Norwegian University
of Life Sciences

Master's Thesis 2017 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science (KBM)

Characterization of cellulases involved in a novel cellulolytic mechanism linked to the Bacteroidetes Type IX Secretion System

Ingrid Heggenes
Biotechnology

Characterization of cellulases involved in a novel cellulolytic mechanism linked to the Bacteroidetes Type IX Secretion System

Master's Thesis 2017

Ingrid Heggenes

Faculty of Chemistry, Biotechnology and Food Science

Norwegian University of Life Sciences

Acknowledgements

The present work was carried out at the Faculty of Chemistry, Biotechnology and Food Science at the Norwegian University of Life Sciences (NMBU) from August 2016 to May 2017, with Adrian Naas, Phil Pope and Vincent Eijsink as supervisors.

First of all, I would like to thank my great main supervisor Adrian for guiding me in all my lab work, always answering my questions and for your generally positive perspective. It has made my master's year a whole lot easier. I am grateful.

Thank you to Phil for support, supervision and encouragement, and for letting me take part in this project. Also, I would like to thank Vincent for letting me write my master's thesis in the Protein Engineering and Proteomics (PEP) group. In addition, thank you for always setting aside time to answer my questions and for your supervision.

Thank you to Magnus and Tina, for your help with the metaproteomics work and Magnus for all your help with the ICS and MS. Thanks also to the rest of the PEP group, everyone has been so helpful and nice during my year here.

Finally, I would like to thank my family, friends and boyfriend for your support, encouragement and love. And my fellow master's students, Heidi and Bjørnar, thank you for fruitful discussions, mutual support and for sharing numerous iced coffees with me. I could not have done this without you.

Ås, May 2017

Ingrid Heggenes

Abstract

Lignocellulose is the most abundant biomass on earth and has a great potential as a source for sustainable production of valuable chemicals and biofuels. Today's depletion of fossil fuel reserves and pollution from their usage creates a need for a more sustainable source of energy. Lignocellulosic biomass is generally regarded as a sustainable and environmentally friendly energy source. However, the production of fuels from biomass is not sufficiently effective to compete fully with fossil fuels on economic terms, mostly due to the low efficiency of the enzymes needed to degrade the biomass. Thus, the need for better enzymes for the conversion of biomass to sugars and eventually biofuels is evident, and a better understanding of how biomass conversion occurs in nature may lead to improvements in the technology.

This thesis was based on a metagenomic study of a cow rumen microbiome sample enriched on switchgrass, where further investigation revealed a putative cellulolytic gene cluster, from a Bacteroidetes-affilated phylotype genome (AGa), with four putative cellulolytic glycoside hydrolase 5 (GH5) family enzymes. The cluster contained both multi-domain enzymes and enzymes with a C-terminal secretion tag for the Type IX secretion system (T9SS). The T9SS has been linked to cellulose degradation in Bacteroidetes. The aim of the present study was to clone, express and purify these GH5 enzymes for biochemical characterization to gain insight into their substrate targets and functional roles in the cluster.

The enzymes, together with truncated versions containing only one domain, were cloned, expressed and purified. Some of the enzymes gave inclusion bodies upon expression and had to be expressed with fusion tags. Further, the enzymes were characterized, through enzymatic assays, in terms of pH and temperature optima, temperature stabilities, substrate specificities, product profiles and cellodextrin cleaving patterns. Enzyme activity was analysed by the dinitrosalicylic acid (DNS) method, high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and high-performance liquid chromatography/mass spectrometry (HPLC-MS). Additionally, the binding to cellulose by a putative carbohydrate-binding domain (CBM) domain from one of the cellulases was investigated, and crystallization trials were executed on an interesting, potentially mutated GH5 domain. Finally, metaproteomics was performed on a cow rumen sample in an attempt to reveal the presence of the enzymes *in vivo*.

Characterization revealed cellulase activity on crystalline cellulose, filter paper, for almost all GH5 enzymes present in the cluster. The enzymes had optimal activity at pH 5.0-6.0 and temperature 50-60 ºC, and temperature stability up to 40 ºC on carboxymethyl cellulose (CMC). The enzymes were active on various substrates with β-1,4-glucosidic linkages in the backbone and on the lignocellulosic substrate switchgrass. Product analysis showed release of mostly cellobiose from cellulosic substrates, and cellodextrin assays showed cleaving of cellopentaose to cellobiose and cellotriose, and cellohexaose to cellobiose, cellotriose and partly cellotetraose for the majority of the enzymes. One GH5 displayed a different cleaving pattern on cellohexaose, cleaving it to solely cellotriose. Combining the enzymes in cocktails gave no synergistic effects. The weakly annotated CBM domain in one of the enzymes was shown to bind to the cellulosic substrate Avicel, indicating a correct annotation as a CBM. Crystallization efforts for one of the domains yielded no diffraction quality crystals and the metaproteomic investigation could not detect peptides specific for the AGa genome.

In conclusion, this thesis describes the characterisation of enzymes from a newly discovered cellulolytic gene cluster, revealing varying activities and substrate preferences, indicating that these enzymes have complementary roles in lignocellulose degradation. The cluster enzyme characterization gives insight into the use of the Type IX secreted multi-domain cellulases. Further work on the gene cluster can provide more insight into the degradation mechanism, and expression of the cluster in a bacterium harbouring the Type IX secretion system could be pursued to improve the expression of the multi-domain cellulases that were difficult to express. Ideally, the isolation of a representative of the AGa phylotype would shed light on its true involvement in cellulose degradation. Collectively, the work presented in this thesis contributes to the understanding of biomass degradation by anaerobic-bacteria, which can help improving the industrial conversion of biomass to valuable products in the future.

Sammendrag

Lignocellulose er den mest forekommende biomassen på jorden og har et stort potensiale som kilde til bærekraftig produksjon av verdifulle produkter som kjemikalier og biodrivstoff. Dagens uttømming av fossile brennstoffreserver og forurensning fra bruken av disse viser behovet for en mer bærekraftig kilde til energi. Lignocellulosisk biomasse blir sett på som en bærekraftig og miljøvennlig energikilde, men produksjonen av for eksempel biodrivstoff fra biomasse er ikke effektiv nok til å fullt ut kunne konkurrere med fossilt brennstoff på økonomiske vilkår, mest på grunn av lite effektive enzymer. Det er derfor et stort behov for bedre enzymer i omgjøringen av biomasse til sukker og tilslutt biodrivstoff, og en bedre forståelse av hvordan dette foregår i naturen kan føre til forbedringer i denne teknologien.

Denne oppgaven ble basert på en metagenomisk studie av en switchgrass prøve fra kumage, hvor videre undersøkelser fant et antatt cellulolytisk genkluster, fra en Bacteroidetes fylotypes genom (AGa), med fire antatt cellulolytiske glykosid-hydrolase familie 5 (GH5) enzymer. Klusteret inneholdt både multi-domene enzymer og enzymer med et C-terminalt sekresjonssignal for Type IX sekresjonssystemet (T9SS). T9SS har blitt koblet til cellulosenedbryting i Bacteroidetes. Målet med denne studien var å klone, uttrykke og rense disse GH5 enzymene for biokjemisk karakterisering for å få innsikt i deres målsubstrater og funksjonelle roller i genklusteret.

Enzymene, i tillegg til avkortede versjoner av enzymene med kun ett domene, ble klonet, uttrykt og renset. Noen av enzymene aggregerte under uttrykking og måtte uttrykkes med «fusion tags». Videre ble enzymene karakterisert, gjennom enzymatiske assays, for å bestemme pH og temperatur optima, temperatur stabiliteter, substrat spesifisiteter, produktdannelse og spaltingsmønster på cellodextriner. Assayene ble analysert med dinitrosalisylsyre (DNS) metoden, høypresisjons-ionebytterkromatografi (HPAEC-PAD) og høypresisjons-væskekromatografi/masse spektrometri (HPLC-MS). I tillegg ble et antatt karbohydrat-bindende domene (CBM) fra en av cellulasene undersøkt for binding til cellulose, og krystalliserings forsøk ble utført på et interessant, potensielt mutert GH5 domene. Til slutt ble metaproteomikk utført på en prøve fra kumage i forsøk på å avsløre tilstedeværelsen av genkluster enzymer *in vivo.*

Karakteriseringen avdekte cellulase aktivitet på krystallinsk cellulose, filterpapir, for nesten alle GH5 enzymene i genklusteret. Enzymene hadde optimal aktivitet ved pH 5.0-6.0 og temperatur 50-60 ºC, og temperaturstabilitet opp til 40 ºC på karboksymetylcellulose (CMC). Enzymene var aktive på diverse substrater med β-1,4-glykosidbindinger i hovedkjeden og på det lignocellulosiske substratet switchgrass. Produktanalyse viste frigjøring av for det meste cellobiose fra cellulosiske substrater og cellodextrin assayer viste kløyving av cellopentaose til cellobiose og cellotriose, og cellohexaose til cellobiose, cellotriose og delvis cellotetraose for mesteparten av enzymene. Et av enzymene viste et annerledes spaltingsmønster på cellohexaose ved å kløyve det til kun cellotriose. Kombinering av enzymene i enzymcocktailer viste ingen synergistisk effekt. Det svakt annoterte CBM domenet i et av enzymene ble vist å binde til det cellulosiske substratet Avicel, som indikerer korrekt CBM annotering. Forsøket på å krystallisere et domenene i et av enzymene ga ingen krystaller med diffraksjonskvalitet og metaproteomikk undersøkelsene kunne ikke detektere peptider spesifikke for AGa genomet.

For å konkludere viser denne oppgaven karakterisering av nye enzymer fra et oppdaget cellulolytisk genkluster, som avslører varierende aktivitet og substratpreferanser for enzymene, og indikerer at enzymene har komplementære roller i nedbryting av lignocellulose. Karakteriseringen av enzymene i klusteret gir innsikt i bruken av Type IX-sekreterte multi-domene cellulaser. Fremtidig arbeid med genklusteret kan føre til mer innsikt i nedbrytingsmekanismen, og uttrykking av klusteret i en bakterie som innehar Type IX sekresjonssystemet kan utføres for å forbedre uttrykkingen av de multi-domene cellulasene som var vanskelige å uttrykke. Ideelt sett skulle en representant fra AGa fylotypen blitt isolert, noe som ville kastet lys på dens sanne involvering i cellulosenedbryting. Samlet sett bidrar arbeidet presentert i denne oppgaven til forståelsen av anaerobe bakterier sin biomassenedbryting, som kan hjelpe til å forbedre den industrielle omgjøringen av biomasse til verdifulle produkter i framtiden.

Abbreviations

$A_{280/540}$	Absorbance at 280/540 nanometres
ACN	Acetonitrile
AmBic	Ammonium bicarbonate
BSA	Bovine Serum Albumin
CAZymes	Carbohydrate-Active Enzymes
CBM	Carbohydrate-Binding Module
CMC	Carboxymethyl cellulose
dGTP	2'-deoxyguanosine 5'-triphosphate
DNS	3,5-dinitrosalicylic acid
DSS	Dissociation solution
DTT	1,4-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GH	Glycoside hydrolase
GH5	Glycoside hydrolase family 5
Glc_x	X numbers of repeating glucose units
HPAEC-PAD	High-Performance Anion-Exchange Chromatography with Pulsed Amperomeric Detection
HPLC-MS	High-Performance Liquid Chromatography/Mass Spectrometry
IAA	Iodoacetamide
ICS	Ion Chromatography System
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
LB	Lysogeny Broth
LDS-PAGE	Lithium dodecyl sulphate polyacrylamide gel electrophoresis

LPMO	Lytic polysaccharide monooxygenase
OD	Optical density
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
PUL	Polysaccharide Utilization Locus
rpm	Revolutions per minute
RT	Room temperature
Sus	Starch utilization system
T9SS	Type IX Secretion System
TAE	Tris-acetate-EDTA
TFA	Trifluoracetic acid
TGS	Tris/Glycine/SDS
UV	Ultra-Violette
w/v	Weight/volume
Xyl_x	X numbers of repeating xylose units

Table of contents

1 Introduction

1.1 Biomass

Utilization of biomass from agricultural waste or non-food crops is regarded as a sustainable way of creating valuable products without competing with food resources. Biomass can be transformed to chemicals or energy products, while at the same time recycle the emitted carbon dioxide through crop growth. The demand for energy and new sources of energy in the world today is evident. The U.S. Energy Information Administration (2016) predicts that the energy consumption worldwide will increase with 48 % from 2012 to 2040, making the production of biofuels and energy from biomass of broad and current interest. Fossil fuels are non-renewable and reserves are predicted to be depleted in the near future (Shafiee & Topal 2009), therefore more environmentally friendly and sustainable energy sources are necessary. The U.S. and EU both have programs for partly replacing transportation fuels with biofuels within 2030 (Himmel et al. 2007). The current production of biofuels is mainly based on conversion of starch from food crops (Liao et al. 2016), termed first generation biofuels. The main disadvantages of first generation biofuels are the usage of food crops and arable land for their production. This paved the way for a new generation of biofuels based on non-food biomass, termed second generation biofuels.

Plant biomasses that cannot be used for food are mainly lignocellulosic biomass. The conversion of lignocellulosic biomass to valuable products is difficult due to its recalcitrant nature. One of the major rate limiting steps in biofuel production is the enzymatic conversion of cellulose to sugars (Himmel et al. 2007), and it is considered the major bottleneck in biofuel production (Vanholme et al. 2013; Viikari et al. 2012). For this reason, new and better enzymes are needed to make the biomass conversion more effective and profitable. One attempt to do so is by engineering enzymes to improve activity, but this approach is difficult and no great results have been reported (Wilson 2009). Another way of finding new and better enzymes for biomass conversion is by searching in microbial environments that specialize on biomass breakdown. These organisms are seen as a “treasure trove of enzymatic tools” (Gilbert et al. 2008). For example, microbes in compost and the cow rumen are efficient degraders of lignocellulosic biomass, and learning from these microbes we may be able to improve the current conversion of biomass to valuable products like biofuels, chemicals and energy.

1.2 Lignocellulose

Lignocellulose is the most abundant organic material on earth and holds large amounts of energy. It is a component of the plant cell wall, contributing to its rigidity, and hence has a strong, recalcitrant structure (Figure 1.1). This recalcitrant structure is due to factors like cellulose crystallinity, lignin hydrophobicity and lignin-hemicellulose-cellulose cross-linkages and interactions (Moraïs et al. 2012). Lignocellulose consists of lignin, hemicellulose and cellulose in various ratios depending on the feedstock source. The cellulose polymers are bound together by hydrogen bonds in microfibrils which are cross-linked with hemicellulose and lignin, making lignocellulose a strong, recalcitrant substrate.

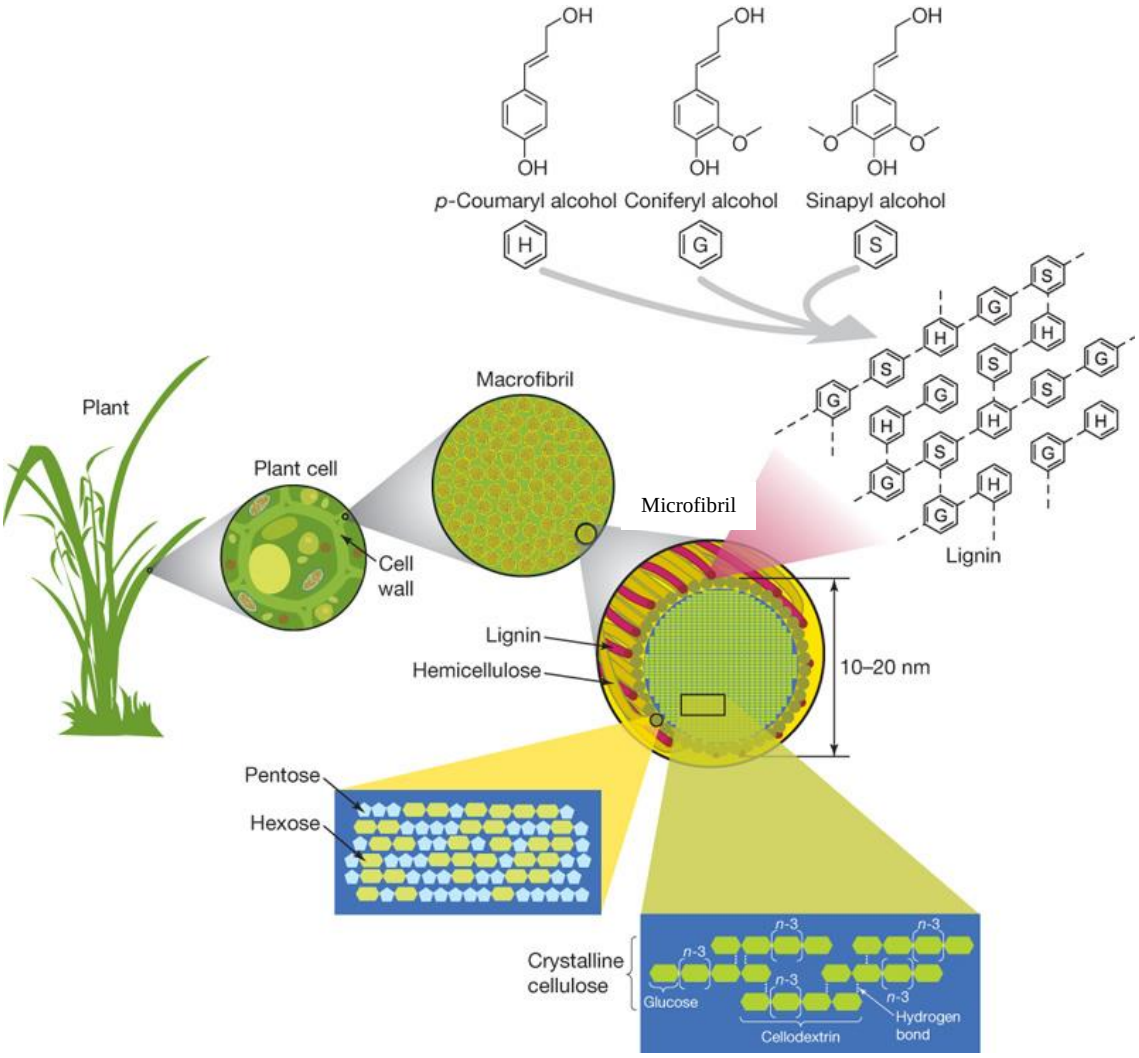


Figure 1.1. Structure of lignocellulose. The plant cell wall is built up of cellulose fibres bound together by hydrogen bonds, and cross-linked with hemicellulose and lignin to a microfibril. Microfibrils are packed together in a macrofibril, which gives structural strength to the plant cell wall. Figure from Rubin (2008), reprinted with permission from Nature Publishing Group.

1.2.1 Lignin

The lignin component of lignocellulose is a highly crosslinked polymer of the monomeric units *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) (de Gonzalo et al. 2016). The rigid nature and hydrophobicity of lignin protects the cellulose and hemicellulose in the plant cell wall from hydrolysis. Lignin is considered the main barrier in deconstruction of the plant cell wall (Chandel et al. 2015), and pre-treatment of lignocellulose is therefore needed to gain access to the energy-rich cellulose.

1.2.2 Hemicellulose

Hemicelluloses are a heterogenous group of non-cellulosic polysaccharides, and not very well defined (Scheller & Ulvskov 2010). They consist of 5- and 6- carbon sugars monomers like arabinose, galactose, glucose, xylose and mannose (Rubin 2008) linked together and contain varying branching and/or substitution. Hemicelluloses have a degree of polymerization of 80-200 (Peng et al. 2009), making them a lower molecular weight substrate than cellulose. Typical hemicelluloses are xylan (β-1,4-xylose), mannan (β-1,4-mannose), xyloglucan (β-1,4-glucose with α-1,6-xylose substitutions), glucomannans (β-1,4-mannose, β-1,4-glucose) and β-(1,3-1,4)-glucans (lichenan, barley β-glucan) and different versions of these with substituted side chains (Scheller & Ulvskov 2010). For example, galactomannan is a mannan with β-1,6-linked galactose sidechains. A special hemicellulose is the β-1,3-glucan pachyman, derived from the cell wall of tree root fungi (Hoffmann et al. 1971).

1.2.3 Cellulose

Cellulose is the main component of lignocellulose and holds a large amount of the energy in plant cell walls. Cellulose is a linear polymer of β-1,4-linked glucose units (Figure 1.2). In the plant cell wall it forms fibril structures with approximately 30-36 polymers, bound together by hydrogen bonds and Van der Waals forces (Somerville et al. 2004). These fibril structures are crystalline, creating a recalcitrant network, but cellulose also contains more disordered, amorphous regions (Quiroz-Castañeda & Folch-Mallol 2013) (Figure 1.2).

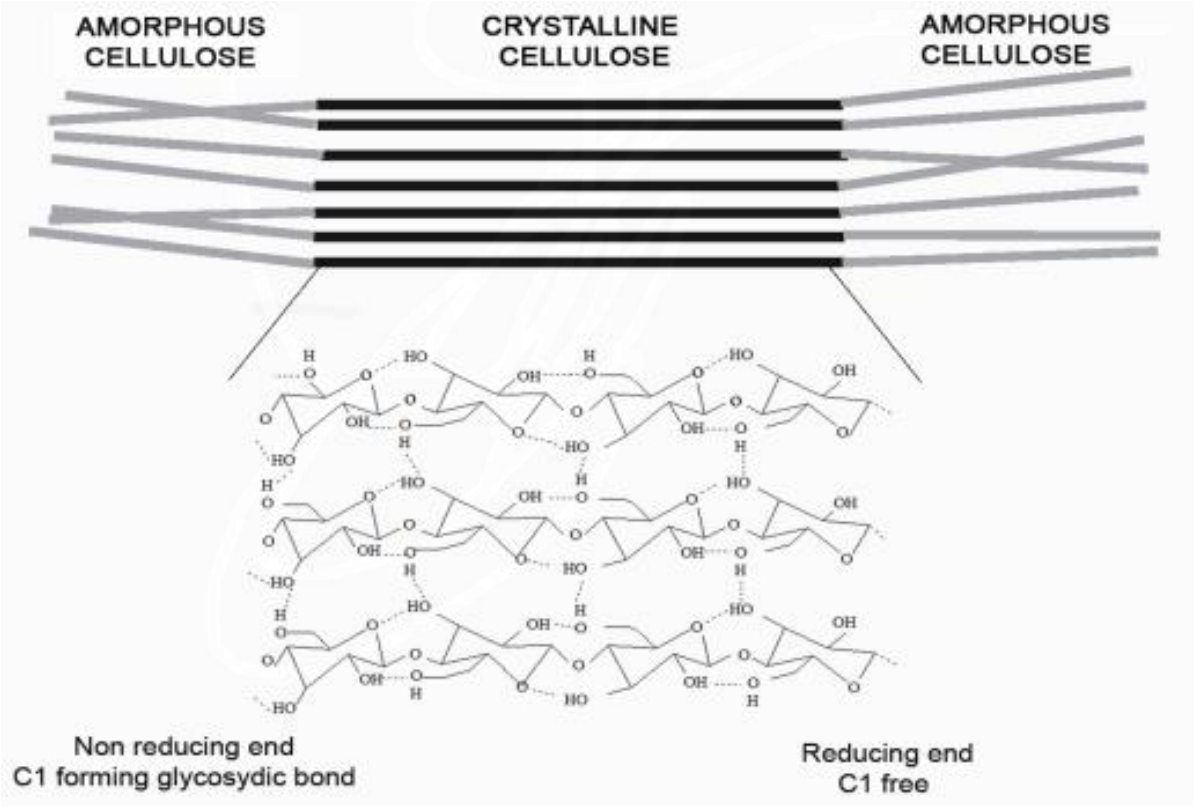


Figure 1.2. Structure of cellulose. Glucose units are connected by β-1,4-linkages, and rotated 180º relative to each other. Repeating units of cellobiose form a linear polymer. Cellulose contains crystalline and amorphous regions. Crystalline cellulose has a highly ordered network with hydrogen bonds and van der Waals forces, while amorphous cellulose has a more irregular structure. Figure taken from Quiroz-Castañeda and Folch-Mallol (2013).

The amorphous regions of cellulose provide access sites for cellulose degrading enzymes. Cleaving of the glycosidic bonds of cellulose leads to a final product of glucose, which can be fermented to products such as ethanol.

1.3 Enzymatic degradation of cellulose

Degradation of cellulose is typically performed by four classes of enzymes; endocellulases, cellobiohydrolases (CBH), lytic polysaccharide monooxygenases (LPMO) and β-glucosidases (Horn et al. 2012). An overview of the classical paradigm for enzymatic degradation of cellulose by these enzymes are shown in Figure 1.3.

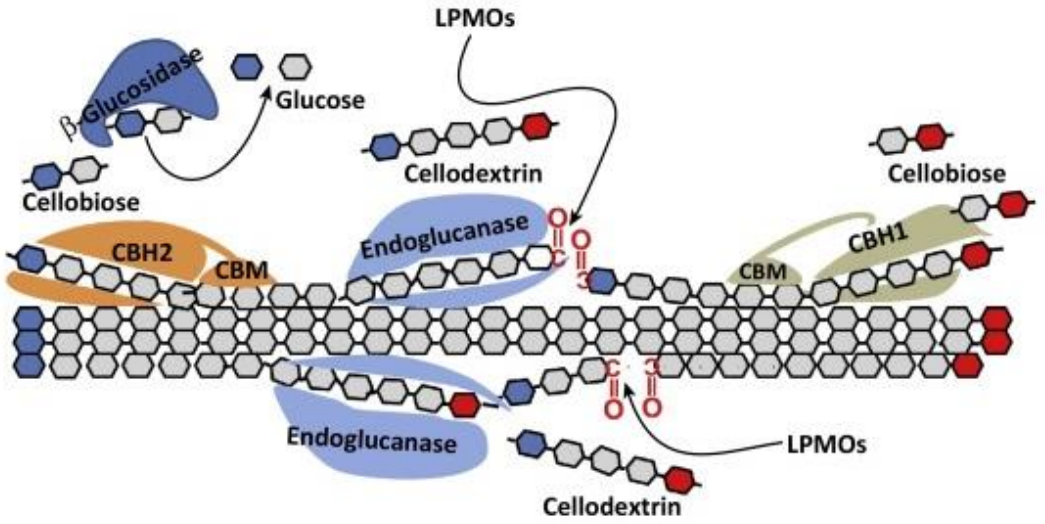


Figure 1.3. Enzymatic degradation of cellulose. Endoglucanases, cellobiohydrolases (CBH1 and CBH2), β-glucosidases and LPMOs act in synergy to degrade cellulose. See text for details. Figure adapted from Gupta et al. (2016), reprinted with permission from Elsevier Ltd.

Endocellulases (endo-1,4-β-glucanases) are enzymes that cleave internal bonds in the amorphous regions of the cellulose chain (Lynd et al. 2002). They can be processive, performing multiple cleavages before dissociating from the substrate, or non-processive, performing one cleavage and dissociating form the substrate before connecting with the substrate at a new site. Processive endocellulases usually produce cellobiose or cellotetraose from the chain ends, while non-processive endocellulases produce new chain ends internally in the cellulose chain (Zhang et al. 2014). Endocellulases typically have a cleft/groove active-site topology, enabling easy binding to random internal sites in the cellulose chain (Davies & Henrissat 1995).

Cellobiohydrolases (exo-1,4-β-glucanases) cleave bonds from the cellulose chain end releasing cellobiose. Cellobiohydrolases can act from the reducing or the non-reducing end of the glycoside chain, and usually act in a processive manner by staying attached to the cellulose chain after release of product (Lynd et al. 2002). The active-site topology of cellobiohydrolases is a tunnel, enabling the cellulose chain to be threaded through the active site (Davies & Henrissat 1995). The tunnel-shaped active site of cellobiohydrolases makes the enzymes able to glide along the chain as products are released (Kipper et al. 2005). Cellobiohydrolases can act on the crystalline parts of cellulose, and are thus thought to be capable of pulling individual cellulose chains out of the crystalline structure (Lynd et al. 2002; Teeri 1997).

β-glucosidases are considered the second line of degradation as they act on cellobiose, the major product of the two previously mentioned enzymes. β-glucosidases cleave the β-1,4-glycosidic bond between the two glucose units of cellobiose, releasing glucose for further utilization.

Recent discoveries add another enzyme to the cellulose degradation mechanism; lytic polysaccharide monooxygenases (LPMOs) (Forsberg et al. 2011; Horn et al. 2012; Vaaje-Kolstad et al. 2010). LPMOs are copper-enzymes that cleave glycosidic bonds in a oxidative manner (Johansen 2016), and thus require oxygen to function as well as a reducing agent. LPMOs attack the crystalline regions of cellulose, making it more accessible for endocellulases and cellobiohydrolases (Kostylev & Wilson 2012).

By working together, these four classes of enzymes can degrade cellulose to glucose. Synergism between endocellulases and cellobiohydrolases have been suggested where endocellulases create new chain ends for cellobiohydrolases to work on, while cellobiohydrolases disrupt the crystallinity and expose new sites on the cellulose fibril for the endocellulase to attack (Kostylev & Wilson 2012). In addition, product inhibition of cellobiohydrolases by cellobiose is seen and proves the importance of β-glucosidases in the degradation system (Teugjas & Väljamäe 2013).

1.4 Carbohydrate-active enzymes

1.4.1 Carbohydrate-active enzymes database (CAZy)

The carbohydrate-active enzymes database (CAZy; http://www.cazy.org) is a collection of all known enzymes that act on carbohydrates (Lombard et al. 2014). The CAZy database groups the enzymes into families based on sequence similarity, and as of 2013 it contained 340 000 carbohydrate-active enzymes (an increase of 225 % from 2008), illustrating the extensive research being done on these enzymes. The database contains 6 protein classes; glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and the non-catalytic carbohydrate-binding modules (CBMs). CBMs are modules of carbohydrate-active enzymes with carbohydrate-binding activity, facilitating hydrolysis by bringing the substrate and catalytic domain closer together (Shoseyov et al. 2006). CBMs can also be a part of a scaffolding subunit in cellulosomes, elaborated further down. Glycoside hydrolases are the largest class of CAZymes and have the function of hydrolysing glycosidic bonds in carbohydrates or between a carbohydrate and a

non-carbohydrate. The cellulases discussed in section 1.3 are all GHs, whereas LPMOs are AAs.

1.4.2 Glycoside hydrolases (GHs)

There are currently 136 GH families in the CAZy database (as of march 2017). Some of the families have been further divided into subfamilies based on phylogenetic analysis, to make the functional annotation of enzymes easier (Lombard et al. 2014). The subfamilies are mostly monospecific (enzymes have the same substrate specificity). Some of the GH families are grouped into clans based on folds rather than the less conserved amino acid sequence.

1.4.3 Glycoside hydrolase family 5 (GH5)

Glycoside hydrolase family 5 is one of the largest GH families, with a variety of specificities and high abundance in different ecological environments (Aspeborg et al. 2012). Within CAZy, the family is a part of the GH-A clan with a typical $(\beta/\alpha)_8$ (TIM barrel) protein fold. The GH5 proteins are known to have a retaining cleavage mechanism, and follows a classical Koshland double-displacement mechanism (Figure 1.4) (Koshland 1953). The two catalytic amino acids, a catalytic nucleophile and a catalytic proton donor, also called the catalytic acid/base, have been experimentally determined to be glutamic acids (Jenkins et al. 1995). Explained by, among others, McIntosh et al. (1996) and Davies and Henrissat (1995), the cleaving mechanism (Koshland double-displacement) starts with the nucleophile amino acid attacking the anomeric carbon of the glycosidic bond forming a glycosyl-enzyme intermediate, while the other catalytic amino acid protonates the glycosidic oxygen of the leaving group. Then the catalytic proton donor (now acid) deprotonates a water molecule which attacks the glycosyl-enzyme bond, cleaving it and releasing the enzyme. The stereochemistry of the anomeric carbon in the glycosidic bond is retained.

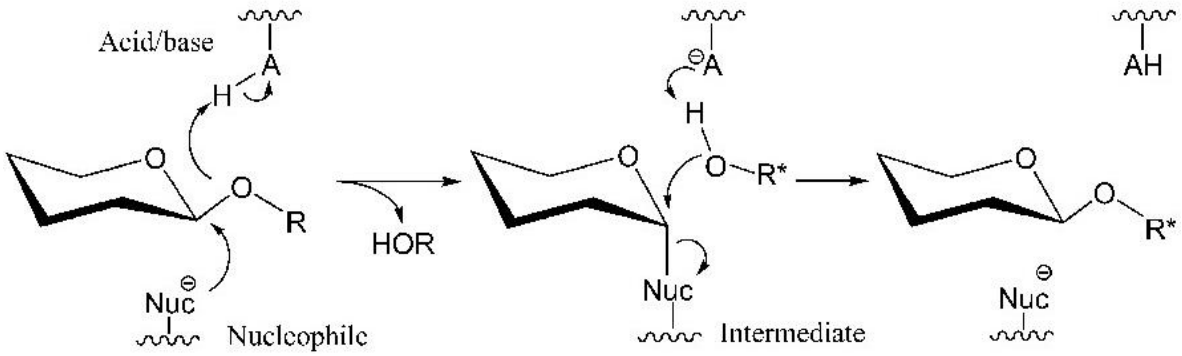


Figure 1.4. Cleaving mechanism of GH5s. The enzymes follow a retaining classical Koshland double-displacement mechanism. The mechanism is explained in the text. R* is most often a H as part of a water molecule. Figure modified from Vuong and Wilson (2010), reprinted with permission from John Wiley & Sons, Inc.

The variety of enzyme specificities in the GH5 family made it hard to predict function of new family members. Therefore, Aspeborg et al. (2012) divided the family into 51 distinct subfamilies by using a multiple sequence alignment to infer an approximate-maximum-likelihood phylogenetic tree. One third of the subfamilies were monospecific, whereas the largest subfamily, subfamily 2 (GH5_2), is polyspecific with many members being extracellular, multi-modular endo-β-1,4-glucanases (Aspeborg et al. 2012).

1.5 Microbial degradation of cellulose

Microbes use different strategies for cellulose degradation. Most known are the three mechanisms referred to as secreted free enzymes, cellulosomes and carbohydrate-targeting polysaccharide utilization loci (PUL). PULs have newly been proposed to include cellulose as a target carbohydrate (Naas et al. 2014). Common for the mechanisms is translocation of the cellulases over the outer cell-membrane, as microbes are unable to transport large molecules such as cellulose into the cell (Wilson 2011).

The first mechanism, secreted free enzymes (Figure 1.5A; see also Figure 1.3), is mostly used by aerobic microorganisms (Wilson 2011). Individual free cellulases are secreted into the cellulose-rich environment where they degrade the cellulose into sugars, which can be taken up and utilized by the microbe. Next to a catalytic domain, cellulases often contain a CBM, to attach the enzyme to cellulose, and may contain additional domains (Wilson 2008).

The second mechanism, cellulosomes (Figure 1.5B), is a known feature of anaerobic microbes (Wilson 2011). Cellulosomes are large complexes of many cellulases, usually attached to the outer surface of the microbe. Cellulosomes consist of a scaffolding protein ("scaffoldin") and cellulases connected by interaction of dockerins in the cellulases and cohesins in the scaffolding unit (Fontes & Gilbert 2010). The cellulosome binds cellulose by CBMs on the scaffolding unit (Wilson 2008; Wilson 2011).

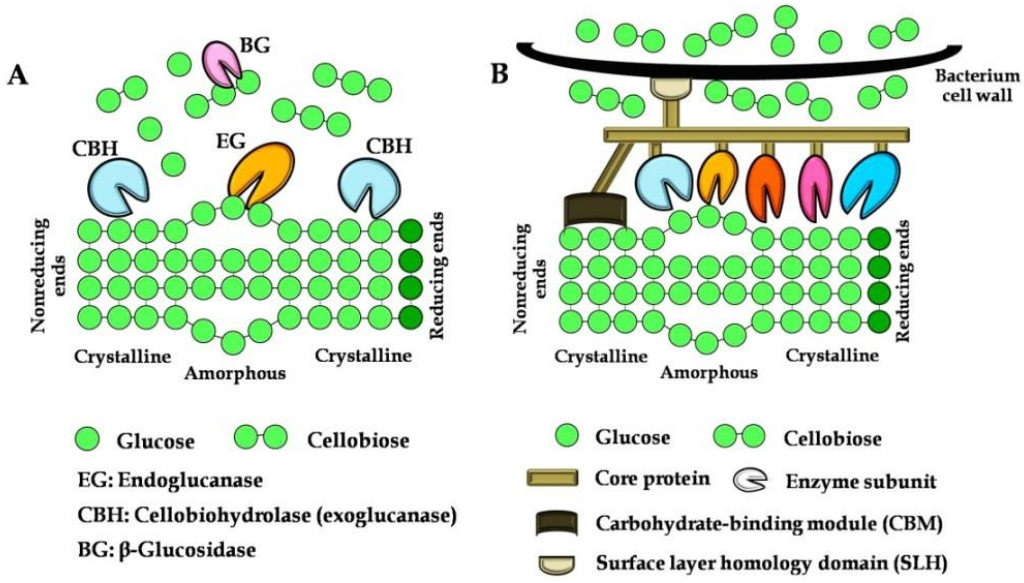


Figure 1.5. Overview of cellulose breakdown by the secreted free enzymes mechanism (A) and the cellulosome mechanism (B). Figure taken from Ratanakhanokchai et al. (2013).

The third well-known mechanism for carbohydrate degradation by microbes involves polysaccharide utilization loci (PULs). PULs were first described by Bjursell et al. (2006). They are substrate specific (one PUL for each carbohydrate substrate) and were proposed to include cellulose by Naas et al. (2014). PULs are a set of co-localized and co-regulated genes that encode different enzymes and proteins needed for polysaccharide degradation (Grondin et al. 2017). PULs are a feature of the phylum Bacteroidetes, a phylum which consists of anaerobic gram-negative bacteria. The gram-negative double membrane creates a periplasmic space and two levels for polysaccharide degradation, where enzymes on the outer membrane cleave carbohydrates into shorter oligomers that are transported to and further broken down in the periplasmic space. This was first described for the starch utilization system (Figure 1.6) where binding proteins (SusD, SusE, SusF), an amylase (SusG) and a transport protein (SusC) on the

cell surface are responsible for binding and cleaving of starch, before transporting oligosaccharides into the periplasmic space (Terrapon et al. 2015). In the periplasmic space, oligosaccharides are broken down to glucose by PUL-encoded neopullanase (SusA) and α-glucosidase (SusB), and the glucose is transported into the cell across the cytoplasmic membrane (Terrapon et al. 2015). In addition, a transmembrane transcriptional regulator (SusR) is present in the cytoplasmic membrane, contributing to the regulation of PUL protein transcription (Martens et al. 2009). A defining feature of PULs is the presence of adjacent *SusC/SusD*-like genes (Martens et al. 2009; Terrapon et al. 2015), involved in transport and binding, respectively.

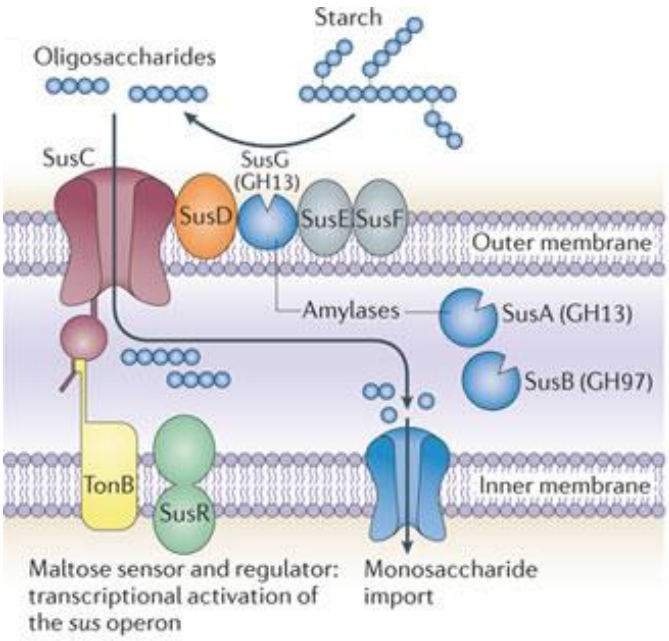


Figure 1.6. Starch utilization system of *Bacteroides thetaiotaomicon* as an example of a PUL system. SusA (neopullanase), SusB (α-glucosidase), SusC (transport protein), SusD, E and F (binding proteins), SusG (amylase) work together for the degradation of starch in regulation of SusR (regulator). TonB provides energy for SusC in the transportation of oligosaccharides. See text for details. Figure modified from Koropatkin et al. (2012), reprinted with permission from Nature Publishing Group.

In addition to the three cellulolytic mechanisms described above, studies on the aerobic soil bacterium *Cytophaga hutchinsonii* (Xie et al. 2007) and the anaerobic rumen bacterium *Fibrobacter succinogenes* (Suen et al. 2011) have revealed the existence of mechanisms differing from these paradigms. Neither of these bacteria have genes encoding cellulose binding

CBMs or exocellulases in their genomes, which are common features of the secreted free enzyme mechanism, and they also lack genes encoding dockerins and cohesins, which are required for the assembly of cellulosomes. Moreover, no *SusC/SusD*-like genes are found in the genome of *F. succinogenes*, and it has been shown that *C. hutchinsonii* does not require its two encoded SusC/SusD pairs for cellulose degradation (Zhu et al. 2015), excluding the PUL mechanism. It has been suggested that *C. hutchinsonii* can glide along the cellulose fibres, using cell surface endocellulases to cleave the cellulose chains (Xie et al. 2007). Further, the oligomers are transported into the periplasmic space where they are depolymerized and utilized. *F. succinogenes* may be able to use a similar mechanism by attaching to the cellulose fibres by so-called fibro-slime proteins (Suen et al. 2011).

The gliding motility of *C. hutchinsonii* and other members of the Bacteroidetes phylum have been coupled to the Por secretion system (PorSS), also called the Type IX secretion system (T9SS) (McBride & Zhu 2013; McBride & Nakane 2015; Sato et al. 2010). The system is highly prevalent in the Bacteroidetes phylum (McBride & Zhu 2013). A few years ago the T9SS was linked to cellulose degradation when it was shown that deletion of a T9SS gene resulted in a cellulose utilization defect in *C. hutchinsonii* (Zhu & McBride 2014). The proteins destined for secretion by T9SS have a common C-terminal domain (CTD) working as a signal peptide, here named Por-tag (Veith et al. 2013). The Por-tag is approximately 72 base pairs long, varying slightly in length between different proteins. The secretion tag is only found in the Bacteroidetes phylum, and common features of the proteins containing the tag seems to be cleaving of the tag during transportation, post translational modifications (such as glycosylation) and attachment to the cell surface (Veith et al. 2013).

An example of a T9SS secreted CAZyme is the chitin-cleaving ChiA from *Flavobacterium johnsoniae*. This bacterium, along with e.g. *Caldicellulosiruptor bescii*, use yet a different mechanism for carbohydrate degradation. The two bacteria both produce multi-modular enzymes, that have an effective technique for degradation of crystalline substrates (chitin and cellulose respectively) (Brunecky et al. 2013; Larsbrink et al. 2016). The key enzyme in each bacterium, the PUL encoded but secreted ChiA from *F. johnsoniae* and CelA from *C. bescii*, consist of exo- and endoacting catalytic domains with one or multiple binding domains in between. This type of enzyme seems to be effective in breakdown of crystalline substrates, due to the cooperation between the exo- and endodomains, as high synergy between the enzyme

domains have been proven for both ChiA (Larsbrink et al. 2016) and CelA (Brunecky et al. 2013).

1.6 The cow rumen as a cellulolytic environment

The cow rumen is a large bio-active environment with a diverse microbial community that metabolizes lignocellulosic biomass. A symbiotic relationship exists between the cow (host) and the microbes, whereby the later degrade lignocellulose and produce short-chain volatile fatty acids which are absorbed by the host for nutrition (Wallace et al. 2015). As these bacteria are specialized in lignocellulose and cellulose breakdown, they are a great source of potential industrial enzymes, and may help deepening our understanding of cellulose degradation in nature. Studies of the microbial diversity in the rumen have indicated that only 6.5-11 % of inherent microbes are representative of previously cultured bacteria (Edwards et al. 2004; Kim et al. 2011), leaving a large fraction of uncultured bacteria and hence a great unexplored resource for further studies and research.

The rumen microbiota is dominated by the three phyla Bacteroidetes, Firmicutes and Proteobacteria (Mao et al. 2015; Petri et al. 2014), with *Fibrobacter succinogenes*, *Ruminococcus albus* and *Ruminococcus flavefaciens* being the major known degraders specialized on rumen biomass utilization (White et al. 2014). Even so, the presence and abundance of different microbes can vary with factors like season and diet (Edwards et al. 2004). The microbial community and the activity of degrading enzymes are also affected by pH and temperature conditions in the rumen. A healthy rumen holds a pH within 5.8-6.4 and a temperature between 37-42 ºC to maintain the microbial environment and secure optimal growth of microbes (Ishler et al. 1996).

1.7 Aim of this study

Enzymatic conversion of cellulose to sugars is considered one of the major bottlenecks in production of second generation biofuels. Finding new and more effective cellulases is therefore crucial and might enable biofuels to compete with traditional fossil fuels. A better understanding of the mechanisms for cellulose turnover in environments that have evolved to specialize in cellulose degradation could prove to be fruitful for this purpose. An example of a

cellulolytic environment is the cow rumen, where symbiotic bacteria break down cellulosic substrates to volatile fatty acids for the host to utilize as energy.

The background for the study described in this thesis was the metagenomic study of the cow rumen by Hess et al. (2011), revealing biomass degrading genes and genomes from bacteria adherent to switchgrass. Further bioinformatical analysis of the dataset (Konietzny et al. 2014; Weimann et al. 2013) led to the prediction of one of the reconstructed genomes as a cellulose degrader. The genome, termed AGa, is a part of the Bacteroidetes phylum. The genome contains a putative cellulolytic gene cluster, which encodes enzymes annotated to target cellulose and its degradation products. No dockerins, cohesins or exocellulases were present in the AGa genome, ruling out the classical mechanisms for cellulose breakdown, secreted free enzymes and cellulosomes. Also, the cluster does not contain *SusC* or *SusD* genes, indicating that the gene cluster is not a classical PUL and might represent a new type of mechanism for cellulose degradation. However, the AGa genome harbours several *SusC/SusD* gene pairs not linked to CAZymes as in the typical PUL gene-organization, which might be involved in cellulose degradation. Due to the putatively novel mechanism for cellulose degradation, the gene cluster was selected for further study. The cluster encodes four putative cellulases containing GH5 domains, of which three contain a Por-tag for secretion through the T9SS, a feature that has been liked to cellulose degradation and gliding motility. One of these enzymes contains two individual GH5 domains, a feature not seen in Bacteroidetes before.

The aim of this study was to clone, express and characterize the four GH5 containing enzymes and their individual domains, to obtain insight into their target substrates, and to elucidate why the AGa gene cluster contains four seemingly similar enzymes. Furthermore, the aim was to study the potential roles of these enzymes in cellulose breakdown, and thus potentially contribute to knowledge on cellulose degradation by the AGa gene cluster and genome, and by microbes in general.

2 Materials

2.1 Laboratory instruments

Laboratory instruments and equipment used in this study, with their supplier, are listed in Table 2.1.

Table 2.1. Instruments. Listed are instruments used in this study, with suppliers.

Instrument/equipment	Supplier
827 pH lab pH meter	Metrohm
ÄKTAprime plus chromatography system with HisTrap™ FF Ni sepharose column (5 ml)	GE Healthcare
Avanti™ J-25 centrifuge with rotor JA-10 and JA-25.50	Beckman Coulter
Benchtop centrifuge 5418 R	Eppendorf
Benchtop UV transilluminator	UVP
Concentrator plus	Eppendorf
D30 Biophotometer	Eppendorf
Dionex™ ICS-3000 Ion Chromatography System with CarboPac™ PA1 column	Thermo Scientific
Dionex™ UltiMate™ 3000	Thermo Scientific
DNA gel electrophoresis: MiniSub® Cell GT Horizontal Electrophoresis System PowerPac™ Basic	 Bio-Rad Bio-Rad
DOPPIO thermal cycler	VWR
FastPrep-24™ 5G homogenizer	MP Biomedicals
Gel Doc™ EZ Imager	Bio-Rad
Harbinger LEX-48 bioreactor system	Harbinger Biotechnology
Heraeus™ Multifuge™ X1R	Thermo Scientific
HisTrap™ FF Ni sepharose column	GE Healthcare
Multiscan™ FC Microplate Photometer	Thermo Scientific
Protein gel electrophoresis: Mini-PROTEAN® Tetra Systems Mini-PROTEAN® TGX Stain-Free™ Gel PowerPac™ 300 power supply	 Bio-Rad Bio-Rad Bio-Rad
Qubit™ fluorometer	Invitrogen
Shaking incubator	New Brunswick Scientific
Sonicator/degaser 3510	Branson
ThermoMixer C	Eppendorf
Vibracell sonicator	Sonics
Water bath	Julabo/Stuart

2.2 Chemicals

Chemicals used in this study are listed in Table 2.2, with the names of the suppliers.

Table 2.2. Chemicals. Listed are chemicals used in this study, with suppliers.

Chemicals	Supplier
1,4-Dithiothreitol (DTT)	Sigma-Aldrich
2-Propanol	Sigma-Aldrich
3,5-Dinitrosalicylic acid	Sigma-Aldrich
Acetic acid	Merck
Acetonitrile (ACN) for HPLC	VWR
Agar powder	VWR
Ammonium bicarbonate (Ambic)	Sigma-Aldrich
Antifoam 204	Sigma-Aldrich
Bacto™ Tryptone	Becton, Dickinson and Co.
Bacto™ Yeast Extract	Becton, Dickinson and Co.
Bug Buster® 10X Protein Extraction Reagent	Novagen
Citric acid buffers: Citric acid Sodium citrate tribasic dihydrate	 Sigma-Aldrich Sigma-Aldrich
Coomassie Brilliant Blue R-250	Bio-Rad
D-Glucose	Sigma-Aldrich
Ethanol (96 %)	VWR
Gel Loading Dye Blue (6X)	New England Biolabs
Glycerol 85 %	Merck
Imidazole	Sigma-Aldrich
Iodoacetamide (IAA)	Sigma-Aldrich
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	Sigma-Aldrich
Kanamycin	Sigma-Aldrich
Methanol for analysis	Merck
Nickel sulfate	Sigma-Aldrich
NuPAGE® LDS Sample Buffer (4X)	Invitrogen
NuPAGE® Sample Reducing Agent (10X)	Invitrogen
PeqGreen	Peqlab
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich
Phosphate buffers: Sodium phosphate dibasic Sodium phosphate monobasic	 Merck Merck
Potassium sodium tartrate	Sigma-Aldrich
SeaKem® LE Agarose	Lonza

Chemicals	Supplier
Sodium acetate anhydrous, BioUltra	Sigma-Aldrich
Sodium chloride	VWR
Sodium hydroxide (NaOH)	Merck
Sodium hydroxide solution 50 % for HPLC	Fluka
Sucrose	Sigma-Aldrich
Tert-butanol	Merck
Titriplex® III ethylenedinitrilotetraacetic acid (EDTA) disodium salt hydrate	Merck
Trifluoracetic acid (TFA)	Sigma-Aldrich
Tris buffers: Tris base Adjusted with HCl/NaOH	Calbiochem
Tris/Glycine/SDS (TGS) buffer (10X)	Bio-Rad
Tris-acetate-EDTA (TAE) buffer (50X)	Pre-made in lab
Tris-HCl	Amresco
Triton X-100	Fluka
Tween® 80	Sigma-Aldrich

2.3 Carbohydrate substrates

Carbohydrate substrates used in this study are shown in Table 2.3.

Table 2.3. Carbohydrate substrates. Carbohydrate substrates were used in assaying enzymatic activity and are listed together with sources and suppliers.

Carbohydrate substrate	Supplier
Arabinoxylan (Wheat)	Megazyme
Avicel® PH-101	Fluka
Carboxymethyl cellulose (CMC) sodium salt, low viscosity	Sigma-Aldrich
Cellodextrins: Cellobiose Cellotriose Cellotetraose Cellopentaose Cellohexaose	 Fluka Megazyme Megazyme Megazyme Megazyme
CM-Pachyman	Megazyme
Filter paper, grade 1 (0.5 mm)	Whatman®
Galactomannan (Carob)	Megazyme
Galactomannan (Guar)	Megazyme
Lichenan (Icelandic moss)	Megazyme
Mannan (Ivory nut)	Megazyme

Carbohydrate substrate	Supplier
Switchgrass (ball milled and washed)	The Noble Foundation, Ardmore, OK, USA
Xylan (Birchwood)	Carl Roth GmbH
Xyloglucan (Tamarind)	Megazyme
β-glucan (barley) medium viscosity	Megazyme

2.4 Enzymes and proteins

Enzymes and proteins used in this study are shown in Table 2.4, with their suppliers.

Table 2.4. Enzymes and proteins. Listed are enzymes and proteins used in this study, with suppliers.

Enzyme/protein	Supplier
Benchmark™ Protein Ladder	Invitrogen
Bovine Serum Albumin (BSA)	New England Biolabs
Lysozyme chloride form from chicken egg white	Sigma-Aldrich
NEBuffer	New England Biolabs
Q5® Hot Start High-Fidelity 2X Master Mix	New England Biolabs
Red Taq DNA polymerase 2x Master Mix	VWR
T4 DNA polymerase	New England Biolabs
TEV protease	Sigma-Aldrich
Trypsin	Promega
β-glucosidase (40 U/ml)	Megazyme

2.5 Cells, plasmid and DNA

Cells, plasmid and DNA used in this study are shown, with suppliers, in Table 2.5. Primers used in this study are discussed in section 3.3.1.

Table 2.5. Cells, plasmid and DNA. Listed are cells, plasmid and DNA used in this study, with suppliers.

Cells/plasmid/DNA	Supplier
2'-deoxyguanosine 5'-triphosphate (dGTP)	Thermo Scientific
GeneRuler™ 1 kb DNA ladder	Thermo Scientific
One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*.	Invitrogen
One Shot® TOP10 Chemically Competent *E. coli*	Invitrogen
pNIC-CH plasmid	Addgene (gift from Opher Gileadi)

2.6 Kits

Different kits used in this study and their suppliers are listed in Table 2.6.

Table 2.6. Kits. Listed are the different kits utilized in this study, with suppliers and what they are used for. The contents of these kits are described in the methods section.

Kit	Supplier	Use
Expresso® solubility and Expression Screening System	Lucigen	Solubility screening of protein expression
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel	Purification of DNA from agarose gel
NucleoSpin® Plasmid Kit	Macherey-Nagel	Plasmid purification
Qubit™ dsDNA BR Assay Kit	Invitrogen	DNA concentration measurement
Wizard™ 2 Screen	Emerald Biosystems	Crystallization trials

2.7 Buffers and other reagents

For immobilized metal affinity chromatography:

Buffer A (binding buffer): 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 5 mM imidazole

Buffer B (elution buffer): 50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 0.5 M imidazole

Strip buffer: 20 mM phosphate buffer pH 7.4, 0.5 M NaCl, 50 mM EDTA

Buffers were filtered through a 0.22 µm pore filter.

For enzyme assays:

Dinitrosalicylic acid (DNS) reagent (1 % (w/v)): 8 g NaOH, 150 g potassium sodium tartrate, 5 g 3,5-dinitrosalicylic acid in 500 ml Milli-Q water stored in dark flask at room temperature.

For high-performance anion-exchange chromatography:

Eluent A: 0.1 M NaOH

Eluent B: 1 M NaOAc, 0.1 M NaOH (0.22 µM filtered)

Eluent C: Milli-Q water

All eluents were sonicated for 20 minutes to degas the solutions. Eluents were made directly prior to use.

For metaproteomics:

Dissociation solution: 1 % MeOH, 1 % tert-butanol, 0.1 % Tween-80, pH 2.0

Cell-wash buffer: 10 mM Tris-HCl, 1 M NaCl, pH 8.0

Lysis buffer: 50 mM Tris-HCl, 200 mM NaCl, 0.1 % Triton X-100, 1 mM DTT, 1 mM EDTA, 1 mM PMSF, pH 7.5

Buffer A: 50 % 2-propanol, 20 % acetic acid, 30 % Milli-Q water
Staining solution: 50 % Coomassie Brilliant Blue R-250 (0.1 %), 50 % buffer A
Destaining solution: 50 % buffer A, 50 % Milli-Q water

DTT solution: 10 µl 1 M DTT (stored frozen), 100 µl 1 M Ammonium bicarbonate (stored frozen), 890 µl Milli-Q water
IAA solution: 10 mg IAA, 100 µl Ammonium bicarbonate (stored frozen), 900 µl Milli-Q water

Trypsin buffer: 25 µl 1 M Ammonium bicarbonate (stored frozen), 875 µl Milli-Q water, 100 µl 100 % ACN
Trypsin solution: 10 ng/µl trypsin in trypsin buffer. The trypsin stock was stored at -80 ºC

3 Methods

3.1 Storage

DNA and primers in solution were stored at -20 ºC. Proteins in protein buffer solution (see section 3.4.6) were stored at 4 ºC. Cells were stored long term as glycerol stocks at -80 ºC. Glycerol stocks were made by mixing 700 µl overnight culture with 300 µl sterile 85 % glycerol.

3.2 Cultivation of bacteria

3.2.1 Lysogeny broth medium

Lysogeny broth (LB) medium was prepared as described by Bertani (1951): 10 g Bacto™ Tryptone, 5 g Bacto™ Yeast Extract and 10 g sodium chloride (NaCl) were added to approximately 800 ml Milli-Q water in a 1 l beaker, and mixed with a magnetic stirrer. Milli-Q water was added to a total volume of 1 l and the solution was poured into a blue cork flask. The medium was then autoclaved.

3.2.2 Lysogeny broth agar

LB agar was prepared as the LB medium, but with addition of 7.5 grams of agar powder per litre prior to autoclaving. The solution was cooled with stirring to 50 ºC, and supplemented with kanamycin to a final concentration of 50 µg/ml. Thereafter, the medium was poured in petri dishes and left to cool and solidify in a laminar flow cabinet. The plates were packed in a plastic bag and stored at 4 ºC. Plates for ligation independent cloning (LIC) selection were made with additional sucrose added after autoclaving to a final concentration of 5 % (w/v).

3.2.3 Kanamycin

A kanamycin solution (50 mg/ml) was made by adding 500 mg kanamycin to 10 ml Milli-Q water in a falcon tube. The solution was sterile filtered to Eppendorf tubes and stored at -20 ºC.

3.2.4 Cultivation of bacteria

Bacterial cultures were started by inoculating 5 ml LB medium, containing 50 µg/ml kanamycin, with a single colony from an agar plate or directly from a glycerol stock. The culture was left at 37 ºC with 220 rpm shaking overnight unless otherwise noted.

3.3 Ligation independent cloning

Ligation independent cloning (LIC) is a technique which eliminates the ligation step from traditional cloning by creating overhangs in both the gene and the vector by using the inherent 3'→5' exo-nuclease activity of T4 polymerase.

All GH5 containing genes were cloned by ligation independent cloning, except *GH5A* (see section 3.5). Some genes had previously been synthesized, without predicted signal peptides (GeneArt® Gene Synthesis (Invitrogen by Thermo Fisher Scientific)), and cloned in the lab and were provided as glycerol stocks of *Escherichia coli (E. coli)* BL21 with the gene in the pNIC-CH vector. These genes were *GH5B*, *GH5C_1*, *GH5C_2*, *GH5C_wt* and *GH5D* (Figure 4.1). The rest of the genes were ordered from GeneArt® Gene Synthesis, without predicted signal peptide, and were subsequently cloned into pNIC-CH. A plasmid map of pNIC-CH is provided in Appendix A, Figure A1. The pNIC-CH plasmid contains a 6xHis-tag that are C-terminally attached to the proteins. An example of a pNIC-CH vector encoding one of the GH5 containing genes is provided in Appendix A, Figure A2.

Removal of the Por-tag from GH5D was done by splicing by overlap extension with 5'-phosphorylated primers. Inverse PCR was performed by Adrian Naas and is not explained in this thesis/chapter.

3.3.1 Primers

To enable LIC specific primers for each gene were designed, with additional bases for pNIC-CH complementarity. Primers were designed using SnapGene® and SnapGene® Viewer software and ordered from Eurofins Genomics. The primers were provided in lyophilized form and resuspended in water to 100 µM. Stock solutions were diluted to 10 µM prior to use. An overview of primers for the different genes is shown in Table 3.1.

Table 3.1. Primers used for cloning of desired GH5 genes. GH5D_nopor primers were used for inverse PCR.

Gene	Forward primer (5'->3')	Reverse primer (5'->3')
GH5A	AATCTGTACTTCCAGGGTGCAGATG TGAGCGATAACTATATCGAAAAATG	GTGGCGGCCGCTCTATTATTTCACGAT GATTTTTGCAAATTTGGAGCC
GH5A_cat	AATCTGTACTTCCAGGGTGCAGATG TGAGCGATAACTATATCGAAAAATG	GTGGCGGCCGCTCTATTAAAAACAAAT ACCACCATAACATGCGGTAAC
GH5A_d2	AATCTGTACTTCCAGGGTAGCAGCG GTATTACCTATTATGATGCA	GTGGCGGCCGCTCTATTAAATGCTCTC GGTATTATTCAGAATTTTTTCCAC
GH5B	TTAAGAAGGAGATATACTATGGGTG GTGGTGGTAATAATGCAAAACAG	AATGGTGGTGATGATGGTGCGCTTTGG TATACAGATTGCTCTGTTTCAGAACG
GH5C_wt	TTAAGAAGGAGATATACTATGGGTG GTTATGATCTGAATGAAGAAGAAAT GGAAC	AATGGTGGTGATGATGGTGCGCTTCTT TCACAATTTCGTTCACGGCCAC
GH5C_wtR	TTAAGAAGGAGATATACTATGGGTG GTTATGATCTGAATGAAGAAGAAAT GGAAC	AATGGTGGTGATGATGGTGCGCGGTAA TCAGCTGTTTATCCAGCAGGATATC
GH5C_1	TTAAGAAGGAGATATACTATGGGTG GTTATGATCTGAATGAAGAAGAAAT GGAAC	AATGGTGGTGATGATGGTGCGCACCAC AGCTTTCATAGCTTTTCGG
GH5C_2	TTAAGAAGGAGATATACTATGAGCA TTCCGACCCAAGAATATGTTGAA	AATGGTGGTGATGATGGTGCGCACCAT GAAAATTATGCGGTTCGCT
GH5C_2R	TTAAGAAGGAGATATACTATGAGCA TTCCGACCCAAGAATATGTTGAA	AATGGTGGTGATGATGGTGCGCACCAT GAAAATTATGCGGTTCGCT
GH5D	TTAAGAAGGAGATATACTATGGGTG GTTATGATATGAATGAAAAAGGCAC C	AATGGTGGTGATGATGGTGCGCTTCTT TCACCAGTTCGCTAACTGCAAC
GH5D_nopor (inverse PCR)	GCGCACCATCATCACCACC	GGTAATCGGCTGTTTATCCAGCAG

3.3.2 Polymerase chain reaction

Polymerase chain reaction (PCR) is a method for amplifying a target DNA sequence. A thermostable polymerase is used, and the reaction mixture is cycled through denaturation, primer annealing and elongation until millions of copies of the desired target DNA are generated.

The genes of interest were amplified by PCR. A 25 µl aliquot of Q5® Hot Start High-Fidelity 2X Master Mix, 1 µl template DNA, 2.5 µl forward primer (10 µM), 2.5 µl reverse primer (10 µM) and autoclaved water to a total of 50 µl were mixed on ice in PCR tubes. The tubes were vortexed, spun down and run in a PCR machine using the program shown in Table 3.2. The

PCR was performed with one non-template control (without the template DNA) for each primer pair.

Table 3.2. PCR program

Repeats	Temperature (ºC)	Time (min:sec)
1	98	00:30
Cycle 35 rounds	98	00:10
	55	00:30
	72	00:30/kb
1	72	10:00
1	10	Hold

PCR products were verified by agarose gel electrophoresis.

3.3.3 Agarose gel electrophoresis

Agarose gel electrophoresis is a separation technique for DNA fragments and can be used to determine fragment size. The negative backbone of DNA will migrate towards the positive pole in an electric field and the agarose gel makes up a network or pores for the DNA to migrate through. Large DNA fragments will take longer time to migrate through the gel pores and therefore the DNA will be separated by size. By comparing the fragment to a standard ladder containing bands with known sizes, the fragment size can be determined.

Agarose gels were made by mixing 0.5 grams of SeaKem® LE Agarose with 50 ml 1x Tris-acetate-EDTA (TAE) buffer in an Erlenmeyer flask. The mixture was microwaved for one minute and cooled to 55 ºC, before 3 µl peqGreen was added and the mixture was poured onto the gel tray for cooling. DNA was prepared for loading by adding 5 µl Gel Loading Dye Blue (6X) to 25 µl sample (PCR product), except for colony PCR (see section 3.3.8). The gel was moved to a MiniSub® Cell GT Cell with TAE buffer and the samples were applied along with GeneRuler™ 1 kb DNA ladder. Gel electrophoresis was performed at 90 volts for 50 minutes using a PowerPac™ Basic, and the DNA was visualized using a Gel Doc™ EZ Imager.

3.3.4 Agarose gel clean-up

Bands in agarose gel electrophoresis were made visible by UV-radiation and verified target bands were cut out of the gel. The gel piece was transferred to an Eppendorf tube and weighed, before purifying DNA from the gel piece using the NucleoSpin® Gel and PCR Clean-up kit. Following the manufacturer's instructions, 200 µl binding buffer NTI was added per 100 mg of gel followed by incubation at 50 ºC for 5-10 minutes with occasional vortexing until the gel material was completely dissolved. A 700 µl aliquot of the solution was added to a column in a collection tube, both included in the kit. The sample was centrifuged at 11000 g for 30 seconds and the flow-through was discarded. A 700 µl aliquot of wash buffer NT3 was then added to the column and centrifugation was conducted as before. This step was repeated once, followed by a centrifugation step at 11000 g for 1 minute to dry the membrane. The column was then moved to a clean Eppendorf tube and 25 µl Milli-Q water was added for elution. After a 1-minute incubation at room temperature (RT), the sample was centrifuged at 11000 g for 1 minute and the eluted DNA was stored at -20ºC for future steps, if further steps were not conducted directly. All centrifugations were performed using an Eppendorf Benchtop centrifuge 5418 R.

3.3.5 Measurement of DNA concentrations

Concentration of purified DNA was measured using the Qubit™ dsDNA BR Assay Kit.

The Qubit™ working solution was prepared by mixing Qubit™ dsDNA BR Reagent with Qubit™ dsDNA BR Buffer 1:200. A 190 µl aliquot of working solution was added to 0.5 ml tubes, provided in the kit, for the two standards and 195 µl for the samples. Standards and samples were added to their respective 0.5 ml tubes to a final volume of 200 µl and vortexed. All tubes were incubated for 2 minutes at RT and measured using Qubit™ fluorometer.

3.3.6 T4 DNA digestion and annealing with pNIC-CH vector

Purified DNA was digested with T4 DNA polymerase to create complimentary overhangs for base-pairing with the digested pNIC-CH vector. pNIC-CH was previously digested with T4 polymerase in the lab, and enables the complimentary overhangs of the DNA and the vector to anneal. By adding dGTP (DNA) and dCTP (vector) to the digestion mixtures, the cleaving of the strands is stopped at complementary sites.

Insert DNA (0.2 pmol) was added to an Eppendorf tube together with 2 µl NEBuffer 2, 1 µl BSA (0.5 mg/ml), 2 µl dGTP (25 mM), 1 µl DTT (100 mM), 1 µl T4 DNA polymerase and Milli-Q water to a final concentration of 50 µl. The mixture was incubated at RT for 1 hour and the enzyme was then heat deactivated at 75 ºC for 21 minutes. A transformation mixture for cloning was prepared by adding 2 µl of T4 digested insert to 1 µl pre-digested pNIC-CH plasmid. Following a 5-minute incubation at RT, 2 µl EDTA (25 mM) was added to the mixture and the mixture was incubated at RT for 10 new minutes before transformation.

3.3.7 Transformation into TOP10 *E. coli*

The pNIC-CH vector contains a kanamycin resistance gene to select for transformed cells. The vector also contains a stuffer fragment of *sacB* which encodes an enzyme that hydrolyses sucrose to high-molecular-weight fructose polymers, levans. Levans can accumulate and lead to toxicity or the hydrolysis leads to inappropriate transfer of fructose residues that can be toxic. Hence, the *sacB* gene is lethal in the presence of sucrose and this allows for selection of bacteria containing the target gene, as the *sacB* gene is disrupted by the target gene upon successful insertion in the vector.

The vector was transformed into One Shot® TOP10 Chemically Competent *E. coli* by addition of 50 µl cells to the transformation mix in an Eppendorf tube. After 10 minutes on ice, the cells were heat shocked at 42 ºC for 35 seconds and replaced on ice for two minutes. The cells were then mixed with 200 µl sterile LB medium and incubated at 37 ºC for 45-60 minutes with 220 rpm horizontal shaking. Finally, the cells were plated out on LB agar with kanamycin and sucrose in a laminar flow cabinet and incubated in 37 ºC cabinet overnight.

3.3.8 Colony PCR

The following day colony PCR was performed to verify the presence of the target gene in the transformants.

A few colonies from the incubated plate were picked with a toothpick and scratched on the inside of a PCR tube. The tubes were microwaved for 2 minutes and 0.5 µl forward and reverse primer (10 pmol), 6 µl Red Taq DNA polymerase 2x Master Mix and 5 µl Milli-Q water were added. Subsequently, a PCR was carried out using the program in Table 3.3.

Table 3.3. Colony PCR program

Repeats	Temperature (ºC)	Time (min:sec)
1	95	00:30
Cycle 35 rounds	95	00:10
	55	00:30
	72	2:00
1	72	10:00
1	10	Hold

PCR products were verified by agarose gel electrophoresis for correct size of insert. Verified colonies were picked for overnight incubation according to section 3.2.4.

3.3.9 Plasmid purification

Glycerol stocks were made of 700 µl overnight cultures of positive clones (see section 3.1) and the rest of the cultures were used for plasmid purification. Cells were harvested by centrifugation at 5000 g for 5 minutes in a Heraeus™ Multifuge™ X1R centrifuge, before plasmid purification using the NucleoSpin® Plasmid Kit. The cells were resuspended in 250 µl buffer A1 and transferred to an Eppendorf tube. Then, 250 µl buffer A2 was added and the tubes inverted 6-8 times followed by incubation at RT for 5 minutes. Subsequently, 300 µl buffer A3 were added and the tubes inverted thoroughly, followed by centrifugation at 13000 rpm for 5 minutes in a Benchtop centrifuge 5418 R. A 750 µl aliquot of supernatant was pipetted to columns in collection tubes, and the tubes were centrifuged at 13000 rpm for one minute. Flow-throughs were discarded and 500 µl buffer AW was added, before a new centrifugation (13000 rpm, one minute). The membrane was washed again with 600 µl buffer A4, followed by a 2-minute centrifugation without buffer (13000 rpm) to dry the membrane. The columns were placed in clean Eppendorf tubes and 50 µl water was added for elution of the plasmids. After one minute of RT incubation, the columns were centrifuged for one minute at 13000 rpm, and plasmids were stored for sequencing and transformation into BL21 *E. coli* cells. Plasmid concentrations were measured by Qubit™ as described in section 3.3.5.

3.3.10 DNA sequencing

All plasmids were verified by DNA sequencing of the complete inserted gene.

Eppendorf tubes with 2 µl sequencing primer (forward or reverse) and 10 µl plasmid DNA were prepared, labelled by barcodes and submitted for sequencing at myGATC. Only the verified plasmids, containing the correct gene sequence, were used further.

3.3.11 Transformation into BL21 *E. coli*

Verified plasmids were transformed into One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*.

A 2 µl aliquot of plasmid solution was mixed with 50 µl BL21 cells that had been thawed on ice. The mixture was placed on ice for 10 minutes and heat shocked at 42 ºC for 35 seconds. After 2 minutes on ice, 200 µl LB medium was added to the cells, under sterile conditions, before incubation at 37 ºC for 45-60 minutes with 220 rpm horizontal shaking. A 100 µl aliquot of cells was plated out on LB agar with kanamycin and the plates were incubated at 37 ºC overnight.

3.4 Protein expression and purification

3.4.1 Protein expression

Overnight cultures of BL21 cells were made from plate colonies (previous section) according to section 3.2.4 and used for protein expression.

A portion of the overnight culture was used to make a glycerol stock and the rest was used to inoculate 500 ml sterile LB medium, containing kanamycin (50 µg/ml) and 100 µl Antifoam 204, in a 1 l Pyrex bottle. The bottle was sealed by aeration caps, and incubated with aeration in a Harbinger LEX bioreactor system using a 37 ºC water bath. Optical density (OD) at 600 nm was measured after 2 hours and every 30 minutes from then, until the OD reached 0.6-0.8. The culture was then moved to room temperature for cooling and 100 µl culture was sampled (un-induced sample). After 15 minutes cooling, protein expression was induced by IPTG to a final concentration of 0.5 mM and the culture was left overnight.

The next day 100 µl culture was sampled (induced sample). The samples (un-induced and induced) were centrifuged in Benchtop centrifuge 5418 R, washed with protein buffer before resuspension in LDS-buffer and run on lithium dodecyl sulphate polyacrylamide gel electrophoresis (LDS-PAGE) gel (see section 3.4.5) to check for protein expression.

3.4.2 Harvest of protein expressing cells

An aliquot of 250 ml culture was centrifuged at 8000 rpm for 10 minutes with a JA-10 rotor in an Avanti™ J-25 centrifuge using 500 ml centrifuge tubes. The cell free supernatant was discarded. This step was repeated to harvest the rest of the culture in the same centrifuge tube. Centrifuge tubes with cell pellet were frozen at -20 ºC for short-term storage until further use.

3.4.3 Lysing of cells for cytosolic protein extraction

Frozen cell pellets were dissolved and lysed to prepare for purification.

A frozen cell pellet was transferred into a 40-ml beaker with 20 ml buffer A (see section 3.4.4) and PMSF protease inhibitor (0.1 mM). A magnetic stirrer was used to stir the mixture for 10 minutes to resuspend the pellet, before addition of 1 ml Bug Buster. The mixture was then stirred for another 10 minutes. The cell suspension was sonicated on ice at 25 % amplitude for 30 seconds with 1-second pulses. This was repeated four times before the lysate was centrifuged for 20 minutes at 20 000 rpm with rotor JA-25.50 in an Avanti™ J-25 centrifuge. The supernatant, with soluble protein, was filtrated through a 0.22 µm filter to a 50-ml falcon tube and kept on ice. Purification with IMAC was conducted subsequently.

3.4.4 Immobilized Metal Affinity Chromatography

Cytosolic protein extract from lysed cells was purified by immobilized metal affinity chromatography (IMAC) to purify target proteins containing 6xHis-tags.

Immobilized metal affinity chromatography is based on the fact that histidine binds divalent metal ions, such as Ni^{2+}, with strong affinity. By recombinantly fusing the target protein with 6 histidine residues (6xHis-tag), the protein can be purified by loading it onto a Ni-NTA column and thereafter eluting when all other proteins have been washed off. This is done by loading the protein with a binding buffer (buffer A) containing a low concentration of imidazole, the imidazole will hinder proteins with low affinity to bind by competitive binding to the resin, allowing weak binding proteins to pass through the column along with proteins with no affinity

for the resin. An elution buffer (buffer B) with high imidazole concentration is later applied, to elute the target protein which has a high affinity for the resin.

The IMAC was performed using an ÄKTAprime plus chromatography system with a 5 ml HisTrap™ FF Ni sepharose column. The column was regenerated for every protein purification. This was done by first washing the whole system and column in water to remove ethanol. Then 5 ml strip buffer (see section 2.7) containing EDTA was injected to strip the column resin for nickel. Nickel fractions were collected and discarded as chemical waste. The column was then washed with water before 3 ml 100 mM $NiSO_4$ was injected. Nickel bound to the column and excess nickel was washed out with water and collected. The column was equilibrated with 5 column volumes of buffer A (see section 2.7) and the system was then ready for a new purification.

The protein sample from section 3.4.3 was loaded on the column (1ml/min) using a sample pump. Unwanted, contaminant proteins were washed off the column using buffer A, and buffer A loading was continued until the UV absorbance reached baseline, meaning that no more protein went through the column. Proteins were eluted with a gradient to 100 % buffer B (see section 2.7) over 100 ml at 2 ml/min, and 2 ml fractions were collected. After a clear peak in UV trace was observed, the gradient was switched to 100 % buffer B and stopped after 2x column volumes to flush out residual bound protein. Samples (5 µl) were taken from the lysate, flow-through, wash and all peak fractions for LDS-PAGE analysis. Fractions containing purified target protein(s) (verified by LDS-PAGE) were collected for buffer exchange and concentration. The Äkta system and columns were washed with water and stored in 20 % ethanol.

3.4.5 Lithium dodecyl sulphate polyacrylamide gel electrophoresis (LDS-PAGE)

LDS-PAGE can separate proteins by size. The proteins are denatured and their disulphide bonds are reduced by adding a sample buffer and reducing agent, and boiling the sample. The sample and a protein ladder are applied on a stain-free gel with pores, and the gel is placed in an electric field. The size of the proteins can be estimated due to varying migration velocities through the porous gel based on molecular weight, as their native structure and charge are disrupted.

Samples were prepared for LDS-PAGE by mixing 5 µl NuPAGE® LDS Sample Buffer (4X) and 2 µl NuPAGE® Sample Reducing Agent (10X) with 13 µl of appropriately diluted protein.

The samples were boiled for 5 minutes. Meanwhile, a Mini-PROTEAN® TGX Stain-Free™ Gel was placed in a Mini-PROTEAN® Tetra Electrode Assembly and the assembly was placed in a buffer tank. Fresh 1x TGS buffer was used to fill the electrode gasket while used 1x TGS filled the rest of the tank. Samples (10 µl for 15 µl wells or 15 µl for 30 µl wells) and 7 µl Benchmark™ Protein Ladder were applied to the gel wells. Gels were run at 280 V for 18 minutes using a PowerPac™ 300 power supply. Finally, the gel was visualized by UV-activation of the proteins in the Stain-Free™ Gels with a Gel Doc™ EZ Imager. The UV induces a reaction between the trihalo compounds in the stain-free gel and the tryptophan residues in the proteins, producing a fluorescent signal.

3.4.6 Protein buffer exchange and concentration

After IMAC purification the target protein was in a buffer with imidazole, which can affect the protein, and buffer exchange was performed.

Protein fractions from IMAC purification were collected and placed in a VIVASPIN 20 filtration column with a 10,000 MWCO PES membrane from Sartorius and centrifuged in a Heraeus Multifuge X1R at 4500 rpm and 4 ºC. The buffer was changed by sequential concentration and dilution with 50 mM Tris-HCl pH 8.0 and 200 mM NaCl. The protein solution was concentrated to a volume of approximately 1 ml and the protein concentration was measured (section 3.4.7) before storage at 4 ºC.

3.4.7 Protein concentration measurement

Protein concentration measurement was conducted using the UV absorbance method and protein extinction coefficients estimated by ProtParam (Gasteiger et al. 2005).

To measure absorbance, UVette® Eppendorf disposable UV cuvettes and an Eppendorf D30 BioPhotometer were used. The sample was diluted to produce an A_{280} value between 0.1 and 1. The absorbance value (A), cuvette length (d = 1 cm) and the proteins' estimated extinction coefficients (ε) were used to calculate the concentration (c) through the formula: $c = \frac{A}{d*\varepsilon}$.

3.5 Expresso® solubility and Expression Screening System

Efforts to express GH5A, GH5A_cat and GH5A_d2 proteins using the pNIC-CH clones and various expression conditions were without success. Expression led to inclusion bodies and the

Expresso® solubility and Expression Screening System from Lucigen® was therefore used to screen for soluble fusion products.

The system contains 7 different fusion tags (Table 3.4) for solubilizing proteins. The fusion tag genes are present in a pSol vector. The vector contains a rhaP$_{BAD}$ promoter for rhamnose induced expression. The fusion tags are N-terminally attached to the desired protein and the tags contain a TEV-protease site for cleaving off the tag and a 6xHis-tag for purification and easy removal after cleavage.

Table 3.4. Expresso Solubility N-terminal fusion tag lengths and sizes.

Fusion tag	Amino acid length	Size (kDa)
AFV	113	13.5
SlyD	210	22.7
Tsf	297	32.2
SUMO	115	13.3
Bla	381	41.3
MBP	382	42.1
GST	233	27.4
Control	14	1.8

Detailed procedures and chemicals are shown in Lucigen®'s protocol for Expresso® Solubility and Expression Screening System.

The desired gene was amplified with traditional PCR method and primers specially designed for the Expresso® system. The primers contain overlaps to the Expresso® vectors. The amplified DNA was verified by agarose gel electrophoresis, the bands were cut out and DNA purified by gel clean-up. The purified DNA, now with complementary overhangs to the vectors, was mixed with supplied linearized, cloning-ready pSol vectors, containing the different fusion tags. The plasmid was transformed into *E. cloni*® 10G Chemically Competent Cells included in the kit. Transformants were screened for the presence of the target gene by colony PCR and overnight cultures were made. The overnight cultures were used to make glycerol stocks, purify plasmids and for inducing protein expression with rhamnose. Purified plasmids were verified for the correct insert by sequencing. Induced cells were harvested, lysed and run on LDS-PAGE to analyse protein solubility. Two positive constructs for each protein were chosen for upscaling to 20 ml, and further upscaling to 500 ml after activity assays had been conducted. Tagged

protein were harvested and purified with IMAC, and run through buffer exchange and concentration as described in section 3.4.6. The fusion tags were cleaved, from the target proteins, at their TEV-protease site by TEV protease from Sigma-Aldrich. The his-tagged protease and fusion tag were removed from the protein solution by retention in the IMAC column in a second purification step (as described in section 3.4.4), while the non-his-tagged protein of interest were not retained. Finally, the target protein, now without fusion tag, was concentrated and stored at 4 ºC.

3.6 Enzyme characterization

3.6.1 Enzyme assays

Enzymatic activity was measured by analysing the products through the DNS method or HPAEC-PAD. In the DNS method, reducing ends are measured by their reaction with 3,5-dinitrosalicylic acid which leads to the formation of the reduced, orange coloured 3-amino-5-nitrosalicylic acid. The intensity of the colour determines the amount of reducing sugars. HPAEC-PAD gives the amount and type of product, through signal intensity and retention time respectively, compared to a standard curve, explained in section 3.6.2.

Determination of initial rapid release of products/initial rates, optimum pH and temperature, temperature stability and comparative activity on CMC (under optimal conditions) for all the enzymes were performed using the standard method described below, with modifications indicated in the subsequent sections.

Reaction mixtures without enzymes were prepared in triplicates in 2 mL reaction tubes. Standard reactions contained substrate (1 % (w/v) CMC or 0.5 % (w/v) β-glucan), 20 mM citric acid buffer pH 5,5 and Milli-Q water to a final volume of 200 µl (including enzyme). The reactions were mixed and pre-heated in a thermomixer for 5 minutes at 40 ºC. Enzymes were added to the reactions and the tubes were inverted to mix. The reactions were incubated in the thermomixer with 800 rpm horizontal shaking at 40 ºC. After the desired incubation time, the amount of reducing sugars was determined using a modified version of the standard DNS method (Miller 1959). A 100 µl aliquot of the reactions was added directly to 100 µl of 1 % DNS reagent (see section 2.7) in a 96-well plate. In addition, triplicates of substrate blinds and water blinds were added, as well as a glucose standards with concentrations 0.1, 0.25, 0.5, 0.75 and 1.0 mg/ml. The 96-well plate was stored on ice until all the reactions had been added. The

plate was then sealed and incubated in a thermomixer at 95 ºC for 20 minutes, before it was cooled on ice. 150 µl heat-treated sample was transferred to a new plate, and absorbance was measured at 540 nm in a Multiscan™ FC Microplate Photometer. A glucose standard was used to calculate the release of glucose equivalents/reducing ends from the A_{540}-value.

3.6.1.1 Assessment of initial rate conditions

Initial rate enzyme assays were performed to determine the conditions (time and enzyme concentrations) at which the reactions were in initial rapid release of products. This was to ensure that subsequent assays were not affected by substrate depletion.

Reaction volumes were increased to 600 µl, and 100 µl was sampled at various consecutive time points (as indicated in Appendix C, Figure C1) to monitor the reaction progress curves. Otherwise, the assay was carried out as described in section 3.6.1.

The determined times and enzyme concentrations for each enzyme that yielded initial rapid release of products (see section 4.3.1.1, Table 4.3) were used in further assays unless otherwise noted.

3.6.1.2 pH optimum

To determine the optimal pH for enzyme activity, reactions were prepared with buffers of pH ranging from 3 to 9.5 in 0.5 intervals. The buffers used were citric acid buffer (pH 3-5.5), phosphate buffer (pH 6-7.5) and Tris buffer (pH 8-9.5). Otherwise, the assay was carried out as described in section 3.6.1. Reactions were run according to the pre-determined reaction times and enzyme concentrations that yielded initial rapid release of products (Table 4.3).

3.6.1.3 Temperature optimum

Optimal temperatures of the enzymes were determined by incubating the reactions at temperatures in 5 ºC intervals between 10-80 ºC. Otherwise, the assay was carried out as described in section 3.6.1. Reactions were run according to the pre-determined reaction times and enzyme concentrations that yielded initial rapid release of products (Table 4.3).

3.6.1.4 Temperature stability

Temperature stability was determined by pre-incubating the enzyme at 40, 50 and 60 ºC for various periods of time, before assaying the enzyme activity to assess residual enzyme activity

after temperature treatment. Otherwise, the assay was carried out as described in section 3.6.1. Enzyme assays were conducted with times and enzyme concentrations determined to yield initial rapid release of products (Table 4.3).

3.6.1.5 Comparative analysis of activity on CMC

Comparative activity on CMC was determined using the optimal pH (5.5) and temperature (40 ºC) elucidated in the previous assays. All enzyme concentrations were 150 nM in this assay for activity comparison on equal terms. Samples were taken after various time points to obtain activity curves. Otherwise, the assay was carried out as described in section 3.6.1.

3.6.1.6 Cellodextrin degradation

The enzymatic activity on cellopentaose (glc_5) and cellohexaose (glc_6) was examined and products were analysed by High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) on a Dionex™ ICS-3000 Ion Chromatography System.

Reaction mixtures with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and Milli-Q water to a total volume of 600 µL (including enzyme) were prepared. Reactions in 2 mL tubes were pre-incubated at 40 ºC for 5 minutes, before adding enzyme to a final concentration of 250 nM or 125 nM. Substrate blinds contained Milli-Q water instead of enzyme, to the same final volume. Samples were taken at various time points, and the reaction was stopped by addition of NaOH to a final concentration of 0.1 M. Samples were stored at -20 ºC before analysis by HPAEC-PAD (see section 3.6.2).

3.6.1.7 Activity on hemicelluloses

Enzymes were assayed for activity on eight hemicellulosic substrates with different composition and linkages to investigate enzyme specificity. Substrates were: CM-pachyman, konjac glucomannan, guar galactomannan, carob galactomannan, xylan, xyloglucan, lichenan and mannan (see section 1.2.2 for details). Reactions were prepared in 96 well plates for all substrates in triplicate, and blanks for each substrate were included. Substrate (1 % w/v), 20 mM citric acid buffer pH 5.5 and water were prepared to a final volume of 100 µl (including enzyme). The reactions were pre-heated at 40 ºC for 5 minutes before 1 µM enzyme was added, and the reactions were incubated at 40 ºC without shaking. After 1 hour, the reaction was stopped by adding 100 µl DNS reagent to each well. Glucose standards were included in the

plate, and the plate was heated and absorbances were measured as described above (section 3.6.1).

3.6.1.8 Filter paper degradation

Filter paper was used as a substrate to assay enzymatic activity on crystalline cellulose. These assays included reactions with multiple enzymes to test for potential synergistic relationships. Reactions with a beta-glucosidase were performed to check for product inhibition.

Reaction mixtures with total volumes of 400 µl in 2 ml tubes were set up in triplicates. Reactions contained 1 % filter paper (particle size 0.5 mm), 20 mM citric acid buffer pH 5.5, 1 µM enzyme and water. The assays were performed at 40 ºC with 800 rpm horizontal shaking. At time points 1, 4, 24, 48 and 72 hours, 50 µl was sampled and mixed with 50 µl 0.2 M NaOH in an Eppendorf tube to stop the reaction. The tubes were centrifuged in benchtop centrifuge 5418 R at 14000 rpm for 5 minutes to pellet the filter paper. A 75 µl aliquot of the supernatant was transferred to 0.3 ml Snap Ring Micro-Vials for HPAEC-PAD analysis (section 3.6.2) and stored in 4 ºC until further use.

3.6.1.9 Switchgrass degradation

Reactions with switchgrass were performed to study the activity of the enzymes on a natural and complex substrate, and explore activity towards other components than cellulose. Switchgrass is a promising biomass crop, primarily composed of cellulose, xylan and lignin (on average 38.0 % glucose, 22.8 % xylose and 22.1 % lignin) (Hu et al. 2010).

Reaction triplicates were set up with 0.2 % switchgrass (ball milled and washed) obtained from The Noble Foundation (Ardmore, OK, USA), 20 mM citric acid buffer pH 5.5 and 1 µM enzyme in a total volume of 150 µl, in 2 ml Eppendorf tubes. Substrate blanks were also included. Reactions were performed at 40 ºC and 800 rpm horizontal shaking for 24 hours. Reactions were stopped by boiling for 10 minutes and filtered using a 96-well filter plate. Aliquots of 100 µl reaction mixture were mixed with the DNS reagent for analysis, whereas the remaining reaction mixture was transferred to vials for HPLC-MS. HPLC-MS was performed with a porous graphitic carbon (PGC) column by experienced researchers in the lab, and is not described further in this thesis. HPLC-MS results were analysed using the Xcalibur™ software (Thermo Scientific). For a detailed description of the analysis see Arntzen et al. (2017).

3.6.1.10 Binding of Cel5A_d2 to Avicel

Assays of binding to Avicel were conducted for the putative carbohydrate-binding module (CBM) domain of Cel5A (Cel5A_d2).

Reactions were prepared in triplicates in 2 ml tubes, in addition to substrate blanks (no Avicel), enzyme blanks (no enzyme) and reaction with lysozyme as a non-binding control. Reactions contained 10 mg/ml Avicel, 20 mM citric acid buffer pH 5.5, 0.04 mg/ml Cel5A_d2 and Milli-Q water to a total of 600 µl. At timepoints 2.5, 5, 15, 30 and 60 minutes, a 75 µl reaction aliquot was transferred to a 96-well filter plate with low protein binding Duapore® membrane (Merck) and filtered to remove Avicel. The OD at 280 nm was then measured and converted to mg/ml of unbound protein using calculated extinction coefficients (ExPASy's ProtParam tool).

3.6.2 High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)

Products obtained after degradation of cellodextrins and filter paper were analysed by HPAEC-PAD to obtain higher sensitivity than DNS analysis and to enable the detection of type of products generated.

HPAEC-PAD is a method where the sample of interest is run through a stationary phase (column) by a mobile phase (eluents), and different constituents are retained by the stationary phase for different time lengths. For the detection of oligosaccharides in this study, high pH was used to deprotonate hydroxyl- and carboxyl groups. Longer, linear oligosaccharides will be more charged than short oligosaccharides and therefore be retained longer by the positive stationary phase. Charged oligosaccharides will be separated based on their length, composition, linkage types and level of formal negative charge. Sodium acetate was used to elute the oligosaccharides, as it has a high ionic strength, and gradient concentrations were used to obtain separation. A gold anode was used for detection of the oligosaccharides as they eluted, by measuring the electric current the oligosaccharides create at the electrode. Known concentrations of oligosaccharide standards were used to determine the type of oligosaccharide and calculate the amount of eluted product based on standard curves.

HPAEC-PAD was conducted using a Dionex™ ICS-3000 Ion Chromatography System equipped with a Dionex™ CarboPac™ PA1 column and the software Chromeleon®.

Aliquots of 75 µl were transferred to 0.3 ml Snap Ring Micro-Vials, avoiding air in the vial bottom, and the vials were capped. Cellodextrin standards as well as the samples were placed in the auto sampler. Eluents were made according to section 2.7 and connected to the correct channels of the ICS. The nitrogen flow was opened and the air in the eluent bottles were changed by opening and closing the bottle lids three times. Next, the channels were purged to the new eluents and afterwards the baseline was monitored and auto zeroed when stable. A pre-setup program in Chromeleon® was run at 0.250 ml/min flow and 30 ºC to analyse the samples, blanks and standards. The program runs 0.1 M NaOH before elution of oligosaccharides with a multi-step linear gradient: 0.1 M NaOH to 0.1 M NaOH/0.1 M NaOAc in 10 minutes; to 0.1 M NaOH/0.14 M NaOAc in 4 minutes; to 0.1 M NaOH/0.3 M NaOAc in 1 minute; to 0.1 M NaOH/1.0 M NaOAc in 2 minutes, before recondition of the column by 0.1 M NaOH for 11 minutes. Data were analysed in Chromeleon®.

3.6.3 Crystallization by hanging drop technique

Efforts were made to crystallize the second catalytic domain of Cel5C (Cel5C_2), in order to obtain the protein structure and gain knowledge of the working mechanism.

In the hanging drop technique, a drop of 50 % protein solution and 50 % precipitant solution is placed on a cover slip and sealed over a reservoir of the precipitant solution. Since the precipitant concentration is lower in the drop, water will vaporize to the reservoir, to obtain the same concentration in the drop as in the reservoir. This will lead to an increase in the protein and precipitant concentration in the drop, and relative supersaturation of protein will slowly increase. Nucleation and crystal growth may occur as the solution reaches supersaturated conditions.

For crystallization, the Emerald Biosystems Wizard™ 2 Screen was used with the hanging drop technique. The screen contains 48 different conditions with varying concentrations and compositions of salts, buffers and precipitation reagents. A 200 µl aliquot of each solution was distributed to respective reservoirs in a 24-well crystallization plate (2 plates). Thereafter, 1 µl of the reservoir solution was mixed with 1 µl 18.68 mg/ml enzyme (Cel5C_2) on a cover slip and the cover slip was turned upside down over the reservoir, which had silicone grease around the top to seal. The crystallization plates were incubated at room temperature and checked regularly using a microscope and score sheet.

3.7 Metaproteomics

Metaproteomic studies were performed on a sample from cow rumen, to investigate the presence of the AGa cluster enzymes. The sample was from a cannulated cow, fed on a mixed diet containing 60 % fibre from Hess et al. (2011). Notably, this sample was not the same as the switchgrass-enriched sample used in the metagenomic study by Hess et al. (2011), from which the AGa genome was reconstructed.

Metaproteomics is the study of all proteins present in a microbial environment. Supplementing metagenomics (study of all gene material), metaproteomics can give information of gene expression and metabolic activity by investigating which proteins are being produced at a given time.

3.7.1 Sample preparation

The cow rumen sample was distributed to 10x 2 ml Eppendorf tubes and centrifuged at 16000 g for 5 minutes in an Eppendorf benchtop centrifuge 5418 R (also used in all following centrifugations unless otherwise noted). The supernatants, containing secreted proteins (secretome), were combined in a tube and kept on ice. Pellets were resuspended in 500 µl dissociation solution (DSS) (see section 2.7) to dissociate the cells from the sample material and centrifuged at 100 g for 30 seconds. Supernatants, with cells, were transferred to a large tube. This step was repeated four times. Next, the large tube with cells was centrifuged at 4300 g for 10 minutes in Heraeus Multifuge X1R. The supernatant was discarded and the cell pellet was washed with 2 ml cell-wash buffer (see section 2.7). The cells were transferred to a bead beating tube and centrifuged at 8000 g for 5 minutes. The supernatant was removed, and 400 µl lysis buffer (see section 2.7) was added, together with glass beads to lyse the cells. After 30 minutes on ice, the cells were disrupted by 3 x 60 second pulses on FastPrep-24™ 5G homogenizer. The disrupted cells were then centrifuged at 16000 g for 15 minutes and the supernatant, with intracellular proteins, was used in further steps.

3.7.2 LDS-PAGE and staining

The secretome and the intracellular protein sample were concentrated on a VIVASPIN 20 filtration column with a 10,000 MWCO PES membrane from Sartorius and concentration was measured as described in section 3.4.7. The samples were analysed using LDS-PAGE, with 19.5 µl sample and 10.5 µl LDS-solution applied to the wells. The LDS-PAGE gel was stained

with Coomassie Brilliant Blue by incubating the gel in staining solution (see section 2.7) for one hour, before rinsing with water. Thereafter, the gel was destained with destaining solution (see section 2.7) for 20 minutes two times. Finally, the gel was left in Milli-Q water and a minor amount of buffer A (see section 2.7) overnight. All staining incubations took place on a slow horizontal shaker.

3.7.3 De-colouring and cleaning of gel pieces

The stained gel was visualized and the intracellular protein and secretome lane were divided in 10 bands and cut in 1x1 mm cubes. The cubes were transferred to Eppendorf Protein LoBind tubes, one tube for each band. 200 µl of Milli-Q water was added to the tubes, which were then incubated at room temperature for 15 minutes. Liquid was removed and 200 µl 50 % acetonitrile (ACN)/25 mM ammonium bicarbonate (Ambic) was added before an additional 15-minute incubation at room temperature, and this process was repeated once. A 100 µl aliquot of 100 % ACN was then added and the gel pieces were incubated at room temperature for 5 minutes. The liquid was removed and the pieces were air dried for 1-2 minutes.

3.7.4 Reduction and alkylation

The dried gel pieces, with 50 µl DTT solution (see section 2.7), were incubated at 56 ºC for 30 minutes in a thermomixer. The dithiothreitol (DTT) solution was removed and 50 µl iodoacetamide (IAA) solution (see section 2.7) was added, before 30-minute room temperature incubation in the dark. The IAA solution was removed and 200 µl 100 % ACN was added. After 5-minutes incubation at room temperature, the liquid was removed and gel pieces were left for air drying.

3.7.5 Digestion with trypsin

The gel pieces were digested by addition of 30 µl trypsin solution (see section 2.7) and incubated on ice for 30 minutes. Additional trypsin buffer (see section 2.7) was added to cover the gel pieces, before incubation at 37 ºC over-night in a thermomixer.

The next day samples were cooled down, combined with 40 µl of 1 % trifluoracetic acid (TFA) and sonicated in a water bath for 15 minutes.

3.7.6 C_{18} solid phase extraction

Snap Ring Micro-Vials (0.3 ml) with 10 µl 70 % ACN/ 0.1 % TFA were prepared in advance. ZipTips with C_{18} material were used to desalt and purify the samples. The tips were conditioned and equilibrated prior to use by; first pipetting and discarding 10 µl 100 % MeOH, then pipetting and discarding 10 µl 70 % ACN/ 0.1 % TFA and finally pipetting and discarding 10 µl 0.1 % TFA. The samples were bound to the C_{18} material by pipetting and releasing 10 µl of the sample. The tip was then wiped on the outside with a tissue and washed with 10 µl 0.1 % TFA. The bound peptides were eluted by repeated pipetting four times in the solution prepared in vials (70 % ACN/ 0.1 % TFA). Tips were re-equilibrated before the next sample, and changed for each lane (secretome/intracellular proteins).

3.7.7 nanoLC-MS/MS preparation and run

The peptide samples were dried using a vacuum concentrator and the dried peptides were dissolved in 10 µl 2 % ACN/ 0.1 % TFA. The peptides were analysed by nanoLC-MS/MS by researchers in the lab (see Arntzen et al. (2017)). Outcome peptide fragments were mapped back to a reference library, made from the annotated ORFs of the AGa genome, by Adrian Naas.

4 Results

Previous analyses of a cow rumen metagenome (Hess et al. 2011) predicted that one of the Bacteroidetes-affilated reconstructed genomes (hereafter AGa) originated from a cellulose degrading bacterium (Weimann et al. 2013). Further investigation of the AGa genome led to the discovery of a possible cellulose-targeting gene cluster (Konietzny et al. 2014) (Figure 4.1), containing putative cellulose targeting GH5s, in addition to a GH3 and a GH94 putatively targeting cellulose degradation products. The aim of this study was to clone, express and characterize the four GH5 containing enzymes of the AGa cellulase cluster, to study their substrate targets and to investigate the potential role of the gene cluster in cellulose degradation. In order to do this, the individual genes were further analysed using bioinformatic tools, before the genes were synthesized and cloned for heterologous production and purification of the enzymes, which then were biochemically characterized.

4.1 Bioinformatics

4.1.1 Gene cluster organization and protein domain structures

The putative cellulose targeting gene cluster was further analysed using BLAST® (Altschul et al. 1990) against the NCBI database of non-redundant protein sequences, dbCAN (Yin et al. 2012) and the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2017). The gene cluster organization is shown in Figure 4.1.

Figure 4.1. AGa cellulase gene cluster organization. GH5 containing genes targeted for production and characterization in this thesis are highlighted in red. Faded genes have not been studied in this thesis, but are of importance for the general cellulolytic mechanism. See text for details on the gene cluster.

The cluster contains four GH5 domain encoding genes (*GH5A-D*) (red), identified through the Conserved Domain Database (CDD), all annotated to subfamily 2 based on closest homologues found using BLAST. In addition, a GH3 family β-glucosidase (yellow) and a GH94 family cellobiose phosphorylase (pink) with a CBM domain are present in the cluster. The cluster also contains genes encoding putative transport proteins (blue); annotated as a Major Facilitator Superfamily (MFS) transporter and an outer membrane (OM) β-barrel protein possibly involved

in sugar transport, in addition to a lipoprotein with unknown function. The individual GH5 genes (*GH5A-D*) have been the focus of this thesis and the encoded proteins (Cel5A-D) are shown in Figure 4.2 with their putative domains. Figure 4.2 also shows truncated protein variants that were made and studied here.

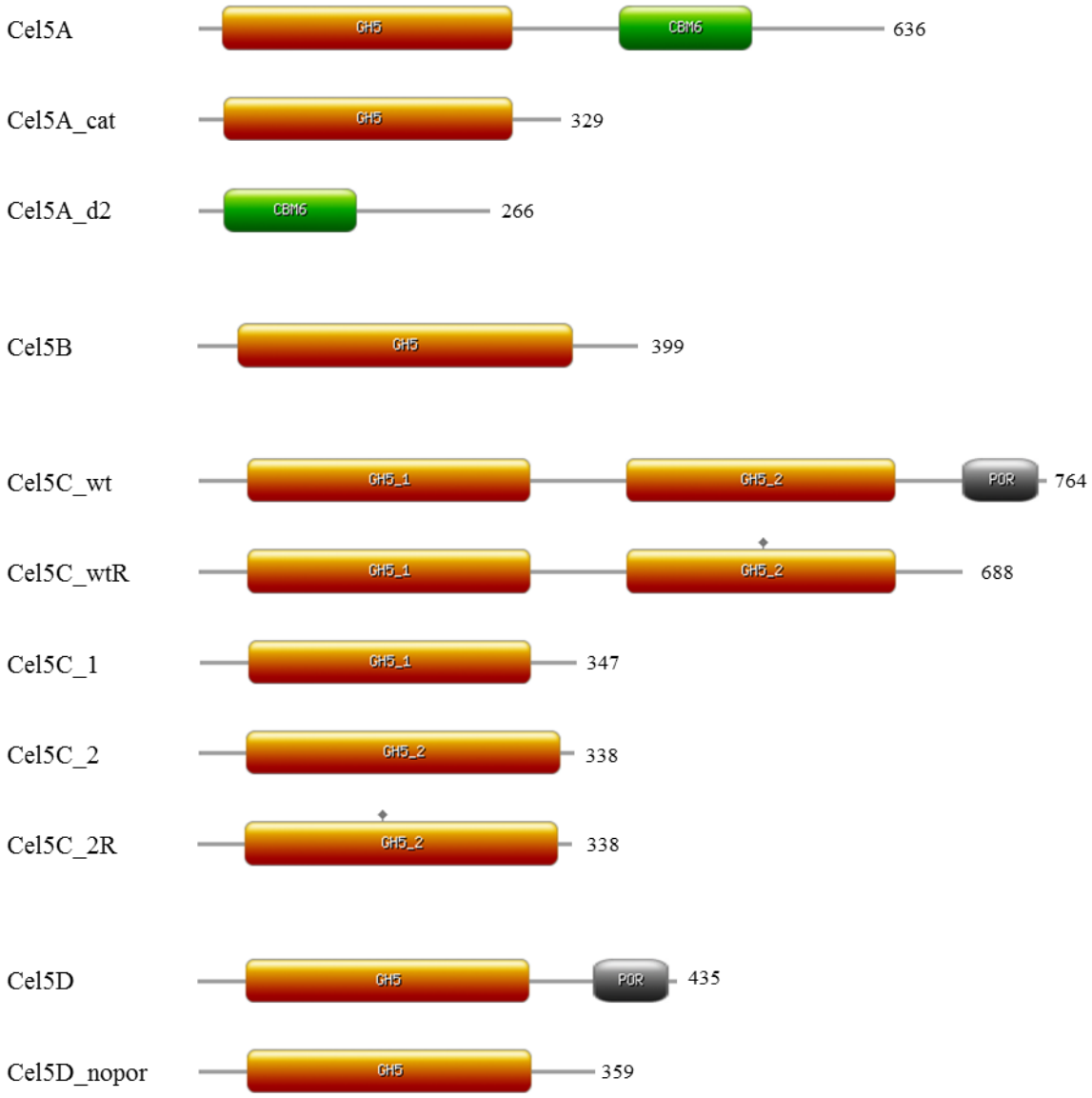


Figure 4.2. Domain structures of wild-type and truncated enzymes used in this study. Grey tags on Cel5C_wtR and Cel5C_2R mark restored amino acids; see text for details. Figures were made with the ExPASy PROSITE MyDomains Image Creator.

Cel5A and Cel5C are multi domain proteins (full-length Cel5C hereafter Cel5C_wt), while Cel5B and Cel5D only contain one GH5 domain. Cel5C_wt has two separate GH5 catalytic domains (Cel5C_1 and Cel5C_2 according to the position from the N-terminal end). Cel5A has one GH5 catalytic domain (Cel5A_cat) and one predicted carbohydrate binding domain

(Cel5A_d2) that has low affilation to CBM6 (low coverage: only 44 amino acids matching the motif, with E-value 2.14 e-04), suggesting a possible novel CBM diverging from the established families. All the enzymes, except Cel5B, contain a C-terminal secretion tag (Por-tag) for the Type IX Secretion System found in Bacteroidetes.

To examine the effect of the different domains of Cel5A and Cel5C_wt, both full-length and truncated versions of the proteins were produced. The effect of retaining the intact Por-tag was also investigated by production of Cel5D with and without the tag.

4.1.2 Protein parameters

The GH5 enzymes from the AGa gene cluster were subjected to bioinformatic analysis. Table 4.1 lists the lengths, protein weights and extinction coefficients of the proteins studied here. These parameters were used to calculate enzyme concentrations and check for correct enzyme weights in LDS-PAGE gels.

Table 4.1. Protein parameters. The number of amino acids and the molecular weights include the 6xHis-tag. Molecular weights and extinction coefficients were calculated using ExPASy's ProtParam tool assuming all cysteine pairs are forming disulphide bridges.

Enzyme	No. of amino acids	Molecular weight (kDa)	Extinction coefficient (M^{-1} cm^{-1})
Cel5A	636	70.2	141150
Cel5A_cat	329	36.2	80830
Cel5A_d2	266	29.3	51840
Cel5B	399	44.8	87165
Cel5C_wt	764	85.7	182755
Cel5C_wtR	688	77.0	179650
Cel5C_1	347	39.3	103290
Cel5C_2	338	37.7	76360
Cel5C_2R	338	37.7	76360
Cel5D	435	49.2	96300
Cel5D_nopor	359	40.7	93320

4.1.3 Multiple sequence alignment

To compare the catalytic domains of the enzymes (orange, Figure 4.2), a multiple sequence alignment (MSA) was made with MUSCLE at EMBL-EBI (Figure 4.3).

```
Cel5A           --ENGEAIQLKGWSSFGF-YGENCLTSANDLESMKGAGANCVRIARYLGNGSGSID----
Cel5B           CDKDGNPVQLAGMSTMGWQWCGDCYTKESIKTMVEEWGINVLRLAMYVEE-GGYNT----
Cel5C2          -----FDVQLRGVSSDNMSIYTRCYSTSSLTALANDWNASLFRISVNTNGKGGYCVNGSD
Cel5C1          VSECGKPVQLRGMSSHGLAWFPKCYTEASLTALVKDWNIDIFRLAIYTHEWGGYTT---N
Cel5D           VNECGKPVQLKGMSSHGLAWFPQCYTEESLSVLVNDWHIDIFRLAIYTHEKGGYCKTDGT
                       :** * *: .      * :  .     :    . .*::      .*

Cel5A           --------DNGIKNWMSWTAQKGMYCVIDWHILEAANGDGNP-GKYTNDAKNFFRMVAQE
Cel5B           ---NPIGFKQRMCEMIDICGELGIYCIVDWHILT----PGNPLDSKYGGAKEFFSFISKK
Cel5C2          QWLSMYDYNDKVDELVRLCGMRGLYCVVDWHLNE----GGDP-NAHLKEATVFWRHMAQR
Cel5C1          QWKSKDDYNAYIDNMVDICAKLGIYCIIDWHVLN--DGSGDP-NYTLDDAIPFWDYMSAK
Cel5D           QWKSKEDYNAYIDELVAICGKLGIYCIIDWHILQ--EGSGNP-KNTLDDAIPFWEYMSAK
                        .  : : :   .  *:**::***:       *:*       *  *:  ::

Cel5A           VADKKYKHIIYELCNEPSG------------------VGWGTIKSYAEDVIKTIVAI---
Cel5B           YANK--EHLLYEICNEPNNCLEKGDPIHPWVCTKETNVTWDMIADYADEIIPAIQGNYDS
Cel5C2          YTKF--THVIFEICDNASG------------------VEWSAIKSYADSMIALIRQF---
Cel5C1          HKDD--KHVLYEICNEPNG----------------FDVKWADVKEYAEAVIPVIRKN---
Cel5D           HKDD--KHVLYEICNEPNG----------------FMVRWSDVKEYADKVIPVIRAN---
                  .    *:::*:*::...                  * *  : .**: :*  *

Cel5A           --DKNKPVVIVGTPNWDQYIYSQV---------ASKGDLIN-----TNDAYVMYAFHLYA
Cel5B           LKVSH-PIVIVGTPQWDQLVDACLKEGMYQGNGKDLCDSLPARDARLKHDNVMYAFHFYA
Cel5C2          --DKN-KVIICGTPSNDREWSSVI---------SN-----P-----LSDSNVMYALHFTV
Cel5C1          --DPD-KIIICGTPTWSQDVDLAA---------QD-----P-----LSYDNVMYTLHFYS
Cel5D           --DPD-KIIICGTPMWSQDVDLAS---------QN-----P-----LSYNNVMYTLHFYS
                       ::* *** ...                .           .   ***::*:

Cel5A           NEAAHVGLESSEFLPAST----------RIPIFVSEWGLSSAQ-----------------
Cel5B           KEHNE-GFEKDGKPDYYNMYAYMYDVLGKLPVFCSEFGLCEANGNGELDPDRTDKWLLLL
Cel5C2          GTDGQ-SLRDKADAAISR----------GLPLFVSEFSLSPSNG-GSVNTTEAEQWITWM
Cel5C1          GTHTQ-YLRDKAQVAINK----------GLALFVTEFGTTQASGDGGVYFDECNTWMDWM
Cel5D           GDHFQ-SLRDKAQTALNN----------GAAIFVTEFGTTKASGDGGVFLDECNRWMEWM
                       : ..                   .:* :*:.   :.

Cel5A           -----------------PEK
Cel5B           SGNNAGKQVVSFCNWSFSDN
Cel5C2          K-----NQGLSWAN------
Cel5C1          D-----ARKISWVNWSFADK
Cel5D           N-----ERKISWVNWSFSDK
```

Figure 4.3. Multiple sequence alignment for the catalytic domains of the AGa GH5s. Catalytic glutamates were identified through MSA with enzymes (not shown) that had experimentally verified catalytic residues, and the catalytic glutamates are marked in yellow. The sequence marked in magenta shows the potentially mutated amino acids in Cel5C_2. The MSA was made with MUSCLE at EMBL-EBI.

There are evident similarities between the catalytic domains of the GH5s, but some enzymes stand out. Cel5C_2 differs from the other domains, lacking the first catalytic glutamate and the consensus sequence around it (magenta in Figure 4.3). To investigate this sequence feature closer and explore if this is a sequencing/assembly error or an evolutionary adaption, both the

wildtype Cel5C_2 and a restored (DNA→NEP) versions of the gene were synthesized, in addition to wildtype and restored versions of full-length protein.

4.1.4 Phylogenetic tree

To further investigate the relationship between the GH5 catalytic domains of the cluster, EMBL-EBI MUSCLE was further used to create a phylogenetic tree (Figure 4.4) of the catalytic domains from the MSA.

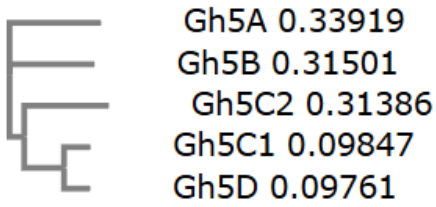


Figure 4.4. Phylogenetic tree of the catalytic domains of AGa GH5s. The tree is a neighbour-joining tree without distance corrections generated from the EMBL-EBI MUSCLE MSA. Numbers indicate number of substitution as a proportion of alignment length.

From the phylogenetic tree, it is clear that the Cel5A and Cel5B catalytic domains separate from Cel5C and Cel5D. Also, Cel5C_1 and Cel5D are the most similar and Cel5C_2 separates from these two. Still, all the catalytic domains belong to subfamily 2 of the GH5 family.

4.2 Cloning, expression and purification

All genes for the AGa GH5 proteins were cloned as described in section 3.3 and the resulting expression plasmids were transformed into chemically competent *E. coli* BL21 (DE3) cells. The proteins were produced by inducing protein expression with IPTG. Some of the proteins (Cel5A, Cel5A_cat, Cel5A_d2, Cel5C_wtR, Cel5C_2R) aggregated and formed inclusion bodies upon expression, and further measures were taken to produce soluble protein as detailed below.

After expression, cells were harvested and lysed to extract cytoplasmic protein, and the his-tagged proteins in the resulting cell-free extract were purified by IMAC. Subsequently

centrifugal tubes were used to exchange buffer and concentrate the proteins. Finally, protein concentrations were measured by the UV-method. Examples of an IMAC chromatogram (Figure B1) and LDS-PAGE of IMAC fractions (Figure B2) are shown in Appendix B. Protein yield for all the enzymes are listed in Table 4.2, and enzyme purity is visualized by an IMAC fraction for each protein in Figure 4.5.

Table 4.2. Protein yield. Listed are protein yields from expression and purification of the AGa enzymes. Protein yield is calculated from final enzyme concentration and volume after purification, buffer exchange and concentration and given as mg/l of culture used for expression.

Enzyme	Protein yield (mg/l culture)
Cel5A	8.1
Cel5A_cat	0.03 (5.1 before TEV cleavage)
Cel5A_d2	1.7 (13.9 before TEV cleavage)
Cel5B	8.4
Cel5C_wt	88.5
Cel5C_wtR	1.1
Cel5C_1	247.5
Cel5C_2	66.0
Cel5C_2R	1.2
Cel5D	58.9
Cel5D_nopor	398.6

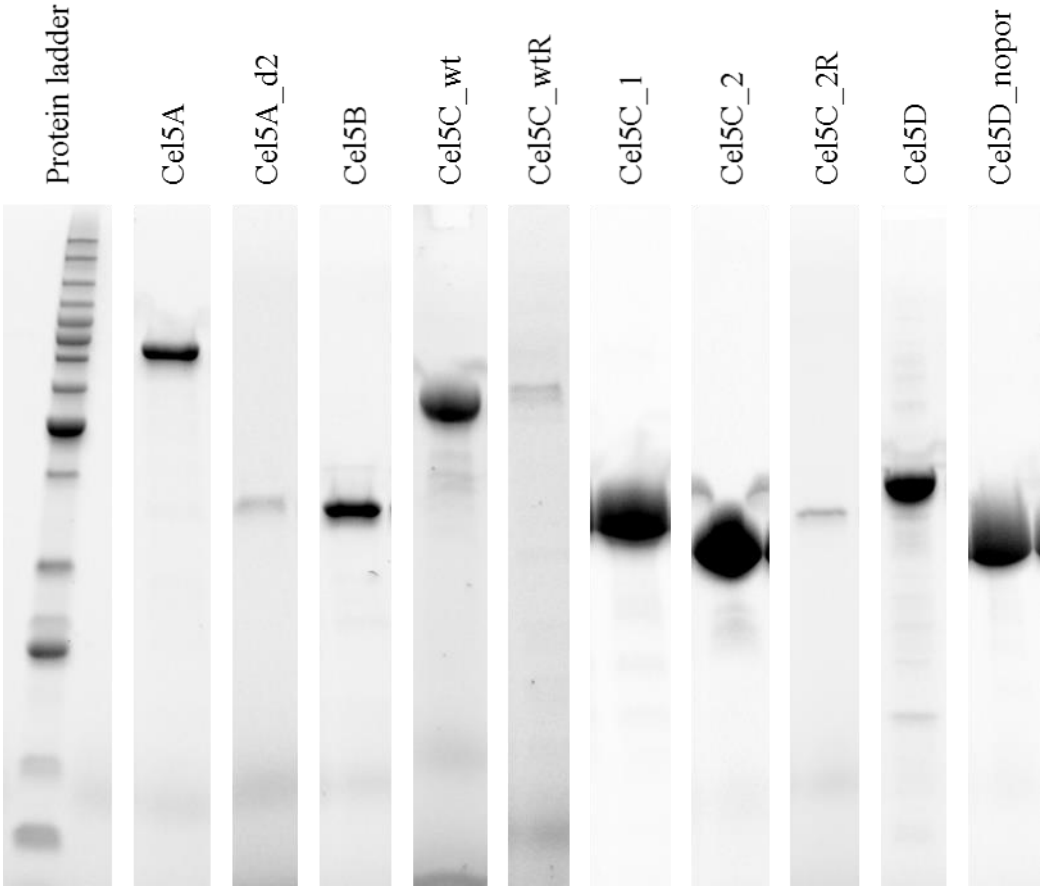


Figure 4.5. IMAC fractions from the purified enzymes. A representative fraction from the LDS-PAGE gel after IMAC purification is presented. Note that the ladder is not given for each protein, and deviations in molecular weight might occur between lanes. Cel5A_cat gave too low yield to be visualized (Table 4.2) and are not included in the figure.

As seen from the table, protein yield varied between the enzymes and this is assessed for each protein below. Figure 4.5 also show different intensities of the different enzyme bands, pointing towards differences in expression and purification. The enzymes contain some contaminant proteins, but appear relatively pure from the fraction gel picture.

The following chapters give an overview of the expression and purification of the individual protein constructs produced for characterization.

4.2.1 Cel5A

Standard expression of the full-length and the separate domains of Cel5A gave inclusion bodies. Therefore, expression at lower temperatures and lower IPTG concentrations was attempted, albeit without success. The Expresso® solubility and Expression Screening System (section

3.5) was then used and small-scale expression was applied to screen for soluble protein for the different fusion tag constructs. The most promising tags, which produced the greatest yield with low levels of insoluble protein, were chosen for upscaling. For Cel5A this was the control vector with a fusion tag containing only an N-terminal his-tag with a cleavable TEV-protease site, whereas for Cel5A_cat and Cel5A_d2 this was the Maltose Binding Protein (MBP) fusion tag. LDS-PAGE analysis of the solubility screening is shown in Figure 4.6.

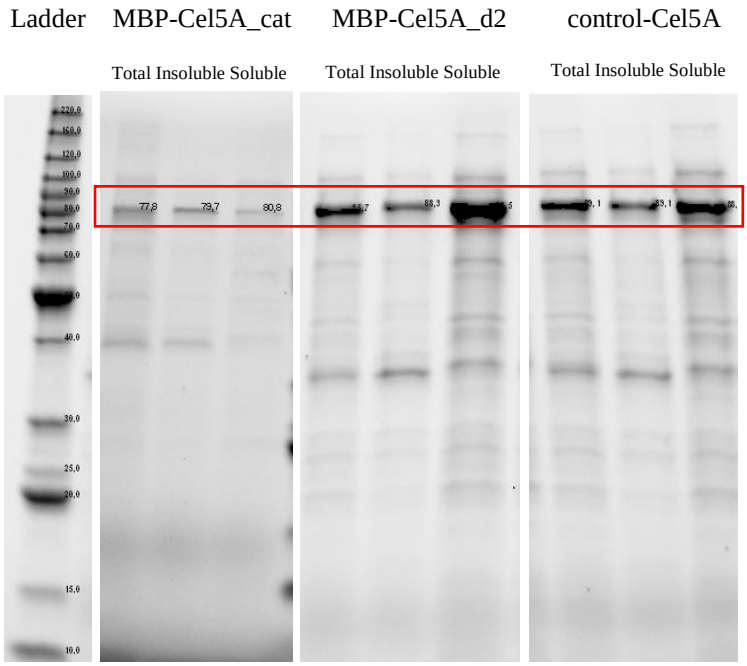


Figure 4.6. Solubility screening of Cel5A_cat, Cel5A_d2 and Cel5A. The figure shows results for the three proteins, with lanes from left to right showing: total protein, insoluble protein fraction and soluble protein fraction (red box). The small numbers indicate molecular weight. The enzymes were screened for all fusion tags, but included in this figure are only protein fusions chosen for upscaling (MBP- Cel5A_cat, MBP- Cel5A_d2 and control- Cel5A).

Upscaling of protein expression after solubility screening led to a higher fraction of insoluble protein than small scale expression. However, in all three cases some soluble protein was produced that could be purified by IMAC as described in the methods. Afterwards, TEV cleavage and purification was performed to remove TEV protease and the cleaved tag. Cel5A was produced with the control tag, and was not subjected to TEV cleavage. Protein yields are shown in Table 4.2.

As seen in Figure 4.6, expression of Cel5A_cat was low, and the final protein yield (Table 4.2) was too low for in-depth characterization. Cel5A_d2 and Cel5A were expressed in larger amounts and most of the protein was in soluble form. TEV cleavage and subsequent second purification step led to a large loss of protein for Cel5A_cat and Cel5A_d2 with the protein yield droping from 5.1 to 0.03 mg/l culture and 13.9 to 1.7 mg/l culture respectively.

4.2.2 Cel5B

Cel5B expression and purification offered no challenges. Yield is shown in Table 4.2.

4.2.3 Cel5C

Cel5C_wt, Cel5C_1 and Cel5C_2 were all successfully expressed and purified. As explained in section 4.1.2, restored catalytic residue-versions of Cel5C_wt and Cel5C_2 were also produced (Cel5C_wtR and Cel5C_2R respectively). Despite only a minor change in sequence, these latter two proteins formed inclusion bodies and only small amounts of soluble protein were obtained. Cel5C_wtR showed a higher amount of soluble protein, but much of the protein did not bind to the column during IMAC, and purification yields were low (Table 4.2). Because of the low yield of both Cel5C_wtR and Cel5C_2R, these could not be assayed for pH optimum, temperature optimum or temperature stability.

4.2.4 Cel5D

Cel5D was produced in two versions, with and without the Por secretion tag. The tag is normally cleaved off during transportation to the cell membrane when produced natively in Bacteroidetes. Expression and purification of both Cel5D and Cel5D_nopor was successful and gave a high yield without inclusion bodies (Table 4.2).

4.3 Characterization

To explore the enzyme activities, substrate targets and optimal reaction conditions, the purified enzymes were subjected to biochemical characterization. All enzymes were characterized, as far as protein yield allowed, with assays determining temperature and pH optima, temperature stability and activity on CMC, cellodextrins, various hemicelluloses, filter paper and switchgrass. In addition, binding of Cel5A_d2 to cellulose was assessed, and Cel5C_2 was subjected to crystallization attempts.

4.3.1 Enzyme assays

In all assays product formation was analysed by the DNS-method, except for the assays with cellodextrin and filter paper, the results of which were analysed by HPAEC-PAD. Product formation in the experiments with switchgrass was analysed by HPLC-MS in addition to the DNS-method.

4.3.1.1 Assessment of initial rate conditions

At the start of the enzyme characterization work, a series of assays, with varying enzyme dosages and incubation times was conducted to establish the conditions for initial rapid release of products. Conditions corresponding to initial rapid release of products were chosen based on two criteria, close to linear product formation over time and an approximately linear enzyme dose-response. Progress curves for each individual enzyme are shown in Appendix C, Figure C1. Table 4.3 shows the selected assay times and enzyme concentrations that were used for the different enzymes.

Table 4.3. Selected conditions for initial rapid release of products for the GH5s when acting on 1 % (w/v) CMC. These assay conditions were used for determination of pH optimum, temperature optimum and temperature stability assay. [1]Enzyme was not available in sufficient amount for in-debt characterization. [2]No enzymatic activity. [3]Enzymes were not available in sufficient amounts for pH and temperature characterization. [4]Too low activity, pH and temperature characterization was deemed unnecessary.

Enzyme	Time (min.)	Concentration (nM)
Cel5A	15	500
Cel5A_cat[1]	NA	NA
Cel5A_d2[2]	NA	NA
Cel5B	5	500
Cel5C_wt	5	50
Cel5C_wtR[3]	NA	NA
Cel5C_1	5	50
Cel5C_2[4]	NA	NA
Cel5C_2R[3]	NA	NA
Cel5D	5	150
Cel5D_nopor	5	50

4.3.1.2 pH optimum

The pH optima for the various protein variants were determined by conducting enzyme assays using buffers with different pH from 3-10 in 0.5 intervals. The pH-activity graphs are shown in Figure 4.7.

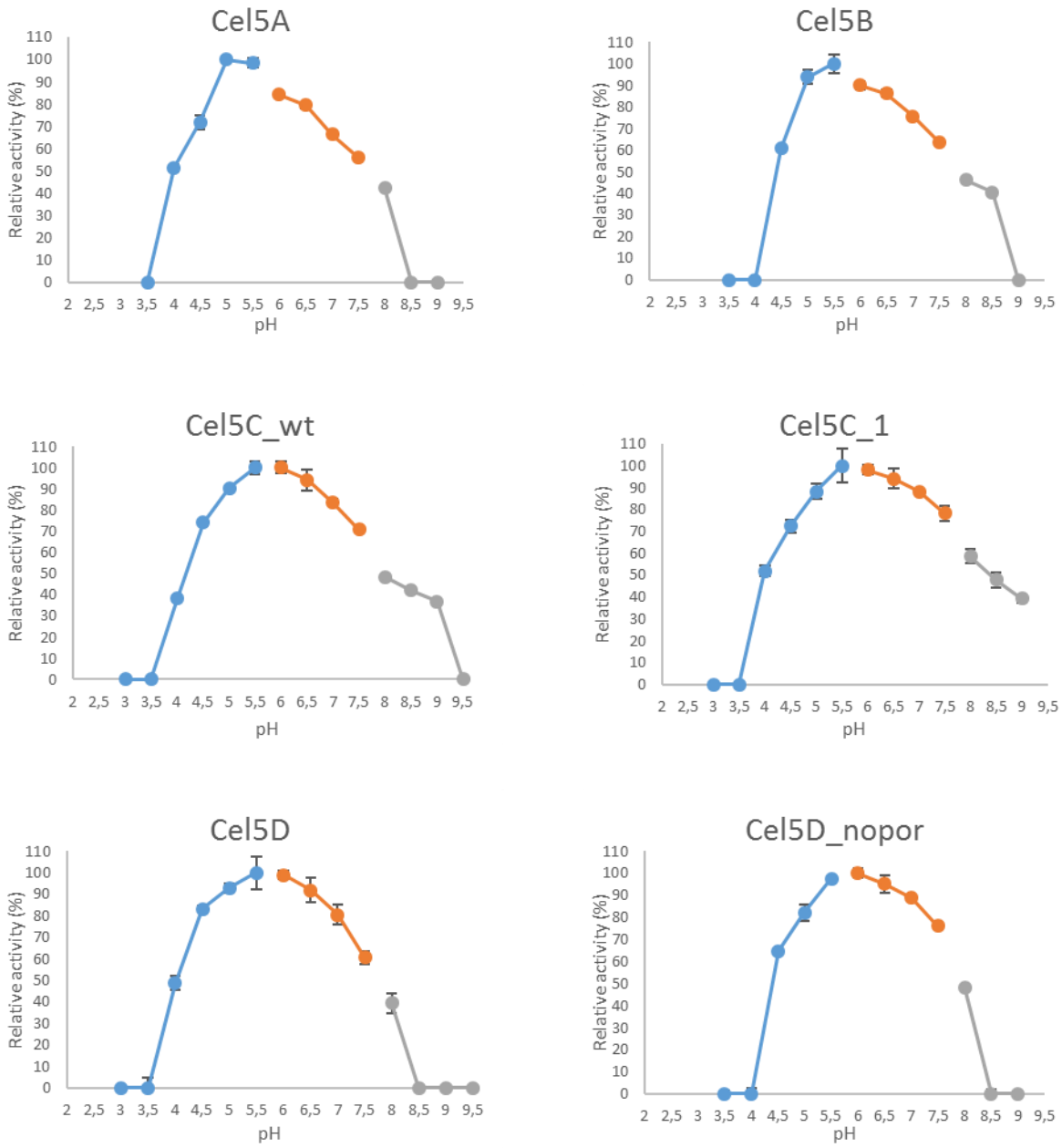


Figure 4.7. pH optimum for the gene cluster enzymes. Assays were performed at 40 ºC, 800 rpm horizontal shaking, 1 % (w/v) CMC, 20 mM buffer and varying enzyme concentration and times according to Table 4.3. Buffers used were citrate buffer (blue), phosphate buffer (orange) and tris buffer (grey). Activity is shown as the relative amount (%) of glucose equivalents (mg/ml). Error bars represent standard deviations between three replicates. Cel5A_cat, Cel5C_wtR, Cel5C_2 and Cel5C_2R were not available in sufficient amounts or not active enough for pH assays.

The data show that all enzymes have their pH optimum within the range of pH 5.0-6.0, which corresponds to the pH in the cow rumen environment (section 1.6). Most of the GH5 enzymes also showed similar pH-activity profile, although some enzymes were more tolerant for acidic conditions and others more tolerant for alkaline conditions.

4.3.1.3 Temperature optimum

To determine the temperature optima for the enzymes, enzyme assays on CMC were performed at different temperatures, ranging from 10-80 ºC in 5 ºC intervals. Relative enzyme activity at different temperatures are shown in Figure 4.8.

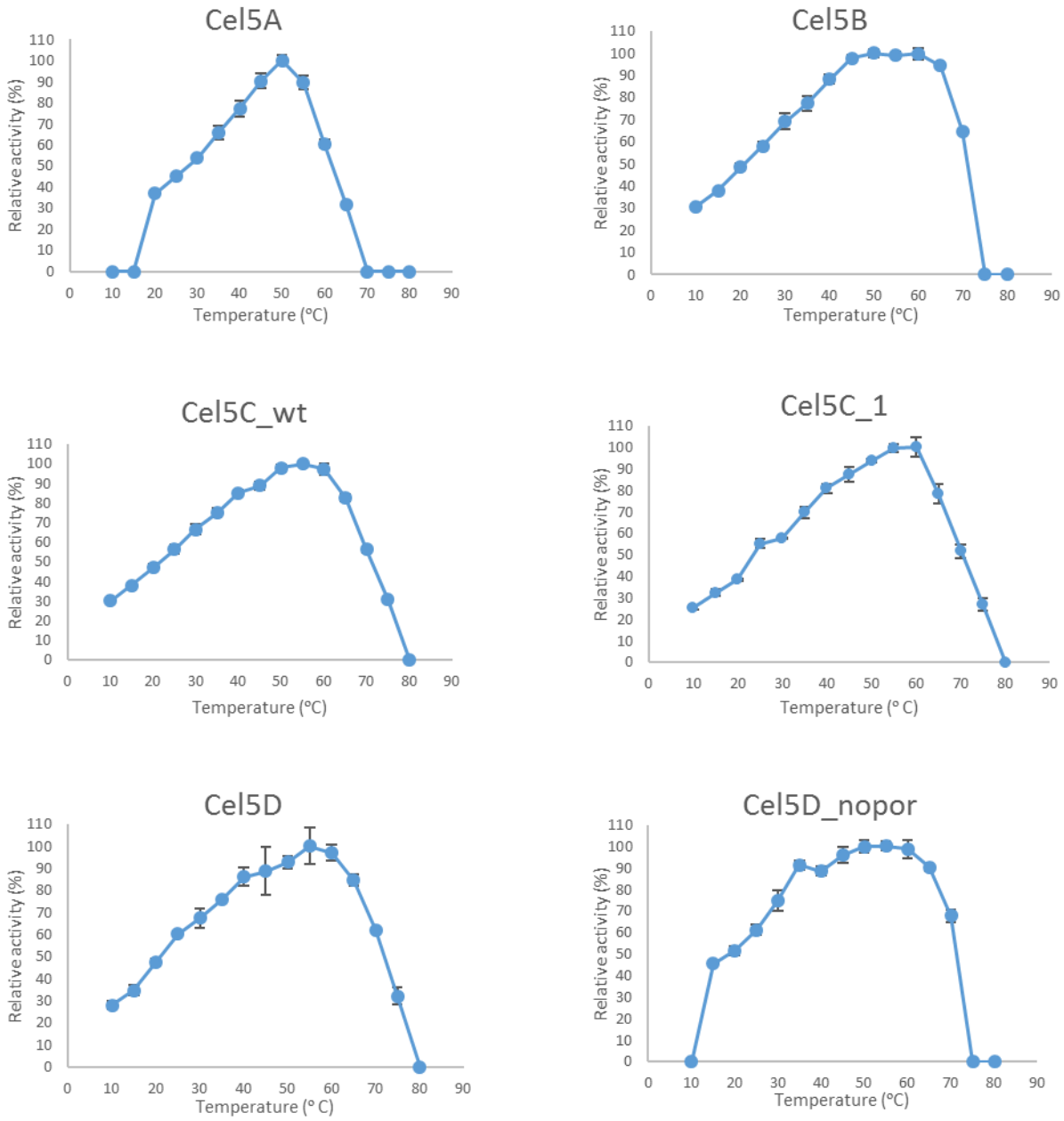


Figure 4.8. Temperature optimum of the AGa enzymes. Assays were performed with 800 rpm horizontal shaking, 1 % (w/v) CMC, 20 mM citrate buffer pH 5.5 and varying enzyme concentration and times according to Table 4.3. Reaction mixtures were pre-heated at given temperatures before adding the enzyme. Activities are shown as relative amount (%) of glucose equivalents (mg/ml). Error bars represent standard deviations between three replicates. Cel5A_cat, Cel5C_wtR, Cel5C_2 and Cel5C_2R were not available in sufficient amounts or not active enough for temperature assays.

All enzymes had an optimum temperature between 50-60 ºC, above the mean temperature in the cow rumen, which is 37-42 ºC. Several of the enzymes showed a wide activity range, with 50 % activity between 25-70 ºC, whereas others showed a narrower profile. To verify if these results could apply over a longer period, temperature stability assays were conducted.

4.3.1.4 Temperature stability

Temperature stability assays were performed by pre-incubating the enzymes at given temperatures for varying time lengths, before performing the standard CMC activity assay (Figure 4.9).

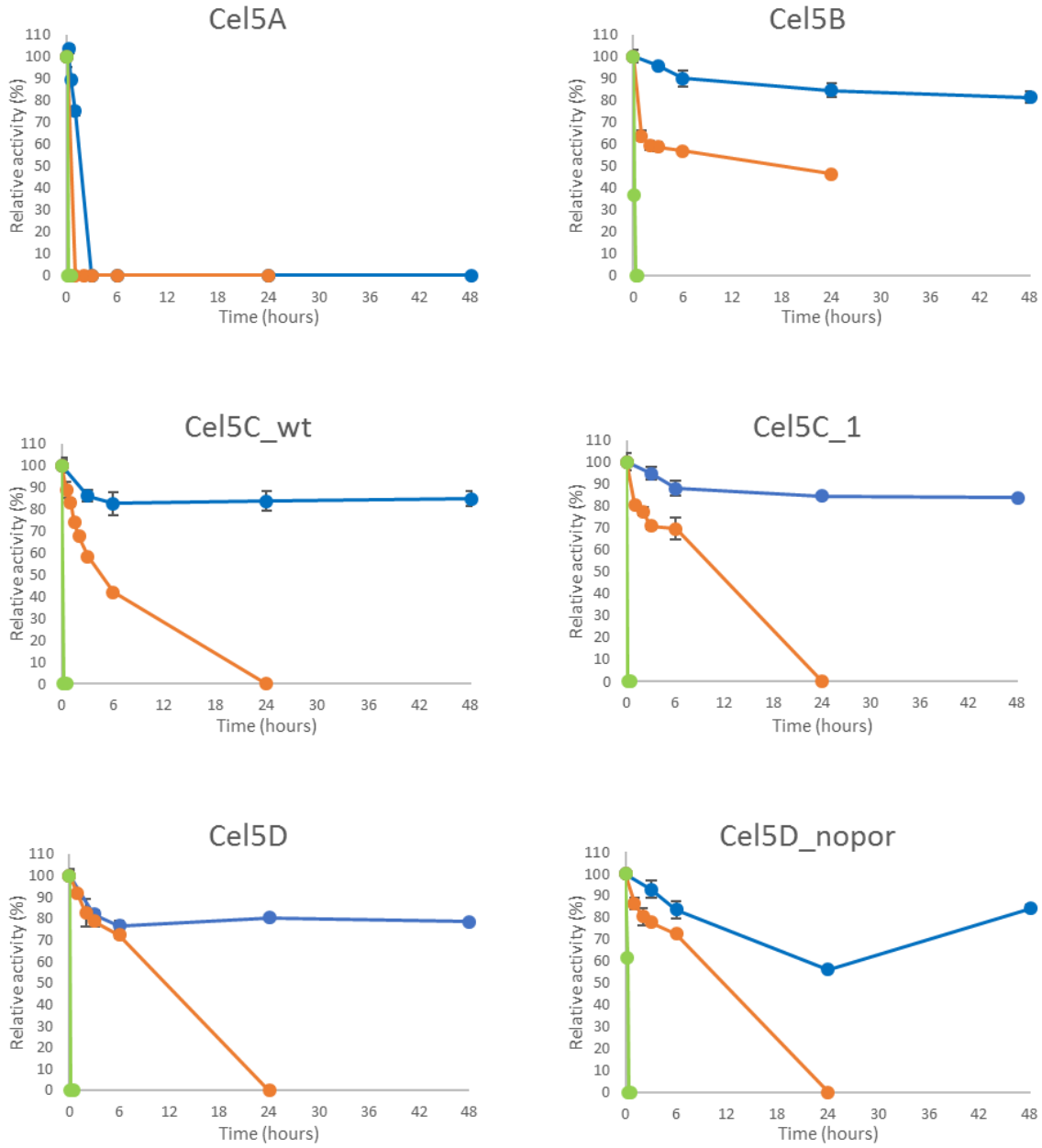


Figure 4.9. Temperature stability of the AGa enzymes. Enzymes were pre-incubated at 40 ºC (blue), 50 ºC (orange) and 60 ºC (green) for varying time points, and assays were performed to check the residual activity after incubation . Assays were performed at 40 ºC and 800 rpm horizontal shaking with 1 % (w/v) CMC, 20 mM citrate buffer pH 5.5 and varying enzyme concentration and times according to Table 4.3. Activities are reported as the relative amount (%) of glucose equivalents (mg/ml) to non-incubated enzyme. Error bars represent standard deviations between three replicates. Cel5A_cat, Cel5C_wtR, Cel5C_2 and Cel5C_2R were not available in sufficient amounts or not active enough for temperature stability assays.

Temperature stability assays revealed that none of the AGa enzymes were stable at 60 ºC and all, except Cel5B, lost their activity after 24 hours at 50 ºC. Most of the enzymes were stable at 40 ºC, except Cel5A, which lost its activity after 3 hours of incubation, and Cel5D_nopor, which had 60 % activity after 24 hours at 40 ºC, but recovers after 48 hours.

4.3.1.5 Comparative analysis of activity on CMC

To obtain an indication of the relative efficiencies of the enzymes, their activity on carboxymethyl cellulose was determined under identical conditions for all enzymes (Figure 4.10).

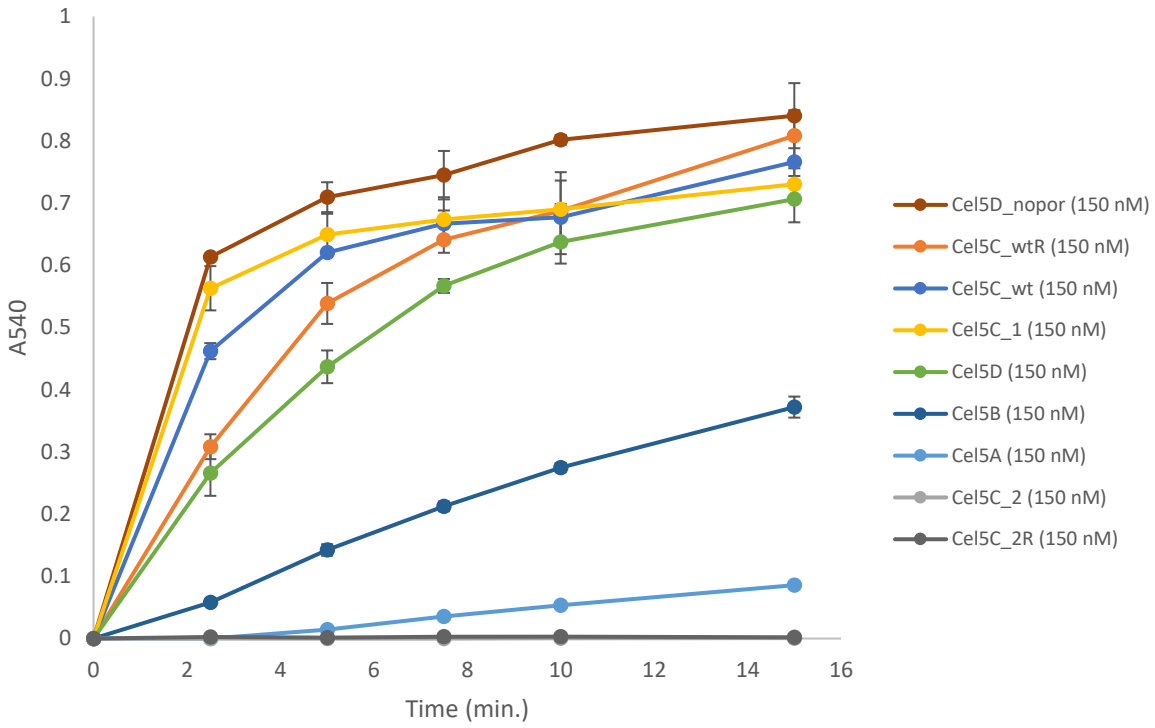


Figure 4.10. Activity of AGa GH5s on CMC. All assays were performed at 40 ºC, 1 % (w/v) CMC, 20 mM citrate buffer pH 5.5, 800 rpm horizontal shaking and 150 nM enzyme. Activity, measured by the DNS assay, is reported as A_{540} for the sake of comparison, as the absorbance values obtained for Cel5A, Cel5C_2 and Cel5C_2R were too low to be quantified using the glucose standard curve. Error bars represent standard deviations between three replicates.

There is a clear separation of enzymes with respect to activity on CMC. Cel5D, Cel5D_nopor, Cel5C_1, Cel5C_wt and Cel5C_wtR have a high activity with minor deviations between them. Cel5B and Cel5A have lower activity on CMC, and Cel5C_2 and Cel5C_2R have almost no activity at the enzyme concentration and incubation time used. The graph shows that Cel5C_wt

has higher activity than the restored version (Cel5C_wtR), but they have nearly the same product formation after 15 minutes. Cel5C_1 alone has a higher activity rate than Cel5C_wt, and is clearly the most active domain of Cel5C_wt, as Cel5C_2 and Cel5C_2R both lack activity under the given conditions. Cel5D_nopor is the enzyme with the highest initial rate and highest product formation, and performs clearly better than native Cel5D, which contains the Por-tag.

A preliminary assay of Cel5A_cat activity on CMC, performed before the MBP fusion tag was cleaved off, showed approximate identical activity to Cel5A (Appendix D, Figure D1). Notice that this is only a preliminary assay, due to low yield.

Since Cel5C_2 and its restored version showed little to no activity on CMC with 150 nM enzyme, these enzymes were also tested for activity on the more easily accessible substrate Barley β-glucan, with higher enzyme loadings and longer incubation times (Figure 4.11). This was done to investigate the effect of the restored catalytic-site mutation.

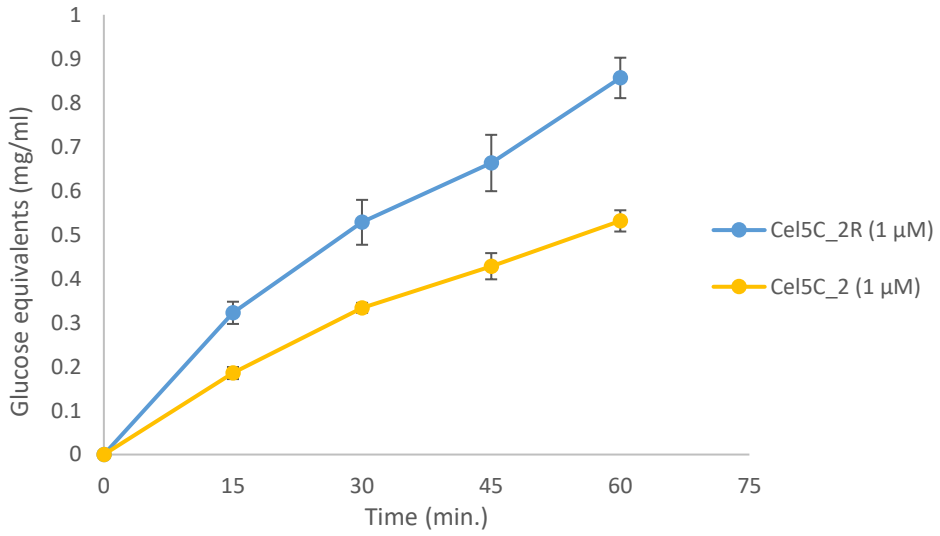


Figure 4.11. β-glucan activity of Cel5C_2 and Cel5C_2R. Assays were performed at 40 ºC and 800 rpm horizontal shaking with 1 % (w/v) β-glucan, 20 mM citrate buffer pH 5.5 and 1 µM enzyme. Error bars represent the standard deviations between three replicates.

The β-glucan activity of Cel5C_2 and Cel5C_2R shows that these are both active enzymes and that the restored version of Cel5C_2 is more active than the wild-type.

4.3.1.6 Activity on cellodextrins

To investigate the potentially complimentary roles of the cluster GH5 enzymes, product formation by the enzymes from cellopentaose (Glc_5) and cellohexaose (Glc_6) was analysed by HPAEC-PAD to investigate cleaving patterns. The results are summarized in Table 4.4, and chromatograms displaying the product distributions over time are shown in Appendix E, Figure E1-Figure E9.

Table 4.4. Activity on cellodextrins. Listed are the products detected after 60 minutes of incubation with the various enzymes, and time before total degradation of substrate. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 250 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

	Cellopentaose		**Cellohexaose**	
Enzymes	**Products after 60 minutes of degradation**	**Time before total degradation of substrate (min.)**	**Products after 60 minutes of degradation**	**Time before total degradation of substrate (min.)**
Cel5A	Glc_2, Glc_3	30	Glc_2, Glc_3, Glc_4	15
Cel5B	Glc_2, Glc_3	60	Glc_3	5
Cel5C_wt	Glc_2, Glc_3	<0.5	Glc_2, Glc_3	<0.5
Cel5C_wtR	Glc_2, Glc_3	1	Glc_2, Glc_3	<0.5
Cel5C_1	Glc_2, Glc_3	<0.5	Glc_2, Glc_3	<0.5
Cel5C_2	Glc_2, Glc_3, Glc_4	>60	Glc_2, Glc_3, Glc_4	60
Cel5C_2R	Glc_2, Glc_3	20	Glc_2, Glc_3	30
Cel5D	Glc_2, Glc_3	<0.5	Glc_2, Glc_3	<0.5
Cel5D_nopor	Glc_2, Glc_3	<0.5	Glc_2, Glc_3	<0.5

The results from Table 4.4 shows that the majority of the enzymes cleave cellopentaose into cellotriose and cellobiose, and cellohexaose into cellobiose and cellotriose. All enzymes except Cel5B also produce cellotetraose from cellohexaose, which is further broken down to cellobiose (Figures in Appendix E). No further degradation of cellotriose is seen. Cel5C_wt, Cel5C_wtR, Cel5C_1, Cel5D and Cel5D_nopor are the fastest enzymes with respect to cellodextrin degradation, degrading the substrates completely in less than 30 seconds (1 minute for Cel5C_wtR on cellopentaose). Cel5C_2 is the slowest enzyme, as it could not fully degrade cellopentaose in the assay time. Cel5B is the enzyme that diverges the most with respect to cellodextrin degradation by degrading cellohexaose to cellotriose and no cellobiose. This is also

the enzyme with the largest deviation in degradation time between cellopentaose and cellohexaose.

4.3.1.7 Activity on hemicelluloses

To check for hemicellulase activity and assess types of bonds cleaved by the cluster enzymes, all enzymes were assayed on various hemicellulosic substrates with different monomers, linkages and monomer-and/or linkage ratios. Product formation was assessed using the DNS method (Figure 4.12).

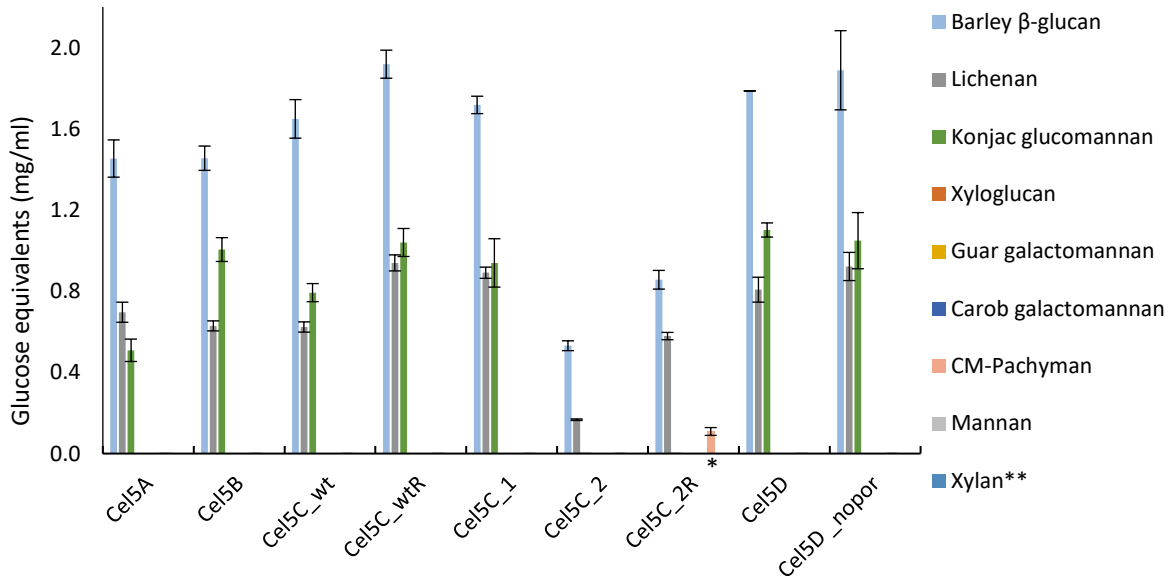


Figure 4.12. Enzyme activity on hemicellulosic substrates. Assays were performed at 40 ºC with 0.5 % (w/v) substrate, 20 mM citrate buffer pH 5.5 and 1 µM enzyme for one hour. These assays were performed without shaking. Error bars represent standard deviations of three replicates. *Value 0,002 A_{540} below the lowest standard curve value **Due to substrate shortage, Cel5C_wtR and Cel5D_nopor were tested on arabinoxylan, which is a substituted xylan.

All enzymes were active on Barley β-glucan (β-1,3-1,4-linked glucose units), konjac glucomannan (β-1,4-linked mannose and glucose residues) and lichenan (β-1,3-1,4-linked glucose units) in various degrees, except Cel5C_2 and Cel5_C2R which had no activity on konjac glucomannan. Only Cel5C_2R showed trace activity on CM-pachyman (β-1,3-linked glucose units substituted with carboxymethyl groups). No activity was seen for any of the enzymes on the remaining hemicelluloses

4.3.1.8 Activity on filter paper

To assess the true cellulase activity of the AGa enzymes, i.e. activity on crystalline substrate, filter paper was used as substrate and product formation was analysed by HPAEC-PAD (Figure 4.13).

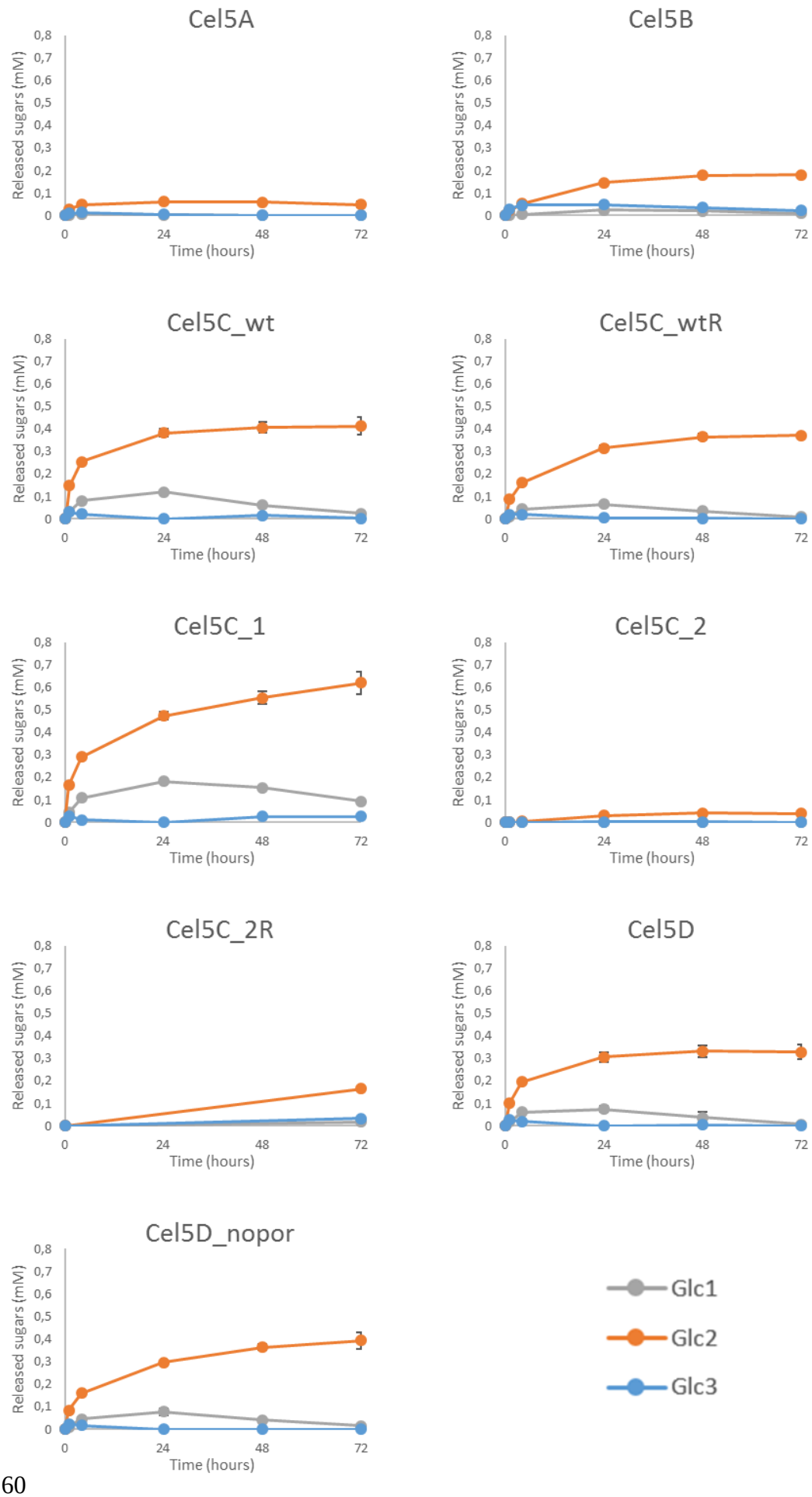

Cel5A
Cel5B
Cel5C_wt
Cel5C_wtR
Cel5C_1
Cel5C_2
Cel5C_2R
Cel5D
Cel5D_nopor
Released sugars (mM)
Time (hours)
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
24
48
72
Glc1
Glc2
Glc3

Figure 4.13. Filter paper degradation by AGa enzymes analysed with HPAEC-PAD. Reactions contained 1 % (w/v) filter paper (particle size 0.5 mm), 20 mM citrate buffer pH 5.5, 1 µM enzyme and were performed at 40 ºC and 800 rpm horizontal shaking. Reactions were stopped with 0.2 M NaOH, then centrifuged and products were analysed by HPAEC-PAC with Chromeleon software. Cel5C_2R was run without replicated and only at one time point (72 hours) due to protein shortage (see section 4.2.3).

Figure 4.13 shows that most of the enzymes seems to be active on filter paper, and these can most likely be classified as cellulases. The enzymes have cellobiose (glc_2) as the main product after filter paper degradation. Small amounts of cellotriose (glc_3) are produced and some glucose (glc_1) is produced by the most active enzymes. Cel5A and Cel5C_2 seem to have little to no activity on filter paper. The glucose concentration in the reactions decreases over time, potentially due to microbial contamination of the assays. The activity difference between the enzymes is similar to what was seen in the activity assay on CMC, except from Cel5C_1 which is by far the most active enzyme on filter paper, while it seems to be less active than at least Cel5D_nopor on CMC.

To investigate if the combination of the full-length enzymes leads to synergy, equimolar amounts of Cel5A, Cel5B, Cel5C_wt and Cel5D were used to degrade filter paper (Figure 4.14). In this same experiment, the effect of β-glucosidase addition to the most active enzyme on filter paper, Cel5C_1, was also investigated, to assess if the enzyme was inhibited by its product, cellobiose.

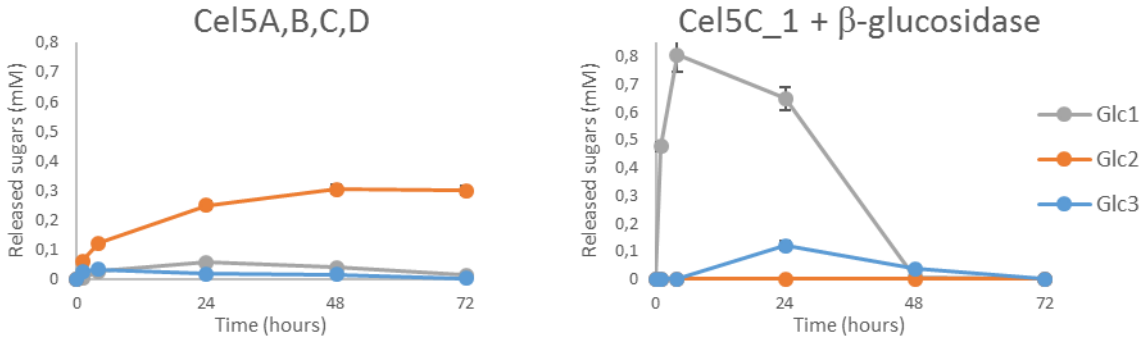


Figure 4.14. Filter paper degradation by AGa enzymes combined and Cel5C_1 with β-glucosidase analysed with HPAEC-PAD. Reactions contained 1 % (w/v) filter paper (particle size 0.5 mm), 20 mM citrate buffer pH 5.5, 1 µM enzyme (total for Cel5A,B,C,D and 1 µM Cel5C_1 + 0.5 U/ml β-glucosidase for Cel5C_1+ β-glucosidase assay) and were performed at 40 ºC and 800 rpm horizontal shaking. Reactions were stopped by adding NaOH to 0.1M, then centrifuged and products were analysed by HPAEC-PAC with Chromeleon software.

The cocktail of full-length AGa enzymes did not increase the breakdown of filter paper. In fact, the enzyme mix had a lower activity than the best enzymes alone. Compared to the best enzyme (Cel5C_1) alone which released approximately 0.6 mM cellobiose (the main product), the cocktail released approximately 0.3 mM cellobiose. Addition of β-glucosidase to Cel5C_1 led to high amounts of glucose only hours after assay commencement and there was no detectable cellobiose. Cellotriose amounts were higher in the reaction with β-glucosidase added. The glucose concentration dropped from 0.8 mM to zero after 48 hours, which might be due to microbial contamination. Due to this, comparison of only the first hours is most relevant, and the β-glucosidase added reaction produced 0.8 mM glucose within the first hours, while CelC_1 alone produced approximately 0.1 mM glucose, 0.3 mM cellobiose and no cellotriose, which would correspond to 0.7 mM glucose when fully degraded to monomers.

4.3.1.9 Activity on switchgrass

To assess the enzyme activity on a natural lignocellulosic substrate, switchgrass was used. Switchgrass was originally used to enrich the biomass degrading consortium in the metagenomic study where the AGa phylotype was reconstructed. Product analysis by the DNS method showed presence of reducing sugars, implying possible cellulose breakdown. However, estimated sugar concentrations were too low for accurate quantification. Differences in amounts of reducing sugars released by the various enzymes were seen, with Cel5C_wt producing the highest absorbance (Appendix F, Figure F1). Analysis of the products by HPLC-MS revealed that both cellodextrin and xylodextrin products were produced (Table 4.5), indicating breakdown of both cellulose and xylan in the switchgrass by the gene cluster enzymes.

Table 4.5. Product analysis of switchgrass assay. The table shows presence of cellodexrins or xylodexrins in the reactions for the different enzymes. Results are analysed by HPLC-MS and Xcalibur software. Relative amounts of products were not quantified. However, the trace amounts of dextrins in the negative control had much weaker signals than enzyme samples.

Enzyme	Cellodextrins	Xylodextrins
Cel5A	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5B	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5C_wt	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5C_wtR	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5C_1	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5C_2	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4
Cel5D	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Cel5D_nopor	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Full-length A+B+C+D	glc_2, glc_3, glc_4	xyl_2, xyl_3, xyl_4, xyl_5
Negative control	Trace glc_2, glc_3	Trace xyl_3, xyl_4

4.3.1.10 Binding of Cel5A_d2 to cellulose

The second domain of Cel5A, Cel5A_d2, is predicted to be a CBM. To analyse this further, binding of Cel5A_d2 to Avicel was analysed. The assay was performed with control reactions lacking substrate (enzyme blank) and a reaction with a protein known to not bind to cellulose (lysozyme). The result is presented in Figure 4.15.

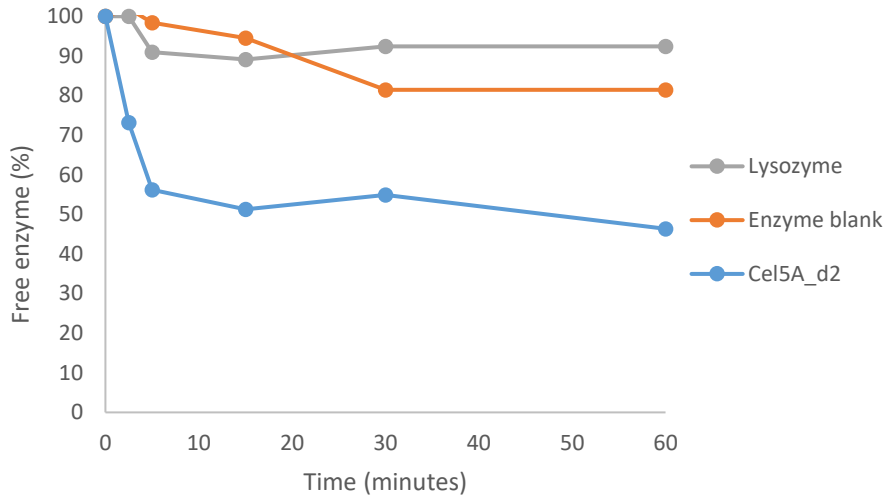


Figure 4.15. Binding of Cel5A_d2 to Avicel. The binding assay was performed at 40 ºC and 800 rpm horizontal shaking with 10 mg/ml Avicel, 20 mM citrate buffer pH 5.5 and 0.04 mg/ml Cel5A_d2. At given time points reaction samples were taken and filtrated, before the flow-through was measured with A_{280}. Protein concentration (mg/ml) in the flow-through was calculated using the proteins extinction coefficient. Values are shown as % free enzyme present in the flow-through after filtration of the reaction incubated with Avicel.

The results of the binding assay with Avicel show that approximately 50 % of Cel5A_d2 binds to the substrate in the assay-time, indicating correct prediction of the domain as a CBM. However further investigation is required to confirm the CBM functionality. Due to the low yield of Cel5A_d2 protein expression, further binding characterization was not performed.

Cel5A_d2 was also tested for Barley β-glucan activity, using 1 µM enzyme for 1 hour, yielding no reducing sugars detected by DNS (data not shown), indicating a non-catalytic function.

4.3.2 Crystallization

Cel5C_2 was screened for crystal formation with the Wizard™ 2 screen, in an effort to get more information about its structure and the topology of the active site with the mutated catalytic amino acids.

Crystal screening plates were checked regularly, but no crystal formation was seen in any of the 48 conditions as per now, 6 months after setting up the screen.

4.4 Metaproteomics

In an effort to identify the AGa phylotype enzymes in a cow rumen sample, and to potentially identify enzymes from the here studied cellulase cluster, a metaproteomic investigation of a different cow rumen sample (fed on a mixed diet with 60 % fibre) was conducted. This sample was another sample than the original, switchgrass-enriched metagenomic sample from Hess et al. (2011).

Peptide fragments from the nanoLC-MS/MS analysis were mapped to a reference library, made from the annotated ORFs of the AGa genome (performed by Adrian Naas). No unique peptides could be mapped to the AGa genome and no matches were found to any of the enzymes or proteins in the AGa gene cluster investigated in this thesis.

5 Discussion

The aim of this study was to characterize the four GH5 enzymes encoded within the AGa cluster and investigate if the cluster could be a contributor to cellulose degradation by the AGa phylotype, to gain more insight into the microbial cellulose degradation mechanisms. The enzymes were subjected to bioinformatical analysis prior to cloning, expression and purification. Finally, characterization was performed for better understanding of the enzymes catalytic mechanisms and potential co-operation.

5.1 Protein domains, parameters and similarity

The AGa cluster GH5 enzymes were subjected to bioinformatic analyses to assess protein domains, parameters and similarities. An overview of the gene cluster organization and protein domains of the cluster GH5s is presented in Figures 4.1 and 4.2. It is clear from these figures that the four GH5 proteins shows differences in domain organization. Four of the same type of enzyme (GH5) in one gene cluster may seem redundant, but the variation in the number of catalytic domains, the presence of a CBM and Por-tag, and protein and linker sizes may indicate different roles or modes of action. The MSA (Figure 4.3) of the catalytic domains of the GH5s showed that the Cel5C_2 domain lacks the first catalytic glutamic acid as well as the two conserved, adjacent amino acids. Whether this represents a sequencing/assembly error or an evolutionary adaption is not known, and was further assessed by production of both wildtype and restored versions. An evolutionary adaption that removes the catalytic amino acid from a glycoside hydrolase seems unorthodox, but has been reported for GHs of other families (Hughes 2012; Rogowski et al. 2015). In the studies, the GHs were active through proposed activation of a water molecule by other residues, or the GHs took on new roles other than catalytic activities.

5.2 Cloning, expression and purification

All AGa cluster GH5 enzymes with truncated versions were cloned, expressed and purified. Cel5B, Cel5C non-restored versions and Cel5D were produced without major problems with the standard procedure. However, the Cel5C restored versions, with only three mutated amino acids, aggregated in inclusion bodies upon expression. We hypothesize that the mutation has changed the folding of the protein or perhaps affected hydrophobic areas, which contributes to

aggregation. The formation of inclusion bodies upon introduction of the restoring mutation does not necessarily mean that the restoration is "wrong", i.e. the true protein is the non-restored, as inclusion bodies are a common problem when expressing protein heterologously. Expressing a Bacteroidetes protein in *E. coli* is in itself an unnatural process and production of soluble and functionally active protein is not guaranteed. Factors like pH, osmolarity, folding mechanisms, cofactors and redox potential can differ from the original cell to the host (Rosano & Ceccarelli 2014). In addition, *E. coli* cannot reproduce native glycosylation of proteins (Demain & Vaishnav 2009), which could affect Cel5A, Cel5C and Cel5D proteins which are thought to be glycosylated through native T9SS secretion. *E. coli* also has problems with toxicity at high cell densities due to acetate formation (Demain & Vaishnav 2009). The above mentioned factors can lead to unstable, misfolded proteins that aggregate to form inclusion bodies (Rosano & Ceccarelli 2014). In addition, high expression levels can lead to aggregation. To prevent aggregation, lowering the production temperature or coupling the protein to a fusion tag could increase solubility (Gonzalez-Montalban et al. 2007). However, soluble protein does not necessarily mean the protein is properly folded and functionally active, as traditionally has been believed (Gonzalez-Montalban et al. 2007). Solubility and protein quality does not always correlate, and fractions of soluble protein might be misfolded and inactive, and conversely, proteins in inclusion bodies can be biologically active (Gonzalez-Montalban et al. 2007). Hence, all the soluble protein obtained by expression of restored Cel5C versions is not necessarily properly folded/functionally active. In addition, Cel5C_wtR was seen to not bind properly to the column during IMAC purification, leading to a low yield. As the protein might not be folded correctly, the His-tag could be buried inside the protein and hence not be exposed to the nickel column. This would lead to the protein not binding during IMAC.

Cel5A was the only enzyme that gave no soluble protein at all with the standard expression protocol. As the protein has two domains, it might be harder to express due to folding issues. However, the Cel5Cwt also has two domains and a larger molecular weight, but was expressed without problems, demonstrating the individuality of which proteins can be properly expressed in *E. coli*. The formation of inclusion bodies in Cel5A might be due to the characteristics of the enzyme in its native form, such as hydrophobic areas or special folding patterns. Expression of Cel5A variants with fusion tags was successful, with moderate amount of soluble protein being produced when using the MBP-tag (for Cel5A_cat and Cel5A_d2) or the control tag (without fusion tag) (Cel5A). Several mechanisms for enhanced solubility by fusion proteins have been proposed throughout the years. This includes: producing micelle-like structures the with target

protein facing inwards to avoid contact with the environment inside the host cell, attracting or acting like chaperones or inhibiting aggregation by being charged and hence repulse each other (Costa et al. 2014). Which mechanism that made it possible for MBP to solubilize Cel5A_cat and Cel5A_d2 is unknown, but MBP has been shown to attract chaperones and also act as one (Costa et al. 2014).

Cel5A full-length protein was, as mentioned, successfully expressed with soluble protein with the control tag. The control tag does not contain a fusion tag, but just the N-terminal His-tag. It can be speculated that the change of expression vector (from pNIC to pSol), or the change of his-tag position from C-terminal in pNIC to N-terminal in pSol, can be the cause.

Neither Cel5A or its truncated versions gave exclusively soluble protein, but instead included inclusion bodies upon expression, suggesting a possibility of poorly folded soluble protein as discussed above.

The TEV cleavage and subsequent purification of MBP-tagged Cel5A_cat and Cel5A_d2 was not too successful as protein yields after cleavage was 0.6 % and 12 % respectively compared to yields before cleavage. During TEV cleavage, only a small amount of Cel5A_cat was cleaved, while nearly all Cel5A_d2 was cleaved. It has been reported that smaller proteins are less amenable to TEV cleavage (Jeon et al. 2005), but Cel5A_d2 is just as small and even smaller than Cel5A_cat, and should have been hard to cleave as well, if this were the case here. The difference in cleavage might be due to other individual factors. The low yield after TEV cleavage and subsequent purification for both proteins might be due to protein degradation or precipitation during TEV cleavage, or weak interactions with the Ni-column during IMAC purification, difficulties encountered in similar TEV cleavage studies (Jeon et al. 2005).

Both Cel5A and Cel5C have C-terminal Por-tags for the T9SS, which have been shown to cleave, modify (glycosylate) and attach proteins to cell surface. It can be possible that these proteins need this transportation path and modification for optimal folding and activity. Beckham et al. (2012) noted that “aggregation upon deglycosylation has been observed for many proteins” and Gupta et al. (2011) have seen that deglycosylation of a cellobiohydrolase from *Trichoderma reesei* reduced the solubility of the enzyme. In the case of Cel5D, which is also predicted to be secreted by the T9SS, the issue of proper folding could be smaller because it is only a single domain protein, leading to better heterologous expression.

5.3 Characterization

Following cloning, expression and purification, the enzymes were subjected to enzymatic characterization to gain information on their substrate targets and mechanisms.

Assays were conducted to determine the reaction conditions where initial rapid rate was obtained, before further characterization was done. Non-linearity in enzymatic product formation is usually due to substrate depletion (Robinson 2015) and needs to be avoided for the reaction to not be affected by substrate concentration. Knowledge of initial rapid rate conditions is required when further assaying pH and temperature optimum, in order to ensure that the difference in activity is due to the factor assayed for, for example pH or temperature. In this study, initial rapid rate was assessed for one pH and temperature, that were thought to be near the optima due to conditions in the enzymes' natural environment. Initial rapid rates would optimally be assessed for each pH and temperature. Linear enzyme dose response of a reaction is also crucial, as same type of enzymes competes for cleaving sites at the substrate. This will lead to an enzyme surplus not giving increased product formation. Therefore, choosing an enzyme concentration within the linear dose-response range is necessary for correct enzyme assay results. In the case of cellulases, optimal initial conditions are almost impossible as the heterogeneity of their substrate causes nonlinear kinetics (Zhang et al. 1999). Based on this, enzyme concentrations and timeframes within approximately constant rates during the time of the assays were found and used in further work in this study. As the cluster GH5 enzymes did not have the same initial rates, different concentrations and times had to be used, but efforts were made to assay the enzymes within the same timeframe. As Cel5A has the lowest initial rate of the enzymes, it had to be assayed for 15 minutes, as opposed to 5 minutes for the other enzymes. This might have affected some results, as enzyme stability will increasingly affect results when using longer timeframes.

For characterization and determination of optimal conditions for subsequent experiments, pH and temperature optima were assessed, as well as temperature stability. Results correlated with rumen conditions, as the pH optima were within 5.0-6.0 and the temperature optima within 50-60 ºC with stability at 40 ºC. The optimal temperature is above temperature in the enzymes' native environment, but the optimal temperature is shown to not be applicable over longer time periods by the stability assays. Although enzymes work faster at higher temperatures (Lesk 2010), the enzymes need to be stable to function over longer time periods. As the assays were conducted in 5 minutes, the enzymes will likely not suffer from destabilization. While this is

useful for when conducting biochemical characterization, such short assaying times may give a wrong impression of the true optimum temperature under more realistic conditions. On the other hand, longer assays would glide over in a temperature stability assay. pH stability assays were not conducted, but the enzymes were shown to remain active over long time periods at the determined optimal pH.

The temperature and pH optima curves for the different enzymes are similar with a broad activity range. The Cel5A temperature optimum curve stands out, with a more narrow and steep curve compared to the other enzymes. This might be due to longer assay time, discussed above, and the enzyme might have been destabilized at the higher temperatures. This can also be seen in the shift of optimum, as the Cel5A is the only enzyme with 50 ºC as optimum temperature compared to 55-60 ºC for the other enzymes.

A larger discrepancy between the enzymes is seen for the temperature stability curves. Surprisingly, Cel5A was not stable at 40 ºC unlike the other enzymes, which could explain its steeper decline in activity at higher temperatures shown in the previous experiment. One possible explanation can be the earlier addressed misfolding of soluble protein, as Cel5A was proven difficult to express, and could harbour misfolded fractions. If time had permitted, the experiment would be repeated to verify the results, and further attempts would be made to improve the expression of Cel5A.

Cel5D_nopor also stood out when assaying temperature stability, with an activity drop to approximately 60 % after 24 hours of incubation at 40 ºC and regaining the activity after 48 hours, for unknown reasons. The assay was repeated to verify results and both replications were run with triplicates giving low standard deviations. The reaction conditions were the same as for the other GH5 enzymes, which do not show this remarkable development of activity over time. The 24-hour sample was incubated together with the 48-hour sample, so an incubation error due to instability in thermomixer temperature is unlikely. It is thus hard to explain the result by faults in the reaction set-up. Searching the literature, no previous examples of this behaviour of thermal stability were found.

Cel5B seems to be the most stable enzyme. The stability differences between Cel5B and the other enzymes might be due to the previously mentioned potential need for glycosylation of the enzymes through the T9SS. Cel5B is the only protein not destined for the T9SS secretion path, and the lack of glycosylation for the T9SS destined enzymes, due to *in vitro* expression, might

contribute to reduced stability. This is exemplified by Borzova et al. (2014) where a deglycosylated glycosidase showed reduced thermal stability relative to glycosylated variants.

All the enzymes, except Cel5A, has a slight drop in activity from non-incubated samples to a few hours of incubation at 40 ºC. This might be due to denaturation of a fraction of the enzyme loading, while the rest of the enzymes in the reaction is not inactivated. Another possible explanation is adsorption of some enzyme molecules to the tube during pre-incubation, hence reducing the enzyme concentration and then reducing the concentration in the enzymatic assays. Incubation of larger reaction volumes in low-bind tubes, or running a control incubation with BSA, could be attempted to obtain straight curves.

For further analysis of enzyme activity on CMC, cellodextrins, filter paper, hemicelluloses and switchgrass, pH 5.5 and 40 ºC were chosen as optimal conditions for all enzymes, as these are the average optima and as standardization facilitates experimental analysis as all enzymes are assayed under the same conditions.

Assays of activity on CMC revealed differences between the cluster enzymes with respect to their catalytic rates. Cel5C and Cel5D, along with their truncated versions, demonstrated a higher activity than Cel5A and Cel5B, except the second domain of Cel5C, which inactive under assay conditions. The seemingly mutated Cel5C_2 and the restored Cel5C_2R needed an increase in enzyme concentration and assay time, and change to a more accessible substrate for any activity to be detected. Cel5C_2R has a slightly higher activity than the wildtype, pointing towards a sequencing or assembly error in the wildtype protein sequence. However, these results are not very conclusive, since one would think that removal of one catalytic amino acid would make the protein inactive, and that subsequent restoration would lead to a greater increase in activity. The residual activity of Cel5C_2, despite lack of the catalytic amino acid, could be explained by the amino acids replacing the consensus sequence. The mutation from Asn-Glu-Pro→Asp-Asn-Ala leads to an aspartic acid adjacent to the catalytic amino acid site, and this amino acid might be able to take the role as the acid/base or nucleophile in the place of glutamic acid. The relatively small increase in activity from Cel5C_2 to Cel5C_2R might be due to troubles with expression of the restored version which, due to this, might not be properly folded.

Cel5C_1 has much higher activity on CMC than Cel5C_2, outperforming the full-length Cel5C_wt, showing that Cel5C_1 is the major catalytic domain of the enzyme. Restoration of

the second domain of the full-length Cel5C did not show an increase in activity on CMC, which could be due to competition for available substrate or the fact that Cel5C_1 is much more active and the added activity of the restored second domain is masked.

Cel5D was produced with and without the Por-tag for T9SS, revealing a clear increase in activity on CMC as the tag was removed. The tag might be of steric hindrance for catalysis or affect the folding.

The relatively low activity on CMC of Cel5A and Cel5B was a bit surprising, but could imply other tasks for these enzymes or a different affinity for substrate, for example shorter oligosaccharides as seen in the cellodextrin assays discussed further down. Also, as discussed above, difficulties associated with Cel5A expression might have caused misfolded fractions, and hence lower enzyme activity. Collectively, the four cluster enzymes seem to have different activity towards the model substrate CMC, suggesting differing roles in biomass degradation.

In light of the phylogenetic analysis conducted, the activities on CMC correlates well with predicted evolutionary relationships as Cel5C_1 and Cel5D are the most active and most similar, while the phylogenetically diverging Cel5C_2, Cel5A and Cel5B are less active (Figure 4.4).

Enzymatic assays on cellopentaose and cellohexaose revealed enzyme cleaving patterns and oligosaccharide affinities, providing new insights into the differing roles of the GH5 enzymes from the AGa cluster. Cel5B was the only enzyme that rapidly cleaved cellohexaose solely into cellotriose, which could imply an affinity for released cellodextrins over polysaccharide substrates, as its activity on CMC is relatively low. The other enzymes produced cellobiose, cellotriose and cellotetraose from the hexamer, and in most cases, the cellotetraose was further broken down to cellobiose. Cellopentaose degradation resulted in cellobiose and cellotriose. No further degradation of cellobiose or cellotriose was seen, and these products are probably further degraded by the GH3 and GH94 also present in the gene cluster. Cel5A showed a slightly different cleaving pattern than most of the enzymes, since it produced mostly cellotriose from cellohexaose, whereas the other enzymes primarily produced cellobiose. Cel5A also showed slow degradation of cellodextrins, as it did on CMC, again pointing towards possible misfolded fractions during expression as discussed above. The high activity of Cel5B on cellohexaose and affinity for shorter cellodextrins could imply a possible role as a second line of degradation

with cellodextrins as main substrates. Other cleaving rates in the cellodextrin assays were as expected from the activity on CMC.

Testing the enzymes on hemicellulosic substrates revealed activity on substrates with β-1,4-glycosidic linkages. All enzymes were active on β-glucan (β-1,3-β-1,4-glucan), konjac glucomannan (β-1,4-glc-β-1,4-man) and lichenan (β-1,3-β-1,4-glucan), except Cel5C_2 and Cel5C_2R which were not active on konjac glucomannan. The enzymes were not active on pure mannan or CM-pachyman (β-1,3-glucan), which rules out their ability to break the β-1,4-mannose linkages in konjac glucomannan and the β-1,3-glucose linkages in β-glucan and lichenan respectively. Trace activity on CM-pachyman was seen for Cel5C_2R, but just below the range of the standard curve and the experiment should be repeated for quantification. However, the reducing sugars detected from CM-pachyman by the active-site restored Cel5C_2R could suggest that this domain plays a role in hemicellulose degradation in the natural environment, possibly making the cellulose more accessible for the first domain. Enzyme activity against xylan could not be detected in this assay, but activity was implied in the HPLC-MS analysis of switchgrass assay discussed below. This could be due to the low sensitivity of the DNS-method used in this assay, as compared to the more sensitive HPLC-MS additionally used in the switchgrass experiment. One should keep in mind that two of the enzymes were tested on arabinoxylan instead of xylan, due to substrate shortage, and xylanase activity of these enzymes cannot be excluded in this assay.

Filter paper assays seem to confirm that most of the AGa GH5 enzymes were cellulases, as assumed, releasing mostly cellobiose, and some cellotriose and glucose. Although, positive controls (enzymes with known cellulase activity) should have been performed for comparison. Note that a negative control is run and values are corrected according to this. Cel5A and Cel5C_2 had low activities, and if they are active on cellulose needs to be further validated. Activity differences between the enzymes on filter paper were similar to those on CMC, with Cel5C_wt, Cel5C_wtR, Cel5C_1, Cel5D and Cel5D_nopor being the most active enzymes. The exception is Cel5C_1, which was the highest yielding enzyme on filter paper, surpassing the best enzyme on CMC, Cel5D_nopor. This might be due to Cel5C_1 being a smaller enzyme than Cel5D_nopor, and hence is more suited for accessing the bonds in the filter paper, while Cel5D_nopor is more dependent on the easier accessible bonds in CMC. Surprisingly, Cel5C_2R had a higher final product concentration than Cel5A and the same as Cel5B, although the low amount of products released by these enzymes need to be noted. Cel5C_2R

was clearly less active than these enzymes on CMC. However, due to the low yield of Cel5C_2R expression, this experiment was performed using only one reaction parallel and one time point, necessitating further validation.

A cocktail of all full-length GH5 containing enzymes was assayed to identify potential synergistic effects between them, but was seen to release less products from filter paper than Cel5C_1 alone (approximately 0.3 mM against 0.6 mM released cellobiose), pointing towards competition for substrate. The lack of synergistic effects might be due to the enzymes *in vivo* being arranged on the outer membrane of the bacteria in a specific order or conformation, leading to a synergistic cooperation not seen *in vitro*.

Addition of beta-glucosidase to the filter paper reaction with Cel5C_1 led to a high initial glucose production as the Cel5C_1 released cellodextrins were further degraded. The amount of produced glucose was compared to the potential glucose amount if the cellobiose from the reaction with Cel5C_1 alone had been converted to glucose and added to the glucose produced in the reaction. This comparison showed a slightly higher glucose amount with beta-glucosidase added (approximately 0.8 mM against 0.7 mM glucose released), but this does not give a clear indication of product inhibition, i.e. that the cellobiose produced by Cel5C_1 alone inhibits the enzyme from producing more. Possible microbial utilization of the produced glucose in the Cel5C_1 + beta-glucosidase assay lead to non-correct values for glucose production and thus removes the possibility to investigate the product inhibition properly. Repeating the experiment and taking more care to ensure sterile conditions might have given a clearer indication of product inhibition. However, due to time constraints this was not completed. In general, product inhibition is a known phenomenon for cellulases and Park et al. (2011) have reported product inhibition of a GH5 by cellobiose. Cellulases and cellobiohydrolases have previously been shown to be inhibited by cellobiose (Holtzapple et al. 1990; Igarashi et al. 1998), at concentrations as low as 250 µM in the study by Igarashi et al. (1998). Hence, product inhibition by cellobiose for the AGa enzymes investigated in this study is not unlikely, as cellobiose concentrations in the Cel5C_1 reaction reached 0.6 mM at the end of the experiment.

As the metagenomic study, which is the basis for the gene cluster in this thesis, was done on the microbial community attached to switchgrass, it was natural to further test the AGa cluster enzymes on this substrate. Importantly, activity on cellulose and model substrates does not necessarily mean activity on lignocellulose, as the cellulose is embedded in lignin and hemicelluloses in these substrates. Assays with switchgrass revealed trace activity detected by

DNS, but DNS is not a very sensitive method and does not distinguish between different types of reducing end sugars (such as glucose or xylose). Results were therefore additionally analysed by HPLS-MS, which surprisingly showed the presence of xylodextrins in addition to cellodextrins, suggesting both xylan and cellulose breakdown, the main polysaccharide constituents of switchgrass (Hu et al. 2010). Enzymes with dual activity of xylanase and cellulase have previously been reported (Maurelli et al. 2008; Ratanachomsri et al. 2006; Xue et al. 1992), including GH5 enzymes (Cai et al. 2010), and have been proposed to make cellulose more accessible (Maurelli et al. 2008). This could presumably be the role of the xylanase activity by the AGa GH5s, but further analysis is required to verify these results and assumptions. Especially due to the presence of both cellodextrins and xylodextrins in the negative control sample, although signals from this sample were much lower than from the enzyme samples.

As none of the enzymes showed activity on xylan or arabinoxylan in the hemicellulosic assays, the xylodextrins observed in the mass spectra from switchgrass degradation were unexpected. This could be explained by the low sensitivity of the DNS assays used for individual hemicellulose testing or perhaps to the model substrates being different from the xylan present in switchgrass. Testing the enzymes on alternative xylan substrates and subsequently analysing the product formation by more sensitive methods like HPAEC-PAD would give more insight into these potential xylan activities, and should be considered in further studies. In addition, as for the filter paper assay, positive controls should be performed for comparison.

The predicted CBM domain of Cel5A was assayed for activity on CMC and checked for binding to Avicel. No activity on CMC was observed. Instead, Cel5A_d2 bound to Avicel with approximately 50 % bound enzyme, supporting its annotation as a CBM. Note that lysozyme (non-binding control) and Cel5A_d2 incubated without Avicel also show a slight drop in free enzyme from non-incubated samples, this might be due to some protein binding to the assay tubes. However, Cel5A_d2 has a considerably smaller amount of free protein after incubation that the controls. In addition, the effect of the CBM was tested on CMC, by assaying both full-length Cel5A and Cel5A_cat, the catalytic domain only. No difference in activity was observed under the given conditions, hence no effect of the CBM for the full-length protein on CMC and no catalytic activity was ascribed to the annotated CBM, as previously tested. The full-length Cel5A and Cel5A_cat were not compared with respect to activity on Avicel, due to inadequate enzyme availability (see above). Activity experiment with Avicel could have indicated if the

CBM enhances the performance of the full-length enzyme, as CBMs have been shown to increase cellulase activity towards crystalline substrates (e.g. Avicel), but not increase activity towards soluble substrates (e.g. CMC) (Bolam et al. 1998; Gilkes et al. 1988; Hall et al. 1995). Although, some endoglucanases have previously been reported to not be affected by the removal of CBM, unlike exoglucanases, in 1 % Avicel degradation (Várnai et al. 2013).

In an effort to obtain structural information on the presumed mutation and active site of Cel5C_2, and to study if the enzyme could be using the inserted Asp as the catalytic amino acid, crystallization screening was performed. However, in the time of this project, no crystal formation could be observed. Troubleshooting the crystallization screening could include assessing many factors that affect the production of crystals, including: protein concentration, concentration of precipitants, temperature, protein denaturation and protein purity (Benvenuti & Mangani 2007). Additional screens could be set up with even higher protein concentration and other conditions, such as different or other concentrations of precipitants. The protein sample could also be further purified by SEC or other methods to obtain a more homogenous protein sample, although Cel5C_2 seemed relatively pure when assessed by LDS-PAGE. Crystallization screening at other facilities using robotic set-ups for more rapid screens was considered, and could be done in the future. Structural characterization of the remaining GH5 domains and full-length enzymes should also be considered, to potentially further discriminate between these enzymes.

5.4 Metaproteomics

The metaproteomic investigation of a new cow rumen sample resulted in no matches to the AGa gene cluster enzymes. The metaproteomics was not performed on the same sample as the metagenomic dataset, as we had no access to this sample. Moreover, the samples used for metaproteomics in this study were not enriched on switchgrass as were the original samples that generated the metagenome and the AGa genome bin, and this might be the reason why no AGa enzymes were observed. However, both the samples were from cows fed with the same diet. As mentioned earlier in section 1.6, the composition of the microbial community in the cow rumen can vary between individuals and the samples used in the metaproteomic experiment can presumably contain less of the bacteria harbouring the AGa genome. In addition, the metabolic state of the cow can affect the expression of different genes, and the metaproteomic sample might not contain AGa proteins due to this. Contaminant proteins during

metaproteomic sample preparation and other proteins present in the original sample can also overshadow the target proteins.

5.5 Summary and concluding remarks

The AGa gene cluster contains both single- and multi-modular enzymes, and several of these are targeted for secretion by the T9SS. All enzymes have optimal activities at pH 5.0-6.0 and 40 ºC, but activities on CMC vary between the enzymes. The GH5s are able to cleave substrates containing β-1,4-glycosidic backbones and exclusively cleave β-1,4-likages. They also cleave cellopentaose to cellobiose and cellotriose, and cellohexaose to cellobiose, cellotriose and cellotetrose (futher broken down to cellobiose). Most of the enzymes are also active on crystalline cellulose (filter paper), producing mostly cellobiose, and the lignocellulosic substrate switchgrass, producing both cellodextrins and xylodextrins.

The enzyme that exerts the greatest structural and mechanistic difference is Cel5B. This enzyme does not contain the secretion tag for the T9SS, as full-length Cel5A, Cel5C and Cel5D does. Cel5B cleaves cellohexaose to solely cellotriose with high affinity, while the other enzymes cleave cellohexaose to cellobiose, cellotriose and cellotetraose. Finally, Cel5B shows a higher thermal stability than the rest of the cluster enzymes. The second domain of Cel5C was restored after discovery of a possible mutated catalytic amino acid. The Cel5C_2 domains have lower activity compared to the other GH5s and restoration did not increase activity as much as expected, but this could be due to Cel5C_2R expression troubles or Asp in Cel5C_2 performing as the catalytic amino acid. Hence, sequencing or assembly error is likely, due to increased activity on beta-glucan, possible activity on CM-pachyman and higher activity on filter paper for the restored protein. The restored version has, as mentioned, trace activity on the β-1,3-linked substrate CM-pachyman, suggesting a role in hemicellulose removal for better access to cellulose for the first domain of the full-length enzyme. This demonstrates inter-domain complementary functions in a single enzyme, which has been proven to be an effective strategy in biomass degradation (Brunecky et al. 2013; Larsbrink et al. 2016). No synergy between the enzymes was seen on filter paper. The putative CBM of Cel5A was shown to bind to cellulose, revealing a binding function for this second domain of Cel5A.

As differences in both the structural organization and activity are seen between the enzymes, even though they all are GH5s, different roles of the enzymes can be envisaged. As Cel5B shows an affinity for cellohexaose, and less affinity for cellulosic substrates, it can be speculated

that this enzyme might have a role as a second line of degradation, where it degrades the products of the other enzymes. As the three remaining enzymes were not produced natively and secreted via their normal T9SS, they are not subjected to further post translational modifications and cell surface attachment, hence the enzymatic activities and (lack of) synergies determined here could be different from *in vivo* actions. The importance of glycosylation was shown for CelA of *C. bescii*, which exerted significant increases in activity when glycosylated (Chung et al. 2015). In addition, troubles with producing soluble protein may have resulted in misfolded or partially misfolded protein, and the observed activities could be affected by this. As Cel5C_2R shows trace activity on β-1,3-linkages (which would need to be further assessed) and the cluster protein releases xylodextrins from switchgrass, this could imply degradation of hemicelluloses by the cluster for access to the cellulose as previously discussed.

The gene cluster is also a part of a greater genome bin (AGa) with other carbohydrate-active enzymes that may contribute to lignocellulosic degradation. Given the presence of several multi-domain cellulases (like Cel5C) in the AGa genome, it is possible that the original bacterium could utilize its collection of multi-domain T9SS endocellulases to degrade cellulose together with the single- and multi domain endocellulases of this cluster. Both *C. hutchinsonii* and *F. succinogenes* are able to degrade cellulose without apparent exocellulases, and this could also be the case for the AGa phylotype.

In conclusion, the analysis of a gene cluster from a Bacteroidetes-affilated phylotype encoding four GH5 enzymes conducted in this thesis gives insight into the non-classical cellulolytic mechanisms not associated with cellulosomes, free secreted enzymes (with cellobiohydrolases) or PULs. The complementary functions of the four GH5 enzymes encoded by this cluster were investigated, and the functional differences between the enzymes were discovered. While further work is needed, the present results contribute to our knowledge on microbial degradation of cellulose in the rumen as well as non-classical cellulolytic mechanisms. This expansion of knowledge on natural cellulose degradation, could eventually lead to improvements in industrial biomass conversion.

However, more work on the AGa genome and the cellulase gene cluster is needed to fully understand the mechanism for cellulose degradation by this phylotype.

5.6 Future perspectives

For future work, assessing the role of Cel5C_2 and the nature of the possible mutation further would be interesting. Assays on CM-pachyman for Cel5C_2R require re-examination with higher enzyme dosages. Expression of the restored version should be additionally optimized, for example using a fusion tag to ensure correctly folded protein and investigate if this increases its activity. New crystallization screens for the second domain of Cel5C should also be set up, in hope of getting more information of the possible mutation from the structure of the domain and the active cleft. The metaproteomics can be repeated with switchgrass-enriched samples, to check if this leads to the detection of AGa proteins.

The xylanase activity of the GH5 enzymes on switchgrass should be further investigated and verified by testing the enzymes on a variety of different xylan substrates, along with more investigation on switchgrass, and more in-depth analysis of the results by HPAEC-PAD or similar methods. Information regarding the enzymes' ability to completely degrade xylan or partially deconstruct it to improve enzyme accessibility to cellulose is desirable.

Further optimization of expression of the T9SS enzymes could be based on heterologous expression of the enzyme in a T9SS harbouring bacterium, for proper transportation and modification of the proteins containing the C-terminal secretion tag for the system. Clearly, efforts should be made to isolate a representative of AGa phylotype, to determine whether it is capable of cellulose utilization. This would also allow for the study of individual knockout mutants to evaluate the function, activity and roles of the different GH5 containing enzymes, and the necessity of each of the enzymes.

6 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215 (3): 403-10.

Arntzen, M. O., Varnai, A., Mackie, R. I., Eijsink, V. G. H. & Pope, P. B. (2017). Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of Carbohydrate-Active Enzymes with versatile polysaccharide-degrading capacity. *Environ Microbiol*, Accepted manuscript posted online 26 April 2017, doi: 10.1111/1462-2920.13770

Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer III, H. & Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology*, 12: 186.

Beckham, G. T., Dai, Z., Matthews, J. F., Momany, M., Payne, C. M., Adney, W. S., Baker, S. E. & Himmel, M. E. (2012). Harnessing glycosylation to improve cellulase activity. *Current Opinion in Biotechnology*, 23 (3): 338-345.

Benvenuti, M. & Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols*, 2 (7): 1633-1651.

Bertani, G. (1951). STUDIES ON LYSOGENESIS. I. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62 (3): 293-300.

Bjursell, M. K., Martens, E. C. & Gordon, J. I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem*, 281 (47): 36269-79.

Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. & Gilbert, H. J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J*, 331 (Pt 3): 775-81.

Borzova, N. V., Gudzenko, O. V. & Varbanets, L. D. (2014). Role of glycosylation in secretion and stability of micromycetes alpha-galactosidase. *Ukr Biochem J*, 86 (6): 31-8.

Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., Yang, S.-J., Resch, M. G., Adams, M. W. W., Lunin, V. V., et al. (2013). Revealing Nature's Cellulase Diversity: The Digestion Mechanism of *Caldicellulosiruptor bescii* CelA. *Science*, 342 (6165): 1513-1516.

Cai, S., Li, J., Hu, F. Z., Zhang, K., Luo, Y., Janto, B., Boissy, R., Ehrlich, G. & Dong, X. (2010). *Cellulosilyticum ruminicola*, a Newly Described Rumen Bacterium That

Possesses Redundant Fibrolytic-Protein-Encoding Genes and Degrades Lignocellulose with Multiple Carbohydrate- Borne Fibrolytic Enzymes. *Applied and Environmental Microbiology*, 76 (12): 3818-3824.

Chandel, A. K., Goncalves, B. C., Strap, J. L. & da Silva, S. S. (2015). Biodelignification of lignocellulose substrates: An intrinsic and sustainable pretreatment strategy for clean energy production. *Crit Rev Biotechnol,* 35 (3): 281-93.

Chung, D., Young, J., Bomble, Y. J., Vander Wall, T. A., Groom, J., Himmel, M. E. & Westpheling, J. (2015). Homologous Expression of the Caldicellulosiruptor bescii CelA Reveals that the Extracellular Protein Is Glycosylated. *PLoS ONE,* 10 (3): e0119508.

Costa, S., Almeida, A., Castro, A. & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in Microbiology,* 5: 63.

Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3 (9): 853-859.

de Gonzalo, G., Colpa, D. I., Habib, M. H. M. & Fraaije, M. W. (2016). Bacterial enzymes involved in lignin degradation. *Journal of Biotechnology*, 236: 110-119.

Demain, A. L. & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, 27 (3): 297-306.

Edwards, J. E., McEwan, N. R., Travis, A. J. & John Wallace, R. (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie van Leeuwenhoek*, 86 (3): 263-281.

Fontes, C. M. & Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem,* 79: 655-81.

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

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In Walker, J. M. (ed.) *The Proteomics Protocols Handbook*, pp. 571-607. Totowa, NJ: Humana Press.

Gilbert, H. J., Stalbrand, H. & Brumer, H. (2008). How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol,* 11 (3): 338-48.

Gilkes, N. R., Warren, R. A., Miller, R. C., Jr. & Kilburn, D. G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *J Biol Chem*, 263 (21): 10401-7.

Gonzalez-Montalban, N., Garcia-Fruitos, E. & Villaverde, A. (2007). Recombinant protein solubility - does more mean better? *Nat Biotechnol*, 25 (7): 718-20.

Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W. & Brumer, H. (2017). Polysaccharide Utilization Loci: Fuelling microbial communities. *J Bacteriol*, Accepted manuscript posted online 30 January 2017, doi:10.1128/JB.00860-16.

Gupta, R., Baldock, S. J., Fielden, P. R. & Grieve, B. D. (2011). Capillary zone electrophoresis for the analysis of glycoforms of cellobiohydrolase. *Journal of Chromatography A*, 1218 (31): 5362-5368.

Gupta, V. K., Kubicek, C. P., Berrin, J.-G., Wilson, D. W., Couturier, M., Berlin, A., Filho, E. X. F. & Ezeji, T. (2016). Fungal Enzymes for Bio-Products from Sustainable and Waste Biomass. *Trends in Biochemical Sciences*, 41 (7): 633-645.

Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Ali, B. R., Hazlewood, G. P. & Gilbert, H. J. (1995). The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. cellulosa is important for the efficient hydrolysis of Avicel. *Biochemical Journal*, 309 (Pt 3): 749-756.

Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., Zhang, T., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 331 (6016): 463-7.

Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. & Foust, T. D. (2007). Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, 315 (5813): 804-807.

Hoffmann, G. C., Simson, B. W. & Timell, T. E. (1971). Structure and molecular size of pachyman. *Carbohydr Res*, 20 (1): 185-8.

Holtzapple, M., Cognata, M., Shu, Y. & Hendrickson, C. (1990). Inhibition of *Trichoderma reesei* cellulase by sugars and solvents. *Biotechnol Bioeng*, 36 (3): 275-87.

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

Hu, Z., Sykes, R., Davis, M. F., Charles Brummer, E. & Ragauskas, A. J. (2010). Chemical profiles of switchgrass. *Bioresource Technology*, 101 (9): 3253-3257.

Hughes, A. L. (2012). Evolution of the βGRP/GNBP/β-1,3-glucanase family of insects. *Immunogenetics*, 64 (7): 549-558.

Igarashi, K., Samejima, M. & Eriksson, K. E. (1998). Cellobiose dehydrogenase enhances *Phanerochaete chrysosporium* cellobiohydrolase I activity by relieving product inhibition. *Eur J Biochem*, 253 (1): 101-6.

Ishler, V. A., Bánné Varga, G. & Heinrichs, A. J. (1996). *From feed to milk: understanding rumen function*. University Park, Pa.: Pennsylvania State University.

Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett*, 362 (3): 281-5.

Jeon, W. B., Aceti, D. J., Bingman, C. A., Vojtik, F. C., Olson, A. C., Ellefson, J. M., McCombs, J. E., Sreenath, H. K., Blommel, P. G., Seder, K. D., et al. (2005). High-throughput Purification and Quality Assurance of *Arabidopsis thaliana* Proteins for Eukaryotic Structural Genomics. *Journal of Structural and Functional Genomics*, 6 (2): 143-147.

Johansen, K. S. (2016). Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation. *Trends in Plant Science*, 21 (11): 926-936.

Kim, M., Morrison, M. & Yu, Z. (2011). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiology Ecology*, 76 (1): 49-63.

Kipper, K., Väljamäe, P. & Johansson, G. (2005). Processive action of cellobiohydrolase Cel7A from *Trichoderma reesei* is revealed as 'burst' kinetics on fluorescent polymeric model substrates. *Biochemical Journal*, 385 (Pt 2): 527-535.

Konietzny, S. G., Pope, P. B., Weimann, A. & McHardy, A. C. (2014). Inference of phenotype-defining functional modules of protein families for microbial plant biomass degraders. *Biotechnol Biofuels*, 7 (1): 124.

Koropatkin, N. M., Cameron, E. A. & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nat Rev Micro*, 10 (5): 323-335.

Koshland, D. E. (1953). STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews*, 28 (4): 416-436.

Kostylev, M. & Wilson, D. (2012). Synergistic interactions in cellulose hydrolysis. *Biofuels*, 3 (1): 61-70.

Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G. H., Koropatkin, N. M., McBride, M. J. & Pope, P. B. (2016). A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnology for Biofuels*, 9 (1): 260.

Lesk, A. M. (2010). *Introduction to Protein Science*. 2nd ed. New York, NY, United States: Oxford University Press Inc.

Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat Rev Micro*, 14 (5): 288-304.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, 42 (Database issue): D490-5.

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.

Mao, S., Zhang, M., Liu, J. & Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Sci Rep*, 5: 16116.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*, 45 (D1): D200-d203.

Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. (2009). Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *The Journal of Biological Chemistry*, 284 (37): 24673-24677.

Maurelli, L., Giovane, A., Esposito, A., Moracci, M., Fiume, I., Rossi, M. & Morana, A. (2008). Evidence that the xylanase activity from *Sulfolobus solfataricus* Oα is encoded by the endoglucanase precursor gene (sso1354) and characterization of the associated cellulase activity. *Extremophiles*, 12 (5): 689-700.

McBride, M. J. & Zhu, Y. (2013). Gliding Motility and Por Secretion System Genes Are Widespread among Members of the Phylum Bacteroidetes. *Journal of Bacteriology*, 195 (2): 270-278.

McBride, M. J. & Nakane, D. (2015). Flavobacterium gliding motility and the type IX secretion system. *Current Opinion in Microbiology*, 28: 72-77.

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996). The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. *Biochemistry*, 35 (31): 9958-66.

Miller, G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, 31 (3): 426-428.

Moraïs, S., Morag, E., Barak, Y., Goldman, D., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D. B. & Bayer, E. A. (2012). Deconstruction of Lignocellulose into Soluble Sugars by Native and Designer Cellulosomes. *mBio*, 3 (6): e00508-12.

Naas, A. E., Mackenzie, A. K., Mravec, J., Schuckel, J., Willats, W. G., Eijsink, V. G. & Pope, P. B. (2014). Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio*, 5 (4): e01401-14.

Park, J. I., Kent, M. S., Datta, S., Holmes, B. M., Huang, Z., Simmons, B. A., Sale, K. L. & Sapra, R. (2011). Enzymatic hydrolysis of cellulose by the cellobiohydrolase domain of CelB from the hyperthermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Bioresource Technology*, 102 (10): 5988-5994.

Peng, F., Ren, J.-L., Xu, F., Bian, J., Peng, P. & Sun, R.-C. (2009). Comparative Study of Hemicelluloses Obtained by Graded Ethanol Precipitation from Sugarcane Bagasse. *Journal of Agricultural and Food Chemistry*, 57 (14): 6305-6317.

Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., McKinnon, J. J. & McAllister, T. A. (2014). Characterization of the Core Rumen Microbiome in Cattle during Transition from Forage to Concentrate as Well as during and after an Acidotic Challenge. *PLOS ONE*, 8 (12): e83424.

Quiroz-Castañeda, R. E. & Folch-Mallol, J. L. (2013). Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars. In *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*, p. Ch. 0. Rijeka: InTech.

Ratanachomsri, U., Sriprang, R., Sornlek, W., Buaban, B., Champreda, V., Tanapongpipat, S. & Eurwilaichitr, L. (2006). Thermostable xylanase from *Marasmius* sp.: purification and characterization. *J Biochem Mol Biol*, 39 (1): 105-10.

Ratanakhanokchai, K., Waeonukul, R., Pason, P., Tachaapaikoon, C., Kyu, K. L., Sakka, K., Kosugi, A. & Mori, Y. (2013). *Paenibacillus curdlanolyticus* strain B-6 multienzyme complex: A novel system for biomass utilization. *Biomass Now-Cultivation and Utilization*: 369-394.

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, 59: 1-41.

Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Baslé, A., Morland, C., Day, A. M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications*, 6: 7481.

Rosano, G. L. & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5: 172.

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature*, 454 (7206): 841-845.

Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G. & Nakayama, K. (2010). A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (1): 276-281.

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

Shafiee, S. & Topal, E. (2009). When will fossil fuel reserves be diminished? *Energy Policy*, 37 (1): 181-189.

Shoseyov, O., Shani, Z. & Levy, I. (2006). Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiology and Molecular Biology Reviews*, 70 (2): 283-295.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., et al. (2004). Toward a Systems Approach to Understanding Plant Cell Walls. *Science*, 306 (5705): 2206-2211.

Suen, G., Weimer, P. J., Stevenson, D. M., Aylward, F. O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N. N., Mikhailova, N., Chertkov, O., et al. (2011). The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist. *PLoS ONE*, 6 (4): e18814.

Teeri, T. T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15 (5): 160-167.

Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. (2015). Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*, 31 (5): 647-655.

Teugjas, H. & Väljamäe, P. (2013). Selecting β-glucosidases to support cellulases in cellulose saccharification. *Biotechnology for Biofuels*, 6 (1): 105.

U.S. Energy Information Administration. (2016). *International Energy Outlook 2016*. Available at: https://www.eia.gov/outlooks/ieo/pdf/0484(2016).pdf (accessed: April 4, 2017).

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

Vanholme, B., Desmet, T., Ronsse, F., Rabaey, K., Van Breusegem, F., De Mey, M., Soetaert, W. & Boerjan, W. (2013). Towards a carbon-negative sustainable bio-based economy. *Front Plant Sci*, 4: 174.

Várnai, A., Siika-aho, M. & Viikari, L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnology for Biofuels*, 6 (1): 30.

Veith, P. D., Nor Muhammad, N. A., Dashper, S. G., Likić, V. A., Gorasia, D. G., Chen, D., Byrne, S. J., Catmull, D. V. & Reynolds, E. C. (2013). Protein Substrates of a Novel Secretion System Are Numerous in the Bacteroidetes Phylum and Have in Common a Cleavable C-Terminal Secretion Signal, Extensive Post-Translational Modification, and Cell-Surface Attachment. *Journal of Proteome Research,* 12 (10): 4449-4461.

Viikari, L., Vehmaanperä, J. & Koivula, A. (2012). Lignocellulosic ethanol: From science to industry. *Biomass and Bioenergy*, 46: 13-24.

Vuong, T. V. & Wilson, D. B. (2010). Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnol Bioeng*, 107 (2): 195-205.

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C.-A., Hyslop, J. J., Ross, D. W., Waterhouse, A., Watson, M. & Roehe, R. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics*, 16 (1): 839.

Weimann, A., Trukhina, Y., Pope, P. B., Konietzny, S. G. & McHardy, A. C. (2013). De novo prediction of the genomic components and capabilities for microbial plant biomass degradation from (meta-)genomes. *Biotechnol Biofuels*, 6 (1): 24.

White, B. A., Lamed, R., Bayer, E. A. & Flint, H. J. (2014). Biomass utilization by gut microbiomes. *Annu Rev Microbiol*, 68: 279-96.

Wilson, D. B. (2008). Three microbial strategies for plant cell wall degradation. *Ann N Y Acad Sci*, 1125: 289-97.

Wilson, D. B. (2009). Cellulases and biofuels. *Curr Opin Biotechnol,* 20 (3): 295-9.

Wilson, D. B. (2011). Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology*, 14 (3): 259-263.

Xie, G., Bruce, D. C., Challacombe, J. F., Chertkov, O., Detter, J. C., Gilna, P., Han, C. S., Lucas, S., Misra, M., Myers, G. L., et al. (2007). Genome Sequence of the Cellulolytic Gliding Bacterium *Cytophaga hutchinsonii*. *Applied and Environmental Microbiology*, 73 (11): 3536-3546.

Xue, G. P., Gobius, K. S. & Orpin, C. G. (1992). A novel polysaccharide hydrolase cDNA (celD) from *Neocallimastix patriciarum* encoding three multi-functional catalytic

domains with high endoglucanase, cellobiohydrolase and xylanase activities. *J Gen Microbiol*, 138 (11): 2397-403.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*, 40 (Web Server issue): W445-51.

Zhang, H., Zhang, J. L., Sun, L., Niu, X. D., Wang, S. & Shan, Y. M. (2014). Molecular dynamics simulation of the processive endocellulase Cel48F from *Clostridium cellulolyticum*: a novel "water-control mechanism" in enzymatic hydrolysis of cellulose. *J Mol Recognit*, 27 (7): 438-47.

Zhang, S., Wolfgang, D. E. & Wilson, D. B. (1999). Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol Bioeng*, 66 (1): 35-41.

Zhu, Y. & McBride, M. J. (2014). Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Applied Microbiology and Biotechnology*, 98 (2): 763-775.

Zhu, Y., Kwiatkowski, K. J., Yang, T., Kharade, S. S., Bahr, C. M., Koropatkin, N. M., Liu, W. & McBride, M. J. (2015). Outer membrane proteins related to SusC and SusD are not required for *Cytophaga hutchinsonii* cellulose utilization. *Appl Microbiol Biotechnol*, 99 (15): 6339-50.

Appendix

Appendix A: pNIC-CH plasmid maps

An over view of the pNIC-CH plasmid with named features including the *sacB* gene is shown in Figure A1. A pNIC-CH vector with an inserted Cel5B-encoding gene, as an example, is shown in Figure A2.

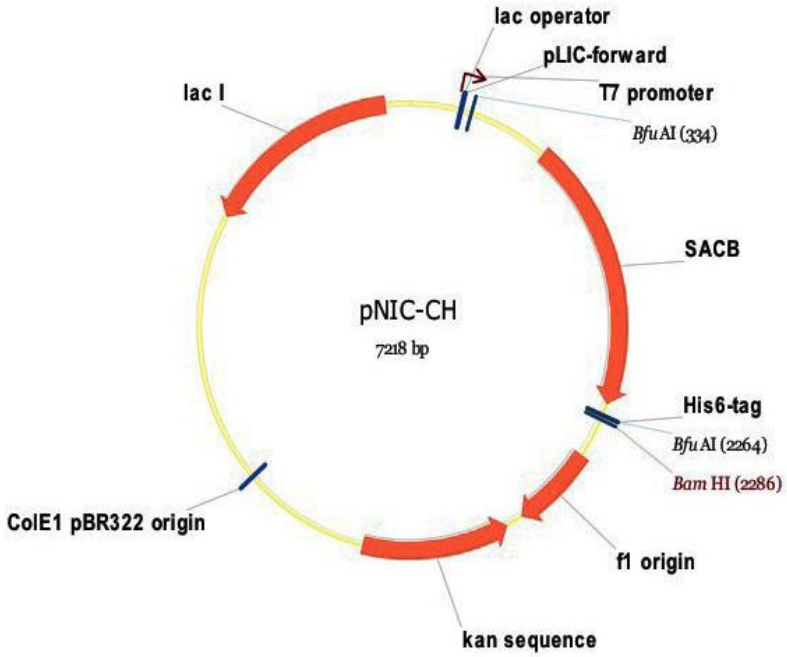


Figure A1. pNIC-CH plasmid map with *sacB* gene. pNIC-CH was a gift from Opher Gileadi (Addgene plasmid # 26117). Figure taken from Addgene.

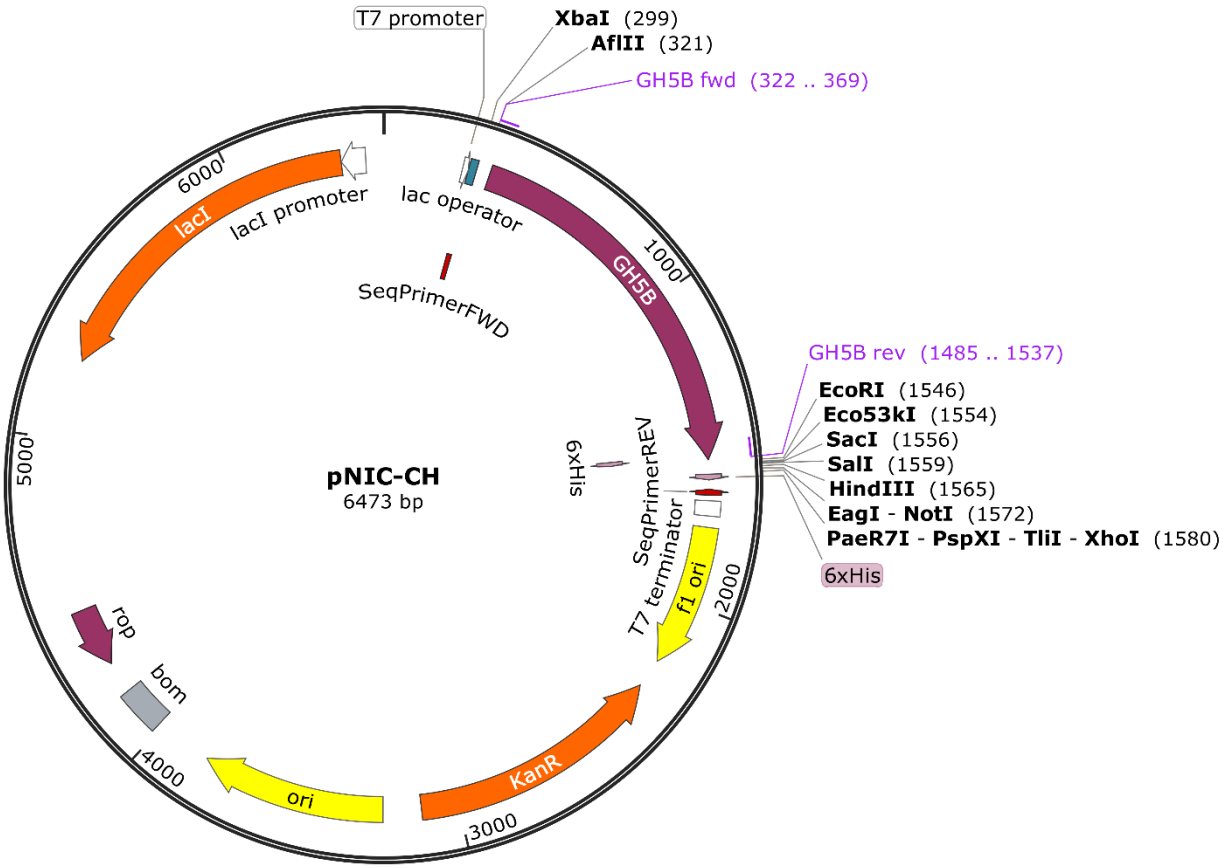


Figure A2. pNIC-CH-based expression vector for production of Cel5B. The picture was made in SnapGene.

Appendix B: IMAC purification example

Purification of recombinantly expressed Cel5B is used as an example of IMAC chromatography, with the chromatogram showed in Figure B1 and LDS-PAGE of fractions after IMAC purification shown in Figure B2.

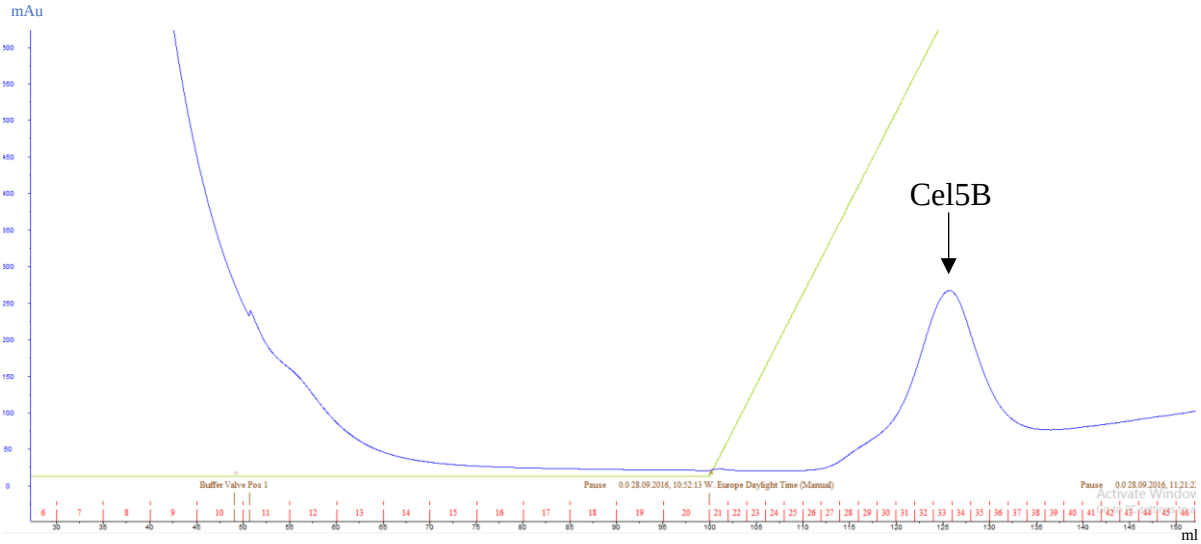


Figure B1. Chromatogram of IMAC purification of Cel5B. The blue line shows UV absorbance (mAu), the green line shows the gradient of elution buffer B, and the red numbers indicate fraction numbers. The flow-through UV peak has been excluded to emphasise the peak of the eluted protein (marked).

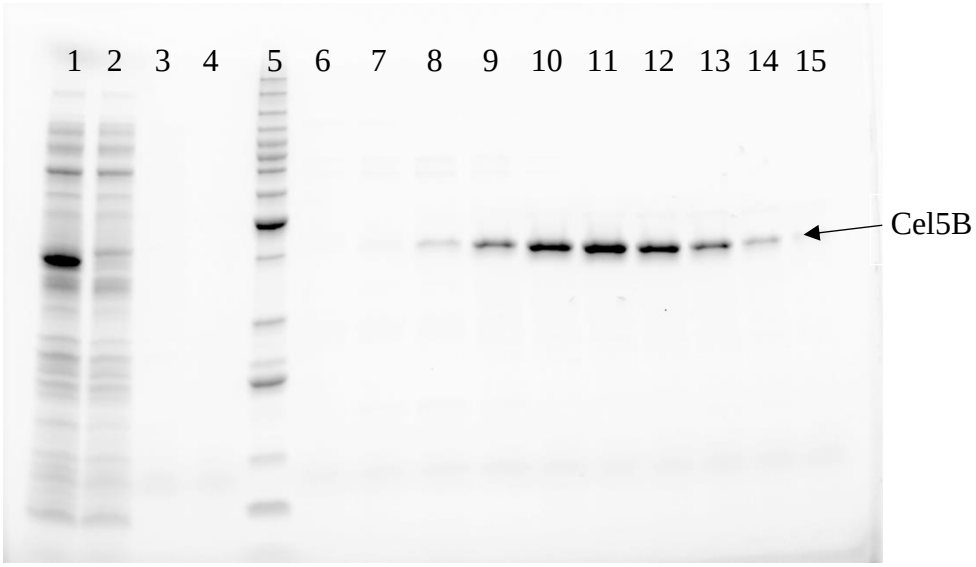


Figure B2. LDS-PAGE fractions after IMAC purification of Cel5B. Fraction samples are from the cell lysate (1), flow-through (2), flow-through tail (3), wash (4) and fraction 28-37 from peak in Figure B1 (6-15). Lane 5 is Benchmark™ Protein Ladder.

Appendix C: Initial rapid rate curves

Figure C1 shows the initial rapid rate curves that were used for determination of assay time and enzyme concentration needed to obtain the conditions for initial rapid release of products that were used in further experiments. Chosen conditions are illustrated in the results chapter (Table 4.3) and circled (red) in Figure C1.

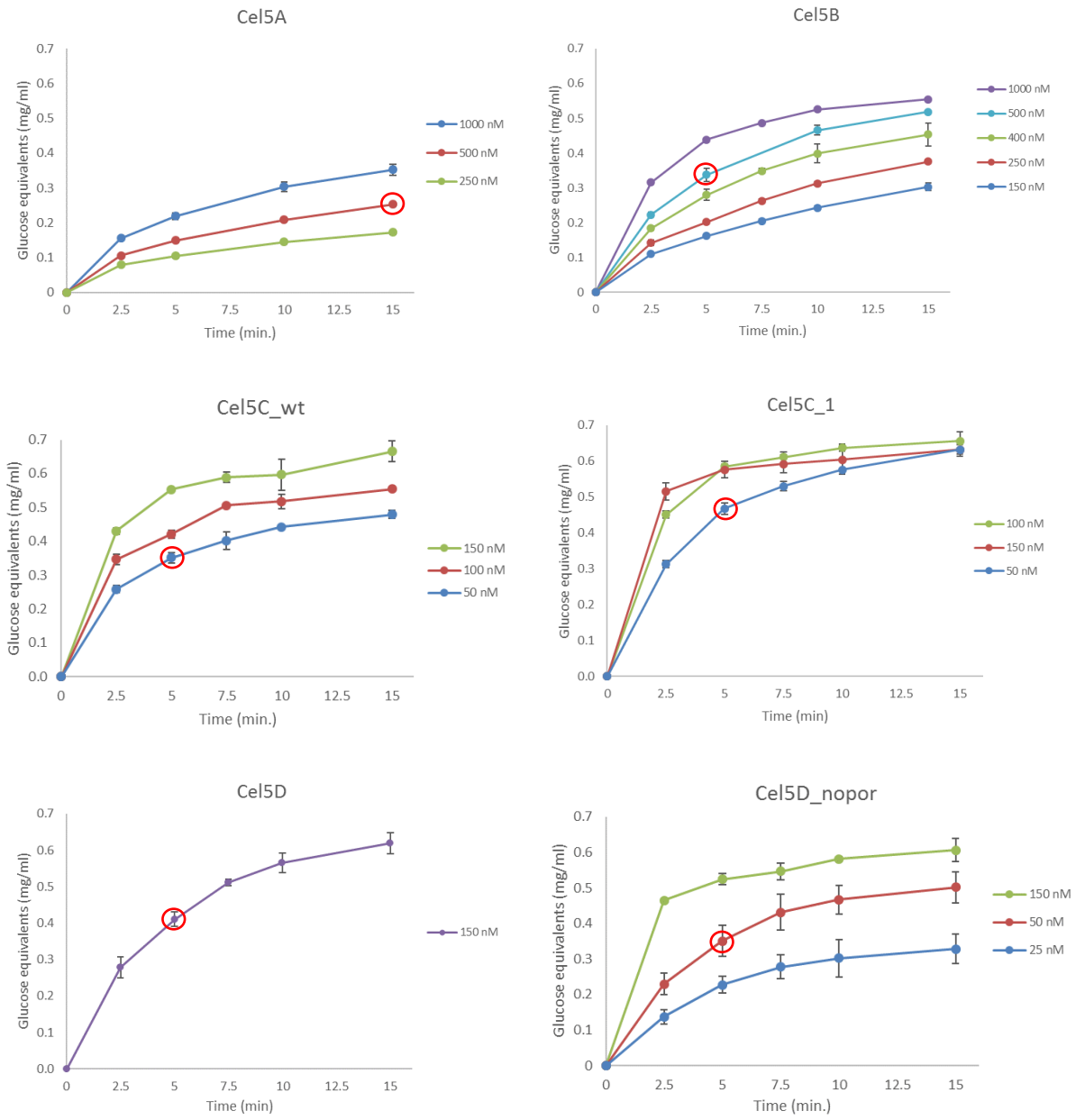


Figure C1. Initial rapid rate curves. Assays were performed at 40 ºC, 1 % (w/v) CMC, 20 mM citrate buffer pH 5.5 and 800 rpm horizontal shaking. Samples were taken at given time points and product formation was analysed by DNS. Error bars represent standard deviations between three replicates. Red circles mark chosen conditions (time and enzyme concentrations) for each enzyme. Cel5A_cat, Cel5C_wtR, Cel5C_2 and Cel5C_2R were not available in sufficient amounts or not active enough for the assay.

Appendix D: Activity on CMC for Cel5A and Cel5A_cat

Preliminary assay on CMC for Cel5A_cat was performed to investigate the effect of the second domain present in the full-length enzyme. Activity on CMC of Cel5A and Cel5A_cat is shown in Figure D1.

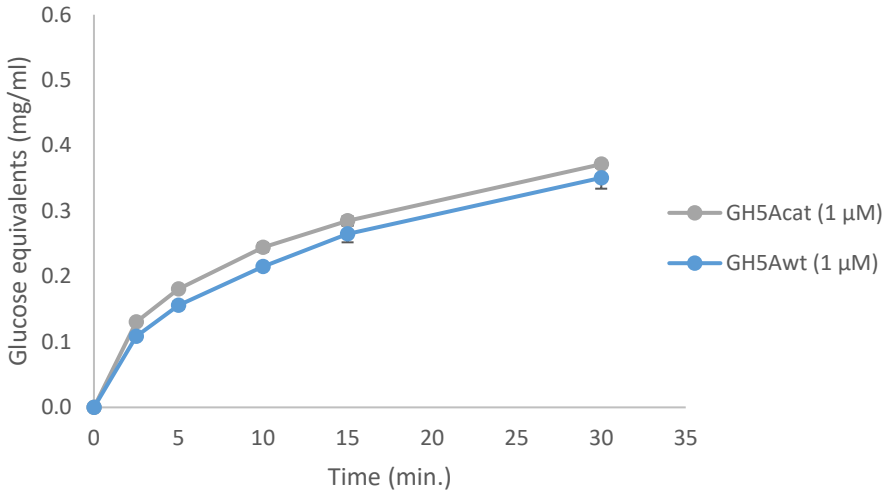


Figure D1. Activity on CMC for Cel5A and Cel5A_cat. Assays were performed at 40 ºC, 1 % (w/v) CMC, 20 mM citrate buffer pH 5.5 and 800 rpm horizontal shaking with 1 µM enzyme. Error bars represent standard deviations between three replicates.

Appendix E: Cellodextrin activity chromatograms

The Figure E1 to Figure E9 shows the chromatograms obtained after HPAEC-PAD analysis of cellodextrin degradation products.

Cel5A

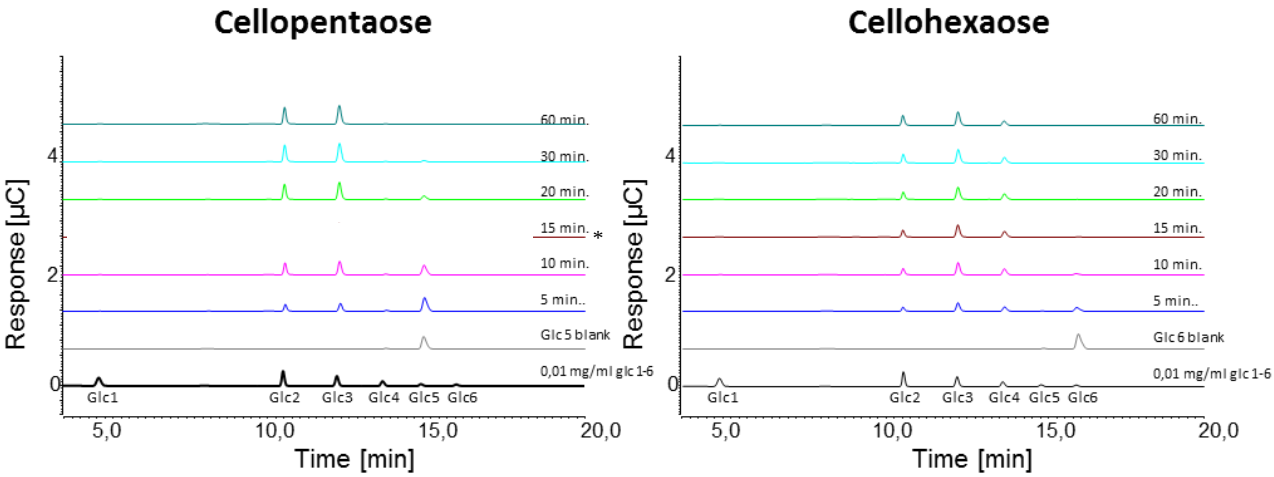


Figure E1. Degradation of cellopentaose and cellohexaose by Cel5A. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 250 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon. *15-minute sample not shown due to lost results during analysis.

Cel5B

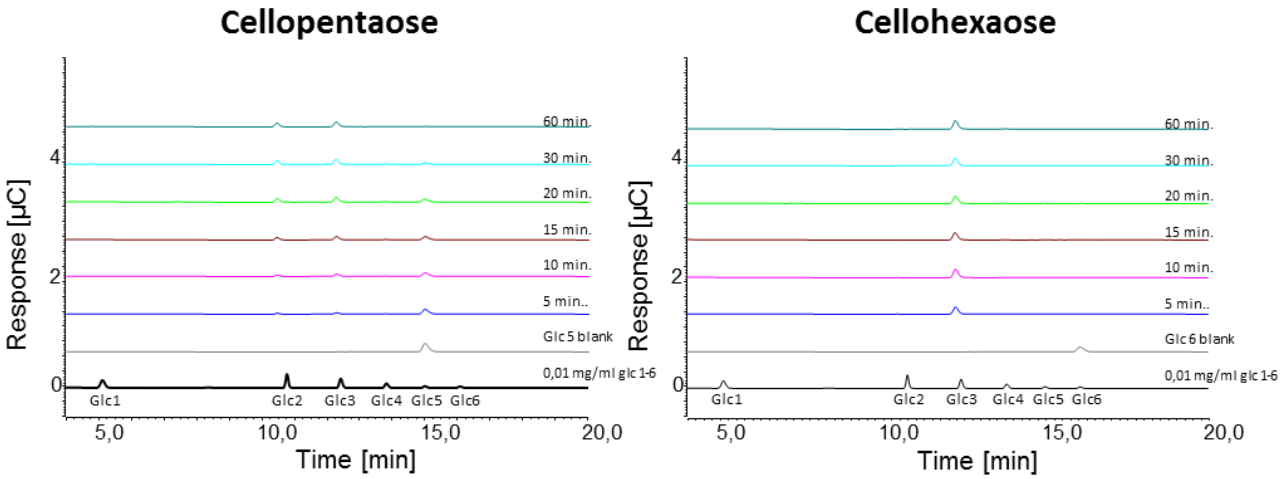


Figure E2. Degradation of cellopentaose and cellohexaose by Cel5B. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 250 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5C_wt

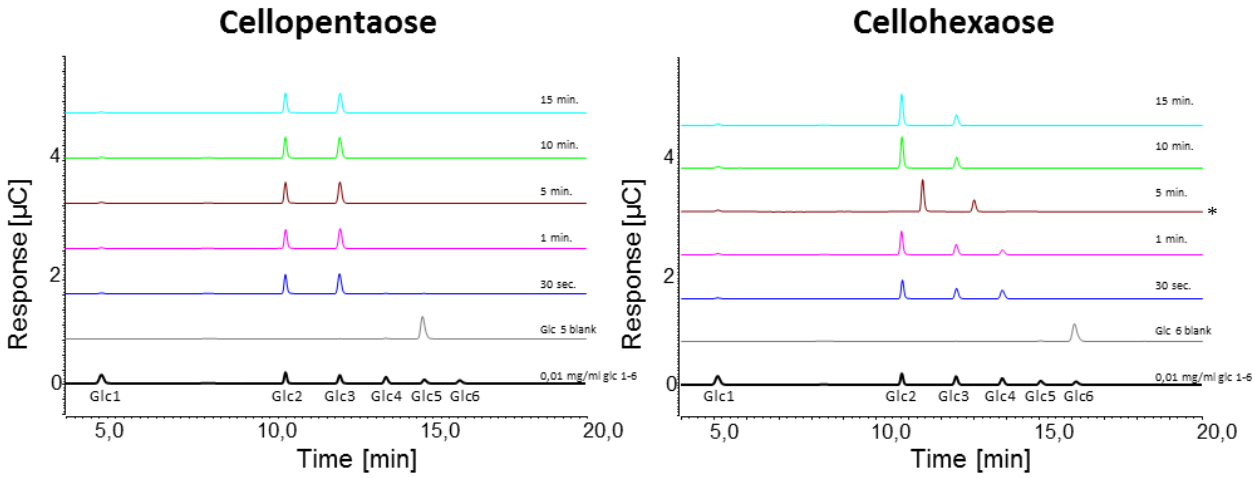


Figure E3. Degradation of cellopentaose and cellohexaose by Cel5C_wt. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 125 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon. *Slightly shifted chromatogram, which sometimes occurs with the ICS-3000 in the lab.

Cel5C_wtR

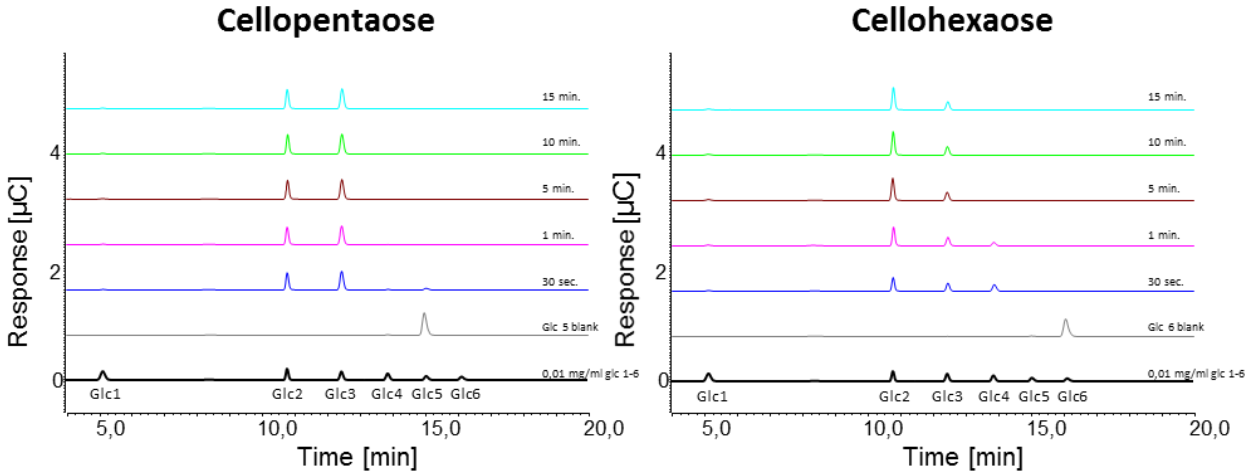


Figure E4. Degradation of cellopentaose and cellohexaose by Cel5C_wtR. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 125 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5C_1

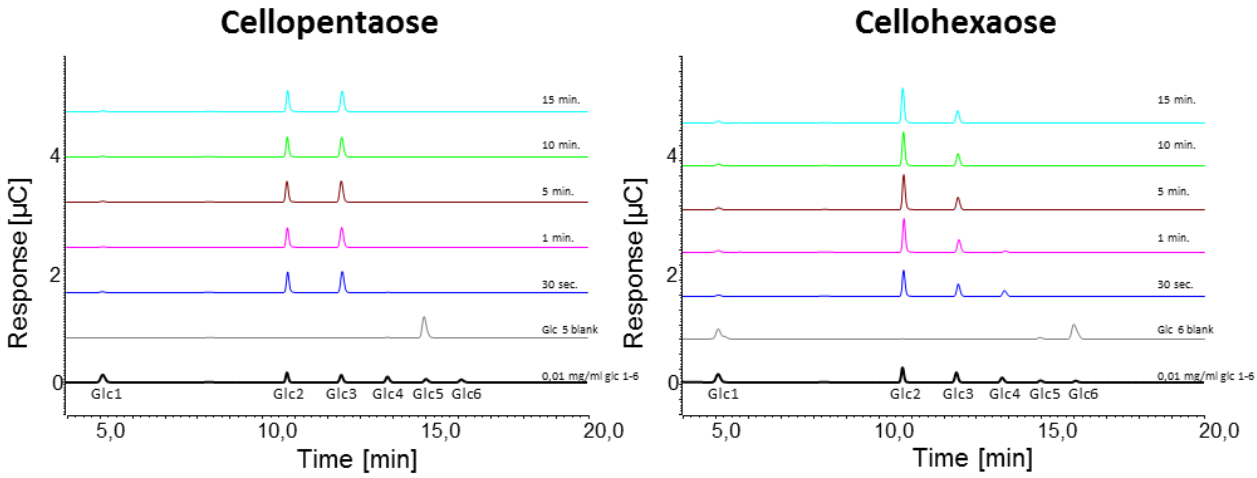


Figure E5. Degradation of cellopentaose and cellohexaose by Cel5C_1. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 125 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5C_2

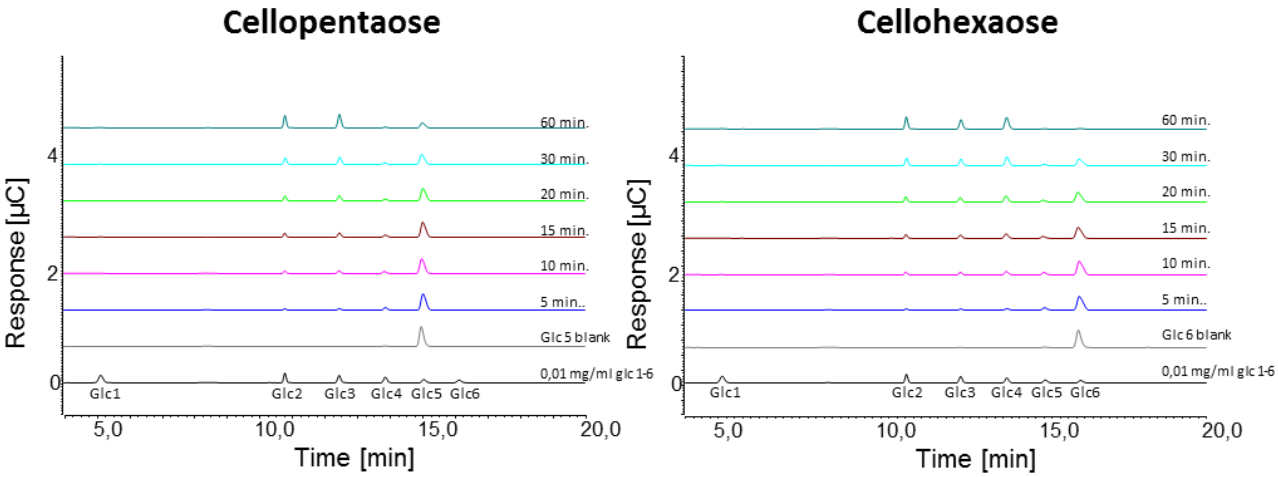


Figure E6. Degradation of cellopentaose and cellohexaose by Cel5C_2. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 250 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5C_2R

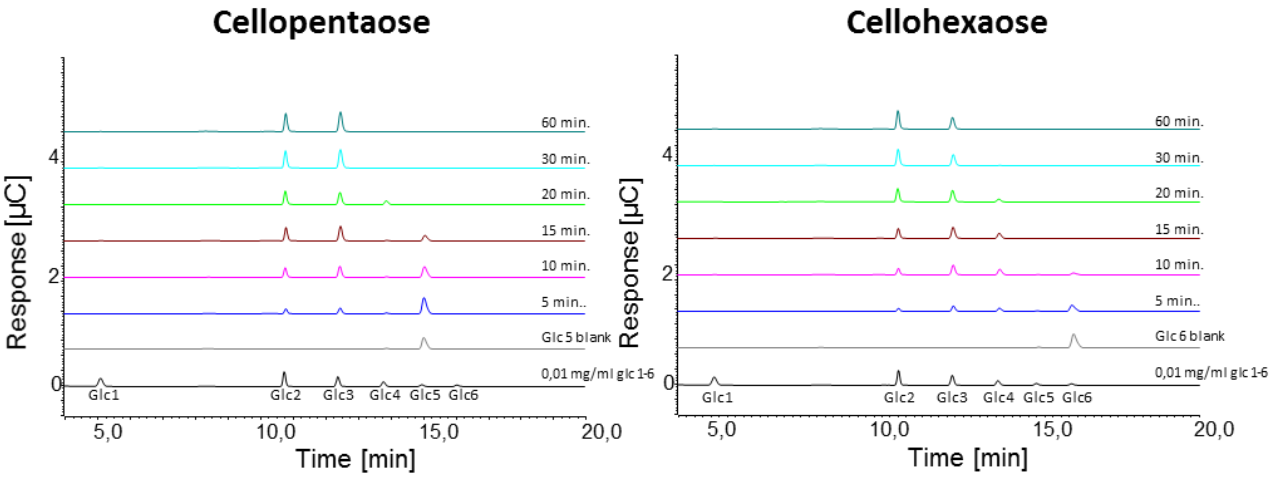


Figure E7. Degradation of cellopentaose and cellohexaose by Cel5C_2R. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 250 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5D

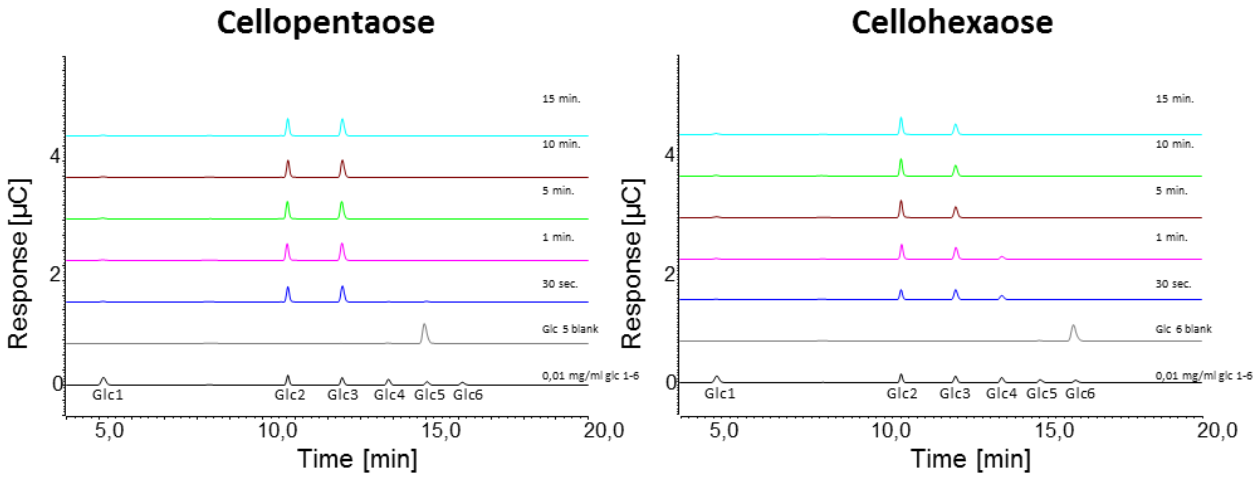


Figure E8. Degradation of cellopentaose and cellohexaose by Cel5D. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 125 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Cel5D_nopor

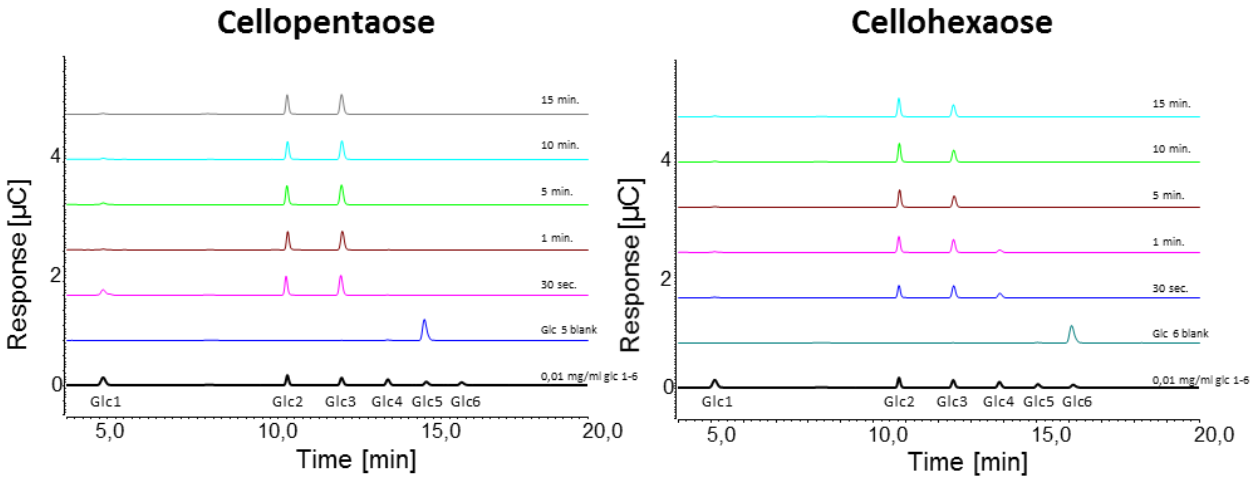


Figure E9. Degradation of cellopentaose and cellohexaose by Cel5D_nopor. Assays were performed at 40 ºC with 0.1 mg/ml substrate, 20 mM citrate buffer pH 5.5 and 125 nM enzyme. Reactions were stopped by adding NaOH to 0.1 M. Reaction products were analysed by HPAEC-PAD and data were analysed in Chromeleon.

Appendix F: Enzymatic activity by AGa enzymes on Switchgrass

Enzyme constructs were tested on switchgrass, a lignocellulosic substrate, to investigate activity on a natural and complex substrate that was also the substrate used to enrich and reconstruct AGa. Product formation by the enzymes were analysed by DNS (Figure F1). As the product formation was low, A_{540} is reported, since quantification by glucose standard curve was not possible.

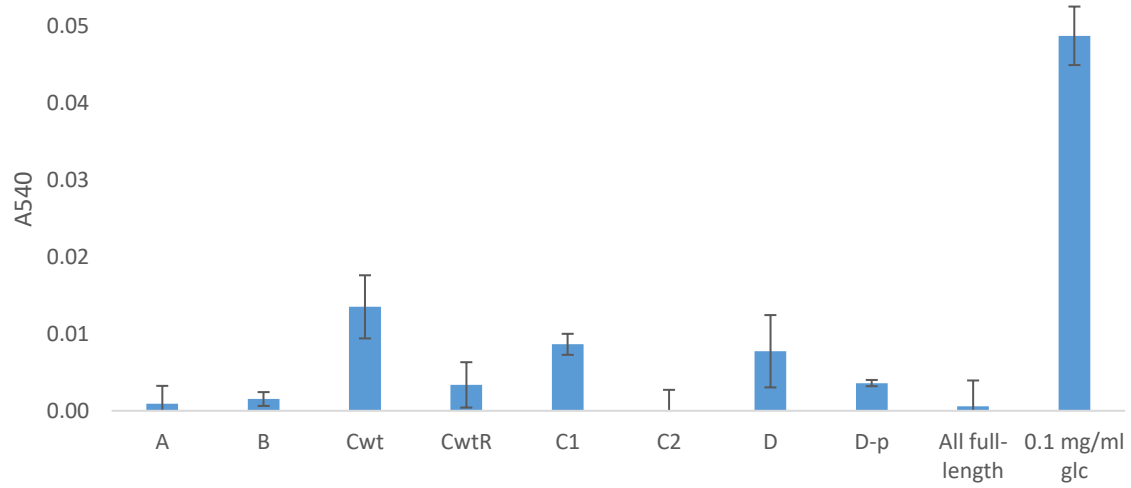
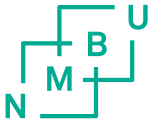


Figure F1. Activity of AGa enzyme constructs on switchgrass. Assays were performed with 0.2 % (w/v) switchgrass, 20 mM citric acid buffer pH 5.5, 1 µM enzyme, 40 ºC and 800 rpm horizontal shaking for 24 hours. Samples were analysed by DNS. Error bars represent standard deviations between three replicates. Values are shown as A_{540}, corrected for background signal, as the product formation was too low to be quantified. The lowest standard curve value is shown on the right.



Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway