a case study for metalloprotein crystallography. *Proc. Natl. Acad. Sci. U S A,* 102(34), pp. 12047-52.

Yip, V.L. & Withers, S.G. (2004). Nature's many mechanisms for the degradation of oligosaccharides. *Org Biomol Chem,* 2(19), pp. 2707-13.

Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., Kitaoka, M., Katayama, T. & Kumagai, H. (2010). Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochemical Journal,* 431, pp. 39-49.

Zechel, D.L. & Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res,* 33(1), pp. 11-8.

Zhao, L.S., Beyer, N.J., Borisova, S.A. & Liu, H.W. (2003). β-Glucosylation as a Part of Self-Resistance Mechanism in Methymycin/Pikromycin Producing Strain *Streptomyces venezuelae*. *Biochemistry,* 42(50), pp. 14794-14804.

Zmudka, M.W., Thoden, J.B. & Holden, H.M. (2013). The structure of DesR from *Streptomyces venezuelae*, a β-glucosidase involved in macrolide activation. *Protein Science,* 22(7), pp. 883-892.

Zverlov, V.V., Volkov, I.Y., Velikodvorskaya, T.V. & Schwarz, W.H. (1997). *Thermotoga neapolitana bglB* gene, upstream of *lamA*, encodes a highly thermostable *β*-glucosidase that is a laminaribiase. *Microbiology,* 143 (Pt 11), pp. 3537-42.

6 Acknowledgements

Välkommen till den delen av avhandlingen som alla kommer att läsa! Och som jag, traditionsenligt, sitter och skriver såhär på natten innan inlämning. And for my international readers: welcome to section that people will actually read. I'm writing this section hours (as is my custom) before submission and still under the influence of a very persistent cold virus.

First of all I would like to thank **Mats Sandgren** for being an understanding and most excellent supervisor. The equally excellent "senior" members of Sandgren/Ståhlberg (or is it Ståhlberg/Sandgren?) lab **Jerry Ståhlberg, Nisse, Henrik, Saeid,** thanks for always being helpful and supportive people. Thanks to my (almost) constant office mates, PhD partners in crime and the two thirds of the 3Ms, **Majid** (will you learn Danish?) and **Miao** for being the smart supportive colleagues you are. Thanks to the remaining young PhD-students: **Anna Borisova**, the smartest Russian I know (you are now the senior crystallography PhD-student (no pressure!)), **Jule** good luck with your PhD! Thanks to all past and present Postdocs, **Takuya Ishida**, **Frida Jacobsson** and **Maria Dimarogona,** for all help and support. Thanks **Johanna B** (är du jäst/gästforskare eller -postdoc?) welcome to group and the office, and **Geniia** (you should also be in the office right?).

Thanks to all my colleagues at the department, the **Torleif Härd** crew: **Christoffer Lendel, Ray, Mahfuz** and **Benjamin.** Also thanks to all new(?) "Chemistry" colleagues.

Thanks to whole BMC/X-ray gang, old and less old, for all the enjoyable lunches, AWs, beer-clubs, fikas and good times: **Christoffer, Lotta, Henrik, Christoffer L, Anna J, Glareh, Annette, Henrik B, Lars B, Anatoly, Avinash, Agata, Yafei, Tiger, Wangshu, Li, Dee, Åsa, Alwyn, Tex, Nina, Sherry, Torsten, Margaretha, Inger, Evalena** and everyone else that I might have forgotten!

I would also like to thank the Toronto people for being fantastic friends for a year, **Marie**, **Julie-Anne**, **Sedric**, **Alex, Mabel, Ruoyu, Tracy** and the rest of the Master lab. You are all awesome and I hope I will see all of you many more times!

Sen vill jag tacka min familj. Tack Irja och Lennart för att ni låtit mig forska mig fram i livet istället för att skaffa mig ett riktigt jobb, ni är de bästa föräldrar man kan ha. Tack lillasyster Johanna som envisas med att vara i landsflykt i Norrland.

Tack Johanna. Jag älskar dig. Du är mitt allt. Tack för att du står ut med mig (framför allt de senaste veckorna).

Praeterea censeo Carthaginem esse delendam.