Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Chemistry, Biotechnology
and Food Science

Master Thesis 2015
60 credits

Expression and Characterization of Lytic Polysaccharide Monooxygenases from Myriococcum thermophilum

Margrethe Gaardløs

Acknowledgements

The following thesis was carried out at the Department of Chemistry, Biotechnology and Food Science at the Norwegian Univeristy of Life Sciences with Prof. Svein Jarle Horn and Trine Isaksen as supervisors.

First, I would like to thank my supervisors for all their help and support during my thesis. Trine, thank you for always being there when I needed help in the lab and with the writing of my thesis, and for all of your support and patience. Svein, thank you for all your excellent advice.

I would also like to thank all the members of the Protein Engingeering and Proteomics (PEP) group for creating a good working environment and helping me in the lab. And thanks to group leader Vincent Eijsink for giving me the opportunity to write my thesis at the PEP group.

At last I want to express my deepest gratitude to my family and friends for their support and encouragement. A huge thanks goes to my parents for always answering the phone when I needed it! Special thanks goes to my fellow master students for always making me smile and for their moral support whenever it was needed, no matter the reason.

Abstract

In the field of biomass utilization, better enzyme technologies need to be developed for more efficient production of biofuels, materials and chemicals. In this regard, the study of the novel oxidative carbohydrate degrading enzymes lytic polysaccharide monooxygenases (LPMOs) are important. In this master thesis three LPMOs from the filamentous fungi *Myriococcum thermophilum*, a cellulose degrading organism with a genome containing over 20 possible LPMOs, are characterized.

One of these three enzymes, together with its truncated version, was successfully cloned and expressed in the yeast *Pichia pastoris*. All three enzymes were characterized using bioinformatics, chromatography, mass spectrometry and colorimetric analysis. Important structural elements and activities were investigated and the three enzymes were compared to each other and to previously characterized LPMOs from other fungi.

By using bioinformatic tools the enzymes were found to most likely belong to three different subclasses of LPMOs. These classes differ in their structures and sequences and use slightly different mechanisms for cellulose degradation, forming different oxidized products. The enzymes showed activity on different substrates and yielded different products, and these results confirmed affiliation to the three subclasses. It is interesting and probably important for cellulose degradation in nature that a wide diversity of these enzymes are found in certain microorganisms. This study adds new information to the as of now quite limited knowledge base that exists for this interesting, newly discovered class of enzymes.

Sammendrag

Industriell utnyttelse av biomasse er avhengig av bedre enzymteknologier for å kunne produsere biodrivstoff, materialer og kjemikalier på en mer effektiv måte. I denne sammenhengen er studiet av de oksidative karbohydratnedbrytende enzymene lytisk polysakkarid monooksygenaser (LPMOer) viktig. I denne masteroppgaven karakteriseres tre LPMOer fra den filamentøse soppen *Myriococcum thermophilum*, en cellulosenedbrytende organisme med et genom som inneholder mer enn 20 mulige LPMOer.

Et av disse enzymene, sammen med en trunkert versjon, ble vellykket klonet og uttrykt i gjæren *Pichia pastoris*. Alle tre enzymene ble deretter karakterisert ved hjelp av bioinformatikk, kromatografi, massespektrometri og kolorimetri. Viktige strukturelle elementer og aktiviteter ble studert, og de tre enzymene ble sammenlignet med hverandre og med tidligere karakteriserte LPMOer fra andre typer sopp.

Enzymene viste seg å mest sannsynlig tilhøre tre ulike underklasser av LPMOer. Disse klassene har små variasjoner i struktur og sekvens, og ved å bruke litt ulike mekanismer for cellulosenedbrytning dannes ulike oksiderte produkter. Enzymene var aktive på ulike substrater og dannet forskjellige produkter, og disse resultatene bekreftet tilhørigheten til de tre underklassene. Det er interessant og antageligvis viktig for cellulosenedbrytning i naturen at det eksisterer et stort mangfold av disse enzymene i enkelte mikroorganismer. Denne studien bygger videre på den så langt begrensede kunnskapsbasen som eksisterer for denne interessante, nyoppdagede klassen enzymer.

Abbreviations

aa Amino Acid

BLAST Basic Local Alignment Search Tool

bp Basepairs

CAZy Carbohydrate-Active enzyme

CBM Carbohydate Binding Module

CDH Cellobiose dehydrogenase

CMC Carboxymethylcellulose

DMC Direct microbial conversion

DNS 3,5-dinitrosalisylic acid

DP Degree of polymerization

FAD Flavin adenine dinucleotide

GH Glycoside hydrolases

HIC Hydrophobic interaction chromatography

HPAEC High-performance anion exchange chromatography

IEC Ion exchange chromatography

kb Kilobases

kDa Kilodalton

LB Luria Bertani

LPMO Lytic Polysaccharide Monooxygenase

MALDI-TOF Matrix-Assisted Laser Desorption and Ionization Time Of Flight

MSA Multiple sequence alignment

NaAc Sodium Acetate

NPP Net primary productivity

ON Over night

PASC Phosphoric acid swollen Avicel

PCR Polymerace Chain Reaction

pI Isoelectric point

rpm Rotations per minute

LDS-PAGE Lithium Dodecyl Sulphate Polyacryl Amide Gel Electrophoresis

w/v Weight/volume

YPDS/YPD Yeast Extract Peptone Dextrose with/ without Sorbitol

Contents

1 Introduction **10**

1.1 Can biomass replace petroleum? 10

1.1.1 Biomass and its human use 10

1.1.2 Two generations of biomass 13

1.2 Carbohydrates in connection with the biomass industry 15

1.2.1 Monosaccharides, the building blocks of polysaccharides 16

1.2.2 Starch 16

1.2.3 Cellulose 17

1.2.4 Hemicelluloses 18

1.2.5 Making value of lignocellulose 19

1.3 Enzymes involved in cellulose degradation 20

1.3.1 Cellulases 21

1.3.2 Lytic Polysaccharide Monooxygenases 22

1.3.3 Classification and phylogeny 24

1.3.4 Reaction mechanism of LPMOs 25

1.3.5 LPMO structure and conserved residues 27

1.3.6 Cellobiose dehydrogenase and other reducing agents for LPMOs 30

1.3.7 Carbohydrate binding modules 31

1.4 Study of three LPMOs from *Myriococcum thermophilum* 32

2 Materials and methods **34**

2.1 Materials 34

2.1.1 Carbohydrate substrates 34

2.1.2 Chemical list 35

2.1.3 Microbial strains and plasmids 36

2.1.4 Primers 36

2.1.5 Software 37

2.1.6 Kits 37

2.1.6.1 NucleoSpin® Plasmid (NoLid) High copy number Miniprep kit (Macherey-Nagel) . . . 37
2.1.6.2 JetStar™ 2.0 Plasmid Midiprep Kit (Genomed) . . . 38
2.1.6.3 Quant-iT™ dsDNA Assay Kit Broad-Range (Invitrogen) 38
2.1.6.4 NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) . 39
2.1.6.5 In-Fusion® HD cloning kit (Clontech) . . . 39
2.1.7 Recipes for agars, media and gels . . . 39
2.1.7.1 Brain Heart Infusion (BHI) medium . . . 39
2.1.7.2 Luria Bertani (LB) . . . 40
2.1.7.3 LB Low Salt . . . 40
2.1.7.4 SOC, Super Optimal Broth . . . 40
2.1.7.5 YPD, Yeast Extract Peptone Dextrose . . . 40
2.1.7.6 YPDS, YPD Sorbitol . . . 40
2.1.8 Other recipes . . . 40
2.1.8.1 Agarosegel 12 % . . . 40
2.1.8.2 DHB matrix . . . 41
2.1.8.3 DNS reagent . . . 41
2.1.8.4 Sodium hydroxide, 1 M . . . 41
2.1.8.5 TAE-buffer 1X . . . 41
2.2 Methods . . . 41
2.2.1 Cloning of the genes . . . 41
2.2.1.1 Glycerol stocks of cultures . . . 41
2.2.1.2 Preparing the gene and plasmid for cloning . . . 42
2.2.1.3 Designing primers . . . 42
2.2.1.4 PCR with Q5 High-Fidelity 2X Master mix . . . 43
2.2.1.5 Restriction cutting . . . 44
2.2.1.6 Agarose gel electrophoresis and DNA-purification from gel . . . 44
2.2.1.7 In-Fusion® cloning of the gene into the plasmid . . . 45
2.2.1.8 Transformation of plasmid into *E.coli* . . . 46
2.2.1.9 Colony PCR with Taq DNA Polymerase . . . 47
2.2.1.10 Purification and linearisation of plasmid before transformation in PichiaPink™ . . . 48
2.2.1.11 Sequencing . . . 48
2.2.1.12 Transformation of PichiaPink™ cells by electroporation 49
2.2.2 Expressing and purifying the proteins . . . 50
2.2.2.1 Testing and analyzing protein expression . . . 50

2.2.2.2 Lithium dodecyl sulfate polyacrylamide gel electrophoresis (LDS-PAGE) . . . 51
2.2.2.3 Large scale protein expression . . . 51
2.2.2.4 Increasing protein concentration . . . 52
2.2.2.5 Precipitation with ammonium sulfate . . . 53
2.2.2.6 Hydrophobic Interaction Chromatography . . . 53
2.2.2.7 Ion Exchange Chromatography . . . 55
2.2.3 Enzyme characterization . . . 56
2.2.3.1 Measuring protein concentrations . . . 56
2.2.3.2 Copper saturation . . . 56
2.2.3.3 Substrate degradation . . . 57
2.2.3.4 High-performance anion-exchange chromatography . . . 58
2.2.3.5 MALDI-TOF mass spectrometry . . . 60
2.2.3.6 Reducing end assay with 3,5-dinitrosalisylic acid . . . 61

3 Results **63**
3.1 Bioinformatics . . . 63
3.1.1 Protein parameters . . . 63
3.1.2 Domain structures and protein modifications . . . 63
3.1.3 Structural studies . . . 65
3.1.4 Multiple sequence alignment . . . 70
3.2 Cloning, expression and characterization . . . 73
3.2.1 Cloning of *Mt*4260 . . . 73
3.2.2 Expression and purification . . . 75
3.2.3 Characterization of enzyme activities . . . 77

4 Discussion **82**
4.1 Expression and purification . . . 82
4.2 Sequence and structure analysis . . . 83
4.3 Future perspectives . . . 87

A Amino acid sequences **98**

B Plasmid maps **100**

C Purification with HIC and IEC **102**

D Chromatograms from characterization **105**

1. Introduction

1.1 Can biomass replace petroleum?

In 1925 Henry Ford saw the possibilities associated with biological material. "We can get fuel from fruit, from that shrub by the roadside, or from apples, weeds, sawdust—almost anything! There is fuel in every bit of vegetable matter that can be fermented. There is enough alcohol in one year's yield of a hectare of potatoes to drive the machinery necessary to cultivate the field for a hundred years. And it remains for someone to find out how this fuel can be produced commercially—better fuel at a cheaper price than we know now" ("Energy from waste and wood," 2015).

Petroleum based products are an important part of today's market, not only for fuel but for the production of chemicals and materials as well. This will not last forever, as the crude oil will eventually run out. There is also reason to be concerned for the security of countries if they are depending on oil from a few nations (Zhang, 2013). In addition, there are huge environmental problems associated with the use of fossil fuels. We live in a world where the energy demand is growing while the old sources of energy are running out and there is a need for new technologies that can develop energy and products from other sources than petroleum. A possible alternative to petroleum is biomass, if we can develop technologies for cheap and effective processing of it.

In this introduction I will focus on lignocellulosic biomass. I will look into some of the prospects and potential problems of increased biomass utilization, technological aspects and challenges and general characteristics of the polysaccharides that constitute the biomass. Then I will introduce the main theme of this thesis, which is the enzymatic breakdown of lignocellulose by one specific class of enzymes, the lytic polysaccharide monooxygenases (LPMOs).

1.1.1 Biomass and its human use

Biomass is a storage of solar energy defined as material that comes from plants, animals and microbes, and was originally produced by CO_2 fixation through photosyntesis. Biomass can replace fossil resources for the production of chemicals and materials,

and it is the only source of renewable liquid fuel with a high energy density. In 2008 renewable energy sources contributed to about 26 % of the worlds energy consumption and biomass was by far the largest source of renewable energy, constituting about 10 % of the global energy supply (Chum *et al.*, 2011). But even though biomass is renewable it is a limited material. The term terrestrial net primary productivity (NPP) refers to the carbon fixated in plants when the energy needed for their respiration is subtracted, and represents the theoretical maximum of plant material that we can utilize. In Figure 1.1 the human demand for biomass is shown as a percentage of the NPP all over the world, illustrating the large global differences in resource supply and demand. Globally, food and feed production uses 2 % of the NPP, burning of biomass for cooking and heating uses 2.3 %, production of construction materials, paper and polymers uses 1 % of NPP and liquid transportation fuels uses 0.2 %. It is thought that potentially about 12.3 % (double of today's consumption, i.e., 130 EJ) of the world's biomass could actually be used, the rest is unavailable or its use raises environmental concerns (Zhang, 2013).

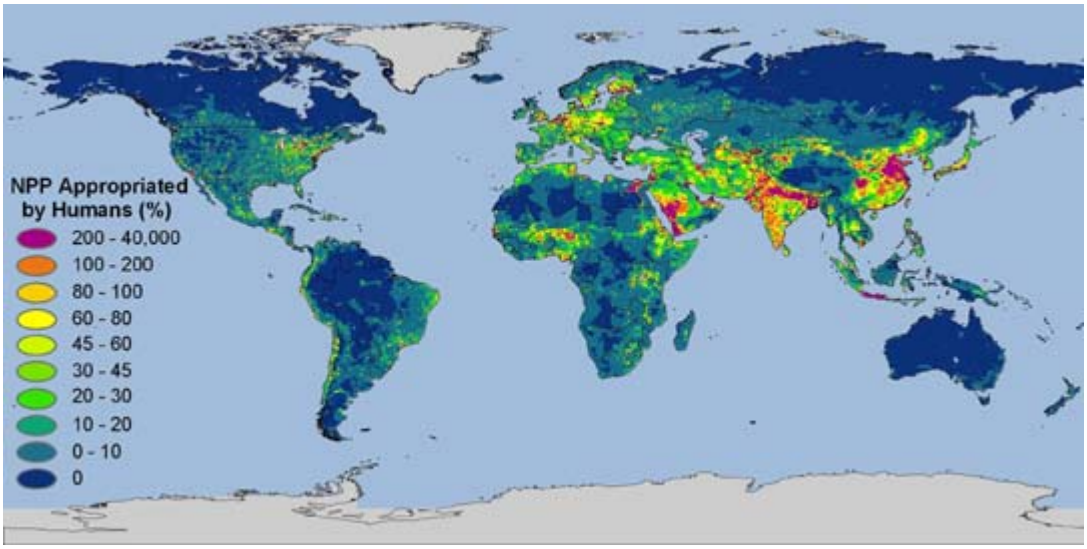

Figure 1.1: In this map the net primary production of plants is compared to the human demand. The colors show the demand as a percentage of the NPP in different areas. Orange and magenta is where the demand exceeds the production, while dark blue shows a low utilization compared to production. Figure source: Imhoff *et al.* (2004)

In the bioenergy industry the products have low value, so the production must be cost efficient and give high yields of the main product while at the same time utilize byproducts as much as possible. The industries will always be pressed to find the best available technologies, and also the cheapest possible starting products that are needed in large volumes. Compared to petroleum the feedstock cost for lignocellulose is so small that it might even out the more expensive processing, making the product price lower for cellulosic products (Lynd *et al.*, 1999). Today, however, the technology for processing

lignocellulosic material is the limiting factor both because of the recalcitrance of the biomass and the diverse intermediate products that results from its breakdown (Lynd *et al.*, 1999; Rubin, 2008). Biofuel production typically generates side streams of protein and lignin. A good biorefinery would be able to utilize these side streams and reduce the fuel prices by producing fuels, chemicals, electricity and feed at the same time. This could also create a market where different industries would not have to compete for the same feedstocks (Lynd *et al.*, 1999; Tuck *et al.*, 2012).

There are two different routes in which products from the biomass industry could enter the market of petroleum based products. Biomass products could directly replace petroleum based products, utilizing already existing infrastructures. But since biomass and petroleum are feedstocks with different chemical compositions, thermolabilities and solvent environments, substitution of the already established products with new ones might be the most sustainable route in long term. A good example is polyethylene produces from carbohydrates, where the yield is low because mass is lost due to the loss of oxygen. Polymers of lactic acid are similar to polyethylene in their characteristics and could be a good substitution at a lower cost compared to petroleum because of cheaper feedstock (Lynd *et al.*, 1999).

In Figure 1.2 the pathways between the human needs and the sustainable energy sources are shown. Potential pathways when the technology improves in the future are also drawn.

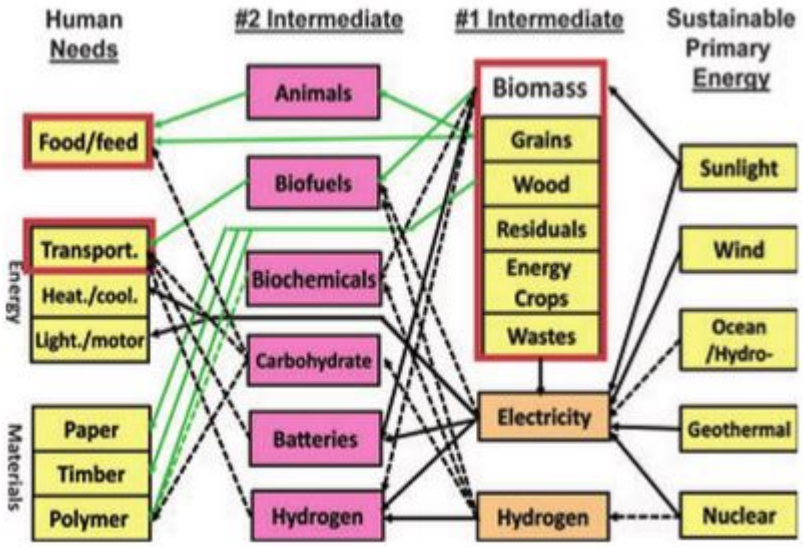

Figure 1.2: Sustainable primary energy sources can meet the human needs in a complex network of pathways. Solid lines show actual conversions while the dashed lines show theoretical ones. Green, solid lines are from biomass sources. Biomass is a large, important intermediate from sun energy with many actual and potential uses and it is the only source of food, feed and renewable materials Figure source: Zhang (2013).

1.1.2 Two generations of biomass

First generation biomass is associated with starch rich materials, plant oils and animal fats and is today used for the production of first generation biofuels such as ethanol and biodiesel. This biomass is relatively easy to process and there is an existing infrastructure for its production. However, the increased use of this biomass for energy will cause problems as both the plants and the land they are grown on have other applications, most importantly for food and animal feed. The human population will probably peak at about 9 billion people, but the wealth is thought to increase and more consumers will be capable of choosing processed and high energy food (Godfray *et al.*, 2010). Together with increased competition for resources and harder weather conditions due to climate change this will put a high pressure on the global food production. There could be a need for as much as 70 to 100 % more food by 2050 (Godfray *et al.*, 2010; You *et al.*, 2013). Competition between food and fuel production could increase food prices and harm poor societies, as is already seen for existing biofuel productions (Tait *et al.*, 2011). Because of the effects of land use change and the amount of resources required to produce the crops even the positive effects for the climate are debatable (Godfray *et al.*, 2010; Kim & Dale, 2005; Searchinger *et al.*, 2008; Shapouri *et al.*, 2010). The use of first generation biomass also cause problems regarding loss of biodiversity, unfair working conditions and land grabbing. Although improvements are made recently (Liska *et al.*, 2009), the world needs new biomass sources that don't compete with food, have high energy content compared with invested energy, don't affect the environment or societies negatively and can be produced in large volumes (Tait *et al.*, 2011; Zhang, 2013).

When wood is used as an energy source it is called wood fuel and the term comprises everything from firewood used for heating or cooking, industrial firewood, charcoal, pellets and biogas to bioethanol (Boucher *et al.*, 2011). Lignocellulose, the structural component of wood and other plant sources, could replace oil as a raw material for fuels, chemicals and materials without most of the problems associated with first generation biomasses. The need for new energy sources has triggered research on what is called second generation biofuels, fuels made from biomass like lignocellulose and algae. Lignocellulosic crops can be grown on lands that can not be used for food production and with less input of resources than starch or sugar crops (You *et al.*, 2013; Zhang, 2013). If the lignocellulosic plants are converted completely to energy they will give substantially higher yields per hectare than most other feedstocks. But new technologies have to be developed before lignocellulose can replace first generation biomass, as it is more difficult to process. Today the lignocellulosic biomass is mostly processed thermochemically (Chum *et al.*, 2011). If this can be switched to biochemical

methods, as is the case for starch and sugar crops, this can be done more sustainably.

There is problems associated with this biomass based economy as well. Wood consumption is thought to increase with 20-50 % by 2050 (Zhang, 2013), and this can lead to increased use of charcoal as wood fuel which has been highly correlated with forest degradation (Boucher *et al.*, 2011). While the global food demand is thought to increase with 50-100 % by 2050, already 30 % of the world's arable land and 70 % of its fresh water withdrawal are for food and feed production. The remainding land should not be cultured as this would have large environmental and sociological consequences (Zhang, 2013). Food production is of course a more important use of land than production of energy and products. Some argue, however, that in poor countries with enough land sustainable bioenergy could lead to new sources of income and development of value from areas that are not suitable for production of food (Lynd & Woods, 2011). Before international policies are implemented that can ensure a sustainable wood fuel use in all parts of the world, both when it comes to human rights and forest protection, the production of any type of wood fuel will have issues.

Alternatives to wood as biomass includes dedicated biofuel crops like *Mischanthus* and switchgrass, or even bamboo. But compared with trees the intensive production of biofuel crops would require large amounts of water and land area (Zhang, 2013) and most of these species do not grow in cold climates (Rubin, 2008). Biomass could also come from agricultural, industrial and municipal waste and this would circumvent the utilization of more land while at the same time taking care of huge waste problems (Tuck *et al.*, 2012; van Wyk, 2001). Even though waste is regarded as unwanted material, almost all organic wastes have some value whether they are polysaccharides, lignin, triglycerides or proteins (Tuck *et al.*, 2012; van Wyk, 2001). Both economics and sustainability must be taken into account when biomass is converted into products.

Renewable natural gas (RNG), which is high quality methane obtained from biomass, is a proposed fuel alternative as it is carbon neutral or even carbon negative, easy to use in already existing energy infrastructures and comes from non-food resources (waste) (Mozaffarian *et al.*, 2004). Because of slow implementation of cellulose fuel production, RNG is today characterized as cellulosic biofuel by the US Environmental Protection Agency to meet political goals (EPA, 2014). The best imaginable scenario could be if waste was used in addition to locally suitable energy crops all over the world, especially as 30-40 % of food goes to waste in both developed and developing countries (Godfray *et al.*, 2010).

Even without the discussion of whether or not wood fuel is the best alternative, it is a fact that the technologies of the industrial degradation of lignocellulosic biomass need development. Use of enzymes in biorefineries is expensive, and the current enzyme technologies must be improved so that the processes used get more effective

(Klein-Marcuschamer *et al.*, 2012). Another important improvement would be genetical optimization of crop species to improve yields, efficiencies and robustness and make them easier to process (Rubin, 2008), but this will not be discussed further here.

1.2 Carbohydrates in connection with the biomass industry

Carbohydrates are a huge and diverse group of organic molecules abundant in nature. They are used in organisms as a source and storage of energy, as mechanical structures, for cell signaling, in enzymes, in ribozymes and in genetic material to name a few examples. When it comes to industrial applications the structure of the plant cell wall makes carbohydrates difficult to work with. It is the outer part of plant cells and contains a complex network of carbohydrates and lignin that protects the plant from the environment and from internal turgidity pressure (Figure 1.3). High lignin content, the diversity of hemicelluloses and the insoluble and crystalline nature of cellulose are large challenges for both natural and industrial enzymatic degradation. But despite its recalcitrance lignocellulose does not accumulate in the biosphere, and its degradation can be understood by studying the organisms and enzyme systems that can break it down and how they cooperate. (Himmel *et al.*, 2007; Lynd *et al.*, 1999).

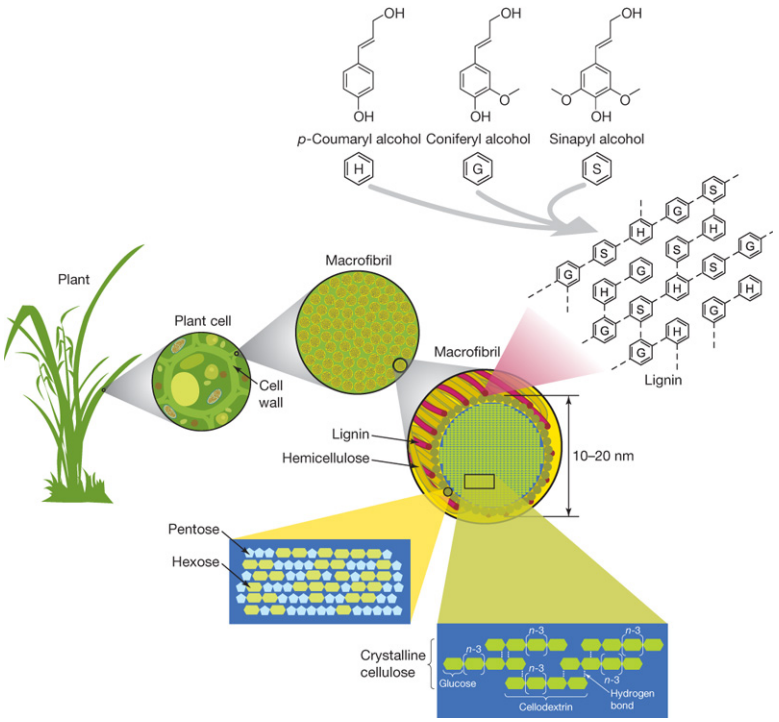

Figure 1.3: Lignocellulose consists of a crystalline cellulosic core surrounded by the more amorphous hemicellulose and aromatic lignin. Macrofibrils consists of microfibrils (mistakenly assigned macrofibrils in the figure) with the three polymers in a complex network. Figure source: Rubin (2008).

1.2.1 Monosaccharides, the building blocks of polysaccharides

Polysaccharides are polymers of monosaccharides, polyhydroxylated ring structures of aldehydes and ketones. These rings can be five- or six membered, called furanose or pyranose respectively, and they can have either five or six carbons, named pentoses or hexoses. The carbon in the aldehyde or ketone group is called the anomeric carbon and is bound to two oxygens. If the hydroxyl group of this carbon is in the axial position the conformation is α, if it is in the equatorial position it is called β. Unbound these two conformations are interchangeable, but in polymers they are fixed as the monosaccharides are linked through the anomeric carbon. Monosaccharides can be in L or D conformation, where the chiral carbon farthest from the anomeric carbon is either on the right side (D) or left side (L) in the Fischer projection of the molecule. The equatorial position is the energetically most favorable position for the hydoxyl groups, making α-D-glucose the most stable hexose and the one found in nature.

Glucose is the most important and abundant monosaccharide. Figure 1.4 shows cellobiose, a disaccharide of glucose occurring in cellulose. Its chemical name is β-D-glucose (1$\rightarrow$4)-D-glucose because the anomeric carbon of the first glucose, assigned C1, is fixed in the β-position and linked to the 4-carbon (C4) of the other D-glucose.

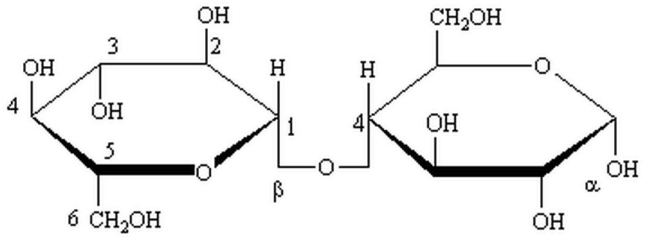

Figure 1.4: A cellobiose molecule, showing the conformation of the linkage and the numbering of the carbon atoms. Figure source: Wikimedia (2015)

The degree of polymerization (DP) of carbohydrate substrates is the number of sugar monomers in the polysaccharides.

1.2.2 Starch

Plants use the glucose polymer starch as energy storage and it is also the most important component of the human diet worldwide. It consists of irregularly branched amylopectin and linear amylose, and it is insoluble and semicrystalline. The glycosidic bonds in starch are α-(1$\rightarrow$4), with α-(1$\rightarrow$6) links creating the branches. Starch is used as an energy storage and both plants and animals contain amylases that can break it down.

The most important starch crops are used as food and feed. These include seeds from rice, maize and wheat, tubers and root crops like potatoes and carrots, and seeds of beans and peas. Nonfood industries are using starch as a renewable material for products like paper, textiles, plastics and first generation biofuels. Amylose has a high value as raw material for a range of products, and synthetic amylose can be made from cellulose to avoid the competition of these products with the food industry (You *et al.*, 2013; Zeeman *et al.*, 2010).

1.2.3 Cellulose

Cellulose is the most abundant biopolymer in the world and is the main component of lignocellulosic biomass. It exists mainly in the cell wall of plants and typically constitutes around 50 % of a tree's dry weight (Brigham *et al.*, 1996). It only contains D-glucose, arranged in long unbranched cellulose molecules with β-(1→4) links. These molecules are arranged in protofibrils consisting of around 30 cellulose chains, which again are arranged in microfibrils (Lynd *et al.*, 2002). This is assembled in insoluble fibers surrounded by a complex mixture of structural polymers as illustrated in Figure 1.3, the most important ones being lignin and hemicellulose (Brigham *et al.*, 1996; Lynd *et al.*, 2002). The glucose chains in cellulose are arranged so that strong hydrogen bonds form between them. These bonds, together with van der Waals forces between sheets of the linked chains and a dense water layer around the sheets, make it hard for enzymes to hydrolyse the polysaccharide. Cellulose in nature exists in different degrees of crystallization, where amorphous cellulose is easier to degrade than crystalline (Himmel *et al.*, 2007; Lynd *et al.*, 2002).

Different choices of substrates for studies of cellulose degrading enzymes can affect the results because of differences in accessibility to the enzymes. Pure cellulose substrates like Avicel are treated with dilute acid to remove hemicelluloses and amorphous cellulose regions and the particle size of these microcrystalline substrates is important for how easy they can be broken down (Lynd *et al.*, 2002). Phosphoric acid swollen cellulose (PASC) is an artificially treated Avicel and an example of a substrate with a low degree of crystallinity (Stålbrand *et al.*, 1998), while acetylated cellulose is an amorphous cellulose substrate. Other, less pure cellulosic substrates contain various degrees of crystallinity of the cellulose and can contain other components like lignin and hemicelluloses. This makes the breakdown more complex than for pure cellulosic substrates and the study of this breakdown is not as straight forward (Lynd *et al.*, 2002). However this study is important, as these substrates are more like the natural ones and the ones the industry have to make use of (Kipper *et al.*, 2005).

1.2.4 Hemicelluloses

Hemicelluloses surround the cellulosic core in plant cell walls and constitute typically around 20 % of the dry weight in trees. They are a complex and diverse group of polysaccharides that gives strength and protection to plants, and the percentage of different hemicelluloses varies a lot in different plant types. Common sugars in hemicellulose are arabinose, xylose, galactose, fucose, mannose and glucose (see Figure 1.5) and non-sugar substitutions can be glucuronic acid and acetyl groups. The backbone usually consists of only one type of monomer while the substitutions can be diverse (Brigham *et al.*, 1996; Mosier *et al.*, 2005). β-D-xylopyranose that can be linked with (1→4) bonds to form xylan constitutes the majority of hemicellulose. Xyloglucan is another important hemicellulose that consists of linear chains of β-(1→4)-D-glucan with xylose substitutions and like most hemicelluloses it is soluble in water when it is not bound to cellulose (Carpita & Gibeaut, 1993). Other examples of hemicelluloses are arabinoxylan, β-glucan, galactomannan, glucomannan and lichenan. The last part of the hemicellulose names describes the backbone, while the substitutions are mentioned first, so for instance arabinoxylan has a xylan backbone with arabinose substitutions. A deviation from this rule is lichenan, a β-glucan named after lichens where it is found.

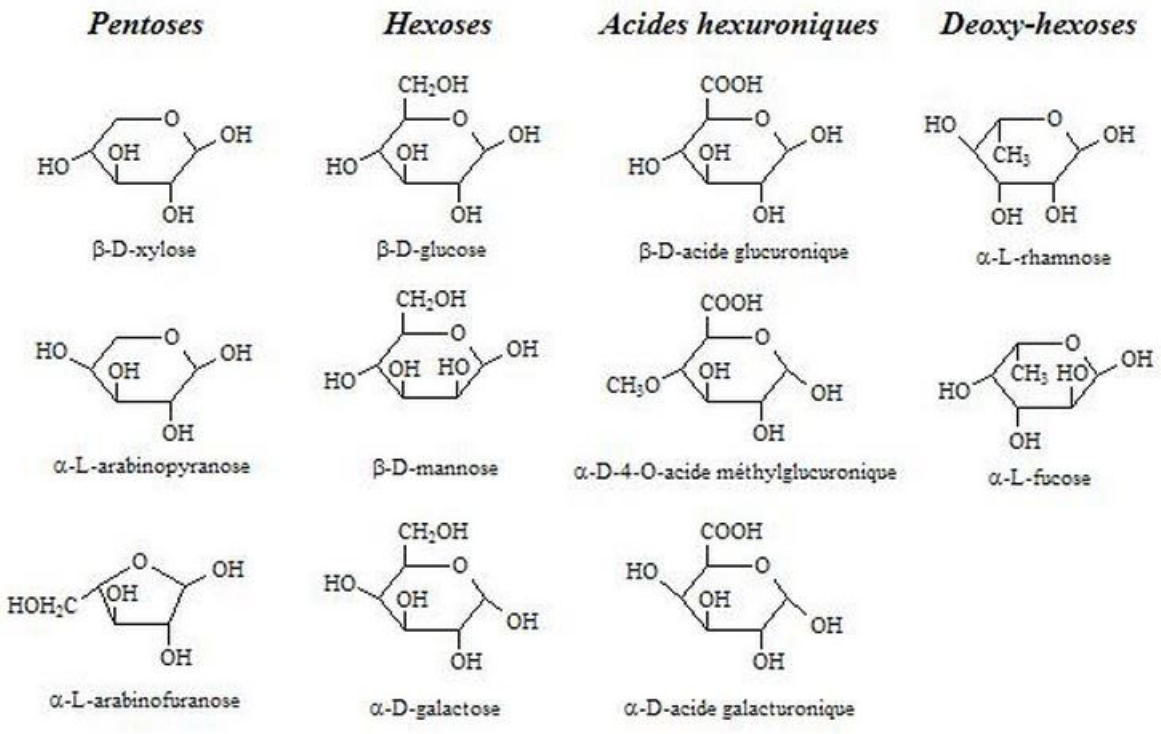

Figure 1.5: Some common hemicellulosic sugars. Figure source: ("Formation d'ingénieurs en Hydraulique et Mécanique des Fluides" 2015)

1.2.5 Making value of lignocellulose

Lignocellulose has to go through four steps in its production to ethanol. First it is pretreated, before it is hydrolysed into monomeric sugars, then the sugars are fermented into ethanol and at last the ethanol is purified (Mosier *et al.*, 2005). It is possible to hydrolyse lignocellulose and ferment the products in the same mixture, a process called simultaneous saccharification and fermentation (SSF). If the cellulases are produced in the same process it is called direct microbial conversion (DMC) (also called consolidated bioprocessing). This method is how nature does it and it could be very efficient, but it needs large amounts of optimization of pretreatment, heterologous enzyme mixtures and fermentation organisms for different substrates (Lynd *et al.*, 1999; Wyman *et al.*, 1992; Zhang, 2013).

Pretreatment of cellulosic substrates before hydrolysis makes the cellulose chains easier accessible by creating more amorphous cellulose and removing lignin and hemicellulose, but it is expensive (Mosier *et al.*, 2005; Zhu *et al.*, 2009). The pretreatment can be biological, chemical or physical, and common pretreatments are steam explosion or treatment with dilute acid or base. Good pretreatment results in high release of sugars from the hemicellulose and yields an easily digestible cellulosic product, it is cheap, and it minimizes the production of harmful waste products (Lynd *et al.*, 1999).

After pretreatment cellulose and hemicellulose must be broken down to their monomers before they can be fermented, and this is done by thermochemical, acidic or enzymatic hydrolysis. Enzymatic hydrolysis is effective and environmental friendly (van Wyk, 2001), and it gives products that are easily fermented. This is in contrast to acid hydrolysis or pyrolysis, which tend to produce fermentation inhibitors (Segato *et al.*, 2012). Thermochemical processes could however be useful for production of electricity from biomass, especially from residues rich in lignin (Lynd *et al.*, 1999). The aromatic lignin is the most important non-carbohydrate part of the cell wall, up to 25 wt.% in trees (Brigham *et al.*, 1996), and contains phenylpropanoid compounds forming a hydrophobic network in the carbohydrate matrix. Its utilization is difficult and will not be discussed further here, but as of today only the production of vanillin is commercial (Lynd *et al.*, 1999; Tuck *et al.*, 2012).

Fermentation of the hydrolyzed sugars by microorganisms produces compounds such as butanol, ethanol or lactic acid, all of which can be converted to important products like polymers and bulk chemicals. Ethanol and butanol can be used as fuels, of which butanol might be the most interesting one in the future as it is less hygroscopic and thus less corrosive, and as it has a higher energy content (Dürre, 2007). Ethanol can however be used to make polymers that could replace plastics from petroleum, like renewable polyethylene (Tuck *et al.*, 2012). Fermentation of the

hexoses glucose, galactose and mannose into ethanol is done naturally by a lot of microorganisms, while natural fermentation of pentoses is rare. The two pentoses xylose and arabinose are important components of hardwood hemicelluloses and genetically modified bacteria have been created to ferment them with high yields (Mosier *et al.*, 2005). Acid-catalyzed dehydration is an alternative to fermentation that can create chemicals with a higher complexity compared to fermentation products. An example is 5-hydroxymethyl-furfural (5-HMF), an attractive bulk chemical that can be converted into products like solvents, transportation fuels or polymers similar to PET polyesters (Tuck *et al.*, 2012). In Figure 1.6 examples of products that could be obtained from the different chemicals in lignocellulosic biomass are depicted.

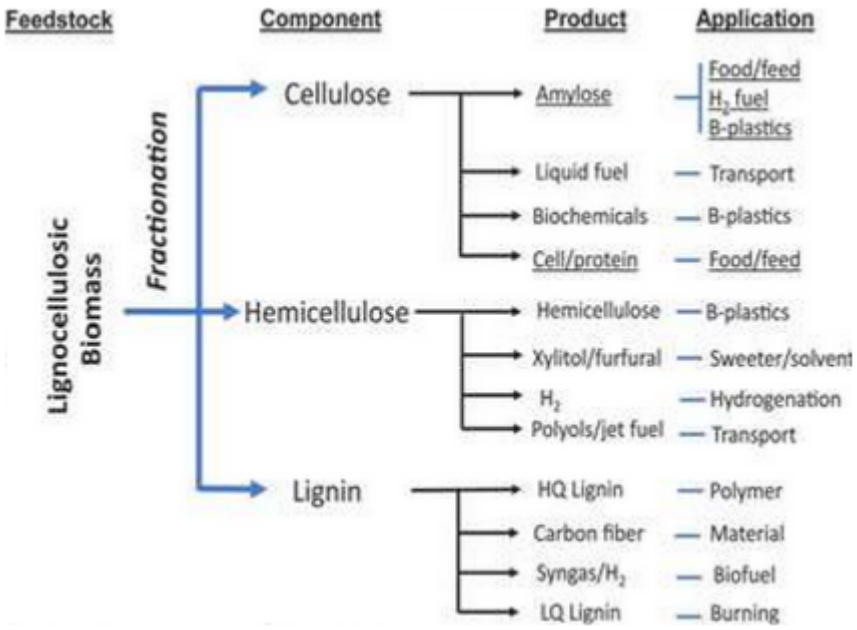

Figure 1.6: A diagram showing how lignocellulosic biomasscould be utilized, giving examples of possible products created from cellulose, hemicellulose and lignin. Figure source: Zhang (2013).

1.3 Enzymes involved in cellulose degradation

The breakdown of carbohydrates is essential for life on earth. Organisms utilize the enormous amount of energy stored in carbohydrates through enzymatic degradation and some of the most complex and interesting enzyme systems are found in organisms that are able to break down lignocellulose. Glycoside bonds in polysaccharides are strong compared to DNA or peptide bonds, making degradation of polysaccharides relatively difficult (Kim *et al.*, 2014). Nature has solved the problems of breaking down carbohydrates by creating large amounts of different enzymes with very specific modes of action, and by combining the actions of these enzymes efficient degradation

of polysaccharides can take place.

1.3.1 Cellulases

A comprehensive sequence, structure and mechanism based classification of carbohydrate active enzymes is found at http://www.cazy.org/ (Lombard *et al.*, 2014) and one of its five enzyme groups is glycoside hydrolases (GHs). Glycoside hydrolases are important for the hydrolysis of glycosidic bonds in carbohydrates and include cellulases that break cellulose bonds. There are three main types of cellulases: endo-acting and exo-acting glucanases and β-glucosidases (or cellobiases). Endoglucanases attack internal glycosidic linkages in amorphous regions by hydrolysis and create new ends. These new ends can then be processed by exoglucanases (also called cellobiohydrolases) into cellobiose (see Figure 1.4) that is cleaved by β-glucosidases into glucose (Himmel *et al.*, 2007). Exoglucanases work processively, which means that once bound to the substrate they make several cuts before dissociating (Kipper *et al.*, 2005). Cellulases can work inside the organism, they can be extracellular, or they can be part of organized structures bound to the cell wall called cellulosomes. Intracellular cellulases and cellulosomes are seen in bacteria, while fungi usually digests the polysaccharides with extracellular enzyme systems (Klyosov, 1990).

Synergism is seen between some of these enzymes, which in the context of cellulose degrading systems means that the rate of product formation and cellulose degradation when the enzymes work together exceeds the sum of the rates when they work by themselves. Synergy has been shown between endo- and exoglucanases, between exo- and exoglucanases and between endo- and endoglucanases as well as intramolecular synergy between binding domains and endoglucanases (Din *et al.*, 1994; Klyosov, 1990).

It has long been known that not all cellulase system can break down crystalline cellulose (Klyosov, 1990), and 65 years ago a hypothesis was formulated that cellulose degradation had a non-hydrolytic initial phase that was responsible for the disruption of cellulose crystal structures (Reese *et al.*, 1950). The authors assigned two systems for the degradation of cellulose, C_1 and C_x, illustrated in Figure 1.7. C_x enzymes are classic hydrolytic enzymes that break the β-(1→4) bonds and although they can hydrolyze some types of amorphous cellulose they can not on their own hydrolyze native cellulosic material. The other part of this theoretical system, C_1, was thought to consist of non-hydrolytic enzymes that facilitates the cellulose degradation with an unknown mechanism. The idea was that only true cellulolytic organisms contained the C_1 enzymes while the C_x enzymes were more common in microbes.

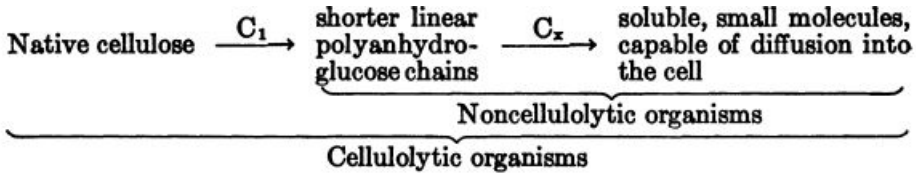

Figure 1.7: A diagram of the proposed two-phase cellulase system. The C_1 system was thought to work by an unknown, non-hydrolytic mechanism, and the reason why some organisms that contained the C_x enzymes could not degrade cellulose was because they were lacking these enzymes. Figure source: Reese *et al.* (1950)

In 1997 a new class of enzymes was reported by Saloheimo *et al.* (1997) that studied an enzyme in *Trichoderma reesei (*now called *Hypocrea jecorina)*, a well known model organism for fungal cellulose degradation. An enzyme called Cel61A (EGIV in the article) was reported to be a weak endoglucanase. It had a C-terminal cellulose binding domain and a N-terminal glycosyl hydrolase domain, it was induced by cellulose and it was suppressed by glucose. Together with a similar enzyme with an unknown function discovered a few years earlier, CELl from *Agaricus bisporus*, it was classified in a new fungal CAZy glycoside hydrolase family called GH61. Later more of these proteins were discovered, and they showed either weak or no endoglucanase activity. What seemed strange was that even though these enzymes apparently had minimal activity they were expressed together with more active cellulases working on the same substrates (Hara *et al.*, 2003; Karlsson *et al.*, 2001; Koseki *et al.*, 2008). Why would the organisms bother to use energy expressing these poor endoglucanases?

1.3.2 Lytic Polysaccharide Monooxygenases

It was work on a different class of enzymes that would solve this question. In 1994 and 2001 evidence was found that the binding modules of some cellulases not only helped binding of the substrate, but also had a non-hydrolytic activity that disrupted crystalline cellulose. It was proposed that the carbohydrate binding modules (CBMs) were the theoretical C_1 components in the cellulase system shown in Figure 1.7 (Din *et al.*, 1994; Gao *et al.*, 2001). In 2005 the structure was solved for a bacterial protein called CBP21 (chitin binding protein 21) from *Serratia marcescens (Vaaje-Kolstad* et al.*, 2005a).* At first it was categorized in the family 33 carbohydrate binding modules (CBM33) in the CAZy database, and it was reported to have no catalytic activity. It did not look like most other CBMs, because its binding surface was made up of hydrophilic instead of aromatic residues and it had a metal ion coordinated by conserved histidines. CBP21 generally enhanced chitinase efficiency by a mechanism that at first was not understood. Mutation experiments indicated that the binding of a metal ion was essential for this enhancing activity (Vaaje-Kolstad *et al.*, 2005b).

When the crystal structure of the fungal GH61 Cel61B from *T. reesei* (*Tr*LPMO9B, or *Hj*LPMO9B) was resolved in 2008 a structural comparison showed that it had a similar overall shape to the bacterial CBP21 (as illustrated in Figure 1.8), it had polar conserved surface residues, conserved histidines (His1 & His89 in Cel61B) and it could bind a metal ion in about the same position (Karkehabadi *et al.*, 2008). Like CBM33s the GH61 enzymes were shown to generally enhance the activity of cellulases without having any measurable glycoside hydrolase activity on their own. It was thought that these two enzyme classes could be using a similar mechanism for enhancing the degradation of chitin or cellulose (Harris *et al.*, 2010; Karkehabadi *et al.*, 2008).

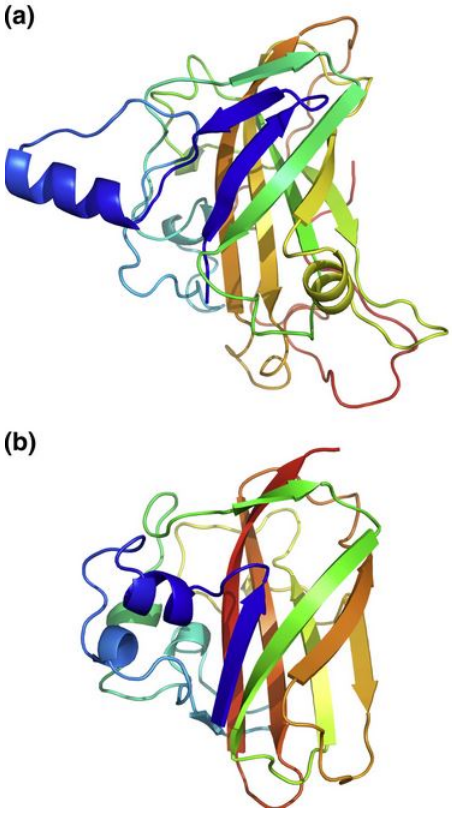

Figure 1.8: The similar folds of the cellulase enhancing enzyme Cel61B from *T. reseei* (a) and the chitinase enhancing enzyme CBP21 from *S. marcescsens* (b) are illustrated. Figure source: Karkehabadi *et al.* (2008)

In 2010 the mechanism of these two classes of proteins was beginning to be a little more understood. CBP21 was found to create chain breaks in chitin and produce chitooligosaccharides with one non-reducing and one aldonic acid end when incubated alone with chitin. Its activity was enhanced by adding reductants and it needed molecular oxygen, making it likely that it used an oxidative mechanism (Vaaje-Kolstad *et al.*, 2010). Both Harris *et al.* (2010) and Dimarogona *et al.* (2012b) found that GH61s had no effect on pure cellulose substrates and required a cellulose substrate with natural reductants like lignin to be active. In 2011 Forsberg *et al.* (2011) reported

that a CBM33 from *Streptomyces coelicolor* produced oxidized products (aldonic acids) when incubated with cellulose, apparently working on crystalline regions. Because the bacterial CBM33s and the fungal GH61s were so similar it was suggested that GH61s also used the same mechanism and further studies confirmed this (Forsberg *et al.*, 2011; Vaaje-Kolstad *et al.*, 2012; Westereng *et al.*, 2011). These findings could give a possible explanation for the C_1 system thought to exist in organisms able to break down crystalline polysaccharides like cellulose and chitin.

Today, CBM33s and GH61s are known as lytic polysaccharide monooxygenases (LPMOs) (Horn *et al.*, 2012). They cleave the $\beta(1\rightarrow4)$ glycosidic bond in chitin, cellulose and hemicellulose (Agger *et al.*, 2014) and the $\alpha(1\rightarrow4)$ (and possibly $\alpha(1\rightarrow6)$) bond in starch (Vu *et al.*, 2014b) by an oxidative mechanism. Their discovery made a fundamental change in the way of looking at carbohydrate degradation since they are lytic enzymes that cleave carbohydrates in an oxidative rather than a hydrolytic way. LPMOs enable the other polysaccharide degrading enzymes by introducing chain breaks in crystalline regions, making the chains more available for further degradation (Payne *et al.*, 2015). Some LPMOs are active on soluble cellooligosaccharides, and this discovery made it possible to show that the products of one LPMO were 4-keto sugars that hydrolysed to gemdiols in solution (Isaksen *et al.*, 2013). This implies that some LPMOs oxidize the C4 in the glycosidic bond, whereas other LPMOs are found to oxidize the C1 (Forsberg *et al.*, 2011; Vaaje-Kolstad *et al.*, 2010) and yet other can oxidize both C1 and C4 (Phillips *et al.*, 2011; Quinlan *et al.*, 2011).

Soon the industrial benefits of using these enzymes were clear. When LPMOs were added to enzyme cocktails for degradation of lignocellulose the activity was increased by a factor of 2, and this would probably reduce the cost by the same factor (Harris *et al.*, 2010). However it is possible that the oxidized products can not be utilized directly by fermentation microbes, and the anaerobic environment of DMC is not well suited for these oxidative enzymes. The reduced cost of using these enzymes have to be balanced with potential losses of carbohydrates as oxidized products, and with the need for oxygen supply in the degradation process (Beeson *et al.*, 2015).

1.3.3 Classification and phylogeny

Since the old classifications of the two enzyme groups into CBMs and GHs was not really appropriate, a new group called polysaccharide monooxygenases, later modified to lytic polysaccharide monooxygenases (LPMOs), was made. Some disagreement still exists as to wether the “lytic” part should be included or not (Beeson *et al.*, 2015; Payne *et al.*, 2015). In the CAZy database LPMOs are placed in three different subclasses in the auxiliary activity (AA) family. Cellulose degrading fungal enzymes formerly known as

GH61s are classified as AA9 (LPMO9), the CBM33 class is called AA10 (LPMO10) and contains chitin and cellulose degrading enzymes from all kingdoms, AA11 (LPMO11) contains chitin degrading fungal enzymes and AA13 (LPMO13) consists of starch degrading fungal enzymes. The auxiliary activity family consists of enzymes that are essential for the degradation of carbohydrates but that do not fit into the other, more established families. It also comprises redox lignin degrading enzymes as these are intertwined and in some cases even the same as carbohydrate degrading enzymes (Levasseur *et al.*, 2013).

The LPMOs can also be classified based on how their phylogeny is related to the carbon they oxidize in the glycosidic bond, in other words their regioselectivity. Type 1 LPMOs (PMO-1) oxidizes C1, type 2 LPMOs (PMO-2) C4 and type 3 (PMO-3) C1 and C4. Some LPMOs resembles the type 3 LPMOs but do not have C4 activity, and these are called type 3* LPMO (PMO-3*) subfamily. (Vu *et al.*, 2014a).

Phylogenetic studies show that fungal genomes often contain up to 30 different AA9 encoding genes (Harris *et al.*, 2010). Since these genes appear in multiple copies in genomes today, after a long time of evolutionary selection, their heterogenicity is probably an important feature for the cellulolytic ability of fungi. Ascomycota and Basidiomycota are two major fungal phyla that both contain LPMOs, indicating that the LPMOs are ancient proteins that could have formed before these phyla diverged more than 600 million years ago (Harris *et al.*, 2010). The fact that LPMOs exist both in bacteria, archaea (Forsberg, 2014) and fungi could suggest an even more ancient origin, or it could be a result of convergent evolution (Harris *et al.*, 2010). Even some viruses contain LPMOs, for instance does some insect viruses have chitin active LPMOs which together with chitinases are important for the colonization of the host (Vaaje-Kolstad *et al.*, 2005b).

1.3.4 Reaction mechanism of LPMOs

Oxidative cleavage of glycosidic bonds by LPMOs is thought to be done by a mechanism where the enzyme's metal center, containing a copper ion (see Section **??**), is activated to hydroxylate either the C1 or C4 carbon of the substrate. This hydroxylation forms an unstable intermediate that will eliminate a water and form either aldonolactones that can be hydrated to aldonic acids (C1) or 4-ketoaldoses that could exists in equilibrium with gemdiols (C4) (See Figure 1.9). C6 oxidation has also been suggested (Quinlan *et al.*, 2011), but it would not lead to cleavage of the glycosidic bond and it is not thought to happen (Vu *et al.*, 2014a).

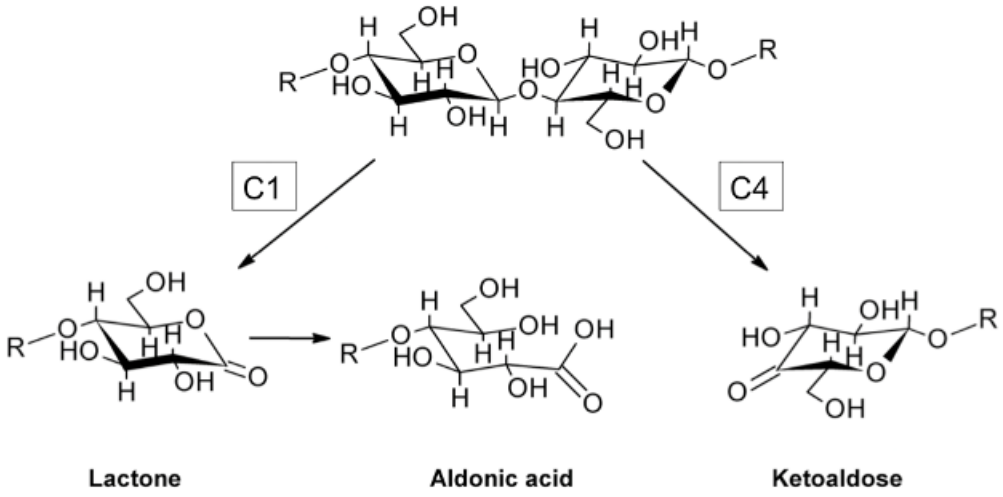

Figure 1.9: Reaction products by C1 and C4 oxidation from cellulose cleavage by LPMOs. Figure source: (Dimarogona *et al.*, 2012a).

Figure 1.10 shows four different paths that the hydrogen abstraction and hydroxylation can take. Common for all the paths is the reduction by an external reducing agent of the bound copper ion from Cu(II) to Cu(I) that in turn can bind molecular oxygen and reduce it to form a Cu(II)-superoxo species. It is not certain if the oxygen binding to the active site happens before or after the enzyme binds the substrate. After the formation of the superoxo complex the reaction can either follow the superoxo pathways (termed S1 and S2 in 1.10) or the oxyl pathways (termed Ox1 and Ox2 in 1.10). In the superoxo mechanisms it is the superoxo that abstracts a hydrogen atom from the substrate, forming a substrate radical and a hydro-superoxo complex. Then the copper center can either be reduced by an external reducing agent before it hydroxylates the substrate (S1), or it can hydroxylate the substrate before it is reduced (S2). In the oxyl pathways the superoxo complex is reduced to form a Cu(III)–oxyl radical and it is this oxyl that abstracts the hydrogen from the substrate. A hydro-superoxo complex is formed by a reducing agent instead of by the substrate, and this is either reduced to an oxyl that abstracts a hydrogen and subsequently hydroxylates the substrate (Ox1), or turned into an O-R species that then abstracts a hydrogen forming a Cu(III) hydroxyl that can hydroxylate the substrate (Kim *et al.*, 2014; Phillips *et al.*, 2011). Energy calculations have favored different pathways depending on the criteria used and it is not clear exactly what happens and with which LPMOs (Beeson *et al.*, 2015). It is possible that different mechanisms exists, as bacterial and fungal LPMOs have different conserved residues in their active sites as explained in Section **??** (Hemsworth *et al.*, 2013).

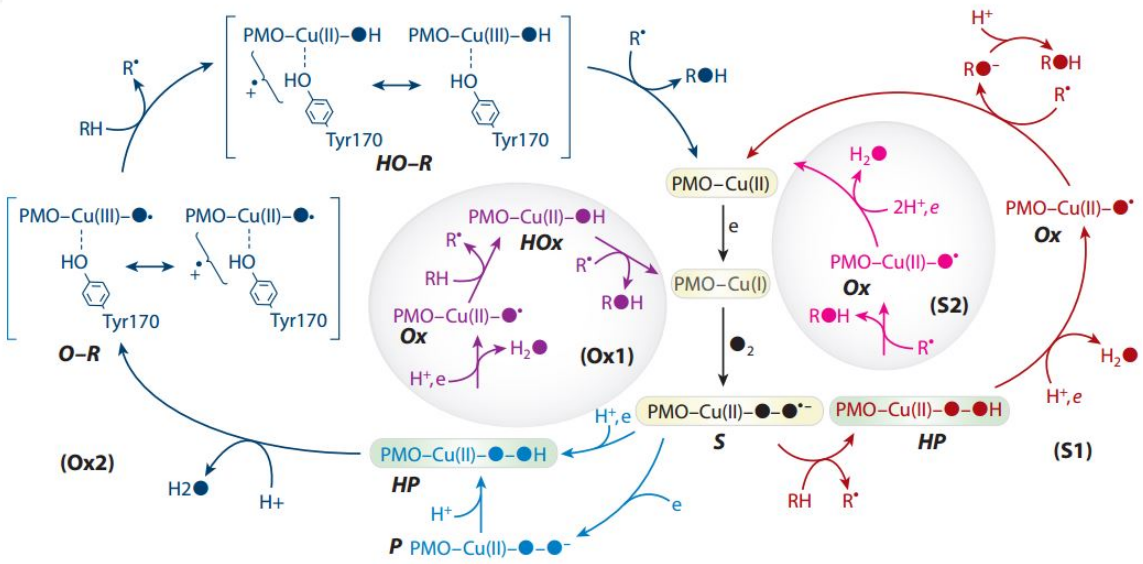

Figure 1.10: Diagram of the four proposed mechanism pathways for the LPMOs. The Ox1, Ox2, S1 and S2 mechanisms are shown in purple, dark blue, red and pink respectively. Figure source: Beeson *et al.* (2015).

1.3.5 LPMO structure and conserved residues

Even though LPMOs generally have a low sequence identity, they are structurally very similar. They all consist of about 200-250 amino acids forming a compact β-sandwich with 8-10 β strands and an immunoglobin-like fold. The binding surface is flat and suited to bind crystalline substrates as shown in Figure 1.11 (Aachmann *et al.*, 2012; Harris *et al.*, 2010). Since LPMOs do not need to pull their substrate chains into a cleft, like cellulases do, they can make cuts in crystalline carbohydrates in an energy efficient way. Fungal LPMOs can be glycosylated and in many cases they contain extra modules with known or unknown functions (Beeson *et al.*, 2015; Harris *et al.*, 2010; Karkehabadi *et al.*, 2008).

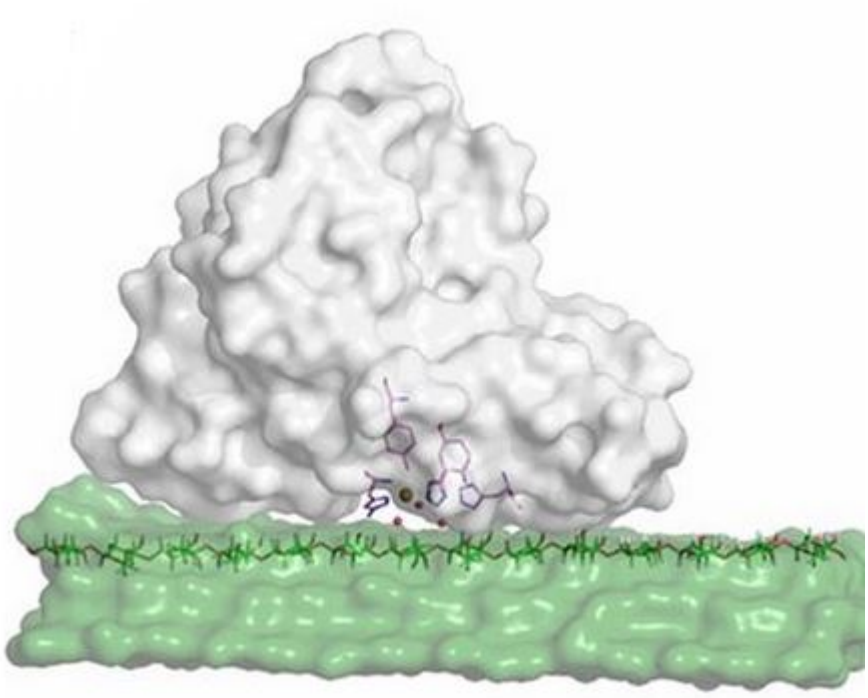

Figure 1.11: A model of a *Thermoascus aurantiacus* LPMO (gray) with cellulose (green) showing the active site and a cellulose chain in stick. Figure source: Kim *et al.* (2014)

The cores of LPMOs are rigid, while the binding surface has more flexible loop regions. These loops have varying secondary structures and conserved residues in different types of LPMOs and are thought to be important for substrate specificity and regioselectivity. Abstraction of hydrogen from the alcohol C4 requires more energy than from the aldehyde C1, and orientation of the active site relative to the bound substrate is probably what defines the enzyme's regioselectivity. The loop region L2 most likely play an important role in this orientation in fungal LPMOs. In type PMO-3s an insertion of about 12 amino acids gives an extended L2 loop that seems to be necessary for C4 oxidation, while in type PMO-2s a conserved lysine and a helix of 9-14 amino acids close to the active site is thought to serve the same purpose (Vu *et al.*, 2014a; Wu *et al.*, 2013). Some bacterial LPMOs can oxidize C4 as well and have a loop region analogous to the L2 loop in fungal LPMOs and differences in the active site that could define their regioselectivity. Borisova *et al.* (Submitted April 2015) also mentions a stretch of residues forming what is termed the L3 loop in C4 oxidizing AA9s and an active site residue that varies depending on the enzymes' regioselectivity. Other structural features that can affect substrate binding and catalysis is posttranslational modifications and depth of the active site groove (Beeson *et al.*, 2015; Forsberg *et al.*, 2014b; Vu *et al.*, 2014a).

Conserved residues in LPMOs are mostly on the binding surface, although there are also conserved aromatic core residues. LPMOs contain two highly conserved histidines forming a T-shaped histidine brace with three nitrogen ligands (Quinlan *et al.*, 2011). Figure 1.12 shows conserved residues in the copper coordinating active site of fungal LPMOs. The oxidation states of the copper ion are thought to be Cu(I) and Cu(II), but there is some uncertainty around this as the x-ray beam will reduce the copper when the crystal structures are elucidated (Hemsworth *et al.*, 2013). The three-ligand histidine brace is thought to coordinate Cu(I), while Cu(II) is probably coordinated by extra water ligands and the active site tyrosine (Gudmundsson *et al.*, 2014).

There are some important differences between fungal and bacterial LPMOs. The N-terminal histidine is conserved in all LPMOs, and in fungal LPMOs it is methylated by the native fungi (Li *et al.*, 2012; Quinlan *et al.*, 2011) but not by the expression yeast *Pichia pastoris* (Kittl *et al.*, 2012; Westereng *et al.*, 2011). The significance of this methylation is still not known and it is not vital for activity, but it might have something to do with copper coordination (Aachmann *et al.*, 2012; Hemsworth *et al.*, 2013). Another important residue is a conserved tyrosine in fungal LPMOs, which in bacteria is usually exchanged for a phenylalanine (Hemsworth *et al.*, 2013) except when the enzyme is C4 oxidative (Forsberg *et al.*, 2014a). AA10s have shown a lower copper affinity than AA9s (Quinlan *et al.*, 2011), maybe due to these differences in their metal binding site. In addition AA10s have a conserved alanine and the glutamine in Figure 1.12 is replaced by a glutamate (Gudmundsson *et al.*, 2014).

Harris *et al.* (2010) reported that mutagenesis of the different residues in the active site of AA9s showed that residues corresponding to His1, His83 and Gln164 in *Nc*LPMO9C (Figure 1.12) were essential for catalysis while Tyr166 (in *Nc*LPMO9C) was important. However the Gln164 are in some LPMOs a glutamate and the Tyr166 is, as already mentioned, sometimes a phenylalanine. The tyrosine residue could be important for C4 oxidation (Vu *et al.*, 2014a).

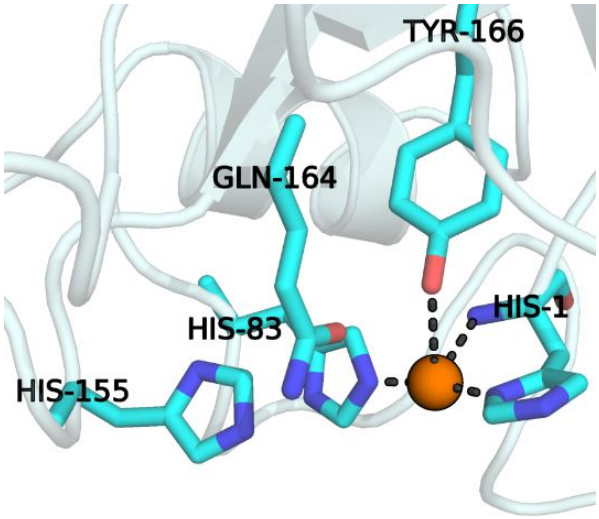

Figure 1.12: The three conserved histidines, the glutamine and the tyrosine in the active site of *Nc*LPMO9C (PDB code 4D7U) (Borisova *et al.*, Submitted April 2015). The three residues Gln164, His155 and Tyr166 forms a hydrogen bonding network with water molecules (not shown) (Vu *et al.*, 2014b). Copper might be coordinated by water molecules in addition to the four ligands shown (Gudmundsson *et al.*, 2014). The N-terminal histidine is not methylated in the model, probably because of expression in *P. pastoris*. Made with PyMOL (Schrödinger, LLC, 2010).

Structures with bound cellulose are difficult to obtain because of fast binding rates (Borisova *et al.*, Submitted April 2015), so it is not certain how the binding occurs. However, conserved aromatic residues on the binding surface in fungal LPMOs are thought to bind cellulose similar to what occurs in CBM1, see Figure 1.14. Based on the distance between aromatic residues the binding could be in one chain or over several (Li *et al.*, 2012). In bacterial LPMOs only one aromatic residue is conserved, making the cellulose binding more unclear (Beeson *et al.*, 2015).

1.3.6 Cellobiose dehydrogenase and other reducing agents for LPMOs

Cellobiose dehydrogenases (CDHs) are enzymes capable of oxidising cellooligosaccharides to cellolactones. They are part of the glucose-methanol-choline oxidoreductase superfamily (AA3 in the CAZy database) and contain a heme domain and a flavin adenine dinucleotide (FAD) domain. The FAD-domain is thought to donate electrons to the heme domain, which can further transfer the electrons to an electron acceptor. CDHs are expressed together with cellulases in most cellulolytic fungi and until recently their biological role has been mysterious (Phillips *et al.*, 2011). Their most likely function was thought to be that they promote cleavage of crystalline cellulose by creating hydroxyl radicals, but experiments by Langston *et al.* (2011) and Phillips *et al.* (2011) implied something else. Synergy was noticed between AA9 LPMOs and CDH

as well as with cellulases, and the CDH gene was shown to be important for cellulose degrading activity of *Neuraspora crassa*.

In fungal systems it is now thought that CDH is a natural electron donor for the LPMO9s, but little is known about this interaction (Langston *et al.*, 2011). Li *et al.* (2012) have proposed a possible conserved CDH binding motif in fungal AA9s, but this is not confirmed experimentally. From this binding surface to the active site there is a path of conserved residues that is hypothezised to work as an electron transfer pathway as shown in 1.13 (Beeson *et al.*, 2015; Li *et al.*, 2012). CDH does not exist in bacteria and can not be the universal electron donor for all LPMOs (Phillips *et al.*, 2011). However, similar, unidentified oxidoreductases could exist in bacteria. Ascorbate, gallic acid (Quinlan *et al.*, 2011) and lignin (Dimarogona *et al.*, 2012b) are examples of other molecules that have been shown to serve as reducing agents for LPMOs when CDH was not present.

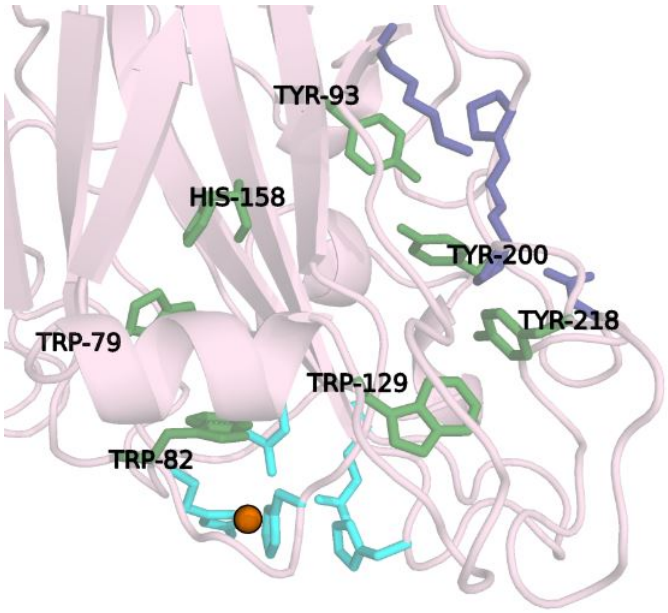

Figure 1.13: Structure of the AA9 *Ta*LPMO9B (PDB code 2YET) showing conserved or partly conserved aromatic amino acids that might play a role in electron transfer (in green) by connecting a possible CDH binding motif (in blue) to the active site of LPMOs (in turquoise). This hypothesis is from Li *et al.* (2012). Made with PyMOL (Schrödinger, LLC, 2010).

1.3.7 Carbohydrate binding modules

The carbohydrate binding modules (CBMs) are domains within enzymes that can bind carbohydrates. Generally CBMs are associated with glycoside hydrolases that are active on insoluble substrates, but large substrate diversity exists. They can bind very spesific ligands or they can bind various ones, and this is important for the way they concentrate the catalytic domains to their substrates and place them in proximity to

their sites of action. In some cases they can even be responsible for processivity (Sakon *et al.*, 1997) and substrate specificity (Araki *et al.*, 2004) of the enzyme. This makes them interesting subjects for studying carbohydrate binding and their effect can be explored by engineering enzymes with removed CBMs (Araki *et al.*, 2004; Din *et al.*, 1994). For instance, truncation of the CBMs from glycoside hydrolases has shown to decrease the enzyme activities on insoluble substrates, although not on soluble ones (Din *et al.*, 1994).

Usually CBMs have a β-sandwich fold, with the ligand recognition site in the same plane. Type A CBMs (Boraston *et al.*, 2004), including family 1 CBMs, bind to the surface of insoluble crystalline cellulose or chitin through three aromatic amino acids that rotate to form a flat binding platform (Mattinen *et al.*, 1997). Because of the spacing between these amino acids it is likely that the module binds to every second glucose unit in a cellulose chain by ring stacking as shown in Figure 1.14. The topography of amino acid side chains and loops are thought to be important for binding specificity, but hydrogen bonds and calcium ions can also play a role (Boraston *et al.*, 2004). One or more CBM1 domains are found in about 20 % of known cellulose active AA9s (Borisova *et al.*, Submitted April 2015), and contains around 40 amino acids (Lombard *et al.*, 2014). Between the modules in multimodular enzymes there are flexible linker regions with high sequence variation that can exhibit different post translational modifications. The length and modifications of these linkers might affect enzyme activity, and the linkers can even bind to the substrate themselves (Payne *et al.*, 2013).

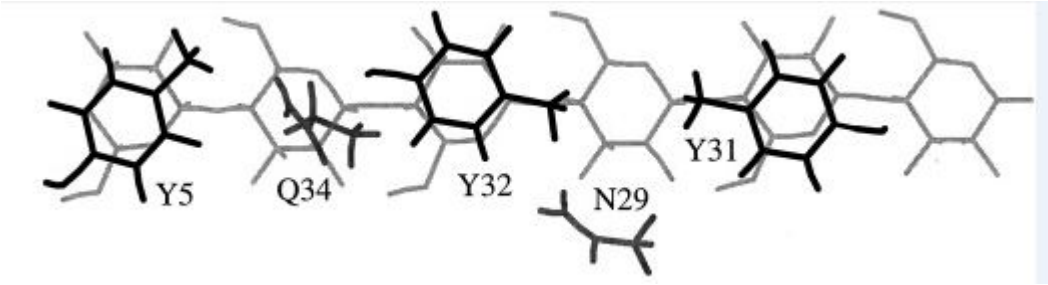

Figure 1.14: Binding model of a CBM1s showing the rings of three conserved aromatic amino acids stacked on top of glucose rings. The distances between the amino acids is about the same as the distance between two glucose units in a cellulose chain. Figure source: Mattinen *et al.* (1997).

1.4 Study of three LPMOs from *Myriococcum thermophilum*

Animals can not process cellulose, but bacteria and fungi can. To find sources of cellulose processing enzymes, it is therefore natural to look at the microbial world. In 1971 a

filamentous thermophilic ascomycete in the order Sordariales called *Myriococcum thermophilum* (other name: *Papulaspora thermophila*) was found in self-heating compost (Fergus, 1971). Sordiareles are saprobic fungi, which means that they digest dead plant and animal matter extracellularly. *N. crassa* is another well-known member of this order. *M. thermophilum* can digest lignocellulose (Chapman *et al.*, 1975), and according to the Genozymes Project Public Genomes database at http://genome.fungalgenomics.ca/ it contains 22 possible AA9 genes. It also has a CDH that has been studied for its applications as a bleaching agent, as a part of biosensors and in biofuel cells (Pricelius *et al.*, 2009), and this is the only protein that has been classified so far from this organism. Thermophilic fungi are especially interesting for the industry because pretreatment of biomass usually occur at high temperatures. However, laboratory studies of thermophilic microorganisms are difficult and few genomes are sequenced (Rubin, 2008). Three AA9 genes from *M. thermophilum* were chosen for characterization in this study, called *Mt*358, *Mt*4260, and *Mt*6403. The *N. crassa* LPMO *Nc*LPMO9C will be used for comparisons, a C4 oxidizing LPMO with activity on β-(1→4) linked glucose units in crystalline cellulose, hemicellulose and soluble substrates (Agger *et al.*, 2014; Isaksen *et al.*, 2013).

In this study the three *M. thermophilum* LPMOs will be expressed, purified and characterized to obtain important information about their sequence, structure, phylogeny, and activity. In the laboratory activity studies will try to determine substrate specificities, regioselectivities and reaction kinetics. Supplemental to the laboratory work are the bioinformatics studies with computers that are essential for everything from aiding in the molecular biology of expressing the proteins to the study of protein sequences, structures and functions.

2. Materials and methods

2.1 Materials

2.1.1 Carbohydrate substrates

Table 2.1: Carbohydrate substrates

Substrate	Origin	Characteristics	Supplier
α-chitin	Crab shell	N-acetyl-D-glucosamine(1-4)	Y. Nagakawa
β-chitin	Squid pen	N-acetyl-D-glucosamine(1-4)	Y. Nagakawa
β-glucan	Barley	β-D-glucan (1-3, 1-4)	Megazyme
Avicel®PH-101		Microcrystalline cellulose	Fluka
Cellulose monoacetate		Acetate ester of cellulose	Bjørge Westereng
Galactomannan	Guar	β-D-mannopyranose(1-4) with α-D-galactopyranose (1-6)	Megazyme
Glucomannan	Konjac	β-D-mannopyranose(1-4) with β-D-glucose(1-6)	Megazyme
Lichenan	Icelandic moss	β-D-glucan(1-3, 1-4)	Megazyme
PASC	Avicel	Phosphoric acid swollen cellulose	
Xylan	Aspen	β-D-xylopyranose(1-4)	Bjørge Westereng
Xylan	Birchwood	β-D-xylopyranose(1-4)	Roth, Karlsruhe, Germany
Xyloglucan	Tamarind seed	β-(1-4)-D-glucan, with substitutions of xylose, arabinose and galactose	Megazyme

2.1.2 Chemical list

Table 2.2: Chemicals

Chemical	Supplier
Acetic acid (glacial) (CH_3COOH)	Merck
Acetonitrile HiPerSolv Chromanorm (CH_3CN)	VWR
Agar-agar	Merck
Agarose, SeaKem®LE	Lonza
Ampicillin	Sigma-Aldrich
Ascorbic acid	Sigma-Aldrich
Bacto™Peptone	Becton, Dickinson and Company
Bacto™Yeast Extract	Becton, Dickinson and Company
Bacto™Tryptone	Becton, Dickinson and Company
BBL™Trypticone™Peptone	Becton, Dickinson and Company
Bis-Tris ($C_8H_{19}NO_5$)	Sigma-Aldrich
Brain Heart Infusion	Oxoid
Copper (II) sulfate pentahydrate ($CuSO_4 \cdot {}_5H_2O$)	Merck
2,5-Dihydroxybenzoic acid ($C_7H_6O_4$)	Sigma-Aldrich
3,5-Dinitrosalicylic acid ($C_7H_4N_2O_7$)	Sigma-Aldrich
D-Sorbitol	Sigma-Aldrich
EDTA (Titriplex®III) di sodium salt ($C_{10}H_{14}N_2Na_2O_8 \cdot 2\,H_2O$)	Merck
Ethanol absolute AnalaR	VWR
Glucose	VWR
Glycerol 85 %	Merck
Hydrochloric acid (HCl)	Merck
Magnesium sulfate ($MgSO_4$)	Sigma-Aldrich
MES hydrate ($C_6H_{13}NO_4S \cdot xH_2O$	Sigma-Aldrich
peqGreen	VWR
Potassium Chloride (KCl)	Merck
Sodium Acetate anhydrous (CH_3COONa)	Sigma-Aldrich
Sodium Chloride (NaCl)	Merck
Sodium Hydroxide Solution (NaOH) 50 %	Sigma-Aldrich
Trizma®base	Sigma-Aldrich
Zeocin™	Invitrogen

2.1.3 Microbial strains and plasmids

Table 2.3: Microbial strains and plasmids

Strain/plasmid	Source
Escherichia coli One Shot®TOP10, chemically competent	Invitrogen
PichiaPink™strain 4	Invitrogen
pPink-GAP-HC	Invitrogen
pPICZ B	Thermo Scientific

2.1.4 Primers

Table 2.4: Primers

Primer name	Sequence (5'-> 3')	Description	Melting temp.,[1] °C
Mt4260 F	TTTCGAAACGGAATTCGAAAC GATGAAGCCATTCTCCTTGATC	Forward cloning primer	69.5
Mt4260 R	ATGGCsCGGCCGGTACCT TATGGCAAACACTGGGAGT	Reverse cloning primer full length	77.2
Mt4260 trunc R	ATGGCCGGCCGGTACCTC AACAAGAAATTGGTCTTGGAC	Reverse cloning primer truncated	74.7
GAP-Seq-F	GTCCCTATTTCAATCAATTGAA	Forward sequencing primer	
Seq R pPink cyc1	GCGTGAATGTAAGCGTGAC	Reverse sequencing primer	

[1] From pDRAW32

2.1.5 Software

Table 2.5: Software

Name	Use	Source
pDRAW32	Display, analysis and construction of plasmids	AcaClone software
SerialCloner	Display, analysis and construction of plasmids	Serial basics
AngularPlasmid	Designing plasmid maps	Vixis
PyMOL	Molecular graphics system, for protein structures	Schrödinger, LLC
Dionex™ Chromeleon™	Operation and analysis for HPAEC	Thermo Scientific
FlexControl	Operation program for MALDI-TOF	Bruker
FlexAnalysis	Data analysis and processing for MALDI-TOF	Bruker
ImageLab	Gel imaging and analysis	Bio-Rad

2.1.6 Kits

2.1.6.1 NucleoSpin® Plasmid (NoLid) High copy number Miniprep kit (Macherey-Nagel)

- Resuspension Buffer A1 with added RNase A
- Lysis Buffer A2, alkaline and with SDS
- Neutralization Buffer A3
- Wash Buffer A4 with added ethanol
- Wash Buffer AW, preheated
- Elution Buffer AE (containing 5 mM Tris/HCl, pH 8.5)
- NucleoSpin® Plasmid (NoLid) column
- Collection Tubes (2 mL)

With this kit plasmids can be purified fast from small scale cell cultures of up to 10 ml. The kit was used according to the manual. First the cells were harvested from the culture by centrifugation. Then they were resuspended, lysed and neutralized by buffers A1, A2 and A3 respectively. The cell material was centrifuged down and

the DNA, which stayed in the solution, was loaded to a silica membrane. The column was washed with a washing buffer before the DNA was eluted with a small amount of elution buffer containing 5 mM Tris/HCl, pH 8.5. All the centrifugations were done at 11 000 x g.

2.1.6.2 JetStar™ 2.0 Plasmid Midiprep Kit (Genomed)

- Cell Resuspending Buffer (E1) with Tris-HCl and EDTA, pH 8.0. With added RNase A to a concentration of 100 µg/ml
- Lysis Buffer (E2) with NaOH and SDS
- Precipitation Buffer (E3) with potassium acetate, pH 5.5
- Equilibration Buffer (E4) with sodium acetate, NaCl and Triton® X-100, pH 5.0
- Wash Buffer (E5) with sodium acetate and NaCl, pH 5.0
- Elution Buffer (E6) with Tris-HCl and NaCl, pH 8.5
- JetStar™ Midi column

This plasmid recovery kit is for higher volumes of cultures, from 15-25 ml. Plasmid isolation was done following the manual. Cells were harvested by centrifugation at 4000 x g for 10 minutes and resuspended, lysed and precipitated by buffers E1, E2 and E3 respectively. Precipitated cell material was removed by centrifugation at 12 000 x g for 10 minutes and the lysate was loaded to the equilibrated column. The column was an anion exchange resin used with gravity flow. After the column was washed twice, bound DNA was eluted with 5 ml of elution buffer.

2.1.6.3 Quant-iT™ dsDNA Assay Kit Broad-Range (Invitrogen)

- Quant-iT™ dsDNA BR reagent
- Quant-iT™ dsDNA BR buffer
- Assay tubes

All of the DNA concentrations was measured with this DNA quantification kit that uses a fluorophoric reagent to measure the sample DNA concentrations. The reagent was diluted 1:200 with the buffer, then the samples were diluted 1:200 (or 1:100 if small amount of DNA was present) with this solution. After an incubation time of 2 minutes in room temperature the fluorescence was measured with Qubit® Fluorometer. The signal is linear with DNA amounts from 2-1000 ng of DNA.

2.1.6.4 NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel)

- Binding Buffer NT1
- Wash Buffer NT3 (Concentrate) with added ethanol
- Warm (50 ºC) Elution Buffer
- NucleoSpin® Gel and PCR Clean-up Columns (yellow rings)
- Collection Tubes (2 ml)

This kit was used for both PCR-clean up and gel extraction, following the kit protocols.

For gel extraction 200 µl of the binding buffer with chaotropic salts was added for each 100 mg of gel, so the gel slices had to be weighed. Samples were incubated with the binding buffer at 50 ºC, then vortexed to dissolve the gel. The columns, with silica membranes, were placed in collection tubes and the samples were loaded. Solutions were run through the membranes by centrifugation. Then the column was washed twice with wash buffer before drying and elution by centrifugation. Samples were eluted in 15 µl elution buffer, which is alkaline and gives the low salt concentrations needed for elution.

For PCR-cleanup one sample volume was mixed with two volumes of binding buffer. Samples were loaded to the column and centrifuged, then wash buffer was run through the column twice with centrifugation. The silica membrane was dried by centrifugation and the samples were eluted with 15 µl elution buffer.

All centrifugations were done at 11 000 x g.

2.1.6.5 In-Fusion® HD cloning kit (Clontech)

- 5X In-Fusion HD Enzyme Premix
- Linearized vector pPink-GAP-HC (50-200 ng)
- Purified PCR fragment (1-2 µl)

The enzyme premix, linearized vector and amplified gene were mixed and the reaction adjusted to 10 µl, then incubated at 50 ºC for 15 minutes. Immediately after it was placed on ice and stored at -20 ºC until transformation.

2.1.7 Recipes for agars, media and gels

2.1.7.1 Brain Heart Infusion (BHI) medium

37 g BHI dissolved in water to a final volume of 1 l.

2.1.7.2 Luria Bertani (LB)

10 g Bacto™ Tryptone, 5 g Bacto™ Yeast extract and 10 g NaCl were mixed and dissolved in 1 l measuring cylinder, adjusted to 1 l with dH_2O and autoclaved.

Agar plates were also made, with 15 g/l agar added right before autoclaving. Then 200 µg/ml ampicillin was added when the temperature was approximately 55 ºC. The medium was distributed to petri dishes on a sterile bench and the plates were cooled down for 20 minutes, then put in a plastic bag, sealed and refrigerated.

2.1.7.3 LB Low Salt

5 g Bacto™ Tryptone, 2.5 g Bacto™ Yeast extract and 2.5 g NaCl were dissolved in approximately 400 ml dH_2O and pH adjusted to 7.5 with 1 M NaOH. The volume was adjusted to 500 ml and the medium was autoclaved.

2.1.7.4 SOC, Super Optimal Broth

2 g Bacto™ Tryptone, 0.5 g Bacto™ Yeast Extract, 0.057 g NaCl , 0.019 g KCl and 0.247 g $MgSO_4$ was dissolved in about 60 ml dH_2O. To room tempered solution, 1 ml 2 M glucose was added and this was adjusted to 100 ml with dH_2O. Then the medium was sterile filtered in 14 ml tubes.

2.1.7.5 YPD, Yeast Extract Peptone Dextrose

10 g Bacto™ Yeast extract and 20 g BBL™ Trypticone™ Peptone were dissolved in dH_2O to 900 ml and autoclaved. Then 100 ml of sterile filtered glucose was added.

2.1.7.6 YPDS, YPD Sorbitol

1 g Bacto™ Yeast extract, BBL™ Trypticone™ Peptone and 18.2 g sorbitol were dissolved in 90 ml dH_2O and autoclaved. After the solution had cooled down 10 ml of 20 % (w/v) sterile filtered glucose was added. The medium was stored at room temperature but put in refrigerator right before use.

2.1.8 Other recipes

2.1.8.1 Agarosegel 12 %

12 g agarose and 1 L 1X TAE running buffer was mixed and stirred with a magnetic stirrer. Then it was autoclaved and stored at 60 ºC

2.1.8.2 DHB matrix

4.5 mg 2,5-dihydroxybenzoic acid (DHB) was vortexed with 150 µl acetronitrile until the DHB was completely dissolved. Then 350 µl dH_2O was added and the solution was stored at 4 ºC for up to a week.

2.1.8.3 DNS reagent

1 g of 3,5-dinitrosalisylic acid (DNS) and 28,2 g Rochelle salt (Na-K-tartrate) were dissolved in 40 ml 1 M NaOH by stirring and heating to 70 ºC. Then the solution was adjusted to 100 ml with dH_2O.

2.1.8.4 Sodium hydroxide, 1 M

26.4 ml of 50 % NaOH with a density of 1.515 g/ml was diluted with water to a total of 1 L.

2.1.8.5 TAE-buffer 1X

4.85 x Trizma base, 1.14 ml 99.8 % acetic acid and 2 ml 0.5 M EDTA was mixed and dissolved in 1 l dH_2O.

2.2 Methods

2.2.1 Cloning of the genes

*Mt*4260 was the only LPMO that was cloned, expressed and purified, together with the truncated version that is denoted *Mt*4260-N. *Mt*358 and *Mt*6403 were expressed earlier and all gene sequences were codon optimized as described by Kittl *et al.* (2012). The genes were amplified by PCR and cloned into the constitutive Pichia Pink™ vector pPINK-GAP-HC (Várnai *et al.*, 2014). This vector was transformed into *E. coli* and the bacteria was allowed to grow. Then the plasmids were extracted from the cells and transformed into the yeast Pichia Pink™ for expression of the proteins.

2.2.1.1 Glycerol stocks of cultures

For storage of cells containing different constructs, 300 µl of sterile glycerol was added to 1 ml cell culture and the cultures were kept in cryo tubes at -80 ºC. When they were to be used again the frozen cells were retrieved by scraping with a toothpick, and inoculated.

2.2.1.2 Preparing the gene and plasmid for cloning

- Low salt LB liquid medium
- BHI medium
- Zeocin™ (Invitrogen)
- Ampicillin
- NucleoSpin® Plasmid (NoLid) High copy number Miniprep kit (Macherey-Nagel)
- JetStar™ 2.0 Plasmid Midiprep Kit (Genomed)
- Isopropanol

mt4260 was already cloned in the vector pPICZB in a glycerol stock of *E. coli* cells, and the vector contained a gene for resistance of the broad spectrum antibiotic Zeocin™ (Invitrogen). The bacteria was grown in low salt LB liquid medium with 25 µg/ml Zeocin™ over night (ON) at 37 ºC and the plasmid was purified from 10 ml cell culture with NucleoSpin® Miniprep kit.

The pPINK-GAP-HC vector was extracted from a 25 ml 37 ºC ON *E. coli* culture in BHI medium containing 50 µg/ml ampicillin by purification with the JetStar™ Midiprep Kit from. After purification, 3.5 ml of isopropanol was added to the 5 ml eluate and this was centrifuged at 12 000 x g for 30 minutes at 4 ºC. The supernatant was removed carefully, 500 µl cold ethanol was added, another centrifugation was done, and the second pellet was air dried and resuspended in 50 µl dH_2O.

2.2.1.3 Designing primers

For PCR-amplification of the gene to be cloned, primers (short DNA strands complementary to the 3' ends of each strand of the gene of interest), were designed with the software pDRAW32. The Kozak sequence used was GAAACG as described in the PichiaPink™ user manual (Invitrogen). Restriction enzymes were chosen so they would not cut in the gene and only once in the plasmid. EcoRI and Acc65I were used and the restriction sites were put in before the Kozak sequence and after the gene, respectively. Then the primers were designed using the specifications in the In-Fusion® cloning manual. The 5' end of the primer had 15 bases specific to the vector, and the 3' end had 18-25 bases specific to the gene. Primers were ordered from GenScript.

2.2.1.4 PCR with Q5 High-Fidelity 2X Master mix

- Q5® High-Fidelity 2X Master Mix (New England BioLabs), 25 µl
- Forward Primer, 2.5 µl
- Reverse Primer, 2.5 µl
- DNA template, 1 µl
- dH_2O, 19 µl
- Mastercycler gradient 120V (Eppendorf)
- 200 µl PCR tubes

To isolate and amplify the gene of interest from the vector, polymerase chain reaction (PCR) was run. A PCR reaction mixture consists of the gene of interest, forward and reverse primers, a thermostable DNA polymerase, dNTPs, cofactor Mg^{2+} and buffers. The reaction mixture is heated so the DNA melts, making both strands available. The primers will pair up with their complementary region at a specific annealing temperature. Then the reaction is heated to the optimum temperature for the DNA polymerase, and this enzyme amplifies the DNA between the forward and reverse primer by using dNTPs in the mixture. This is run in several cycles, normally around 30, amplifying the gene exponentially. PCR was run with Q5 High-Fidelity DNA Polymerase (New England BioLabs) to amplify the genes. This polymerase copies the DNA with low error, which is important since the amplified gene must be without mutations. The annealing temperatures used were 2 ºC lower than the melting temperatures for the primers, which were found by the DNA analysis software pDRAW32. Q5® High-Fidelity 2X Master Mix, primers and DNA template were mixed and placed in the Mastercycler gradient using the following program:

- Initial denaturation at 98 ºC, 30 seconds
- 30 cycles of denaturation at 98 ºC for 10 seconds, annealing at 53 ºC (54 ºC for *mt4260-N*) for 10 seconds and elongation at 72 ºC for 20 seconds
- Final elongation at 72 ºC for 2 minutes

Since the two PCR reaction mixtures were for full length and truncated proteins of the same gene, the only thing that differed was that the reverse primer annealed to the truncation site for *mt4260-N*. The annealing temperatures were slightly different, 53 ºC for the full length gene and 54 ºC for the truncated. 36 ng of plasmid was added to a total of 50 µL of reaction volume. After the PCR-reaction agarose gel electrophoresis was used to isolate the gene from the rest of the PCR-reagents, see section 2.2.1.6.

2.2.1.5 Restriction cutting

- Restriction enzymes from Thermo Scientific FastDigest:
  - EcoRI, 2.5 µl
  - Acc65I, 2.5 µl
- FastDigest Green Buffer, 5 µl (Thermo Scientific)
- Isolated plasmid with a concentration of 0.78 ng/µL, 4µL
- dH_2O, up to 50 µl

The vector the genes were to be inserted into, pPink-GAP-HC, was cut with sequence-specific enzymes (restriction endonucleases) to remove the gene already inserted and to linearize it for the cloning of the gene of interest. To find the restriction enzymes needed to cut out the old insert, EcoRI and Acc651, pDRAW32 was used. These enzymes needed to incubate at 37 ºC for one hour to cut the plasmid. The mix was vortexed and incubated on water bath at 37 ºC for one hour, then put on ice before the plasmid was isolated with agarose gel electrophoresis.

2.2.1.6 Agarose gel electrophoresis and DNA-purification from gel

- Agarose, 1.2 %
- peqGreen (peqlab)
- 1X TAE running buffer
- Quick Load® 1 kb DNA ladder (New England BioLabs)
- FastDigest Green Buffer (Thermo Scientific)
- NucleoSpin® Gel and PCR Clean-up kit

Gel electrophoresis is a separation technique to separate molecules according to their size and charge by placing them in a gel in an electric current. DNA molecules have a negative charge proportional to their mass, and will migrate towards the anode and be separated according to their size. Their runs in the gel can be compared with a ladder with known sizes to determine approximate lengths. Agarose gel is the most common gel used for separation of DNA and it contains a network with pores that the DNA-molecules has to cross. The smaller the molecules the easier they can go through these pores, so the smallest molecules will wander the furthest. DNA conformation

will also affect the distance it can travel through the gel, so it is important that all DNA-fragments are linearized.

1.2 % agarose gel was used with the dye peqGreen, a non toxic and sensitive alternative to ethidium bromide. 60 ml of hot, liquid agarose was mixed with 3 µL of peqGreen dye and poured into a mold with a well comb to set. When solid the gels were placed in the electrophoresis chamber and covered with TAE buffer. Green buffer was added to the samples and they were loaded in the wells together with the ladder. Gels were run on 90 V for about 30 minutes and visualized by UV-light with Gel Doc™ (Bio-Rad).

After the run the plasmids and genes were cut out of the gel and cleaned with NucleoSpin® Gel and PCR Clean-up kit if the DNA was to be used further. Purified samples were eluted with 15 µL elution buffer and their DNA concentrations were measured with Quant-iT™ dsDNA assay kit.

2.2.1.7 In-Fusion® cloning of the gene into the plasmid

The gene was cloned into the plasmid using the In-Fusion® HD cloning kit from Clontech, see 2.1.6.5. The In-Fusion® enzyme recognizes a 15 bp overlap between the vector and the gene that is made by primer design (see Section 2.2.1.3). Figure 2.1 gives an overview of the whole procedure, from restriction cutting and primer design to the cloning reaction and transformation.

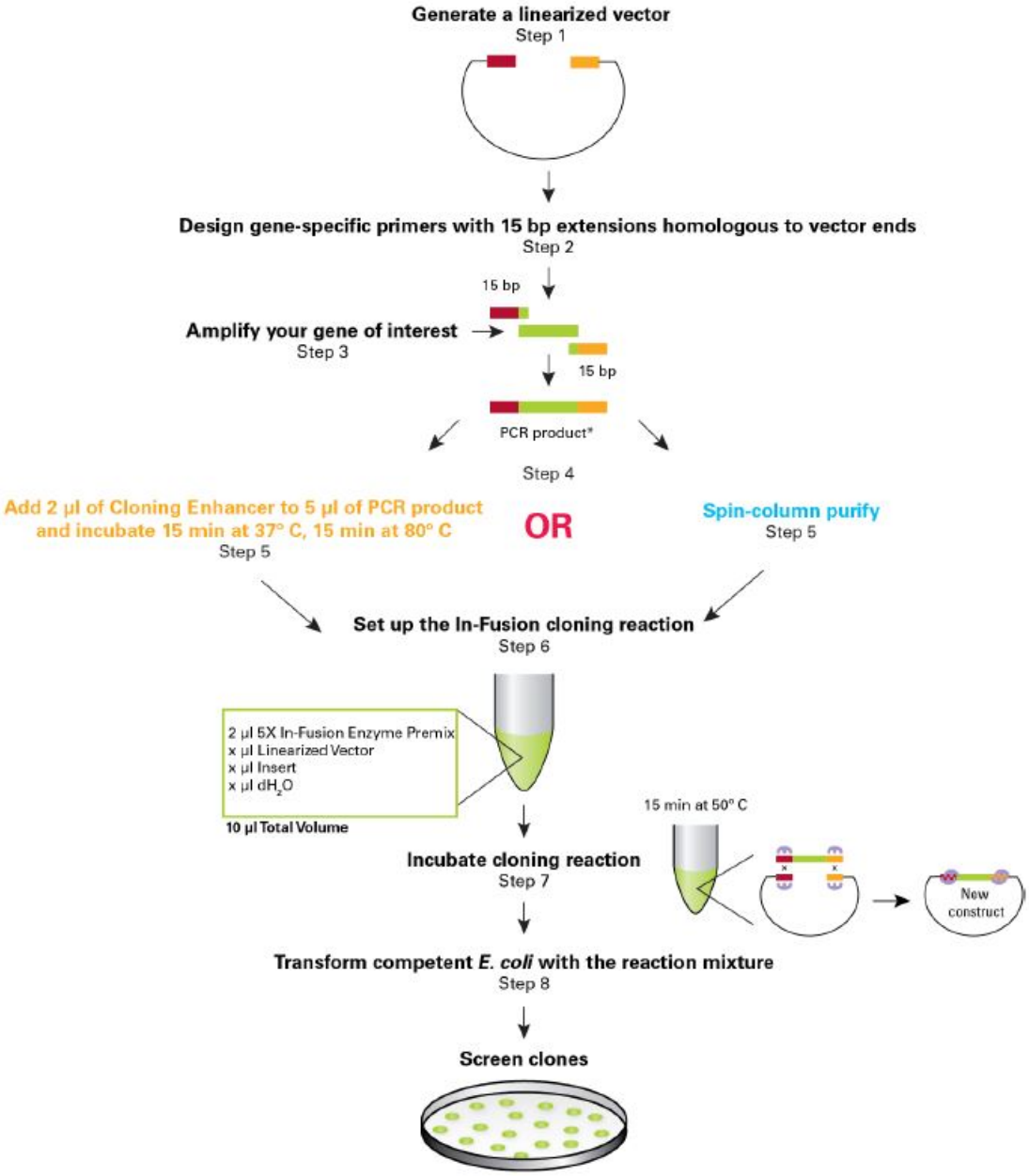

Figure 2.1: The cloning of the construct into *E. coli* using the InFusion® cloning procedure. Taken from InFusion® HD Cloning Kit User Manual (Clontech® Laboratories Inc, 2014).

To calculate the amounts of vector and insert for the reaction the In-Fusion® calculator at http://bioinfo.clontech.com/infusion/molarRatio.do was used.

2.2.1.8 Transformation of plasmid into *E.coli*

- SOC-medium
- LB agar plates
- LB with 200 µg/ml ampicillin

- In-Fusion® reaction, containing vector with desired gene
- Chemically competent TOP10 *E. coli* cells

Cells that are used in transformation must be treated so they will take up the plasmids. The treated cells are called competent cells and they can be either chemically competent or electrocompetent. Cells that are treated with calsium chloride and other salts obtain holes in the cell membrane, making it possible for the plasmid to enter. These chemically competent cells are cheaper and easier to work with than the electrocompetent cells, treated with an electric pulse to disrupt the cell membrane, but the electrocompetent cells have a higher transformation efficiency.

Reactions with 100 µL chemically competent cells and 10 µL In-Fusion® reaction were mixed carefully and placed on ice for 30 minutes. Directly after they were placed on 42 ºC water bath for exactly 1 minute to give the cells a heat shock. Then the tubes were cooled down for approximately 2 minutes on ice, and 230 µL of SOC medium was added to each tube. The tubes were incubated with shaking for at least one hour at 37 ºC. Then they were plated out on LB agar plates and incubated over night at 37 ºC.

Seven colonies were picked from the plates and incubated with shaking in 50 ml LB with 200 µg/ml ampicillin overnight.

2.2.1.9 Colony PCR with Taq DNA Polymerase

- Taq DNA Polymerase Master Mix, 25 µl (VWR)
- Forward & reverse primer, 1 µl
- PCR tubes smeared with the colonies

Colony PCR was run on the transformed colonies to see if the transformation had been successful. Since high accuracy was not required the polymerase used was Taq DNA polymerase (VWR) instead of the Q5 polymerase used for the amplification of the gene. DNA polymerase and primers were mixed and the volume adjusted to 50 µl with dH_2O. A positive control was also made using the original vector with the gene instead of the colonies. Then PCR was run with the following program:

- Initial denaturation at 94 ºC, 2 minutes
- 30 cycles of denaturation at 94 ºC for 20 seconds, annealing at 53 ºC (54 ºC for *Mt*4260-N) for 25 seconds and elongation at 72 ºC for 30 seconds
- Final elongation at 72 ºC for 5 minutes

PCR reactions were checked on agarose gel and glycerol stocks were made of two successfully transformed colonies. One of these colonies was incubated in an Erlenmeyer flask with 25 ml LB-medium and 200 µg/ml ampicillin at 37 ºC with shaking ON.

2.2.1.10 Purification and linearisation of plasmid before transformation in PichiaPink™

- NEBuffer 4 (10X), 4 µl (New England Biolabs)
- BSA (10X), 5 µl
- AflII restriction enzyme, 4 µl (New England Biolabs)

The transforned 25 ml *E.coli* culture was purified with JetStar™ 2.0 Plasmid Midiprep Kit (Genomed) and precipitated with isopropanol as described in 2.2.1.2 before the plasmid concentration was measured fluorometrically. Then the plasmids were linearized by cutting with the restriction enzyme AflII that cuts once in the TRP2 region of the plasmid and not in the gene. The reaction was mixed and adjusted to 50 µl with dH_2O in an eppendorf tube, then placed on a 37 ºC water bath for 2.5 hours.

After linearization the plasmid was cleaned for restriction components by the NucleoSpin® Gel and PCR Clean-up kit with the PCR clean-up protocol. The concentration of the cleaned, linearized plasmid was measured, then the plasmid was compared with non-linearized plasmid to check for correct restriction cutting on an agarose gel.

2.2.1.11 Sequencing

- Sequencing primers:
  - GAP-Seq-F (50 µM) A-59, 0.5 µl
  - Seq R pPink cyc1, 0.5 µl
  - Forward primer for the gene (10 pmol/µl), 2.5 µl
  - Reverse primer for the gene (10 pmol/µl), 2.5 µl
- Plasmid with insert, 2 µl

To make sure that there were no mutations in the genes before transformation into PichiaPink™ cells the genes were sequenced, that is, the order of the bases in the gene was determined. The plasmid was mixed with the four different primers in four reactions and the volumes were adjusted to 11 µl with dH_2O. Then the vials were sent by mail to GATC biotech (Germany) where the sequencing was done using Sanger sequencing and their LIGHTrun™ sequencing service. Sequences were analyzed by Geir Mathiesen.

2.2.1.12 Transformation of PichiaPink™ cells by electroporation

- YPD medium
- PichiaPink™ cells strain 4
- Linearized plasmid
- Sorbitol, 1 M
- 0.2 cm electroporation cuvettes (GenePulser®)
- Sterile shake flasks and centrifugation tubes
- YPDS medium
- PAD (Pichia Adenine Dropout) agar plates (Life Technologies)
- GenePulser® II (Bio-Rad), Pulse controller PLUS (Bio-Rad)

P. pastoris is a well-known yeast that is commonly used as an expression system of eukaryote proteins (Daly & Hearn, 2005). Even though it makes post translational modifications of fungal enzymes, the methylation of the N-terminal histidine in AA9s described in Section 1.3.5 is not observed. PichiaPink™ is a strain of *P. pastoris* with modifications to facilitate protein expression that was used to express the two proteins. An effective selection system selects recombinants based on the red color of purine precursors. Mutation of the ADE2 gene, important for biosynthesis of purine nucleotides, render the cells unable to grow on adenine lacking medium unless they are successfully transformed. The color of the transformants can be used to indicate the relative expression levels of the protein of interest as the white colonies contain higher copy numbers of the plasmid with the ADE2 gene and the gene of interest than the red colonies (Invitrogen, 2010).

Linearized plasmids were transformed into the PichiaPink™ cells. The incorporation of plasmids into the yeast's chromosomal DNA is done at the TRP2 region, which is complementary to the same region in the plasmid. No antibiotics were used for selection, so the work was done as sterile as possible to avoid contamination. To prepare the cells they were incubated in an Erlenmeyer flask with 10 ml YPD medium at 30 ºC and shaking for two days. This starter culture was then added to a total of 100 ml YPD media in a sterile 1 liter flask and grown until the OD was 1.3-1.5, about 12 hours. At the correct OD the culture was transferred to a 300 ml centrifuge bottle and centrifuged at 1500 g at 4 ºC for 5 minutes. The pellet was re-suspended with 250 ml of ice cold, sterile water and centrifuged again in the same way. Next, 50 ml of ice

cold, sterile water was added before a third centrifugation. Finally 10 ml of ice cold sorbitol was added and the cells were centrifuged for a fourth time, then 300 µL of ice cold sorbitol was added to a final volume of approximately 500 µL. The cells were stored on ice and electroporated the same day.

To make the prepared cells electrocompetent transformation was done with an electroporator. 80 µL of the PichiaPink™ cells were mixed with 5-10 µg of the linearized plasmid DNA, and this was transferred to an ice-cold 0.2 cm electroporation cuvette and incubated for 5 minutes on ice. The cells were pulsed in the electroporator at 1.51 kV, 25 µF and 200 Ω for 5 ms. 1 ml of ice cold YPDS medium was added immediately after pulsing and this was mixed carefully by pipetting up and down. The cells were incubated for 2-12 hours without shaking. 150 µL of the cell mixture were spread out on three PAD selection plates and incubated at 30 ºC for about a week until colonies appeared. Controls with untransformed cells were also spread on plates.

White colonies were streaked out on new PAD-plates and incubated at 30 ºC until growth. Plates were stored at 4 ºC for maximum a week until expression.

2.2.2 Expressing and purifying the proteins

*Mt*4260 and *Mt*4260-N were expressed in 1.5 l cultures with the PichiaPink™ system after initial tests and purified with a membrane system, centrifugation filters and chromatographic methods. Before injection on the chromatographic columns, all samples were centrifuged to make sure no particles would clog the columns. Protein content of samples from different stages was analyzed using LDS-PAGE protein gels. After purification procedures the concentrations of the samples were measured with UV detection at 280 nm, see Section 2.2.3.1.

2.2.2.1 Testing and analyzing protein expression

- YPD medium

Initial tests and glycerol stock preparations were done with 3-5 white colonies from the PAD plates that were inoculated in 10 ml YPD medium and grown for two days with shaking at 30 ºC. Glycerol stocks were made of the grown colonies, and the cultures were centrifuged at 1500 g for 5 minutes before supernatants were checked for protein expression with LDS-PAGE (see Section 2.2.2.2). When protein was detected 3-5 10 ml ON cultures were started, inoculated from the glycerol stocks. These cultures were added to 50 ml YPD medium and incubated with shaking at 30 ºC. Samples were collected from these cultures at 24 hours intervals and centrifuged at 1500 g for 5 minutes. Both supernatants and pellets were kept, but in different tubes, and frozen

until analysis. OD was measured at each time point. At the end the pellet samples were dissolved in dH_2O and all the samples were run on LDS-PAGE gel.

2.2.2.2 Lithium dodecyl sulfate polyacrylamide gel electropohoresis (LDS-PAGE)

- NuPAGE® LDS sample buffer 4x (Invitrogen)
- NuPAGE® Sample reducing agent 10x (Invitrogen)
- Mini-PROTEAN® TGX Stain-Free™ gel (Bio Rad)
- Tris/Glycine/SDS (TGS) electrophoresis buffer, 10x (Bio Rad)
- BenchMark™ protein ladder (Invitrogen)

Protein expression was analyzed to see if the desired protein was expressed by using lithium dodecyl sulfate polyacrylamide gel electrophoresis (LDS-PAGE). Proteins are denatured by lithium dodecyl sulfate (LDS) and a reducing agent that reduces disulphide bridges. The anionic LDS reagent will bind to denatured polypeptide chains, giving them a negative charge that depends on their mass, so the charge-to-mass ratio is almost equal for all the proteins. Sample proteins will wander through a polyacrylamide gel placed in an electric field, and the larger proteins will move slower than the smaller because they are more retained. By comparing with a protein ladder the molecular size of the proteins can be measured.

A Mini-PROTEAN® TGX Stain-Free™ gel with trihalo compounds that react with tryptophans in the proteins when induced with UV-light was used. The LDS reagent was made mixing LDS Sample Buffer (4x), Reducing Agent (10x) and dH_2O in small volumes to 2x the reagents and an equal amount of 2x LDS reagent and sample was mixed before gel application. Samples were mixed in eppendorf tubes with the 2x LDS reagent, boiled for ten minutes, centrifuged and applied on the stain free gel together with 5 µL BenchMark™ ladder. Then LDS-PAGE was run for 17 minutes and 300 V.

2.2.2.3 Large scale protein expression

- YPD medium
- Antifoam 204 (Sigma-Aldrich)
- 20 % Glucose, wt./V
- Baffled shake flasks

- Harbinger LEX bubbling system

When the tests showed expression of the desired proteins, cultures were scaled up and the Harbinger LEX bubbling system or baffled shake flasks were used. In the LEX system, 500 ml cell cultures were grown in 1 liter flasks in a water bath with a regulated temperature. A continuous flow of air was directed into the flasks through a filtered air inlet ending in a porous stone hanging into the culture from the stopper of the flask. Conditions for growth were therefore good and the cultures grew quickly. The PichiaPink™ cultures were grown at 28 ºC for about 24 hours. 150 µl antifoam was added to prevent large foam formations.

The cultures were also grown in shake flasks. This took longer time but was a more known method and was done because some problems occurred with expression with the LEX system. 300 ml cultures were incubated at 28 ºC and 200 rpm shaking for 72 hours. Samples were taken out every day and OD was measured, then supernatants were run on gel.

Sterile glucose was added to the bottles and shake flasks at 24 hour intervals up to a concentration of 1 %.

Centrifugation of the cultures was done at 8700 rpm for 12 minutes before the supernatants were centrifuged at 75465 x g for 15 minutes. Before purification, supernatants were sterile filtrated.

2.2.2.4 Increasing protein concentration

- Vivaflow 200, 10,000 MWCO polyetersulfone, crossflow cassette (Sartorius) with pump and tubing
- 25 µM sodium acetate buffer pH 5.0
- Amicon® Ultra - 15, Ultracel® 10K
- Macrosep® Advance Centrifugal Devices, Omega™ 10K

Vivaflow 200 crossflow cassette was used to concentrate and desalt protein solutions, and to change buffer. A pump connected to the filter system pumps the sample over the polyethersulfone membrane with 10 kDa pores and the smaller molecules pass through the membrane and go to waste, while the larger go back into the sample reservoir. To desalt the solution and to change buffer into the one used in further chromatographic purification steps, the sterile filtrated supernatants were placed in a sealed container connected to a buffer reservoir. Small molecules were removed from the solutions through filtration and a vacuum was created in the container, sucking in buffer to replace the missing liquid Sartorius Stedim Biotech (2014).

When volumes of the protein solutions were small enough, centrifugal tubes with filters were used for further concentration and buffer change. The filters had pore sizes of 10 kDa retaining molecules larger than this in the filter when centrifuged. Two types of tubes were used: Amicon® Ultra - 15, Ultracel® 10K with cellulosic filter and Macrosep® Advance Centrifugal Devices, Omega™ 10K with an Omega™ membrane. The latter was used when there was a possibility of cellulases in the solutions to be concentrated, as these could break down the cellulose membrane. The tubes were centrifuged at 4500 x g for as long as it took to reduce the volume to a few milliliters for purification.

2.2.2.5 Precipitation with ammonium sulfate

- 4 M ammonium sulfate

The protein solutions were precipitated with up to 1 M of ammonium sulfate and centrifuged, so that some of the unwanted proteins would precipitate while the desired protein would stay in solution. Ammonium sulfate was added carefully drop by drop under constant stirring to avoid unwanted precipitation due to high local salt concentrations. An anti-chaotropic salt like ammonium sulfate is a salt that promotes order in a solution, which means that proteins can easier form hydrophobic interactions between themselves that can be so strong that they precipitate. To check that the wanted proteins did not precipitate SDS-gels were run with both the supernatants and the pellets.

2.2.2.6 Hydrophobic Interaction Chromatography

- HiTrap™ phenyl FF column (GE Healthcare) used with ÄKTAdesign™ chromatographic system
- 1 M ammonium sulfate in 25 mM sodium acetate pH 5.0
- 25 mM sodium acetate pH 5.0
- 1 M NaOH

The ammonium sulfate precipitation also prepared the solutions for hydrophobic interaction chromatography (HIC). In HIC, proteins are separated based on their hydrophobic interactions with the column and this is affected by the mobile phase buffer. Anti-chaotropic salts are used to ensure that the protein of interest binds to the column and is not washed out in the flow-through. It is crucial to add just the right amount of salt so that the protein of interest do not precipitate or denature, while at

the same time it binds to the column. This amount has to be decided experimentally for each protein. After the sample is run through the column and some of the proteins are bound to the stationary phase, the running buffer is gradually changed from one containing ammonium sulfate to one that does not. Proteins bound to the column will elute at different points in the gradient elution according to their hydrophobicity, with the most hydrophobic proteins eluting at the end. Very hydrophobic proteins might not elute at all and must be washed out of the column with sodium hydroxide (GE Healthcare, 2006).

In this experiment, the column was a HiTrap™ Phenyl FF (fast flow, as opposed to HP - high performance) High Sub with a column volume (CV) of 5 ml. High sub means that there is a high amount of substituted phenyl groups, making the interactions between the column media and the proteins strong. "Fast flow" is because the particles are relatively large, allowing higher speed because of reduced pressure but at the expense of the resolution. The start buffer was 25 mM sodium acetate pH 5.0 with 1 M ammonium sulfate and the elution buffer was the same without ammonium sulfate. ÄKTAdesign™ was the cromatographic system used to control elution and fraction collection (GE Healthcare, 2006).

Before sample application, the column was washed with start buffer and elution buffer at 1 ml/min to remove preservatives or contaminations from previous use, and then it was equilibrated with the starting buffer by running it through the system until UV and conductivity were stable. Sample loops and pumps were also washed with buffer. Then the protein solution was loaded in the sample loop (2 ml loops were used) with a syringe and injected onto the column, 1.5 ml at a time. The flow rate of injection was 0.5-1 ml/min and fractions of 2 ml were collected. Start buffer was sent through the column until the UV and conductivity signals were stable, then the buffer was gradually changed over 5 CV from 100 to 0 % start buffer. Samples were taken of the fractions where a UV peak was observed and the fractions were run on a LDS-gel to check for the desired protein.

For *Mt*4260-N a stepwise elution was developed to save time and buffer. The final, most successful elution only needed two steps, one with 20 % start buffer where the protein eluted and one with 0 % start buffer to remove the rest of the bound proteins.

Fractions appearing to contain relatively pure proteins were combined and concentrated, then buffer was changed with centrifugal tubes to 20 mM Tris HCl pH 8.0 before the next purification step, ion exchange chromatography.

2.2.2.7 Ion Exchange Chromatography

- HiTrap™ Q Sepharose FF column (GE Healthcare) with ÄKTAdesign™ chromatographic system
- 20 mM Tris pH 8.0, start buffer
- 1 M NaCl in 20 mM Tris pH 8.0, elution buffer

Ion exchange chromatography (IEC) was used as a polishing step of the sample with *Mt*4260 after HIC. This type of chromatography uses reversible ionic bonds to attach the target molecules to charged molecules called ion exchangers that are fixed to the stationary phase of the column. The stationary phase usually consists of spherical, porous particles with a high surface area. When the ion exchangers are cationic and the molecules they attract are anions the system is called anion exchange. Proteins will carry a net negative charge when the pH in the solution is above their pI and the separation is done based on the molecules' net surface charge. The start buffer, in this case Tris HCl, has a low ionic strength, and the anionic target protein will bind to the ion exchangers. Elution is done with a buffer with higher ionic strength due to added NaCl, and the Cl^- ions in this buffer will bind to the ion exchangers instead of the target protein. The higher the net charge of the proteins is the more strongly they will bind to the ion exchangers, so in this case the least negative proteins will elute first. By careful optimization of conditions like pH and ionic strength in the buffers it is possible to separate very similar proteins based on their charge properties (GE Healthcare, 2010).

The column was a HiTrap™ Q Sepharose FF column with a CV of 5 ml, with substitutions of a strong anionic exchanger of quaternary ammonium. A strong ion exchanger do not change its binding properties at differing pH, which gives an easier method development than using a weak one (GE Healthcare, 2010). As start buffer 20 mM Tris pH 8.0 was used and the elution buffer was the same but with 1 M NaCl.

*Mt*4260 have a theoretical pI of about 5.2, see Table 3.1. Since the buffers had a pH of 8.0, the protein had a negative surface charge and would bind to the ion exchangers. First the column was washed with start buffer and elution buffer at 5 ml/min. Then it was equilibrated with start buffer to make sure that all the charged groups attached to the stationary phase was bound to counter ions. Sample was applied to the equilibrated column at 1 ml/min in fractions of 0.5-1.5 ml. This was washed through until the UV and conductivity signals were stable, then elution started with fractionation of 2 ml. The column was regenerated by washing with both buffers between sample applications. A 3 CV gradient elution from 0-100 % elution buffer was done first with some of the sample, then a stepwise elution was done based on the results from this

first run with the rest of the sample. The most successful stepwise elution had a 6 % elution buffer concentration, followed by 20 % and 100 %. Elution at 20 % gave less pure substances than at 6 %, but much larger amounts of the protein eluted. IEC did not yield a significantly purer protein solution than HIC, and was skipped for *Mt*4260-N.

Optimization of elution buffer concentration steps and sample preparations were done with both HIC and IEC to make sure the proteins were as pure as possible. The samples were applied in small fractions and each elution was analyzed with UV and LDS-PAGE before subsequent injections.

Both columns were cleaned and stored in accordance with their protocols.

2.2.3 Enzyme characterization

2.2.3.1 Measuring protein concentrations

- Spectrophotometer with UV lamp, Ultraspec 2100 pro specrophotometer (GE Healthcare)
- Disposable cuvettes, 1.5 ml (Brand)
- Sample buffers

Protein concentrations were mostly measured spectrophotometrically by the absorption of light at 280 nm (A280) by aromatic amino acids. The amount of light absorbed by the sample is measured, and because of the linear relationship (from 0.1 to 1 absorbance units) between absorbance and protein concentration the concentration can be calculated. This is a quick, but not very accurate method of concentration measurement as different characteristics of proteins also affect the absorbance.

Measurements were done by first calibrating to zero with a blank containing the components of the sample except the protein. Then the samples were measured, making sure the absorbance signal was within the linear range. Protein concentrations were estimated by using the formula $c = \frac{A_{280}}{\varepsilon_{280} \cdot l}$ where c is protein concentration, $A_{280}$is absorbance at 280 nm, $\varepsilon_{280}$is the absorbance coefficient at 280 nm and l is the length of the cuvette (normally 1 cm). This gives the concentration in moles per liter. To get the concentration in g/l the molar concentration is multiplied by the molar mass of the protein. Molar masses and absorbance coefficients are listed in Table 3.1.

2.2.3.2 Copper saturation

- 1 M $CuSO_4$

- PD MidiTrap™ G-25 prepacked columns (GE Healthcare)
- 20 mM MES buffer

LPMOs need copper in their active site to work properly, so the purified enzymes were saturated with copper in the form of $CuSO_4$. To remove the unbound copper gel filtration with Sephadex™ G-25 medium was used. Molecules too large to enter the pores in this gel matrix like proteins will elute first, while smaller molecules like copper ions will penetrate the pores and use longer time to pass the column.

The column was equilibrated with 20 mM MES buffer, which was the buffer used later in the reactions. This was done by running 5 ml of the buffer through the column three times. The column always retain approximately 1 ml of liquid, so after the equilibration the column was filled with buffer. Samples were upconcentrated to about 200 µl and the protein concentrations were measured. $CuSO_4$ was added to an amount five times that of the enzyme and the samples were mixed, then incubated in room temperature for 30 minutes. After incubation they were run through the equilibrated column. A maximum of 300 µl sample was applied, then equilibration buffer was added up to 1 ml. This was allowed to enter the column, and flow through was discarded before 1 ml of buffer was added. A test tube was placed under the column to collect the largest sample molecules. To elute the rest of the protein and all the copper, another 500 µl buffer was added. This elution was stored for later purification. Protein concentration in the first elution was measured and this enzyme solution was used in further experiments.

2.2.3.3 Substrate degradation

- Soluble and insoluble substrates, see Table 2.1, at a concentration of 1 mg/ml
- Enzymes:
  - *Mt*358
  - *Mt*4260 and *Mt*4260-N
  - *Mt*6403
- Buffers (reactions concentrations)
  - Sodium acetate pH 5.6, 0.1 M
  - MES pH 6.5, 0.5 M
- Ascorbic acid, 0.1 M

The products of cellulose cleavage by LPMOs are aldonic acids for C1 oxidation and 4-ketoaldoses for C4 oxidation. confirmed by chromatography, mass spectrometry and NMR (Forsberg *et al.*, 2011; Isaksen *et al.*, 2013). When cellulose is degraded by LPMOs, each cleavage site gives rise to one oxidized and one native end. Comparisons of the products with standards of native and oxidized oligosaccharides and their retention times in a HPLC column makes it possible to identify the products formed. For C4 oxidized products no standards are available. Thus, mass spectrometry and NMR are important for confirming the nature of these products.

To map substrate specificities and characterize the type of products created by the three enzymes used in this study, reactions were set up with different soluble and insoluble substrates and analyzed by high-performance anion exchange chromatography (HPAEC) or matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS).

Reactions were made with enzyme concentrations of either 30 mg/g substrate or 2 µM, substrate concentrations of 1 mg/ml, ascorbic acid concentrations of 2 mM and buffer concentrations as listed above. To check for possible background activity of other enzymes, identical reactions were made but without ascorbic acid as LPMOs require reducing agents to function. Control samples were also made with all the substrates without enzymes and, if needed, with only one reaction component so signals from the different chemicals could be differentiated from enzyme product signals. The reactions were usually made in volumes of 100 µL in eppendorf tubes, incubated over night or for specific time periods at 50 ºC and 1000 rpm shaking in table top incubators, and stopped by boiling for 1 minute.

2.2.3.4 High-performance anion-exchange chromatography

- Reactions from Section 2.2.3.3
- Standards: Mix of glucose, cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose with concentrations of 0.1, 0.05 and 0.025 mg/ml
- CarboPac®PA1 Carbohydrate Column (Thermo Scientific)
- HPLC vials
- Sodium Acetate, anhydrous
- NaOH, 50 %
- Milli-Q water

To analyze reactions for soluble carbohydrate products high-performance anion exchange chromatography (HPAEC) was used. The system was a Dionex ICS 3000 with an anion exchange resin and a PAD-detector (Pulsed Amperometric Detection). This was able to separate native and oxidized oligosaccharides.

For separation of oligosaccharides produces by the LPMOs a CarboPac®PA1 (2 x 50 mm) Carbohydrate Column was used, with MicroBead™pellicular structure and CarboPac PA1 guard column (2 x 50 mm). This column is made of nonporous polymers with attached MicroBeads bound to the functional groups. At alkaline conditions the hydroxyl groups of carbohydrates will be ionized and can be separated by anion exchange as described in Section 2.2.2.7. Eluents with sodium hydroxide are used in the chromatographic system to produce the anions. For native or oxidized oligosaccharides the hydroxyl groups have different pKas because they experience different chemical environments. Oligosaccharides with different degrees of polymerization and linkages also have different acidities. How easily available the sugar anions created are to the functional groups in the stationary phase also affects the retention times of the sugars. This means that it is possible to distinguish between different oxidation states and degrees of polymerization (DP) of the oligosaccharides in the samples. For instance a 4 DP sugar will have a longer retention time than a 3 DP sugar and the aldonic acids created from C1 oxidation (see Section 1.3.4) will have longer retention times than their native (non oxidized) counterparts due to the C1 carboxylic acid.

The PAD detector is an electrochemical detector with a gold electrode that oxidizes or reduces the analyte and measures the change in potential between two electrodes for the oxidation. The detector applies a pulsed sequence of potential variations for milliseconds, then another potential is pulsed to remove the analyte from the electrode surface. It is one of the most sensitive detection methods and requires only small amounts of analyte.

As eluents 0.1 M NaOH (A) and 1 M sodium acetate in 0.1 M NaOH (B) was used. The sodium acetate was filtrated to remove particles. All eluents were sonicated before use to remove dissolved gas, as carbonate in the eluents can bind to the stationary phase of the column and reduce separation. New eluents were made every 48 hours.

Elution was done with a flow rate of 0.25 ml/min for 75 minutes as follows:

1. Linear gradient of 100 % A to 10 % B for 35 minutes
2. Dionex curve 6 gradient of 10 % B to 30 % B in 25 minutes
3. Dionex curve 6 gradient of30 % B to 100 % B in 5 minutes
4. 100 % B to 100 % A in 1 minute

5. Reequilibration for 9 minutes

Reaction mixtures were centrifuged, and only the supernatants were transferred to HPLC-vials to make sure no particles were added on the column. The vials were put in the autosampler and the program was run, injecting 2-2.5 µl of sample. Subsequently, the results were analyzed with Chromeleon™ software. For identification of peaks standards were run at regular intervals in the sequence, but HPAEC was not used for quantification.

2.2.3.5 MALDI-TOF mass spectrometry

- DHB matrix (2.1.8.2)
- Reactions as made in Section 2.2.3.3
- MALDI-TOF, Ultraflex TOF-TOF (Bruker)
- Blow dryer

Samples were run through a MALDI-TOF mass spectrometer. MALDI stands for matrix-assisted laser desorption/ionization and is an ionization technique well suited for macromolecules. The molecules to be analyzed are dissolved in a matrix and dried on a plate. What type of matrix is used depends on the sample molecules and must absorb the laser wavelength strongly. A commonly used matrix for proteins and carbohydrates is 2,5-dihydroxybenzoic acid (also called gentisic acid, abbreviated DHB). In the ion source the matrix with dissolved sample is shot with short, intense laser pulses and the dry droplet is vaporized. Since it is the matrix, not the samples, that absorb the laser the wavelength does not have to change according to the sample composition, and the method is universal. The sample molecules will ionize (the exact mechanism is not fully understood) and accelerate through an electrostatic field to the analyzer.

Commonly used with the MALDI ion source is the TOF (time-of-flight) mass analyzer. It separates the molecules by their velocities in a field free tube, which in turn depends on their mass-to-charge ratios. The name time-of-flight comes from the fact that it is the time the molecules take through the flight tube that is measured when they reach the detector. This is a sensitive and fast analyzer with principally no upper limit for the size of the molecules it can separate (de Hoffman & Stroobant, 2007).

Droplets of 2 µL of DHB solution was placed on the MALDI plate, then 1 µL samples were placed directly into this droplet. After drying, the plate was loaded into the mass spectrometer and laser beams with different intensities were shot at the samples. Different mass ranges were analyzed and the best mass window for each sample

was found experimentally. At the end the results were analyzed using FlexAnalysis (Bruker).

2.2.3.6 Reducing end assay with 3,5-dinitrosalisylic acid

- DNS reagent (made as described in 2.1.8.3)
- Enzymes: *Mt*4260 and *Mt*4260-N
- PASC, Avicel and tamarind xyloglucan
- Ascorbic acid
- MES buffer
- Cellobiose
- XG7, a xyloglucan heptasaccharide

Breakdown of cellulose can be detected by monitoring new reducing ends that are formed at cleavage points. The reducing ends can reduce colorimetric reagents and make a color change that can be measured spectrophotometrically by comparing with known standards. The reagent 3,5-dinitrosalicylic acid (DNS, 2-hydroxy-3,5-dinitrobenzoic acid) contains a nitro group that can be reduced to an amino group by the hemiacetal or aldehyde group at the reducing end of the sugar. 3-amino-5-nitrosalicylic acid is formed and can absorb light strongly at 540 nm, which makes this reaction ideal for spectrophotometric measurement. To quantify the reaction, suitable standards with known concentrations are used.

All reactions were made using copper saturated *Mt*4260 and *Mt*4260-N in different concentrations, 40 mM MES buffer pH 6.5 with and without 2 mM ascorbic acid and with a substrate concentration of close to 5 mg/ml. Avicel, PASC and tamarind xyloglucan were the substrates and the reactions were incubated for two hours at 50 ºC and 1000 rpm. 40 µl samples were taken after 5, 15, 30, 60 and 120 minutes of incubation and immediately added to 60 µl of the alkaline DNS reagent, quenching the reactions. Then the samples were boiled for ten minutes to develop the color and centrifuged to remove insoluble substrates, before 50 µl of the supernatants were put in a 384-well plate for absorbance measurement at 540 nm.

For quantification of samples with PASC and Avicel, cellobiose was used. A glucose tetramer with three xyloses (XXXG, the glucose at the reducing end being the only one without xylose) were used for the xyloglucan. Four known concentrations of standards were made in triplicates to make standard curves. The concentrations in the samples were calculated using these curves and correcting for background noise by blanks with

only substrate or substrate and ascorbic acid. A blank with only water and DNS was also made. Blanks with ascorbic acid and substrate were incubated with the reactions and samples were taken at the same time points, as ascorbic acid can change the results when incubated at a time period. The blanks with substrate or water were only boiled as previous experiments show that they do not change over time.

3. Results

3.1 Bioinformatics

3.1.1 Protein parameters

Using the theoretical Expasy ProtParam tool from http://web.expasy.org/protparam/ (Gasteiger, E et al.), properties for the three proteins included in this study were calculated. The properties are shown in Table 3.1 and are based on mature proteins without signal peptides. Genes of full length and truncated *Mt*4260 are denoted *mt4260* and *mt4260-N* respectively, while the corresponding proteins will be called *Mt*4260 and *Mt*4260-N.

Table 3.1: Key properties of the four studied proteins. The properties were calculated by the ProtParam server, and N-terminal signal peptides were omitted. Number of basepairs of the respective genes is also listed.

Protein	Basepairs	Amino acids	Mol. weight, DA	pI	ε^1
Mt358	693	208	22514.8	5.37	44140
Mt4260	930	289	29775.7	5.21	45880
Mt4260-N	747	227	24039.8	5.80	39670
Mt6403	1038	319	33164.6	5.76	39670

[1] Assuming all pairs of cysteines form sulfide bonds

3.1.2 Domain structures and protein modifications

Pfam (Finn *et al.*, 2014) gives a general idea about protein domains in sequences, based on Hidden Markov Models and the CAZy database. AA9 domains were found in all three protein sequences by the Pfam server, and a C-terminal CBM1 domain was discovered in *Mt*4260. CBM1 domains contain four conserved cysteines that can form two sulfide bonds and three conserved aromatic amino acids that binds to cellulose (entry PS51164 from http://prosite.expasy.org/, (Sigrist *et al.*, 2012)). *Mt*6403 does

not have a CBM domain, instead it has an extended C-terminal region. The domain structures of the three enzymes are shown in Figure 3.1.

Proteins will be secreted from cells if they have a signal sequence targeting them to the secretory pathway. If the signal peptide native to the expressed enzyme is recognized by the expression host this will be sufficient to secrete the enzyme. SignalP (Petersen *et al.*, 2011) was used to determine the length of the native secretion signal peptides that all three proteins contain in their N-terminus. The amino acid sequences of the three proteins are shown in Appendix A together with the sequence of a fourth AA9 protein, the C4 oxidizing *Nc*LPMO9C (Isaksen *et al.*, 2013), as this protein was used for comparisons in this study. The truncated version of *Nc*LPMO9C will be denoted *Nc*LPMO9C-N.

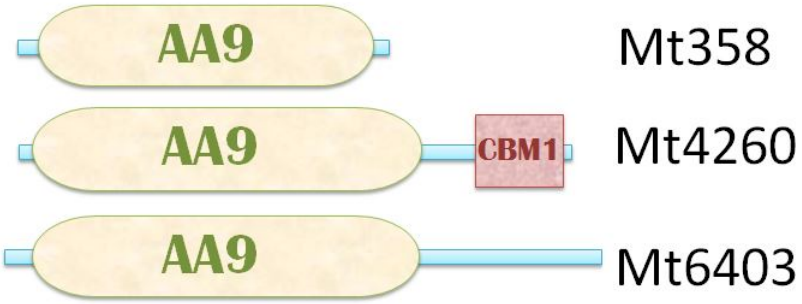

Figure 3.1: Pfam assigned modules for *Mt358*, *Mt*4260 and *Mt*6403. All three enzymes contain a short signal peptide and one AA9 domain, while *Mt*4260 has a CBM1. *Mt*6403 has a longer C-terminal region with an unknown function

Between the catalytic domain and the CBM domain in *Mt*4260 is a linker region rich in glycine, proline, serine and asparagine residues. *Nc*LPMO9C is also a two domain protein, with a similar but longer linker with a large amount of alanine, proline, serine, threonine and glycine. The long C-terminal, disordered region of *Mt*6403 is about as long as the linker of *Nc*LPMO9C and contains many threonines, alanines, serines and prolines. These are typical amino acid compositions for disordered, flexible regions.

Protein glycosylation is the most important post translational modification and exists in various forms. O-glycosylation is usually found in disordered regions of proteins and gives resistance to proteases, but it can also affect enzyme activity in ways that are so far not clear (Payne *et al.*, 2013). Five mucin type GalNAc O-glycosylation sites was found in *Mt*4260 using the server NetOGlyc (Steentoft *et al.*, 2013), all at serines in or close to the linker peptide. *Mt358* has two possible O-glycosylation sites in the AA9 domain and *Mt*6403 has 8 in the AA9 domain and 33 in its C-terminal tail. *Nc*LPMO9C has 2 O-glycosylation sites in the AA9 domain and 28 in the linker region. It also has two N-glycosylation sites. It is important to note that predictions made by the NetOGlyc server are based on mammalian proteins and might not be convertible to

the fungal proteins. There is no universal consensus motif for O-glycosylation and its regulation is complex and not yet understood. A little more is known about asparagine linked N-glycosylation, but no sites were predicted by the NetNGlyc server (Gupta *et al.*, 2004) in any of the *M. thermophilum* proteins.

3.1.3 Structural studies

A look in the CAZy database at http://www.cazy.org/ shows that six AA9 structures, eight AA10 structures, one AA11 structure and one AA13 structure are published. The six AA9 structures are listed in Table 3.2 with their PDB structure codes, names, regioselectivities and published articles. In addition the structure of *Nc*LPMO9C-N is solved and is soon to be published (Borisova *et al.*, Submitted April 2015), so this structure is also listed in the table and will be used for further comparisons. This table illustrates that two of the structures are PMO-1s that oxidize C1 upon cellulose cleavage, two of them are PMO-2 oxidizing C4 and three are PMO-3s oxidizing both C1 and C4.

Table 3.2: AA9 enzymes with published crystal structures

Structure	Name	Regioselectivity	Source
4B5Q	*Pc*LPMO9D	C1	Wu *et al.* (2013)
3EII/3EJA	*Tt*LPMO9E	C1	Harris2010
4EIR	*Nc*LPMO9D	C4	Li *et al.* (2012)
4D7U	*Nc*LPMO9C-N	C4	Borisova *et al.* (Submitted April 2015)
4EIS	*Nc*LPMO9M	C1 and C4	Li *et al.* (2012)
2YET/3ZUD	*Ta*LPMO9B	C1 and C4	Quinlan *et al.* (2011)
2VTC	*Hj*LPMO9B	C1 and C4	Karkehabadi *et al.* (2008)

Homology models of *Mt358*, *Mt*4260-N and *Mt*6403 were made using the SwissModel server (Biasini *et al.*, 2014). The template for *Mt*358 was 3EJA, and these two proteins had a very high sequence identity of 79 %. 4EIR served as a template for *Mt*4260-N with a 46 % sequence identity, and 2YET was the template for *Mt*6403 with a sequence identity of 58 %. The sequence identities between all these pairs are sufficient for relatively reliable models.

In Figure 3.2 PyMOL graphics of the seven published AA9 structures are compared to the homology models regarding their secondary structures, L2 loops and regioselectivities. It is apparent from the figure that the enzymes that oxidize both C1 and C4

have an extended surface area due to a larger L2 loop, as described in Section 1.3.5. Figure 3.2 illustrates that *Mt*358 has a similar L2 loop to the PMO-1s, the L2 loop of *Mt*4260-N has the same shape as in the PMO-2s while *Mt*6403 exhibits the same long L2 loop as the PMO-3s.

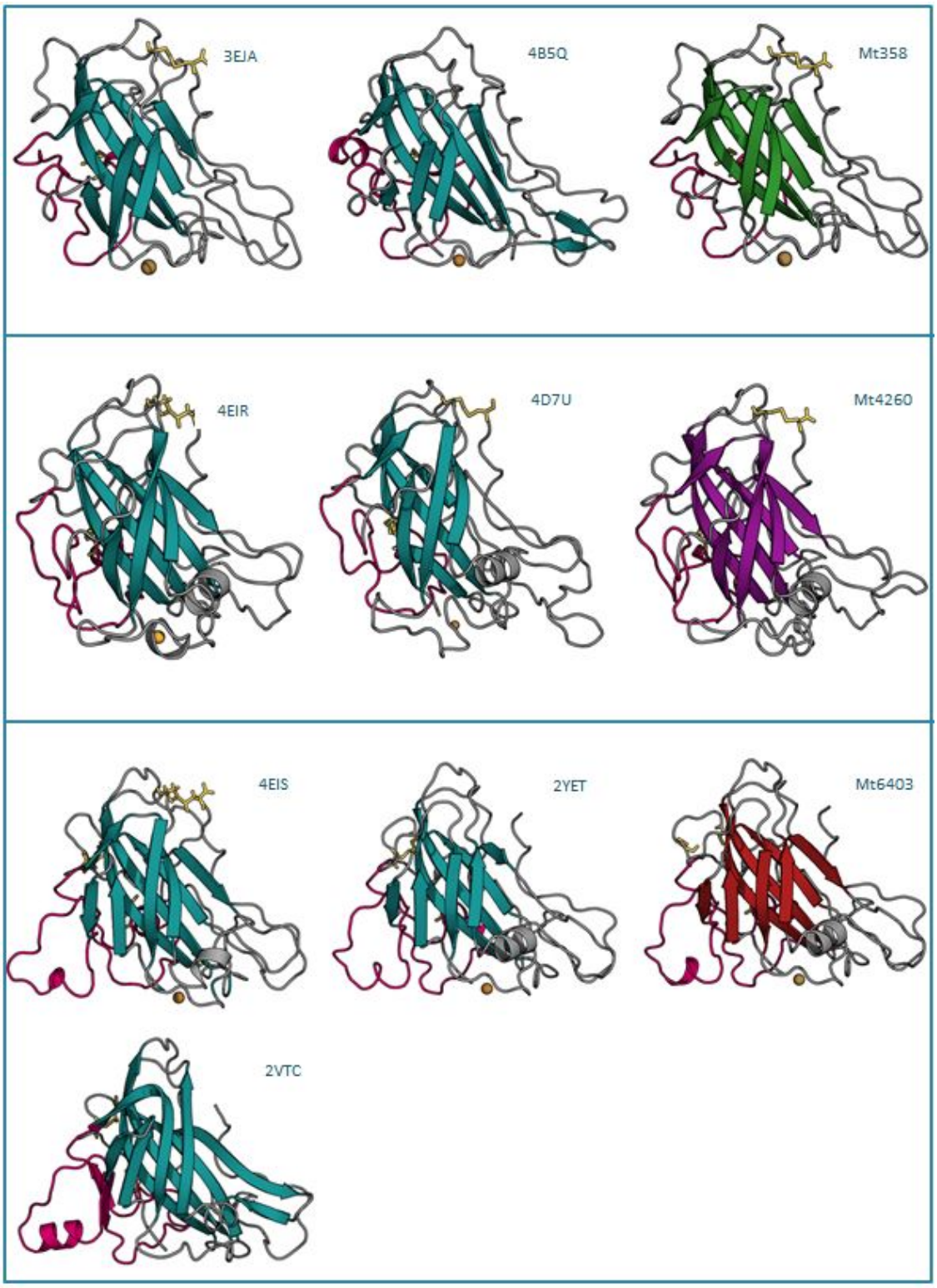

Figure 3.2: Comparison of seven known structures of AA9 proteins to the three *M. thermophilum* homology models made in this study. L2 loops are colored pink, beta sheets teal and cystines in yellow sticks. If the structure was published with a divalent ion this is shown as a golden sphere. The upper models are C1 oxidizing, the middle ones are C4 oxidizing and the bottom ones oxidize both C1 and C4. Proteins from this study are placed besides the group most similar based on their structure, with different colored beta sheets: green in *Mt*358, purple in *Mt*4260-N and red in *Mt*6403.. Made with PyMOL (Schrödinger, LLC, 2010).

In Figure 3.3 the three *M. thermophilum* homology models are structurally aligned in PyMOL, emphasizing the L2 and L3 loop (described by Wu *et al.* (2013) and Borisova *et al.* (Submitted April 2015), see Section 1.3.5). The L3 loop is only present in *Mt*4260-N, and only *Mt*6403 has an extended L2-loop. Topographies of the putative binding surfaces in these three proteins are also shown superimposed, to compare differences that could play a role in binding to different substrates.

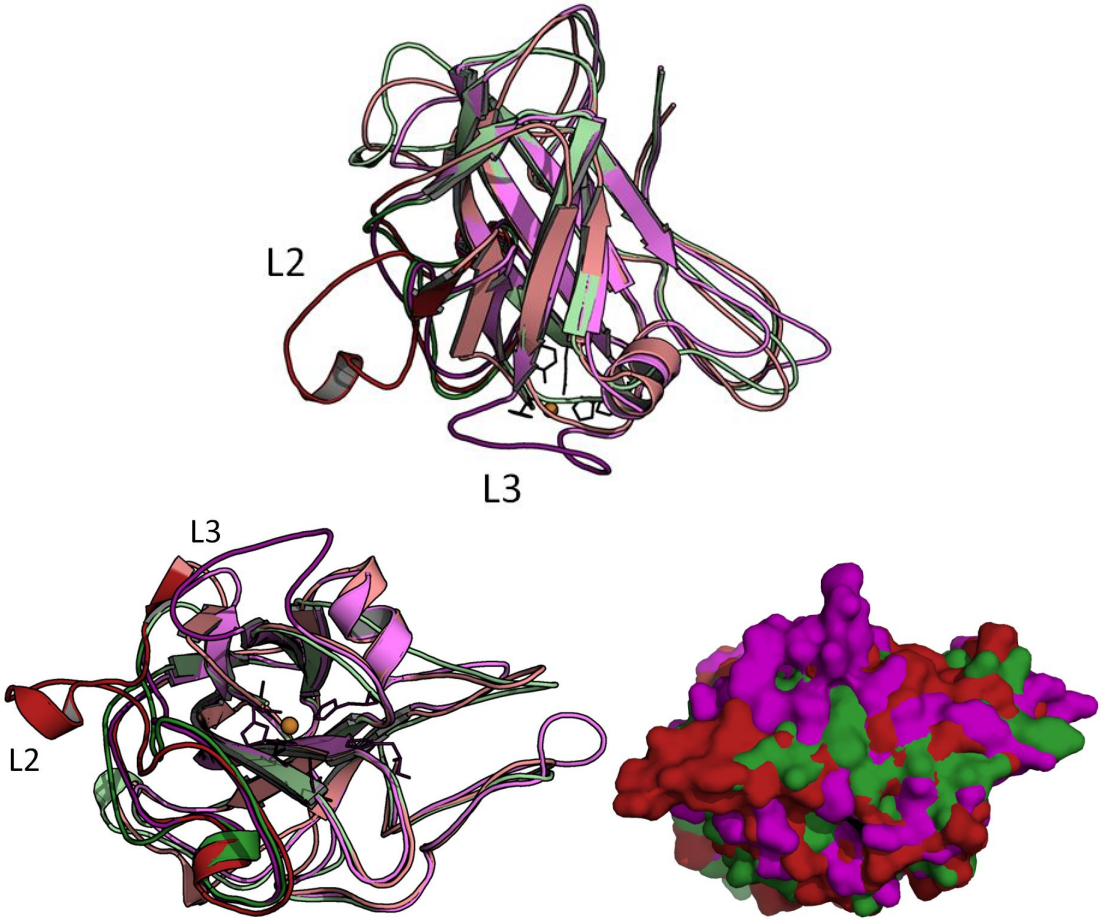

Figure 3.3: *Upper part:* Structural alignment of the three *M. thermophilum* LPMOs illustrating the L2 loop that is found in all AA9s but extended in PMO-3s, and the L3 loop (Borisova *et al.*, Submitted April 2015) found in PMO-2s. *Mt*358 is shown in green, *Mt*4260-N in purple and *Mt*6403 in red. The active site of *Mt*4260 is shown in lines with a copper in orange. *Lower part:* The same alignment rotated 90 degrees to show the putative LPMO binding surfaces, shown in both cartoon and surface to illustrate secondary structures as well as topography. Made with PyMOL (Schrödinger, LLC, 2010)

Putative binding surfaces of all known AA9 structures differ in their topography as shown in Figure 3.4. They have varying widths, lengths and protruding regions that could accommodate different substrates in ways that are still not clear. To illustrate different substrates that the AA9s might have to bind a cellulose pentamer and a hemicellulose nonamer (called XLLG after xyloglucan nomenclature where G = β-D-glc, X = α-D-Xyl(1-6)-β-D-Glc, and L = β-D-Gal-(1-2)α-D-Xyl-(1-6)-β-D-Glc (Agger *et al.*, 2014)), are shown in the figure.

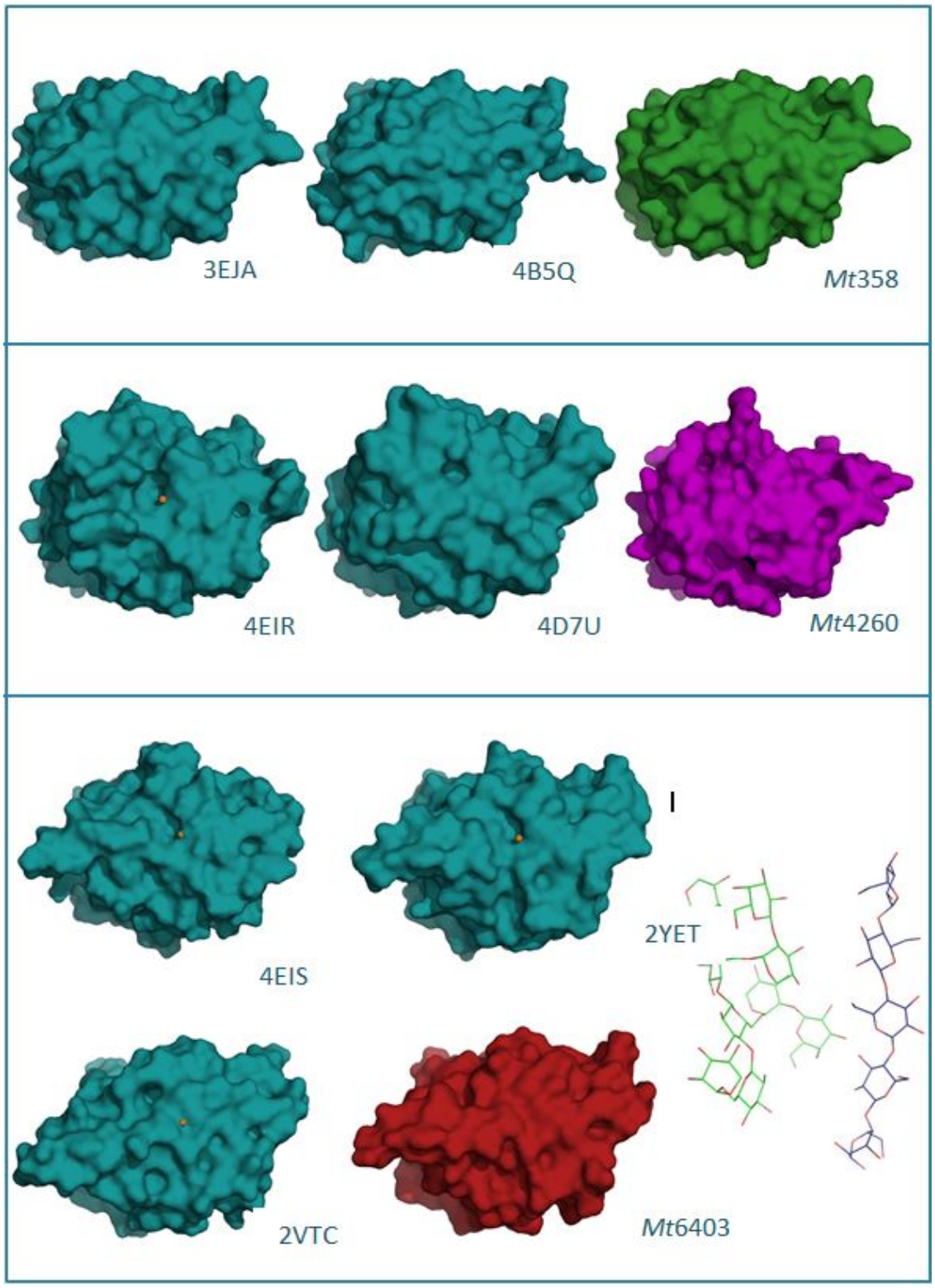

Figure 3.4: Surfaces of the same ten AA9 structures shown in Figure 3.2, showing the putative binding surface as in Figure 3.3 (bottom). The seven crystal structures are colored teal and the *M. thermophilum* proteins are colored green (*Mt*358), purple (*Mt*4260-N) and red (*Mt*6403). Divalent metal ions are shown as golden spheres. The upper models are C1 oxidizing, the middle ones are C4 oxidizing and the bottom ones oxidize both C1 and C4. The structural models of the proteins from this study are placed besides the group most similar based on their structure. In the right corner are cellulose (Glc5, blue) and xyloglucan (XLLG, green) oligosaccharides, showing the linearity of cellulose compared to the branched structure of xyloglucan. Made with PyMOL (Schrödinger, LLC, 2010).

The active copper site is conserved in all LPMOs and has five essential residues that are shown in Figure 3.5. It consists of two histidines coordinating the copper with three ligands in a T-shaped histidine brace. Water molecules and a tyrosine can also take part in the coordination, depending on the copper ionic state. Two other residues in the active site, a histidine and a glutamine, participates in a hydrogen bonding network together with water molecules. The bonds in this network are illustrated in Figure 1.12 in Section 1.3.5.

In the active site, close to the copper, there is a residue that differs between the three LPMOs in this study. It is hypothesized that this residue is important for regioselectivity as it could restrict positioning of the substrate to allow oxidation of either C1 or C4 (Borisova *et al.* (Submitted April 2015), Trine Isaksen (personal communication)). Figure 3.5 illustrates how this residue possibly restricts positioning of the substrate more in *Mt*358 than in *Mt*4260, with *Mt*6403 being something in between. C1 oxidizing AA9s mentioned earlier also have a tyrosine and C1/C4 oxidizing AA9s mentioned have a proline, while the two C4 oxidizing enzymes have either an alanine or an aspartate (see Table 3.2).

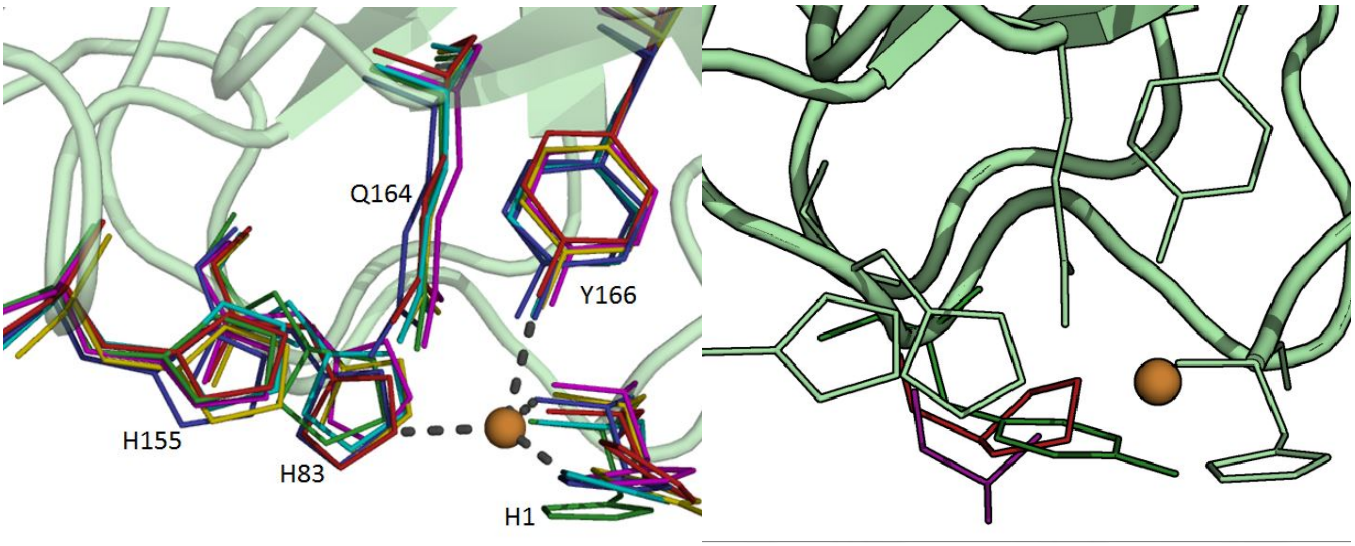

Figure 3.5: *Left part:* Aligned residues in the active site of six AA9s with five absolutely conserved amino acids (numbering of *Mt*4260). *Mt*358 (green), *Mt*4260-N (purple), *Mt*6403 (red), *Nc*LPMO9C-N (blue), *Ncr*PMO-3 (yellow), and *Pc*LPMO9D (teal). *Right part:* An active site residue that differs in the three *M. thermophilum* LPMOs is shown in lines: a tyrosine in *Mt*358 (green), an alanine for *Mt*4260 (purple) and a proline in *Mt*6403 (red). Active site residues are shown in lines for *Mt*358, in the same position as in the upper image. The homology model of Mt358 is shown surrounding the residues in both figures, in pale green. The figures were made and aligned with PyMOL (Schrödinger, LLC, 2010).

3.1.4 Multiple sequence alignment

A multiple sequence alignment (MSA) of the six published AA9 structures, together with the three *M. thermophilum* proteins and *Nc*LPMO9C-N, summarizes conserved

features that are described so far (Figure 3.6). For *Mt*4260 and *Mt*6403 only the N-terminal AA9 domains were aligned. In all ten proteins 24 residues are fully conserved, and five of these are the three histidines, the glutamate and the tyrosine in the catalytic site that are illustrated in Figure 3.5. Cysteines that form disulfide bridges are conserved and important for the stability of the structures. Together the AA9s have three conserved disulfide bridges.

The L2 loop is extended in the three PMO3s compared to the other two classes, and this together with 10-12 inserted residues in PMO2s is thought to be important for C4 oxidation. Another residue that could be important in regioselectivity is P83 in *Mt*6403, described above.

Based on the sequence alignment *Mt358* has a similar sequence to the PMO-1s, *Mt*4260-N has the same features as the PMO-2s, and *Mt*6403 resembles the PMO-3s.

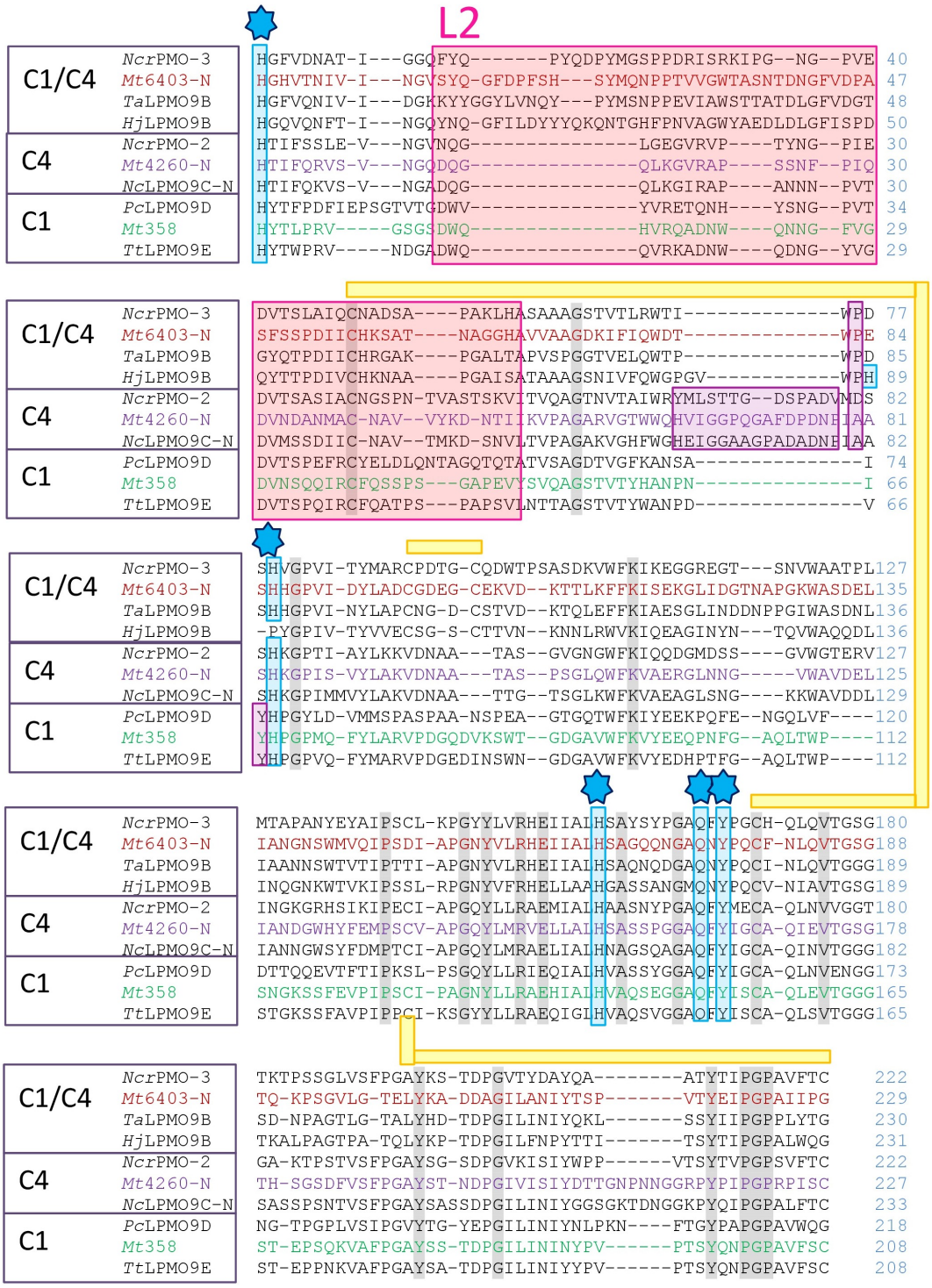

Figure 3.6: MSA for ten AA9s, including the three from *M. thermophilum*. *Ncr*PMO-3, *Ta*LPMO9B, *Hj*LPMO9B, *Ncr*PMO-2, *Pc*LPMO9D, *Tt*LPMO9E and *Nc*LPMO9C-N are shown in black letters, *Mt*358 in green, *Mt*4260-N in purple and *Mt*6403-N in red.. Fully conserved residues are highlighted in grey unless they are important residues in the copper site (highlighted turquoise with a star). Features important for regioselectivity is the L2 loop (highlighted pink), a few residues around the second coordinating histidine (in purple), and the residue that might restrict substrate access to the active site (also in purple). Yellow lines mark cysteine pairs forming disulfide bridges. The alignment is made with Clustal Omega (Sievers *et al.*, 2011) and the order of the enzymes is assigned by this program based on sequence similarity. Li *et al.* (2012), Wu *et al.* (2013) and Borisova *et al.* (Submitted April 2015) are used as references to find conserved residues and structural features in the published structures and in *Nc*LPMO9C-N.

3.2 Cloning, expression and characterization

Full length and truncated version of one of the *M. thermophilum* proteins, *Mt*4260, was cloned and expressed. *Mt*358 and *Mt*6403 were expressed earlier as described by Kittl *et al.* (2012).

3.2.1 Cloning of *Mt*4260

To study the importance of the CBM1 domain in *Mt*4260 a protein with only the AA9 domain was expressed in addition to the full length protein. The site of truncation was determined using sequence alignment. First the protein sequence was aligned with the very similar, two-domain protein *Nc*LPMO9C with Clustal Omega (Sievers *et al.*, 2011). These two proteins show a high sequence similarity and both have an AA9 domain and a CBM1 domain. *Nc*LPMO9C has already been truncated successfully (Borisova *et al.*, Submitted April 2015), so its truncation point was used as a reference in the resulting alignment shown in Figure 3.7. A clear difference evident from the alignment is the longer linker region of *Nc*LPMO9C between the domains. The truncation site was determined between C242 and S243 (the numbering includes the signal peptide).

Figure 3.7: Sequence alignment of *Mt*4260 and *Nc*LPMO9C. Conserved residues are marked in purple, and boxes over the sequences are marked with the following color code: Signal peptides are grey, the AA9 domain tan, the linker peptide blue and the CBM1 domain red.

For cloning of *mt4260* and *mt4260-N* the expression vector pPICZB containing *mt4260* was extracted from ON cultures of *E. coli* as described in 2.2.1.2. PCR amplification of both versions of this plasmid, described in 2.2.1.4, gave fragments that are shown in Figure 3.8 and that corresponds well to their predicted amounts of basepairs (930 and 747 respectively, see Table 3.1). Both gene versions were cloned into pPink-GAP-HC for expression in *E. coli* and PichiaPink™. Maps of the two vectors are found in Appendix B.

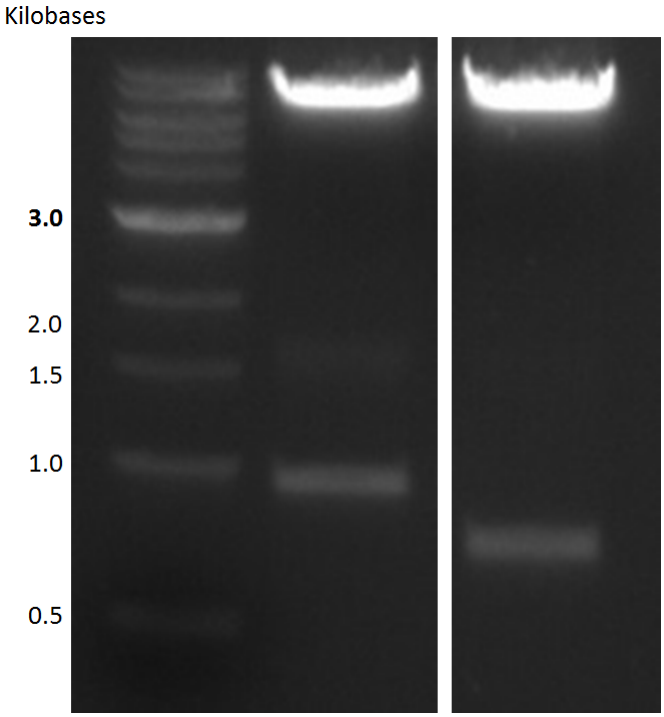

Figure 3.8: PCR amplification products of the *mt4260* genes. The left lane shows Quick-Load® 1 kb DNA Ladder, the middle lane shows *mt4260* (930 bp) and the right lane shows *mt4260-N* (747 bp).

pPink-GAP-HC was extracted, purified and cut as described in Sections 2.2.1.2 and 2.2.1.5 before the genes were cloned into the vector with In-Fusion® cloning. Then the plasmids were transformed into *E.coli* and the colonies were tested with colony PCR to confirm transformation. One transformed colony was cultured and the vector was purified from this culture and linearized. The genes were sequenced to make sure they were cloned correctly and without mutations. Finally, the plasmid was transformed into PichiaPink™. This transformation was succesful for both *mt4260* and *mt4260-N*, as white colonies grew on the selection plates after about a week while nothing grew on negative control plates. Figure 3.9 shows a picture of a plate with red and white PichiaPink™ colonies, with the white colonies expressing larger amounts of the protein as described in Section 2.2.1.12. Several white colonies were picked and cultured for expression of the desired proteins.

Figure 3.9: PAD selection plate with white and red PichiaPink™ colonies with transformed *mt4260-N*. The white ones are the ones expressing most protein.

3.2.2 Expression and purification

When expressed extracellularly, the proteins are cleaved after their signal sequences and the mature *Mt*4260 and *Mt*4260-N were predicted to have molecular masses of 29,776 and 24,040 Dalton, respectively (Table 3.1). However, the molecular masses of the expressed proteins were somewhat larger than expected as shown in Figure 3.10, about 27,000 Da for truncated and 35,000 Da for full length. This could be because of O-glycosylations as explained in Section 3.1.2. One of the five O-glycosylations predicted at serine residues in *Mt*4260 is in the AA9 domain and four are in the linker region, which means that *Mt*4260 has more potential O-glycosylation sites than *Mt*4260-N. This could explain why *Mt*4260 is 5,000 Da larger than expected while *Mt*4260-N is only 3,000 Da larger.

Initial tests of protein expression were successful, so large scale expression of the enzymes were perfomed. Expression of *Mt*4260 was done with PichiaPink™ cultures both in regular shake flasks and with the Harbinger LEX system as described in Section 2.2.2. With the LEX system the cell cultures grew very thick and expressed protein was not detected the first time, so the expression had to be done twice. Figure 3.10 also shows the expressed *Mt*4260 from both methods. OD of cell cultures were monitored and when harvested (after 23 hours with the LEX system and 42 hours in shake flasks) they all had an OD of about 40. This corresponds to cell densities of approximately 2×10^9 since the cell density at one OD is about 5×10^7 (Invitrogen, 2010). *Mt*4260-N was only expressed in shake flasks as this gave more protein and was a more reliable method. In the *Mt*4260-N culture OD was 27 after 48 hours, which corresponds to a cell density of about 1.4×10^9.

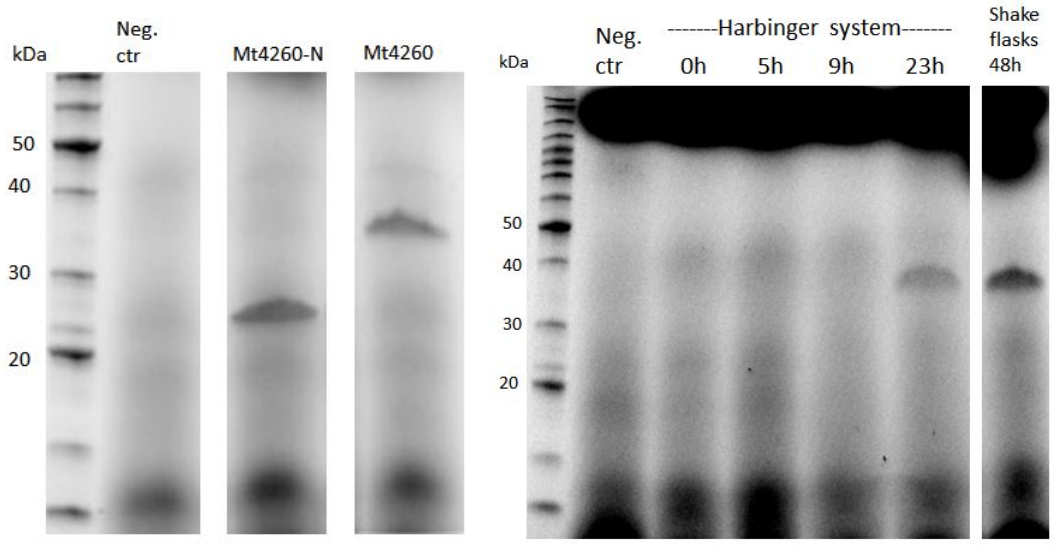

Figure 3.10: *Left:* Both truncated and full length M4260 were expressed in test cultures and the two protein versions from supernatants of centrifuged cultures is shown as clear bands in an LDS-PAGE gel. *Right*: LDS-PAGE gel comparing *Mt*4260 in cultures from the LEX bubbling system and shake flasks. Samples expressed with the Harbinger LEX system is from a 500 ml culture and is more diluted than the shake flask sample from 300 ml. From the Harbinger system samples were taken out at four different time points, and all are shown in the gel. The negative control was untransformed PichiaPink™ cells.

HIC (and IEC for *Mt*4260) successfully purified the proteins after some optimization. In Appendix C chromatograms and LDS-PAGE gels from the purification steps of both proteins are shown. Purification of 1.5 liter cultures resulted in 25.6 mg of *Mt*4260 and 17 mg of *Mt*4260-N. Figure 3.11 shows the protein solutions before and after purification, with most of the contaminations being removed in purification steps. After HIC *Mt*4260-N was about as pure as *Mt*4260 was after HIC and IEC, so a IEC step was not used for the truncated protein. The figure also shows a sample from the filtrate from *Mt*4260-N after pre-purification steps, which has small amounts of the protein as well. Since the filtrates had quite large volumes a lot of the protein is actually lost in these steps.

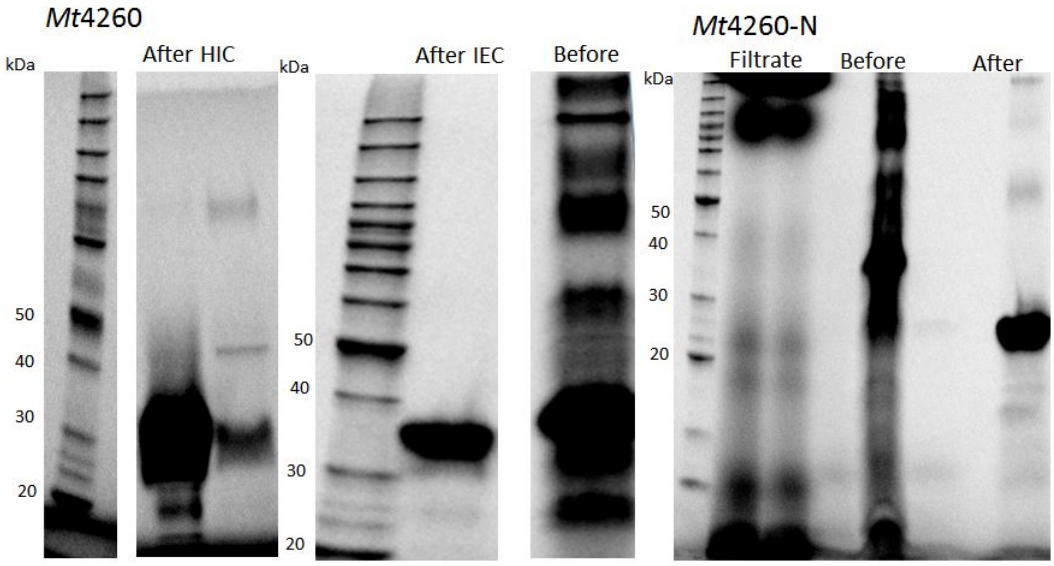

Figure 3.11: Purification results of full length and truncated *Mt*4260. *Left: Mt*4260 before and after HIC and IEC. *Right: Mt*4260-N before and after HIC purification steps. Filtrates are also shown, indicating that some protein got lost in filtration steps.

3.2.3 Characterization of enzyme activities

Activity assays were used to characterize the purified enzymes. First, their substrate specificities were investigated by HPAEC and MALDI-TOF as described in Sections 2.2.3.3, 2.2.3.5 and 2.2.3.4. Table 3.3 shows which substrates the three enzymes showed activity on. Appendix D includes chromatograms for reactions where activity was detected (except for *Mt*4260-N's activity on tamarind xyloglucan, which was confirmed only by DNS assay as described below), while chromatograms are not included for reactions where no activity was found. Notable results from Table 3.3 is that both *Mt*4260 and *Mt*6403 are active on xyloglucan, and that *Mt*4260 activity on soluble cellooligosaccharides. Avicel, cellulose monoacetate and several other hemicelluloses than the ones listed were also tested as substrates, but the results were inconclusive.

Table 3.3: A summary of substrate specificities of the three *M. thermophilum* enzymes. *Mt*4260-N is in a separate column from *Mt*4260 as it was not tested on every substrate that *Mt*4260 was tested on. + means activity was detected, - means activity was not detected and blank means the substrate-enzyme combination was not tested.

Substrate	Mt358	Mt4260	Mt4260-N	Mt6403
Glc5	-	+	+	-
PASC	+	+		+
Tamarind xyloglucan	-	+	+	+
Aspen xylan	-	-		-
Birchwood xylan	-	-		-
α-chitin	-	-		-
β-chitin	-	-		-

Regioselectivities and product formations of the three enzymes were also studied by HPAEC and mass spectrometry. In Figure 3.12 the different peaks of all three enzymes incubated with PASC are assigned to their most likely products as described by Agger *et al.* (2014), based on MALDI-TOF analysis and earlier assigned elution orders of oligosaccharides. Peaks in the ICS chromatograms corresponding to C1 oxidized products elute slightly after native oligosaccharides while peaks diagnostic for C4 products elute even later. Reaction with *Mt358* made products typical for C1 oxidizing LPMOs (ladder of products eluting from 37-52 minutes), *Mt*4260 made later eluding peaks (46-53 minutes) corresponding to typical C4 oxidized products. *Mt*6403 have relatively small peaks eluting in a wide time range, but it is not easy to conclude on regioselectivity based on this chromatogram. C4 oxidized products are most likely observed, though.

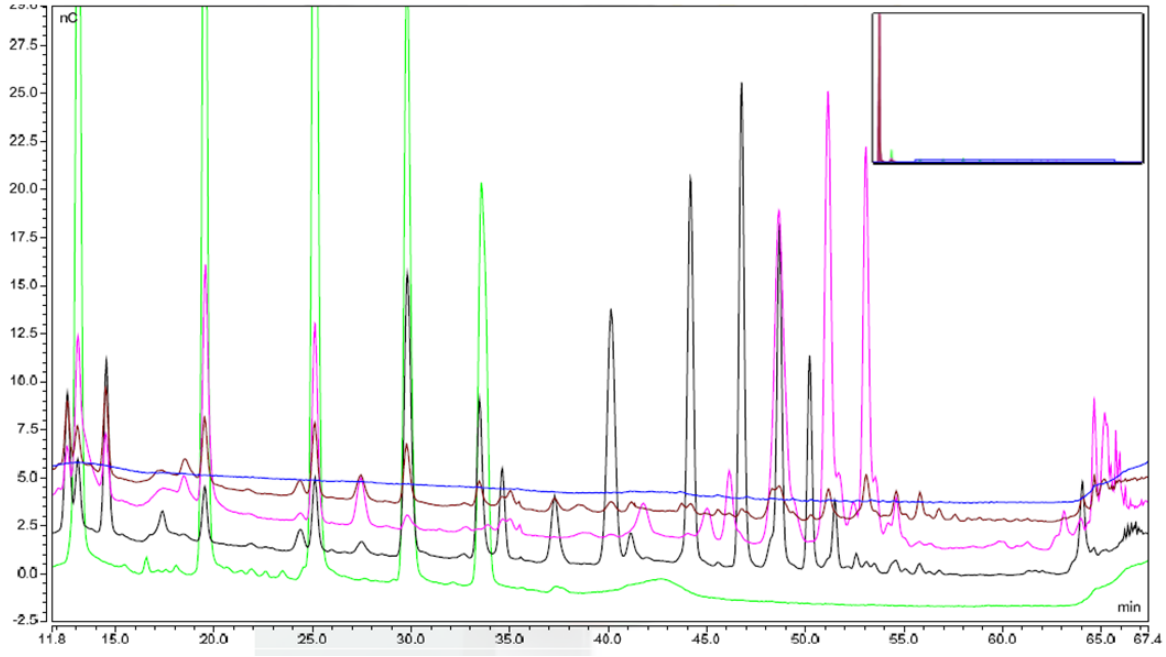

Figure 3.12: HPAEC chromatogram of products generated from 1 mg/ml PASC incubated at 50 ºC for eight hours with all three enzymes in a concentration of 30 mg enzyme per g of substrate, 2 mM ascorbic acid and 5 mM sodium acetate pH 5.6. The reaction with *Mt358* is colored in black and shows a cluster of products corresponding to C1 oxidized oligomers (37-52 minutes), pink shows the reaction with *Mt*4260 yielding products diagnostic for C4 oxidation (46-53 minutes) while *Mt*6403 gives the brown, short peaks indicating both C1 and C4 oxidized products. The green peaks are standards with 0.05 mg/ml native cellooligosaccharides DP 2-6, and the blue line is a negative control without enzymes.

Mass spectrometry results were used to confirm the chromatographic results as the mass profiles differ slightly for the different oxidized products. Figure 3.13 shows mass spectra from incubation with PASC for *Mt358* and with tamarind xyloglucan for *Mt*4260 and *Mt*6403. Together with the chromatogram the MALDI-TOF data clearly show that *Mt358* is C1 oxidizing because it creates products with masses that corresponds to lactones and aldonic acids of the cellooligosaccharides. *Mt*4260 is confirmed C4 oxidizing because of characteristic peaks at m/z = 1245.1 and m/z = 1407.3 that could be G_{ox}XXL and G_{ox}XLL, respectively (see Section 3.1.3 for xyloglucan nomenclature). For *Mt*6403 the chromatograms and mass spectra does not give clear answers to its oxidation form, but the enzyme seems to give C4 oxidized products with a larger DP than *Mt*4260 as evident both from the later eluting peaks in the chromatogram and the larger product masses in the mass spectrum. Mass spectra for PASC were not used for *Mt*4260 and *Mt*6403 because they gave very few visible peaks.

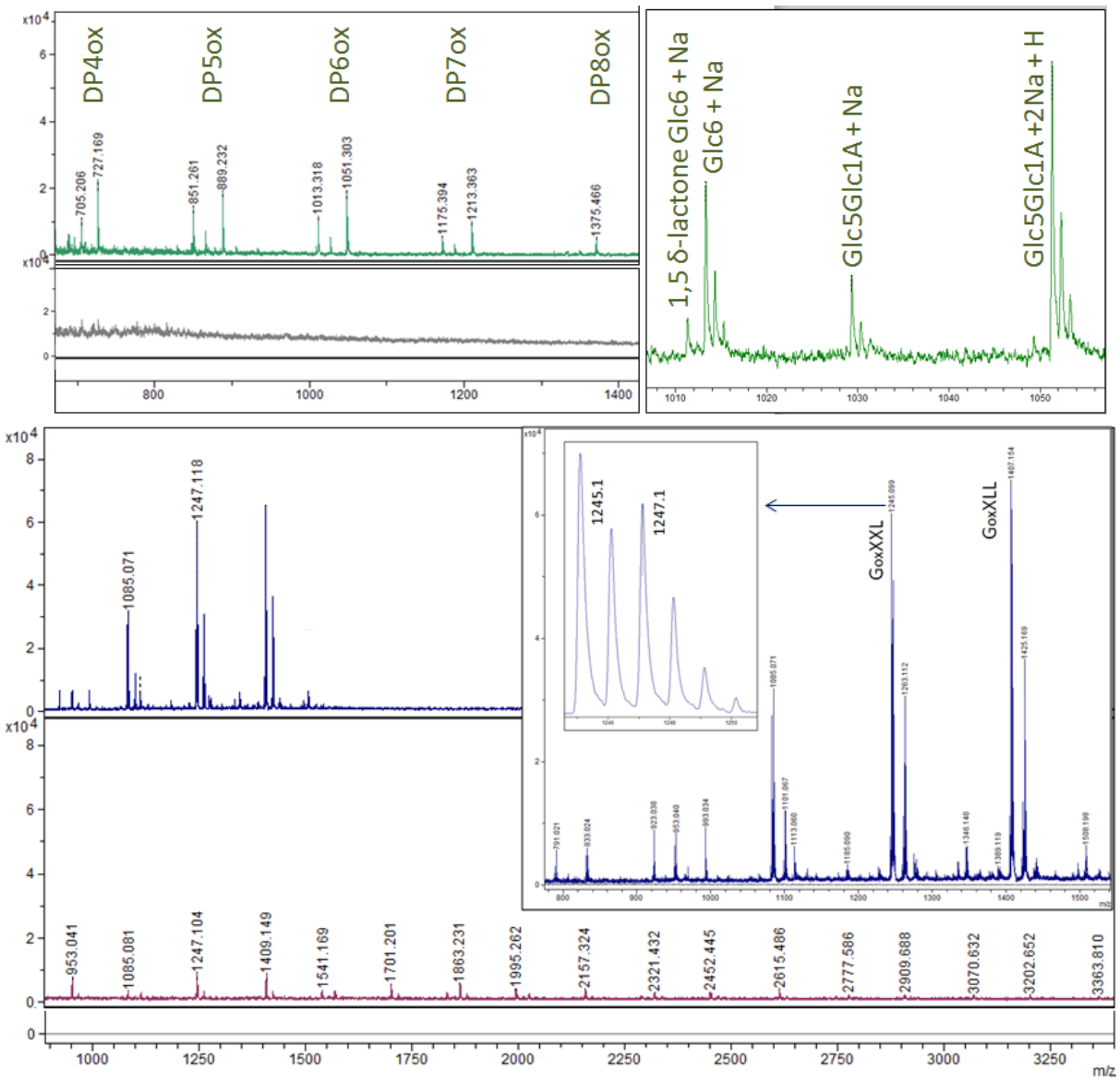

Figure 3.13: *Upper panels:* MALDI-TOF mass spectra of products generated from 1 mg/ml PASC incubated with *Mt*358 (green) and a negative control (in grey) at 50 ºC for eight hours in a concentration of 30 mg enzyme per g of substrate, 2 mM ascorbic acid and 5 mM sodium acetate pH 5.6. Included is a close up of one of the peak clusters. m/z is hard to see in the photo, but for 1,5 δ-lactone Glc6 m/z = 1011.3, for Glc6 (M+Na) m/z = 1013.3, for Glc5Glc1A (M+Na) m/z = 1029.3 and for Glc5Glc1A (M+2Na+H) m/z = 1051.3. Annotations in A and B are based on data from Forsberg *et al.* (2014b) where they are thought to be evident of C1 oxidation. *Lower panel:* MALDI-TOF spectra of products from *Mt*4260 (blue) and *Mt*6403 (red) incubated for 18 hours with tamarind xyloglucan and the same other reaction components as for the PASC reaction. Negative control is shown as a grey line beneath the mass spectra of *Mt*6403. A close up of the peaks from *Mt*4260 is shown, where one peak at m/z = 1247.1 is shown in a box. This peak probably confirms C4 oxidation as described by Agger *et al.* (2014) and could consist of a product called $G_{ox}XXXG_{OH}$. For *Mt*6403 possible products are not annotated but it seems to give weaker signals than *Mt*4260, and products with a higher DP. A periodicity of 162 is apparent in the peaks, corresponding to a glucose unit.

To quantify product formation in the reactions DNS assays were conducted with *Mt*4260 and *Mt*4260-N as described in Section 2.2.3.6. DNS assay could not be used

with *Mt*358 since this enzyme does not produce reducing ends (it seems to oxidize the C1 position), and if Mt6403 oxidizes both C1 and C4 this assay could not be used to measure its kinetics either. The substrates PASC, Avicel and tamarind xyloglucan were tested. However, Avicel gave so low product concentrations that detection was not possible with this method and PASC gave varying results. Because of this only the reaction rates with tamarind xyloglucan are calculated and the results for the other two substrates are not included. Different concentrations of the enzymes were tried out and the concentration that gave the best results for both was 0.8 µM. A graph showing the reaction rate with tamarind xyloglucan is shown in Figure 3.14. The reaction rate with *Mt*4260 was calculated to be 0.3 s^{-1} while the rate with *Mt*4260-N was 0.09 s^{-1}.

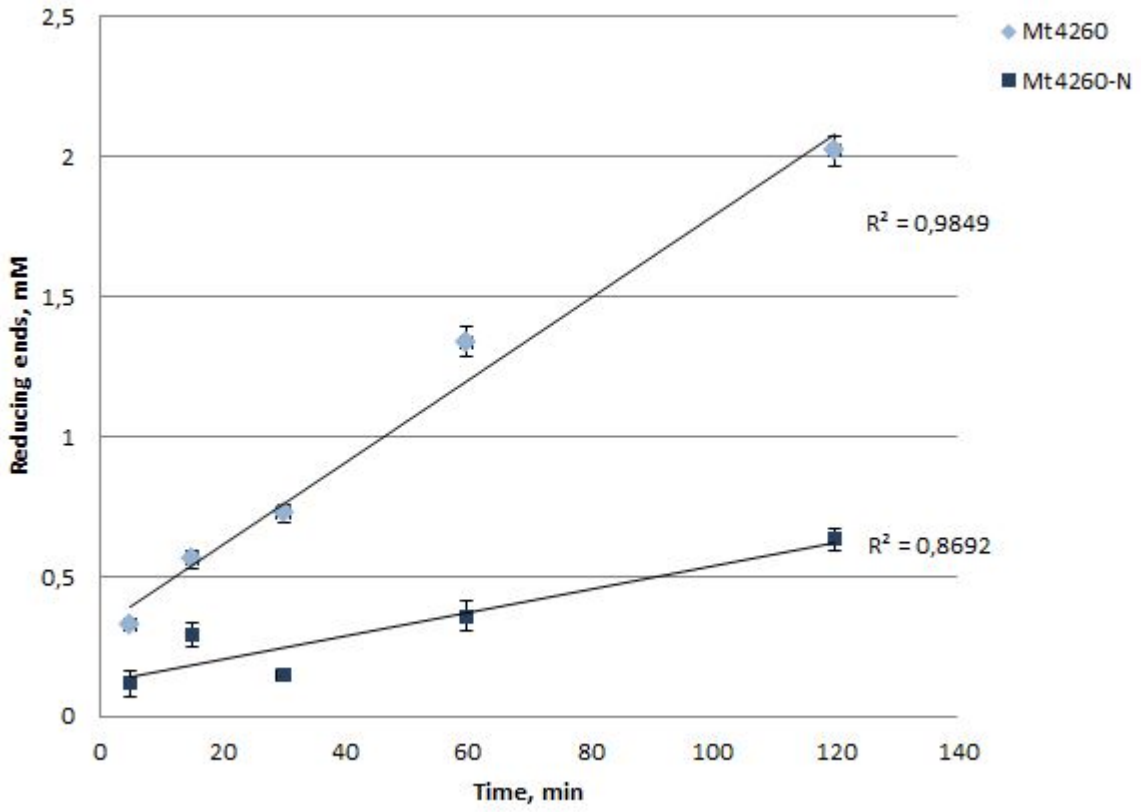

Figure 3.14: Reaction rate of 0.8 µM *Mt*4260 and *Mt*4260-N with 5 mg/ml tamarind xyloglucan, in 40 mM MES buffer pH 6.5 with 2 mM ascorbic acid at 50 ºC. Reducing ends are measured with DNS assay. The points are calculated based on triplicates of the reactions. Control reactions without ascorbic acid gave no measurable products. Equations of the regression lines were y=0.0147x+0.3204 for *Mt*4260 and y=0.0042x+0.119 for *Mt*4260-N.

4. Discussion

The goal of this study was to characterize the activity of three AA9 LPMOs from the filamentous fungi *M. thermophilum* after having cloned, expressed and purified them. Unfortunately, there was only time to express one of the proteins, the two domain enzyme *Mt*4260. It was however cloned and expressed in both full length and truncated version. Characterization studies were done and homology models were made for all three enzymes, and their differences and similarities were investigated and compared to AA9s with known structures.

4.1 Expression and purification

*Mt*4260 and *Mt*4260-N were both expressed and secreted successfully in *P. pastoris* with their native signal peptides, as confirmed by activity tests in addition to LDS-PAGE gels. If the N-terminal signal peptides were not processed accurately the N-terminal histidine would not be the first amino acid, something that is essential for activity (Harris *et al.*, 2010).

Harbinger LEX expression system was tested for expression in *P. pastoris* but it seemed like the system, that has proven very effective for protein expression in *E. coli*, was not suitable for expression in the yeast. The cultures grew very thick and lumpy, probably to the death phase, and the procedure had to be repeated a few times before the desired protein was expressed. Further optimization of this method is needed when it comes to how long the cells should be incubated. Also, the rate of air bubbled in, potential stirring of the culture to prevent yeast lumps, and if nutrients should be added during incubation needs to be optimized. However, there was not enough time during the work of this thesis. For this reason shake flasks were mostly used for the expressions as they gave more reliable results.

Expression of fungal LPMOs in *P. pastoris* does not give methylation of the N-terminal histidine (Westereng *et al.*, 2011), which is known to happen in native fungi. Since this methylation in the active site might affect enzyme activity (Aachmann *et al.*, 2012; Hemsworth *et al.*, 2013) in ways that are unclear, it is important to note. An

alternative to expression in *P. pastoris* is expression in *N. crassa*, as described by Vu *et al.* (2014a). However, due to the lack of an established *N. crassa* expression system in our laboratory, the well-adapted yeast expression system PichiaPink™ was applied instead. Activity comparisons of methylated and non methylated AA9s should be done in the future to show if this is important or not.

Purifications gave very pure protein samples and it appears that the purification with ammonium sulfate precipitation and HIC, followed by IEC when necessary, was a successful method. Very small amounts of contaminations can be noticed in the gels, but for the timescale used for most experiments (8-18 hours) no other activity was detected. Size exclusion chromatography could have been used to further purify the proteins, assuming other proteins present had sufficiently different masses from the proteins of interest.

A lot of protein was lost during filtrations, especially for *Mt*4260-N, and protein was observed in flow-throughs both from membrane filtering and filter tubes. Filters used had a cut-off of 10 kDa which should be small enough for both proteins, but all filters give a percentage of recovery below 100 %. Recovery of the tubes were 90 %, so 10 % of the proteins will be lost during every centrifugation. Since large volumes of culture had to be concentrated to just a few milliliters, quite a few centrifugations were performed. This problem could have been reduced if larger columns that could take higher volumes were used for purification. Slower centrifugations and lower pressures for membrane filtering could also have reduced loss.

4.2 Sequence and structure analysis

All three proteins are AA9 LPMOs based on sequence analysis. *Mt*4260 has a CBM1 domain with a linker region rich in possible O-glycosylation sites, while *Mt*6403 has a long C-terminal region with a large amount of O-glycosylation sites. Observed molecular weights of *Mt*4260 and *Mt*4260-N after expression were larger than what would be expected from their sequences, which could mean that they were glycosylated. The function of this possible glycosylation is not clear, but glycosylated linker regions in some GHs are though to be associated with processivity and binding of this linker to the substrate is modelled by Payne *et al.* (2013). O-glycosylation in CBM1 domains could also be important for binding because of interactions with the substrate (Taylor *et al.*, 2012). In general, length and modifications of linkers are thought to have an impact on enzyme activity, which could be because the linkers actually bind to the substrate or because the space and flexibility between the modules are important (Srisodsuk *et al.*, 1993). It is not unlikely that the linker region is more than just a connection between two domains, and that the long C-terminal end of *Mt*6403 has a function as well. Only

speculations can be made so far about the function of these regions, but one theory is that they could affect substrate binding in some way. N-glycosylations in AA9s might also affect binding directly (Li *et al.*, 2012), but this will not be discussed further as none of the three *M. thermophilum* enzymes contain any potential N-glycosylation sites.

Based on relatively similar templates, homology models were made of the three enzymes. Structure and sequence analysis seem to imply that each of the enzymes belong to a different class of AA9s: *Mt*358 is similar to PMO-1s, *Mt*4260 to PMO-2s and *Mt*6403 to PMO-3s. This is based on conserved residues and structural similarities to already characterized proteins with known structures, like the L3 loop in *Mt*4260 and the extended L2 loop in *Mt*6403. Based on this classification it would be expected that *Mt*358 produces C1 oxidized products, *Mt*4260 produces C4 oxidized products and *Mt*6403 produces both. Of the two subclasses of PMO-3s, *Mt*6403 most likely belong to PMO-3 and not PMO-3*, an assumption that is based on the presence of the residue Y24 (Vu *et al.*, 2014a).

Homology models of the active sites in these three LPMOs align with the active sites in known structures of AA9s, and this supports further the theory that all AA9s use the same basic mechanism for oxidative cleavage of carbohydrates. The residues depicted in Figure 3.5 (*Right part*) differs in the different classes of AA9s, as pointed out by Borisova *et al.* (Submitted April 2015). The tyrosine in PMO-1s (and *Mt*358) is large and would restrict positioning of the substrate in this specific position, the proline in PMO-3s (and *Mt*6403) is somewhat smaller and could partly accommodate a substrate here, while the small amino acids found in PMO-2s (and *Mt*4260) could give complete access to this site. If positioning of the substrate relative to the active site is what makes the different regioselectivities, this residue could be essential.

It is interesting that the genome of *M. thermophilum* and other filamentous, cellulose degrading fungi contain over 20 possible LPMOs, as described in Section 1.3.3. With the diversity of carbohydrates it is likely that this diversity of LPMOs in lignocellulose degrading fungi is designed to break down a lot of different substrates. As illustrated in Figure 3.3 (*Lower part*), *Mt*358 has the smallest surface area on the binding surface and it is relatively flat. The homology model of *Mt*4260 has a bulky and quite broad surface with protruding areas. *Mt*6403 is also bulky and with a larger surface area, like C1/C4 oxidizing AA9s. The putative binding surfaces could accommodate different substrates in different LPMOs depending on their topography, which varies between AA9s as shown in Figure 3.4. For instance the relatively flat, narrow surfaces of *Pc*LPMO9D and *Mt*358 could be suitable for binding a few glucan chains in crystalline cellulose, while the broader, more bumpy surfaces of *Nc*LPMO9C, *Mt*4260 and *Mt*6403 look like they could bind to more bulky substrates. *Nc*LPMO9C has been

shown to cleave hemicelluloses and soluble cellulose oligosaccharides in addition to crystalline cellulose (Agger *et al.*, 2014), supporting this hypothesis. Compared to the outline of the binding surface of *Mt*4260, *Nc*LPMO9C has a more even outline with less protruding regions.

Crystal structures with bound substrates have not been possible to obtain for LPMOs as of today, probably because of low affinities for soluble substrates (Borisova *et al.*, Submitted April 2015). In addition to the topography, the types of amino acids on the binding surface are important for substrate binding. Aromatic amino acids could for instance bind substrates in a similar fashion as CBM1s as described by Harris *et al.* (2010). This was not investigated further here, but modeling of this binding could be very interesting. It is possible that these residues differs to accommodate different substrates in AA9s, and that is why conserved binding residues in all AA9s have not been easy to find (Harris *et al.*, 2010; Li *et al.*, 2012; Wu *et al.*, 2013). The importance of the topographic differences and differing amino acid compositions in LPMO binding surfaces is not known, but it could be investigated further by binding studies and molecular simulations with substrates for all AA9s with known structures. A good development of this field would be very useful for the biomass industry, as the effect of LPMOs on different substrates could be predicted just by studying their binding surfaces.

Activity characterization

ICS and MALDI were used to confirm regioselectivites and they point to the same conlusions as sequence and structure analyses. However, there were some problems with instruments and method optimization. What substrates they were active on were even harder to confirm, as the signals were usually very low or there were a lot of noise in both chromatograms and mass spectra. *Mt*358 only showed activity on cellulose, while both *Mt*4260 and *Mt*6403 showed activity on the bulky hemicellulose tamarind xyloglucan. This resonates well with what has been described earlier about the more narrow and flat binding surface of *Mt*358. *Mt*4260 was also active on soluble oligosaccharides with DP of 5, as has been shown earlier for *Nc*LPMO9C. It would be interesting to do more thorough activity tests to get more information about substrate preferences for all three enzymes, and to compare this to their structures. Since these three enzymes probably belong to each of the three AA9 classes (PMO-1, PMO-2 and PMO-3) while coming from the same organism, comparison of their substrate profiles are interesting in regard to mechanisms used in nature for effective polysaccharide degradation.

None of the enzymes showed any measurable activity on the crystalline cellulose

substrate Avicel. This is perhaps strange as LPMOs are thought to be acting on crystalline substrates, but cuts in the crystal regions do not necessarily give soluble products that can be measured. In later studies synergy experiments with cellulases could be performed on crystalline substrates to see if the LPMOs have any enhancing effects on degradation, as was described in Section 1.3.2.

Ascorbic acid was used as a reducing agent in all activity studies. Other possible reducing agents are CDH, lignin or gallic acid as described in Section 1.3.6. A problem with using CDH as a reducing agent is that it also oxidizes cellulose and cellooligosaccharides, which can make the detection of oxidized products from LPMO action difficult. Ascorbic acid is a very strong reducing agent, and if too much electrons are donated to the LPMOs a futile side reaction producing H_2O_2 occur (Kittl *et al.*, 2012). This might affect enzyme activity and lead to inactivation of the enzyme, and more should be known about the effects of using different reductants on activity.

An assay with 3-amino-5-nitrosalicylic acid (DNS) was used to determine enzyme rates of *Mt*4260 with and without its binding domain. It was evident from DNS assays that removal of the CBM1 domain in *Mt*4260 reduced the degradation rate on tamarind xyloglucan, so the binding domain seems to be important for the enzyme's activity. It would be interesting to compare its degradation rates with and without the CBM1 domain on xyloglucan with rates on crystalline and semi crystalline cellulose as well, as this could give insight into the effect the domain has on substrate specificities. However, good enough results were not obtained on these substrates because of problems with detection and reproducibility.

The degradation rates of 0.3 s^{-1} for *Mt*4260 and 0.09 s^{-1} for *Mt*4260-N was higher than what was observed for *Nc*LPMO9C and *Nc*LPMO9C-N, which had degradation rates of 0.1 and 0.04 respectively (Borisova *et al.*, Submitted April 2015). Initial rates were much faster, so much so that a concentration of 0.8 µM enzyme had to be used instead of the 4 µM used for *Nc*LPMO9C. *Nc*LPMO9C has a longer linker region between the two domains than *Mt*4260. It is hypothezised by Payne *et al.* (2013) that a longer linker region in GH6s compared to GH7s could give a lower processivity because the protein is better anchored to the substrate. If this was the case here it would be expected that the effect of removing the CBM1 domain would be greater in *Nc*LPMO9C, but instead the opposite was reported.

DNS assay is a commonly used method for measurements of activity in cellulases. But it has problems because of possible interferences from sugar decomposition products and buffer, and different signals depending on oligosaccharide length (Miller, 1959). Also, the alkaline reagent destroys part of the reducing sugars, and the method is not sensitive at sugar concentrations below 1000 µM (Horn & Eijsink, 2004). In addition, since no reducing ends are created from oxidation of the C1 position, this

method can not be used for C1 oxidizing LPMOs. Little is known about LPMO kinetics so far and better methods should be developed so more certainty can be placed on the few results obtained. This would aid greatly in characterization of LPMOs and in experiments where degradation rates are important, like synergy studies.

4.3 Future perspectives

In the future more applied work on these three LPMOs and on other LPMOs from thermophilic fungi could look at synergy between LPMOs and with added cellulases. Since LPMOs differ in regioselectivities and substrate specificities their activities probably complement each other. The genome of *M. thermophilum* contains as much as 22 possible LPMOs, something that is not uncommon for lignocellulose degrading fungi. There must be reasons for this, and this study gives evidence that different activities in the LPMOs from one organism could be one. Further work in LPMO characterisation could contribute to the field of improving technogies for the biomass industry as little is known so far about this large and diverse group of lignocellulose degrading enzymes.

References

Aachmann, Finn L., Sørlie, Morten, Skjåk-Bræk, Gudmund, Eijsink, Vincent G. H., & Vaaje-Kolstad, Gustav. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(46), 18779–18784.

Agger, Jane W., Isaksen, Trine, Várnai, Anikó, Vidal-Melgosa, Silvia, Willats, William G. T., Ludwig, Roland, Horn, Svein J., Eijsink, Vincent G. H., & Westereng, Bjørge. 2014. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proceedings of the National Academy of Sciences*, **111**(17), 6287–6292.

Araki, Rie, Ali, Mursheda K, Sakka, Makiko, Kimura, Tetsuya, Sakka, Kazuo, & Ohmiya, Kunio. 2004. Essential role of the family-22 carbohydrate-binding modules for β-1,3-1,4-glucanase activity of *Clostridium stercorarium* Xyn10B. *FEBS Letters*, **561**, 155–158.

Beeson, William T., Vu, Van V., Span, Elise A., Phillips, Christopher M., & Marletta, Michael A. 2015. Cellulose degradation by polysaccharide monooxygenases. *Annual Review of Biochemistry*, **84**(1). PMID: 25784051.

Biasini, Marco, Bienert, Stefan, Waterhouse, Andrew, Arnold, Konstantin, Studer, Gabriel, Schmidt, Tobias, Kiefer, Florian, Cassarino, Tiziano Gallo, Bertoni, Martino, Bordoli, Lorenza, & Schwede, Torsten. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**(July), 252–258.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*, **382**(3), 769–781.

Borisova, Anna, Isaksen, Trine, Dimarogona, Maria, Kognole, Abhishek, Mathiesen, Geir, Várnai, Anikó, Røhr, Åsmund, Payne, Christina, Sørlie, Morten, Sandgren, Mats, & Eijsink, Vincent. Submitted April 2015. Insights into LPMO diversity from structural and functional characterization of *Nc*LPMO9C, a broad-specificity

lytic polysaccharide monooxygenase. *The Journal of Biological Chemistry*.

Boucher, Doug, Elias, Pipa, Lininger, Katherine, May-Tobin, Calen, Roquemore, Sarah, & Saxon, Earl. 2011. *The Root of the Problem. What's driving tropical deforestation today? Chapter 8 Wood for Fuel*. Tech. rept. Union of Concerned Scientists.

Brigham, John S., Adney, William S., & Himmel, Michael E. 1996. *Handbook on Bioethanol: Production and Utilization Chapter 6. Biomass Feedstock Resources and Composition*. Applied Energy Technology Series. Taylor & Francis.

Carpita, Nicholas C., & Gibeaut, David M. 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal*, **3**(1), 1–30.

Chapman, E. S., Evans, E., Jacobelli, M. C., & Logan, A. A. 1975. The cellulolytic and amylolytic activity of *Papulaspora thermophila*. *Mycologia*, **67**(3), 608–615.

Chawdry, Rehan. 2015 (April). *Angular Plasmid,* `http://angularplasmid.vixis.com/`.

Chum, H., Faaij, A., Moreira, J., Berndes, G., Dhamija, P., Dong, H., Gabrielle, B., Eng, A. Goss, W. Lucht, M. Mapako, Cerutti, O. Masera, McIntyre, T., Minowa, T., & Pingoud, K. 2011. *Bioenergy. IPCC special report on renewable energy sources and climate change mitigation*. Tech. rept. IPCC.

Daly, Rachel, & Hearn, Milton T. W. 2005. Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production. *Journal of Molecular Recognition*, **18**(2), 119–138.

de Hoffman, Edmond, & Stroobant, Vincent. 2007. *Mass Spectrometry, Principles and Applications*. 3rd edn. John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester: John Wiley & Sons, Ltd.

Dimarogona, Maria, Topakas, Evangelos, & Christakopoulos, Paul. 2012a. Cellulose degradation by oxidative enzymes. *Computational and Structural Biotechnology Journal*, **2**(3).

Dimarogona, Maria, Topakas, Evangelos, Olsson, Lisbeth, & Christakopoulos, Paul. 2012b. Lignin boosts the cellulase performance of a GH-61 enzyme from *Sporotrichum thermophile*. *Bioresource Technology*, **110**(0), 480 – 487.

Din, Neena, Damude, Howard G., Gilkes, Neil R., Robert C. Miller, Jr., Warren, R. Antony J., & Kilburn, Douglas G. 1994. C1-Cx revisited: Intramolecular synergism in a cellulase. *Proceedings of the National Academy of Sciences*, **91**(November), 11383–11387.

Dürre, Peter. 2007. Biobutanol: An attractive biofuel. *Biotechnology Journal*, **2**(12), 1525–1534.

"Energy, from waste & wood". 2015 (April). *Retrieved from* `http://energyfromwasteandwood.weebly.com/`.

EPA. 2014 (July). *EPA issues final rule for renewable fuel standard (RFS) pathways II and modifications to the RFS program, ultra low sulfur diesel requirements, and E15 misfueling mitigation requirements*. Regulatory Announcement.

Fergus, C.L. 1971. The temperature relationships and thermal resistance of a new thermophilic Papulaspora from mushroom compost. *Mycologia*, **63**(2), 426–431.

Finn, Robert D., Bateman, Alex, Clements, Jody, Coggill, Penelope, Eberhardt, Ruth Y., Eddy, Sean R., Heger, Andreas, Hetherington, Kirstie, Holm, Liisa, Mistry, Jaina, Sonnhammer, Erik L. L., Tate, John, & Punta, Marco. 2014. Pfam: the protein families database. *Nucleic Acids Research*, **42**(D1), D222–D230.

"Formation, d'ingénieurs en Hydraulique et Mécanique des Fluides". 2015 (April). *Retrieved from* `http://hmf.enseeiht.fr/travaux/CD0506/bei/bei_ere/2/html/Binome1/projet/general_fichiers/images/hemicellulose.JPG`.

Forsberg, Zarah, Vaaje-Kolstad, Gustav, Westereng, Bjørge, Bunæs, Anne C., Stenstrøm, Yngve, MacKenzie, Alasdair, Sørlie, Morten, Horn, Svein J., & Eijsink, Vincent G.H. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Science*, **20**(9), 1479–1483.

Forsberg, Zarah, Røhr, Åsmund Kjendseth, Mekasha, Sophanit, Andersson, K. Kristoffer, Eijsink, Vincent G. H., Vaaje-Kolstad, Gustav, & Sørlie, Morten. 2014a. Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry*, **53**(10), 1647–1656. PMID: 24559135.

Forsberg, Zarah, Mackenzie, Alasdair K., Sørlie, Morten, Røhr, Åsmund K., Helland, Ronny, Arvai, Andrew S., Vaaje-Kolstad, Gustav, & Eijsink, Vincent G. H. 2014b. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences*, **111**(23), 8446–8451.

Forsberg, Zarah Kristina. 2014. *Discovery and characterization of cellulose-active lytic polysaccharide monooxygenase*. PhD thesis, Norwegian University of Life Sciences.

Gao, Pei-Ji, Chen, Guan-Jun, Wang, Tian-Hong, Zhang, Ying-Shu, & Liu, Jie. 2001. Non-hydrolytic disruption of crystalline structure of cellulose by cellulose binding domain and linker sequence of cellobiohydrolase I from *Penicillium janthinellum*. *Acta Biochimica et Biophysica Sinica*, **33**(1), 13–18.

Godfray, H. Charles J., Beddington, John R., Crute, Ian R., Haddad, Lawrence, Lawrence, David, Muir, James F., Pretty, Jules, Robinson, Sherman, Thomas, Sandy M., & Toulmin, Camilla. 2010. Food security: The challenge of feeding 9 billion people. *Science*, **327**(5967), 812–818.

Gudmundsson, Mikael, Kim, Seonah, Wu, Miao, Ishida, Takuya, Momeni, Majid Hadadd, Vaaje-Kolstad, Gustav, Lundberg, Daniel, Royant, Antoine, Ståhlberg, Jerry, Eijsink, Vincent G. H., Beckham, Gregg T., & Sandgren, Mats. 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. *Journal of Biological Chemistry*, **289**(27), 18782–18792.

Gupta, R., Jung, E., & Brunak., S. 2004. Prediction of N-glycosylation sites in human proteins. *In Preparation*.

Hara, Yukari, Hinoki, Yumi, Shimoi, Hitoshi, & Ito, Kiyoshi. 2003. Cloning and sequence analysis of endoglucanase genes from an industrial fungus, *Aspergillus kawachii*. *Bioscience, Biotechnology, and Biochemistry*, **67**(9), 2010–2013.

Harris, Paul V., Welner, Ditte, McFarland, K. C., Re, Edward, Navarro Poulsen, Jens-Christian, Brown, Kimberly, Salbo, Rune, Ding, Hanshu, Vlasenko, Elena, Merino, Sandy, Xu, Feng, Cherry, Joel, Larsen, Sine, & Lo Leggio, Leila. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase Family 61: Structure and function of a large, enigmatic family. *Biochemistry*, **49**(15), 3305–3316. PMID: 20230050.

Healthcare, GE. 2006. *Hydrophobic interaction and reversed phase chromatography, Principles and methods*. GE Healthcare Bio-Sciences AB.

Healthcare, GE. 2010. *Ion exchange chromatography & chromatofocusing: Principles and methods*. GE Healthcare Bio-Sciences AB.

Hemsworth, Glyn R., Taylor, Edward J., Kim, Robbert Q., Gregory, Rebecca C., Lewis, Sally J., Turkenburg, Johan P., Parkin, Alison, Davies, Gideon J., & Walton, Paul H. 2013. The copper active site of CBM33 polysaccharide oxygenases. *Journal of the American Chemical Society*, **135**(16), 6069–6077.

Himmel, Michael E., Ding, Shi-You, Johnson, David K., Adney, William S., Nimlos, Mark R., Brady, John W., & Foust, Thomas D. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, **315**(February), 804–807.

Horn, Svein J., & Eijsink, Vincent G. H. 2004. A reliable reducing end assay for chito-oligosaccharides. *Carbohydrate Polymers*, **56**(1), 35–39.

Horn, Svein J., Vaaje-Kolstad, Gustav, Westereng, Bjørge, & Eijsink, Vincent GH. 2012. Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels*, **5**(45).

Imhoff, Marc L., Bounoua, Lahouari, Ricketts, Taylor, Loucks, Colby, Harriss, Robert, & Lawrence, William T. 2004. Global patterns in human consumption of net primary production. *Nature*, **429**, 870–873.

Invitrogen. 2010 (August). *PichiaPink™Expression System. User Manual*. Invitrogen.

Isaksen, Trine, Westereng, Bjørge, Aachmann, Finn L., Agger, Jane W., Kracher, Daniel,

Ludwig, Roland, Haltrich, Dietmar, Eijsink, Vincent G. H., & Horn, Svein J. 2013. A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *The Journal of Biological Chemistry*, **289**(December), 2632–2642.

Karkehabadi, Saeid, Hansson, Henrik, Kim, Steve, Piens, Kathleen, Mitchinson, Colin, & Sandgren, Mats. 2008. 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), 144 – 154.

Karlsson, Johan, Saloheimo, Markku, Siika-aho, Matti, Tenkanen, Maija, Penttila, Merja, & Tjerneld, Folke. 2001. Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*. *European Journal of Biochemistry*, **268**(24), 6498–6507.

Kim, Seonah, Ståhlberg, Jerry, Sandgren, Mats, Paton, Robert S., & Beckham, Gregg T. 2014. Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proceedings of the National Academy of Sciences*, **111**(1), 149–154.

Kim, Seungdo, & Dale, Bruce E. 2005. Environmental aspects of ethanol derived from no-tilled corn grain: Nonrenewable energy consumption and greenhouse gas emissions. *Biomass and Bioenergy*, **28**, 475–489.

Kipper, Kalle, Väljamäe, Priit, & Johansson, Gunnar. 2005. Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as "burst" kinetics on fluorescent polymeric model substrates. *Biochemical Journal*, **385**, 527–535.

Kittl, Roman, Kracher, Daniel, Burgstaller, Daniel, Haltrich, Dietmar, & Ludwig, Roland. 2012. Production of four *Neurospora crassa* lytic polysaccharide monooxygenases in *Pichia pastoris* monitored by a fluorimetric assay. *Biotechnology for Biofuels*, **5**(1), 79.

Klein-Marcuschamer, Daniel, Oleskowicz-Popiel, Piotr, Simmons, Blake A., & Blanch, Harvey W. 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering*, **109**(4), 1083–1087.

Klyosov, Anatole A. 1990. Trends in biochemistry and enzymology of cellulose degradation. *Biochemistry*, **29**(47).

Koseki, Takuya, Mese, Yuichiro, Fushinobu, Shinya, Masaki, Kazuo, Fujii, Tsutomu, Ito, Kiyoshi, Shiono, Yoshihito, Murayama, Tetsuya, & Iefuji, Haruyuki. 2008. Biochemical characterization of a glycoside hydrolase family 61 endoglucanase from *Aspergillus kawachii*. *Applied Microbiology and Biotechnology*, **77**, 1279–1285.

Langston, James A., Shaghasi, Tarana, Abbate, Eric, Xu, Feng, Vlasenko, Elena, & Sweeney, Matt D. 2011. Oxidoreductive cellulose depolymerization by the enzymes

cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and Environmental Microbiology*, **77**(19), 7007–7015.

Levasseur, Anthony, Drula, Elodie, Lombard, Vincent, Coutinho, Pedro, & Henrissat, Bernard. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, **6**(1), 41.

Li, Xin, Beeson, William T., Phillips, Christopher M., Marletta, Michael A., & Cate, Jamie H. D. 2012. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure*, **20**(6), 1051–1061.

Liska, Adam J., Yang, Haishun S., Bremer, Virgil R., Klopfenstein, Terry J., Walters, Daniel T., Erickson, Galen E., & Cassman, Kenneth G. 2009. Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol. *Journal of Industrial Ecology*, **13**(1), 58–74.

Lombard, Vincent, Golaconda Ramulu, Hemalatha, Drula, Elodie, Coutinho, Pedro M., & Henrissat, Bernard. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**(D1), D490–D495.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, **66**(3), 506–577.

Lynd, Lee R., & Woods, Jeremy. 2011. Perspective: A new hope for Africa. *Nature*, **474**, S20–S21.

Lynd, Lee R., Wyman, Charles E., & Gerngross, Tillman U. 1999. Biocommodity engineering. *Biotechnology Progress*, **15**(5), 777–793.

Mattinen, Maija-Liisa, Linder, Markus, Teleman, Anita, & Annila, Arto. 1997. Interaction between cellohexaose and cellulose binding domains from *Trichoderma reesei* cellulases. *FEBS Letters*, **407**(3), 291–296.

Miller, Gail Lorenz. 1959. Use of dinitrosaIicyIic acid reagent for determination of reducing sugar. *Analytical Chemistry*, **31**, 426–428.

Mosier, Nathan, Wyman, Charles, Dale, Bruce, Elander, Richard, Lee, Y.Y., Holtzapple, Mark, & Ladisch, Michael. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, **96**(6), 673 – 686.

Mozaffarian, M., Zwart, R.W.R., Boerrigter, H., Deurwaarder, E.P., & Kersten, S.R.A. 2004. "Green gas" as SNG (synthetic natural gas): A renewable fuel with conventional quality. *Contribution to the "Science in Thermal and Chemical Biomass Conversion" Conference*, August.

Payne, Christina M., Resch, Michael G., Chen, Liqun, Crowley, Michael F., Himmel, Michael E., Taylor, Larry E., Sandgren, Mats, Ståhlberg, Jerry, Stals, Ingeborg, Tan, Zhongping, & Beckham, Gregg T. 2013. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceedings of*

the National Academy of Sciences, **110**(36), 14646–14651.

Payne, Christina M., Knott, Brandon C., Mayes, Heather B., Hansson, Henrik, Himmel, Michael E., Sandgren, Mats, StÃ¥hlberg, Jerry, & Beckham, Gregg T. 2015. Fungal cellulases. *Chemical Reviews*, **115**(3), 1308–1448. PMID: 25629559.

Petersen, Thomas Nordahl, Brunak, Søren, von Heijne, Gunnar, & Nielsen, Henrik. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**, 785–786.

Phillips, Christopher M., Beeson, William T., Cate, Jamie H., & Marletta, Michael A. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chemical Biology*, **6**(12), 1399–1406. PMID: 22004347.

Pricelius, S., Ludwig, R., Lant, N., Haltrich, D., & Guebitz, G. M. 2009. Substrate specificity of *Myriococcum thermophilum* cellobiose dehydrogenase on mono-, oligo-, and polysaccharides related to in situ production of H_2O_2. *Applied microbiology and Biotechnology*, **85**, 75–83.

Quinlan, R. Jason, Sweeney, Matt D., Lo Leggio, Leila, Otten, Harm, Poulsen, Jens-Christian N., Johansen, Katja Salomon, Krogh, Kristian B. R. M., Jørgensen, Christian Isak, Tovborg, Morten, Anthonsen, Annika, Tryfona, Theodora, Walter, Clive P., Dupree, Paul, Xu, Feng, Davies, Gideon J., & Walton, Paul H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences*, **108**(37), 15079–15084.

Reese, Elwyn T., Siu, Ralph G. H., & Levinson, Hillel S. 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of Bacteriology*, **59**(4), 485–497.

Rubin, Edward M. 2008. Genomics of cellulosic biofuels. *Nature*, **454**(August), 841–845.

Sakon, Joshua, Irwin, Diana, Wilson, David B., & Karplus, P. Andrew. 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology*, **4**, 810–818.

Saloheimo, M., Nakari-Setälä, T., Tenkanen, M., & Penttilä, M. 1997. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *European Journal of Biochemistry*, **249**(2), 584–591.

Sartorius Stedim Biotech, Gmbh. 2014 (March). *Directions for Use. Vivaflow 50/50R/200*.

Schrödinger, LLC. 2010 (August). *The PyMOL Molecular Graphics System, Version 1.3r1*.

Searchinger, Timothy, Heimlich, Ralph, Houghton, R.A., Dong, Fengxia, Elobeid, Amani, Fabiosa, Jacinto, Tokgoz, Simla, Hayes, Dermot, & Yu, Tun-Hsiang. 2008.

Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, **319**(5867), 1238–1240.

Segato, Fernando, Damasio, Andre R, Goncalves, Thiago, Murakami, Mario, Squina, Fabio, Polizeli, MariadeLourdesTM, Mort, Andrew, & Prade, Rolf. 2012. Two structurally discrete GH7-cellobiohydrolases compete for the same cellulosic substrate fiber. *Biotechnology for Biofuels*, **5**(1), 21.

Shapouri, H., Gallagher, Paul W., Nefstead, Ward, Schwartz, Rosalie, Noe, Stacey, & Conway, Roger. 2010. *2008 energy balance for the corn-ethanol industry*. Tech. rept. United States Department of Agriculture.

Sievers, Fabian, Wilm, Andreas, Dineen, David, Gibson, Toby J, Karplus, Kevin, Li, Weizhong, Lopez, Rodrigo, McWilliam, Hamish, Remmert, Michael, Söding, Johannes, Thompson, Julie D, & Higgins, Desmond G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**(1).

Sigrist, CJA, de Castro, E, Cerutti, L, Cuche, BA, Hulo, N, Bridge, A, Bougueleret, L, & Xenarios, I. 2012. New and continuing developments at PROSITE. *Nucleic Acids Research*, **10**.

Srisodsuk, M, Reinikainen, T, Penttilä, M, & Teeri, T T. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *Journal of Biological Chemistry*, **268**(28), 20756–61.

Stålbrand, Henrik, Mansfield, Shawn D., Saddler, John N., Kilburn, Douglas G., Warren, R. Antony J., & Gilkes, Neil R. 1998. Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β-1,4-glucanases. *Applied and Environmental Microbiology*, **64**(7), 2374–2379.

Steentoft, Catharina, Vakhrushev, Sergey Y., Joshi, Hiren J., Kong, Yun, Vester-Christensen, Malene B., Schjoldager, Katrine T-B G., Lavrsen, Kirstine, Dabelsteen, Sally, Pedersen, Nis B., Marcos-Silva, Lara, Gupta, Ramneek, Bennett, Eric Paul, Mandel, Ulla, Brunak, Søren, Wandall, Hans H., Levery, Steven B., & Clausens, Henrik. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The Embo Journal*, **32**(10), 1478–1488.

Tait, Joyce, Adcock, Mike, Barker, Guy C., Caney, Simon, Chataway, Joanna, Gill, Robin, Hutton, Jon, Leyser, Ottoline, Mortimer, Nigel, Raines, Christine, Smale, Ian, & Watson, Jim. 2011 (April). *Biofuels: Ethical issues*. Tech. rept. Nuffield Council on Bioethics.

Taylor, Courtney B., Talib, M. Faiz, McCabe, Clare, Bu, Lintao, Adney, William S., Himmel, Michael E., Crowley, Michael F., & Beckham, Gregg T. 2012. Computational investigation of glycosylation effects on a family 1 carbohydrate-binding module.

Journal of Biological Chemistry, **287**(5), 3147–3155.

Tuck, Christopher O., Pérez, Eduardo, Horváth, István T., Sheldon, Roger A., & Poliakoff, Martyn. 2012. Valorization of biomass: Deriving more value from waste. *Science*, **337**(6095), 695–699.

Vaaje-Kolstad, G, Houston, DR, Riemen, AH, Eijsink, VG, & van Aalten, DM. 2005a. Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *Journal of Biological Chemistry*, **280**(12), 11313–11319.

Vaaje-Kolstad, Gustav, Horn, Svein J., van Aalten, Daan M. F., Synstad, Bjørnar, & Eijsink, Vincent G. H. 2005b. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *Journal of Biological Chemistry*, **280**(31), 28492–28497.

Vaaje-Kolstad, Gustav, Westereng, Bjørge, Horn, Svein J., Liu, Zhanliang, Zhai, Hong, Sørlie, Morten, & Eijsink, Vincent G. H. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, **330**(6001), 219–222.

Vaaje-Kolstad, Gustav, BÃžhle, Liv Anette, GÃ¥seidnes, Sigrid, Dalhus, BjÃžrn, BjÃžrÃ¥s, Magnar, Mathiesen, Geir, & Eijsink, Vincent G.H. 2012. Characterization of the Chitinolytic Machinery of *Enterococcus faecalis* {V583} and High-Resolution Structure of Its Oxidative {CBM33} Enzyme. *Journal of Molecular Biology*, **416**(2), 239 – 254.

van Wyk, Jacobus P.H. 2001. Biotechnology and the utilization of biowaste as a resource for bioproduct development. *Trends in Biotechnology*, **19**(5), 172–177.

Várnai, Anikó, Tang, Campbell, Bengtsson, Oskar, Atterton, Andrew, Mathiesen, Geir, & Eijsink, Vincent G. H. 2014. Expression of endoglucanases in *Pichia pastoris* under control of the GAP promoter. *Microbial Cell Factories*, **13**(57).

Vu, Van V., Beeson, William T., Phillips, Christopher M., Cate, Jamie H. D., & Marletta, Michael A. 2014a. Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *Journal of the American Chemical Society*, **136**(2), 562–565. PMID: 24350607.

Vu, Van V., Beeson, William T., Span, Elise A., Farquhar, Erik R., & Marletta, Michael A. 2014b. A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences*, **111**(38), 13822–13827.

Westereng, Bjørge, Ishida, Takuya, Vaaje-Kolstad, Gustav, Wu, Miao, Eijsink, Vincent G. H., Igarashi, Kiyohiko, Samejima, Masahiro, Ståhlberg, Jerry, Horn, Svein J., & Sandgren, Mats. 2011. The Putative Endoglucanase PcGH61D from <italic>Phanerochaete chrysosporium</italic> Is a Metal-Dependent Oxidative Enzyme that Cleaves Cellulose. *PLoS ONE*, **6**(11), e27807.

Wikimedia. 2015 (April). *Retrieved from* `http://upload.wikimedia.org/`

wikipedia/commons/b/bc/Structur_der_Cellobiose.

Wu, Miao, Beckham, Gregg T., Larsson, Anna M., Ishida, Takuya, Kim, Seonah, Payne, Christina M., Himmel, Michael E., Crowley, Michael F., Horn, Svein J., Westereng, Bjørge, Igarashi, Kiyohiko, Samejima, Masahiro, Ståhlberg, Jerry, Eijsink, Vincent G. H., & Sandgren, Mats. 2013. Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the Basidiomycota fungus *Phanerochaete chrysosporium*. *Journal of Biological Chemistry*, **288**(18), 12828–12839.

Wyman, Charles E., Spindler, Diane D., & Grohmann, Karel. 1992. Simultaneous saccharification and fermentation of several lignocellulosic feedstocks to fuel ethanol. *Biomass and Bioenergy*, **3**(5), 301–307.

You, Chun, Chen, Hongge, Myung, Suwan, Sathitsuksanoh, Noppadon, Ma, Hui, Zhang, Xiao-Zhou, Li, Jianyong, & Zhang, Y.-H. Percival. 2013. Enzymatic transformation of nonfood biomass to starch. *Proceedings of the National Academy of Sciences*, **110**(18), 7182–7187.

Zeeman, Samuel C., Kossmann, Jens, & Smith, Alison M. 2010. Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annual Review of Plant Biology*, **61**, 209–234.

Zhang, Y.-H Percival. 2013. Next generation biorefineries will solve the food, biofuels, and environmental trilemma in the energy-food-water nexus. *Energy Science & Engineering*, **1**(1), 27–41.

Zhu, J.Y., Pan, X.J., Wang, G.S., & Gleisner, R. 2009. Sulfite pretreatment (SPORL) for robust enzymatic saccharification of spruce and red pine. *Bioresource Technology*, **100**(8), 2411 – 2418.

A. Amino acid sequences

The amino acid sequences for the three proteins from *M. thermophilum* are shown below together with *Nc*LPMO9C from *N. crassa*. Signal peptides are in green, AA9 domains in red and the CBM1 domain in blue. The linker between the AA9 domain and the CBM1 domain is in magenta. Sources: Pfam, http://pfam.xfam.org/search/sequence for the domains and SignalP for signal peptide predictions. The C-terminal border for the AA9 domain for *Mt*4260 was determined exactly by multiple sequence alignments, see Section 3.2.1.

***Mt*358** MLTTTLALLT AALGASAHYT LPRVGSGSDW QHVRQADNWQ NNGFVGDVNS QQIRCFQSSP SGAPEVYSVQ AGSTVTYHAN PNIYHPGPMQ FYLARVPDGQ DVKSWTGDGA VWFKVYEEQP NFGAQLTWPS NGKSSFEVPI PSCIPAGNYL LRAEHIALHV AQSEGGAQFY ISCAQLEVTG GGSTEPSQKV AFPGAYSSTD PGILININYP VPTSYQNPGP AVFSC

225 amino acids. Signal peptide 1-17, AA9 domain 18-217.

***Mt*4260** MKPFSLIALA TAVSGHTIFQ RVSVNGQDQG QLKGVRAPSS NFPIQDVNDA NMACNAVVYK DNTIIKVPAG ARVGTWWQHV IGGPQGAFDP DNPIAASHKG PISVYLAKVD NAATASPSGL QWFKVAERGL NNGVWAVDEL IANDGWHYFE MPSCVAPGQY LMRVELLALH SASSPGGAQF YIGCAQIEVT GSGTHSGSDF VSFPGAYSTN DPGIVISIYD TTGNPNNGGR PYPIPGPRPI SCSGSGGNNG GGNDNGGGNN NGGGSSGGSV PLYGGQCGGIG YTGPTSCAEG TCTATNEYYS QCLP

304 amino acids. Signal peptide 1-15, AA9 domain 16-242, CBM1 domain 272-299.

***Mt*6403** MSSFTSKGFL SALMGAAMVA AHGHVTNIVI NGVSYQGFDP FSHSYMQNPP TVVGWTASNT DNGFVDPASF SSPDIICHKS ATNAGGHAVV AAGDKIFIQW DTWPESHHGP VIDYLADCGD EGCEKVDKTT LKFFKISEKG LIDGTNAPGK WASDELIANG NSWMVQIPSD IAPGNYVLRH EIIALHSAGQ QNGAQNYPQC FNLQVTGSGT QKPSGVLGTE LYKADDAGIL ANIYTSPVTY EIPGPAIIPG ASAVE-QTTSA ITSSASAITG SATAAPPAAT TTAAPTTTTA GSDVTATPST GGSDSSAQPT ATATSSTPRP TRCPGLKKRR RHARDVRVSL

340 amino acids. Signal peptide 1-23, AA9 domain 24-242.

***Nc*LPMO9C** MKTGSILAAL VASASAHTIF QKVSVNGADQ GQLKGIRAPA NNNPVT-
DVMSSDIICNAVTM KDSNVLTVPA GAKVGHFWGH EIGGAAGPND ADNPIAASHKG-
PIMVYLAKV DNAATTGTSG LKWFKVAEAG LSNGKWAVDD LIANNGWSYFDMPT-
CIAPGQ YLMRAELIAL HNAGSQAGAQ FYIGCAQINV TGGGSASPSN TVSF-
PGAYSA SDPGILINIY GGSGKTDNGG KPYQIPGPAL FTCPAGGSGG SSPAP-
ATTAS TPKPTSASAP KPVSTTASTP KPTNGSGSGT GAAHSTKCGG SKPAAT-
TKAS NPQPTNGGNS AVRAAALYGQ CGGKGWTGPT SCASGTCKFSNDWYSQCLP

359 amino acids. Signal peptide 1-16, AA9 domain 17-235, CBM1 domain 326-354

B. Plasmid maps

Mt358, Mt4260 and Mt6403 were cloned into the pPICZB vector by Roman Kittl et al. and the plasmids with inserts were isolated from E. coli NEB5alpha. The pPICZB vector map is shown in Figure B.1.

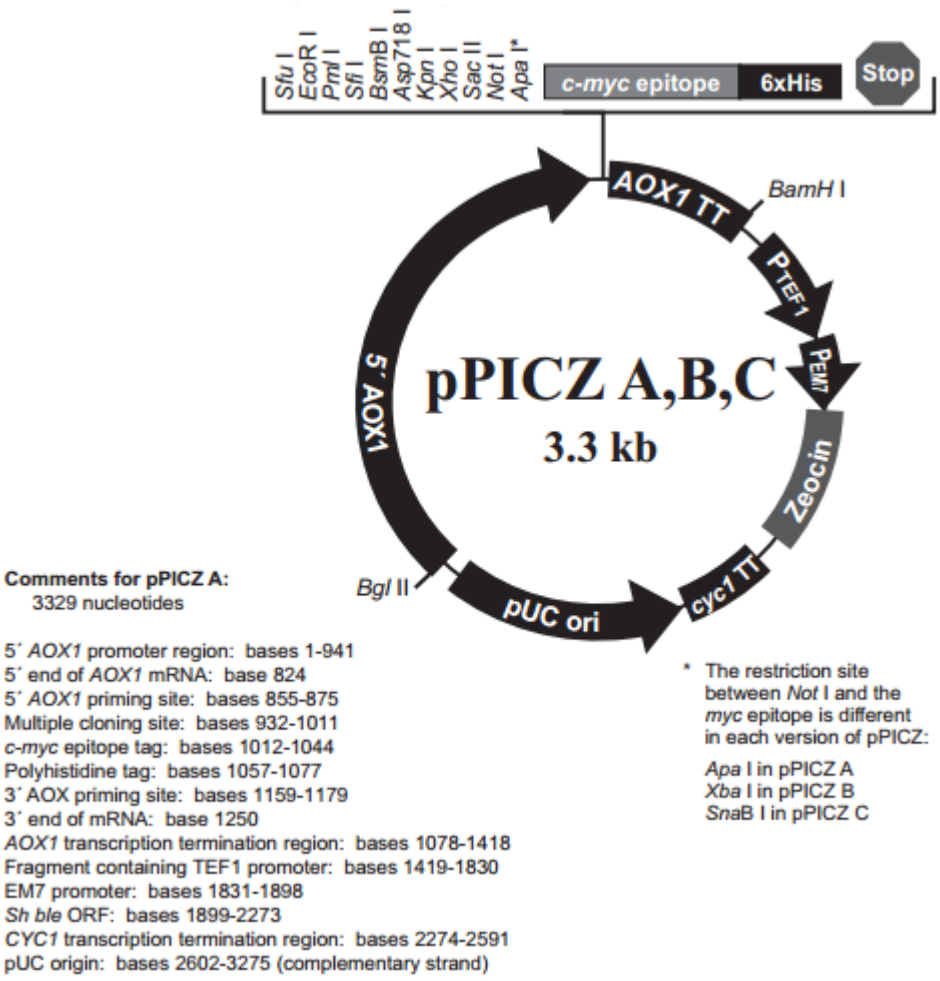

Figure B.1: pPICZ A, B and C vector from Invitrogen (http://tools.lifetechnologies.com/content/sfs/manuals/ppicz_man.pdf). The restriction sites BstBI and XbaI were used for cloning of the three *M. thermophilum* genes.

pPink-GAP-HC (Invitrogen) with inserted *mt4260* is shown in Figure B.2. The high copy expression vector is based on the pUC19 vector and contains ampicillin resistance

gene, the *Pichia pastoris* ADE2 gene, the GAP promoter driving the expression of the desired gene product, the TRP2 gene, pUC ori and CYC1 transcription termination region. It also contains a multiple cloning site with restriction sites *Eco*RI, *Rsr*II, *Sph*I, *Stu*I, *Kpn*1, *Fse*I, *Nae*I and *Swa*I. This vector was used for expression in Pichia Pink™. Transformation of PichiaPink™ is done by integration of the vector into the TRP2 locus of the yeast's chromosomal DNA

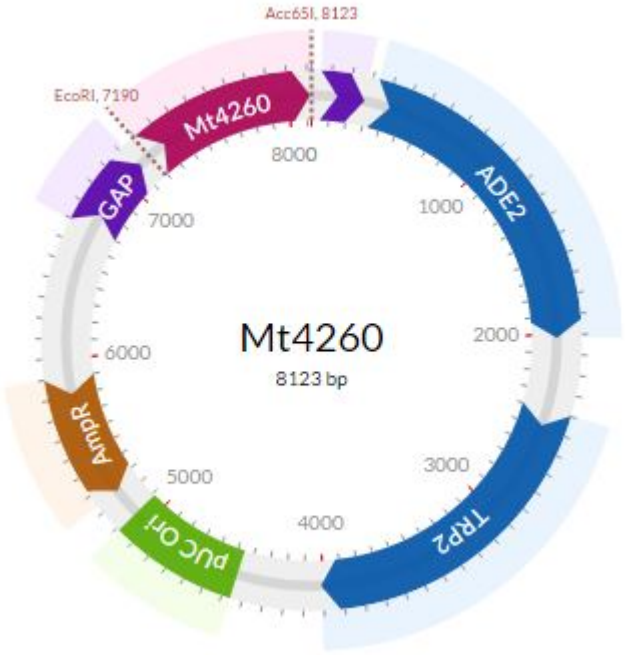

Figure B.2: Plasmid map of pPink-GAP-HC with *mt4260*. The graphic is made with Angular Plasmid Chawdry (2015).

C. Purification with HIC and IEC

Figures C.1 shows chromatograms from the purification of *Mt*4260 samples by HIC and IEC and LDS-PAGE gels of the purification fractions. Figure C.2 shows purification of a *Mt*4260-N sample by HIC, with chromatogram and LDS-PAGE gel.

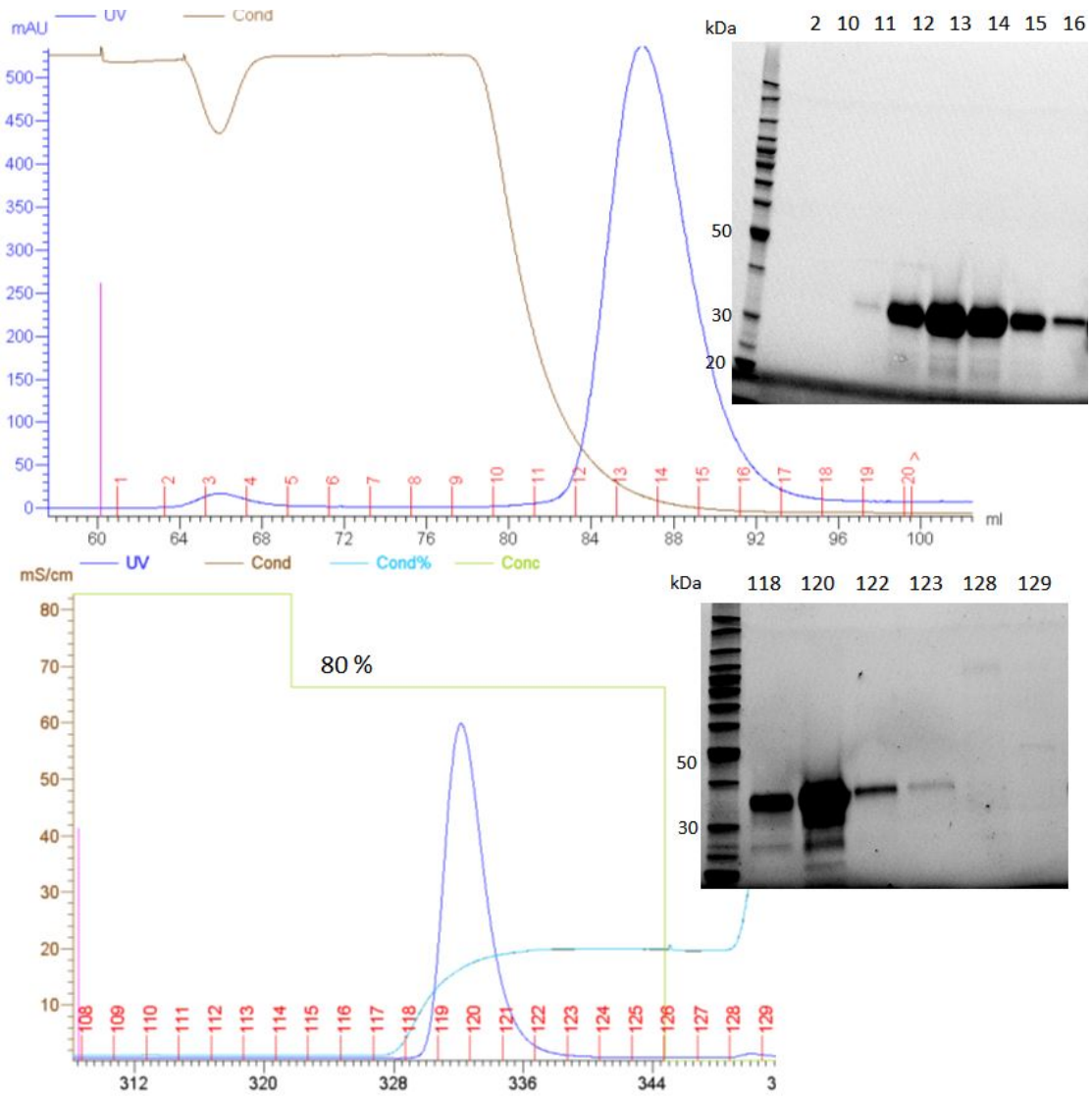

Figure C.1: *Upper:* HIC of *Mt*4260 with a gradient elution. The protein elutes in one relatively pure peak around 88 ml, as is illustrated in the picture from the gel where fractions 2 and 10-16 are run. The small flowthrough peak at about 3.5 ml does not seem to contain any of the protein. *Below*: Stepwise IEC purification of *Mt*4260 with the protein eluting at 80 % start buffer.

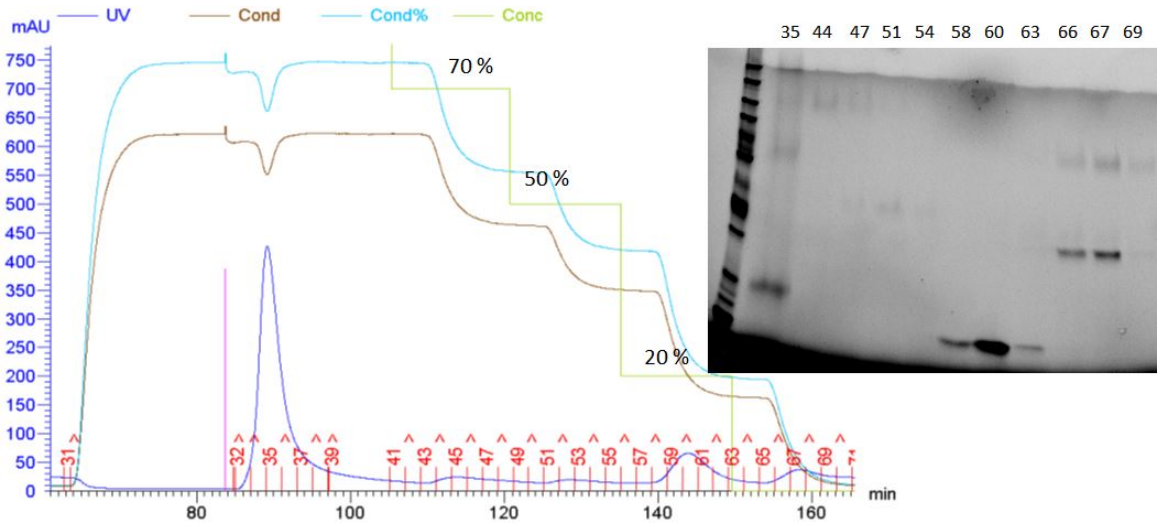

Figure C.2: HIC of *Mt*4260-N with a stepwise elution of 100 %, 70 %, 50 %, 20 % and 0 % start buffer. *Mt*4260-N elutes at 20 % start buffer, so the final stepwise elution had steps of 20 and 0 % start buffer.

D. Chromatograms from characterization

Given below are chromatograms used to find substrate profiles of the three enzymes. Only the ones with detected products are shown.

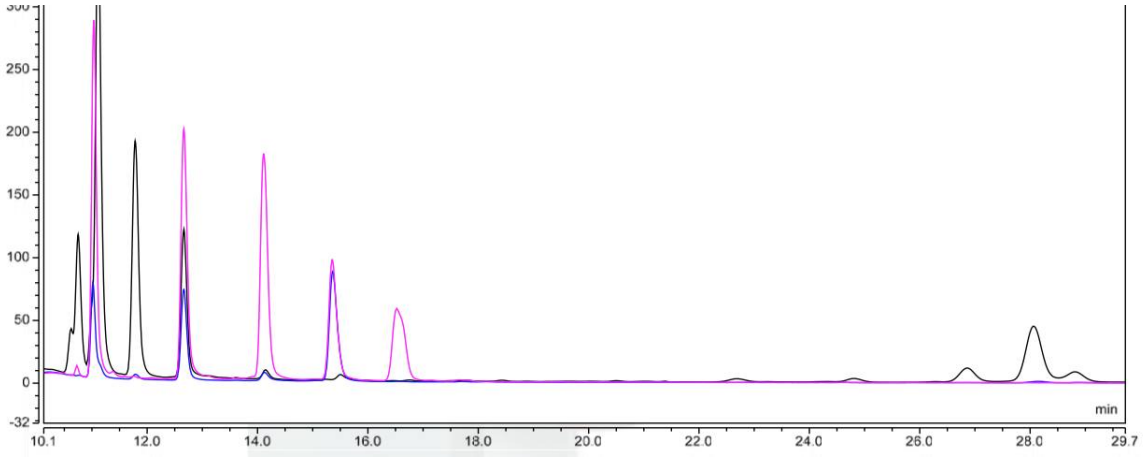

Figure D.1: HPAEC chromatogram from the reaction of 4 µM Mt4260 with 1 mg/ml Glc5 (glucose pentamer) incubated at 50 ºC ON in 20 mM sodium acetate buffer pH 5.6, shown in black. In blue a control without ascorbic acid is shown, and this does not contain the peaks indicating C4 oxidized products. It also contains a larger amount of Glc5, indicating that it was not used in reactions. A standard is shown in pink and consists of 0.1 mg/ml glucose oligosaccharides with DP 2-6. A peak around 28 minutes is diagnostic for a C4 oxidized product. At about 27 minutes a product eluted that could be ascorbic acid as it appears in samples containing this reductant without enzyme (not shown). A shorter elution gradient was used than in other analyses, which makes elution times different than in Figure 3.12 and Figure D.3.

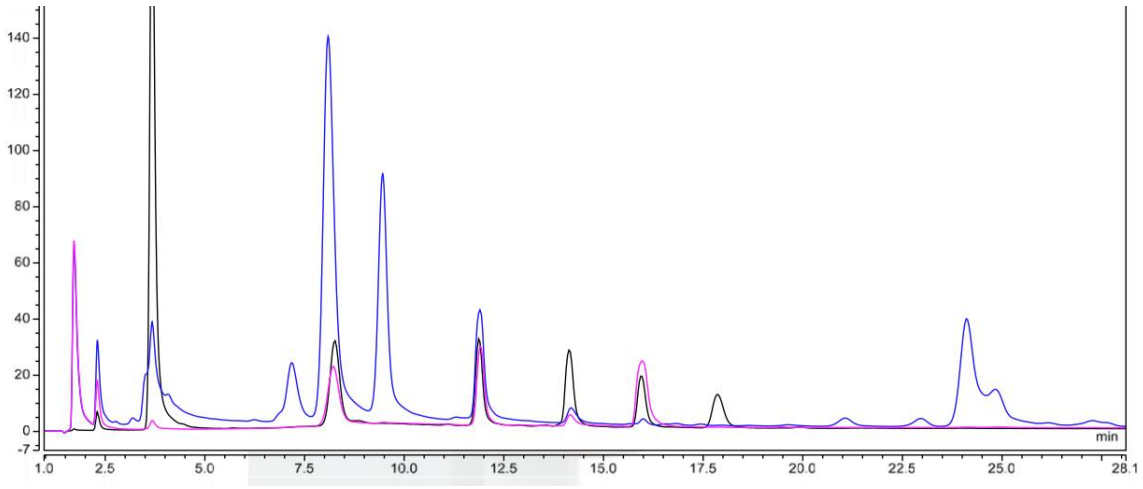

Figure D.2: HPAEC chromatogram from the reaction of 4 µM *Mt*4260-N with 1 mg/ml Glc5 (glucose pentamer) incubated at 50 ºC ON in 20 mM sodium acetate buffer pH 5.6, shown in blue. In pink a control without ascorbic acid is shown, and this does not contain the peaks indicating C4 oxidized products. It also contains a larger amount of Glc5, indicating that it was not used in reactions. A standard is shown in black and consists of 0.05 mg/ml glucose oligosaccharides with DP 1-6. A peak around 24-25 minutes is diagnostic for a C4 oxidized product. At about 23 minutes a product eluted that could be ascorbic acid as it appears in samples containing this reductant without enzyme (not shown). A shorter elution gradient was used than in other analyses, which makes elution times different than in Figure 3.12 and Figure D.3.

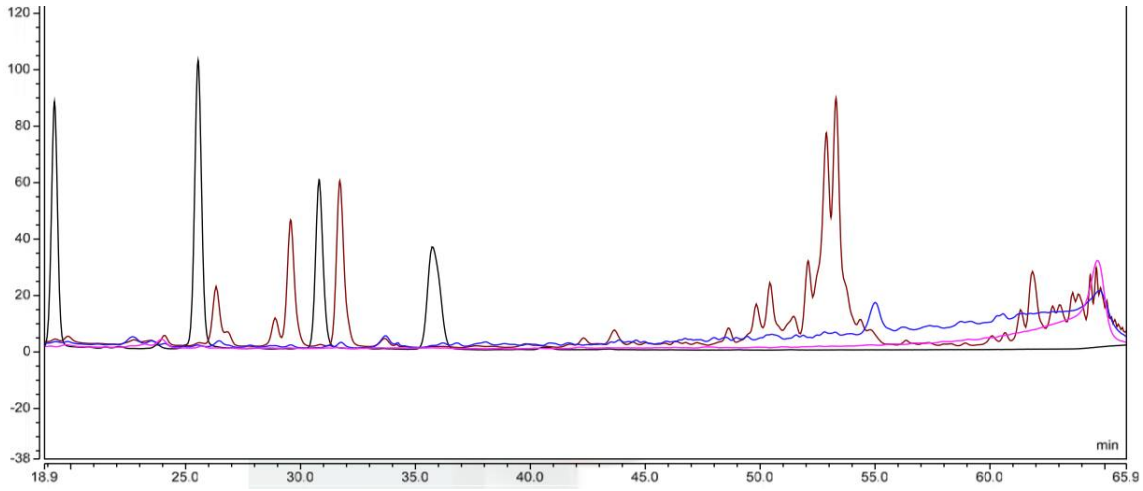

Figure D.3: HPAEC chromatogram showing products generated from 5 mg/ml tamarind xyloglucan incubated at 50 ºC for 18 hours with an enzyme concentration of 30 mg enzyme per g of substrate, 2 mM ascorbic acid and 5 mM sodium acetate pH 5.6. The reaction with *Mt*4260 is colored brown and shows a cluster diagnostic for C4 oxidized products from 48-55 minutes, while *Mt*6403 is colored blue and shows much smaller peaks around the same retention time. A negative control without enzymes is shown in pink and a standard with 0.1 M of cellooligosaccharides DP 3-6 is shown in black.