Norwegian University of Life Sciences
Faculty of Chemistry, Biotechnology
and Food Science

Philosophiae Doctor (PhD)
Thesis 2019:62

Discovery and characterization of bacterial enzymes for processing of lignocellulosic biomass

Oppdagelse og karakterisering av bakterielle enzymer til prosessering av lignocellulose

Marianne Slang Jensen

Discovery and characterization of bacterial enzymes for processing of lignocellulosic biomass

Oppdagelse og karakterisering av bakterielle enzymer til prosessering av lignocellulose

Philosophiae Doctor (PhD) Thesis

Marianne Slang Jensen

Norwegian University of Life Sciences
Faculty of Chemistry, Biotechnology and Food Science

Ås 2019

Thesis number 2019:62
ISSN 1894-6402
ISBN 978-82-575-1622-2

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The work presented in this thesis was carried out in the Protein Engineering and Proteomics (PEP) group at the Faculty of Chemistry, Biotechnology and Food Science at the Norwegian University of Life Sciences in the period from 2015 to 2019. The project was funded by the Norwegian Research Council through the NorZymeD project (project number 221568).

Firstly, I would like to express my gratitude to my main supervisor **Prof. Vincent Eijsink** for giving me the opportunity to work on this exciting project. I could not have wished for a better supervisor and I am truly grateful for the amount of work you have put into guiding me. Thank you for all your patience and inspiring discussions.

I would also like to thank my co-supervisor **Prof. Gustav Vaaje-Kolstad**. Thank you for encouraging me and for readily sharing your insights and ideas. Your enthusiastic guidance and thorough explanations of complex issues are much appreciated.

To my second co-supervisor **Dr. Zarah Forsberg**, thank you for sharing your impressive knowledge with me and for fun times both in the lab and outside work. I am grateful for all the advice and support I have received from you.

A special thanks to **Lasse Fredriksen** for encouragement and for fun teamwork in the lab. Your humor has really been a highlight of every day at work and I am so happy that you were also assigned to the NorZymeD project. Your positive attitude and your motivational talks have made all the difference when preparing for dreaded presentations and when coping with failed experiments.

Furthermore, I would like to thank **Tina Rise Tuveng** for being an exceptional colleague and friend, especially during the last hectic months. Not only have you been invaluable in finishing the experimental work, but you also let me stay with you and your family whenever I needed to work long nights.

A big thanks to **Ellen Hasle Kokkim** and **Anne Cathrine Bunæs** for managing the lab and for always providing me with answers and solutions to any problem. **Anne Cath**, I have thoroughly enjoyed collaborating with you in the student lab courses, and I am grateful for this experience.

I would also like to thank the rest of the PEP-group for providing a fun and interesting working environment. I have learnt a lot from all of you. I also wish to thank colleagues at NMBU, The University of Bergen, The University of Tromsø, and SINTEF for collaborations that have led to joint publications.

To my family and friends, thank you for always being supportive and for helping me remember the things that matter most in life. **Ole**, thank you for always being there and believing in me; I am lucky to have you.

Marianne Slang Jensen

Ås, June 2019

SUMMARY

Utilization of non-edible biomass as a source of renewable carbon can assist the transition to a sustainable bioeconomy. Such biomass includes non-edible plant materials commonly referred to as lignocellulose. Lignocellulose is a complex material mainly composed of cellulose, a variety of additional polysaccharides collectively referred to as hemicellulose, and lignin, which evolved in plants to confer rigidity and durability against mechanical and biological damage. The resistance to degradation can be overcome by employing specialized biocatalysts (enzymes) produced by microorganisms that exploit the biomass as a source of energy. Biocatalysts acting on the polysaccharides include hydrolytic enzymes (glycoside hydrolases; GHs), such as endoglucanases, cellobiohydrolases and hemicellulases, and oxidative enzymes called lytic polysaccharide monooxygenases (LPMOs). The catalytic modules of the enzymes are often attached to carbohydrate-binding modules (CBMs) that adhere to the substrate. Some lignocellulose-degrading organisms thrive in harsh environments that may resemble desired conditions in industrial bioprocessing, and bioprospecting of such organisms can provide enzymes with traits that are suitable for industrial applications. The performance of enzymes at industrially relevant conditions and other parameters such as substrate preference may be enhanced by enzyme engineering. This thesis describes three studies aimed at finding novel enzymes for processing of lignocellulosic biomass and one study aimed at engineering enzyme properties.

Paper I describes a GH6-CBM2 endoglucanase (mgCel6A) discovered through metagenome mining of a bacterial community isolated from rice straw inoculated with compost and incubated at high temperature. The characteristics of the enzyme did indeed reflect the high temperature of its origin, including an apparent melting temperature (T_m) of approximately 76 °C, optimal activity in overnight reactions at 60 °C, and retention of 90% residual activity after exposure to 65 °C for 24 hours. The crystal structure of the catalytic domain was solved and showed a catalytic cleft consistent with an endoglucanase mode of action. Removal of the CBM2 proved to be beneficial at high substrate concentrations (i.e. industrially relevant conditions), showing that the presence of CBMs needs consideration when optimizing enzymes for industrial use.

Paper II describes a novel hemicellulase, namely a GH10 xylanase (AMOR_GH10A), identified from a metagenome collected from hot sediments near the deep-sea vents of the Arctic Mid-Ocean Ridge. The enzyme was clearly thermostable with an apparent T_m of 85 °C and optimal activity at 80 °C, but was also dependent on NaCl at levels roughly equivalent to seawater. In the absence of salt, thermal stability decreased and catalytic activity was virtually abolished. AMOR_GH10A displayed an unusually broad substrate specificity and high activity towards cellulosic substrates. The enzyme contains an additional domain the function of which could not be predicted. Binding studies with various polysaccharides showed that this domain is a novel type of CBM, and the domain became the founding member of the CBM85 family.

Paper III describes the characterization of an LPMO (mgLPMO10A) derived from a rice straw/compost metagenome and with oxidative activity on cellulosic substrates at high temperatures (at least up to 80 °C) and an apparent T_m of 83 °C. The enzyme displayed a clear synergistic relationship with two cellobiohydrolases and a commercial cellulase cocktail. LPMOs are copper dependent enzymes that can only oxidize the substrate when they coordinate a copper ion in the active site. Interestingly, the experiments showed that, while the non copper-saturated LPMO sample was about 20-fold less active compared to the copper-saturated sample, the two samples provided similar synergistic effects when combined with cellulases. This observation may have implications for optimizing the amount of LPMOs used in industrial bioprocessing of biomass.

Paper IV describes the engineering and characterization of chitin-active mutants of a naturally cellulose-active LPMO. Tuning of the substrate specificity was achieved through the design of a rational mutant library consisting of 4320 mutants that were screened for chitinolytic activity using a high-throughput MS-based method. The chitin-active mutants displayed high activity in the initial phase of chitin-degradation, but they seemed to be more rapidly inactivated compared to naturally chitin-active LPMOs. While the selected mutants showed reduced thermal stability, activity studies led to the conclusion that the low operational stability of the mutants is due to self-inactivation by oxidative damage. This sensitivity to oxidative damage may reflect non-optimized binding to chitin, which could lead to imprecise coordination of the oxidizing reactive oxygen species, thus increasing the risk for non-productive LPMO reactions.

SAMMENDRAG

Anvendelse av ikke-spiselig biomasse som en kilde til fornybart karbon kan bidra til å etablere en bærekraftig bioøkonomi. Disse karbonkildene omfatter plantebasert biomasse som ofte betegnes som lignocellulose. Lignocellulose er et komplekst materiale som hovedsakelig består av cellulose, hemicellulose, og lignin. Planter utviklet disse komponentene for å skaffe seg strukturell styrke og motstandsdyktighet i møte med mekaniske og biologiske trusler. Spesialiserte biokatalysatorer (enzymer) som produseres av mikroorganismer som utnytter denne biomassen som en energikilde kan imidlertid benyttes for å overvinne denne motstandsdyktigheten. Biokatalysatorer som bryter ned strukturelle polysakkarider omfatter hydrolytiske enzymer (glykosid hydrolaser; GHer), slik som endoglukanaser, cellobiohydrolaser og diverse hemicellulaser, og oksidative enzymer kjent som lytiske polysakkarid-monooksygenaser (LPMOer). Det katalytiske domenet av disse enzymene er ofte knyttet til en karbohydratbindende modul (CBM) som binder til substratet. Noen lignocellulose-nedbrytende organismer trives i tøffe miljøer som ligner forholdene i industriell bioprosessering, og enzymer med egenskaper som er særlig egnet for industrielt bruk kan dermed oppdages via bioprospektering av slike organismer. Enzymenes potensiale ved industrielt relevante forhold og på industrielle substrater kan optimaliseres ytterligere ved hjelp av enzym «engineering». Denne avhandlingen inneholder tre studier av lignocellulose-aktive enzymer oppdaget ved hjelp av bioprospektering, samt et studie hvor egenskaper i et enzym ble modifisert ved hjelp av «engineering».

Artikkel I beskriver en GH6-CBM2 endoglukanase (mgCel6A) som ble oppdaget i metagenomiske data hentet fra et bakterielt samfunn isolert fra risstrå inokulert med kompost og inkubert ved høy temperatur. Enzymets egenskaper gjenspeiler miljøet det stammer fra, med blant annet et smeltepunkt (T_m) på ca. 76 °C, optimal aktivitet i overnatt reaksjoner ved 60 °C, samt at det beholdt 90 % aktivitet etter å ha vært eksponert for 65 °C i 24 timer. Krystallstrukturen til det katalytiske domenet ble bestemt og viste en katalytisk kløft som samsvarer med at enzymet virker som en endoglukanase. Å fjerne CBMen viste seg å være fordelaktig ved høye substratkonsentrasjoner (dvs. industrielt relevante forhold), noe som tilsier at tilstedeværelsen av CBMer bør vurderes ved optimalisering av enzymer til industrielt bruk.

Artikkel II beskriver en hemicellulase (en GH10 xylanase; AMOR_GH10A), som ble oppdaget i et metagenom isolert fra varme sedimenter langs den arktiske Midthavsryggen. Enzymet hadde termostabile egenskaper, inkludert en T_m på 85 °C og optimal aktivitet ved 80 °C, men var samtidig avhengig av NaCl i konsentrasjoner på nivå med sjøvann. I fravær av salt ble stabiliteten redusert og katalytisk aktivitet nærmest forsvant. AMOR_GH10A viste en usedvanlig bred substratspesifisitet og høy aktivitet mot cellulose-baserte substrater. Enzymet inneholdt tilsynelatende ytterligere et domene, men dets funksjon kunne ikke predikeres på grunn av lav sekvensidentitet med kjente proteiner. Bindingsstudier med ulike polysakkarider viste at dette domenet er en ny type CBM, og dette domenet danner grunnlaget for CBM85 familien.

Artikkel III beskriver karakteriseringen av en LPMO (mgLPMO10A) hentet fra et risstrå/kompost metagenom og med oksidativ aktivitet på cellulosesubstrater ved høye temperaturer (opp til minst 80 ° C), samt en T_m på 83 °C. Studier med forskjellige enzymkombinasjoner viste tydelige synergistiske effekter når LPMOen ble kombinert med to cellobiohydrolaser eller en kommersiell cellulasecocktail. LPMOer er kobberavhengige enzymer som kun kan oksidere substratet når et kobberion koordineres i det aktive setet, og en ikke-kobbermettet LPMO-prøve var omtrent 20 ganger mindre aktiv enn en kobbermettet LPMO-prøve. Til tross for dette viste eksperimentelle data at ikke-kobbermettet LPMO ga like store synergieffekter med cellulaser som kobbermettet LPMO. Dette er en interessant observasjon som kan ha betydning for optimalisering av mengden LPMO i enzympreparater for bioprosessering.

Artikkel IV beskriver kitinaktive mutanter konstruert med utgangpunkt i en naturlig celluloseaktiv LPMO. Modifisering av substratspesifisitet ble oppnådd ved å lage et rasjonelt mutantbibliotek for det celluloseaktive enzymet bestående av 4320 mutanter, som deretter ble screenet for kitinaktivitet ved bruk av en «high-throughput» metode basert på massespektrometri (MS). Det ble påvist relativ høy initialaktivitet for et utvalg av de kitinaktive mutantene, men disse viste seg å bli raskere inaktivert sammenlignet med en naturlig kitinaktiv LPMO. Mutantene hadde redusert termisk stabilitet, men aktivitetsstudiene med kitin antydet at den lave stabiliteten trolig skyldes selv-inaktivering som følge oksidativ skade. Sensitiviteten mot oksidativ inaktivering kan gjenspeile ikke-optimalisert binding til kitin, ettersom dette kan føre til unøyaktig koordinering av den reaktive oksygenforbindelsen som dannes under katalyse, og dermed øke risikoen for ikke-produktive (selvdestruktive) LPMO reaksjoner.

ABBREVIATIONS

AA	Auxiliary activity
AscA	Ascorbic acid
BG	β-glucosidase
CAZyme	Carbohydrate-active enzyme
CBH	Cellobiohydrolase
CBM	Carbohydrate-binding module
Chi	Chitinase
DM	Dry matter
DP	Degree of polymerization
DSC	Differential scanning calorimetry
e.g.	For example
EG	Endoglucanase
GH	Glycoside hydrolase
Glc	Glucose
GlcNAc	*N*-Acetylglucosamine
ICP-MS	Inductively coupled plasma mass spectrometry
i.e.	That is
LPMO	Lytic polysaccharide monooxygenase
MS	Mass spectrometry
PASC	Phosphoric acid-swollen cellulose
PDB	Protein Data Bank
PUL	Polysaccharide utilization loci
Sc	*Streptomyces coelicolor*
Sm	*Serratia marcescens*
Tf	*Termobifidia fusca*
Tm	Melting temperature

LIST OF PAPERS

Paper I

Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome.
Jensen, M. S., Fredriksen, L., MacKenzie, A. K., Pope, P. B., Leiros, I., Chylenski, P., Williamson, A. K., Christopeit, T., Østby, H., Vaaje-Kolstad, G., Eijsink, V. G. H. (2018). *PLOS ONE*, 13 (5): e0197862.

Paper II

Discovery of a thermostable GH10 xylanase with broad substrate specificity from the Arctic Mid-Ocean Ridge vent system.
Fredriksen, L., Stokke, R., Jensen, M. S., Westereng, B., Jameson, J.-K., Steen, I. H., Eijsink, V. G. H. (2019). *Applied and Environmental Microbiology*, 85 (6): e02970-18.

Paper III

A thermostable bacterial AA10 LPMO with high operational stability in a wide temperature range.
Jensen, M. S., Forsberg, Z., Tuveng, T. R., Fredriksen, L., Vaaje-Kolstad, G., Eijsink, V. G. H.
Manuscript in preparation.

Paper IV

Tuning the substrate specificity of an AA10 LPMO from cellulose to chitin.
Jensen, M. S., Klinkenberg, G., Bissaro, B., Chylenski, P., Vaaje-Kolstad, G., Kvitvang, H. F., Nærdal, G. K., Forsberg, Z., Eijsink, V. G. H.
Submitted to The Journal of Biological Chemistry.

Other publications by the author

The leaderless bacteriocin Enterocin K1 is highly potent against *Enterococcus faecium*: A study on structure, target spectrum and receptor.
Ovchinnikov, K. V., Kristiansen, P. E., Straume, D., Jensen, M. S., Aleksandrzak-Piekarczyk, T., Nes, I. F., Diep, D. B. (2017). *Frontiers in Microbiology*, 8: 774.

Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation.
Petrovic, D. M., Bissaro, B., Chylenski, P., Skaugen, M., Sørlie, M., Jensen, M. S., Aachmann, F. L., Courtade, G., Varnai, A. & Eijsink, V. G. H. (2018). *Protein Science*, 27 (9): 1636-1650.

1 INTRODUCTION

1.1 Biomass

The term biomass encompasses all organic matter associated with living or recently living organisms.

Microfossil records and geochemical traces offer convincing evidence for the presence of structurally complex life on Earth that date back to 3400-3800 million years ago (Lowe and Tice, 2006, Awramik, 1992, Schopf, 1993, Mojzsis et al., 1996), suggesting that the earliest and simplest forms of life came into existence at a time point prior to this. Earth was formed approximately 4500 million years ago, but harsh surface conditions due to the absence of liquid water that first formed around 4300 million years ago (Mojzsis et al., 2001) and the "late heavy bombardment" era of massive meteor impacts that dominated the planet until approximately 3800 million years ago, could easily have hampered the establishment of permanent life (Nisbet and Fowler, 1996, Nisbet and Sleep, 2001, Sleep, 2010). However, recent carbon isotopic measurements indeed suggest that a biosphere harbouring ancient life existed 4100 million years ago (Bell et al., 2015) and putative fossilised microorganisms from seafloor hydrothermal vents could be as old as 4290 million years (Dodd et al., 2017).

The earliest living organisms likely were single-celled anaerobes that inhabited aquatic environments (Martin and Sousa, 2016, Canfield et al., 2006). Some hypothesize that the ancestral organisms were thermophiles originating from deep-sea hydrothermal vents where volcanic activity supplied organic molecules, because such environments were most likely to survive "hot ocean bottlenecks" caused by meteor impacts (Nisbet and Fowler, 1996, Nisbet and Sleep, 2001).

Marine cyanobacterium-like microorganisms had acquired a complex metabolic pathway known as anoxic photosynthesis by 3500 million years ago (Tice and Lowe, 2004, William Schopf, 2011), which gave them the ability to exploit the unlimited energy of sunlight to reduce inorganic compounds into organic compounds (i.e. biomass). Photosynthesis revolutionized life on Earth and the course of evolution when phototrophic microbes gained the ability to use water as an electron donor for reduction of carbon dioxide, an evolutionary event that turned photosynthesis into an oxygenic

process with molecular oxygen as the by-product of biomass formation. In essence, oxygenic photosynthesis employs solar energy and water to fixate carbon dioxide into organic molecules known as carbohydrates, where the initial splitting of water molecules delivers reducing power and simultaneously releases oxygen (Nowicka and Kruk, 2016, Janssen et al., 2014, Barber, 2017). Over hundreds of millions of years, the phototrophs changed the composition of the Earth's atmosphere by accumulating molecular oxygen through photosynthesis (Buick, 2008, Dietrich et al., 2006), which in turn enabled highly efficient aerobic metabolisms to develop and accelerated the evolution of advanced multicellular organisms (Dismukes et al., 2001).

Complex organisms that arose in the wake of these events included the ancestors of land plants, fungi, insects, and crustaceans, which were all exposed to a selection pressure for developing structural components that could provide resilience to mechanical (e.g. the forces of gravity and wind) and biological challenges (e.g. invading organisms) in their environment (Duchesne and Larson, 1989). In multicellular land plants, which arose around 500 million years ago (Morris et al., 2018), the ability to stretch vertically towards the sunlight in the competition for energy provided an evolutionary advantage. This advantage drove the evolution of complex plant cell walls, including cell-wall embedded carbohydrates known as cellulose and hemicellulose (Duchesne and Larson, 1989). In parallel evolutionary events (the oldest known fossil being 505 million years old; Ehrlich et al., 2013), organisms such as fungi and arthropods (e.g. insects and crustaceans; i.e. invertebrates with an exoskeleton) underwent evolutionary processes where a conspicuously similar but nitrogen-containing carbohydrate called chitin was incorporated to provide structural integrity.

1.1.1 Structural carbohydrates embedded in biomass

Carbohydrates, initially recognized as "hydrates of carbon" with the general but not absolute molecular formula of $C_n(H_2O)_n$, are organic compounds also known as saccharides or sugars (Hon, 1994). The basic constituents are referred to as monosaccharides or simple sugars, which attain distinct ring-shaped configurations that usually consist of five carbons (pentoses) or six carbons (hexoses), as exemplified in Fig. 1A. Examples of monosaccharides include glucose, galactose, fructose, ribose, mannose, xylose, and arabinose. Monosaccharides can assemble into disaccharides

such as sucrose, maltose, and lactose (Fig. 1), where two sugar units are joined by a covalent bond known as a glycosidic linkage (Hon, 1994). Glycosidic linkages can occur at various positions of the sugar rings and are named according to which carbons of the individual sugars are adjacent to the glycosidic linkage, and according to the orientation of the linkage (i.e. α or β, which refers to the configuration of the anomeric carbon, C1; see Fig. 1). Importantly, in glucose polymers, the β-orientation generates a much more stable polymer than the α-orientation, as every α-glycosidic bond bends slightly and ultimately generates a helical (spiral-like) structure, whereas the straight β-glycosidic bond allows more stable packing of multiple individual chains with a linear structure (Winger et al., 2009).

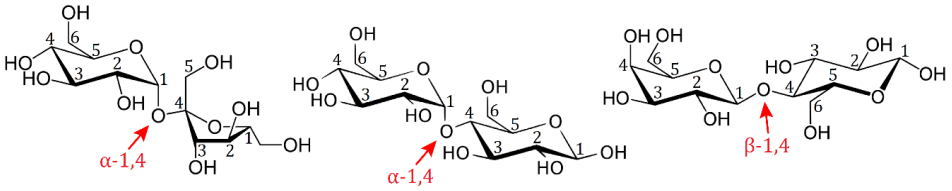

Figure 1. Carbohydrate structures and the glycosidic bond. Three common disaccharides are displayed with numbered carbon atoms to denote which carbons of the sugar rings are adjacent to the glycosidic linkage (the glycosidic oxygen is marked with a red arrow). A) Sucrose: an α-1,4-glycosidic linkage joins a hexose sugar (glucose) and a pentose sugar (fructose). B) Maltose: an α-1,4-glycosidic bond joins two glucose units. C) Lactose: a β-1,4-glycosidic linkage joins two hexose sugars; galactose and glucose. Note that the β-glycosidic linkages entails that neighbouring sugar molecules are rotated 180° relative to one another, a structural trait that is not observed for disaccharides with α-glycosidic linkages.

Polysaccharides (i.e. carbohydrates consisting of up to several thousand monosaccharides) may contain α- and/or β-glycosidic bonds. Starch is a well-known and abundant polymer of α-linked glucose and is one of Nature's primary compounds for energy storage that can be converted to monosugars relatively easily (Merino and Cherry, 2007). Polysaccharides that are based on β-glycosidic bonds are frequently found embedded in recalcitrant biomasses where they serve a structural function in the organism. While the intra-chain glycosidic linkages confer rigidity to individual polysaccharides, the chains of β-linked polysaccharides additionally tend to arrange into larger structures (fibrils) where multiple chains interact with one another through inter-chain hydrogen bonding and van der Waals forces (Brett, 2000, Kubicki

et al., 2018). This phenomenon adds to the complexity and recalcitrance of structural carbohydrate polymers encountered in Nature, and their remarkable resistance to degradation makes them ideal as structural components in living organisms.

1.1.1.1 Cellulose

Cellulose is the major structural component of the plant cell wall, and the most abundant polysaccharide on Earth with an estimated annual production rate of 1500 billion tons (Klemm et al., 2005). Cellulose is a linear polysaccharide comprised of single glucose units rotated 180° relative to each other and connected by β-1,4-glycosidic linkages, making cellobiose (Glc_2) the repeating unit (Cocinero et al., 2009; Fig 2).

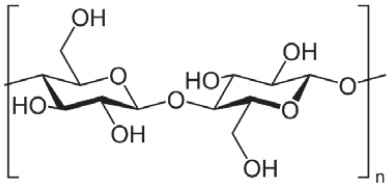

Figure 2. Cellobiose. The repeating unit of cellulose is cellobiose, which consists of two glucose molecules rotated 180° relative to one another and joined by a β-1,4-glycosidic linkage.

The assembly of individual cellulose chains can be amorphous (non-ordered, loosely packed) although cellulose chains more often arrange into crystalline fibrils where multiple cellulose chains are tightly interlinked and stabilized by extensive hydrogen bonding (Figure 3; Somerville et al., 2004). The advantage of the fibril assembly is demonstrated by comparing the estimated half-life of a glycosidic bond, which is approximately five million years (Wolfenden et al., 1998), with the half-life of a cellulose fibril, which is estimated to be one hundred million years (Wilson, 2008). Crystalline cellulose occurs naturally as the two allomorphs I_β and I_α. Both crystal forms have parallel arrangements of the individual cellulose chains, but I_β cellulose is the more stable form as I_α cellulose is irreversibly converted to I_β cellulose when exposed to heat (Gross and Chu, 2010). I_β cellulose is the form commonly found in higher plants. Various methods exist for converting cellulose to other forms, including forms referred to as “amorphous” and alternative crystalline forms such as Cellulose II, where the cellulose chains are organized in an antiparallel fashion. One cellulose form commonly

used in enzyme research is phosphoric-acid swollen cellulose (PASC), which is considered as amorphous or as a low-crystallinity form of cellulose II (Atalla, 1993).

The intrinsic recalcitrance of cellulose fibrils is reinforced in the plant cell wall due to the fibrils being intertwined with other carbohydrate polymers collectively known as hemicelluloses (see section 1.1.1.2 for details) and the complex aromatic polymer lignin. The resulting complex co-polymeric structures are commonly referred to as lignocellulose (Fig. 3). Lignocellulose acts as a barrier to invading organisms while enabling the plant body to grow vertically and bend towards the sunlight. The combined properties of the lignocellulose constituents make plant bodies both rigid and flexible. Different plant species and different plant parts have distinct ratios of each component that determine the appearance and properties of the plant. Generally, in woody plant biomass the ratio is 40-50% cellulose, ~30% hemicellulose and 20-30% lignin (Pauly and Keegstra, 2008). It is worth noting that there is huge variation in the chemical composition of hemicelluloses among plants, as outlined below.

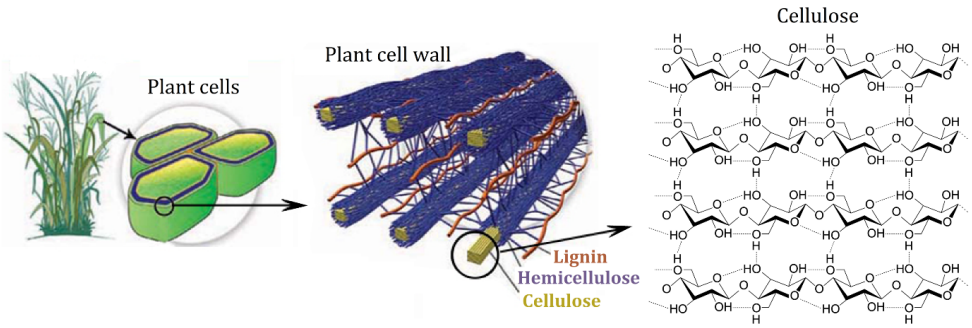

Figure 3. Cellulose fibrils embedded in the lignocellulosic matrix of the plant cell wall. The cell wall is mainly composed of lignocellulose, a complex matrix of cellulose (a glucose polymer where bundles of individual chains order into crystalline fibrils), hemicelluloses (amorphous sugar polymers that differ from cellulose), and lignin (an amorphous and aromatic polymer). The crystallinity of the cellulose fibrils is conferred by extensive inter-chain hydrogen bonding between individual chains. This image was adapted from the Genome Management Information System, Oak Ridge National Laboratory (public.ornl.gov/site/gallery).

1.1.1.2 Hemicellulose

In addition to cellulose, the lignocellulose of the plant cell wall is composed of other carbohydrate polymers known as hemicelluloses. Hemicelluloses constitute a diverse group of amorphous polysaccharides where the β-1,4-linked backbone can encompass different monosaccharides (e.g. glucose, mannose, or xylose) and may have branching sugars connected through a variety of α- and β-glycosidic linkages. Additionally, hemicelluloses frequently have decorating groups (e.g. acetylations) attached to the sugars.

Plants vary in terms of the composition of their hemicellulose fraction. The hemicelluloses xyloglucan, glucuronoarabinoxylan and galactoglucomannan are some of the most abundant hemicelluloses in the cell walls of terrestrial plants and their basic structure is illustrated in Fig. 4.

Xyloglucan is an abundant hemicellulose in land-based plants and consists of β-1,4-linked glucose units as a backbone with branching α-1,6-linked xylose units. Galactose and fucose units may also branch from the xyloses, and the galactose units frequently carry acetylations (i.e., a hydroxyl group (-OH) of the sugar ring has been converted to -O-CO-CH_3).

Glucuronoarabinoxylan consist of β-1,4-linked xylose units that are decorated with α-1,2-linked glucuronosyl units (glucuronic acids) and α-1,3-linked arabinose units. Glucuronoarabinoxylan is one of the hemicelluloses in conifers (i.e. softwoods) and grasses, while a similar form without branching arabinoses (i.e. glucuronoxylan) is abundant in dicots (i.e. hardwoods and flowering plants).

Galactoglucomannan is an abundant hemicellulose in softwoods and has a backbone consisting of β-1,4-linked mannose and β-1,4-linked glucose units at varying ratios. The backbone is decorated with α-1,6-linked galactose residues that are primarily attached to the mannose residues, and is often acetylated (Scheller and Ulvskov, 2010). Fig. 4 shows several other types of hemicelluloses, including mixed linkage β-glucan, glucomannan, and galactomannan.

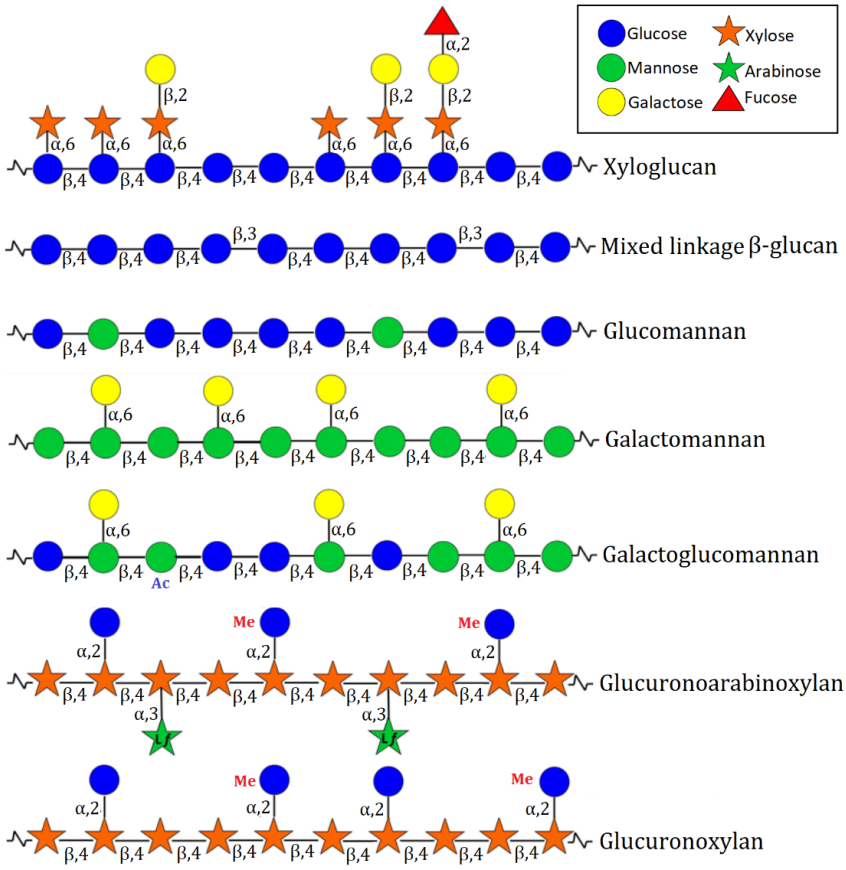

Figure 4. Common hemicelluloses in lignocellulosic biomass. The picture shows xyloglucan, mixed linkage β-glucan, glucomannan, galactomannan, galactoglucomannan, glucuronoarabinoxylan, and glucuronoxylan. The figure was adopted from Glycopedia (https://www.glycopedia.eu/e-chapters/the-plant-cell-walls/article/introduction), and is based on the Symbol Nomenclature for Graphical Representations of Glycans (Varki et al., 2015). Ac, acetylation; Me, methylation; Lf, feruloylated L-arabinose.

1.1.1.3 Chitin

Chitin, a polysaccharide structurally similar to cellulose, represents the second most abundant biomass of the biosphere, with an annual production rate estimated to 100 billion tons (Tharanathan and Kittur, 2003). Chitin is an essential structural component in many organisms and occur, for example, in the fungal cell wall and the protective exoskeletons of arthropods (i.e. organisms such as insects and crustaceans).

The structural similarity of chitin to cellulose lies in the linear homo-polymeric nature of the two polysaccharides, with monosaccharides being rotated 180° relative to each other and joined by β-1,4-glycosidic linkages. Like cellulose, chitin is insoluble and individual chitin chains arrange into recalcitrant fibrils. The difference lies in the monosaccharide composition. Chitin consists of a glucose derivative, *N*-acetyl glucosamine (GlcNAc), that has a nitrogenous group rather than a hydroxyl group at the C2 carbon of the sugar ring, making the disaccharide chitobiose $(GlcNAc)_2$ the repeating unit of chitin (Fig. 5).

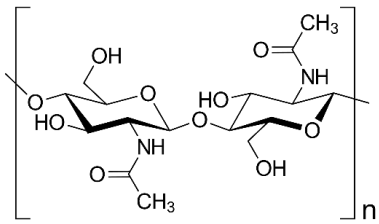

Figure 5. Chitobiose. The repeating unit of chitin is chitobiose, which consists of two molecules of the glucose derivative *N*-acetyl glucosamine that are rotated 180° relative to one another and joined by a β-1,4-glycosidic linkage.

Chitin fibrils are not embedded in a matrix equivalent to the hemicellulose and lignin that surround cellulose fibrils, but rather exist in complex with proteins and minerals such as calcium carbonate (Younes and Rinaudo, 2015). Chitin is crystalline in its native state, and can attain three distinct allomorphs based on the arrangement of individual chitin chains relative to each other (Rinaudo, 2006; Fig. 6). The most recalcitrant chitin allomorph is α-chitin, which is also the most abundant in Nature as it is part of the cell walls of fungi and the exoskeletons of insects and crustaceans. In α-chitin, the individual chitin chains are arranged in an antiparallel fashion and are held together by hydrogen bonding (Minke and Blackwell, 1978). A less abundant form of chitin is the β-chitin allomorph found in squid pens and the spine of some diatoms and tubeworms, in which the individual chitin chains are oriented in parallel to each other (Saito et al., 2000, Gardner and Blackwell, 1975). This arrangement lacks much of the inter-chain hydrogen bonding and is thus more loosely packed than α-chitin, allowing water molecules to enter the structure (Sawada et al., 2012). The third chitin allomorph, where two chains are oriented parallel to each other followed by one anti-parallel chain (Jang et al., 2004),

is γ-chitin, which has been identified in beetle cocoon fibres and in the stomach of certain squids.

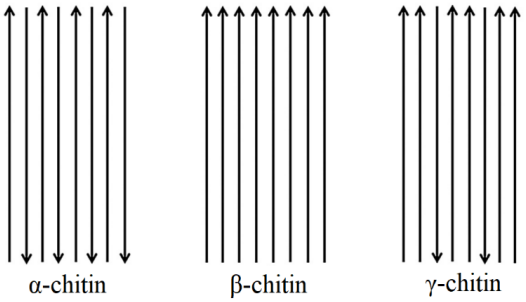

Figure 6. Chain arrangement in chitin allomorphs. Individual chitin chains can arrange into fibrils in three distinct manners. α-chitin is the most abundant in Nature (anti-parallel chains that attain the most impenetrable chitin form), followed by β-chitin (parallel chains, resulting in a more loosely packed structure), and the rare allomorph γ-chitin (pairs of parallel chains interspersed with single anti-parallel chains). The figure was taken from (Anitha et al., 2014).

The tensile strength of cellulose fibrils is more than twice the tensile strength of chitin fibrils (Duchesne and Larson, 1989). This likely made cellulose a more appropriate cell wall polymer in the development of land-based plants since gravitational forces are higher on land compared to aquatic environments, where chitin seems to be more abundantly used. In addition, chitin synthesis requires large amounts of reduced nitrogen, which can be accessed in aquatic environments but is scarce in terrestrial environments. Thus, a nitrogen-dependent cell wall synthesis could be inefficient in terrestrial environments, as it would result in slower growth and less rigid plant bodies (Duchesne and Larson, 1989).

1.2 Biomass degradation in Nature

Recalcitrant polysaccharide-rich biomasses are continuously synthesized in enormous quantities, yet they evidently do not accumulate in the biosphere. As a counteraction to organisms evolving rigid structural components to protect them from mechanical and biological harm, other organisms evolved strategies that enabled them to overcome or even benefit from their presence. The interplay between producers and degraders of

biomass secures the crucial recycling of carbon and nitrogen. Degraders of biomass-associated polysaccharides are particularly prominent in fungi and bacteria, which often exist in symbiotic relationships with higher organisms such as ruminants, insects, and molluscs. Some insects and crustaceans are known to degrade recalcitrant polysaccharides on their own (Wilson, 2008, Cragg et al., 2015, Lynd et al., 2002).

1.2.1 Microbial strategies for biomass degradation

Biomass-degrading microorganisms inhabit vastly different environments including the anaerobic rumen of animals, soil, seawater, cold environments, and hot environments, and are thus dependent on degradation machineries that are specialized not only for handling different substrates, but also for different conditions (e.g. temperature, pH, salinity, oxygen level, redox potential, pressure; Cragg et al., 2015). The key and ultimate step in biomass conversion by microbes is the enzymatic depolymerization of biomass polysaccharides to mono- or oligosaccharides that can be assimilated by the microbe, and Nature thus contains a plethora of enzymes and enzyme systems for this purpose. Additionally, multiple strategies for decreasing the recalcitrance of plant-based biomass have been identified (Cragg et al., 2015), including mechanical disruption in the animal rumen, oxidative attacks by brown rot fungi through Fenton chemistry (Arantes and Goodel, 2014), and deployment of oxidative enzymes (e.g. peroxidases and laccases) that modify or remove lignin by white rot fungi and certain ligninolytic bacteria (Pollegioni et al., 2015, Bugg et al., 2011).

Key enzymatic strategies for degradation of recalcitrant biomass include the production of free enzymes secreted into the environment, the use of multi-enzyme complexes called cellulosomes that are attached to the microbial cell wall, the orchestrated production of dedicated degradative enzymes and transporters, encoded by so-called polysaccharide utilization loci that are attached to the inner and outer membrane, and the secretion of enzyme cocktails packed in secreted outer membrane vesicles.

1.2.1.1 Free enzymes

Aerobic bacteria and fungi express and secrete enzymes into their environment upon exposure to a polysaccharide carbon source. The secretion of enzymes to outside the cell occurs because the organisms are unable to transport insoluble material across the

cell membrane, and thus the polysaccharides must be depolymerized before the basic components can be absorbed for metabolism. Enzyme production is an energy-consuming process, and organisms using the strategy of producing free enzymes risk that the sugars released by their enzymes are consumed by competing organisms. The free enzymes strategy is employed by both cellulolytic and chitinolytic organisms and has been especially well studied in the cellulolytic fungus *Trichoderma reesei* (Bischof et al., 2016), the cellulolytic bacterium *Thermobfidia fusca* (Gomez del Pulgar and Saadeddin, 2014), and the chitinolytic bacterium *Serratia marsecens* (Vaaje-Kolstad et al., 2013).

The secreted enzymes are specialized towards targeting specific substrates (e.g. cellulose, hemicellulose, or chitin), and have synergistic modes of action where they attack different parts of the substrate to secure efficient turnover. The free enzymes operate independently of each other outside the producing cell and are dependent on locating and binding to the target substrate before depolymerization can be initiated. For this purpose, the enzymes often contain dedicated binding modules that direct the catalytic module to the proper substrate. Figure 7 shows an overview of a typical free enzyme system for degradation of chitin or cellulose. The general view is that endo-acting hydrolytic enzymes and oxidative enzymes generate new chain ends within the most crystalline parts of the substrate, while hydrolytic exo-acting enzymes attach to the chain ends and release disaccharides from the substrate. Endo-acting hydrolytic enzymes attack randomly in amorphous regions of the substrate and are not limited to the chain ends for production of disaccharides or other oligomeric products. Dedicated hydrolytic enzymes called β-glucosidases are finally responsible for converting the solubilized oligomeric sugars (mainly dimers) to monomers (Horn et al., 2012). A detailed description of different enzyme activities relevant to free enzyme systems in general and the work described in this thesis in particular is provided in section 1.3.

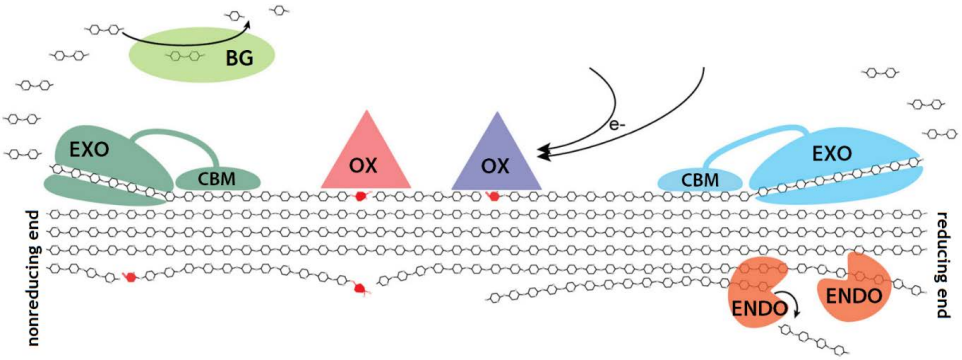

Figure 7. Free enzyme system for aerobic degradation of structural polysaccharides. The picture shows a schematic overview of a cellulolytic machinery attacking a cellulose fibril using independent enzymes secreted from the host cell. Oxidative enzymes (OX) generate new chain ends within the crystalline parts of the polysaccharide to loosen the structure and facilitate easier access for endo-acting enzymes (ENDO) that attack randomly in amorphous cellulose regions and exo-acting enzymes (EXO) that attach to the chain ends and processively release disaccharides. The solubilized sugars are finally degraded to monomers by a β-glucosidase (BG). Note that similar enzymes with corresponding activities are used to degrade chitin. The figure was adapted from (Horn et al., 2012).

1.2.1.2 Cellulosomes

The use of free enzymes secreted to the environment is not a common strategy among microorganisms that inhabit anaerobic environments, which have rather developed alternative strategies where the organism remains attached to the enzymes and thus in proximity of any solubilized sugars. A reason for this could be that microbes living in anaerobic environments need to limit their energy use (e.g. production of enzymes), compared to aerobic producers (Bayer et al., 2004).

The most common strategy employed by anaerobes entails the use of cellulosomes and has only been observed in connection to conversion of celluloses and never in connection to chitin degradation. Distinct for these anaerobic cellulolytic bacteria is that the cellulolytic enzymes are kept attached to a molecular scaffold called a scaffoldin subunit that is bound to the bacterial surface, which enables the enzymes to operate as a large multi-enzyme complex that degrades the carbohydrates in proximity of the producing organism (Bayer et al., 2004). The different enzymes (i.e. the catalytic

modules of the cellulosome) are attached to the scaffolding subunit through cohesin modules (Fig. 8). The scaffoldin subunit additionally holds a binding module that enables the cellulosome (and thus the host cell) to adhere to the appropriate substrate. These enzyme systems were first described in the anaerobic thermophilic soil bacterium *Clostridium thermocellum* (Lamed et al., 1983, Bayer et al., 2008). Since then, a wide diversity of cellulosome structures has been unravelled, including cellulosomes containing hemicellulases, next to cellulases (Artzi et al., 2017). The proximity of multiple catalytic modules that execute different tasks in the degradation process promotes synergistic interplay that makes polysaccharide turnover more efficient. On the other hand, it is likely that, compared to free enzymes, such a large protein complex is inefficient in terms of finding and interacting with suitable access points in the substrate.

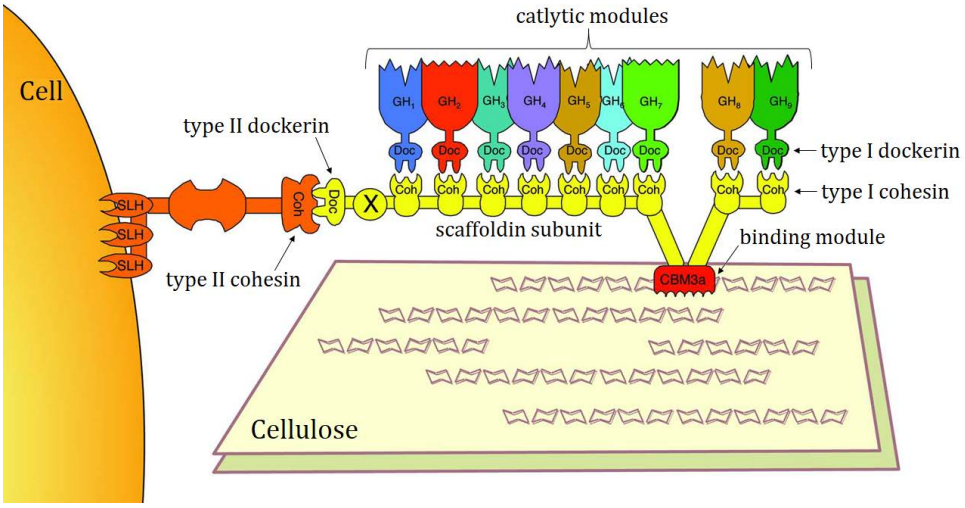

Figure 8. Arrangement of a cellulosome attached to the cell wall of an anaerobic cellulolytic bacterium. A type II cohesin is directly attached to the cell wall of the producer, and a scaffoldin subunit containing a type II dockerin is anchored to the type II cohesin module. The scaffoldin contains a dedicated carbohydrate-binding module (CBM) to facilitate adhesion of the cell to the proper substrate, and type I cohesins act as anchoring points for various catalytic modules that hold different roles in cellulose/hemicellulose degradation. These catalytic modules are anchored to the type I cohesins through type I dockerins. (Yaniv et al., 2014).

1.2.1.3 Polysaccharide utilization loci

Another strategy adopted by organisms that inhabit anaerobic environments, predominantly bacteria that have been identified in highly anaerobic gut environments, is the use of polysaccharide utilization loci (PULs). PULs contain co-localized and co-regulated genes that encode all the enzymes and proteins the microbe needs to digest a specific polysaccharide (Fig. 9A). Once expressed, the enzymes are arranged onto the outer and inner membrane of the host, where the enzymes attached to the outer membrane convert the substrate to soluble oligomers that can be transported to the periplasm, where the remaining enzymes complete the conversion to monomeric sugars (Fig. 9B). PULs were first identified during studies of starch conversion by *Bacteroidetes thetaiotaomicron* in the human gut (Bjursell et al., 2006) and have since been found in many Bacteroidetes species specialized towards a variety of substrates, such as hemicelluloses (Larsbrink et al., 2014, Cuskin et al., 2015), cellulose (Naas et al., 2014, Mackenzie et al., 2015), and chitin (Larsbrink et al., 2016).

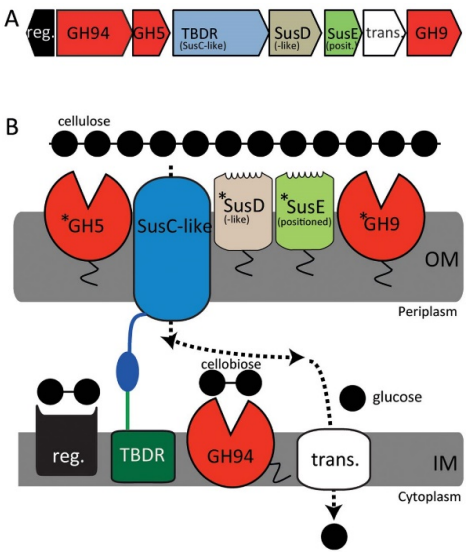

Figure 9. A cellulose-targeting polysaccharide Utilization Locus (PUL) of a *Bacteroidetes* bacterium present in the cow rumen. Panel A displays the gene cluster (locus) and panel B shows the arrangement of the various gene products on the outer membrane (OM) and inner membrane (IM) of the cell. Insoluble cellulose is depolymerized by enzymes (GH5 & GH9) on the outer membrane and transported to the periplasm (involving the Sus proteins) where enzymes attached to the inner membrane (GH94) convert the solubilized components to monomers, which are finally transported to the cytoplasm. The figure was taken from (Naas et al., 2014).

1.2.1.4 Outer membrane vesicles

Bacteria may employ another strategy for cellulose-degradation that entails secretion of membranous compartments called outer membrane vesicles (OMVs; Figure 10). OMVs contain various cellular components and periplasmic proteins that can improve the invasive abilities of the bacterium (Jan, 2017). OMVs are spherical structures made from the outer membrane of Gram-negative bacterial host cells (Beveridge, 1999), and potentially function as long distance delivery tools (Kulp and Kuehn, 2010). This strategy has been described for *Fibrobacter succinogenes* (one of the major inhabitants of the herbivore rumen), which releases membranous fragments that can degrade cellulosic substrates (Groleau and Forsberg, 1981, Forsberg et al., 1981), and these OMVs have later been shown to be enriched with carbohydrate-active enzymes (Arntzen et al., 2017). It is still debated whether the release of OMVs is simply a consequence of aging cells and passive membrane disintegration, or reflects a true biological function. OMVs equipped with hydrolytic enzymes have also been identified in Bacteroides species (Elhenawy et al., 2014).

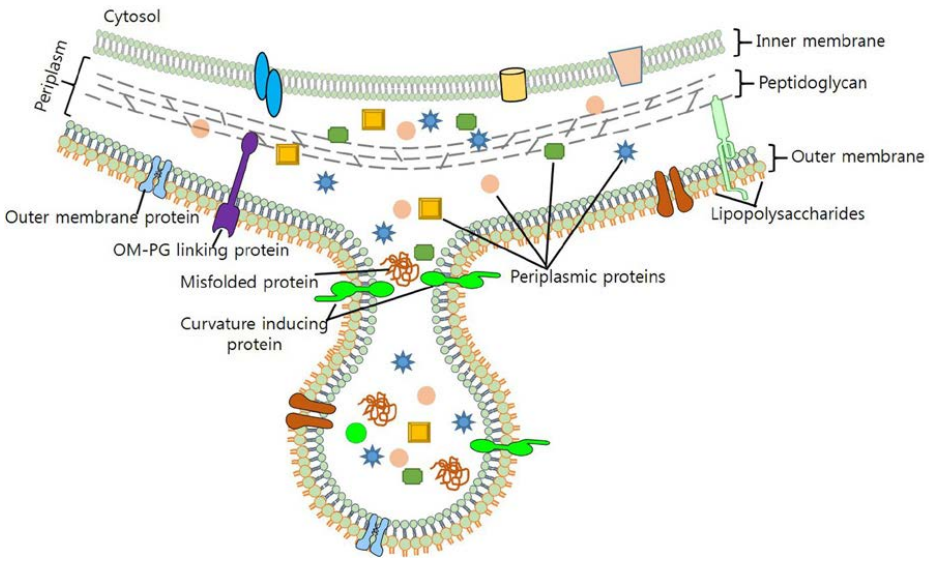

Figure 10. Secretion of enzymes in outer membrane vesicles (OMVs). Periplasmic enzymes and other cell components are incorporated into an OMV through a budding event where part of the outer membrane will eventually separate from the host cell to form an independent spherical entity. The enzyme-enriched OMV is subsequently deployed into the extracellular environment. The figure was taken from (Jan, 2017).

1.3 Biomass-active enzymes

The microbial strategies for cellulose, hemicellulose and chitin decomposition discussed in section 1.2 are based on carbohydrate-active enzymes (CAZymes) with various modes of action and different roles in the degradation process. CAZymes are classified and grouped into protein families in the CAZy database (www.cazy.org), which is based on sequence similarities that mainly reflect structural features and not necessarily substrate specificity (Lombard et al., 2014, Cantarel et al., 2009). The CAZy database currently comprises five classes of CAZymes and one class of non-catalytic binding modules that are frequently associated with the CAZymes, and together these six classes comprise more than 400 protein families.

The protein class of glycosyltransferases (GTs) holds 107 families (as of June 2019) and comprises enzymes that catalyze biosynthesis (rather than biodegradation) of oligo- and polysaccharides by catalyzing the formation of glycosidic bonds. GTs employ an activated donor carbohydrate (i.e. a sugar molecule with a phosphate-containing leaving group) and an acceptor molecule that is most commonly another carbohydrate, but may also be one of several other biomolecules such as lipids, proteins, or nucleic acids (Lairson et al., 2008).

Carbohydrate esterases (CEs) currently comprise 16 protein families that remove ester-based decorations (e.g. acetylations) from sugars. These decorations may protect the plant cell wall from degradation by preventing enzymes from accessing and cleaving the glycosidic bonds. CEs thus assist the other CAZymes involved in polysaccharide depolymerisation by removing these obstacles (Biely, 2012, Sista Kameshwar and Qin, 2018).

The class of polysaccharide lyases (PLs) consists of 37 families that cleave glycosidic bonds in uronic acid-containing polysaccharides [i.e., sugars where the C_6 carbon (CH_2OH) has been oxidized to a carboxylic acid (COOH), making the polysaccharide acidic]. PLs employ a β-elimination mechanism that differs from the hydrolytic mechanisms commonly used by other CAZymes that cleave glycosidic bonds (Sutherland, 1995).

This thesis includes experimental work with members of the three remaining protein classes of the CAZy database, namely the glycoside hydrolases (GHs), the non-catalytic

carbohydrate binding modules (CBMs), and oxidative enzymes classified as auxiliary activities (AAs) and called lytic polysaccharide monooxygenases (LPMOs). These will be discussed in depth in section 1.3.1-1.3.3.

1.3.1 Glycoside hydrolases (GHs)

Glycoside hydrolases (GHs), also known as glycosyl hydrolases, are enzymes that depolymerize carbohydrates by using a hydrolytic mechanism to cleave the glycosidic bond between sugar moieties (Davies and Henrissat, 1995). There are currently 165 GH families in the CAZy database. GHs target various substrates such as amorphous or crystalline cellulose, oligosaccharides, hemicelluloses, or chitin. Many GH families display a relatively wide substrate specificity, while others seem to be specialized towards a single substrate. GHs are generally named according to substrate specificity (e.g. cellulase, xylanase, chitinase, etc.) and their mode of action (i.e. whether they cleave the substrate randomly (endo-acting) or attack from the chain ends (exo-acting). Additionally, GHs include enzymes that convert dimers and oligomers to their monomeric constituents (e.g. β-glucosidases for cellobiose and chitobiases for chitobiose).

Structure and mode of action

The mode of action of GHs is reflected in the enzyme structure (Fig. 11). Tunnel-shaped active sites are characteristic for exo-acting enzymes that thread the polymer while making multiple catalytic cuts and releasing dimers, in what is referred to as processive action (Fig 11A). Open-cleft active sites are present in endo-acting enzymes that attach to random cleavage sites along the polymer and that likely detach from the substrate in between catalytic steps (Fig. 11B). Exo-acting GHs tend to have a preference for attacking the polymer from either the reducing chain end or non-reducing chain end, and differ in terms of the degree of processivity (i.e. how many catalytic cuts the enzyme generates before detaching from the substrate). β-glucosidases and chitobiases are recognized by a pocket-like active site (Fig. 11C) where dimers and soluble oligomers are converted to monomers (Davies and Henrissat, 1995).

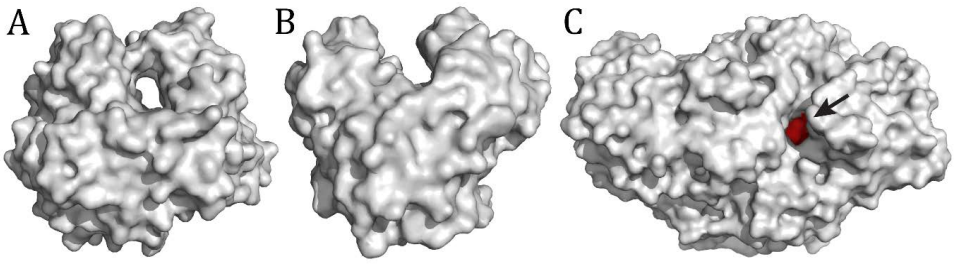

Figure 11. Topology of glycoside hydrolases (GHs). A) Surface representation of the crystal structure of a cellobiohydrolase (exo-acting GH) from *Trichoderma reesei* (PDB ID: 1CB2; Koivula et al., 1996) showing a tunnel-shaped active site for threading individual cellulose chains while making consecutive catalytic cuts. B) An endoglucanase from *Thermobifidia fusca* (PDB ID: 1TML; Spezio et al., 1993) showing a catalytic cleft that allows the enzyme to attach and detach from the cellulose chain at random cleavage sites. C) A β-glucosidase from *Trichoderma reesei* (PDB ID: 4I8D; Karkehabadi et al., 2014) where cellobiose is converted to glucose. The structure has a shallow pocket colored in red and marked with an arrow. The figure was made using Pymol.

Mechanism

GHs attack the glycosidic bond of carbohydrates after using binding energy to force the substrate (sugar chain) into a conformation where the scissile bond is weakened (substrate distortion), which lowers the energy barrier of the hydrolytic reaction (Davies and Henrissat, 1995, Speciale et al., 2014, Vocadlo and Davies, 2008). A hydrolytic attack is then initiated, which involves two essential amino acids located on the enzyme surface and in proximity of the scissile bond. The cleavage results in either retention or inversion of the stereochemistry of the hydroxyl group of the anomeric center (i.e. the C1 carbon that was previously adjacent to the glycosidic bond).

The inverting mechanism typically relies on a catalytic acid and a catalytic base located approximately 10 Å apart (Davies and Henrissat, 1995). Glycosidic bond cleavage follows a single displacement mechanism that passes through an oxocarbenium ion-like transition state (Fig. 12). The catalytic acid promotes leaving group departure by protonating the glycosidic oxygen, while the catalytic base positions and activates a water molecule to carry out a nucleophilic attack on the anomeric (C1) carbon. This results in cleavage of the glycosidic bond where the configuration of the hydroxyl group on the anomeric carbon attains opposite stereochemistry, relative to the starting

situation (Sinnott, 1990, Davies and Henrissat, 1995, Rye and Withers, 2000, Vuong and Wilson, 2010).

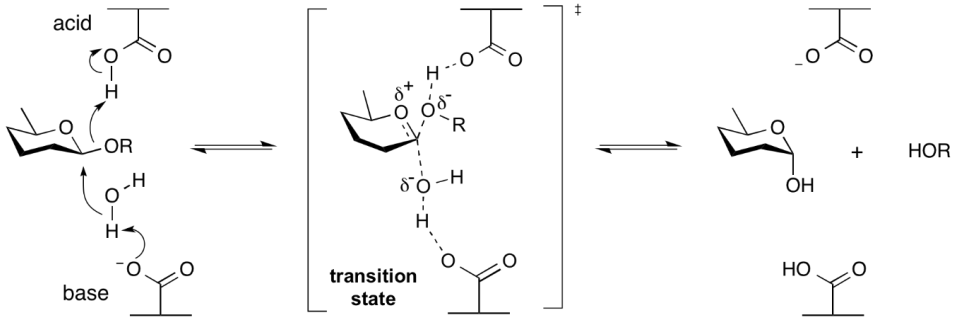

Figure 12. Inverting mechanism for a glycoside hydrolase acting on a cellulosic substrate. The configuration of the hydroxyl group on the anomeric carbon (i.e. the C1 carbon adjacent to the glycosidic oxygen) is inverted when the glycosidic bond is cleaved. The catalytic acid donates a proton to the glycosidic oxygen while the catalytic base activates a water molecule to carry out a nucleophilic attack on the anomeric carbon. Consequently, an OH from this water molecule becomes attached to the anomeric carbon, while the glycosidic oxygen and the proton donated by the catalytic acid are linked to the C4 carbon of the leaving group (HOR). This image was taken from cazypedia.com (Consortium, 2017).

In GHs that employ a retaining mechanism, a general base/acid catalyst and a nucleophile are located approximately 5.5 Å apart. The distance is less than in inverting enzymes because retaining enzymes do not need to accommodate a water molecule in addition to the sugar chain (Davies and Henrissat, 1995). The retaining mechanism rather occurs via a double displacement mechanism where each step passes through an oxocarbenium ion-like transition state. The general acid/base catalyst first acts as an acid by donating a proton to the glycosidic oxygen while the nucleophile simultaneously forms a covalent intermediate with the anomeric carbon as a result of nucleophilic attack (Fig. 13). The leaving group (ROH; i.e., the sugar that does not participate in the covalent intermediate with the nucleophile) is released. The deprotonated acid/base catalyst will in the next step act as a general base where it activates a water molecule to carry out a nucleophilic attack on the sugar-nucleophile (i.e. glycosyl enzyme) intermediate, which leads to retention of the stereochemistry at the anomeric carbon.

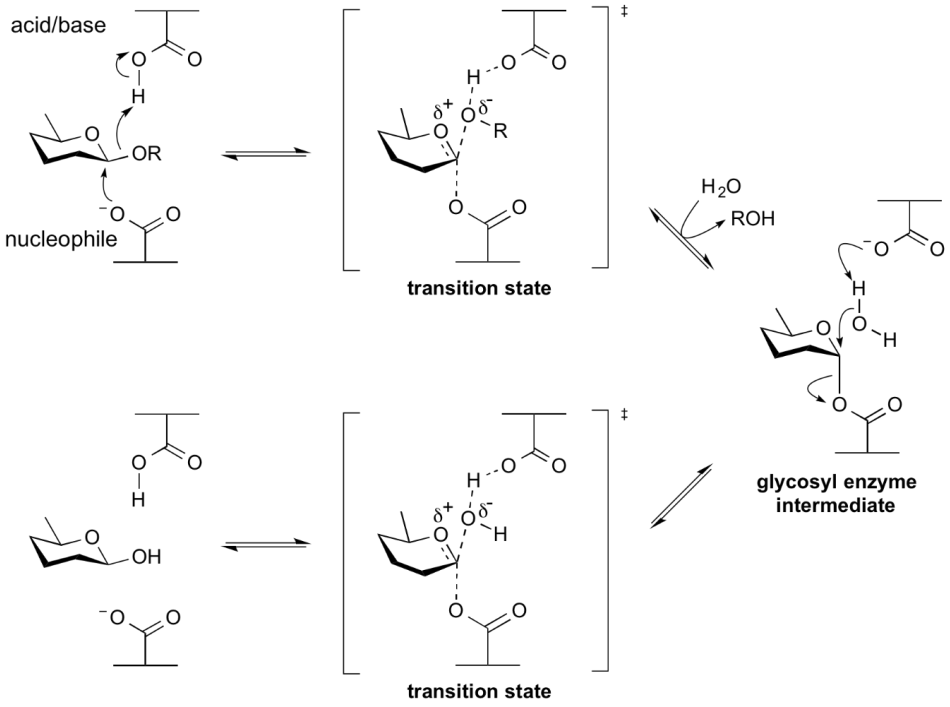

Fig 13. Retaining mechanism for a glycoside hydrolase acting on a cellulosic substrate. The configuration of the hydroxyl group on the anomeric carbon (i.e. the C1 carbon adjacent to the glycosidic oxygen) is retained when the glycosidic bond is cleaved. In the first step of the double displacement mechanism, the acid/base acts as an acid and donates a proton to the glycosidic oxygen, while the nucleophile simultaneously forms a covalent intermediate with the anomeric sugar carbon (glycosyl-enzyme intermediate). The leaving group (ROH) can then be released, followed by the acid/base acting as a base in the second step of the mechanism, where it activates a water molecule that goes on to initiate a nucleophilic attack on the anomeric carbon, thereby releasing the sugar from the glycosyl enzyme intermediate. Image from cazypedia.com (Consortium, 2017).

GHs involved in the degradation of lignocellulose

The enzymatic depolymerization of lignocellulosic materials involves hydrolytic enzymes including hemicellulases such as xylanases, mannanases, and xyloglucanases that degrade and remove hemicelluloses that cover the cellulose fibrils, making the cellulose more susceptible to attack by cellulases. The subsequent turnover of cellulose is generally considered to depend on endocellulases (also known as endoglucanases,

e.g., members of GH5, GH6 and GH9) that attack randomly in amorphous areas of the cellulose fibrils to loosen the structure further and generate new chain ends. The chain ends are attacked from either the reducing end or non-reducing end by exocellulases (also known as cellobiohydrolases, e.g. GH6s from the non-reducing end and GH7s or GH48s from the reducing end), which release dimers (cellobiose), usually in a processive manner. The released dimers and oligomers are finally degraded to the monomeric end-product (glucose) by β-glucosidases (e.g. GH3s), that are responsible for cleaving carbohydrates with a low degree of polymerization (DP) (Horn et al., 2012). Considering the distinct activities of these various enzymes, it is not surprising that they show synergistic relationships when they are combined. These synergistic relationships are not only based on one enzyme type generating new access points for another enzyme type , but also on the removal of obstacles that hinder processive action by other enzymes (Jalak et al., 2012, Igarashi et al., 2011).

A thermostable GH6 endoglucanase is the topic of the first paper of this thesis, while paper II describes investigations of a thermostable GH10 xylanase.

The GH6 family includes both endoglucanases and cellobiohydrolases of bacterial or fungal origin. GH6 enzymes employ an inverting mechanisms to hydrolyze glycosidic bonds, but the catalytic base that participates in bond cleavage remains elusive within this family (Payne et al., 2015, Mayes et al., 2016). Structurally characterized GH6 endoglucanases display an open catalytic cleft that differ from the tunnel-shaped catalytic sites of the GH6 cellobiohydrolases. Shortening and displacement of two surface loops in the endoglucanases relative to the corresponding loops in the cellobiohydrolases are believed to be the reason for this structural difference, as the elongated loops of the cellobiohydrolases fold over the active site cleft to form a tunnel (Meinke et al., 1995).

The GH10 family represents one of the major xylanase families together with the GH11 family, and most members are of bacterial or fungal origin. Xylanases cleave xylan (i.e. backbones consisting of β-1,4-linked xylose units), which is the second most abundant polysaccharide in plant biomass after cellulose (Biely et al., 2016). The GH10 family holds endo-xylanases that differ from GH11 xylanases in that they tolerate higher levels of substitutions (Collins et al., 2005, Chakdar et al., 2016). Some GH10 xylanases exhibit a weak side activity toward cellulose, which is believed to be related to specific

structural features that distinguish cellulose-active xylanases from non-cellulose-active xylanases (Notenboom et al., 1998). This is discussed in more detail in Paper II, which describes a GH10 xylanase with remarkably high activity on glucans.

GHs involved in the degradation of chitin

The degradation of chitin depends on fewer enzyme activities compared to lignocellulose, which reflects the lower relative complexity of the substrate. Dedicated chitinases (GH18s and GH19s) carry out tasks that are equivalent to the tasks accomplished by cellulases, with endo-acting chitinases loosening the crystalline structure and generating new chain ends, while exo-acting chitinases can attach to either the reducing or non-reducing chain ends (Igarashi et al., 2014). The exo-acting chitinases cleave dimers (chitobiose) from the chain ends in a processive manner (Horn et al., 2006). Chitobiases (GH20s) finally degrade the dimers to the monomeric compound *N*-acetyl glucosamine . While GH18s employ a retaining mechanism where the anomeric configuration is conserved, notably through a unique substrate-assisted mechanism, GH19s employ an inverting mechanism (Eijsink et al., 2010).

1.3.2 Carbohydrate binding modules (CBMs)

The CAZy database also contains non-catalytic modules that interact with the substrate without cleaving it. These are called carbohydrate-binding modules (CBMs). To date, 85 CBM families have been deposited in the CAZy database, including CBM85 that originates from the studies presented in Paper II of this thesis.

CBMs display a variety of substrate preferences, ranging from cellulosic substrates to hemicelluloses and other polysaccharides such as starch and chitin. Some also bind to non-carbohydrate components that are present in biomass, such as lignin (e.g. Strobel et al., 2015). CBMs may exhibit substrate promiscuity that allows them to bind to several different substrates and not be hindered by e.g. decorating groups on the substrate (Boraston et al., 2004). The structures of CBMs show that they commonly consist of beta-sheets and have an abundance of aromatic amino acids on their surfaces that interact with the substrate. The substrate-binding surface can be either flat or contain a groove or sometimes even a pocket (Fig. 14).

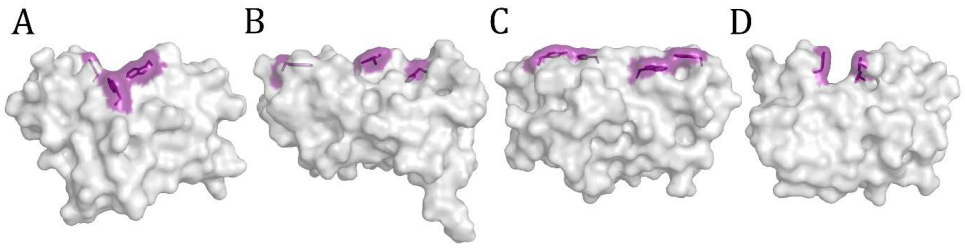

Figure 14. Topology of representatives of the families CBM1-4. A) CBM1 with a shallow groove, PDB ID: 4QPW. B) CBM2 with a flat substrate-binding surface, PDB ID: 6F7E. C) CBM3 with a flat substrate-binding surface, PDB ID: 1NBC. D) CBM4 with a pocket, PDB ID: 1GU3. The substrate-binding surface is facing upwards, and surface-located aromatic residues putatively involved in substrate-binding are colored in purple and shown as sticks. The figure was made using Pymol.

CBMs are attached to GHs through linkers of varying length, ranging from a few amino acids to over one hundred. The linkers are often rich in proline, threonine and serine residues, where proline and serine may be glycosylated, which grants the linker with protease resistance (Sammond et al., 2012, Poon et al., 2007, Payne et al., 2013). Many of these linkers are extended and flexible (e.g. Courtade et al., 2018), but exceptions exist (e.g. van Aalten et al., 2000). CBMs function by adhering to substrates, thus placing the catalytic module in proximity of substrate binding sites (Fig. 15; Boraston et al., 2004). This creates a proximity effect on carbohydrate turnover as the CBMs increase the effective enzyme (catalytic module) concentration near the substrate. Such proximity effects are particularly important in systems with low substrate concentration. When the substrate is scarce, the CBMs greatly enhance the degradation rate, but studies of fungal enzymes containing CBM1s have shown that this effect is abolished when the substrate concentration is increased (Varnai et al., 2013). The substrate concentration will eventually reach a level where the catalytic modules are as likely to encounter substrate binding sites without the assistance of the CBM, and at that point the CBM will rather hamper the degradation efficiency as it decreases the desorption rate from the substrate. Similar conclusions about the substrate-dependency of the CBM effect were reached in the study of a bacterial GH6 described in Paper I of this thesis and in a recent study of a CBM-containing LPMO (Courtade et al., 2018; Fig 15B).

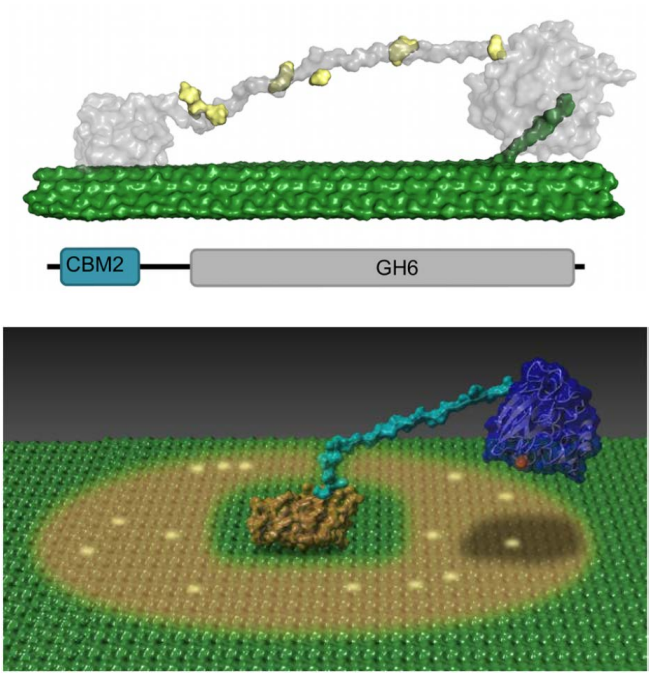

Figure 15. Carbohydrate binding modules (CBMs). The upper panel shows a CBM2 anchored to the surface of a cellulose fibril and attached to a catalytic module (a GH6 cellulase) through a flexible linker. The linker is glycosylated and the attached sugars are shown in yellow. The figure was taken from (Sammond et al., 2012). The lower panel shows an oxidative enzyme (LPMO; shown in blue) docked on the surface of crystalline cellulose through a CBM2 (orange) and a linker (cyan) that allows the catalytic module to move around while being kept in proximity of substrate binding sites. The CBM enables the catalytic LPMO domain to make random cuts on multiple cellulose chains in proximity of the anchoring point (within the area marked in brown). The figure was taken from (Courtade et al., 2018).

Many catalytic modules do not have a CBM attached and must locate and bind the substrate on their own, which, as noted above, can be challenging if the substrate concentration is low. Obviously, the catalytic domains themselves also have the ability to interact with and bind to the substrate, and the extent to which a CBM contributes likely varies. Importantly, when it comes to CBMs, there are many variations in the world of CAZymes. Some catalytic modules may have several different CBMs attached, that potentially can bind different substrates. It has also been pointed out that the so-called “proximity” effect (i.e., when a CBM brings its appended catalytic domain close to

the substrate), may be indirect, in the sense that the CBM binds to a non-substrate polysaccharide in the plant cell wall. For example, Hervé et al. 2010 showed that the capacity of a xylanase to degrade xylan in plant cells could be augmented by appending a cellulose-binding CBM.

Some studies indicate that certain CBMs have a disruptive effect on the substrate, for example meaning that they open up crystalline structures to make them more accessible to the action of cellulases (Hall et al., 2011, Eijsink et al., 2008, Bernardes et al., 2019). Such CBMs, and other potentially disruptive proteins with no known catalytic function, such as swollenins and expansins (Eijsink et al., 2008), may perhaps be explored in developing more efficient cellulase cocktails. However, recent work in the field has been mostly focusing on the substrate-disruptive action of the lytic polysaccharide monooxygenases (LPMOs), which are discussed below.

1.3.3 Auxiliary activities (AAs)

The most recent enzyme class added to the CAZy database comprises the auxiliary activities (AAs) and currently holds 16 families. AAs assist other CAZymes in the degradation of biomass and include members that are not active on carbohydrates, but act or non-carbohydrates such as lignin, which needs to be removed or modified before other CAZymes (such as GHs) can efficiently carry out depolymerisation. The common characteristic of AAs is that they are redox enzymes that act in conjunction with other CAZymes (Levasseur et al., 2013). The AA protein families comprise various oxidases (e.g. laccases), peroxidases, dehydrogenases (e.g. cellobiose dehydrogenases), reductases, and lytic polysaccharide monooxygenases (LPMOs). LPMOs, which make up families AA9, AA10, AA11, AA13, AA14, AA15 and AA16, will be discussed in detail below.

1.3.3.1 Lytic polysaccharide monooxygenases (LPMOs)

The hydrolytic enzymes involved in biomass degradation have been known for over 100 years. The limited ability of these enzymes to efficiently degrade crystalline substrates has led to suggestions that supplementary strategies for enabling easier access to individual polysaccharide chains are involved in the biomass degradation process. In

1950, Reese and colleagues hypothesized that the conversion of native cellulose to soluble sugars must depend on at least two enzyme systems (Fig. 16), where the initial system (C_1) functions as a sort of pre-treatment that allows the second enzyme system (C_x) to carry out complete solubilisation of the polysaccharides (Reese et al., 1950). In 1974, Eriksson and co-workers suggested that an oxidative mechanism likely is involved in cellulose depolymerisation. They based this suggestion on the observation of a significant improvement in cellulose turnover by a culture supernatant of a cellulose-degrading fungus in the presence of oxygen compared to under anaerobic conditions. Eriksson et al. also observed that oxygen was consumed upon adding a cellulosic substrate to such culture supernatants (Eriksson et al., 1974).

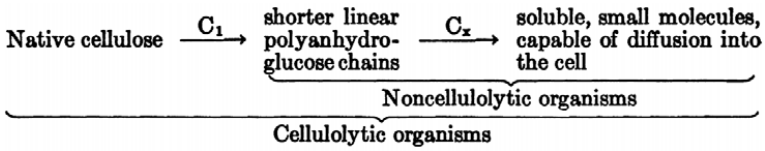

Figure 16. The C_1C_x hypothesis for degradation of native cellulose. The first enzyme system (C_1) fragments the crystalline cellulose and acts as a sort of pre-treatment that is essential for enabling the second enzyme system (C_x) to efficiently convert the cellulose to soluble sugars that can diffuse into the cell. The figure was taken from (Reese et al., 1950).

The question of whether a C_1-type enzyme system and/or oxidative activity was involved in polysaccharide degradation remained unsolved for decades, until the recent discovery that CAZymes previously believed to be CBM33s or endoglucanases with very weak hydrolytic activity (GH61s) are in fact enzymes that cleave polysaccharides using an oxidative mode of action (Vaaje-Kolstad et al., 2010).

Early studies on members of the CBM33 family supported a function as CBMs, since these proteins were found to bind to chitin and cellulose and showed a high content of conserved aromatic residues putatively involved in substrate binding. Surprisingly, structural data on CBP21, a single domain CBM33 from the chitinolytic bacterium *Serratia marcescens*, showed that the aromatic residues are not located on the surface but rather buried in the core of the protein (Vaaje-Kolstad et al., 2005a). Furthermore, in what in retrospect is a landmark study, studies of chitin degradation demonstrated that the presence of CBP21 dramatically enhanced polysaccharide turnover by GHs

(Vaaje-Kolstad et al., 2005b). The first structure of a GH61 (Karkehabadi et al., 2008) showed that this putative endoglucanase lacked the typical catalytic features of a GH and revealed a remarkable structural resemblance with CBM33s, suggesting that the GH61s were not really cellulases and that CBM33s and GH61s have similar functions. Indeed, in 2010, Harris et al (Harris et al., 2010) showed that GH61 proteins boost the activity of cellulases, just like CBM33 proteins boost the activity of chitinases.

The oxidative activity of these (until then) enigmatic CBM33 and GH61 proteins was for the first time demonstrated for CBP21 on chitin, and oxidative activity on cellulose was shortly after demonstrated for another CBM33 (Forsberg et al., 2011) and several GH61s (Quinlan et al., 2011, Phillips et al., 2011). These two protein families were renamed as lytic polysaccharide monooxygenases (LPMOs) and placed in the Auxiliary Activities class of the CAZy database, as families AA10 (CBM33) and AA9 (GH61).

The biological roles of LPMOs are thought to primarily relate to their ability to act on recalcitrant polysaccharide structures during biomass turnover, as outlined in more detail below. However, studies have also suggested alternative functions, primarily functions as virulence factors in opportunistic pathogens not associated with biomass degradation (Kirn et al., 2005, Paspaliari et al., 2015, Agostoni et al., 2017). It is conceivable that in such cases LPMOs enable colonization by breaching protective polysaccharide-rich barriers in the host organism, but experimental evidence for such a function is lacking.

LPMO substrate specificity

Since the discovery of their oxidative chemistry, many LPMOs have been studied, mostly originating from bacteria and fungi. The LPMOs previously known as GH61s are of fungal origin and are classified as AA9s. These LPMOs have proven activity on cellulose (e.g. Phillips et al., 2011, Quinlan et al., 2011), various β-glucan hemicelluloses (Fanuel et al., 2017, Agger et al., 2014), xylan (Frommhagen et al., 2015) and soluble cello-oligosaccharides (Isaksen et al., 2014, Bennati-Granier et al., 2015). Of note, not all AA9 LPMOs have all these activities and not all functionally characterized LPMOs have been tested with all these substrates.

The LPMOs previously known as CBM33s are mainly of bacterial origin and are classified as AA10s. AA10s can display cellulolytic activity (e.g. Forsberg et al., 2011) or

chitinolytic activity (e.g. Vaaje-Kolstad et al., 2010). Very recently, claims of an AA10 being active on xylan have appeared in the literature (Corrêa et al., 2019).

Based on amino acid sequence similarity, additional AA families have been established in recent years, and these families presently include AA11s (active on chitin and mainly of fungal origin; Hemsworth et al., 2014), AA13s (active on starch and of fungal origin; (Vu et al., 2014b, Lo Leggio et al., 2015), AA14s (active on xylan and of fungal origin; Couturier et al., 2018), AA15s (active on cellulose and chitin and isolated from insects; Sabbadin et al., 2018), and AA16 (active on cellulose and of fungal origin; Filiatrault-Chastel et al., 2019).

Accumulating data on natural substrate diversity indicate that LPMOs have evolved to facilitate degradation of all the major carbohydrate polymers frequently present in biomass. It is worth noting that LPMO-producing organisms can contain up to 40 LPMO-encoding genes, suggesting that they employ specialized LPMOs for different substrates, substrate allomorphs and co-polymeric structures in plant cell walls (Busk and Lange, 2015, Horn et al., 2012, Agger et al., 2014).

The third paper of this thesis describes a novel cellulolytic and thermostable AA10 LPMO, and the fourth paper describes attempts to engineer the substrate specificity of an AA10 LPMO from cellulose-specific to chitin-specific in an attempt to shed light on the (largely unknown) structural determinants of LPMO substrate specificity.

LPMO structures

There are currently 39 solved LPMO structures in the PDB database (www.rcsb.org; accessed on 12.06.19), and these are mainly representatives of the AA9 and the AA10 family. The characterized members of these two families share a similar cone-shaped tertiary structure with a flat substrate-binding surface that contains the copper-binding active site (Fig. 17). The core of the protein consists of beta-strands that are connected by loops and a varying amount of short alpha-helices, and some of these connecting elements are involved in the formation of the substrate binding surface. These loops and helices show large variation and are proposed to contribute to substrate specificity of LPMOs, as they constitute large areas of the substrate-binding surface (Li et al., 2012, Book et al., 2014, Vu et al., 2014a, Bissaro et al., 2018a, Courtade et al., 2016, Frandsen et al., 2016).

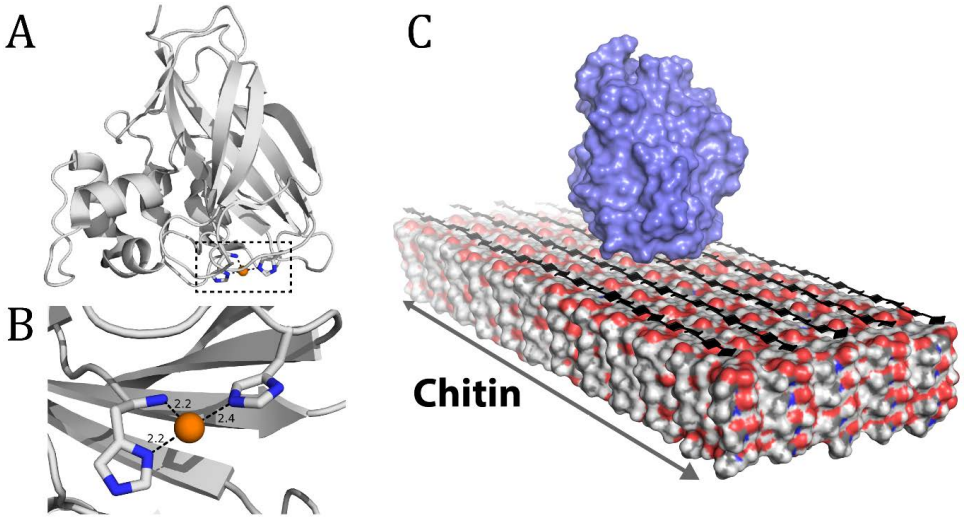

Figure 17. LPMO structure and substrate interaction. Panel A shows the X-ray structure of a cellulose-active AA10 LPMO from *Streptomyces coelicolor* (*Sc*LPMO10C; Forsberg et al., 2014a) in grey cartoon representation with a dotted box denoting the location of the catalytic center, and the substrate-binding surface facing downwards. Panel B shows a close-up view of the catalytic center where a single copper ion (orange sphere) is coordinated by two conserved histidines (grey and blue sticks) in a so-called histidine brace. Note the copper-coordinating function of the N-terminal amino group of the protein, i.e., the amino group of His1. Panel A and B were made using Pymol. Panel C shows a MD simulation of a chitin-active LPMO from *Serratia marcescens* (*Sm*LPMO10A) interacting with crystalline chitin. The Figure was taken from Bissaro et al., 2018a.

The more recently discovered LPMO families have only one solved structure each (and none for AA16 so far). The members of these families (AA11, AA13 & AA14 & AA15) show large structural similarity with LPMOs in families AA9 and AA10, including a surface-exposed copper-binding site on a relatively flat binding surface. There are however variations in how flat the surface is, and Figure 18 shows that LPMOs do exhibit variations in surface shape and ruggedness. For example, structural studies of the starch-active AA13 LPMO revealed a shallow groove across the copper site (PDB ID: 4OPB; Lo Leggio et al., 2015). Likewise, Couturier et al., 2018 showed that the surface of an AA14 LPMO that acts specifically on cellulose-bound xylan has a more rippled shape, compared to the flat surfaces of many AA9 LPMOs.

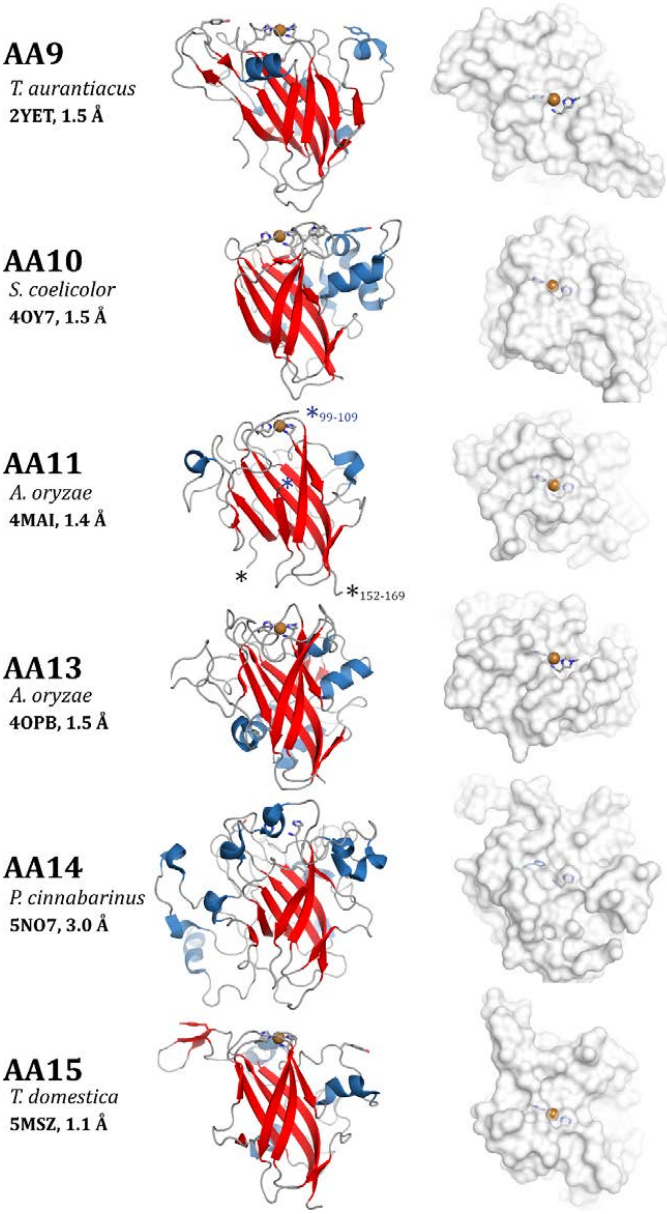

Figure 18. Three-dimensional structures of representatives of six of the seven currently known LPMO families. Left panels: The proteins are depicted with secondary structure representation and with their substrate-binding surface facing upwards. The catalytic copper (orange sphere) and histidine brace (sticks) are marked (note that the AA14 structure does not contain a copper). The blue and black stars denote stretches of sequence that could not be modelled in the AA11 structure. Right panels: The substrate binding surface is displayed with surface representation. The PDB codes are indicated in the figure. The figure was made in Pymol.

Catalytic mechanism

X-ray crystallographic data have provided insights into the binding of single oligomeric substrates to LPMOs (Simmons et al., 2017, Frandsen et al., 2016), whereas LPMO surfaces interacting with polymeric substrates have been experimentally mapped by using NMR (Aachmann et al., 2012, Courtade et al., 2016) and modelling (Bissaro et al., 2018a). The binding surfaces contain a highly characteristic "histidine-brace" (Quinlan et al., 2011; Fig. 17) made up of two conserved histidines of which one is always the N-terminal residue of the protein. This histidine brace binds a single copper (Fig. 17B) and constitutes the active site of the enzyme. Of note, the N-terminal histidine is methylated in fungal LPMOs. Recent studies (Petrovic et al., 2018) indicate that this methylation has little effect of the catalytic properties of the LPMO but may have an effect on enzyme stability.

The catalytic mechanism of LPMOs remains somewhat enigmatic and its details are beyond the scope of its thesis. In their original 2010 study, Vaaje-Kolstad and colleagues demonstrated that LPMOs need reducing power to function and it is well established that the copper must be reduced from Cu(II) to Cu(I) state for oxidative cleavage to occur. Electrons for driving LPMO reactions may be derived from a variety of small molecule reductants, including ascorbic acid (Frommhagen et al., 2016, Vaaje-Kolstad et al., 2010, Quinlan et al., 2011, Kracher et al., 2016), lignin and lignin fragments (Muraleedharan et al., 2018, Hu et al., 2016, Westereng et al., 2015), as well as various redox enzymes, of which cellobiose dehydrogenase is the best known (Phillips et al., 2011, Kracher et al., 2016, Langston et al., 2011, Tan et al., 2015).

What may happen after reduction of the copper has been discussed in multiple research papers and reviews (e.g. Chylenski et al., 2019, Beeson et al., 2015, Phillips et al., 2011, Bissaro et al., 2017, Wang et al., 2018, Walton and Davies, 2016). The key step in LPMO catalysis is the formation of an intermediate oxygen species that is a sufficiently strong oxidant to abstract a hydrogen atom from one of the carbons in the scissile glycosidic bond. It is generally assumed that hydrogen abstraction is followed by hydroxylation of the resulting substrate radical, which leads to destabilization and spontaneous cleavage of the glycosidic bond (Phillips et al., 2011, Chylenski et al., 2019).

The LPMO reaction may lead to oxidation of either the C1 or the C4 carbon in the scissile glycosidic bond. If the oxidation resides on the C1-carbon of the newly formed upstream chain end, the products are referred to as C1-oxidized, and if the C4-carbon of the newly formed downstream chain end carries the oxidation, then the product is referred to as C4-oxidized (Fig. 19). The primary products of C1-oxidation are 1,5-δ-lactones, which are spontaneously converted to their more stable aldonic acid forms. The primary products of C4-oxidation are 4-ketoaldoses, which are spontaneously hydrated to their corresponding gemdiol forms (Isaksen et al., 2014). Different LPMOs can generate either C1- or C4-oxidized products, while some LPMOs generate mixtures of both.

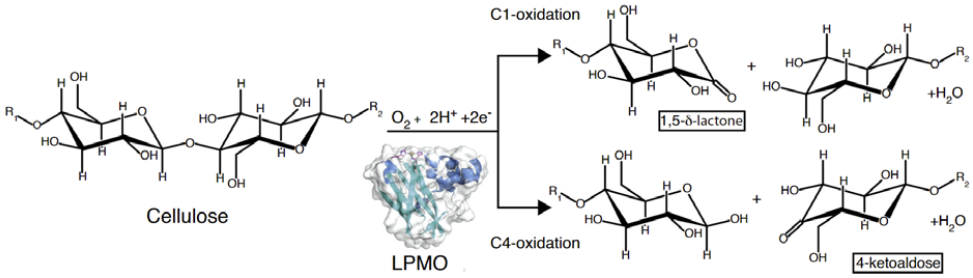

Figure 19. Regioselectivity of LPMOs. LPMOs oxidize either the C1 or C4 carbon adjacent to the glycosidic bond, or they may be able to introduce the oxidation at either site. C1-oxidation results in a lactone that is spontaneously converted to an aldonic acid. C4 oxidation generates 4-ketoaldoses that are spontaneously converted to the corresponding gemdiol form. The picture was taken from (Vaaje-Kolstad et al., 2017).

It was initially assumed that the oxidizing species in LPMO catalysis was generated from molecular oxygen, which thus was considered the natural co-substrate of LPMOs and leading to these enzymes being named monooxygenases (Fig. 20, equation A). Recent studies however challenge this view by suggesting that hydrogen peroxide is the true co-substrate of LPMOs (Bissaro et al., 2018b, Bissaro et al., 2017), and that these enzymes thus should be referred to as peroxygenases (Fig. 20, equation B).

Figure 20 shows an overview of LPMO reactions involving O_2 and H_2O_2. The oxygen-driven reaction (Fig. 20A) depends on the supply of two electrons and two protons per catalytic cycle. Importantly, this entails stoichiometric consumption of reducing equivalents, which may limit the LPMO reaction under certain conditions. The

monooxygenase reaction involves a reduction of the catalytic copper from Cu(II) to Cu(I) after which the reduced LPMO binds and reduces molecular oxygen, leading to the formation of a superoxide-LPMO intermediate (Kjaergaard et al., 2014). While the first electron needed in the catalytic cycle is stored on the LPMO after the reductant reduces the copper, a major question related to the O_2–driven reaction is how the second electron and the two protons are delivered to the active site while catalysis is ongoing and the active site is confined by the substrate. Exactly how the electron and protons are acquired has been debated and may perhaps involve some sort of electron transport channel or chain (Beeson et al., 2015, Walton and Davies, 2016). Of note, it has been shown by NMR, that cellobiose dehydrogenase, which is assumed to drive the O_2-mediated LPMO reaction by delivering electrons, cannot interact with the copper site when the LPMO binds a substrate (Courtade et al., 2016).

The delivery of the second electron is a non-issue when considering the H_2O_2–driven reaction depicted in Fig. 20B. Bissaro et al. demonstrated in 2016 that addition of H_2O_2 under anaerobic conditions could drive LPMO reactions and significantly boost the reaction rate compared to aerobic conditions without addition of H_2O_2. Competition experiments with labelled H_2O_2 clearly indicated that the LPMOs preferably use the available H_2O_2 during catalysis, even in the presence of a ten-fold molar surplus of O_2 (Bissaro et al., 2016, Bissaro et al., 2017). Bissaro et al. also showed that a H_2O_2 consuming enzyme (horseradish peroxidase) inhibited LPMO activity. Subsequent studies support the ability of LPMOs to use H_2O_2 instead of O_2 (Hangasky et al., 2018, Kuusk et al., 2018, Wang et al., 2018, Hedegard and Ryde, 2018, Hegnar et al., 2019), but whether H_2O_2 or O_2 is the preferred or "natural" co-substrate remains debated (Bissaro et al., 2017, Bissaro et al., 2018b, Chylenski et al., 2019, Forsberg et al., 2019, Hangasky et al., 2018, Eijsink et al., 2019).

The H_2O_2-driven reaction requires only a single electron for a priming reduction from Cu(II) to Cu(I) state, after which the enzyme can carry out consecutive catalytic cuts. The exact reaction mechanism has not been established yet, but the reduced LPMO is believed to interact with a H_2O_2 molecule that is coordinated between the catalytic copper and the scissile bond. In any case, the H_2O_2-driven reaction regenerates the Cu(I) state which enables the LPMO to directly initiate a new catalytic cycle (Fig. 20).

Importantly, reduced LPMOs (Cu(I) state) act as oxidases when not bound to the proper substrate, and in this situation they will reduce O_2 to H_2O_2 (Isaksen et al., 2014, Kittl et al., 2012), meaning that H_2O_2 can potentially be generated in any LPMO reaction (step 3 in Fig. 20). LPMOs are easily reduced (Kracher et al., 2016), which means that when assuming O_2-driven catalysis, the transfer of the second electron likely is the rate limiting step of the reaction. However, if one assumes that the LPMOs only use H_2O_2, the reaction rate is likely limited by the rate at which the LPMO can generate H_2O_2 from O_2 (in reactions without addition of exogenous H_2O_2, i.e. most LPMO reactions described to date). Additionally, the presence of reductant that has not yet been used to reduce the LPMO may lead to formation of O_2-derived intermediates such as superoxide and H_2O_2, in particular in the presence of trace amounts of transition metals (step 5 in Fig. 20). To complicate things further, some reductants may also react with H_2O_2 (step 5' in Fig. 20).

Next to the multiple processes involving reductant and LPMO and that can affect the levels of H_2O_2 in the reaction, H_2O_2 consuming and H_2O_2 generating enzymes may play a role in LPMO-mediated degradation of biomass in nature (Bissaro et al., 2018b). H_2O_2 consuming enzymes include for example peroxidases acting on lignin, whereas H_2O_2 generating enzymes include members of the AA3 family that are frequently produced by biomass degrading fungi, including cellobiose dehydrogenases and other members of this family that are true oxidases, such as glucose oxidase. All in all, it would seem that H_2O_2 is a readily available co-substrate both in the natural setting of LPMOs and in common laboratory settings.

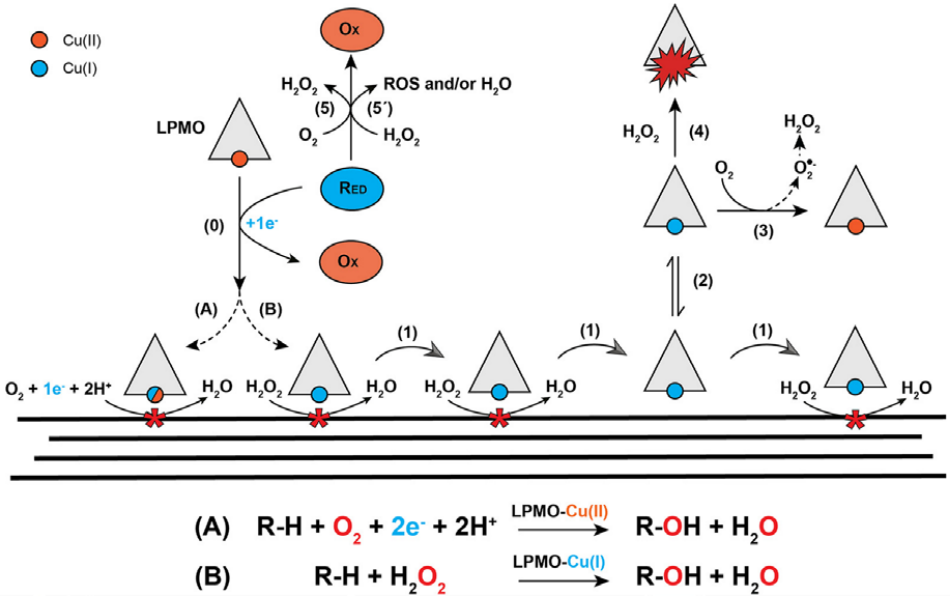

Figure 20. Proposed reactions involving LPMOs. Reaction (A) shows the oxygen-mediated reaction and is depicted in the left side of the Figure. In this reaction, the LPMO needs to be reduced from the Cu(II) to Cu(I) state [step (0)] prior to every catalytic event, while a second electron and two protons need to be delivered in each catalytic cycle. Reaction (B) shows the H_2O_2 -dependent reaction, where only a priming reduction [(0)] is needed before the LPMO can carry out consecutive catalytic events [(1)], without additional supply on electrons or protons. In this case, electrons and protons are supplied from the hydrogen peroxide. Note that reduced LPMOs in solution may react with O_2 to generate H_2O_2 [(3)] and are prone to oxidative self-inactivation when reacting with H_2O_2 in the absence of the proper substrate [(4)]. H_2O_2 may also be formed by reactions involving O_2 and the reductant [(5)] in the presence of trace amounts of transition metals and may react further with these reductants to generate other reactive species [(5')]. The figure was taken from (Hegnar et al., 2019).

Importantly, reduced LPMOs are prone to autocatalytic inactivation, especially at low substrate concentration (Bissaro et al., 2017, Courtade et al., 2018, Forsberg et al., 2018, Loose et al., 2018); step 4 in Fig. 20). The reduced copper of an LPMO in solution will react with O_2 or H_2O_2, and in the absence of substrate to create a "caged" environment for the generated reactive oxygen species (Wang et al., 2018, Bissaro et al., 2016, Bissaro et al., 2017, Bissaro et al., 2018b, Frandsen et al., 2016, Simmons et al., 2017), this species may damage the enzyme itself, as has been indeed documented (Loose et al.,

2018, Bissaro et al., 2017). It has been claimed that enzyme inactivation is more prominent in H_2O_2-driven reaction relative to O_2-driven reactions (Hangasky et al., 2018), but this is disputed (Eijsink et al., 2019, Chylenski et al., 2019).

Despite the large attention to LPMOs in recent years and despite major progress, many aspects of these enzymes remain unknown, for example, the structural determinants of substrate specificity. Of note, for LPMOs, little is known about the variation in and the structural determinants of more "simple" but industrially highly relevant properties, such as the pH and temperature optima for activity and thermal stability. Paper III of this thesis describes a thermostable LPMO with interesting properties, which may be used in further studies on the effects of temperature on LPMO functionality. Paper IV describes attempts to unravel structural determinants of the substrate specificity of LPMOs.

1.4 The emerging bioeconomy

The world economy has since the industrial revolution of the 18th century relied heavily on non-renewable carbon sources (i.e., fossilized carbon such as coal, petroleum, and natural gas) to sustain the demand for energy, chemicals, and fuels (Liu et al., 2012). Fossilized carbon encompasses complex hydrocarbons that originate from ancient biomass that was deposited millions of years ago and exposed to high pressure and temperature over a prolonged period (Gupta, 2018, Schobert, 2013). Importantly, unlike the biomass of living organisms, the fossilized carbon reservoirs are finite and will eventually be depleted (Höök et al., 2010). Additionally, the use of fossilized carbon sources is considered hazardous to the environment, as it is a major source of air pollutants, contributes to global climate changes through the release of climate gases, and disrupts natural landscapes during the extraction process (Vanholme et al., 2013).

The term bioeconomy refers to economic activities using renewable biological resources to produce energy and commodities (Guo and Song, 2019). A transition towards a sustainable bioeconomy is indeed ongoing, with the aim of exploiting renewable biological sources (e.g. from agriculture, forestry, aquaculture, and municipal biowaste) in an environment friendly manner to sustain our consumption. Non-biological renewable energy sources, including wind power, solar power, geothermal power, and hydropower, will undoubtedly play an important role in the future of

society, but these sources are mainly used to generate electricity, not chemicals and consumables (Henrich et al., 2015). “Green carbon” from biomass represents an attractive alternative to fossilized carbon, and can be converted to value-added products such as food, feed, fiber, chemicals, materials, and fuels, with the potential of replacing most if not all of today’s petroleum-based products (Vanholme et al., 2013). The production of plant biomass additionally captures CO_2 from the atmosphere during the growth phase. While there are many types of biomass that could be considered, this thesis focuses on plant-derived “lignocellulosic” biomass.

As discussed in sections 1.2 and 1.3, specialized biocatalysts have developed in Nature to function as tools in the turnover of continuously replenished biomasses. These tools can be implemented in industrial processes to enable enzyme-driven bioprocessing in biorefineries (Fig. 21). Of note, while this thesis focuses on the use of enzymes and enzyme mixtures, Nature’s catalytic power may also be exploited in the form of microbes or microbial communities with abilities to convert biomass to useful products. One example is the production of biogas (methane) through anaerobic digestion. An area of research referred to as “synthetic biology” aims at developing microbes that are genetically engineered to carry out specific biotransformations.

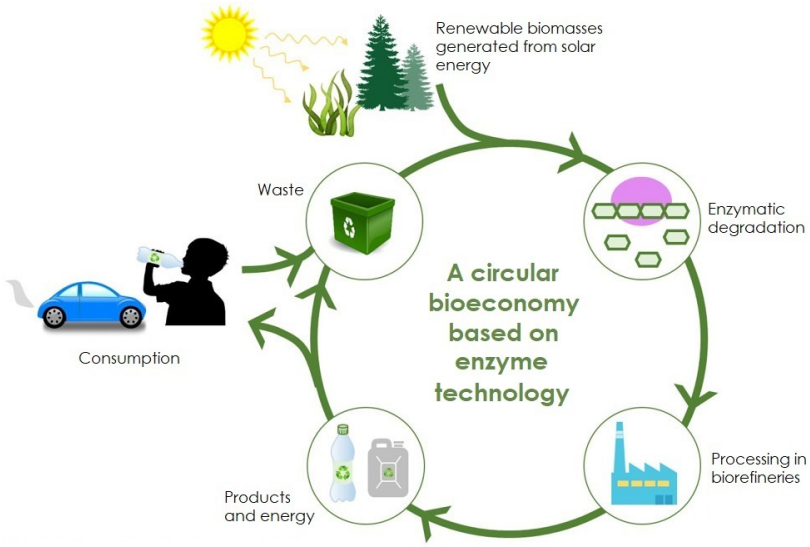

Figure 21. Schematic overview of the bioeconomy. The use of specialized enzymes as biocatalysts to degrade recalcitrant biomasses is crucial for achieving a circular bioeconomy. Please note that biomass processing involves more steps than enzymatic degradation, as outlined in the main text, and that production of certain biomass-derived products, such as cellulose-based materials, do not necessarily involve enzymes. The figure was adapted from www.bioteknologiradet.no (the Norwegian Biotechnology Advisory Board).

1.4.1 Enzyme technology for bioprocessing in biorefineries

"Biorefinery" is a concept based on the conversion of biomass to chemicals, energy and materials that can replace the need for petroleum, coal, natural gas, and other non-renewable energy and chemical sources (Liu et al., 2012). The most well-known products of biorefineries are biofuels, in particular bioethanol obtained by fermentation of biomass-derived sugars.

So-called "1st generation biofuels", which are made from food crops that mostly consist of the easily degradable carbohydrate starch, can already be produced in an economically competitive process (Vanholme 2013). However, 1st generation biofuels are not considered a truly sustainable source of energy because they compete directly with human food (i.e. deplete food crops and farming land) and because the energy balance of their production and use is not particularly favorable, compared to cellulosic ethanol (Van Noorden, 2013).

Cellulose, which constitutes a non-edible and much more rigid carbohydrate than starch, has traditionally been used for firewood, building material, cotton, and paper to provide energy and useful materials. However, cellulosic biomass can alternatively be converted to sugars that may be fermented to high-value products and fuels, referred to as 2nd generation biofuels and chemicals, and this possibility has received much attention in recent years. For instance, agriculture and forestry generate large amounts of cellulosic by-products that are currently being disposed of as waste, and bioprocessing of these by-products would ensure complete utilization of the biomass feedstock (Saini et al., 2015). One major bottleneck related to making 2nd generation biofuels and chemicals lies in overcoming the recalcitrance of lignocellulosic biomass. Despite major progress in recent years, in part due to the discovery of LPMOs, the enzyme cocktails currently used are not always sufficiently effective. Thus, the required enzyme dosages may prevent processes from being economically viable (Taha et al., 2016).

One way to decrease the recalcitrance prior to adding enzymes, is to reduce the particle size and disrupt the substrate matrix in order to make the polysaccharides more accessible (Arantes and Saddler, 2011), as visualized in Fig. 22. Such pre-treatments include milling, chipping, grinding or freezing, followed by physiochemical (e.g. steam explosion), chemical (e.g. dilute acid or ionic liquids) or biological (e.g. microbes or enzymes) pre-treatments that disrupt the lignocellulosic matrix by removing or modifying hemicellulose and/or lignin and loosening the crystalline structure of the cellulose (Harun et al., 2011). A detailed discussion of these technologies (Galbe and Zacchi, 2012, Chandra et al., 2007, Chaturvedi and Verma, 2013, Maurya et al., 2015, Haghighi Mood et al., 2013) is beyond the scope of this thesis. Importantly, for most, if not all lignocellulosic biomasses, the current status is that relatively harsh pre-treatments are needed to reach acceptable yields during subsequent enzymatic depolymerization.

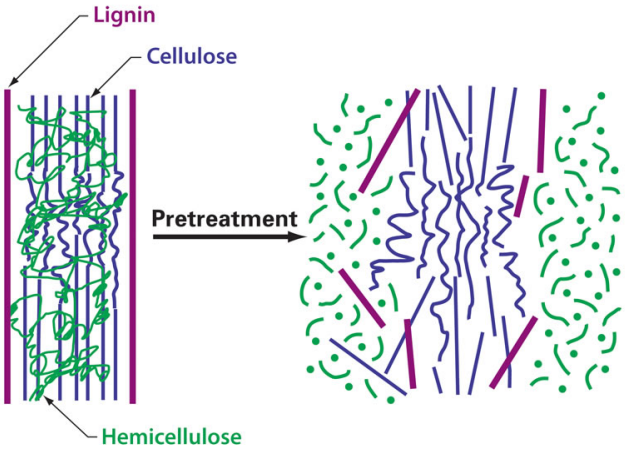

Figure 22. Pre-treatment on lignocellulosic biomass. Different pre-treatment methods are used to reduce particle size, increase accessible surface area, and remove or modify the hemicellulose and lignin that surround the cellulose fibrils. The figure was taken from www.glbrc.org (Great Lakes Bioenergy Research Center).

After pre-treatment, the biomass can be depolymerized to fermentable sugars using the CAZymes described in section 1.3. The monomeric sugars can subsequently be processed to desired compounds (such as bioethanol or single cell protein) by fermenting organisms or used as a precursor to be further processed into advanced chemicals. In order to maximize the yields during the enzymatic depolymerization, the CAZymes are applied together in carefully determined optimal ratios, commonly referred to as enzyme cocktails. Enzymes are costly, and it is thus desirable to use enzymes that are easily produced in order to lower the total enzyme cost, and to employ enzymes that are highly stable at the processing conditions in order to ensure that the enzymes stay active throughout the incubation time. Thermal stability is a factor of particular importance in many biorefinery processes. High temperature limits contamination of the bioreactor by mesophilic microbes, increases the solubility of organic compounds and thus allows processing of the substrate at high dry matter concentration, and speeds up the reaction rate (Haki and Rakshit, 2003).

1.4.2 Strategies towards better enzymes

Enzymes suited for industrial use may be discovered through bioprospecting of environments where the conditions resemble the desired processing conditions. Bioprospecting generally involves either sequence-based selection of (putative) enzyme-encoding genes, based on sufficient sequence similarity to previously characterized enzymes (genome or metagenome mining), or activity-based screening, where one looks for desired activities, followed by a targeted search for the encoding gene (functional screening). Importantly, the recent (meta)-omics revolution, makes it possible to access many biodiversities without ever culturing a microbe. By combining metagenomic and metaproteomic information, active metabolic pathways in an industrially interesting microbial ecosystem may be identified (e.g. Hagen et al., 2017). Enzymes can be further optimized for industrial bioprocessing conditions and optimal performance on industrially relevant substrates through enzyme engineering, which can be based on strategies such as rational design and directed evolution.

The bioprospecting approach implies that one searches organisms and environments for useful enzymes. In an industrial perspective, one would be looking for enzymes that may work well under industrial conditions, which include uncommon temperature, pH, or salinity. Biomass-degrading organisms and their enzymes are abundant in Nature, including in ecological niches with extreme conditions, and thus hold almost unlimited possibilities for discovery of novel enzymes with interesting properties. Mining of microbial metagenomes generated from special ecological niches may yield enzymes that are adapted to these niches, as shown in Papers I and II of this thesis. The deposition of metagenomes from relevant environments in publicly accessible databases has eased modern bioprospecting by making enzyme-encoding sequences easily accessible (Tiwari et al., 2018, Li et al., 2011, Madhavan et al., 2017).

It is important to note that bioprospecting based on metagenome mining can only reveal enzymes that are similar to previously characterized enzymes (i.e., their sequences share enough identity with enzymes of known function to relatively safely predict that the structure and activity must be similar). By comparison, functional screening of microbial communities only addresses the actual performance of the enzymes and is not biased by the sequence, which can lead to discovery of completely novel enzymes. One method for carrying out functional screening is to express random fragments of a

metagenome in a so-called fosmid library, and subsequently explore the sequences of fosmids that encode desired enzyme activities (Armstrong et al., 2019). Fosmids (F-plasmid-based vectors usually introduced to E. coli host cells) can express inserts of up to 40 kb, and detection of activity therefore usually requires subcloning of the individual reading frames present on the fragment in order to establish exactly which gene encodes the novel enzyme with the desired properties.

While bioprospecting provides scientists and industry with enzymes that have evolved over millions of years to carry out tasks under specific conditions, there are still remaining challenges that need to be met to make industrial enzymatic bioprocessing economically feasible. As discussed above, industrial biocatalysts need to be produced in a cheap manner, which is not always possible. Furthermore, key operational properties of candidate enzymes, in particular the pH-activity profile, the temperature-activity profile, as well as thermal and pH stability, may not be optimal for the envisaged industrial application.

To improve the efficiency of natural enzymes at process conditions, different strategies for enzyme engineering can be applied to alter traits. These can be based on rational design and mutagenesis of a few selected residues (e.g. Zhang et al., 2015, Eijsink et al., 2004), or on generating mutant libraries that can be searched for improved variants using high-throughput screening (directed evolution) (Schmidt-Dannert and Arnold, 1999, Lutz and Iamurri, 2018, Eijsink et al., 2005). One key method for generating diversity in mutant libraries is “gene shuffling” (e.g. Dana et al., 2012, Stemmer, 1994), which implies the generation of libraries of hybrid enzymes. While an in-depth discussion of these strategies and of the structure-function relationships that underlie enzyme parameters such as thermal stability is beyond the scope of this thesis, a short summary of key strategies is provided below. Of note, in the context of lignocellulose processing, engineering efforts may also be directed at the biomass itself, by engineering for instance plants to be more easily degradable (Mahon and Mansfield, 2019).

Rational design, also referred to as structure-based design, for improving enzyme properties normally requires that one knows the crystal structure of the enzyme and understands the structural basis of enzyme properties. However, multiple sequence alignments of enzymes with characterized traits (but potentially unknown structure) can also be a starting point for rational design (sequence-based design), in which

sequences of related enzymes are aligned and compared to identify amino acid patterns that correlate with the desired trait. Sequence alignments may also be used to identify a "consensus sequence", which in some cases may represent a protein of improved stability (Lehmann and Wyss, 2001). Rational design employs site-directed mutagenesis to introduce the amino acid mutations that are predicted to have implications for the desired trait. Thermostability is often of particular interest because it may allow prolonged use of the enzyme at harsh conditions that may be desired in industrial applications (Han et al., 2019, Watanabe et al., 2018). Strategies for thermostabilization through rational design include so-called entropic stabilization by rigidifying mutations such as the removal of glycines, or the introduction of prolines or sulfide bridges, which may be particularly effective when introduced in surface loops in order to reduce loop flexibility. Other stabilizing mutations could be aimed at stabilizing secondary structure elements (e.g., "helix capping") or strengthening hydrophobic interactions within the protein (Watanabe et al., 2018, Eijsink et al., 2004).

The first directed evolution of an enzyme (Chen and Arnold, 1993) was the start of a journey that led to Frances Arnold being awarded the Nobel prize in Chemistry in 2018. Directed evolution mimics natural evolution by introducing an artificial selection pressure or selection criteria while speeding up the DNA mutation rate, and was enabled by the high-throughput screening (HTS) revolution, i.e., the ability to carry out massive amounts of parallel biochemical reactions using robots (Pereira and Williams, 2007). Directed evolution usually relies on error-prone PCR that introduces random mutations, followed by a screening step to identify variants with improved properties, and is thus not dependent on structural data. Since, error-prone PCR will generate huge libraries (and since screening is costly), more sophisticate knowledge-based methods for library generation are often used, for example methods that only mutate a (relevant) part of the enzyme or a sub-set of codons for selected residues that are thought to be important (see below). Gene shuffling may be used to construct hybrid enzymes, in which a set of related sequences are fragmented and randomly recombined (Dalby, 2011, Stemmer, 1994, Dana et al., 2012). Directed evolution generally depends on iterative cycles where the best enzymes from each cycle are used as starting point for the next cycle (e.g. Eijsink et al., 2005).

There are several downsides to directed evolution (Watanabe et al., 2018). Very few of the random mutants are likely to display the desired characteristics. Thus, if one does not have any information that may be used to limit library size (see above), huge random libraries must be produced to increase the chances of generating only a few improved mutants. Also, there is a possibility that the introduction of multiple mutations results in beneficial mutations being masked by mutations with negative effects. Multiple mutations may also hamper the process of determining exactly which mutations led to the improved traits. Finally, error-prone PCR will produce many mutations on the DNA-level, but not all of these will affect the protein at the amino acid level, as amino acid substitutions often require more than one mutation within each codon.

An approach that employs a combination of the abovementioned strategies may be the best method for altering enzyme traits efficiently (Lehmann and Wyss, 2001, Lutz, 2010, Steiner and Schwab, 2012). Semi-rational design (or focused directed evolution) results in smaller, more high-quality libraries focusing on pre-selected target sites. A common approach is to carry out saturation mutagenesis of selected sites, which implies that the libraries are constructed such that they include all 20 possible amino acids at these selected sites. An early example of combining enzyme engineering methods successfully (site-directed mutagenesis, error-prone PCR, and finally random shuffling of the best mutations) was demonstrated for a heme peroxidase for application in laundry detergents, where one mutant showed a 174-fold increase in thermal stability and a 100-fold increase in oxidative stability relative to the wild-type enzyme (Cherry et al., 1999).

Paper IV of this thesis describes a study in which high-throughput screening of a semi-rationally designed mutant library was used to tackle one of the major outstanding issues in the LPMO field, namely identification of the structural determinants of substrate specificity.

2 THE PURPOSE OF THIS STUDY AND OUTLINE OF THE THESIS

Enzymes are likely to play an important role as biocatalysts in the transition to a sustainable bioeconomy. CAZymes can be used in processing of renewable carbon sources such as biomass from forestry, agriculture, and aquaculture to generate energy and commodities in an environmentally friendly manner. However, these biomasses are often highly resistant to enzymatic conversion and the efficiencies of several processing steps, including the performance of the enzymes themselves, remain major bottlenecks in making industrial bioprocessing economically viable.

Better enzymes for improved bioprocessing can be discovered through bioprospecting of relevant environments where biomass turnover occurs in nature. In addition, naturally evolved enzymes may be optimized for industrial process conditions or industrial substrates by enzyme engineering.

The work described in this thesis was done within the framework of the national NorZymeD project in Norway. The project involved Universities, research institutions and industrial actors and was aimed at establishing a national pipeline for the development of enzyme technology with the potential of increasing value creation from various biomasses in Norway.

In the NorZymeD project, metagenomic data from various biodiversities was searched by metagenome mining to discover novel enzymes. The metagenome sources included high-temperature environments, such as bioreactors and compost, hot sediments from deep-sea vents of the Arctic Mid-Ocean Ridge, gill-symbionts of the arctic wood-eating shipworm, and the microbiome of the Svalbard reindeer rumen. Selected enzyme candidates were produced and characterized in lab-scale to evaluate their potential for industrial application, using industrially relevant process conditions and substrates provided by the industrial partners. Selected enzymes were tested in pilot scale at industrial facilities. Enzyme engineering through directed evolution was employed in an effort to alter enzyme traits, which implied the generation of mutant libraries, followed by screening using high throughput mass spectrometry-based methods developed within the project.

This thesis comprises three scientific papers that report discovery and characterization studies of three different CAZymes with potential applications in the processing of lignocellulosic biomass. Paper I describes a thermostable GH6 endocellulase from a rice straw/compost metagenome, whereas Paper III describes a thermostable LPMO from the same environment. Paper II describes a GH10 xylanase, with a novel CBM (CBM85) that was derived from deep-sea hot vents of the Arctic Mid-Ocean Ridge. Paper IV describes the first directed evolution study of an LPMO and was carried out to assess the opportunities provided by MS-based high throughput screening technologies. As an initial assessment of these opportunities, a mutant library was generated and then screened for variants with altered substrate specificity.

The conducted work is mainly fundamental research on prominent enzymes for bioprocessing of lignocellulosic biomass (i.e., cellulases, hemicellulases, LPMOs, and carbohydrate binding modules), but with emphasis on industrially relevant aspects such as thermal stability and substrate specificity.

3 MAIN RESULTS AND DISCUSSION

3.1 Paper I – Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome

Depolymerization of cellulose to glucose is a common main goal in biorefining of plant biomass. Glucose is an ideal starting platform for fermentation processes that generate value-added products such as bioethanol or single cell protein. Degradation of cellulose can be achieved by deploying cellulases, which commonly entails the combined action of endoglucanases, cellobiohydrolases and β-glucosidases (see section 1.3).

After metagenome mining of publicly available metagenomic data collected from rice straw inoculated with compost and incubated at high temperature (55 °C) (Reddy et al., 2013), 20 CAZyme-encoding genes were synthesized for expression in *E. coli*. Subsequent cloning, expression and purification studies yielded 14 soluble proteins of which 8 showed significant activity on an industrial cellulose substrate (sulfite-pulped Norway spruce from Borregaard AS): three GH5s, two GH6s, two GH48s, and one GH9. Based on a combined assessment of the activity of the individual enzymes on sulfite-pulped spruce and the ease of protein production and purification, a putative endoglucanase (mgCel6A) belonging to the GH6 family was subjected to in-depth studies, which are described in Paper I. In addition to the GH6 catalytic domain, the mature protein encompasses an extended linker region and a carbohydrate-binding module (CBM2; Fig. 23).

Figure 23. Domain structure of mgCel6A. The full-length enzyme [omitting the signal peptide (SP)] consists of a catalytic domain (GH6) connected to a binding module (CBM2) through a 40-residue linker rich in prolines and threonines.

C-terminally His-tagged versions of full-length mgCel6A and the catalytic domain alone were expressed in *E. coli* and the enzymes were purified using immobilized metal affinity chromatography (IMAC). The full-length enzyme was easily produced despite its

40 residue linker, with approximately 500 mg enzyme obtained per liter *E. coli* culture. A simple and high-yielding production process is desirable for any enzyme as it lowers the total cost of enzyme production, and is an important aspect when evaluating the potential of enzymes in applied settings.

Functional characterization showed that mgCel6A is thermostable enzyme, a trait that is often desired in industrial bioprocessing, for example to avoid contamination issues or increase product solubility. The apparent melting temperature (T_m) was 76 °C, and the highest product yields from the industrial spruce substrate pre-treated by sulfite-pulping were obtained at 75 °C in short reactions (15 minutes). In overnight reactions, mgCel6A performed best at 60 °C. In stability assays, the enzyme retained an impressive 90 % activity after being pre-incubated at 65 °C for 24 hours.

The enzyme exhibited a clear preference for amorphous cellulosic substrates such as PASC (phosphoric acid swollen cellulose) and the sulfite-pulped spruce, compared to crystalline cellulose, such as Avicel and filter paper. Approximately 40 % of the chain ends generated during the enzymatic reaction resided in the non-solubilized fraction of the substrate, indicating that the enzyme does not attack from the chain ends and consistent with an endoglucanase mode of action. Analysis of products generated from sulfite-pulped spruce showed considerable production of trimeric products, which also indicates endoglucanase action.

Protein crystals were obtained for the catalytic domain and the solved structure supported an endoglucanase mode of action by showing an open substrate-binding cleft and an overall close resemblance to the nearest homologue with a solved structure, the endoglucanase *Tf*Cel6A from *Thermobifidia fusca* (Spezio et al., 1993), with 68% sequence identity. The structure of mgCel6A showed a putative flexible loop region in which two residues could not be modelled (Fig. 24). This is consistent with observations made for *Tf*Cel6A, where the corresponding loop is thought to alter its conformation upon substrate binding (Larsson et al., 2005).

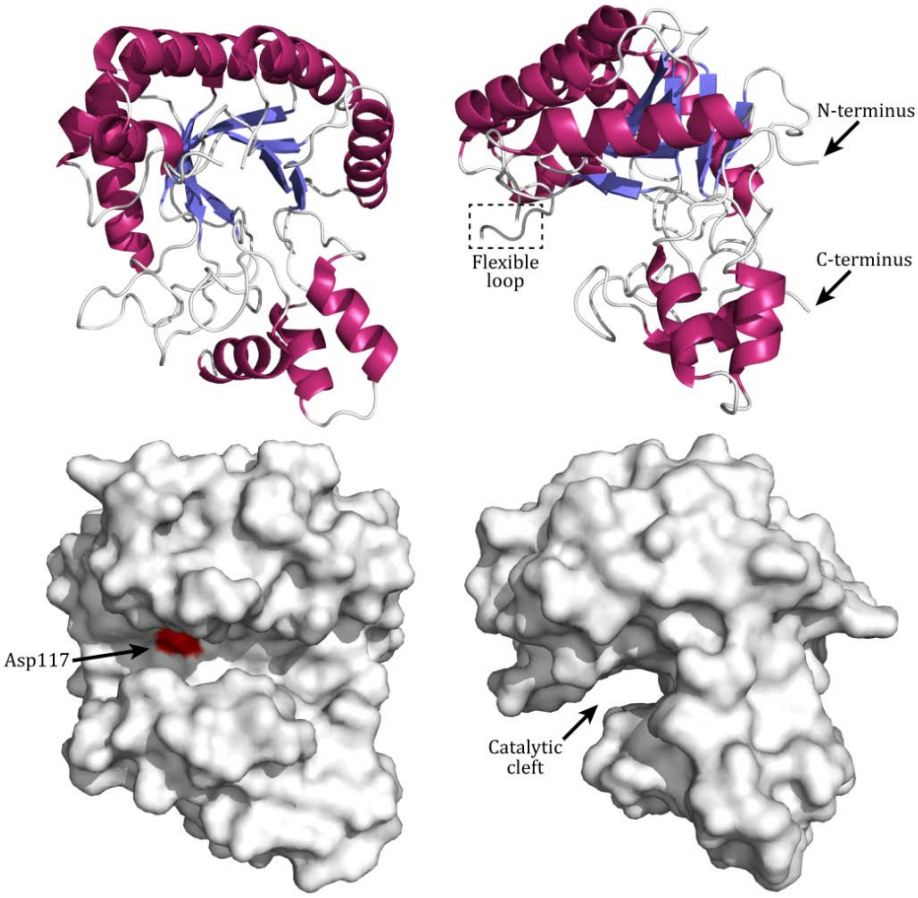

Figure 24. Structure of the catalytic GH6 domain of mgCel6A. The enzyme is viewed looking directly onto the catalytic cleft (left panels), and from the side (right panels). The upper panels show the structure as cartoon representation with β-sheets in blue, α-helices in pink, and loops in white. The lower panels show a surface representation of the structure. The structure was solved at 1.88 Å (PDB ID: 6FAO) and shows a deep catalytic cleft. Ser84 and Ser85 that reside in a glycine-rich loop (GASSGG) could not be modelled and indicate a flexible region that might undergo conformational changes upon substrate binding. The location of the putative catalytic acid, Asp117, is indicated in red in the lower left panel (note that the catalytic base remains elusive in GH6s; Payne et al., 2015).

The closely related *Tf*Cel6A is a well-studied member of the GH6 family (Calza et al., 1985), and comparison of the two enzymes using previously published data for *Tf*Cel6A indicates that mgCel6A is more stable and easier to produce. Comparison of their

activity on sulfite-pulped spruce (this study) showed that both enzymes could only convert approximately 15 % of the substrate in 48 hours when working alone (loaded at 8 mg/g cellulose), and thus, evidently, need synergistic partners such as cellobiohydrolases, LPMOs or β-glucosidases to fully depolymerize the substrate.

The effect of the CBM2 on the efficiency of the GH6 domain was also investigated and it was shown that the presence of the CBM hampered product yields when the substrate loading exceeded 50 g/L dry matter (Fig. 25). This important observation is similar to observations made for fungal cellulases that are connected to CBM1s (Varnai et al., 2013), and underpins that the role of CBMs in industrial processes may require some reconsideration. CBMs confer a "proximity effect" (Boraston et al., 2004) that increases the effective enzyme concentration on the substrate in dilute systems, but this effect seems to be abolished at high substrate concentrations when the catalytic domain is likely to encounter substrate-binding sites regardless of the CBM. The negative effect of the CBM under these conditions is presumably because the CBM hampers the rate of substrate desorption. The substrate loading in industrial settings is usually high, and the present results thus indicate that it could be beneficial to employ enzymes without CBMs in industrial bioprocessing.

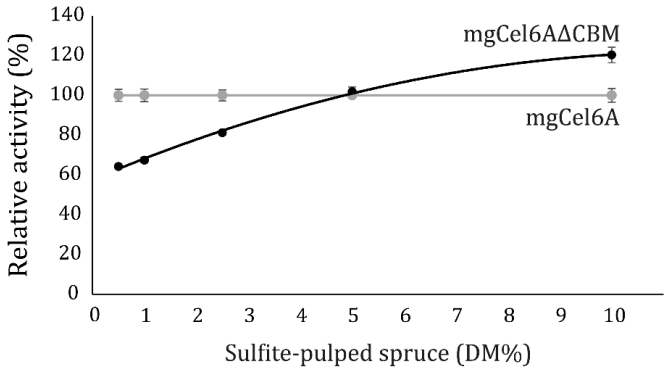

Figure 25. Effect of the CBM2 at different dry matter concentrations. The yield after 48 hours of the reaction with the full-length enzyme (mgCel6A) at each dry matter concentration (DM%) is defined as 100 % activity. The catalytic domain alone (mgCel6AΔCBM) performed worse at low substrate concentrations, but as the dry matter content in the reaction was increased, the catalytic domain eventually outperformed the full-length enzyme. The figure was adapted from Fig. 8 of Paper I.

3.2 Paper II – Discovery of a thermostable GH10 xylanase with broad substrate specificity from the Arctic Mid-Ocean Ridge vent system

Cellulose fibrils are frequently associated with different hemicelluloses in the plant cell wall. The hemicelluloses are usually partly removed by different pre-treatment methods, but residual hemicelluloses still represent obstacles that hamper the action of the cellulases. Hemicellulases can be employed to enzymatically remove different hemicelluloses from cellulosic substrates, and thereby work in synergy with the cellulases.

A putative xylanase from the GH10 family was discovered in a metagenome collected from hot vent sediments at the Arctic Mid-Ocean Ridge (AMOR). This deep-sea vent system represents an interesting biodiversity that has not yet been properly explored, and is thus likely to hold enzymes with novel properties. A chamber filled with an industrial sulfite-pulped spruce substrate was incubated in hot sediments (approximately 70 °C) near the deep-sea vents and collected after one year of deployment at the sea floor. After extraction of DNA from the incubation chamber, the metagenome of the bacteria that had been able to grow on the substrate was determined.

Metagenome mining revealed a putative GH10 (AMOR_GH10A) with relatively low sequence similarity to other GH10s (42 %) and only 31% sequence identity with the closest characterized relative (a cellulosomal xylanase, Xyn10D, from *Clostridium thermocellum;* Zverlov et al., 2005). Interestingly, the complete AMOR_GH10A protein seemed to contain a novel domain upstream of the GH10 (Fig. 26) that could not be classified into any of the established CAZy families using the dbCAN CAZyme annotation tool. The enzyme also had an inter-domain sequence that did not resemble a linker, as well as a C-terminal sequence without known function. These parts of the enzyme could perhaps affect enzyme activity and/or stability, as similar sequence stretches were also identified upstream and downstream of phylogenetically related GH10s.

Figure 26. Domain structure of AMOR_GH10A. The reading frame encodes a N-terminal signal peptide followed by a domain of unknown function and a catalytic GH10 domain separated by an inter-domain sequence that does not resemble a linker (i.e. lacks regions of low sequence complexity that are commonly found in linkers). The C-terminal region downstream of the GH10 domain could also not be assigned any specific function.

Full-length AMOR_GH10A as well as the domain of unknown function were expressed in *E. coli* with a C-terminal His-tag. The novel domain was tested for CBM-like characteristics and this showed that it was able to bind to various insoluble and soluble (hemi)cellulosic substrates (Fig. 27), but with a surprising preference for glucose backbones (of the tested substrates; see also below). BLASTp searches with the novel domain sequence revealed many "unknown" domains with sequence identities up to 39 % that are associated with GH10s and in some cases other GH-families. The proteins containing homologues of the AMOR_GH10A CBM frequently originated from halophilic and/or aquatic environments, including geothermal environments. This novel domain of AMOR_GH10A became the founding member of a novel CBM-family, CBM85 (www.cazy.org/CBM85.html).

AMOR_GH10A produced substituted and unsubstituted xylooligosaccharides from birchwood glucuronoxylan, confirming its activity as an (endo-)xylanase. The conditions at the sea floor near the vent system are characterized by high temperature and salinity. Using birchwood glucuronoxylan as substrate, it was shown that AMOR_GH10A expectedly exhibited high thermostability and halotolerance with an apparent T_m of 85 °C in the presence of 0.5 M NaCl, while in the absence of salt the apparent T_m decreased to 75 °C. Optimal product yields from glucuronoxylan were obtained at 80 °C and the activity was virtually abolished in the absence of NaCl, demonstrating that the enzyme is dependent on NaCl to function properly (at least at high temperatures).

The enzyme was shown to have a remarkably broad substrate specificity, also showing good activity on xyloglucan, glucomannan and PASC. The activity on β-glucan backbones was confirmed by kinetic analysis of oligosaccharide degradation, which showed much

better performance of AMOR_GH10A toward cellopentaose compared to xylopentaose, with k_{cat}/K_m values of 22.9 $min^{-1}mM^{-1}$ and 0.12 $min^{-1}mM^{-1}$, respectively. It is not uncommon for GH10 xylanases to display activity on e.g. cellulose but this activity is usually clearly lower, compared to xylan substrates, which makes AMOR_GH10A a unique exception. However, quantitatively, the activity of AMOR_GH10A was poor compared to other characterized xylanases, which could indicate that this enzyme, which originates from a largely unexplored biodiversity, may have evolved to interact with entirely different substrates, or that the consequence of having such a wide substrate specificity is overall lower activity.

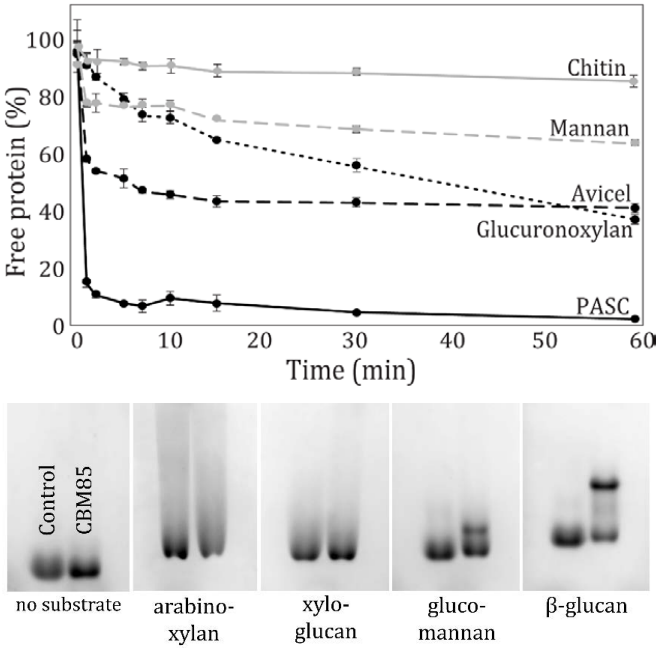

Figure 27. Binding properties of the founding member of the CBM85 family. The upper panel displays binding to insoluble substrates (the substrate names have been simplified, see details in Paper II) for the novel domain, and show a preference for glucose backbones (PASC and Avicel) followed by xylan backbones (glucuronoxylan). The lower panel shows binding to soluble substrates assessed by gel-shift assays using non-denaturing gels. The control gel to the left shows the migration of RNAse A (control protein) and the CBM in the absence of substrate. The other gels (where the respective substrates have been incorporated into the gels) show clear binding for the CBM to a glucose backbone (β-glucan), weak binding to glucomannan (mixed glucose and mannan backbone), and no binding to arabinoxylan and xyloglucan. The figure was adapted from Fig. 8 and Fig. 9 of Paper II.

The well-known xylanase A from *Cellulomonas fimi* (CfXyn10A, formerly known as Cex) has a weak cellulolytic side-activity for which two structural features have been suggested to be involved (Notenboom et al., 1998). Firstly, the conformation of a conserved glutamine (Gln87) in the -2 subsite differs compared to non-cellulose-active GH10s in that this residue attains a position that likely allows the accommodation of a glucose unit. While this glutamine residue is conserved in many GH10 xylanases with and without a cellulolytic side-activity, it is not conserved in AMOR_GH10A, which has a tryptophan in this position (Trp325), suggesting that the –2 subsite of AMOR_GH10A may look quite different. Secondly, a sequence element involved in shaping the –1 subsite and conserved in all structurally characterized GH10 xylanases to date is lacking in AMOR_GH10A. One of the residues in this region of CfXyn10A, Trp281, is thought to sterically hinder binding of glucosides and must rearrange to allow binding of such substrates. The absence of this entire region in AMOR_GH10A suggests that the -1 subsite looks quite different, and could imply that sterical hindrance of glucosides is not a problem in the –1 subsite of AMOR_GH10A.

The binding properties of the novel CBM85 partly coincide with the observed catalytic activity of the appended GH10 domain in that the CBM85 binds well to β-glucans and birch glucuronoxylan. However, we were not able to detect binding of the CBM to wheat arabinoxylan or tamarind xyloglucan (TXG), although AMOR_GH10A is active towards these substrates. The closest relatives of the CBM85s are certain CBM22 domains, such as CtCBM22-2 appended to a *C. thermocellum* xylanase. This CBM22 interacts with a wide range of xylans and xylooligosaccharides and residues that are key to substrate binding by this CBM, Trp53 and Tyr103, correspond to Trp73 and Trp122 in the CBM85 of AMOR_GH10A (see Fig. 7 in Paper II). While this conservation of aromatic surface residues suggests similarities, other surface residues of CBM22 are not all conserved in the CBM85, which could explain the differences in observed substrate-binding preferences. Considering the apparent discrepancies between the binding preferences of the CBM85 and the activity profile of AMOR_GH10A, it would be of interest to study substrate binding by the GH10 domain alone. However, repeated efforts to express the AMOR_GH10A catalytic domain alone failed.

3.3 Paper III – A thermostable bacterial AA10 LPMO with high operational stability in a wide temperature range.

LPMOs are important enzymes in bioprocessing as they can disrupt crystalline polysaccharide fibrils, thereby granting hydrolytic enzymes access to binding sites in parts of the substrate that these enzymes would otherwise struggle to reach. The powerful oxidative chemistry of LPMOs is a relatively recent discovery and multiple aspects of these enzymes remain insufficiently explored. LPMOs suffer from inactivation, not only due to usual inactivating factors such as extremes of pH or high temperatures, but also due to autocatalytic oxidative inactivation, as discussed in section 1.3 of the thesis. Advances in LPMO stability are thus of great interest.

Paper III describes the production and purification of a putative bacterial AA10 LPMO that had been identified as overexpressed in a compost/rice straw metagenome (Simmons et al., 2014; note that the LPMO is referred to as a CBM33 in this paper) . The LPMO was only overexpressed when the rice straw was incubated at high temperature (55 °C), and not during incubation at mesophilic temperature (35 °C). The protein consists of a catalytic AA10 domain that is linked to a CBM2 via a 38 residue threonine- and proline-rich linker region (Fig. 29).

Figure 29. Domain structure of mgLPMO10A. The enzyme comprises a 192-residue catalytic AA10 domain connected to a carbohydrate-binding module (CBM2) through a threonine- and proline-rich linker region.

The full-length LPMO (residues 32-363) and the catalytic domain (residues 32-223) were expressed using the native signal peptide of *Sm*LPMO10A to facilitate periplasmic expression. Before use, fractions of the purified LPMOs were copper-saturated by incubation with a three-fold molar access of $CuSO_4$, followed by removal of surplus copper by gel filtration chromatography (Loose et al., 2014).

Activity assays under “standard” LPMO conditions (1 mM ascorbic acid, aerobic) showed that mgLPMO10A was active on cellulose (PASC and Avicel). Analysis of

products generated from PASC showed that the LPMO oxidizes C1 in the scissile glycosidic bonds. During the further characterization of mgLPMO10A, a well-studied C1-oxidizing cellulose-active LPMO from the mesophilic bacterium *Streptomyces coelicolor* was used for comparison (Courtade et al., 2018, Forsberg et al., 2014a). Like mgLPMO10A, this LPMO, known as CelS2 or *Sc*LPMO10C, contains a CBM2. The two LPMOs share 62 % sequence identity.

Unfolding studies using differential scanning calorimetry (DSC) indicated that the apparent melting temperature of full-length mgLPMO10A was approximately 83 °C (Fig. 30). To the best of our knowledge, this is the highest apparent melting temperature recorded for an LPMO, and it is significantly higher than the apparent melting temperature of *Sc*LPMO10C, which is approximately 64 °C (see Paper IV). The apparent T_m decreased to approximately 73 °C when the copper was removed (by incubation with EDTA) to generate an apo-form of full-length mgLPMO10A. This demonstrates that the LPMO is stabilized when the transition metal is coordinated in the histidine brace. The unfolding curve for full-length mgLPMO10A in Fig. 30 shows a single transition which means that the two domains do not unfold independently or that they have approximately the same melting temperature. Removal of the CBM2 from mgLPMO10A affected the melting temperature negatively and yielded a T_m of approximately 76 °C. This indicates that the two domains may interact with each other in a way that confers protection at high temperatures, but could also be a result of an unfavorable configuration at the C-terminus caused by non-optimal truncation.

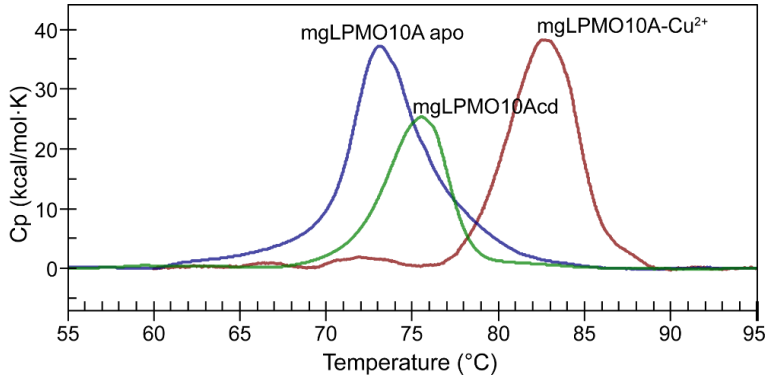

Figure 30. Thermal stability of mgLPMO10A. The Figure shows melting curves, as determined by differential scanning calorimetry (DSC), for copper-saturated mgLPMO10A, as well as the apo-form and the truncated version ("cd"). The protein was heated at 1 °C/min in 50 mM sodium phosphate buffer, pH 6.0. The figure is identical to Fig. 2 of Paper III.

Monitoring the degradation of Avicel at 40-80 °C showed that the copper-saturated mgLPMO10A is highly active up to temperatures of at least 80 °C (Fig. 31). Both *Sc*LPMO10C and mgLPMO10A produced oxidized products at a faster rate as the temperature was increased, but *Sc*LPMO10C became unstable at 70 °C and virtually no product was detected at 80 °C. The reactions with mgLPMO10A showed only minor enzyme inactivation at 80 °C. Of note, we do not know exactly how elevated temperature affects the reductant, the co-substrate, and the H_2O_2-generating oxidase activity of LPMOs. Thus, the product yields and apparent enzyme inactivation observed in Fig. 8 do not necessarily reflect only thermal stability or higher catalytic rates at elevated temperatures. For example, increased LPMO activity could result in faster H_2O_2 generation at the higher temperatures. The temperature could also have implications for substrate binding, which would affect both the oxidase activity of the LPMO and the sensitivity to autocatalytic inactivation. Further experiments are needed to resolve these issues.

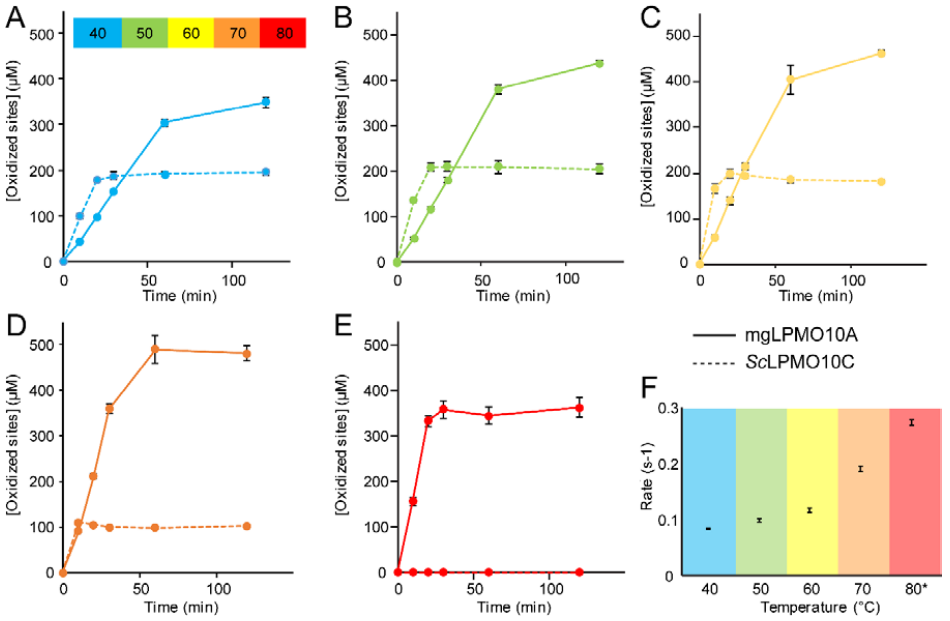

Figure 31. Activity at elevated temperatures. Panels A-E show product generation in reactions containing 10 g/L Avicel and 1 μM copper-saturated LPMO (mgLPMO10A or *Sc*LPMO10C) in the presence of 1 mM AscA, in 50 mM sodium phosphate buffer, pH 6.0, at temperatures ranging from 40 °C to 80 °C. Soluble oxidized products were converted to oxidized dimer and trimer by treatment with an endoglucanase prior to quantification. Panel F shows a summary of the effect of temperature on the initial rate of mgLPMO10A. The figure is identical to Fig. 3 of Paper III.

A thermostable LPMO could be valuable in industrial applications, and such applications would likely involve hydrolytic enzymes (cellulases and hemicellulases) that work in synergy with the LPMO. Metatranscriptomic analysis of the compost/rice straw sample from which mgLPMO10A was derived (Simmons et al., 2014) showed that mgLPMO10A was overexpressed together with a single cellulase, a putative GH48, when the community was incubated at high temperature. This GH48 (mgCel48A) and a putative GH6 (mgCel6B) were both produced as part of the initial phase of the study that led to paper I, as described above. Both cellulases are putative CBHs with CBM2s, supposedly attacking cellulose chains from the reducing and non-reducing end, respectively, and both showed a clear synergistic effect with mgLPMO10A at elevated temperature (60 °C).

MgLPMO10A was also tested in a spiking experiment with a commercial cellulase cocktail (Celluclast) that is assumed to be largely devoid of LPMO activity. In the experiment, 15 % of the enzyme cocktail (consisting of Celluclast mixed with a β-glucosidase in a 10:1 ratio on a protein basis) was replaced by mgLPMO10A. This setup generated clearly more products than 100 % of the cocktail, meaning that the presence of mgLPMO10A enables the cellulases to release more sugars (Fig. 32).

The synergy experiments were carried out with copper-saturated mgLPMO10A, but also with an untreated mgLPMO10A sample taken directly from the purification (referred to as mgLPMO10A-Cu^{2+} and mgLPMO10A^{pure}, respectively). Interestingly, both synergy experiments showed that the untreated LPMO was as beneficial to the product formation as the copper-saturated LPMO after industrially relevant processing times. However, the effect was smaller for the untreated LPMO at the earlier time points, suggesting that this enzyme works more slowly than the copper-saturated enzyme.

These results suggest that perhaps it could be beneficial to employ non copper-saturated LPMOs in bioprocessing settings. Industrial processing with cellulases is usually conducted over several days, while copper-saturated LPMOs often generate all the products within a few hours and are then inactivated, and are thus unable to assist the cellulases in attacking the remaining cellulose towards the end of the reaction. Copper-saturated LPMOs may be more exposed to autocatalytic inactivation because the ratio between reduced (and thus potentially self-destructive) LPMOs in solution and available substrate may become unfavourable.

The results described above drew our attention to the copper saturation procedure that is commonly used in LPMO research. The effect of copper saturation was further investigated in additional experiments with both the copper-saturated enzyme and the untreated enzyme, “as is”, i.e., used directly after the final purification step. Copper saturation is a common procedure prior to LPMO activity assays that aims to secure that every LPMO has a copper coordinated in the histidine brace and is thus capable of oxidative cleavage of polysaccharides.

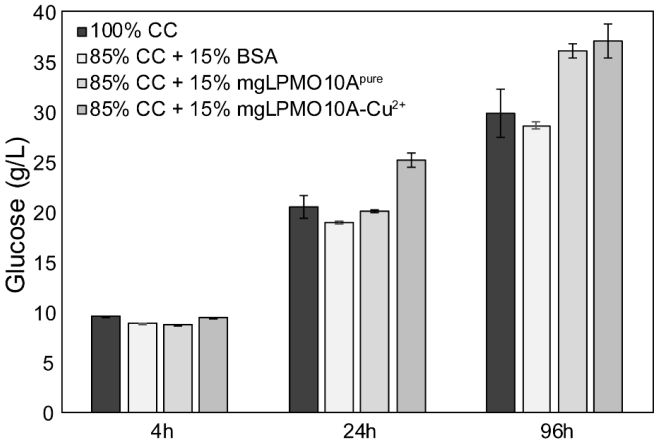

Figure 32. Synergy with cellulases and the effect of copper saturation. CC stands for the commercial enzyme cocktail Celluclast supplemented with a β-glucosidase (BG) in a 10:1 ratio. In the experiments, 15 % of the total loading of CC (4 mg/g glucan) was replaced by mgLPMO10A (copper-saturated or untreated) and used to degrade 50 g/L Avicel at 50 °C and pH 5.0 (i.e. the optimal conditions for the cocktail enzymes). The glucose yields (i.e. not counting oxidized LPMO products) showed that mgLPMO10A (both copper-saturated and untreated) significantly increased cellulose efficiency. The figure is identical to Fig. 5 of Paper III.

Time course experiments with the untreated LPMO (Fig. 33) showed that product generation was slow but steady (approximately 20 times slower than the reaction with copper-saturated LPMO at 60 °C; Fig. 31). The untreated LPMO was prone to inactivation at the higher temperatures (70 and 80 °C), thus supporting the DSC results (Fig 30), which showed that copper saturation improves thermostability. In the time course experiments, the untreated LPMO was almost immediately inactivated at 80 °C (Fig. 33), while the copper-saturated LPMO was highly active at 80 °C (Fig. 31). Note that the seemingly lower stability at 70 °C and 80 °C, compared to Figure 31 may be due to both much longer incubation times and the lack of the stabilizing effect provided by copper binding.

All in all, these observations suggest that the purified mgLPMO10A contains very little copper, meaning that most of the enzyme is in an inactive state. The synergy data discussed above show that this in fact may be favorable in certain settings. The presence of copper in mgLPMO10A was investigated using ICP-MS, which showed that the copper-saturated LPMO bound approximately 1 copper per molecule, but that there is

also a significant amount of free copper (approximately 0.35 per molecule) left after the copper saturation. It is important to note that the presence of a small amount free copper could in principle affect the LPMO reactions. Further studies on the effects of remaining free copper are currently in progress. Importantly, the ICP-MS data also showed that the untreated LPMO had only 0.07 copper per LPMO, which is consistent with the observations described above.

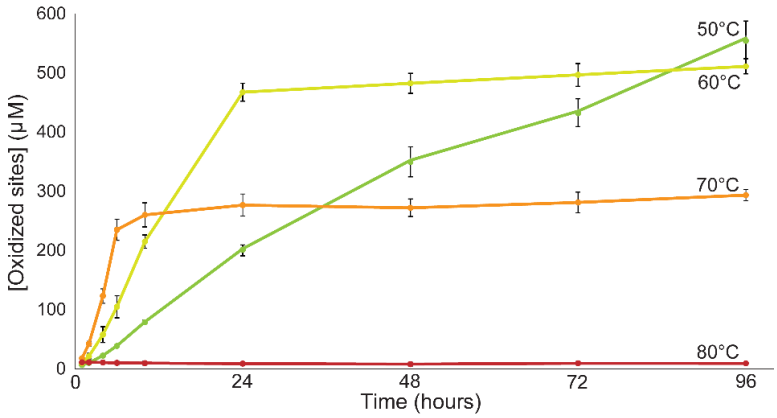

Figure 33. Progress curves using mgLPMO10A without copper saturation. The graphs show oxidized products generated by untreated mgLPMO10A from 10 g/L Avicel at 50 °C to 80 °C. The Figure shows that using an LPMO that had not been copper-saturated yielded stable and slow reaction kinetics. The reaction rate increased with the temperature but was, in all cases much lower than when using the copper-saturated enzyme. At 80 °C LPMO inactivation was almost immediate. The figure is identical to Fig. 6 of Paper III.

3.4 Paper IV – Tuning the substrate specificity of an AA10 LPMO from cellulose to chitin

The aim of the study described in Paper IV was to use a directed evolution approach to change the LPMO substrate specificity from cellulose to chitin, using a well-studied AA10 LPMO, *Sc*LPMO10C, which is produced naturally by the soil bacterium *Streptomyces coelicolor* A3 (Forsberg et al., 2014a), as a template. This study was a first step in the development of a pipeline for engineering properties of LPMOs, with the ultimate goal of increasing the applicability of these enzymes in biomass processing. The study involved the development of novel methods for high-throughput screening of LPMO action by SINTEF in Trondheim, Norway.

Residues that based on previous studies and sequence alignments seemed likely to be involved in substrate specificity were targeted in the design of the library, which comprised 4320 unique mutants, involving five residues in the substrate-binding surface. 11 520 colonies were screened using a mass spectrometry-based high-throughput assay and 37 positive (i.e., likely chitin-active) mutants were selected. The mutant genes were sequenced to reveal the mutations that conferred chitinolytic activity, leading to the identification of 27 unique mutation patterns. The mutation patterns were not necessarily straightforward to explain, demonstrating that large mutant libraries and high throughput screening methods can identify successful solutions that would most likely not have been considered otherwise.

Experimental work on a selection of the putatively chitin-active variants verified the chitinolytic activity. For the most active mutant (M18), the activity on β-chitin was 70-fold higher than in wild-type *Sc*LPMO10C (note that this is a rough estimate as the wild-type chitinolytic activity was close to the lower detection limit). The same mutant showed a 16-fold decrease in cellulose-activity (Fig. 34). None of the tested mutants showed activity towards α-chitin, the more crystalline (i.e. more challenging to degrade) and abundant chitin-allomorph. Perhaps α-chitin should also have been used to screen the library.

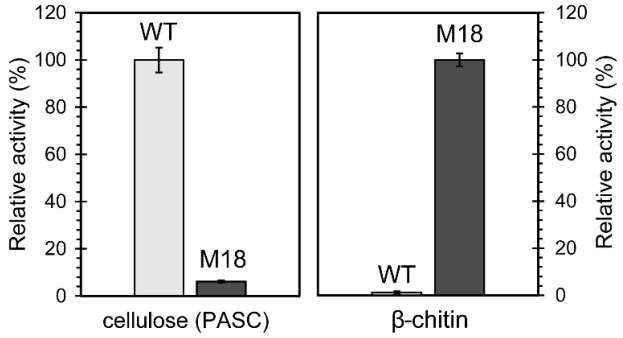

Figure 34. Shift in cellulolytic and chitinolytic activity in the best *Sc*LPMO10C mutant (M18). The figure illustrates that M18 showed a 16-fold decrease in cellulolytic activity and a 70-fold increase in chitinolytic activity relative to *Sc*LPMO10C wild-type. Product formation was assessed after four hours of incubation with 0.5 % (w/v) PASC or 1 % (w/v) β-chitin, respectively. The figure is identical to Fig. 8 of Paper IV.

When compared to a naturally chitinolytic AA10 LPMO (*Sm*LPMO10A), the mutants proved to be quite active during the initial phase of the reaction, however they had lower operational stability both in O_2 and H_2O_2 driven reactions (shown here for M18; Fig. 35). Based on the experimental data and other mutational studies on LPMOs (Loose et al., 2018, Forsberg et al., 2018, Bissaro et al., 2017), the increased rate of inactivation in M18 is likely related to suboptimal substrate binding. As discussed in section 1.3.3.1, LPMOs that are reduced but not interact productively with the substrate are prone to auto-oxidative damage due to the reaction of Cu(I) in the catalytic center with O_2, H_2O_2, or other oxygen species. It would thus seem that M18, while representing a considerable step forward in the engineering of LPMO substrate specificity, is not yet a "perfect" chitin-active LPMO.

The intrinsic capability of the wild-type enzyme for cleaving chitin is interesting although the product yield is close to the lower detection limit (likely explaining why this activity had not been detected before). It is likely that the catalytic domain of the wild-type binds much worse to chitin than the mutants generated in this study, and thus is inactivated very rapidly in the reactions by auto-oxidation as the enzyme is unable to coordinate the H_2O_2 correctly relative to the chitin scissile bond.

Fig. 35 shows that the O_2 driven reaction (A) is slower compared to the reaction where H_2O_2 is added at 15-minute intervals (B), as both the mutant and *Sm*LPMO10A have converted all the H_2O_2 to products before 7.5 minutes have passed (and most likely much faster than this, based on recent kinetic studies of H_2O_2 driven LPMO reactions; Kuusk et al., 2018). Assuming that H_2O_2 is the co-substrate of LPMOs, the rate of product generation in the O_2-driven reaction is likely limited by the rate of H_2O_2 formation during side reactions with O_2 and the LPMO or reductant, whereas the H_2O_2 driven reaction rate is limited by the amount of H_2O_2 added, although too high H_2O_2 amounts can lead to H_2O_2 being used to inactivate the LPMO during unproductive binding rather than product formation. Of note, experiments with higher dosages of H_2O_2 indeed led to increased enzyme inactivation, especially for M18, further highlighting the stability difference between this mutant and *Sm*LPMO10A.

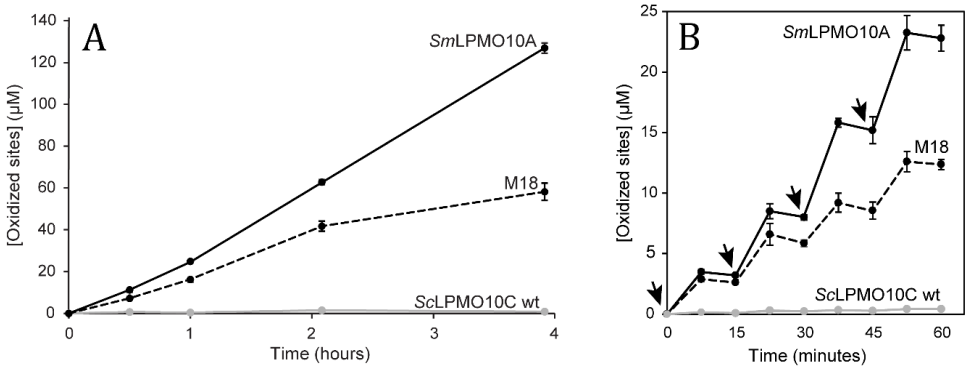

Figure 35. Progress curves and operational stability. The graphs show progress curves for the naturally chitin-active *Sm*LPMO10A, the naturally cellulose-active *Sc*LPMO10C wt, and a chitin-active mutant of *Sc*LPMO10C (M18). The panels show an O_2-driven reaction with 1 mM ascorbic acid added at the beginning of the reaction (A) and a H_2O_2-driven reaction with 15 μM H_2O_2 and 100 μM ascorbic acid added every 15 minutes (B). Note that these graphs only show soluble products and that the fraction of soluble versus insoluble oxidized products may change over time. Also note that, in the experiment of panel B, accumulating ascorbic acid may contribute to product formation, since the experiments were conducted aerobically. See Paper IV for further discussion. The figure was adapted from Fig. 3 and Fig. 5 of Paper IV.

Melting point assays showed that the thermal stability of the mutants had decreased significantly as a result of the mutations. This was a somewhat surprising result because there are only four mutations relative to the wild-type, and they all reside on the surface of the enzyme. Of note, incubation temperatures were such that it is not to be expected that thermal stability plays a role in shaping the progress curves depicted in Fig. 35.

In all three mutants selected for characterization (referred to as M2, M5 and M18), four of the five targeted residues were mutated, as illustrated in Fig. 36. Tyr79 was the only residue that was not altered in the chitin-active mutants, suggesting that the wild-type residue is preferred also for chitinolytic activity. The four mutations all contributed to the chitinolytic activity, as mutation of the individual residues back to wild-type sequence in all cases reduced (but did not abolish) the chitinolytic activity. This suggests that when combined, the mutations generate a substrate-binding surface that is more suited for binding correctly to chitin.

Three of the four M18 mutants still displayed a drastic decrease in cellulosic activity relative to the wild-type enzyme. However, one residue (Trp141 in the wild-type and Gln141 in M18) stood out because mutation of this residue back to wild-type sequence recovered much of the activity on cellulose.

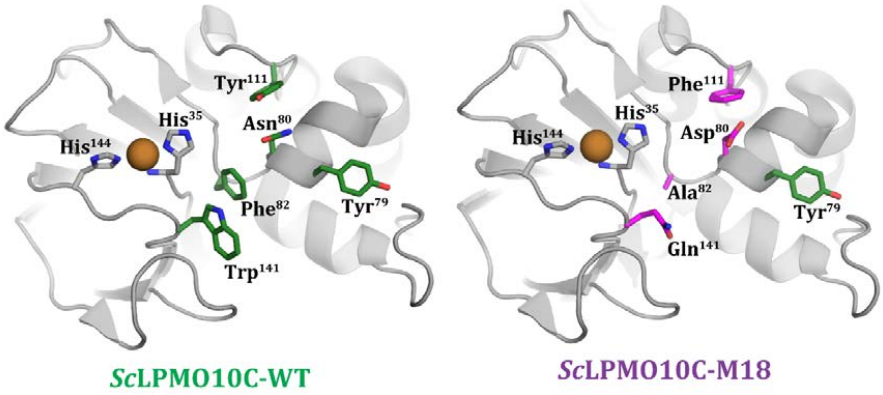

Figure 36. Structural differences between the *Sc*LPMO10C wild-type and M18. The figure displays the substrate-binding surface of *Sc*LPMO10C in the wild-type enzyme (left) and the best mutant M18 (right). The five targeted residues, as well as the two histidines (His35 and His144) that coordinate the copper (orange sphere), are displayed as sticks. Note that the two other characterized mutants, M2 and M5, differed slightly from M18 in that Ala82 was replaced by Gln82 in M2 and His82 in M5. The figure was adapted from Fig. S1 of Paper IV.

In early work on *Sc*LPMO10C it was shown that this enzyme binds strongly to chitin, in fact stronger than to its preferred substrate cellulose (Forsberg et al., 2014b). As shown in Paper IV, this binding is primarily mediated by the CBM as the catalytic domain of *Sc*LPMO10C bound much weaker that the full-length enzyme. Thus, it was not unexpected that M18, which contains the CBM, bound equally well to chitin as wild-type *Sc*LPMO10C. Considering the potential impact of substrate-binding on LPMO functionality, in particular operational stability, it would be of major interest to study the substrate binding properties of the catalytic domains of *Sc*LPMO10C and M18, as well as the control LPMO (single domain chitin-active *Sm*LPMO10A). Perhaps the strong binding of the CBM to chitin is actually unfavourable. If the catalytic domain is anchored to chitin through the CBM, but moves around without sufficiently rapidly finding productive binding sites, it could perhaps be a cause of rapid inactivation and explain why M18 is less stable than e.g. CBP21. Single domain M18 might not encounter the same problem. Unfortunately, despite multiple attempts, it has so far not been possible to recombinantly produce the catalytic domain of M18.

The study described in paper IV was successful in generating chitinolytic activity in a naturally cellulolytic LPMO, and it demonstrated how a directed evolution approach can reveal combinations of mutations that give a desired change in LPMO performance. Even though the main goal of altering the substrate specificity was achieved, the mutants do not seem to be optimized in regards to substrate binding, which affects the operational stability, and somehow the mutations also came at the expense of thermostability. Targeting additional residues could perhaps have provided better and more "complete" chitin-active variants without these stability issues. Importantly, the results described in Paper IV show a proof-of concept for MS-based screening of LPMO functionality, providing opportunities for directed evolution studies targeting other properties of these enzymes, such as stability.

4 CONCLUDING REMARKS

In conclusion, the work presented in this thesis demonstrates how modern bioprospecting approaches that employ metagenome mining can uncover novel enzymes with interesting activities and potential applications in industry. The properties of the enzymes described in Papers I-III reflect the habitats they were sampled from, including traits such as thermal stability and halotolerance.

Novel enzyme families can be discovered using this bioprospecting approach, as completely novel domains may be encoded together with domains that are recognized due to similarity to known enzymes. This was demonstrated by the identification of the founding member of the CBM85 family, which was discovered based on its association with a GH10 domain detected by the dbCAN metaserver for protein annotation.

Enzyme discovery could be further improved by functional screening of biodiversities, for example by screening fosmid libraries for biomass-degrading abilities, since this would allow discovery of unknown enzymes that are not yet in the databases and thus cannot be recognized by bioinformatics approaches. This approach was pursued in the NorZymeD project by functional screening of a fosmid library expressing random fragments of metagenomic DNA collected from an anaerobic high-temperature (65 °C) fermentation of sulfite-pulped spruce using an inoculum from a biogas reactor. The attempt was however not successful, because sequencing of the positive fosmids revealed only well-known cellulolytic enzymes.

The work included in this thesis offers some interesting data regarding aspects relevant to applied enzyme settings, in particular, the effect of CBMs on enzyme efficiency (Paper I), the effect of copper saturation on LPMO performance (Paper III) and the synergy between a novel, stable LPMO and both individual cellulases and a cellulase cocktail (Paper III).

Still, more could have been done. In work preceding the study described in Paper I, several thermally stable bacterial cellulases were produced that are not described in this thesis, except for two that were used in the work described Paper III. Of note, preliminary studies not reported in this thesis showed that several combinations of these bacterial enzymes show expected synergistic effects but did not reveal enzyme

combinations that, in terms of efficiency, could compete with current commercial enzyme cocktails composed primarily of fungal cellulases and LPMOs.

The xylanase described in Paper II shows peculiar characteristics, whereas the LPMO described in Paper III shows unprecedented thermal stability and boosts cellulase activity. Further exploration of the potential of these enzymes in biomass processing still needs to be done.

LPMOs are crucial for the efficiency of modern commercial cellulose cocktails (Müller et al., 2018, Müller et al., 2015, Johansen, 2016, Chylenski et al., 2019) but these enzymes retain many secrets and people have been claiming recently that their potential is not yet fully exploited (Bissaro et al., 2017, Eijsink et al., 2019, Forsberg et al., 2019). The NorZymeD project invested considerable resources in developing high-throughput methods for functional screening of LPMOs, and in the study described in Paper IV the developed protocols were successfully used to alter the substrate specificity of an LPMO. The altered substrate specificity of the mutants was accompanied by lower operational stability, revealing yet again the complexity of understanding LPMO functionality.

Importantly, the developed screening method will allow future screening efforts aimed at understanding more LPMO parameters, and at developing LPMO variants with improved industrially relevant properties, such as improved operational stability.

5 REFERENCES

AACHMANN, F. L., SØRLIE, M., SKJÅK-BRÆK, G., EIJSINK, V. G. & VAAJE-KOLSTAD, G. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 18779-18784.

AGGER, J. W., ISAKSEN, T., VARNAI, A., VIDAL-MELGOSA, S., WILLATS, W. G., LUDWIG, R., HORN, S. J., EIJSINK, V. G. & WESTERENG, B. 2014. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 6287-6292.

AGOSTONI, M., HANGASKY, J. A. & MARLETTA, M. A. 2017. Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases. *Microbiology and Molecular Biology Reviews,* 81.

ANITHA, A., SOWMYA, S., KUMAR, P. T. S., DEEPTHI, S., CHENNAZHI, K. P., EHRLICH, H., TSURKAN, M. & JAYAKUMAR, R. 2014. Chitin and chitosan in selected biomedical applications. *Progress in Polymer Science,* 39**,** 1644-1667.

ARANTES, V. & GOODEL, B. 2014. *C*urrent Understanding of Brown-Rot Fungal Biodegradation Mechanisms: A Review. Deterioration and Protection of Sustainable Biomaterials*. American Chemical Society.*

ARANTES, V. & SADDLER, J. N. 2011. Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. *Biotechnology for Biofuels,* 4**,** 3.

ARMSTRONG, Z., LIU, F., KHEIRANDISH, S., CHEN, H.-M., MEWIS, K., DUO, T., MORGAN-LANG, C., HALLAM, S. J. & WITHERS, S. G. 2019. High-Throughput Recovery and Characterization of Metagenome-Derived Glycoside Hydrolase-Containing Clones as a Resource for Biocatalyst Development. *mSystems,* 4**,** e00082-19.

ARNTZEN, M. Ø., VÁRNAI, A., MACKIE, R. I., EIJSINK, V. G. H. & POPE, P. B. 2017. Outer membrane vesicles from F*ibrobacter succinogenes* S85 contain an array of carbohydrate-active enzymes with versatile polysaccharide-degrading capacity. *Environmental Microbiology,* 19**,** 2701-2714.

ARTZI, L., BAYER, E. A. & MORAIS, S. 2017. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology,* 15**,** 83-95.

ATALLA, R. 1993. Trichoderma reesei cellulases and other hydrolases. *The structures of native cellulases. Foundation for biotechnical and industrial fermentation, Helsinki***,** 25-39.

AWRAMIK, S. M. 1992. The oldest records of photosynthesis. *Photosynthesis Research,* 33**,** 75-89.

BARBER, J. 2017. A mechanism for water splitting and oxygen production in photosynthesis. *Nature Plants,* 3, 17041

BAYER, E. A., BELAICH, J. P., SHOHAM, Y. & LAMED, R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Reviews in Microbiology,* 58**,** 521-554.

BAYER, E. A., LAMED, R., WHITE, B. A. & FLINT, H. J. 2008. From cellulosomes to cellulosomics. *The Chemical Record,* 8**,** 364-377.

BEESON, W. T., VU, V. V., SPAN, E. A., PHILLIPS, C. M. & MARLETTA, M. A. 2015. Cellulose degradation by polysaccharide monooxygenases. *Annual Reviews in Biochemistry,* 84**,** 923-946.

BELL, E. A., BOEHNKE, P., HARRISON, T. M. & MAO, W. L. 2015. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 14518-14521.

BENNATI-GRANIER, C., GARAJOVA, S., CHAMPION, C., GRISEL, S., HAON, M., ZHOU, S., FANUEL, M., ROPARTZ, D., ROGNIAUX, H., GIMBERT, I., RECORD, E. & BERRIN, J. G. 2015. Substrate

specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina. Biotechnology for Biofuels,* 8**,** 90.

BERNARDES, A., PELLEGRINI, V. O. A., CURTOLO, F., CAMILO, C. M., MELLO, B. L., JOHNS, M. A., SCOTT, J. L., GUIMARAES, F. E. C. & POLIKARPOV, I. 2019. Carbohydrate binding modules enhance cellulose enzymatic hydrolysis by increasing access of cellulases to the substrate. *Carbohydrate Polymers,* 211**,** 57-68.

BEVERIDGE, T. J. 1999. Structures of gram-negative cell walls and their derived membrane vesicles. *Journal of bacteriology,* 181**,** 4725-4733.

BIELY, P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advanves,* 30**,** 1575-1588.

BIELY, P., SINGH, S. & PUCHART, V. 2016. Towards enzymatic breakdown of complex plant xylan structures: State of the art. *Biotechnology Advances,* 34**,** 1260-1274.

BISCHOF, R. H., RAMONI, J. & SEIBOTH, B. 2016. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories,* 15**,** 106.

BISSARO, B., ISAKSEN, I., VAAJE-KOLSTAD, G., EIJSINK, V. G. H. & RØHR, Å. K. 2018a. How a Lytic Polysaccharide Monooxygenase Binds Crystalline Chitin. *Biochemistry,* 57**,** 1893-1906.

BISSARO, B., RØHR, A. K., MÜLLER, G., CHYLENSKI, P., SKAUGEN, M., FORSBERG, Z., HORN, S. J., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2017. Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. *Nature Chemical Biology,* 13**,** 1123-1128.

BISSARO, B., RØHR, Å. K., SKAUGEN, M., FORSBERG, Z., HORN, S. J., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2016. Fenton-type chemistry by a copper enzyme: molecular mechanism of polysaccharide oxidative cleavage. *bioRxiv***,** 097022.

BISSARO, B., VARNAI, A., RØHR, Å. K. & EIJSINK, V. G. H. 2018b. Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass. *Microbiology and Molecular Biology Reviews,* 82.

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. *Journal of Biological Chemistry,* 281**,** 36269-36279.

BOOK, A. J., YENNAMALLI, R. M., TAKASUKA, T. E., CURRIE, C. R., PHILLIPS, G. N. & FOX, B. G. 2014. Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases. *Biotechnology for Biofuels,* 7**,** 109.

BORASTON, A. B., BOLAM, D. N., GILBERT, H. J. & DAVIES, G. J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical journal,* 382**,** 769-781.

BRETT, C. 2000. Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. *Int Rev Cytol,* 199, 161-99

BUGG, T. D., AHMAD, M., HARDIMAN, E. M. & RAHMANPOUR, R. 2011. Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep,* 28**,** 1883-1896.

BUICK, R. 2008. When did oxygenic photosynthesis evolve? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 363**,** 2731-2743.

BUSK, P. K. & LANGE, L. 2015. Classification of fungal and bacterial lytic polysaccharide monooxygenases. *BMC genomics,* 16**,** 368-368.

CALZA, R. E., IRWIN, D. C. & WILSON, D. B. 1985. Purification and characterization of two .beta.-1,4-endoglucanases from *Thermomonospora fusca. Biochemistry,* 24**,** 7797-7804.

CANFIELD, D. E., ROSING, M. T. & BJERRUM, C. 2006. Early anaerobic metabolisms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 361**,** 1819-1836.

CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research,* 37**,** D233-8.

CHAKDAR, H., KUMAR, M., PANDIYAN, K., SINGH, A., NANJAPPAN, K., KASHYAP, P. L. & SRIVASTAVA, A. K. 2016. Bacterial xylanases: biology to biotechnology. *3 Biotech,* 6**,** 150-150.

CHANDRA, R. P., BURA, R., MABEE, W. E., BERLIN, A., PAN, X. & SADDLER, J. N. 2007. Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Advanced Biochemical Engineering Biotechnology,* 108**,** 67-93.

CHATURVEDI, V. & VERMA, P. 2013. An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. *3 Biotech,* 3**,** 415-431.

CHEN, K. & ARNOLD, F. H. 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences,* 90**,** 5618-5622.

CHERRY, J. R., LAMSA, M. H., SCHNEIDER, P., VIND, J., SVENDSEN, A., JONES, A. & PEDERSEN, A. H. 1999. Directed evolution of a fungal peroxidase. *Nature Biotechnology,* 17**,** 379-384.

CHYLENSKI, P., BISSARO, B., SØRLIE, M., RØHR, Å. K., VÁRNAI, A., HORN, S. J. & EIJSINK, V. G. H. 2019. Lytic Polysaccharide Monooxygenases in Enzymatic Processing of Lignocellulosic Biomass. *ACS Catalysis,* 9**,** 4970-4991.

COCINERO, E. J., GAMBLIN, D. P., DAVIS, B. G. & SIMONS, J. P. 2009. The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes. *Journal of the American Chemical Society,* 131**,** 11117-11123.

COLLINS, T., GERDAY, C. & FELLER, G. 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews,* 29**,** 3-23.

CONSORTIUM, T. C. 2017. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology,* 28**,** 3-8.

CORRÊA, T. L. R., JÚNIOR, A. T., WOLF, L. D., BUCKERIDGE, M. S., DOS SANTOS, L. V. & MURAKAMI, M. T. 2019. An actinobacteria lytic polysaccharide monooxygenase acts on both cellulose and xylan to boost biomass saccharification. *Biotechnology for Biofuels,* 12**,** 117.

COURTADE, G., FORSBERG, Z., HEGGSET, E. B., EIJSINK, V. G. H. & AACHMANN, F. L. 2018. The carbohydrate-binding module and linker of a modular lytic polysaccharide monooxygenase promote localized cellulose oxidation. *Journal of Biological Chemistry,* 293**,** 13006-13015.

COURTADE, G., WIMMER, R., RØHR, A. K., PREIMS, M., FELICE, A. K., DIMAROGONA, M., VAAJE-KOLSTAD, G., SØRLIE, M., SANDGREN, M., LUDWIG, R., EIJSINK, V. G. & AACHMANN, F. L. 2016. Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 5922-5927.

COUTURIER, M., LADEVEZE, S., SULZENBACHER, G., CIANO, L., FANUEL, M., MOREAU, C., VILLARES, A., CATHALA, B., CHASPOUL, F., FRANDSEN, K. E., LABOUREL, A., HERPOEL-GIMBERT, I., GRISEL, S., HAON, M., LENFANT, N., ROGNIAUX, H., ROPARTZ, D., DAVIES, G. J., ROSSO, M. N., WALTON, P. H., HENRISSAT, B. & BERRIN, J. G. 2018. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nature Chemical Biology,* 14**,** 306-310.

CRAGG, S. M., BECKHAM, G. T., BRUCE, N. C., BUGG, T. D., DISTEL, D. L., DUPREE, P., ETXABE, A. G., GOODELL, B. S., JELLISON, J., MCGEEHAN, J. E., MCQUEEN-MASON, S. J., SCHNORR, K., WALTON, P. H., WATTS, J. E. & ZIMMER, M. 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology,* 29**,** 108-119.

CUSKIN, F., LOWE, E. C., TEMPLE, M. J., ZHU, Y., CAMERON, E., PUDLO, N. A., PORTER, N. T., URS, K., THOMPSON, A. J., CARTMELL, A., ROGOWSKI, A., HAMILTON, B. S., CHEN, R., TOLBERT, T. J., PIENS, K., BRACKE, D., VERVECKEN, W., HAKKI, Z., SPECIALE, G., MUNOZ-MUNOZ, J. L., DAY, A., PENA, M. J., MCLEAN, R., SUITS, M. D., BORASTON, A. B., ATHERLY, T., ZIEMER, C. J., WILLIAMS, S. J., DAVIES, G. J., ABBOTT, D. W., MARTENS, E. C. & GILBERT, H. J. 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature,* 517**,** 165-169.

DALBY, P. A. 2011. Strategy and success for the directed evolution of enzymes. *Current Opinion in Structural Biology,* 21**,** 473-480.

DANA, C. M., SAIJA, P., KAL, S. M., BRYAN, M. B., BLANCH, H. W. & CLARK, D. S. 2012. Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnology and Bioengineering,* 109**,** 2710-2719.

DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure,* 3**,** 853-859.

DIETRICH, L. E., TICE, M. M. & NEWMAN, D. K. 2006. The co-evolution of life and Earth. *Current Biology,* 16**,** R395-400.

DISMUKES, G. C., KLIMOV, V. V., BARANOV, S. V., KOZLOV, Y. N., DASGUPTA, J. & TYRYSHKIN, A. 2001. The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences,* 98**,** 2170-2175.

DODD, M. S., PAPINEAU, D., GRENNE, T., SLACK, J. F., RITTNER, M., PIRAJNO, F., O'NEIL, J. & LITTLE, C. T. S. 2017. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature,* 543**,** 60.

DUCHESNE, L. C. & LARSON, D. W. 1989. Cellulose and the Evolution of Plant Life. *BioScience,* 39**,** 238-241.

EHRLICH, H., RIGBY, J. K., BOTTING, J. P., TSURKAN, M. V., WERNER, C., SCHWILLE, P., PETRÁŠEK, Z., PISERA, A., SIMON, P., SIVKOV, V. N., VYALIKH, D. V., MOLODTSOV, S. L., KUREK, D., KAMMER, M., HUNOLDT, S., BORN, R., STAWSKI, D., STEINHOF, A., BAZHENOV, V. V. & GEISLER, T. 2013. Discovery of 505-million-year old chitin in the basal demosponge V*auxia gracilenta*. *Scientific Reports,* 3**,** 3497.

EIJSINK, V., HOELL, I. & VAAJE-KOLSTAD, G. 2010. Structure and function of enzymes acting on chitin and chitosan. *Biotechnology & Genetic Engineering Reviews,* 27**,** 331-66.

EIJSINK, V. G., GASEIDNES, S., BORCHERT, T. V. & VAN DEN BURG, B. 2005. Directed evolution of enzyme stability. *Biomolecular Engineering,* 22**,** 21-30.

EIJSINK, V. G., VAAJE-KOLSTAD, G., VARUM, K. M. & HORN, S. J. 2008. Towards new enzymes for biofuels: lessons from chitinase research. *Trends in Biotechnology,* 26**,** 228-235.

EIJSINK, V. G. H., BJØRK, A., GÅSEIDNES, S., SIREVÅG, R., SYNSTAD, B., BURG, B. V. D. & VRIEND, G. 2004. Rational engineering of enzyme stability. *Journal of Biotechnology,* 113**,** 105-120.

EIJSINK, V. G. H., PETROVIC, D., FORSBERG, Z., MEKASHA, S., RØHR, A. K., VARNAI, A., BISSARO, B. & VAAJE-KOLSTAD, G. 2019. On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). *Biotechnology for Biofuels,* 12**,** 58.

ELHENAWY, W., DEBELYY, M. O. & FELDMAN, M. F. 2014. Preferential Packing of Acidic Glycosidases and Proteases into *Bacteroides* Outer Membrane Vesicles. *mBio,* 5**,** e00909-14.

ERIKSSON, K. E., PETTERSSON, B. & WESTERMARK, U. 1974. Oxidation: an important enzyme reaction in fungal degradation of cellulose. *FEBS Letters,* 49**,** 282-285.

FANUEL, M., GARAJOVA, S., ROPARTZ, D., MCGREGOR, N., BRUMER, H., ROGNIAUX, H. & BERRIN, J.-G. 2017. The *Podospora anserina* lytic polysaccharide monooxygenase PaLPMO9H catalyzes oxidative cleavage of diverse plant cell wall matrix glycans. *Biotechnology for Biofuels,* 10**,** 63.

FILIATRAULT-CHASTEL, C., NAVARRO, D., HAON, M., GRISEL, S., HERPOËL-GIMBERT, I., CHEVRET, D., FANUEL, M., HENRISSAT, B., HEISS-BLANQUET, S., MARGEOT, A. & BERRIN, J.-G. 2019. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnology for Biofuels,* 12**,** 55.

FORSBERG, C. W., BEVERIDGE, T. J. & HELLSTROM, A. 1981. Cellulase and Xylanase Release from *Bacteroides succinogenes* and Its Importance in the Rumen Environment. *Applied and Environmental Microbiology,* 42**,** 886-896.

FORSBERG, Z., BISSARO, B., GULLESEN, J., DALHUS, B., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2018. Structural determinants of bacterial lytic polysaccharide monooxygenase functionality. *Journal of Biological Chemistry,* 293**,** 1397-1412.

FORSBERG, Z., MACKENZIE, A. K., SØRLIE, M., RØHR, Å. K., HELLAND, R., ARVAI, A. S., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2014a. Structural and functional characterization of a

conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 8446-8451.

FORSBERG, Z., RØHR, A. K., MEKASHA, S., ANDERSSON, K. K., EIJSINK, V. G., VAAJE-KOLSTAD, G. & SØRLIE, M. 2014b. Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry,* 53**,** 1647-1656.

FORSBERG, Z., SØRLIE, M., PETROVIC, D., COURTADE, G., AACHMANN, F. L., VAAJE-KOLSTAD, G., BISSARO, B., RØHR, A. K. & EIJSINK, V. G. 2019. Polysaccharide degradation by lytic polysaccharide monooxygenases. *Current Opinion in Structural Biology,* 59**,** 54-64.

FORSBERG, Z., VAAJE-KOLSTAD, G., WESTERENG, B., BUNÆS, A. C., STENSTRØM, Y., MACKENZIE, A., SØRLIE, M., HORN, S. J. & EIJSINK, V. G. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Science,* 20**,** 1479-1483.

FRANDSEN, K. E., SIMMONS, T. J., DUPREE, P., POULSEN, J. C., HEMSWORTH, G. R., CIANO, L., JOHNSTON, E. M., TOVBORG, M., JOHANSEN, K. S., VON FREIESLEBEN, P., MARMUSE, L., FORT, S., COTTAZ, S., DRIGUEZ, H., HENRISSAT, B., LENFANT, N., TUNA, F., BALDANSUREN, A., DAVIES, G. J., LO LEGGIO, L. & WALTON, P. H. 2016. The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nature Chemical Biololgy,* 12**,** 298-303.

FROMMHAGEN, M., KOETSIER, M. J., WESTPHAL, A. H., VISSER, J., HINZ, S. W., VINCKEN, J. P., VAN BERKEL, W. J., KABEL, M. A. & GRUPPEN, H. 2016. Lytic polysaccharide monooxygenases from *Myceliophthora thermophila* C1 differ in substrate preference and reducing agent specificity. *Biotechnology for Biofuels,* 9**,** 186.

FROMMHAGEN, M., SFORZA, S., WESTPHAL, A. H., VISSER, J., HINZ, S. W. A., KOETSIER, M. J., VAN BERKEL, W. J. H., GRUPPEN, H. & KABEL, M. A. 2015. Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnology for Biofuels,* 8**,** 101.

GALBE, M. & ZACCHI, G. 2012. Pretreatment: The key to efficient utilization of lignocellulosic materials. *Biomass and Bioenergy,* 46**,** 70-78.

GARDNER, K. H. & BLACKWELL, J. 1975. Refinement of the structure of beta-chitin. *Biopolymers,* 14**,** 1581-1595.

GOMEZ DEL PULGAR, E. M. & SAADEDDIN, A. 2014. The cellulolytic system of *Thermobifida fusca*. *Critical Reviews in Microbiology,* 40**,** 236-247.

GROLEAU, D. & FORSBERG, C. W. 1981. Cellulolytic activity of the rumen bacterium *Bacteroides succinogenes*. *Canadian Journal of Microbiology,* 27**,** 517-530.

GROSS, A. S. & CHU, J.-W. 2010. On the Molecular Origins of Biomass Recalcitrance: The Interaction Network and Solvation Structures of Cellulose Microfibrils. *The Journal of Physical Chemistry B,* 114**,** 13333-13341.

GUO, M. & SONG, W. 2019. The growing U.S. bioeconomy: Drivers, development and constraints. *New Biotechnology,* 49**,** 48-57.

GUPTA, T. 2018. Coal, the Black Carbon. *In:* GUPTA, T. (ed.) *Carbon: The Black, the Gray and the Transparent.* Cham: Springer International Publishing.

HAGEN, L. H., FRANK, J. A., ZAMANZADEH, M., EIJSINK, V. G. H., POPE, P. B., HORN, S. J. & ARNTZEN, M. Ø. 2017. Quantitative Metaproteomics Highlight the Metabolic Contributions of Uncultured Phylotypes in a Thermophilic Anaerobic Digester. *Applied and Environmental Microbiology,* 83.

HAGHIGHI MOOD, S., HOSSEIN GOLFESHAN, A., TABATABAEI, M., SALEHI JOUZANI, G., NAJAFI, G. H., GHOLAMI, M. & ARDJMAND, M. 2013. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Reviews,* 27**,** 77-93.

HAKI, G. D. & RAKSHIT, S. K. 2003. Developments in industrially important thermostable enzymes: a review. *Bioresource Technology,* 89**,** 17-34.

HALL, M., BANSAL, P., LEE, J. H., REALFF, M. J. & BOMMARIUS, A. S. 2011. Biological pretreatment of cellulose: Enhancing enzymatic hydrolysis rate using cellulose-binding domains from cellulases. *Bioresource Technology,* 102**,** 2910-2915.

HAN, H., LING, Z., KHAN, A., VIRK, A. K., KULSHRESTHA, S. & LI, X. 2019. Improvements of thermophilic enzymes: From genetic modifications to applications. *Bioresource Technology,* 279**,** 350-361.

HANGASKY, J. A., IAVARONE, A. T. & MARLETTA, M. A. 2018. Reactivity of O_2 versus H_2O_2 with polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America,* 115**,** 4915-4920.

HARRIS, P. V., WELNER, D., MCFARLAND, K. C., RE, E., NAVARRO POULSEN, J. C., BROWN, K., SALBO, R., DING, H., VLASENKO, E., MERINO, S., XU, F., CHERRY, J., LARSEN, S. & LO LEGGIO, L. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry,* 49**,** 3305-3316.

HARUN, M. Y., DAYANG RADIAH, A. B., ZAINAL ABIDIN, Z. & YUNUS, R. 2011. Effect of physical pretreatment on dilute acid hydrolysis of water hyacinth (*Eichhornia crassipes*). *Bioresource Technology,* 102**,** 5193-5199.

HEDEGÅRD, E. D. & RYDE, U. 2018. Molecular mechanism of lytic polysaccharide monooxygenases. *Chemical Science,* 9**,** 3866-3880.

HEGNAR, O. A., PETROVIC, D. M., BISSARO, B., ALFREDSEN, G., VARNAI, A. & EIJSINK, V. G. H. 2019. pH-Dependent Relationship between Catalytic Activity and Hydrogen Peroxide Production Shown via Characterization of a Lytic Polysaccharide Monooxygenase from *Gloeophyllum trabeum*. *Applied and Environmental Microbiology,* 85**,** e02612-18.

HEMSWORTH, G. R., HENRISSAT, B., DAVIES, G. J. & WALTON, P. H. 2014. Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature Chemical Biology,* 10**,** 122-126.

HENRICH, E., DAHMEN, N., DINJUS, E. & SAUER, J. 2015. The Role of Biomass in a Future World without Fossil Fuels. *Chemie Ingenieur Technik,* 87**,** 1667-1685.

HERVÉ, C., ROGOWSKI, A., BLAKE, A. W., MARCUS, S. E., GILBERT, H. J. & KNOX, J. P. 2010. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 15293-15298.

HON, D. N.-S. 1994. Cellulose: a random walk along its historical path. *Cellulose,* 1**,** 1-25.

HÖÖK, M., BARDI, U., FENG, L. & PANG, X. 2010. Development of oil formation theories and their importance for peak oil. *Marine and Petroleum Geology*, 27, 1995-2004

HORN, S. J., SØRBOTTEN, A., SYNSTAD, B., SIKORSKI, P., SØRLIE, M., VÅRUM, K. M. & EIJSINK, V. G. H. 2006. Endo/exo mechanism and processivity of family 18 chitinases produced by Serratia marcescens. *The FEBS Journal,* 273**,** 491-503.

HORN, S. J., VAAJE-KOLSTAD, G., WESTERENG, B. & EIJSINK, V. G. 2012. Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels,* 5**,** 45-45.

HU, J., PRIBOWO, A. & SADDLER, J. N. 2016. Oxidative cleavage of some cellulosic substrates by auxiliary activity (AA) family 9 enzymes influences the adsorption/desorption of hydrolytic cellulase enzymes. *Green Chemistry,* 18**,** 6329-6336.

IGARASHI, K., UCHIHASHI, T., KOIVULA, A., WADA, M., KIMURA, S., OKAMOTO, T., PENTTILÄ, M., ANDO, T. & SAMEJIMA, M. 2011. Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science,* 333**,** 1279-1282.

IGARASHI, K., UCHIHASHI, T., UCHIYAMA, T., SUGIMOTO, H., WADA, M., SUZUKI, K., SAKUDA, S., ANDO, T., WATANABE, T. & SAMEJIMA, M. 2014. Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. *Nature Communications,* 5**,** 3975.

ISAKSEN, T., WESTERENG, B., AACHMANN, F. L., AGGER, J. W., KRACHER, D., KITTL, R., LUDWIG, R., HALTRICH, D., EIJSINK, V. G. & HORN, S. J. 2014. A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *Journal of Biological Chemistry,* 289**,** 2632-2642.

JALAK, J., KURAŠIN, M., TEUGJAS, H. & VÄLJAMÄE, P. 2012. Endo-exo synergism in cellulose hydrolysis revisited. *The Journal of biological chemistry,* 287**,** 28802-28815.

JAN, A. T. 2017. Outer Membrane Vesicles (OMVs) of Gram-negative Bacteria: A Perspective Update. *Frontiers in microbiology,* 8**,** 1053-1053.

JANG, M.-K., KONG, B.-G., JEONG, Y.-I., LEE, C. H. & NAH, J.-W. 2004. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. *Journal of Polymer Science Part A: Polymer Chemistry,* 42**,** 3423-3432.

JANSSEN, P. J., LAMBREVA, M. D., PLUMERE, N., BARTOLUCCI, C., ANTONACCI, A., BUONASERA, K., FRESE, R. N., SCOGNAMIGLIO, V. & REA, G. 2014. Photosynthesis at the forefront of a sustainable life. *Frontiers in Chemistry,* 2**,** 36.

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

KARKEHABADI, S., HANSSON, H., KIM, S., PIENS, K., MITCHINSON, C. & SANDGREN, M. 2008. The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 A resolution. *Journal of Molecular Biology,* 383**,** 144-154.

KARKEHABADI, S., HELMICH, K. E., KAPER, T., HANSSON, H., MIKKELSEN, N. E., GUDMUNDSSON, M., PIENS, K., FUJDALA, M., BANERJEE, G., SCOTT-CRAIG, J. S., WALTON, J. D., PHILLIPS, G. N., JR. & SANDGREN, M. 2014. Biochemical characterization and crystal structures of a fungal family 3 beta-glucosidase, Cel3A from *Hypocrea jecorina*. *Journal of Biological Chemistry,* 289**,** 31624-31637.

KIRN, T. J., JUDE, B. A. & TAYLOR, R. K. 2005. A colonization factor links *Vibrio cholerae* environmental survival and human infection. *Nature,* 438**,** 863-866.

KITTL, R., KRACHER, D., BURGSTALLER, D., HALTRICH, D. & LUDWIG, R. 2012. Production of four Neurospora crassa lytic polysaccharide monooxygenases in *Pichia pastoris* monitored by a fluorimetric assay. *Biotechnology for Biofuels,* 5**,** 79.

KJAERGAARD, C. H., QAYYUM, M. F., WONG, S. D., XU, F., HEMSWORTH, G. R., WALTON, D. J., YOUNG, N. A., DAVIES, G. J., WALTON, P. H., JOHANSEN, K. S., HODGSON, K. O., HEDMAN, B. & SOLOMON, E. I. 2014. Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 8797-8802.

KLEMM, D., HEUBLEIN, B., FINK, H. P. & BOHN, A. 2005. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition,* 44**,** 3358-93.

KOIVULA, A., REINIKAINEN, T., RUOHONEN, L., VALKEAJARVI, A., CLAEYSSENS, M., TELEMAN, O., KLEYWEGT, G. J., SZARDENINGS, M., ROUVINEN, J., JONES, T. A. & TEERI, T. T. 1996. The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. *Protein Engineering,* 9**,** 691-699.

KRACHER, D., SCHEIBLBRANDNER, S., FELICE, A. K., BRESLMAYR, E., PREIMS, M., LUDWICKA, K., HALTRICH, D., EIJSINK, V. G. & LUDWIG, R. 2016. Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science,* 352**,** 1098-1101.

KUBICKI, J. D., YANG, H., SAWADA, D., O'NEILL, H., OEHME, D. & COSGROVE, D. 2018. The Shape of Native Plant Cellulose Microfibrils. *Scientific Reports,* 8**,** 13983.

KULP, A. & KUEHN, M. J. 2010. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol,* 64**,** 163-184.

KUUSK, S., BISSARO, B., KUUSK, P., FORSBERG, Z., EIJSINK, V. G. H., SØRLIE, M. & VÄLJAMÄE, P. 2018. Kinetics of H_2O_2-driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase. *Journal of Biological Chemistry,* 293**,** 523-531.

LAIRSON, L. L., HENRISSAT, B., DAVIES, G. J. & WITHERS, S. G. 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry,* 77**,** 521-555.

LAMED, R., SETTER, E. & BAYER, E. A. 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of bacteriology,* 156**,** 828-836.

LANGSTON, J. A., SHAGHASI, T., ABBATE, E., XU, F., VLASENKO, E. & SWEENEY, M. D. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and Environmental Microbiology,* 77**,** 7007-7015.

LARSBRINK, J., ROGERS, T. E., HEMSWORTH, G. R., MCKEE, L. S., TAUZIN, A. S., SPADIUT, O., KLINTER, S., PUDLO, N. A., URS, K., KOROPATKIN, N. M., CREAGH, A. L., HAYNES, C. A., KELLY, A. G., CEDERHOLM, S. N., DAVIES, G. J., MARTENS, E. C. & BRUMER, H. 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature,* 506**,** 498-502.

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**,** 260.

LARSSON, A. M., BERGFORS, T., DULTZ, E., IRWIN, D. C., ROOS, A., DRIGUEZ, H., WILSON, D. B. & JONES, T. A. 2005. Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: the role of tyrosine Y73 in substrate ring distortion. *Biochemistry,* 44**,** 12915-12922.

LEHMANN, M. & WYSS, M. 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology,* 12**,** 371-375.

LEVASSEUR, A., DRULA, E., LOMBARD, V., COUTINHO, P. M. & HENRISSAT, B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels,* 6**,** 41.

LI, L.-L., TAGHAVI, S., MCCORKLE, S. M., ZHANG, Y.-B., BLEWITT, M. G., BRUNECKY, R., ADNEY, W. S., HIMMEL, M. E., BRUMM, P., DRINKWATER, C., MEAD, D. A., TRINGE, S. G. & LELIE, D. V. D. 2011. Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases. *Biotechnology for Biofuels,* 4**,** 23.

LI, X., BEESON, W. T. T., PHILLIPS, C. M., MARLETTA, M. A. & CATE, J. H. 2012. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure,* 20**,** 1051-1061.

LIU, S., ABRAHAMSON, L. P. & SCOTT, G. M. 2012. Biorefinery: Ensuring biomass as a sustainable renewable source of chemicals, materials, and energy. *Biomass and Bioenergy,* 39**,** 1-4.

LO LEGGIO, L., SIMMONS, T. J., POULSEN, J.-C. N., FRANDSEN, K. E. H., HEMSWORTH, G. R., STRINGER, M. A., VON FREIESLEBEN, P., TOVBORG, M., JOHANSEN, K. S., DE MARIA, L., HARRIS, P. V., SOONG, C.-L., DUPREE, P., TRYFONA, T., LENFANT, N., HENRISSAT, B., DAVIES, G. J. & WALTON, P. H. 2015. Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature communications,* 6**,** 5961-5961.

LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research,* 42**,** D490-D495.

LOOSE, J. S. M., ARNTZEN, M. O., BISSARO, B., LUDWIG, R., EIJSINK, V. G. H. & VAAJE-KOLSTAD, G. 2018. Multipoint Precision Binding of Substrate Protects Lytic Polysaccharide Monooxygenases from Self-Destructive Off-Pathway Processes. *Biochemistry,* 57**,** 4114-4124.

LOWE, D. R. & TICE, M. M. 2006. Hydrogen-based carbon fixation in the earliest known photosynthetic organisms. *Geology,* 34**,** 37-40.

LUTZ, S. 2010. Beyond directed evolution--semi-rational protein engineering and design. *Current opinion in biotechnology,* 21**,** 734-743.

LUTZ, S. & IAMURRI, S. M. 2018. Protein Engineering: Past, Present, and Future. *Methods Mol Biol,* 1685**,** 1-12.

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 : MMBR,* 66**,** 506-577.

MACKENZIE, A. K., NAAS, A. E., KRACUN, S. K., SCHÜCKEL, J., FANGEL, J. U., AGGER, J. W., WILLATS, W. G. T., EIJSINK, V. G. H. & POPE, P. B. 2015. A Polysaccharide Utilization Locus from an Uncultured Bacteroidetes Phylotype Suggests Ecological Adaptation and Substrate Versatility. *Applied and Environmental Microbiology,* 81**,** 187-195.

MADHAVAN, A., SINDHU, R., PARAMESWARAN, B., SUKUMARAN, R. K. & PANDEY, A. 2017. Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. *Applied Biochemistry and Biotechnology,* 183**,** 636-651.

MAHON, E. L. & MANSFIELD, S. D. 2019. Tailor-made trees: engineering lignin for ease of processing and tomorrow's bioeconomy. *Current Opinion in Biotechnology,* 56**,** 147-155.

MARTIN, W. F. & SOUSA, F. L. 2016. Early Microbial Evolution: The Age of Anaerobes. *Cold Spring Harbor perspectives in biology,* 8**,** a018127-a018127.

MAURYA, D. P., SINGLA, A. & NEGI, S. 2015. An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol. *3 Biotech,* 5**,** 597-609.

MAYES, H. B., KNOTT, B. C., CROWLEY, M. F., BROADBELT, L. J., STÅHLBERG, J. & BECKHAM, G. T. 2016. Who's on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolases. *Chemical science,* 7**,** 5955-5968.

MEINKE, A., DAMUDE, H. G., TOMME, P., KWAN, E., KILBURN, D. G., MILLER, R. C., JR., WARREN, R. A. & GILKES, N. R. 1995. Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *Journal of Biological Chemistry,* 270**,** 4383-4386.

MERINO, S. T. & CHERRY, J. 2007. Progress and challenges in enzyme development for biomass utilization. *Advances in Biochemical Engineering/Biotechnology,* 108**,** 95-120.

MINKE, R. & BLACKWELL, J. 1978. The structure of α-chitin. *Journal of Molecular Biology,* 120**,** 167-181.

MOJZSIS, S. J., ARRHENIUS, G., MCKEEGAN, K. D., HARRISON, T. M., NUTMAN, A. P. & FRIEND, C. R. 1996. Evidence for life on Earth before 3,800 million years ago. *Nature,* 384**,** 55-59.

MOJZSIS, S. J., HARRISON, T. M. & PIDGEON, R. T. 2001. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature,* 409**,** 178-181.

MORRIS, J. L., PUTTICK, M. N., CLARK, J. W., EDWARDS, D., KENRICK, P., PRESSEL, S., WELLMAN, C. H., YANG, Z., SCHNEIDER, H. & DONOGHUE, P. C. J. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences,* 115**,** E2274-E2283.

MÜLLER, G., CHYLENSKI, P., BISSARO, B., EIJSINK, V. G. H. & HORN, S. J. 2018. The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail. *Biotechnology for Biofuels,* 11**,** 209.

MÜLLER, G., VARNAI, A., JOHANSEN, K. S., EIJSINK, V. G. & HORN, S. J. 2015. Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. *Biotechnology for Biofuels,* 8**,** 187.

MURALEEDHARAN, M. N., ZOURARIS, D., KARANTONIS, A., TOPAKAS, E., SANDGREN, M., ROVA, U., CHRISTAKOPOULOS, P. & KARNAOURI, A. 2018. Effect of lignin fractions isolated from different biomass sources on cellulose oxidation by fungal lytic polysaccharide monooxygenases. *Biotechnology for Biofuels,* 11**,** 296.

NAAS, A. E., MACKENZIE, A. K., MRAVEC, J., SCHÜCKEL, J., WILLATS, W. G. T., EIJSINK, V. G. H. & POPE, P. B. 2014. Do Rumen *Bacteroidetes* Utilize an Alternative Mechanism for Cellulose Degradation? *mBio,* 5**,** e01401-14.

NISBET, E. G. & FOWLER, C. M. R. 1996. Some liked it hot. *Nature,* 382**,** 404-405.

NISBET, E. G. & SLEEP, N. H. 2001. The habitat and nature of early life. *Nature,* 409**,** 1083-1091.

NOTENBOOM, V., BIRSAN, C., WARREN, R. A., WITHERS, S. G. & ROSE, D. R. 1998. Exploring the cellulose/xylan specificity of the beta-1,4-glycanase cex from *Cellulomonas fimi* through crystallography and mutation. *Biochemistry*, 37, 4751-4758.

NOWICKA, B. & KRUK, J. 2016. Powered by light: Phototrophy and photosynthesis in prokaryotes and its evolution. *Microbiological Research,* 186-187**,** 99-118.

PASPALIARI, D. K., LOOSE, J. S., LARSEN, M. H. & VAAJE-KOLSTAD, G. 2015. *Listeria monocytogenes* has a functional chitinolytic system and an active lytic polysaccharide monooxygenase. *The FEBS Journal,* 282**,** 921-936.

PAULY, M. & KEEGSTRA, K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal,* 54**,** 559-568.

PAYNE, C. M., KNOTT, B. C., MAYES, H. B., HANSSON, H., HIMMEL, M. E., SANDGREN, M., STÅHLBERG, J. & BECKHAM, G. T. 2015. Fungal Cellulases. *Chemical Reviews,* 115**,** 1308-1448.

PAYNE, C. M., RESCH, M. G., CHEN, L., CROWLEY, M. F., HIMMEL, M. E., TAYLOR, L. E., SANDGREN, M., STÅHLBERG, J., STALS, I., TAN, Z. & BECKHAM, G. T. 2013. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceedings of the National Academy of Sciences,* 110**,** 14646-14651.

PEREIRA, D. A. & WILLIAMS, J. A. 2007. Origin and evolution of high throughput screening. *British journal of pharmacology,* 152**,** 53-61.

PETROVIC, D. M., BISSARO, B., CHYLENSKI, P., SKAUGEN, M., SØRLIE, M., JENSEN, M. S., AACHMANN, F. L., COURTADE, G., VARNAI, A. & EIJSINK, V. G. H. 2018. Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation. *Protein Science,* 27**,** 1636-1650.

PHILLIPS, C. M., BEESON, W. T., CATE, J. H. & MARLETTA, M. A. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chemical Biology,* 6**,** 1399-1406.

POLLEGIONI, L., TONIN, F. & ROSINI, E. 2015. Lignin-degrading enzymes. *The FEBS Journal,* 282**,** 1190-1213.

POON, D. K., WITHERS, S. G. & MCINTOSH, L. P. 2007. Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi Xylanase Cex through NMR spectroscopic analysis. *Journal of Biological Chemistry,* 282**,** 2091-2100.

QUINLAN, R. J., SWEENEY, M. D., LO LEGGIO, L., OTTEN, H., POULSEN, J. C., JOHANSEN, K. S., KROGH, K. B., JØRGENSEN, C. I., TOVBORG, M., ANTHONSEN, A., TRYFONA, T., WALTER, C. P., DUPREE, P., XU, F., DAVIES, G. J. & WALTON, P. H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences,* 108**,** 15079-15084.

REDDY, A. P., SIMMONS, C. W., D'HAESELEER, P., KHUDYAKOV, J., BURD, H., HADI, M., SIMMONS, B. A., SINGER, S. W., THELEN, M. P. & VANDERGHEYNST, J. S. 2013. Discovery of Microorganisms and Enzymes Involved in High-Solids Decomposition of Rice Straw Using Metagenomic Analyses. *PLOS ONE,* 8**,** e77985.

REESE, E. T., SIU, R. G. H. & LEVINSON, H. S. 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of bacteriology,* 59**,** 485-497.

RINAUDO, M. 2006. Chitin and chitosan: Properties and applications. *Progress in Polymer Science,* 31**,** 603-632.

RYE, C. S. & WITHERS, S. G. 2000. Glycosidase mechanisms. *Current Opinion in Chemical Biology,* 4**,** 573-580.

SABBADIN, F., HEMSWORTH, G. R., CIANO, L., HENRISSAT, B., DUPREE, P., TRYFONA, T., MARQUES, R. D. S., SWEENEY, S. T., BESSER, K., ELIAS, L., PESANTE, G., LI, Y., DOWLE, A. A., BATES, R., GOMEZ, L. D., SIMISTER, R., DAVIES, G. J., WALTON, P. H., BRUCE, N. C. & MCQUEEN-MASON, S. J. 2018. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nature communications,* 9**,** 756-756.

SAINI, J. K., SAINI, R. & TEWARI, L. 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech,* 5**,** 337-353.

SAITO, Y., OKANO, T., GAILL, F., CHANZY, H. & PUTAUX, J. L. 2000. Structural data on the intra-crystalline swelling of beta-chitin. *International Journal of Biological Macromolecules,* 28**,** 81-88.

SAMMOND, D. W., PAYNE, C. M., BRUNECKY, R., HIMMEL, M. E., CROWLEY, M. F. & BECKHAM, G. T. 2012. Cellulase Linkers Are Optimized Based on Domain Type and Function: Insights from Sequence Analysis, Biophysical Measurements, and Molecular Simulation. *PLOS ONE,* 7**,** e48615.

SAWADA, D., NISHIYAMA, Y., LANGAN, P., FORSYTH, V. T., KIMURA, S. & WADA, M. 2012. Water in Crystalline Fibers of Dihydrate β-Chitin Results in Unexpected Absence of Intramolecular Hydrogen Bonding. *PLOS ONE,* 7**,** e39376.

SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. *Annual Review of Plant Biology,* 61**,** 263-289.

SCHMIDT-DANNERT, C. & ARNOLD, F. H. 1999. Directed evolution of industrial enzymes. *Trends in Biotechnology,* 17**,** 135-136.

SCHOBERT, H. 2013. Chemistry of Fossil Fuels and Biofuels (Cambridge Series in Chemical Engineering). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511844188

SCHOPF, J. W. 1993. Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science,* 260**,** 640-646.

SIMMONS, C. W., REDDY, A. P., D'HAESELEER, P., KHUDYAKOV, J., BILLIS, K., PATI, A., SIMMONS, B. A., SINGER, S. W., THELEN, M. P. & VANDERGHEYNST, J. S. 2014. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnology for Biofuels,* 7**,** 495.

SIMMONS, T. J., FRANDSEN, K. E. H., CIANO, L., TRYFONA, T., LENFANT, N., POULSEN, J. C., WILSON, L. F. L., TANDRUP, T., TOVBORG, M., SCHNORR, K., JOHANSEN, K. S., HENRISSAT, B., WALTON, P. H., LO LEGGIO, L. & DUPREE, P. 2017. Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates. *Nature Communications,* 8**,** 1064.

SINNOTT, M. L. 1990. Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews,* 90**,** 1171-1202.

SISTA KAMESHWAR, A. K. & QIN, W. 2018. Structural and functional properties of pectin and lignin–carbohydrate complexes de-esterases: a review. *Bioresources and Bioprocessing,* 5**,** 43.

SLEEP, N. H. 2010. The Hadean-Archaean environment. *Cold Spring Harbor perspectives in biology,* 2**,** a002527-a002527.

SOMERVILLE, C., BAUER, S., BRININSTOOL, G., FACETTE, M., HAMANN, T., MILNE, J., OSBORNE, E., PAREDEZ, A., PERSSON, S., RAAB, T., VORWERK, S. & YOUNGS, H. 2004. Toward a systems approach to understanding plant cell walls. *Science,* 306**,** 2206-2211.

SPECIALE, G., THOMPSON, A. J., DAVIES, G. J. & WILLIAMS, S. J. 2014. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Current Opinion in Structural Biology,* 28**,** 1-13.

SPEZIO, M., WILSON, D. B. & KARPLUS, P. A. 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry,* 32**,** 9906-9916.

STEINER, K. & SCHWAB, H. 2012. Recent advances in rational approaches for enzyme engineering. *Computational and structural biotechnology journal,* 2**,** e201209010-e201209010.

STEMMER, W. P. C. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature,* 370**,** 389-391.

STROBEL, K. L., PFEIFFER, K. A., BLANCH, H. W. & CLARK, D. S. 2015. Structural insights into the affinity of Cel7A carbohydrate-binding module for lignin. *The Journal of Biological Chemistry,* 290**,** 22818-22826.

SUTHERLAND, I. W. 1995. Polysaccharide lyases. *FEMS Microbiology Reviews,* 16**,** 323-347.

TAHA, M., FODA, M., SHAHSAVARI, E., ABURTO-MEDINA, A., ADETUTU, E. & BALL, A. 2016. Commercial feasibility of lignocellulose biodegradation: possibilities and challenges. *Current Opinion in Biotechnology,* 38**,** 190-197.

TAN, T. C., KRACHER, D., GANDINI, R., SYGMUND, C., KITTL, R., HALTRICH, D., HALLBERG, B. M., LUDWIG, R. & DIVNE, C. 2015. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nature Communications,* 6**,** 7542.

THARANATHAN, R. N. & KITTUR, F. S. 2003. Chitin — The Undisputed Biomolecule of Great Potential. *Critical Reviews in Food Science and Nutrition,* 43**,** 61-87.

TICE, M. M. & LOWE, D. R. 2004. Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature,* 431**,** 549-552.

TIWARI, R., NAIN, L., LABROU, N. E. & SHUKLA, P. 2018. Bioprospecting of functional cellulases from metagenome for second generation biofuel production: a review. *Critical Reviews in Microbiology,* 44**,** 244-257.

VAAJE-KOLSTAD, G., FORSBERG, Z., LOOSE, J. S. M., BISSARO, B. & EIJSINK, V. G. H. 2017. Structural diversity of lytic polysaccharide monooxygenases. *Current Opinion in Structural Biology,* 44**,** 67-76.

VAAJE-KOLSTAD, G., HORN, S. J., SØRLIE, M. & EIJSINK, V. G. 2013. The chitinolytic machinery of *Serratia marcescens*--a model system for enzymatic degradation of recalcitrant polysaccharides. *The FEBS Journal,* 280**,** 3028-3049.

VAAJE-KOLSTAD, G., HOUSTON, D. R., RIEMEN, A. H., EIJSINK, V. G. & VAN AALTEN, D. M. 2005a. Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *Journal of Biological Chemistry,* 280**,** 11313-11319.

VAAJE-KOLSTAD, G., HORN, S. J., VAN AALTEN, D. M., SYNSTAD, B. & EIJSINK, V. G. 2005b. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *The Journal of Biological Chemistry,* 280**,** 28492-28497.

VAAJE-KOLSTAD, G., WESTERENG, B., HORN, S. J., LIU, Z., ZHAI, H., SØRLIE, M. & EIJSINK, V. G. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science,* 330**,** 219-222.

VAN AALTEN, D. M., SYNSTAD, B., BRURBERG, M. B., HOUGH, E., RIISE, B. W., EIJSINK, V. G. & WIERENGA, R. K. 2000. Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9-A resolution. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 5842-5847.

VAN NOORDEN, R. 2013. EU debates U-turn on biofuels policy. *Nature,* 499**,** 13-14.

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. *Frontiers in Plant Science,* 4**,** 174.

VARKI, A., CUMMINGS, R. D., AEBI, M., PACKER, N. H., SEEBERGER, P. H., ESKO, J. D., STANLEY, P., HART, G., DARVILL, A., KINOSHITA, T., PRESTEGARD, J. J., SCHNAAR, R. L., FREEZE, H. H., MARTH, J. D., BERTOZZI, C. R., ETZLER, M. E., FRANK, M., VLIEGENTHART, J. F., LÜTTEKE, T., PEREZ, S., BOLTON, E., RUDD, P., PAULSON, J., KANEHISA, M., TOUKACH, P., AOKI-KINOSHITA, K. F., DELL, A., NARIMATSU, H., YORK, W., TANIGUCHI, N. & KORNFELD, S. 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology,* 25**,** 1323-1324.

VARNAI, 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**,** 30.

VOCADLO, D. J. & DAVIES, G. J. 2008. Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology,* 12**,** 539-555.

VU, V. V., BEESON, W. T., PHILLIPS, C. M., CATE, J. H. & MARLETTA, M. A. 2014a. Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *Journal of the American Chemical Society,* 136**,** 562-565.

VU, V. V., BEESON, W. T., SPAN, E. A., FARQUHAR, E. R. & MARLETTA, M. A. 2014b. A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences,* 111**,** 13822-13827.

VUONG, T. V. & WILSON, D. B. 2010. Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnology and Bioengineering,* 107**,** 195-205.

WALTON, P. H. & DAVIES, G. J. 2016. On the catalytic mechanisms of lytic polysaccharide monooxygenases. *Current Opinion in Chemical Biology,* 31**,** 195-207.

WANG, B., JOHNSTON, E. M., LI, P., SHAIK, S., DAVIES, G. J., WALTON, P. H. & ROVIRA, C. 2018. QM/MM Studies into the H_2O_2-Dependent Activity of Lytic Polysaccharide Monooxygenases: Evidence for the Formation of a Caged Hydroxyl Radical Intermediate. *ACS Catalysis,* 8**,** 1346-1351.

WATANABE, M., MATSUZAWA, T. & YAOI, K. 2018. Rational protein design for thermostabilization of glycoside hydrolases based on structural analysis. *Applied Microbiology and Biotechnology,* 102**,** 8677-8684.

WESTERENG, B., CANNELLA, D., WITTRUP AGGER, J., JØRGENSEN, H., LARSEN ANDERSEN, M., EIJSINK, V. G. H. & FELBY, C. 2015. Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Scientific Reports,* 5**,** 18561.

WILLIAM SCHOPF, J. 2011. The paleobiological record of photosynthesis. *Photosynthesis research,* 107**,** 87-101.

WILSON, D. B. 2008. Three microbial strategies for plant cell wall degradation. *Annals of the New York Academy of Sciences,* 1125**,** 289-297.

WINGER, M., CHRISTEN, M. & VAN GUNSTEREN, W. F. 2009. On the Conformational Properties of Amylose and Cellulose Oligomers in Solution. *International Journal of Carbohydrate Chemistry,* 2009**,** 8.

WOLFENDEN, R., LU, X. & YOUNG, G. 1998. Spontaneous Hydrolysis of Glycosides. *Journal of the American Chemical Society,* 120**,** 6814-6815.

YANIV, O., FICHMAN, G., BOROVOK, I., SHOHAM, Y., EDWARD, B., LAMED, R., SHIMON, L. & FROLOW, F. 2014. Fine-structural variance of family 3 carbohydrate-binding modules as extracellular biomass-sensing components of *Clostridium thermocellum* anti-σI factors.

YOUNES, I. & RINAUDO, M. 2015. Chitin and Chitosan Preparation from Marine Sources. Structure, Properties and Applications. *Marine Drugs,* 13**,** 1133-1174.

ZHANG, J., SHI, H., XU, L., ZHU, X. & LI, X. 2015. Site-Directed Mutagenesis of a Hyperthermophilic Endoglucanase Cel12B from *Thermotoga maritima* Based on Rational Design. *PLoS One,* 10**,** e0133824.

ZVERLOV, V. V., SCHANTZ, N., SCHMITT-KOPPLIN, P. & SCHWARZ, W. H. 2005. Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. *Microbiology,* 151**,** 3395-3401.

Paper I

PLOS ONE

RESEARCH ARTICLE

Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome

Marianne S. Jensen[1], Lasse Fredriksen[1], Alasdair K. MacKenzie[1], Phillip B. Pope[1], Ingar Leiros[2], Piotr Chylenski[1], Adele K. Williamson[2], Tony Christopeit[2], Heidi Østby[1], Gustav Vaaje-Kolstad[1], Vincent G. H. Eijsink[1]*

1 Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway, **2** The Norwegian Structural Biology Centre, Department of Chemistry, UiT-The Arctic University of Norway, Tromsø, Norway

* vincent.eijsink@nmbu.no

OPEN ACCESS

Citation: Jensen MS, Fredriksen L, MacKenzie AK, Pope PB, Leiros I, Chylenski P, et al. (2018) Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome. PLoS ONE 13(5): e0197862. https://doi.org/10.1371/journal.pone.0197862

Editor: Jean-Guy Berrin, Institut National de la Recherche Agronomique, FRANCE

Received: January 12, 2018

Accepted: May 9, 2018

Published: May 24, 2018

Data Availability Statement: All relevant data are within the paper.

Funding: This work was funded by the Research Council of Norway through the NorZymeD project, project number 221568. PBP is supported by a European Research Council Starting Grant Fellowship (336355—MicroDE).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Enzymatic depolymerization of recalcitrant polysaccharides plays a key role in accessing the renewable energy stored within lignocellulosic biomass, and natural biodiversities may be explored to discover microbial enzymes that have evolved to conquer this task in various environments. Here, a metagenome from a thermophilic microbial community was mined to yield a novel, thermostable cellulase, named mgCel6A, with activity on an industrial cellulosic substrate (sulfite-pulped Norway spruce) and a glucomannanase side activity. The enzyme consists of a glycoside hydrolase family 6 catalytic domain (GH6) and a family 2 carbohydrate binding module (CBM2) that are connected by a linker rich in prolines and threonines. MgCel6A exhibited maximum activity at 85˚C and pH 5.0 on carboxymethyl cellulose (CMC), but in prolonged incubations with the industrial substrate, the highest yields were obtained at 60˚C, pH 6.0. Differential scanning calorimetry (DSC) indicated a $T_{m(app)}$ of 76˚C. Both functional data and the crystal structure, solved at 1.88 Å resolution, indicate that mgCel6A is an endoglucanase. Comparative studies with a truncated variant of the enzyme showed that the CBM increases substrate binding, while not affecting thermal stability. Importantly, at higher substrate concentrations the full-length enzyme was outperformed by the catalytic domain alone, underpinning previous suggestions that CBMs may be less useful in high-consistency bioprocessing.

Introduction

Cellulose is a major component of plant cell walls, where it contributes to the rigidity that enables the plant to stretch towards the sunlight. Estimates of worldwide annual plant biomass production amount to a staggering 170–200 billion tons, of which approximately 70% is cell wall material often referred to as lignocellulose [1]. Accordingly, lignocellulose represents the most abundant biomass on Earth and thus our largest renewable carbon reservoir, and holds great potential for sustainable energy production [2, 3]. Utilization of lignocellulose as a "green carbon" feedstock to replace oil-based commodities can generate second-generation

biofuels and higher-value products from low-value waste products in agriculture and forestry, and efficient enzymatic depolymerization is a key step in such biomass-based value creation [4–6].

Lignocellulosic biomass is composed of a complex matrix of polysaccharides (cellulose and various hemicellulose) and the aromatic polymer lignin. The ratio of these compounds differs between plant species, but generally amounts to 40–50% cellulose, 20–35% hemicellulose and 15–30% lignin [1]. Cellobiose is the repeating unit of cellulose, and consists of two glucose units rotated 180˚ relative to one another and joined by a β-1,4 glycosidic linkage [7]. Individual cellulose chains arrange into crystalline fibrils that are stabilized by extensive hydrogen bonding, and in the plant cell wall these fibres are embedded in a matrix of hydrophobic lignin and amorphous hemicelluloses [8, 9]. As a result, lignocellulose constitutes a recalcitrant structure that is highly resistant to degradation.

Microorganisms have evolved the ability to degrade lignocellulose by producing specialized enzymes. Various catalytic activities with different roles in lignocellulose conversion have been identified, reflecting the complexity and diversity of naturally occurring substrates. Deconstruction of the most recalcitrant of the polysaccharides, cellulose, is achieved by cellulases, which are hydrolytic enzymes belonging to various glycoside hydrolase (GH) families in the Carbohydrate Active enZymes database (CAZy; www.cazy.org). Cellulases are usually classified further as endoglucanases, which cleave internal glycosidic bonds, and cellobiohydrolases, which release cellobiose from accessible reducing or non-reducing chain ends. The cellobiose is finally converted to glucose by β-glucosidases [10]. Efficient degradation of lignocellulose requires additional enzymes including lytic polysaccharide monooxygenases (LPMOs), which employ an oxidative mechanism to break internal bonds in crystalline cellulose regions [11–15], and hemicellulases that degrade various hemicelluloses (e.g. xyloglucans, xylans, mannans and glucomannans) surrounding the cellulose fibrils [9]. All the enzymes involved may contain one or more carbohydrate binding modules (CBMs), which are thought to increase the effective enzyme concentration on the substrate surface by positioning the catalytic domains in proximity to the substrate [16, 17].

Due to the recalcitrance of lignocellulosic biomass, thermochemical pre-treatments (e.g. steam explosion or acidic/alkaline treatments) are implemented prior to enzymatic treatments, to accelerate industrial lignocellulose turnover in biorefineries. Such pre-treatments often require harsh conditions, which may lead to destruction of sugars and the production of various enzyme inhibitory substances [6, 18]. Because of this, and because it is generally considered favourable to carry out the enzymatic conversion processes at higher temperatures [19], there is a need for robust enzymes that originate from harsh natural environments, or that have been made robust by enzyme engineering. Today's commercial market for cellulases is dominated by fungal enzymes from species such as the mesophilic fungus *Trichoderma reesei* [4, 20]. The temperature optima of many fungal cellulases are in the 40–50˚C range, which is lower than what is commonly desired in industry [21]. Even today, enzyme costs remain a significant factor in industrial biomass processing [22, 23], so there is a general interest in discovering better enzymes.

In an attempt to find thermostable enzymes for cellulose processing, we have explored the metagenome of a thermostable microbial community derived from rice straw inoculated with compost and incubated at 55˚C [24]. This led us to express and characterize a 45 kDa two-domain thermostable bacterial cellulase comprised of a GH6 domain and a C-terminal CBM2 domain. We present the functional and structural characteristics of this enzyme, called mgCel6A, and assess its potential for use in high-temperature industrial degradation of sulfite-pulped lignocellulosic biomass (Norway spruce). With a view on the potential industrial application of mgCel6A, we have also assessed the role of the CBM and how this role depends on the dry matter concentration in the enzymatic reaction.

Materials and methods

Enzyme origin and homology modelling

Publicly available metagenome data accessible in the Joint Genome Institute IMG/M database (https://img.jgi.doe.gov/cgi-bin/m/main.cgi; IMG genome ID 2199352008) were mined for putatively lignocellulose active enzymes using dbCAN (csbl.bmb.uga.edu/dbCAN; [25]). The 1404 bp mgCel6A encoding gene (IMG gene ID 2200387098) was extracted from the metagenome, and was annotated using InterPro (www.ebi.ac.uk/interpro). BLASTp (blast.ncbi.nlm.nih.gov) and the PDB database (rcsb.org) were used to investigate similarities to known cellulases and to check for occurrence of expected active site residues, and structures of homologous proteins were visualized and inspected in PyMol (pymol.org).

Cloning, expression and purification of mgCel6A and mgCel6AΔCBM

The *mgCel6A* gene (codon-optimized for *Escherichia coli* expression) was synthesized (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and bp 106–1404 (omitting the predicted 35 amino acid signal peptide sequence) was amplified by PCR using Phusion DNA polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and suitable primers (Eurofins, Ebersberg, Germany). To facilitate further subcloning, the forward primer was 5'TAAGAAGGAGATATACTATGGCAGATAGCGCATTTTATGTTGAT3' where the underlined nucleotides represent an over-hang sequence for ligation-independent cloning (LIC; [26]). Two different reverse primers were employed, one to amplify the full-length gene (*mgcel6A*) with sequence 5'AATGGTGGTGATGATGGTGCGCGCTGGTACATGCACTACCATTCAG3', and one to amplify the catalytic domain alone (*mgcel6AΔCBM*) with sequence 5'AATGGTGGTGATGATGGTGCGCTGCTGCAATTGCCAGTTCATAT3'. The PCR products were purified using a Nucleospin Gel and PCR Clean-Up kit (Macherey-Nagel, Düren, Germany) and inserted into the pNIC-CH expression vector (AddGene, Cambridge, Massachusetts, USA) by LIC. As a result of this cloning strategy, the N-terminus of the (signal peptide-free) protein is extended with a methionine, while a seven residue His-tag (AHHHHHH) is added at the C-terminus. LIC was followed by heat shock transformation into chemically competent OneShot *E. coli* TOP10 cells (Invitrogen, Carlsbad, California, USA). The host strain was allowed to proliferate in Super Optimal broth with Catabolite repression (SOC) for 60 minutes prior to plating on Lysogenic Broth (LB) agar containing 50 μg/ml kanamycin and 5% sucrose, followed by incubation overnight at 37˚C. Single transformant colonies were inoculated in liquid LB containing 50 μg/ml kanamycin and incubated overnight at 37˚C. The plasmid was isolated from transformants using a NucleoSpin Plasmid kit (Macherey-Nagel), and the cellulase-encoding gene sequences were verified by Sanger sequencing (GATC, Konstanz, Germany). The isolated plasmids were subsequently transformed by heat shock into chemically competent OneShot BL-21 Star™ (DE3) *E. coli* cells (Invitrogen) and grown in SOC media as described above, before plating on LB agar containing 50 μg/ml kanamycin and overnight incubation at 37˚C. Transformant colonies were inoculated and grown in Terrific Broth (TB) containing 50 μg/ml kanamycin using a Harbinger system (Harbinger Biotechnology & Engineering, Markham, Canada) at 22˚C overnight. Protein expression was subsequently induced by addition of 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the cultures were incubated for another 24 hrs at 22˚C. The cell pellets were harvested by centrifugation at 5000 x *g* for 15 minutes (Beckman Coulter Brea, California, USA), followed by rapid cooling to minus 80˚C, after which the cells were resuspended in 50 mM Tris pH 8.0 with 500 mM NaCl and 5 mM imidazole. The cells were lysed using a Vibracell sonicator (Sonics & Materials Inc., Newtown, Connecticut, USA) with 5 seconds on/off pulses for 3 minutes at 30% amplitude while kept on ice, and the cell debris

was removed by centrifugation at 15,000 x *g* for 15 minutes. The cell-free protein extracts were filtrated using 0.45 μm syringe filters (Sarstedt, Nümbrecht, Germany) after which the proteins were purified by immobilized metal affinity chromatography (IMAC) using an Äkta pure chromatography system (GE HealthCare, Chicago, USA) equipped with a Ni^{2+} affinity His-Trap™ HP 5 ml column (GE HealthCare). The His-tagged proteins were eluted using a linear gradient of 5–500 mM imidazole in 50 mM Tris pH 8.0, 500 mM NaCl. Protein fractions were examined by SDS-PAGE (Bio-Rad, Hercules, California, USA), after which relevant fractions were pooled and concentrated using 10,000 MWCO (Molecular Weight Cut-Off) Vivaspin ultrafiltration tubes (Sartorius, Göttingen, Germany), with simultaneous buffer exchange to 20 mM Tris-HCl, pH 8.0. The purified proteins were stored at 4˚C. The protein concentration was estimated using the Bio-Rad Protein Assay (Bio-Rad) based on the Bradford method [27] or by measuring the A_{280} and using theoretical extinction coefficients (web.expasy.org/protparam) for calculating the concentrations. In both cases, a Biophotometer (Eppendorf, Hamburg, Germany) was used for measuring absorbance.

Crystallization, data collection, structure determination, and model refinement

Crystallization experiments were performed with a stock solution of the (His-tagged) GH6 catalytic domain (12 mg/ml, estimated by A_{280}) in 20 mM Tris-HCl pH 8.0. Initial crystallization conditions were screened using the vapour diffusion sitting drop method using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, California, USA). The plates were set up with 60 μl reservoir solutions and sitting drops with equal amounts of reservoir solution and protein stock solution in a total drop volume of 1 μl. During incubation at 20˚C, crystals appeared after about 5 weeks at conditions containing 1 M $(NH4)_2SO_4$, 0.1 M BisTris pH 5.5, and 1% PEG 3350. Crystals were harvested, transferred to a cryoprotectant solution consisting of the reservoir solution containing 27% Ethylene glycol, and subsequently flash cooled in liquid N_2. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) beamline ID30B. Data collection and processing statistics are presented in Table 1. The crystal structure was solved by molecular replacement using MolRep in the CCP4 program package [28] with 2boe as a search model (this is a single mutant of Cel6A from *Thermobifida fusca*; this is the *Tf*Cel6A structure with the highest available resolution, 1.15 Å; [29]). The initial refinement was executed in Refmac [30] followed by automated model improvement in Buccaneer [31]. The manual building was done in Coot [32] interspersed by cycles of refinement in Refmac and resulted in a final R_{work}/R_{free} of 15.83/20.44. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB; www.rcsb.org) with accession code 6FAO (S1). Figs presented in the results section were generated using Pymol (pymol.org). The DALI server (http://ekhidna2.biocenter.helsinki.fi/dali) was used to generate a structure-based alignment of mgCel6A (PDB ID: 6FAO), the homologous endoglucanase *Tf*Cel6A (PDB ID: 1TML; [33]), and the cellobiohydrolase *Tf*Cel6B (PDB ID: 4B4H; [34]). Residues missing from the PDB files due to poorly defined electron density were manually inserted in the structurally aligned sequences based on visual inspection of superimposed structures in Pymol. ESPript [35] was employed to visualize the final structure-based alignment and highlight features such as conserved residues and specific loop regions.

Apparent melting temperature ($T_{m(app)}$)

A Nano-Differential Scanning Calorimeter III (Calorimetry Sciences Corporation, Lindon, USA) was employed to determine the apparent melting temperatures of mgCel6A and mgCel6AΔCBM. The sample solutions contained approximately 1.5 mg/ml enzyme dialyzed

overnight at 4˚C against 75 mM phosphate-citrate buffer, pH 6.0, and were degassed (5 min) and filtered (0.22 μm) prior to loading the sample cell. Buffer from the dialysis, also degassed and filtered, was used to record buffer baselines prior to the protein scans. A scan rate of 1˚C/min from 20˚C to 90˚C was employed, and the experiments were carried out in duplicate, using freshly dialyzed enzyme for each scan. The data were analysed using the NanoAnalyze software (tainstruments.com); buffer baselines were subtracted from the protein scans.

Substrates

Enzyme activity was primarily evaluated on an industrial substrate derived from unbleached Norway spruce (*Picea abies*) through a sulfite pulping pre-treatment termed the BALI^{TM} process [36, 37], developed at Borregaard AS (Sarpsborg, Norway). The substrate had a glucan content of 88%, while hemicelluloses and acid insoluble lignin comprised the remaining 12%.

Table 1. Data collection and refinement statistics.

Space group	$P2_12_12_1$
Wavelength (Å)	0.976
Resolution range (Å)	57.91–1.88 (1.92–1.88)
Completeness (%)	99.3 (94.4)
Mean I/σ(I)	10.5 (1.8)
$R_{p.i.m.}$	0.064 (0.534)
Total No. of reflections	135306 (6962)
No. of unique reflections	27338 (1636)
R_{work}/R_{free}	15.83/20.44
Cell dimensions	
a, b, c (Å)	48.74, 78.68, 85.53
α, β, γ (˚)	90, 90, 90
Number of non-H atoms	
Protein	2108
Water	379
Other*	9
Total	2496
R.m.s. deviations	
Bonds (Å)	0.019
Angles (˚)	1.738
Average B factors (Å^2)	
Overall	17.34
Protein	16.60
Water	29.23
Other[a]	29.50
Ramachandran plot (# / %)	
Preferred	258 / 93.48
Allowed	17 / 6.16
Outliers[b]	1 / 0.36

Values in parentheses are for the outermost shell.

[a]A sulfate ion and one molecule of ethylene glycol were identified in electron density and added. The sulfate ion was likely introduced during crystallisation and the ethylene glycol molecule was probably introduced due to cryo protection.

[b]Residues in the Ramachandran plot were categorized according to the nomenclature used in WinCoot version 0.8.6.1.

https://doi.org/10.1371/journal.pone.0197862.t001

The substrate was dried at 40˚C overnight and the particle size was reduced in a planetary ball mill PM 100 (Retsch, Haan, Germany) followed by sieving through a 0.85 mm screen, to make the substrate amenable to use in small-scale reactions. Enzyme activity was also assessed on the cellulosic model substrates carboxymethyl cellulose (CMC) (Sigma-Aldrich, St. Louis, Missouri, USA), Avicel PH-101 (Sigma-Aldrich), filter paper (Whatman no.1) and phosphoric acid swollen cellulose (PASC) prepared from Avicel PH-101 according to [38], as well as on the hemicellulosic substrates konjac glucomannan (KGM), xylan and tamarind xyloglucan (all from Megazyme, Wicklow, Ireland). Cello-oligomers with degree of polymerization (DP) DP2-DP6 (Megazyme) were used as substrates for evaluating cleavage patterns.

Activity assays

Phosphate-citrate buffers in the pH range pH 3.0–8.0 were used in all activity assays. Activity assays were carried out in 96-well microtiter plates (Thermo Fisher Scientific) with plastic sealing for short incubations, and in 2 mL screw cap micro tubes (Sarstedt) for overnight incubations. The reaction mixtures were incubated in thermomixers (Eppendorf, Hamburg, Germany), and the enzyme concentration in activity assays was 1 μM unless stated otherwise. Reactions with insoluble substrates were stopped by boiling the samples for 10 minutes, and the soluble reaction products were collected by vacuum filtration using 0.45 μm 96-well filter plates (Merck Millipore, Darmstadt, Germany). Reactions with soluble substrates were stopped by addition of a double volume of 3,5-dinitrosalicylic acid (DNS reagent) or an equal volume of 0.1 M NaOH depending on the method used for product analysis (DNS or HPLC, respectively). Hydrolysis yields in reactions were analysed using the DNS reagent for detection of reducing ends [39], and standard solutions of cellobiose (the main end-product generated by mgCel6A from cellulose) were used for quantification. Experiments were performed in triplicates.

Optimal conditions

The optimal conditions for hydrolysis were initially assessed with 1% (w/v) CMC as substrate. The optimal temperature for activity was determined by comparing the yields from 10 minute incubations at temperatures ranging from 50˚C to 90˚C, and the optimal pH for activity was determined by comparing the yields from 10 minute incubations at pH 3.0–8.0. The optimal conditions for hydrolysis of the industrial sulfite-pulped spruce substrate were estimated in the same manner, using two different incubation times (15 minutes and 24 hours).

Thermal and pH stability

To assess thermostability and tolerance to acidic pH, the enzyme was pre-incubated at 4 μM concentration for up to 24 hours in phosphate-citrate buffers ranging from pH 4.0 to pH 6.0, at 60˚C, 65˚C or 70˚C. The enzyme samples were kept at 4˚C after pre-incubation. The residual activities at various time points were estimated by diluting the pre-incubated enzyme two-fold in 2% (w/v) CMC in water followed by 10 minutes of hydrolysis at the optimal temperature (85˚C) for the enzyme. The pHs were not adjusted after pre-incubation. Half-lives were determined by estimating the incubation time needed to reduce activity by 50%, compared to the activity of non-pre-incubated enzyme at the same pH.

Reactions with various substrates

Degradation of various substrates (CMC, Avicel, PASC, sulfite-pulped spruce, KGM, xylan, and tamarind xyloglucan), all at 1% (w/v) in phosphate-citrate buffer pH 6.0 by 1 μM mgCel6A was examined by incubating reaction mixtures at 60˚C and 600 rpm for 72 hours.

To determine the distribution of reducing ends in the soluble versus insoluble fraction after hydrolysis of cellulose, a filter paper assay based on the method described by Irwin et al. [40] was used. Disks of filter paper (4.0 mg, generated by a paper punch) were incubated with enzyme (0.5 μM) for 24 hours at 60˚C in phosphate citrate buffer, pH 6.0, before the soluble fraction (supernatant) and insoluble fraction (remaining filter paper) were separated by centrifugation (12.000 x *g* for 5 minutes). 400 μl of the supernatant was removed and boiled with DNS reagent. The remaining filter paper was washed five times in distilled water, after which it was resuspended in 400 μl of distilled water and boiled with DNS reagent.

The efficiency of 1 μM mgCel6A or mgCel6AΔCBM in degrading sulfite-pulped spruce at varying dry matter (DM) loading (0.5–10% DM) was assessed after 48 hrs of incubation at 60˚C and 1000 rpm. The conversion of 10% DM sulfite-pulped spruce by mgCel6A was determined with an enzyme loading of 8 mg enzyme per g cellulose for 72 hours at 1000 rpm, and the product yield was calculated according to Kristensen et al. [41]. The reactions were stopped by boiling the samples for 10 minutes, and the soluble products were separated from the insoluble substrates by vacuum filtration, as above. Activity on soluble cello-oligomers (DP2-6) was assessed using substrate concentrations of 0.1% (w/v) and an enzyme concentration of 1 μM; reaction mixtures were incubated at 60˚C for 18 hrs without shaking and the reactions were stopped by addition of an equal volume of 0.1 M NaOH. Products were stored at 4˚C until analysis by DNS [39], HPAEC-PAD, or MALDI-TOF MS (see below).

Product analysis by HPAEC-PAD

Glucose and soluble cello-oligomers (DP2-6) were analysed by high-performance anion-exchange chromatography (HPAEC) using an ICS3000 system (Dionex, ThermoScientific, San Jose CA, USA) equipped with pulsed amperometric detection (PAD) and a CarboPac PA1 column (Dionex). A multistep linear gradient was used to elute the products at 0.25 ml/min, going from 0.1 M NaOH to 0.1 M NaOH, 0.1 M sodium acetate (NaOAc) in 10 minutes, to 0.1 M NaOH, 0.14 M NaOAc in 5 minutes, to 0.1 M NaOH, 0.3 M NaOAc in 1 minute, and to 0.1 M NaOH, 1.0 M NaOAc in 2 minutes, before column reconditioning by applying 0.1 M NaOH for 11 minutes. Data collection and analysis were carried out with the Chromeleon 7.0 software, and a DP1-DP6 cello-oligomer standard was used to quantify the products.

Product analysis by MALDI-TOF MS

Product formation from hemicellulosic substrates was assayed qualitatively using an matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) UltrafleXtreme mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) equipped with a Nitrogen 337-nm laser. Reaction products (1 μl) were applied to an MTP 384 ground steel target plate TF (Bruker Daltonics) together with 2 μl of 9 mg/ml of 2,5-dihydroxybenzoic acid (DHB) dissolved in 30% acetonitrile, followed by air-drying. Spectra were collected using Bruker FlexControl software and analysed with Bruker flexAnalysis software.

Substrate binding

Enzyme binding to Avicel was examined using a slightly modified variant of the A_{280} method described by Vaaje-Kolstad and co-workers [42]. The binding mixtures contained 1% (w/v) Avicel and 0.1 mg/ml mgCel6A or mgCel6AΔCBM in phosphate-citrate buffer pH 6.0 and were incubated at 22˚C and 1000 rpm. Substrate binding was monitored by determining the A_{280} of the liquid fraction at various time points, using aliquots that were taken from the binding reactions and vacuum filtered over a 0.45 μm filter to separate unbound protein from

protein bound to substrate. Protein concentrations were calculated by using theoretical extinction coefficients.

Results and discussion

Sequence analysis and enzyme production

Metagenome mining led to the identification of mgCel6A in metagenomic data originating from a high-temperature (55˚C) rice straw/compost bioreactor. Studies have been published on the metagenome [24] and its metatranscriptome [43]. Studying this and a related mesophilic metagenome, Reddy and colleagues [24] showed that cellulases containing CBM2s, such as mgCel6A, were significantly overrepresented in the thermophilic rice straw/compost community. They also found that these CBM2-containing cellulase-encoding genes primarily belonged to the genus *Micromonospora*. Using the InterPro server, mgCel6A was predicted to comprise a signal peptide, a GH6 catalytic domain, a 40 residue proline- and threonine-rich linker, and a CBM2 (Fig 1A). Previously identified GH6 cellulases include both endoglucanases and cellobiohydrolases of mainly bacterial and fungal origin (CAZy). CBM2s are known to bind cellulose, and in some cases chitin or xylan [16].

Employing BLASTp, the closest relative of the GH6 domain was found to be an unpublished putative endoglucanase (KXK58956.1) from *Micromonospora rosaria*. The closest relative with a known structure in the PDB database is an endo-1,4-glucanase from *Thermobifidia fusca* (*Tf*Cel6A, formerly known as *Thermomonospora fusca* endocellulase E2; PDB ID: 1TML; 68% sequence identity; [33]). The PDB contains several more recent structures of *Tf*Cel6A, including a complex with a substrate analogue and single-mutation variants, with higher resolution (PDB ID: 2BOD, 2BOE, 2BOF, 2BOG; [29]).

The C-terminal CBM2 domain, which is connected to the GH6 domain by a 40-residue P/T-rich linker region (Fig 1A), has 67% sequence identity with the CBM of an unpublished endoglucanase from *Micromonospora echinaurantiaca* (SCG62086.1). As for homologues with known structure, the closest relative is a CBM associated with the bifunctional beta-1,4-xylanase/glucanase C_{ex} from *Cellulomonas fimi* (45% sequence identity; PDB ID: 1EXH; [44]). Notably, the 40-residue P/T linker in mgCel6A is long compared to its closest homologue with known structure, *Tf*Cel6A, where the catalytic domain and the CBM2 are connected by a 21-residue P/T/N linker.

Both mgCel6A and mgCel6AΔCBM were expressed in *E. coli*. SDS-PAGE showed that most of the protein was soluble after cell lysis, with only small remains in the cell pellet. The two proteins were easy to produce and purify and the yields of purified protein were approximately 500 mg and 350 mg per litre culture, respectively.

Structure of the GH6 catalytic domain

The crystal structure of the catalytic domain of mgCel6A was solved to 1.88 Å (PDB ID: 6FAO) and is a typical representative of the seven-stranded TIM barrel α/β fold (Fig 1C). Statistics are shown in Table 1. The final model contained residues 2–288 of the expressed 288-residue catalytic domain plus an extra alanine from the C-terminal His-tag. Ser84 and Ser85 in a glycine-rich surface loop (SGASSGGM) could not be modelled.

Fig 1B shows a structure-based sequence alignment of mgCel6A, the endocellulase *Tf*Cel6A (PDB ID: 1TML; [33]) and the cellobiohydrolase *Tf*Cel6B (PDB ID: 4B4H; [34]), highlighting conserved residues and the location of secondary structure elements in mgCel6A. Previous studies have shown that GH6 endoglucanases owe their open cleft structure to shortening and displacement of two surface loops compared to the corresponding loops in GH6 cellobiohydrolases. In the latter, elongated loops fold over the active site cleft and close it off to form a

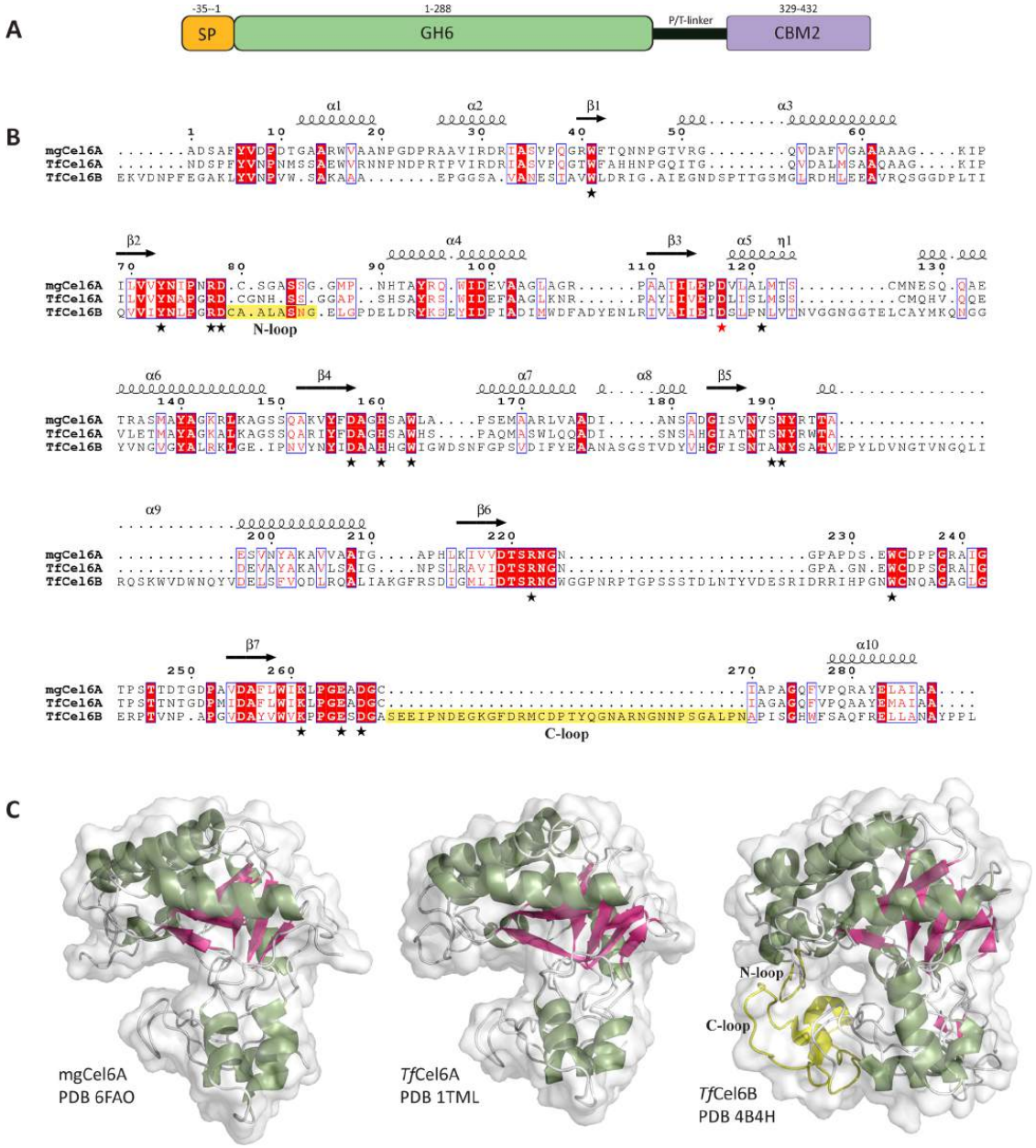

Fig 1. Sequence and structure of mgCel6A. (A) The domain structure of mgCel6A including a 35 residue signal peptide (SP), a 288 residue catalytic domain (GH6), a 40 residue proline- and threonine-rich linker, and a 104 residue carbohydrate binding module family 2 (CBM2). (B) Structure-based sequence alignment of the catalytic domains of mgCel6A (PDB ID: 6FAO), *Tf*Cel6A (PDB ID: 1TML), and *Tf*Cel6B (PDB ID: 4H4B). The secondary structure elements of mgCel6A are depicted above the aligned sequences. Red boxes show residues that are conserved in the three enzymes, while blue boxes indicate similar residues. A red star marks the catalytic acid, while black stars denote other residues potentially involved in catalysis and/or substrate-binding [29, 46, 47]; all these are shown in Fig 2A, 2C and 2E. Residues highlighted in

yellow indicate two loops that differ between endocellulases and cellobiohydrolases in the bacterial GH6 family (see text for more details). (C) Side-view of the catalytic domain of mgCel6A (PDB ID: 6FAO), *Tf*Cel6A (PDB ID: 1TML), and *Tf*Cel6B (PDB ID: 4B4H), coloured by secondary structure (green α-helices and pink β-sheets), and with transparent surface representation. The amino-proximal (N) and carboxyl-proximal (C) loop regions of *Tf*Cel6B that are responsible for tunnel formation are displayed in yellow. Note that one end of the active site tunnel in *Tf*Cel6B can be closed off by a flexible loop that is not visible in the structure displayed here; for more details see [34].

https://doi.org/10.1371/journal.pone.0197862.g001

tunnel, which likely restricts the enzyme to attacking the chain ends of cellulose, while promoting processivity [45]. Fig 1B and 1C highlight two loop regions termed the N-loop (amino-proximal) and C-loop (carboxyl-proximal). The N-loop is highly flexible (Fig 2) in both mgCel6A and *Tf*Cel6A to the extent that two of its residues could not be modelled in the mgCel6A structure. The N-loop has a different conformation and is slightly extended in *Tf*Cel6B, which together with a drastically extended C-loop (Fig 1B) gives the active site of this enzyme a tunnel-like character (Fig 1C). MgCel6A has a deep but open substrate-binding cleft similar to the one observed in *Tf*Cel6A, which suggests that mgCel6A, like *Tf*Cel6A, is an endoglucanase.

Residues known to be involved in catalysis and exposed residues that shape the substrate-binding catalytic cleft of *Tf*Cel6A [29, 46, 47] are marked with stars in the structure-based sequence alignment of Fig 1B, which shows that most of these residues are conserved in mgCel6A and to a somewhat lesser extend in *Tf*Cel6B. Based on the analogy with *Tf*Cel6A, Asp117 is the catalytic acid in mgCel6A, while the catalytic base remains elusive, as is the case for GH6 cellulases in general [48].

Fig 2 presents a closer view on the catalytic clefts of structures of *Tf*Cel6A and mgCel6A and includes the structure of *Tf*Cel6A bound to the substrate analogue methyl cellobiosyl-4-thio-beta-cellobioside [29]. Generally, the clefts look similar except for Arg78 and Asp79 whose conformations vary between the three structures. Notably, these two residues are adjacent to the flexible N-loop (high B-factors; Fig 2B, 2D and 2F). Asp79 has been proposed as the catalytic base in *Tf*Cel6A [47], even though this aspartate does not reside within hydrogen-bonding distance of a water molecule that could act as a nucleophile on the scissile bond. Based on simulations of the cellobiohydrolase Cel6A from *Trichoderma reesei* (*Tr*Cel6A), Mayes and co-workers recently proposed that Asp175 in *Tr*Cel6A (corresponding to Asp79 in mgCel6A) may act as a more "remote" catalytic base by coordinating a short water wire that comprises two water molecules aligned between the scissile bond and the aspartate [49].

The variation in the conformations of Arg78 and Asp79 and the high B-factors of the N-loop clearly suggest that these residues and the N-loop change their conformation during substrate binding and/or catalysis. However, determination of crystal structures of *Tf*Cel6A with and without a bound substrate analogue did not show conformational changes nor rigidification of this loop upon substrate-binding [29] (Fig 2B, 2D and 2F). In fact, as visible in Fig 2, in the structure of *Tf*Cel6A with a bound substrate analogue, as many as six residues of the N-loop could not be modelled, as opposed to only two or even zero in the ligand-free structures of mgCel6A and *Tf*Cel6A, respectively. Modelling of the substrate-analogue into the catalytic clefts of ligand-free mgCel6A and *Tf*Cel6A (Fig 2) further emphasized the structural variation and the remaining uncertainty concerning the role of Asp79. In the structure of the complex, the closest distance between Asp79 and the scissile bond is 10,6 Å compared to 9,1 Å and 7,7 Å in the models of ligand binding *Tf*Cel6A and mgCel6A, respectively.

Apparent melting temperature ($T_{m(app)}$)

The apparent melting points of the two enzyme variants mgCel6A and mgCel6AΔCBM were estimated with differential scanning calorimetry (DSC) by monitoring the change in heat

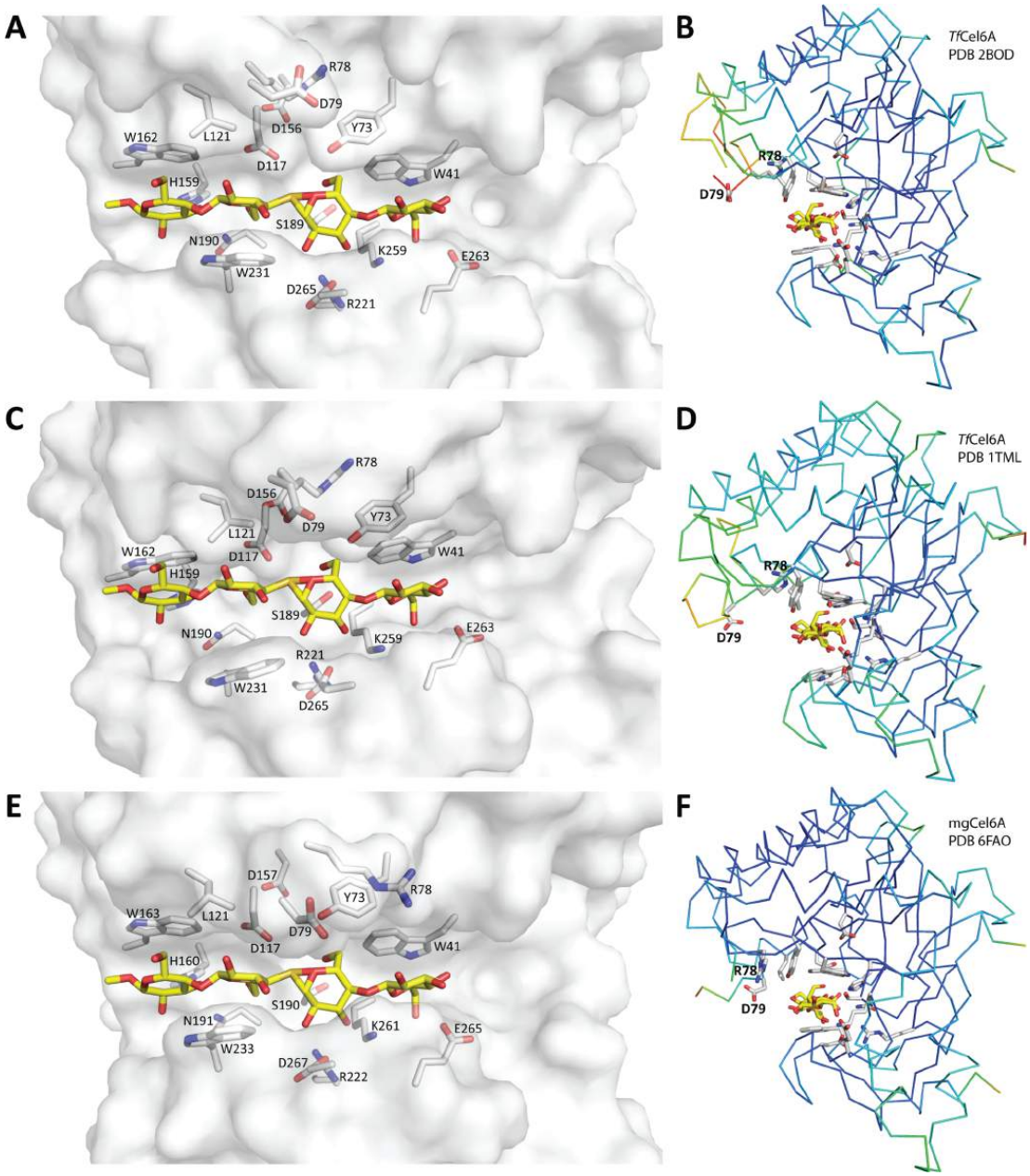

Fig 2. Comparison of the catalytic clefts of mgCel6A and *Tf*Cel6A. All pictures contain the substrate analogue methyl cellobiosyl-4-thio-beta-cellobioside which was modelled on the basis of superimposition with the crystal structure of *Tf*Cel6A in complex with this analogue (panel A, D; PDB ID: 2BOD), using Pymol. Panels A, C and E show transparent surface representations of the catalytic clefts of *Tf*Cel6A crystallized with ligand, apo *Tf*Cel6A with modelled ligand and apo mgCel6A

with modelled ligand. The side chains of residues in the catalytic cleft (marked in Fig 1B) are represented by grey sticks, while the substrate analogue is represented by yellow sticks. Panels B, D and F show ribbon representations of the same three proteins, coloured according to B-factor (blue means low; red means high). Note that the N-loop lacks residues in two of the three structures, explaining the additional chain ends in this region (see text for details).

https://doi.org/10.1371/journal.pone.0197862.g002

capacity along a temperature gradient. Both enzyme variants showed irreversible unfolding, which was not fully two-state, as shown by the shoulders at approximately 65˚C (Fig 3). While removal of the CBM had some effect on the shape of this shoulder, the main unfolding event was not affected and both enzyme variants exhibited a $T_{m(app)}$ of 76˚C. Since the "shoulder peak" was present for both mgCel6A and mgCel6AΔCBM, it cannot be attributed to independent denaturation of the CBM. It is possible, however, that the shoulder seen for mgCel6A relates to the CBM, whereas the different shoulder seen for mgCel6AΔCBM reflects another irregularity, for example related to fraying at the newly generated C-terminal end. Although DSC data derived from irreversible unfolding processes must not be over-interpreted [50, 51], the data clearly show that mgCel6A and mgCel6AΔCBM are stable enzymes. Notably,

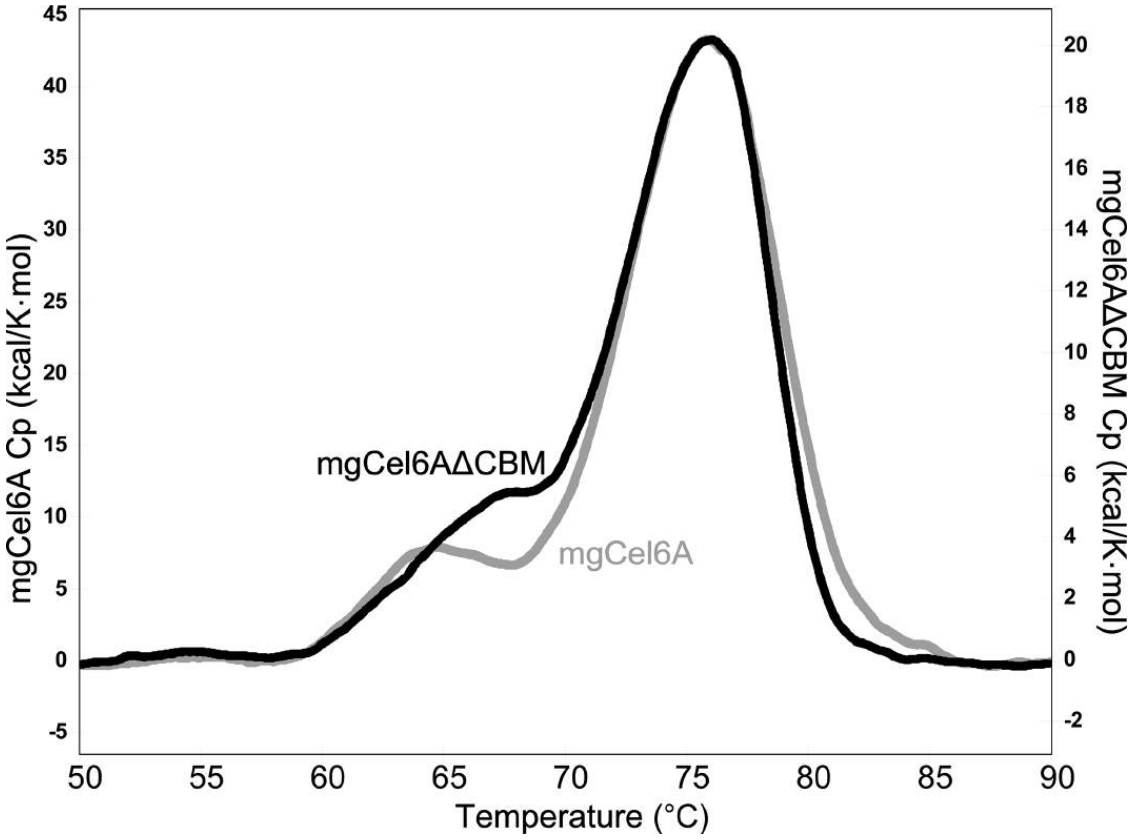

Fig 3. DSC thermograms for mgCel6A and mgCel6AΔCBM. The molar heat capacity (Cp) after buffer baseline subtraction and sigmoidal baseline fitting is plotted against the temperature. The thermograms are displayed as overlay graphs with individual y-axes. The protein samples (pH 6.0) were heated at a rate of 1˚C/min.

https://doi.org/10.1371/journal.pone.0197862.g003

preliminary unfolding experiments based on measuring intrinsic fluorescence indicated similar $T_{m(app)}$ values (data now shown).

Optimal conditions for activity

When using CMC and short incubation times (10 min), the full-length enzyme performed best at 85˚C and pH 5.0 (Fig 4A and 4B). For sulfite-pulped spruce, the optimal reaction conditions were 75˚C and pH 6.0 (Fig 4C and 4D). The differences in the optimal temperature and pH displayed in in Fig 4 and the results of similar assays using different conditions (pH, incubation time; not shown) suggest that stability comes into play at temperatures as high as 75˚C and pH values below 6.0, as confirmed by the results described below. In any case, the results depicted in Fig 4 confirm that the enzyme is highly thermoactive.

Operational enzyme stability

Hydrolysis of sulfite-pulped spruce in overnight reactions (24 hours) showed the highest yields for reactions run at pH 6.0 and 60˚C (Fig 5A and 5B). The lower optimum temperature observed here shows that stability comes into play when using more industrially relevant

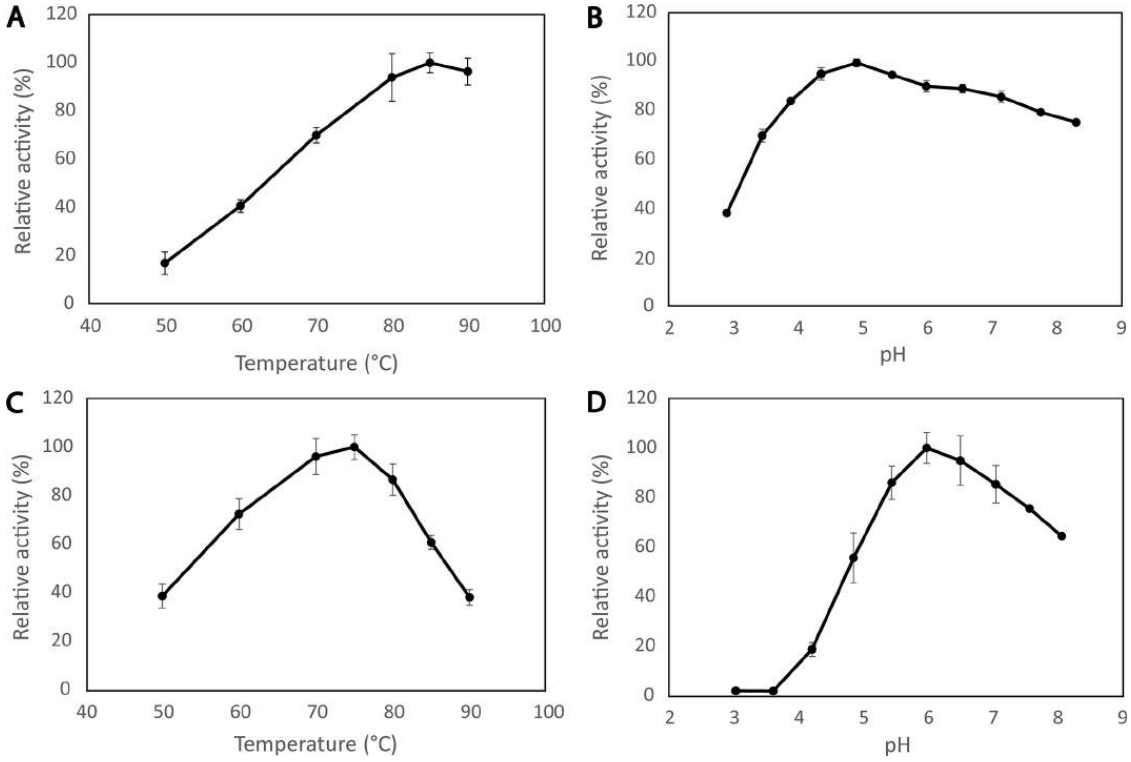

Fig 4. Temperature and pH optimum of mgCel6A. Panels A & B show the optimal temperature at pH 5.0 (A) and the optimal pH at 85˚C (B) for activity on CMC, assessed using 10 min reaction times. Panels C & D show the optimal temperature at pH 6.0 (C) and the optimal pH at 75˚C (D) for activity on sulfite-pulped spruce, assessed using 15 min reaction times.

https://doi.org/10.1371/journal.pone.0197862.g004

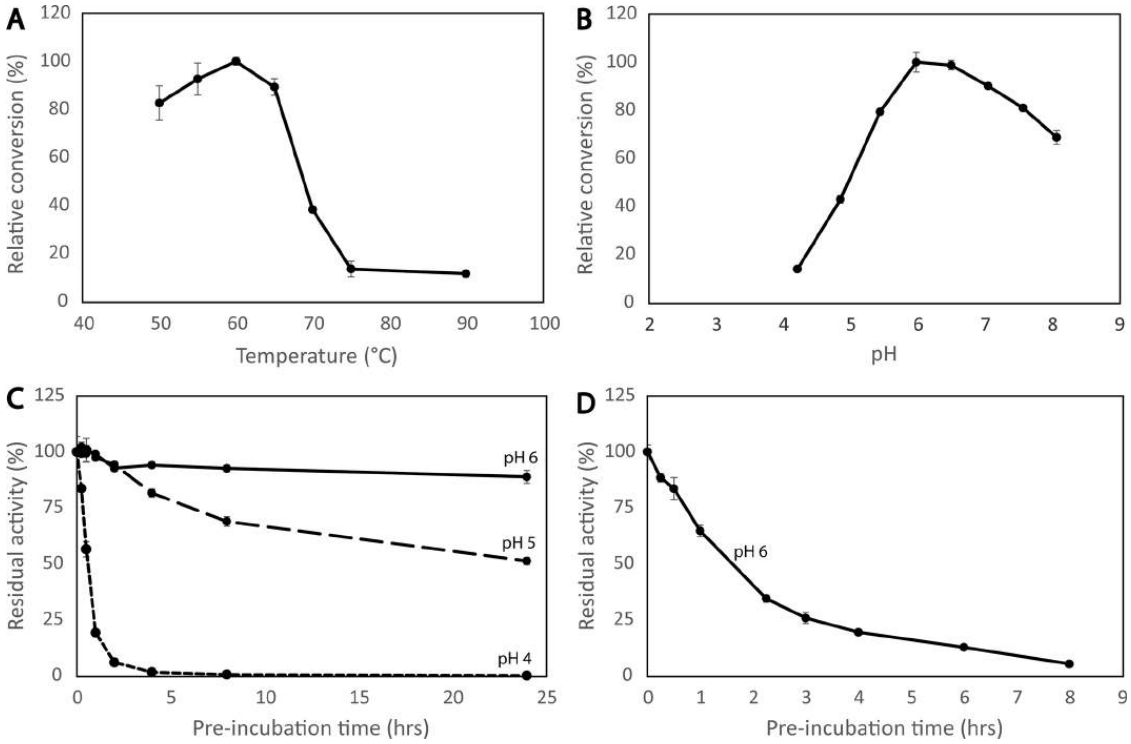

Fig 5. Operational enzyme stability. Panels A & B show the optimal temperature at pH 6.0 (A) and the optimal pH at 60˚C (B) for 24 hours reactions with sulfite-pulped spruce. Panel C shows half-life curves which were obtained by pre-incubating the enzyme at pH 4.0, 5.0 or 6.0 at 65˚C for 24 hours, followed by measurement of remaining activity in a 10 minute assay with CMC. Panel D shows remaining activity against CMC after incubation of mgCel6A at 70˚C, pH 6.0, for various periods of time.

https://doi.org/10.1371/journal.pone.0197862.g005

reaction times. Thermal and pH stability were further investigated by constructing half-life curves after pre-incubating the enzyme in various conditions prior to carrying out an activity assay. These experiments showed that the enzyme retained full activity after pre-incubation for 24 hours at 60˚C, pH 6.0 and pH 5.0 (results not shown). At higher temperatures or lower pH, the enzyme became unstable. For example, at 65˚C the enzyme retained 0%, 50% and 90% activity after 24 hours at pH 4.0, 5.0 or 6.0, respectively (Fig 5C), and at 70˚C the half-life of mgCel6A was less than 2 hours when pre-incubated at pH 6.0 (Fig 5D).

Further studies of enzyme activity

The enzyme was active on all the tested cellulosic model substrates (Fig 6A), including CMC, PASC and Avicel which represent various cellulose structures and degrees of crystallinity. CMC is considered an endoglucanase substrate due to the carboxymethyl substitutions not being compatible with the catalytic site (tunnel) of cellobiohydrolases [52]. While indeed active on CMC, mgCel6A did not exhibit high activity on this substrate considering that CMC is a soluble cellulose that should be easily accessible (Fig 6A). Therefore, an additional experiment according to Irwin et al. [40] was carried out, showing that after degradation of filter paper,

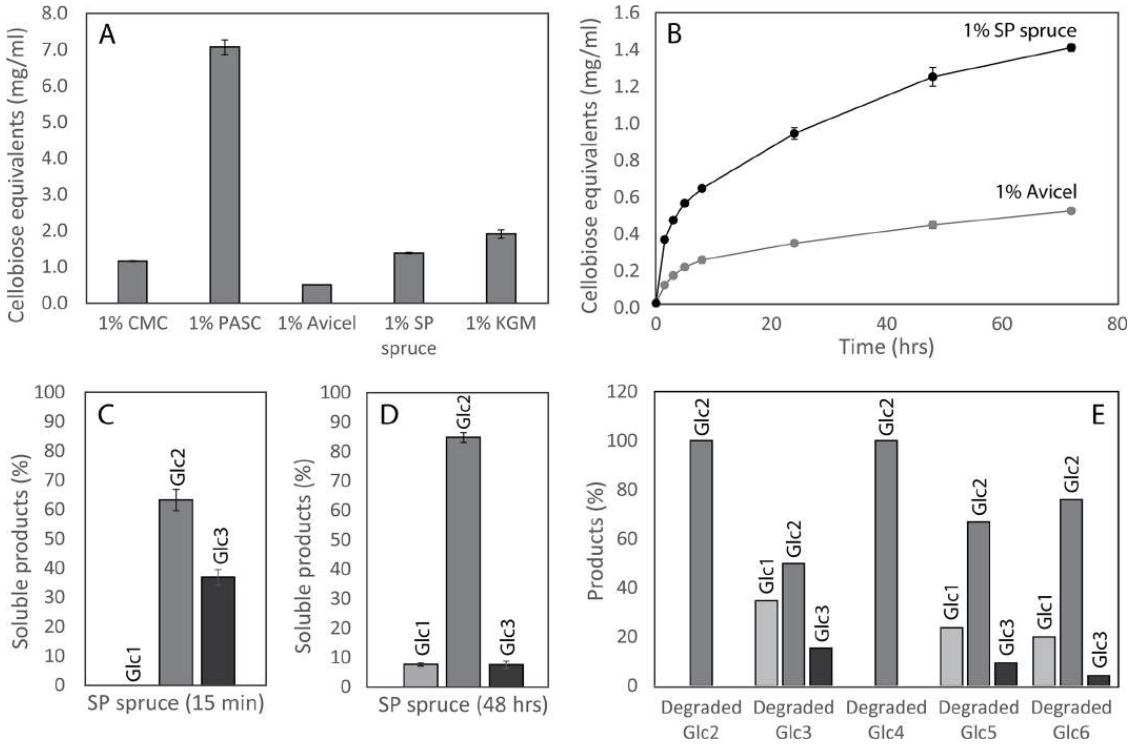

Fig 6. Degradation reactions and hydrolysis products. Panel A shows total solubilization yields after incubating 10 mg/ml CMC, PASC, Avicel, sulfite-pulped spruce and KGM with 1 μM mgCel6A for 72 hours (60˚C, pH 6.0). Panel B shows progress curves for hydrolysis of 10 mg/ml sulphite-pulped spruce or Avicel by 1 μM mgCel6A (60˚C, pH 6.0). Panel C & D show product profiles from sulfite-pulped spruce (conditions as in A) after 15 minutes (C) and 48 hours (D). Panel E shows product profiles obtained after degrading various soluble cello-oligomers (DP2-DP6, at 0.1 mg/ml), for 18 hours.

https://doi.org/10.1371/journal.pone.0197862.g006

approximately 40% of the newly generated reducing ends resided in the insoluble substrate. This high fraction of insoluble reducing ends, together with the structural data, clearly show that mgCel6A is an endocellulase. The nature of the endoglucanase mode of action implies that most cuts will be made in the internal parts of the cellulose chains, thereby allowing other enzymes, such as cellobiohydrolases to access the substrate more easily. Although such internal cleavage is crucial for efficient depolymerization of cellulose, endoglucanases alone do not usually lead to a high degree of substrate solubilization, meaning that hydrolysis yields often seem low. MgCel6A alone was able to convert 13.8±0.3% of sulfite-pulped industrial spruce (10% DM) into soluble cello-oligomers during a 72 hours reaction at 60˚C, using an enzyme load of 8 mg enzyme per 1 g cellulose.

Progress curves for hydrolysis of sulfite-pulped spruce and Avicel showed that mgCel6A degrades the industrial substrate more easily than Avicel (Fig 6B). While Avicel is predominantly crystalline, the sulfite-pulped spruce likely contains a higher ratio of amorphous regions. Accordingly, compared to Avicel, the enzyme was considerably more efficient on PASC (Fig 6A), a substrate that is primarily amorphous [38]. Thus, mgCel6A seems more active on amorphous cellulose regions, while degrading crystalline cellulose at a slower rate. As

commonly observed in enzymatic hydrolysis of cellulose, the reaction rate for mgCel6A became drastically impaired within only a few hours (Fig 6B), suggesting that the substrate rapidly becomes less accessible and degradable after initial fast conversion of the easily accessible parts [53].

Hydrolysis products

Analysis of reaction products obtained from the industrial substrate showed that mgCel6A generates cellobiose and cellotriose during the initial phase of the reaction (Fig 6C), whereas cellobiose accumulates as the main end-product in longer incubations (Fig 6D). Initial formation of a considerable amount of trimeric products fits well with the notion that mgCel6A is an endo-acting enzyme, whereas the later dominance of the dimer suggests that the trimer is slowly converted. The latter was indeed demonstrated in experiments using cellotriose as substrate (Fig 6E). While the dimer was not cleaved by mgCel6A, all other tested cello-oligomers, from DP3 to DP6, were converted to various amounts of DP1-3, consistent with cellobiose being the dominating end-product (Fig 6E). Notably, Fig 6E also shows that the longer oligomers were converted faster than the trimer.

Activity on hemicellulose

The activity of mgCel6A towards various hemicelluloses showed that the enzyme can degrade konjac glucomannan (KGM) (Figs 6A & 7), a hemicellulose consisting of β(1–4) linked glucose and mannan units (with an approximately 40:60 ratio), with backbone acetylations that make the polymer soluble. A mass spectrum of mgCel6A-generated products (Fig 7) showed a variety of products in the DP3-DP13 range, carrying zero, one or two acetylations [54]. KGM is a soluble substrate, which probably offers an explanation to why mgCel6A effectively degrades KGM, relative to the insoluble cellulosic substrates (Fig 6A). Using MS for highly sensitive product analysis, no activity was observed towards xylan or tamarind xyloglucan.

The effect of the CBM on substrate binding and enzyme efficiency

Substrate binding assays with Avicel showed that removal of the CBM from mgCel6A drastically reduced substrate affinity (Fig 8A). Whereas most of the full-length enzyme was bound to the substrate after a few minutes of incubation, most of the truncated variant remained in solution even after longer incubations. This is an expected result, that has, for example, also been described for fungal cellulases [55].

The closest characterized homolog of mgCel6A, *Tf*Cel6A, also has a CBM2 (although with only 34% sequence identity to the CBM2 discussed in this paper), which has been shown to increase binding affinity to insoluble substrates and to increase hydrolysis yields at low substrate concentrations (~1% DM) [40]. Likewise, at low substrate concentrations (0.5–2.5% DM), hydrolysis yields obtained with mgCel6AΔCBM were only 60–80% of the yields obtained with the full-length enzyme. However, at higher substrate concentrations, there was no difference and at 10% DM, mgCel6AΔCBM outperformed mgCel6A (Fig 8B and 8C). Thus, at industrially relevant substrate concentrations, it seems advantageous to employ this GH6 without its CBM2 in order to maximize the yield.

Interestingly, a previous study on fungal CBM1-containing fungal cellulases has shown that, while the CBM1-containing enzymes are clearly more efficient at low substrate concentrations, they are outperformed by their truncated, CBM1-free variants at high substrate concentrations (which imply low water contents). The present data, for a bacterial CBM2, indicate that the substrate concentration-dependence of the impact of CBMs is a general phenomenon. One explanation is that at high substrate concentrations the "proximity effects" ascribed to

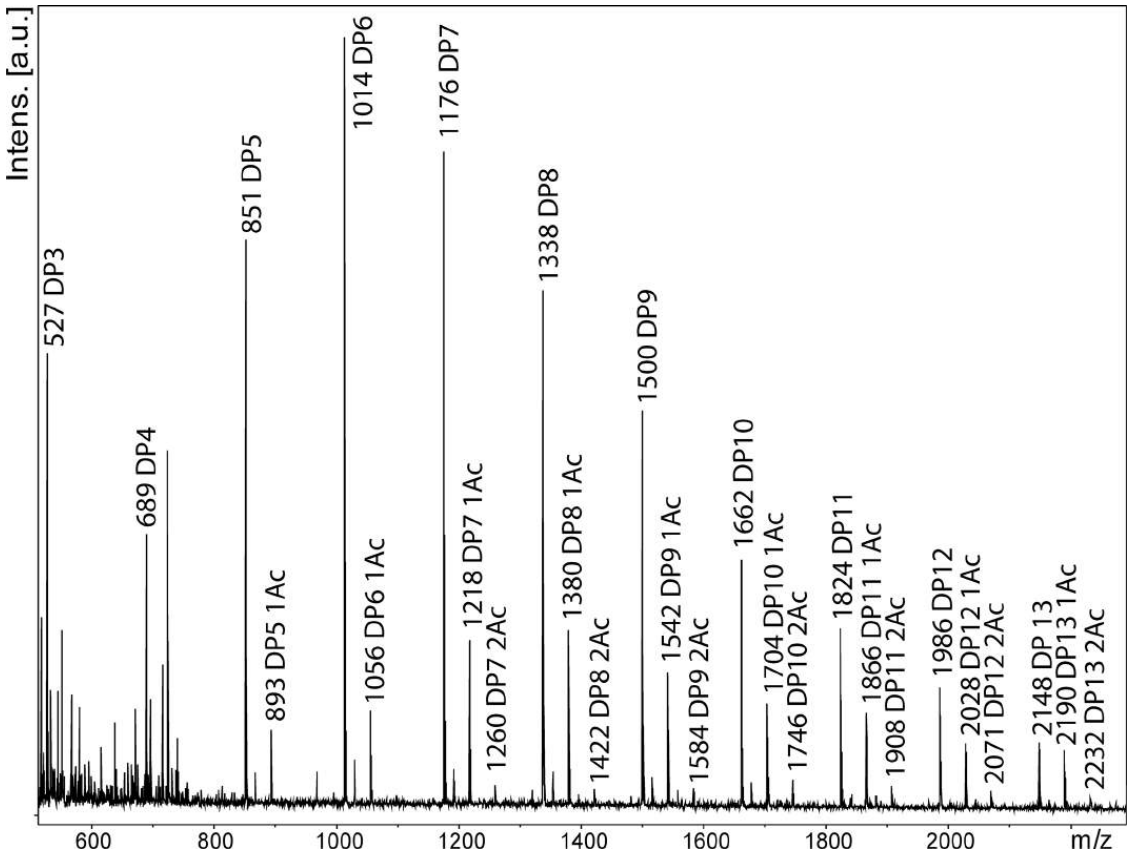

Fig 7. Products generated by mgCel6A from konjac glucomannan (KGM). The indicated m/z values are for sodium adducts of products carrying zero, one (1Ac) or two (2Ac) acetylations. All labelled peaks were not observed in the negative control (= a reaction without enzyme).

https://doi.org/10.1371/journal.pone.0197862.g007

CBMs are superseded by the substrate concentration being saturating. In addition, CBMs likely hamper the rate of substrate desorption, which, under conditions where the CBM no longer contributes to the rate of substrate binding will lead to a net reduction in the overall reaction rate [55, 56].

Concluding remarks

By mining metagenomic data originating in a relevant natural biodiversity, we have obtained and characterized a thermostable bacterial GH6 cellulase, for which structural and functional data demonstrate endoglucanase activity. The enzyme is active on all tested cellulosic substrates, including industrial sulfite-pulped spruce. The overall solubilization yield for the latter was approximately 14% when loaded at 8 mg/g cellulose, which is a promising yield for an endoglucanase acting alone on insoluble cellulose [57, 58].

MgCel6A resembles the well-known *Tf*Cel6A endoglucanase in several ways, but has a seemingly higher temperature optimum than the 55–58˚C that has been reported for the *T*.

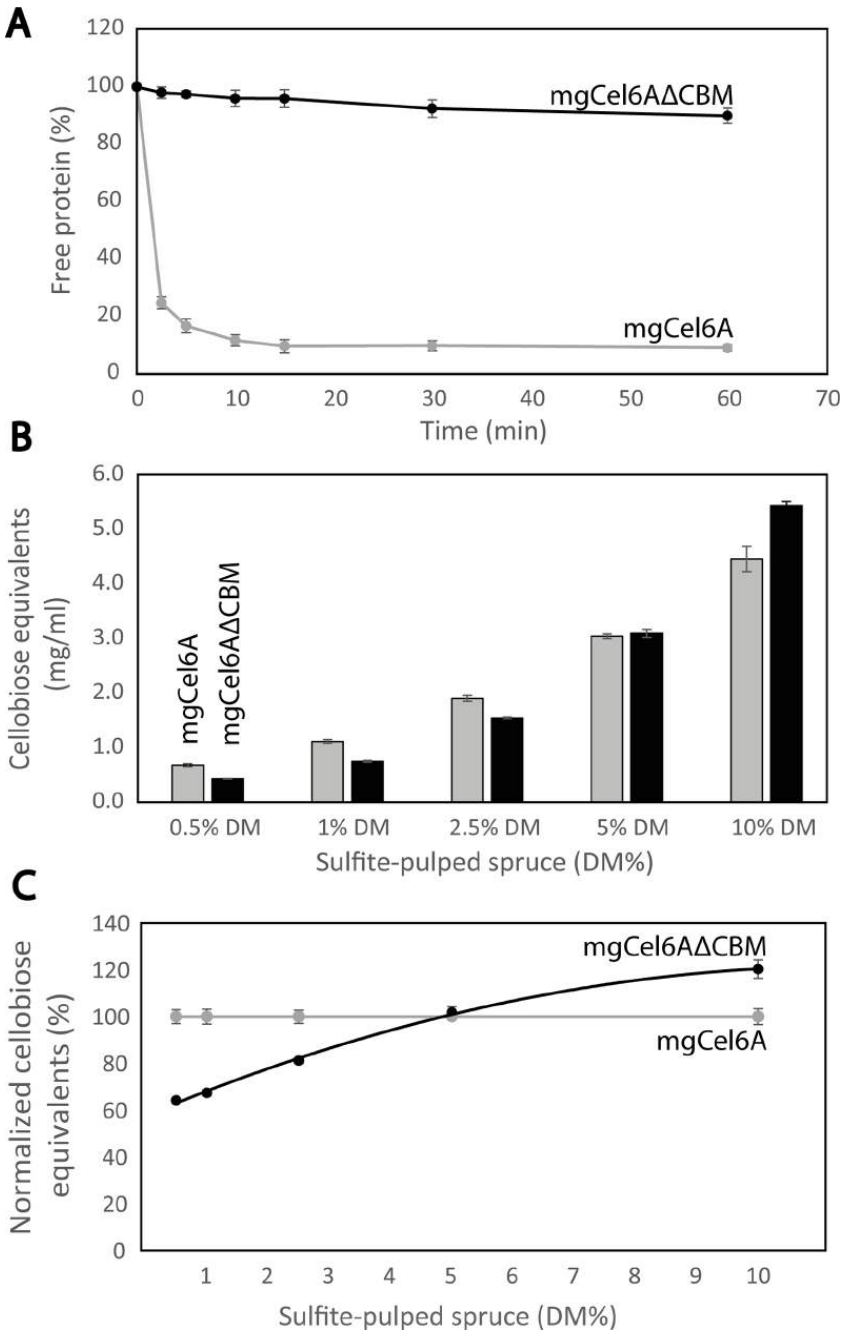
A
Free protein (%)
Time (min)
mgCel6AΔCBM
mgCel6A
B
Cellobiose equivalents (mg/ml)
mgCel6A
mgCel6AΔCBM
0.5% DM
1% DM
2.5% DM
5% DM
10% DM
Sulfite-pulped spruce (DM%)
C
Normalized cellobiose equivalents (%)
mgCel6AΔCBM
mgCel6A
Sulfite-pulped spruce (DM%)

Fig 8. The effect of the CBM on the efficiency of mgCel6A. Panel A shows binding of mgCel6A and mgCel6AΔCBM to Avicel. Panel B shows hydrolysis yields obtained with mgCel6A (grey bars) and mgCel6AΔCBM (black bars), when degrading sulfite-pulped spruce in reactions with various dry matter concentrations (pH 6.0, 60˚C, 48 hrs). The enzyme concentration was kept constant at 1 μM, independently of substrate concentration. Panel C shows a comparison of normalized hydrolysis yields obtained with mgCel6AΔCBM or mgCel6A at various substrate concentrations. At each dry matter concentration, the yield obtained with the full-length enzyme was set to 100%.

https://doi.org/10.1371/journal.pone.0197862.g008

fusca enzyme [33, 59] (note that assay conditions affect the apparent temperature optimum, complicating direct comparison with literature data). Importantly, mgCel6A seems much easier to produce in E. coli (500 mg/L for the full-length enzyme, versus 30–40 mg/L reported for *Tf*Cel6A; [60]). The latter is quite remarkable considering that mgCel6A contains a 40-residue P/T linker. *Tf*Cel6A is known to be active on CMC, PASC and filter paper; it is not known whether the enzyme, like mgCel6A, is also active on glucomannan. Enzyme assays with *Tf*Cel6A are commonly performed at 50–55˚C [40, 46, 47, 61]. MgCel6A shows good operational stability at 60˚C in overnight reactions with sulfite-pulped spruce, a $T_{m(app)}$ of 76˚C according to DSC, and the enzyme retains 90% activity after pre-incubation at 65˚C for 24 hours (Calza et al showed that *Tf*Cel6A loses less than 20% of its activity after 18 hours at 56˚C [59]).

In conclusion, it seems that mgCel6A is a useful enzyme for conducting studies of cellulose degradation at elevated temperatures. This novel endoglucanase exhibits desirable properties compatible with industrial process conditions and has activity on an industrial lignocellulosic substrate, making it a promising enzyme for development of industrial cellulose conversion processes. When employing mgCel6A at industrially relevant substrate concentrations (≥10% DM), the hydrolysis yield was enhanced by removal of the CBM. The true potential of mgCel6A will be explored in follow-up studies, where the enzyme will be assessed in the context of synergistic enzyme cocktails.

Acknowledgments

Provision of beam time at the European Synchrotron Radiation Facility (ESRF) is highly valued.

Author Contributions

Conceptualization: Marianne S. Jensen, Lasse Fredriksen, Alasdair K. MacKenzie, Phillip B. Pope, Gustav Vaaje-Kolstad, Vincent G. H. Eijsink.

Data curation: Marianne S. Jensen.

Formal analysis: Marianne S. Jensen.

Funding acquisition: Vincent G. H. Eijsink.

Investigation: Marianne S. Jensen, Lasse Fredriksen.

Methodology: Marianne S. Jensen, Lasse Fredriksen, Alasdair K. MacKenzie, Ingar Leiros, Piotr Chylenski, Adele K. Williamson, Tony Christopeit, Heidi Østby.

Project administration: Vincent G. H. Eijsink.

Resources: Piotr Chylenski.

Software: Marianne S. Jensen, Ingar Leiros, Adele K. Williamson.

Supervision: Gustav Vaaje-Kolstad, Vincent G. H. Eijsink.

Validation: Vincent G. H. Eijsink.

Visualization: Marianne S. Jensen.

Writing – original draft: Marianne S. Jensen.

Writing – review & editing: Marianne S. Jensen, Lasse Fredriksen, Phillip B. Pope, Ingar Leiros, Adele K. Williamson, Vincent G. H. Eijsink.

References

1. Pauly M. & Keegstra K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J*, 54 (4): 559–568. https://doi.org/10.1111/j.1365-313X.2008.03463.x PMID: 18476863
2. Himmel M. E., Ding S. Y., Johnson D. K., Adney W. S., Nimlos M. R., Brady J. W. et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315 (5813): 804–807. https://doi.org/10.1126/science.1137016 PMID: 17289988
3. Zhang Z., Donaldson A. A. & Ma X. (2012). Advancements and future directions in enzyme technology for biomass conversion. *Biotechnol Adv*, 30 (4): 913–919. https://doi.org/10.1016/j.biotechadv.2012.01.020 PMID: 22306162
4. Harris P. V., Xu F., Kreel N. E., Kang C. & Fukuyama S. (2014). New enzyme insights drive advances in commercial ethanol production. *Curr Opin Chem Biol*, 19: 162–170. https://doi.org/10.1016/j.cbpa.2014.02.015 PMID: 24681544
5. Horn S. J., Vaaje-Kolstad G., Westereng B. & Eijsink V. G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels*, 5 (1): 45. https://doi.org/10.1186/1754-6834-5-45 PMID: 22747961
6. Merino S. T. & Cherry J. (2007). Progress and challenges in enzyme development for biomass utilization. *Adv Biochem Eng Biotechnol*, 108: 95–120. https://doi.org/10.1007/10_2007_066 PMID: 17594064
7. Cocinero E. J., Gamblin D. P., Davis B. G. & Simons J. P. (2009). The building blocks of cellulose: the intrinsic conformational structures of cellobiose, its epimer, lactose, and their singly hydrated complexes. *J Am Chem Soc*, 131 (31): 11117–11123. https://doi.org/10.1021/ja903322w PMID: 19722675
8. Duchesne L. C. & Larson D. W. (1989). Cellulose and the Evolution of Plant Life. *BioScience*, 39 (4): 238–241.
9. Scheller H. V. & Ulvskov P. (2010). Hemicelluloses. *Annu Rev Plant Biol*, 61: 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315 PMID: 20192742
10. Kostylev M. & Wilson D. (2012). Synergistic interactions in cellulose hydrolysis. *Biofuels*, 3 (1): 61–70.
11. Bissaro B., Røhr A. K., Müller G., Chylenski P., Skaugen M., Forsberg Z. et al. (2017). Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2. *Nat Chem Biol*, 13 (10): 1123–1128. https://doi.org/10.1038/nchembio.2470 PMID: 28846668
12. Vaaje-Kolstad G., Westereng B., Horn S. J., Liu Z., Zhai H., Sorlie M. et al. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*, 330 (6001): 219–222. https://doi.org/10.1126/science.1192231 PMID: 20929773
13. Quinlan R. J., Sweeney M. D., Lo Leggio L., Otten H., Poulsen J. C., Johansen K. S. et al. (2011). Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A*, 108 (37): 15079–15084. https://doi.org/10.1073/pnas.1105776108 PMID: 21876164
14. Phillips C. M., Beeson W. T., Cate J. H. & Marletta M. A. (2011). Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol*, 6 (12): 1399–1406. https://doi.org/10.1021/cb200351y PMID: 22004347
15. Forsberg Z., Vaaje-Kolstad G., Westereng B., Bunaes A. C., Stenstrom Y., MacKenzie A. et al. (2011). Cleavage of cellulose by a CBM33 protein. *Protein Sci*, 20 (9): 1479–1483. https://doi.org/10.1002/pro.689 PMID: 21748815
16. Boraston A. B., Bolam D. N., Gilbert H. J. & Davies G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, 382 (3): 769–781.
17. Herve C., Rogowski A., Blake A. W., Marcus S. E., Gilbert H. J. & Knox J. P. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci U S A*, 107 (34): 15293–15298. https://doi.org/10.1073/pnas.1005732107 PMID: 20696902
18. Galbe M. & Zacchi G. (2007). Pretreatment of lignocellulosic materials for efficient bioethanol production. *Adv Biochem Eng Biotechnol*, 108: 41–65. https://doi.org/10.1007/10_2007_070 PMID: 17646946

19. Haki G. D. & Rakshit S. K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresour Technol*, 89 (1): 17–34. PMID: 12676497

20. Nidetzky B., Steiner W., Hayn M. & Claeyssens M. (1994). Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: a new model for synergistic interaction. *Biochem J*, 298 (3): 705–710.

21. Viikari L., Alapuranen M., Puranen T., Vehmaanpera J. & Siika-Aho M. (2007). Thermostable enzymes in lignocellulose hydrolysis. *Adv Biochem Eng Biotechnol*, 108: 121–145. https://doi.org/10.1007/10_2007_065 PMID: 17589813

22. Kumar R., Tabatabaei M., Karimi K. & Sárvári Horváth I. (2016). Recent updates on lignocellulosic biomass derived ethanol—A review. *Biofuel Research Journal*, 3 (1): 347–356.

23. Liu G., Zhang J. & Bao J. (2016). Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling. *Bioprocess Biosyst Eng*, 39 (1): 133–140. https://doi.org/10.1007/s00449-015-1497-1 PMID: 26541585

24. Reddy A. P., Simmons C. W., D'Haeseleer P., Khudyakov J., Burd H., Hadi M. et al. (2013). Discovery of microorganisms and enzymes involved in high-solids decomposition of rice straw using metagenomic analyses. *PLoS One*, 8 (10): e77985. https://doi.org/10.1371/journal.pone.0077985 PMID: 24205054

25. 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 Research*, 40 (Web Server issue): 445–451.

26. Aslanidis C. & de Jong P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research*, Vol. 18 (20): 6069–6074. PMID: 2235490

27. Bradford M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72: 248–254. PMID: 942051

28. Winn M. D., Ballard C. C., Cowtan K. D., Dodson E. J., Emsley P., Evans P. R. et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography*, 67 (Pt 4): 235–42. https://doi.org/10.1107/S0907444910045749 PMID: 21460441

29. Larsson A. M., Bergfors T., Dultz E., Irwin D. C., Roos A., Driguez H. et al. (2005). Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: the role of tyrosine Y73 in substrate ring distortion. *Biochemistry*, 44 (39): 12915–12922. https://doi.org/10.1021/bi0506730 PMID: 16185060

30. Murshudov G. N., Vagin A. A. & Dodson E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica. Section D, Biological Crystallography*, 53 (Pt 3): 240–255. https://doi.org/10.1107/S0907444996012255 PMID: 15299926

31. Cowtan K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica. Section D, Biological Crystallography*, 62 (Pt 9): 1002–1011. https://doi.org/10.1107/S0907444906022116 PMID: 16929101

32. Emsley P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, 60 (Pt 12 Pt 1): 2126–2132.

33. Spezio M., Wilson D. B. & Karplus P. A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry*, 32 (38): 9906–9916. PMID: 8399160

34. Sandgren M., Wu M., Karkehabadi S., Mitchinson C., Kelemen B. R., Larenas E. A. et al. (2013). The structure of a bacterial cellobiohydrolase: the catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. *J Mol Biol*, 425 (3): 622–635. https://doi.org/10.1016/j.jmb.2012.11.039 PMID: 23220193

35. Robert X. & Gouet P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42 (W1): W320–W324.

36. Rødsrud G., Lersch M. & Sjöde A. (2012). History and future of world's most advanced biorefinery in operation. *Biomass and Bioenergy*, 46: 46–59.

37. Sjöde, A., Frölander, A., Lersch, M. & Rødsrud, G. (2013). Lignocellulosic biomass conversion by sulfite pretreatment., Patent EP2376642 B1, Borregaard, AS.

38. Wood T. M. (1988). Preparation of crystalline, amorphous, and dyed cellulase substrates. *Methods in Enzymology*, 160: 19–25.

39. Miller G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.*, 31 (3): 426–428.

40. Irwin D. C., Spezio M., Walker L. P. & Wilson D. B. (1993). Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnol Bioeng*, 42 (8): 1002–1013. https://doi.org/10.1002/bit.260420811 PMID: 18613149

41. Kristensen J. B., Felby C. & Jorgensen H. (2009). Determining yields in high solids enzymatic hydrolysis of biomass. *Appl Biochem Biotechnol*, 156 (1–3): 127–132. https://doi.org/10.1007/s12010-008-8375-0 PMID: 18836690

42. Vaaje-Kolstad G., Houston D. R., Riemen A. H., Eijsink V. G. & van Aalten D. M. (2005). Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J Biol Chem*, 280 (12): 11313–11319. https://doi.org/10.1074/jbc.M407175200 PMID: 15590674

43. Simmons C. W., Reddy A. P., D'Haeseleer P., Khudyakov J., Billis K., Pati A. et al. (2014). Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol Biofuels*, 7 (1): 495. https://doi.org/10.1186/s13068-014-0180-0 PMID: 25648696

44. Xu G. Y., Ong E., Gilkes N. R., Kilburn D. G., Muhandiram D. R., Harris-Brandts M. et al. (1995). Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. *Biochemistry*, 34 (21): 6993–7009. PMID: 7766609

45. Meinke A., Damude H. G., Tomme P., Kwan E., Kilburn D. G., Miller R. C. et al. (1995). Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *J Biol Chem*, 270 (9): 4383–4386. PMID: 7876202

46. Wolfgang D. E. & Wilson D. B. (1999). Mechanistic studies of active site mutants of Thermomonospora fusca endocellulase E2. Biochemistry, 38 (30): 9746–9751. https://doi.org/10.1021/bi990401v PMID: 10423254

47. André G., Kanchanawong P., Palma R., Cho H., Deng X., Irwin D. et al. (2003). Computational and experimental studies of the catalytic mechanism of Thermobifida fusca cellulase Cel6A (E2). Protein Eng, 16 (2): 125–134. PMID: 12676981

48. Payne C. M., Knott B. C., Mayes H. B., Hansson H., Himmel M. E., Sandgren M. et al. (2015). Fungal cellulases. Chem Rev, 115 (3): 1308–1448. https://doi.org/10.1021/cr500351c PMID: 25629559

49. Mayes H. B., Knott B. C., Crowley M. F., Broadbelt L. J., Ståhlberg J., & Beckham G. T. (2016). Who's on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolases", Journal. Chem. Sci., 7 (9): 5955–5968.

50. Duy C. & Fitter J. (2005). Thermostability of irreversible unfolding alpha-amylases analyzed by unfolding kinetics. J Biol Chem, 280 (45): 37360–37365. https://doi.org/10.1074/jbc.M507530200 PMID: 16150692

51. Sanchez-Ruiz J. M., Lopez-Lacomba J. L., Cortijo M. & Mateo P. L. (1988). Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. Biochemistry, 27 (5): 1648–1652. PMID: 3365417

52. Hall M., Bansal P., Lee J. H., Realff M. J. & Bommarius A. S. (2010). Cellulose crystallinity—a key predictor of the enzymatic hydrolysis rate. FEBS J, 277 (6): 1571–1582. https://doi.org/10.1111/j.1742-4658.2010.07585.x PMID: 20148968

53. Chandra R. P., Bura R., Mabee W. E., Berlin A., Pan X. & Saddler J. N. (2007). Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Eng Biotechnol, 108: 67–93. https://doi.org/10.1007/10_2007_064 PMID: 17530205

54. Albrecht S., van Muiswinkel G. C., Schols H. A., Voragen A. G. & Gruppen H. (2009). Introducing capillary electrophoresis with laser-induced fluorescence detection (CE-LIF) for the characterization of konjac glucomannan oligosaccharides and their in vitro fermentation behavior. J Agric Food Chem, 57 (9): 3867–3876. https://doi.org/10.1021/jf8038956 PMID: 19296676

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

56. Le Costaouec T., Pakarinen A., Varnai A., Puranen T. & Viikari L. (2013). The role of carbohydrate binding module (CBM) at high substrate consistency: comparison of Trichoderma reesei and Thermoascus aurantiacus Cel7A (CBHI) and Cel5A (EGII). Bioresour Technol, 143: 196–203. https://doi.org/10.1016/j.biortech.2013.05.079 PMID: 23796604

57. Chylenski P., Forsberg Z., Ståhlberg J., Várnai A., Lersch M., Bengtsson O. et al. (2017). Development of minimal enzyme cocktails for hydrolysis of sulfite-pulped lignocellulosic biomass. J Biotechnol, 246: 16–23. https://doi.org/10.1016/j.jbiotec.2017.02.009 PMID: 28219736

58. Medve J., Karlsson J., Lee D. & Tjerneld F. (1998). Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes. Biotechnol Bioeng, 59 (5): 621–634. PMID: 10099380

59. Calza R. E., Irwin D. C. & Wilson D. B. (1985). Purification and characterization of two beta-1,4-endoglucanases from Thermomonospora fusca. Biochemistry, 24 (26): 7797–7804.

60. Zhang S., Barr B. K. & Wilson D. B. (2000). Effects of noncatalytic residue mutations on substrate specificity and ligand binding of Thermobifida fusca endocellulase cel6A. Eur J Biochem, 267 (1): 244–252. PMID: 10601873

61. Ghangas G. S. & Wilson D. B. (1988). Cloning of the Thermomonospora fusca Endoglucanase E2 Gene in Streptomyces lividans: Affinity Purification and Functional Domains of the Cloned Gene Product. Appl Environ Microbiol, 54 (10): 2521–2526. PMID: 16347759

Paper II

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENZYMOLOGY AND PROTEIN ENGINEERING

Discovery of a Thermostable GH10 Xylanase with Broad Substrate Specificity from the Arctic Mid-Ocean Ridge Vent System

L. Fredriksen,[a] R. Stokke,[b] M. S. Jensen,[a] B. Westereng,[a] J.-K. Jameson,[a] I. H. Steen,[b] V. G. H. Eijsink[a]

[a]Faculty of Chemistry, Biotechnology, and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway
[b]Department of Biological Sciences and K. G. Jebsen Centre for Deep Sea Research, University of Bergen, Bergen, Norway

ABSTRACT A two-domain GH10 xylanase-encoding gene (*amor_gh10a*) was discovered from a metagenomic data set, generated after *in situ* incubation of a lignocellulosic substrate in hot sediments on the sea floor of the Arctic Mid-Ocean Ridge (AMOR). AMOR_GH10A comprises a signal peptide, a carbohydrate-binding module belonging to a previously uncharacterized family, and a catalytic glycosyl hydrolase (GH10) domain. The enzyme shares the highest sequence identity (42%) with a hypothetical protein from a *Verrucomicrobia* bacterium, and its GH10 domain shares low identity (24 to 28%) with functionally characterized xylanases. Purified AMOR_GH10A showed thermophilic and halophilic properties and was active toward various xylans. Uniquely, the enzyme showed high activity toward amorphous cellulose, glucomannan, and xyloglucan and was more active toward cellopentaose than toward xylopentaose. Binding assays showed that the N-terminal domain of this broad-specificity GH10 binds strongly to amorphous cellulose, as well as to microcrystalline cellulose, birchwood glucuronoxylan, barley β-glucan, and konjac glucomannan, confirming its classification as a novel CBM (CBM85).

IMPORTANCE Hot springs at the sea bottom harbor unique biodiversity and are a promising source of enzymes with interesting properties. We describe the functional characterization of a thermophilic and halophilic multidomain xylanase originating from the Arctic Mid-Ocean Ridge vent system, belonging to the well-studied family 10 of glycosyl hydrolases (GH10). This xylanase, AMOR_GH10A, has a surprisingly wide substrate range and is more active toward cellopentaose than toward xylopentaose. This substrate promiscuity is unique for the GH10 family and could prove useful in industrial applications. Emphasizing the versatility of AMOR_GH10A, its N-terminal domain binds to both xylans and glycans, while not showing significant sequence similarities to any known carbohydrate-binding module (CBM) in the CAZy database. Thus, this N-terminal domain lays the foundation for the new CBM85 family.

KEYWORDS CBM, GH10, deep-sea, thermostable, xylanase

Xylan is the most abundant polymer in hemicelluloses, which, together with cellulose and lignin, are considered to be among the most important renewable resources on earth (1). Xylan consists of a backbone of β-1,4-D-xylan decorated with O-acetyl, β-L-arabinofuranosyl, α-D-glucuronic acid, and phenolic acid at various positions (2). Complete degradation of this complex structure requires synergistic action of multiple enzymes, including endo-1,4-β-D-xylanases (EC 3.2.1.8) and β-D-xylosidases, which catalyze hydrolysis of the xylan main chain, and α-L-arabinofuranosidases, α-D-glucuronidases, acetylxylan esterases, ferulic acid esterases, and *p*-coumaric acid esterases which target side groups (3).

Citation Fredriksen L, Stokke R, Jensen MS, Westereng B, Jameson J-K, Steen I, Eijsink VGH. 2019. Discovery of a thermostable GH10 xylanase with broad substrate specificity from the Arctic Mid-Ocean Ridge vent system. Appl Environ Microbiol 85:e02970-18. https://doi.org/10.1128/AEM.02970-18.

Editor Eric V. Stabb, University of Georgia

Address correspondence to V. G. H. Eijsink, vincent.eijsink@nmbu.no.

Received 12 December 2018
Accepted 3 January 2019
Accepted manuscript posted online 11 January 2019
Published 6 March 2019

Of these enzymes, endo-1,4-β-D-xylanases (EC 3.2.1.8) have been grouped into the glycosyl hydrolase families 5, 7, 8, 10, 11, 30, 43, 51, 98, and 141 according to the CAZy classification system (4). GH10 and GH11 xylanases are the most common and, while both act on the xylan main chain, these two enzyme types have different folds, substrate specificities, and mechanisms of action (5). GH10 xylanases exhibit higher catalytic versatility and can catalyze the cleavage of the xylan backbone at the nonreducing side of substituted xylose residues, whereas GH11 enzymes require unsubstituted regions of the xylan backbone (6, 7). As a result, GH10 xylanases generally produce shorter xylooligosaccharides than do members of the GH11 family (6). Some GH10 xylanases have been shown to have a side activity toward cellulose, an activity that has been linked to specific structural features in the active site (8, 9). Such multifunctional enzymes active on plant polysaccharides could be of great interest to the biofuel and biorefinery industries in order to reduce enzyme cost and consequently increase the economical viability of plant polysaccharide conversion.

Next to a catalytic domain, many carbohydrate-active enzymes, including GH10 xylanases, often contain one or more carbohydrate-binding modules (CBMs) (10). CBMs are thought to enhance enzyme efficiency by targeting the substrate to the appended catalytic domain (10, 11). CBMs are currently divided into 84 families in the CAZy database (4).

Xylanases have attracted attention due to their application in the food, animal feed, pulp and paper, and biofuel industries (7, 12). For example, the saccharification of xylan-rich lignocellulosic biomass by cellulases may benefit from the presence of xylanases that help to remove hemicelluloses that cover the cellulose fibrils (13). Thermostable xylanases are of special interest due to their ability to function under harsh conditions, and several such enzymes have been described, originating from compost (14), the termite gut (15), a solfataric spring (16), and the deep sea (17). Hot environments in the deep sea form an attractive source for enzymes, but there are few reports on thermophilic xylanases from the deep sea.

Hydrothermal vent fields in the deep sea are populated by thermophilic microbial communities of significant metabolic and phylogenetic diversity and therefore constitute an interesting biotope for metagenome mining (18). In the present study, in search of glycosyl hydrolases (GHs) for conversion of lignocellulosic biomass, a chamber filled with an industrial lignocellulosic substrate was deployed in hot sediments, 570 m below the sea level at the Jan Mayen hydrothermal vent field. Subsequently, the resulting enriched microbial DNA was sequenced and mined for the presence of carbohydrate-active enzymes. The resulting metagenomic data set revealed a gene encoding a two-domain xylanase (*amor_gh10a*), which was synthesized and cloned in *Escherichia coli*, after which the recombinant enzyme (AMOR_GH10A) was characterized in detail. The results show that AMOR_GH10A is a thermophilic and halophilic xylanase with a surprisingly high activity on cellulosic substrates and containing a previously uncharacterized CBM.

RESULTS

Domain structure of AMOR_GH10A and bioinformatic analysis of its GH10 domain. A chamber filled with industrial lignocellulose was submerged in hot sediments at the Arctic Mid-Ocean Ridge (AMOR), and the DNA material present in this chamber after approximately 1 year of incubation was subjected to metagenomic sequencing. The resulting metagenomic data set was mined for the presence of carbohydrate-active enzymes (CAZymes), which led to the identification of the 1,767-bp *amor_gh10a* gene (see Fig. S1A in the supplemental material). The encoded protein, AMOR_GH10A, comprises 589 amino acids (Fig. S1B), including an 18-residue signal peptide, a putative CBM domain (residues 19 to 180), and a C-terminal GH10 domain (residues 239 to 529; Fig. 1A). The CBM domain was not found by dbCAN, but more sensitive methods for detection of remote homologies implemented in Swiss-Model predicted a possible relationship (<20% sequence identity) with CBM22 domains, which are known to bind xylan (see below for further discussion of the CBM). The

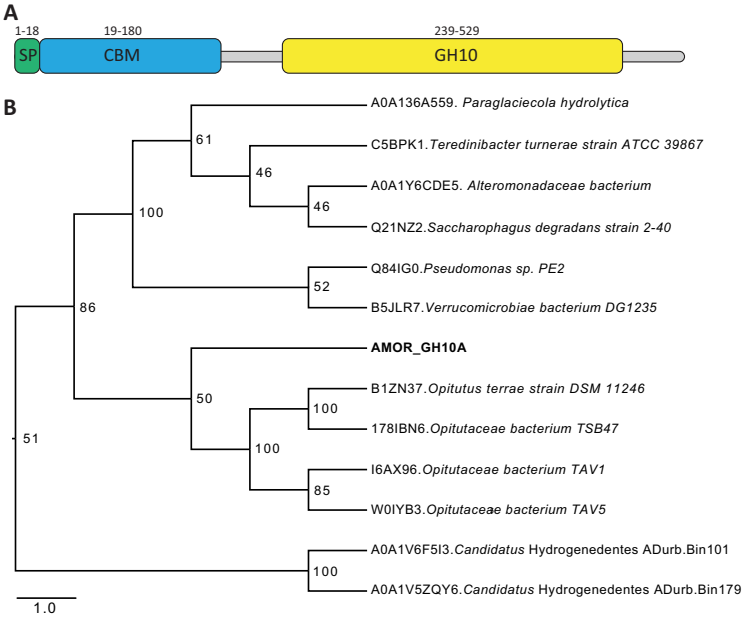

FIG 1 Domain architecture and phylogenetic analysis of the GH10 domain of AMOR_GH10A. (A) Domain architecture of AMOR_GH10A drawn to scale. The full-length enzyme contains a predicted signal peptide (SP; residues 1 to 18), followed by a putative CBM domain (residues 19 to 180) and a GH10 xylanase domain (residues 239 to 529). (B) Phylogenetic analysis of the GH10 domain. The tree was constructed using the maximum-likelihood method inferred by RAxML (v8.2.9). The numbers at the nodes indicate bootstrap values based on 1,000 bootstrap replications. The scale bar represents 1.0 amino acid substitutions per position.

C-terminal cutoff point for the CBM domain was set using the structure of a CBM22 from *Clostridium thermocellum* (PDB 1H6Y [19]) as a template.

BLAST searches (20) revealed that full-length AMOR_GH10A shares the highest sequence identity (42%) with a hypothetical protein from a *Verrucomicrobia* bacterium. A search in UniProt yielded 42 AMOR_GH10A related proteins, none of which have been characterized experimentally, and Fig. 1B shows a phylogenetic tree for the twelve of these proteins that are most similar to AMOR_GH10A. The putative GH10 proteins shown in Fig. 1B originate from organisms associated with marine environments (the NCBI taxonomy ID numbers are indicated in parentheses): *Paraglaciecola hydrolytica* (1799789), *Teredinibacter turnerae* (377629), *Alteromonadaceae* bacterium (1304903), *Saccharophagus degradans* (203122), *Pseudomonas* sp. PE2 (147642), *Verrucomicrobiae* bacterium (382464), *Opitutus terrae* strain DSM 11246 (452637), *Opitutaceae* bacterium TSB47 (1184151), *Opitutaceae* bacterium TAV1 (278956), *Opitutaceae* bacterium TAV5 (794903), "*Candidatus* Hydrogenedentes bin 101" (1852847), and "*Candidatus* Hydrogenedentes bin 179" (1852849). Of these, the GH10 domain from AMOR_GH10A is most similar to the predicted GH10 domain from UniProt entry A0A178IBN6, originating from *Opitutaceae* bacterium TSB47 (sequence identity, 45%). Compared to characterized enzymes, the GH10 domain of AMOR_GH10A is most similar to the cellulosomal xylanase Xyn10D (sequence identity, 31%) from *C. thermocellum* (21).

The closest relatives of the GH10 domain of AMOR_GH10A with a known three-dimensional structure are Xyn10B from *Cellvibrio mixtus* (*Cm*Xyn10B [22]; sequence identity, 28.4%), XynA from *Streptomyces lividans* (*Sl*XynA [23]; sequence identity, 24.6%), and TmxB from *Thermotoga maritima* (24; sequence identity, 23.9% [Fig. 2]).

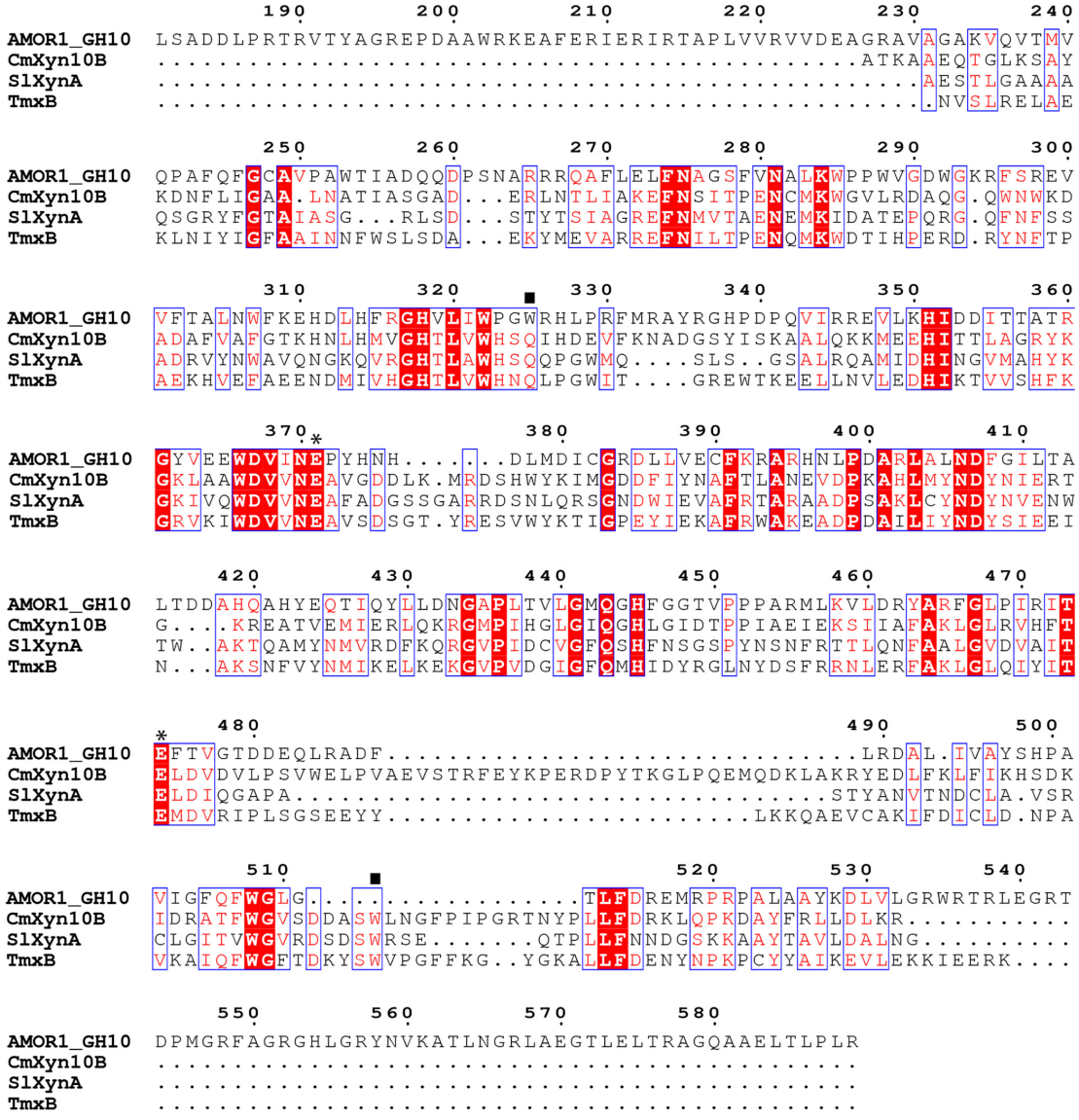

FIG 2 Sequence alignment of the C-terminal part of AMOR_GH10A and GH10 domains with known structures. The alignment includes *Cm*Xyn10B (PDB 1UQY) from *Cellvibrio mixtus*, *Sl*XynA from *Streptomyces lividans* (PDB 1E0V), and TmxB from *Thermotoga maritima* (PDB 1VBR). The alignment is numbered according to AMOR_GH10A. Fully conserved residues appear in white on a red background, whereas less-conserved residues appear as red letters. The presumed catalytic residues are indicated by black stars. Subsite residues potentially involved in determining substrate specificity (see Discussion section) are marked by black squares.

While *Cm*Xyn10B and *Sl*XynA are mesophilic enzymes originating from soil bacteria (25, 26), TmxB is a thermophilic enzyme stable up to 100°C (24).

The catalytic acid-base and nucleophile residues for *Cm*Xyn10B, *Sl*XynA, and TmxB have previously been identified (24, 27, 28) and correspond to residues E371 and E474 in AMOR_GH10A, respectively (Fig. 2). Interestingly, a conserved sequence element involved in shaping the −1 subsite present in all structurally characterized GH10

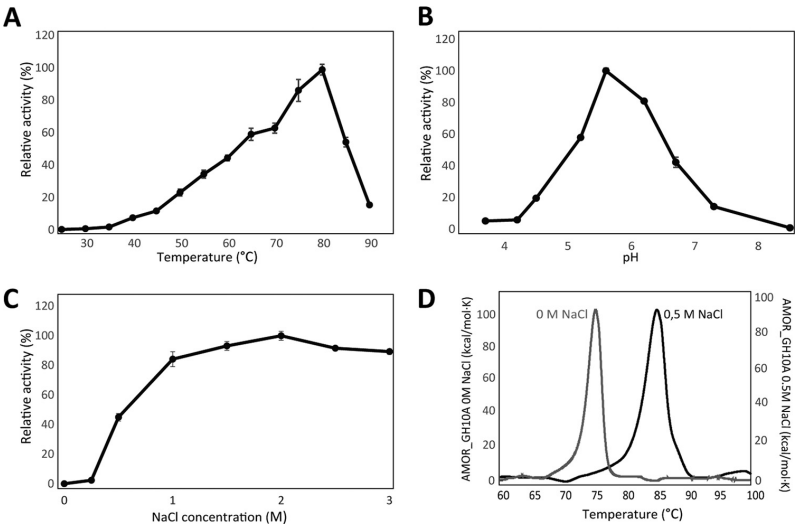

FIG 3 Effect of temperature, pH, and NaCl on the activity and thermostability of AMOR_GH10A. (A) Temperature dependency of activity at pH 5.6. (B) pH dependency of activity at 80°C. Reaction mixtures contained 0.5% (wt/vol) BGX, 2 μM AMOR_GH10A, and buffer supplemented with 0.5 M NaCl, and the reaction time was 10 min. (C) Effect of the NaCl concentration on enzyme activity at pH 5.6 and 80°C. (D) DSC thermograms with or without NaCl supplemented in the buffer. The protein samples (0.5 mg/ml) were heated at a rate of 1°C/min in 25 mM sodium acetate buffer (pH 5.6). The buffer baseline was subtracted.

xylanases is lacking in AMOR_GH10A (between residues 511 and 512 in AMOR_GH10A; Fig. 2). Another unique feature of the catalytic domain of AMOR_GH10A, is the high arginine/lysine ratio, which could be associated with halophilicity (29). The Arg/Lys ratios for the four catalytic domains shown in Fig. 2 are 3.7 (AMOR_GH10A), 0.6 (*Cm*Xyn10B), 2 (*Sl*XynA), and 0.5 (TmxB). One striking feature concerns the insertion of three consecutive arginines at positions 265 to 267 in AMOR_GH10A (Fig. 2). The sequence immediately upstream of the AMOR_GH10A catalytic domain (residues 181 to 238) has similarities to the corresponding parts of the 12 phylogenetically related enzymes depicted in Fig. 1 (data not shown; sequence identities, 44 to 64%). The C-terminal extension of the GH10 domain of AMOR_GH10A (residues 530 to 589) is most similar to the C-terminal parts of GH10s from *Opitutaceae* bacterium TSB47 (sequence identity, 46%) and *Opitutus terrae* (sequence identity, 45%).

Cloning, expression, and purification of AMOR_GH10A. A gene fragment encoding the mature part of AMOR_GH10A (omitting the 18-residue signal peptide) was synthesized, amplified by PCR, and cloned by ligation-independent cloning (LIC) into the pNIC-CH vector. By doing so, a sequence encoding a C-terminal hexahistidine tag (His tag) was added to the *amor_gh10a* open reading frame (ORF). The resulting plasmid was transformed into *E. coli* and protein production was achieved by IPTG (isopropyl-β-D-thiogalactopyranoside) induction of the T7 promoter. The protein was purified to electrophoretic homogeneity (see Fig. S2 in the supplemental material), and the final yield of purified AMOR_GH10A was approximately 50 mg of protein per liter of *E. coli* culture.

Optimal conditions, halophilicity, and apparent melting temperature of AMOR_GH10A. The optimal conditions for AMOR_GH10A activity were analyzed using commercially available birch glucuronoxylan (BGX) as the substrate, and the results are shown in Fig. 3. The optimal temperature for AMOR_GH10A activity is 80°C (Fig. 3A), which is slightly higher than the temperature measured at the deep-sea sampling site (70°C). The optimal pH for AMOR_GH10A is approximately 5.6 (Fig. 3B), which is more typical of fungal xylanases than bacterial ones (7). The enzyme showed a pronounced salt-dependent increase in activity up to 1 M NaCl (Fig. 3C), which was also observed in

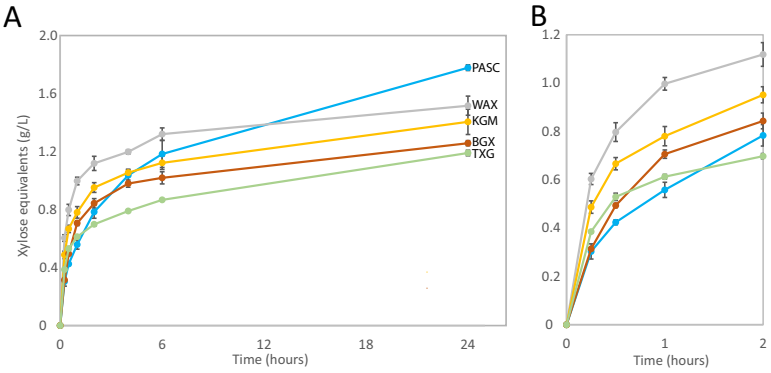

FIG 4 Substrate specificity of AMOR_GH10A. (A) Time courses for the degradation of 0.5% (wt/vol) of phosphoric acid-swollen cellulose (PASC; blue), wheat arabinoxylan (WAX; gray), konjac glucomannan (KGM; yellow), birchwood glucuronoxylan (BGX; orange), or tamarind xyloglucan (TXG; green) by 2 μM AMOR_GH10A over 24 h. The reaction mixtures were incubated at 60°C in 25 mM NaOAc buffer supplemented with 0.5 M NaCl (pH 5.6). Background signals for reactions without enzyme were subtracted before calculating the amount of reducing ends using the DNS method and xylose as standard. (B) Zoom-in view of the first 2 h, revealing differences in the initial reaction rates. More details on the initial rates are provided in Fig. S3 and the main text.

reactions performed at lower temperatures (60 and 50°C; data not shown). Upon heating, the protein unfolded irreversibly, leading to precipitation. The differential scanning calorimetry (DSC) scans shown in Fig. 3D show apparent melting temperatures of 75.3 and 85.1°C in the absence of NaCl and in the presence of 0.5 M NaCl, respectively. Clearly, NaCl has a major effect on both enzyme activity and stability.

Substrate specificity. AMOR_GH10A was incubated with several model polysaccharide substrates—birchwood glucuronoxylan (BGX), wheat arabinoxylan (WAX), phosphoric acid-swollen cellulose (PASC), tamarind xyloglucan (TXG), and konjac glucomannan (KGM)—to assess substrate specificity (Fig. 4A). Remarkably, AMOR_GH10A converted all substrates, albeit at different apparent rates. Initial time points (Fig. 4B) show that the enzyme was most active on (soluble) WAX and least active on (insoluble) PASC. Of note, after 24 h of incubation, the highest product yield was nevertheless obtained for PASC. Determination of initial rates at the optimum reaction temperature of 80°C, yielded rates of 4.3 s^{-1} for WAX, 2.2 s^{-1} for BGX, and 2.4 s^{-1} for PASC (Fig. S3). From these results it is clear that AMOR_GH10A can cleave both at glucose (as in KGM, PASC, and TXG) and xylose (as in BGX and WAX) residues (KGM consists of β-1,4-linked mannose and glucose residues in a 60:40 ratio; TXG comprises a β-1,4-linked glucan backbone with substitutions). The enzyme showed no detectable activity toward ivory nut mannan (INM) and crystalline cellulose (Avicel; data not shown).

Hydrolysis products. Products generated from BGX were also analyzed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Fig. 5A) and matrix-assisted laser desorption ionization−time of flight (MALDI-TOF) mass spectrometry (MS) (Fig. 5B). The HPAEC-PAD analysis of unsubstituted xylooligosaccharide products (XOS) showed that similar amounts of xylotriose (X_3), xylotetraose (X_4), xylopentaose (X_5), and xylohexaose (X_6) and only negligible amounts of xylose (X) and xylobiose (X_2) were generated within the early phase of the reaction. After 24 h, the reaction products were predominantly X_2 to X_5. Further analysis of the reaction products using MALDI-TOF showed that AMOR_GH10A also generates oligomers substituted with a methylated glucuronic acid (Fig. 5B). Incubation of AMOR_GH10A with wheat arabinoxylan (WAX) resulted in lower yields, but a similar profile of nonsubstituted xylooligosaccharides, whereas MALDI-TOF MS-based analysis of substitutions was not accomplished due to the identical masses of xylose and arabinose (results not shown).

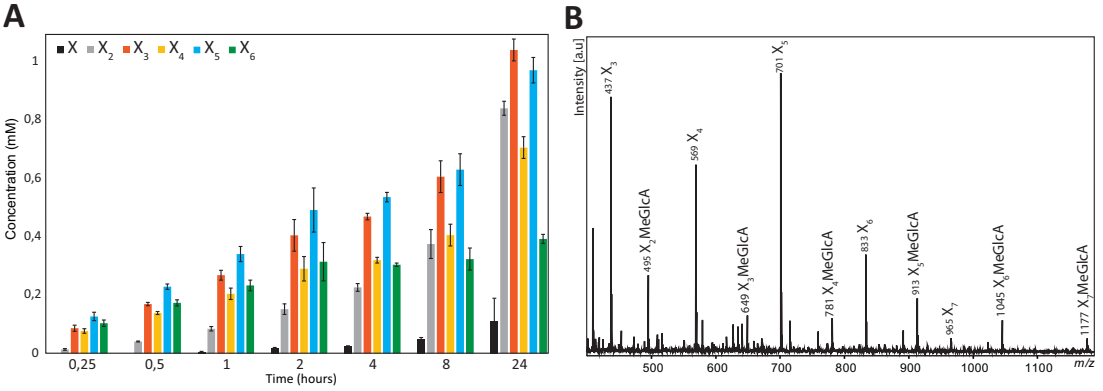

FIG 5 Action of AMOR_GH10A on BGX. (A) Production of unsubstituted xylooligosaccharides (X to X_6) generated from 0.5% (wt/vol) BGX by 2 μM AMOR_GH10A at 60°C (pH 5.6), determined by HPAEC-PAD. (B) Both unsubstituted and MeGlcA-substituted xylooligosaccharides in the reaction mixture after 24 h. The annotated *m/z* values indicate sodium adducts. None of the labeled peaks were observed in the negative control (i.e., a reaction without added enzyme).

To investigate the role of acetylation on the activity of AMOR_GH10A, acetylated and deacetylated versions of BGX were generated in-house from steam-exploded birch (30) and subsequently incubated with AMOR_GH10A (note that the commercial BGX used in all other experiments described in the present study is deacetylated). Product analysis by HPAEC-PAD and MALDI-TOF MS (Fig. 6) showed that acetylation reduces the efficiency of AMOR_GH10A, as reflected by lower production of unsubstituted XOS such

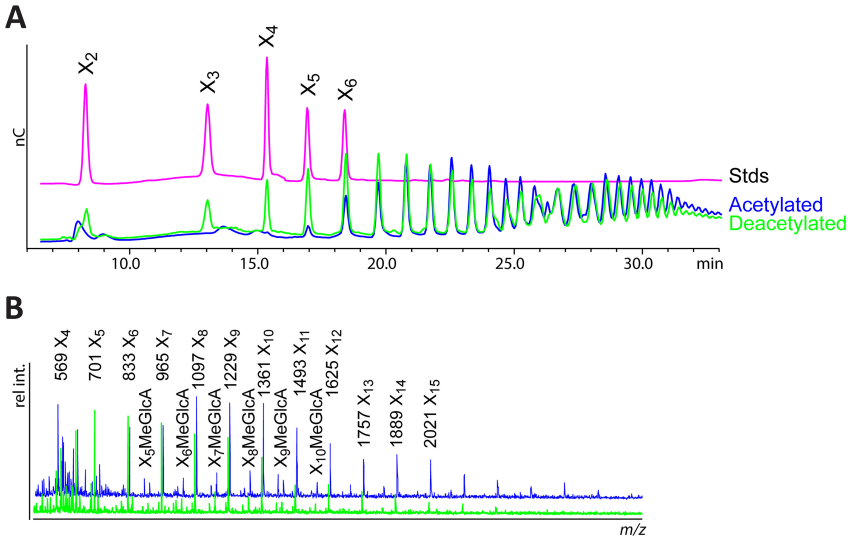

FIG 6 Effect of acetylation on hydrolysis of birchwood glucuronoxylan by AMOR_GH10A. (A) HPAEC-PAD analysis of products generated by AMOR_GH10A from acetylated (blue) and deacetylated (green) birchwood glucuronoxylan. These substrates were generated in house and differ from the commercial BGX used in all other experiments in this study. Note that, to facilitate analysis, products generated from the acetylated substrate were subjected to a deacetylation protocol, prior to HPAEC-PAD. A standard sample containing xylo-oligosaccharides (XOS; X_2 to X_6) is shown in pink. Reaction conditions: 6 μM AMOR_GH10A, 5% (wt/vol) substrate in NaOAc buffer (pH 5.0), incubated for 24 h at 37°C at 800 rpm. (B) MALDI-TOF analysis of the products. The annotated *m/z* values indicate sodium adducts. None of the labeled peaks were observed in the negative control (i.e., a reaction without added enzyme).

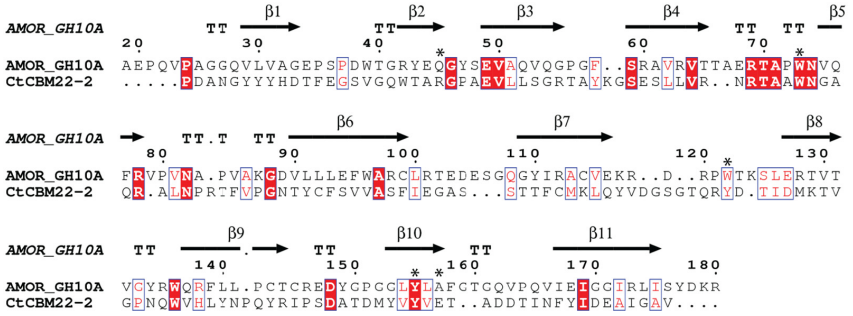

FIG 7 Sequence alignment of the CBM of AMOR_GH10A and *Ct*CBM22-2. The alignment was obtained by superposing a structural model of the CBM of AMOR_GH10A and the crystal structure of *Ct*CBM22-2. The sequence numbering refers to AMOR_GH10A. Conserved residues are printed in white on a red background, and conservatively substituted residues are printed in red. The predicted secondary structure elements of AMOR_GH10A are shown above the sequence alignment, where β-strands are rendered as arrows, and strict β-turns are represented by the letters TT. Positions marked by black stars represent *Ct*CBM22-2 residues previously shown to be involved in ligand binding (66). The figure was constructed using ESPript 3.0. For clarity, the wild-type glutamic acid (E) is shown in position 138 of *Ct*CBM22-2 (corresponding to A157 in AMOR_GH10A), thus replacing the alanine residue present in the E138A mutant of *Ct*CBM22-2 that was crystallized (PDB 1H6Y).

as X_3, X_4, X_5, and X_6 (Fig. 6A). MALDI-TOF analysis (inset) confirmed that glucuronosylated XOS (X_5MeGlcA-X_{10}MeGlcA) were generated from both acetylated and deacetylated substrates, in addition to unsubstituted XOS.

Confirming the remarkably wide substrate specificity shown in Fig. 4, Fig. S4 shows that AMOR_GH10A is capable of converting both xylopentaose (X_5) and cellopentaose (G_5) to a mixture of monomeric to tetrameric products.

Kinetics of oligosaccharide degradation. As suggested by the data provided in Fig. S4, steady-state kinetics showed much better performance of AMOR_GH10A toward cellopentaose compared to xylopentaose (Fig. S5). The experimental data fitted excellently to the Michaelis-Menten equation ($R^2 > 0.98$; Fig. S5) and yielded K_m values of 0.48 and 2.43 mM and k_{cat} values of 11 and 0.3 min^{-1} for cellopentaose and xylopentaose, respectively.

Identification and characterization of the CBM. In addition to the GH10 catalytic domain, AMOR_GH10A contains an N-terminal domain whose function could not be predicted through BLAST searches nor by dbCAN. However, a protein domain similar to residues 19 to 180 was identified using the SWISS-MODEL server (31), which identifies remote homologs based on both sequence and predicted structural similarities. The number 1 hit was a type 2 CBM22 domain (CBM22-2) from the *C. thermocellum* xylanase Xyn10B (*Ct*CBM22-2, PDB 1H6Y; 13.9% sequence identity), followed by the CBM22 domain from *Paenibacillus barcinonensis* xylanase Xyn10C (*Pb*CBM22-2, PDB 4XUN). Building a homology model based on the *Ct*CBM22-2 structure resulted in a relatively low SWISS-MODEL global model quality estimation score of 0.49. Figure 7 shows a sequence alignment of the CBM from AMOR_GH10A and *Ct*CBM22-2.

To investigate the possible role of the putative CBM in AMOR_GH10A, a gene fragment encoding residues 19 to 180 from AMOR_GH10A was cloned, and the gene product was expressed as a C-terminally His-tagged protein. The resulting protein was purified to electrophoretic homogeneity (Fig. S2), and the final yield was in the order of 35 mg of purified protein per liter *E. coli* culture. Figure 8 shows the binding of the recombinantly expressed putative CBM to various insoluble substrates. The putative CBM bound very efficiently to PASC (~80% binding after 1 min of coincubation with the substrate) and also showed clear binding to Avicel, birchwood glucuronoxylan (BGX) and, less so, ivory nut mannan (INM). Binding to chitin was very weak. This shows that the region from residues 19 to 180 of AMOR_GH10A indeed has a CBM function and defines a novel CBM family in CAZy. Binding experiments with full-length AMOR_

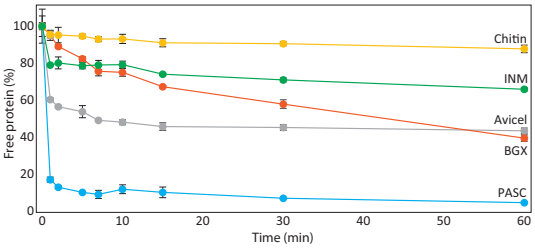

FIG 8 Binding of the putative CBM from AMOR_GH10A to various insoluble substrates. The figure shows binding of the CBM to phosphoric acid-swollen cellulose (PASC) in light blue, Avicel in gray, birchwood glucuronoxylan (BGX) in orange, chitin in yellow, and ivory nut mannan (INM) in green. The CBM of AMOR_GH10A at 0.1 mg/ml was incubated with the substrates at 0.1% (wt/vol) for 60 min. The incubation was carried out at 22°C and 1,000 rpm in NaOAc buffer (pH 5.6) supplemented with 0.5 M NaCl. Bound protein was removed by filtration, and the remaining unbound protein was measured by A_{280}. The absorbance values are shown as the percentage of the absorbance value for a control reaction without added substrate.

GH10A gave similar results, showing that binding affinity is primarily determined by the CBM (data not shown).

Studies of binding to soluble polysaccharides using affinity polyacrylamide gel electrophoresis (affinity PAGE) (Fig. 9) showed that the CBM binds well to barley β-glucan (BG) and konjac glucomannan (KGM). Somewhat surprisingly, in light of the activity data presented in Fig. 4, binding to wheat arabinoxylan (WAX) or tamarind xyloglucan (TXG) was not detected using this approach.

DISCUSSION

There is an increasing industrial interest in xylanases for applications in animal nutrition, wood, pulp biobleaching, and biomass degradation. From an economical standpoint, there is an advantage in using multifunctional enzymes capable of operating at harsh conditions, such as elevated temperatures. We describe here the characterization of AMOR_GH10A, a thermostable GH10 xylanase from a deep-sea metagenomic data set originating from the hot vents at the Arctic Mid-Ocean Ridge. The enzyme consists of a previously uncharacterized CBM domain and a downstream GH10 catalytic domain. The temperature and pH optima of AMOR_GH10A were 80°C and 5.6, respectively. Similar optima have been observed for other carbohydrate-active enzymes from the deep sea (17, 32, 33). AMOR_GH10A is halophilic, and Fig. 3C and D show that the beneficial effect of salt is due to effects on both activity and stability.

Reported features of halophilic enzymes include increased usage of negatively charged amino acids on the surface, and an increased arginine/lysine ratio (29, 34). The

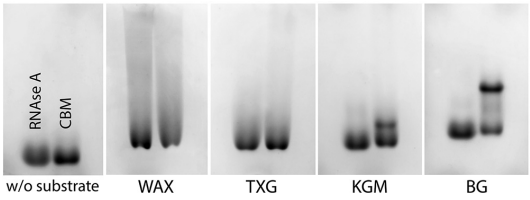

FIG 9 Binding of the CBM from AMOR_GH10A to soluble substrates. The pictures show the result of acidic affinity PAGE with the soluble polysaccharides wheat arabinoxylan (WAX), tamarind xyloglucan (TXG), konjac glucomannan (KGM), and barley β-glucan (BG) and a control gel without added substrate. The proteins loaded were AMOR_GH10A_CBM (2 μg, right lane in each of the gels) and 2 μg of the reference protein, RNase A (left lane in each gel). The four gels to the right contained 0.1% (wt/vol) of the indicated polysaccharide. The ratio of the migration distance of the CBM in the presence of polysaccharide to that in the absence of polysaccharide was divided by the corresponding ratio for the reference protein and was below the cutoff value of 0.85 (67) in two cases, 0.67 for β-glucan and 0.82 for konjac glucomannan, indicating significant binding of the CBM to these two substrates.

number of negatively charged residues predicted to be located on the surface of the AMOR_GH10A catalytic domain is not significantly different compared to the structurally related enzymes shown in Fig. 2. However, the GH10 domain of AMOR_GH10A shows an unusually high arginine/lysine ratio (26:7 compared to on average 16:20 in the three GH10s shown in Fig. 2).

AMOR_GH10A produced a variety of substituted and unsubstituted xylooligosaccharides from birchwood glucuronoxylan, a typical feature of GH10 xylanases with an endo mode of action (35). Acetylations have a slight inhibiting effect on AMOR_GH10A activity (Fig. 6), which has been observed previously for GH10 xylanases (30). Overall, products generated by AMOR_GH10A from BGX (Fig. 5 and 6) corresponded well with expected GH10 endoxylanase products. However, unlike classical GH10 xylanases, which have no or very low activity on β–1,4-linked glucans, AMOR_GH10A was surprisingly active on glucomannan (KGM), amorphous cellulose (PASC), and tamarind xyloglucan (TXG; Fig. 4). Steady-state kinetics with oligomeric substrates yielded low k_{cat} values (0.3 to 11 min^{-1}) compared to the apparent initial rates observed for polymeric substrates (in the order of 2 to 4 s^{-1}). Confirming the β-glucanase activity of AMOR_GH10A, the enzyme was much more active on cellopentaose (G_5) compared to xylopentaose (X_5), with a 35-fold-higher k_{cat} and a 5-fold-lower K_m.

Some GH10 xylanases display activity on cellulose but this activity is clearly lower, compared to the activity on xylan (36–38). In this respect, AMOR_GH10A is a clear and unique exception. Notenboom et al. (37) have proposed that two structural features are responsible for the cellulolytic side activity of xylanase A from *Cellulomonas fimi* (CfXyn10A, formerly known as Cex), which hydrolyzes xylan 40-fold more efficiently than cellulose. In the –2 subsite, Gln87 moves to a position different from that in non-cellulose-active GH10s to accommodate the exocyclic C-6 OH of glucose. This glutamine residue is conserved in *Cm*Xyn10B, *Sl*XynA, and TmxB, but not in AMOR_GH10A, which has a tryptophan in this position (Fig. 2; residue 325). Neighboring residues also differ quite substantially between AMOR_GH10A and the other three xylanases, suggesting that the –2 subsite of enzymes may look quite different. In the –1 subsite of CfXyn10A, Trp281 is thought to sterically hinder the binding of glucosides, and this residue must rearrange to allow binding of the C-6 OH group of glucose. Strikingly, the region including this conserved tryptophan residue and present in all known GH10 xylanases characterized to date appears to be lacking in AMOR_GH10A, meaning that its –1 subsite must differ substantially from other GH10 xylanases. While structural studies are needed to shed light on the structural determinants of the broad substrate specificity of AMOR_GH10A, the sequence comparisons suggest that the unique properties of this enzyme are accompanied by considerable structural deviations relative to common GH10 xylanases.

It must be noted that the broad specificity of AMOR_GH10A, possibly caused by atypical features of the –2 and –1 subsites, as discussed above, comes at a cost. The observed catalytic rates toward xylan at a substrate concentration in the same range as typically observed K_m values for mesophilic and thermophilic GH10 xylanases (15, 17, 38–41) are at least an order of magnitude lower compared to most (15, 17, 38–40), albeit not all (41), previously described mesophilic and thermophilic xylanases.

An N-terminal domain of AMOR_GH10A of unknown function was predicted to have a CBM-like structure and was experimentally shown to bind to amorphous cellulose (PASC), microcrystalline cellulose (Avicel), and birchwood glucuronoxylan (Fig. 8), as well as the soluble polysaccharides barley β-glucan (BG) and konjac glucomannan (KGM) (Fig. 9). Consequently, this domain represents a novel CBM and has been annotated as CBM85 in the CAZy database. Compared to CBMs with a known structure, the CBM85 from AMOR_GH10A seems to be most similar to a CBM22 domain (*Ct*CBM22-2) from *C. thermocellum* xylanase (19). *Ct*CBM22-2 interacts with a wide range of xylans and xylooligosaccharides, and experiments have revealed the roles of several residues in the ligand-binding cleft in binding of xylooligosaccharides (Fig. 7). Among these are Trp53 and Tyr103, which are residues that typically interact with carbohydrate ligands through hydrophobic stacking interactions. These residues

correspond to Trp73 and Trp122 in the CBM85 (Fig. 7), meaning that hydrophobicity and aromaticity are conserved. Furthermore, residues Arg25, Tyr136, and Glu138 have been implicated in ligand binding through hydrogen bonding. Of these, only Tyr136 is conserved in the CBM85 (Tyr155). Arg25 and Glu138 in the CBM22 align with Gln45 and Ala157 in the CBM85 (Fig. 7).

It is not uncommon for CBMs to have broad binding specificities. For example, xylan-binding CBM4 and CBM6 also bind to amorphous cellulose (42, 43), while a CBM22 has been shown to bind microcrystalline cellulose (44). Uniquely for CBM85, its affinity to amorphous cellulose seems exceptionally high (Fig. 8), whereas binding to Avicel is much weaker. Binding to soluble β-glucan and konjac glucomannan, both unsubstituted polysaccharides, was also observed (Fig. 9). Thus, it would seem that CBM85, like CBM22 (44, 45), preferentially interacts with individual glucan chains, and binds to the more amorphous regions of crystalline cellulose. While AMOR_GH10A showed activity toward wheat arabinoxylan (WAX) or tamarind xyloglucan (TXG) we could not detect binding of the CBM to these substrates in the affinity PAGE experiments. This could reflect a technical limitation in the affinity PAGE experiment or a temperature effect, since binding studies were done at a temperature quite different from the temperature used in the activity assays. It is also possible that for these two substrates, the catalytic domain dominates binding affinity. Since repeated efforts to express the AMOR_GH10A catalytic domain alone failed, the substrate-binding affinity of this domain alone remains unknown.

To shed light on the possible distribution of CBM85 domains, and the potential conserved association of the CBM of AMOR_GH10A and GH10 domains similar to the catalytic domain of AMOR_GH10A, the amino acid sequences of the 12 putative GH10 enzymes shown in Fig. 1B were aligned and investigated for the presence of conserved protein domains. Of these, 10 sequences were found to contain putative domains with a significant similarity (>25% sequence identity) to CBM85 and the aligned sequences are shown in the supplemental material (Fig. S6). The sequences of these putative CBMs found in AMOR_GH10A homologues are 28 to 34% identical to the CBM85 of AMOR_GH10A. The large majority of hits obtained upon BLAST searches with the CBM85 sequence concerned domains of unknown function (closest hit: 39% sequence identity to a hypothetical protein from a *Verrucomicrobia* bacterium), which supports the conclusion that CBM85 represents a novel CBM family. BLASTp searches with the CBM85 sequence gave >150 hits and the large majority of the detected proteins contained a GH10 domain, although other domains (GH5 and GH16) were also encountered. The majority of microbes harboring these proteins were from aquatic environments, and several were classified as halophilic or halotolerant.

The function of the cryptic sequence separating the CBM and catalytic domain of AMOR_GH10A remains elusive due to the lack of similarities to sequences of known function and because it lacks the repetitive patterns commonly found in linkers (46). However, this part of AMOR_GH10A has similarities to the corresponding parts of the 12 phylogenetically related enzymes depicted in Fig. 1, implying that these sequences could play a role in AMOR_GH10A functionality. A comparison of the C-terminal part downstream of the GH10 domain to the same set of 12 homologs revealed that this part was most similar to the C-terminal parts of *Opitutaceae* bacterium TSB47 and *Opitutus terrae* (sequence identities: 46 and 45%, respectively). More detailed studies are needed to elucidate the function of these parts of the enzyme, which could affect enzyme activity and/or stability.

In conclusion, AMOR_GH10A represents one of few well-characterized carbohydrate-active enzymes isolated from a deep-sea hot vent and displays several interesting features such as high stability, halophilicity, a novel type of CBM, and a broad substrate range. Some of these features are unique compared to other GH10 xylanases characterized to date. First, AMOR_GH10A contains the founding member of a new CAZy family of CBMs, CBM85, which binds to both xylan and glucan substrates, with a preference for amorphous material. Second, and in line with the broad binding specificity of the CBM85, AMOR_GH10A has a broad substrate

specificity. GH10 enzymes are generally considered xylanases, and the much higher activity of AMOR_GH10A on cellopentaose compared to xylopentaose is truly remarkable and unprecedented. Further studies are needed to unravel the structural features underlying the special properties of AMOR_GH10A and to shed more light on the largely uncharacterized biodiversity that targets polysaccharides in deep-sea hot vents.

MATERIALS AND METHODS

Sampling, DNA extraction, and sequencing. The substrate used for generating the metagenomics data set in this study was derived from unbleached Norway spruce (*Picea abies*) which had been subjected to a sulfite pulping pretreatment method termed the BALI process (47, 48), developed at Borregaard AS (Sarpsborg, Norway). The substrate was dried at 40°C overnight and the particle size was reduced in a planetary ball mill PM 100 (Retsch, Haan, Germany), followed by sieving through a 0.85-mm screen. This substrate had a glucan content of 85%, while hemicelluloses (3% mannan and xylan, respectively) and acid insoluble lignin comprised 11%. At the Arctic Mid-Ocean Ridge (AMOR), the Jan Mayen hydrothermal vent field is situated on the Mohn's Ridge north of the Island Jan Mayen at 71°N (49, 50). During a research cruise at the Mohn's Ridge in 2014, a titanium chamber (length, 2.5 cm; volume, 16 ml; 1-mm pores) was filled with the industrial substrate (1 g of substrate mixed with ~16 ml of sediment sampled from the same vent site) and deployed in ~70°C hot sediments at ~570 m below sea level (71°17.904′N 5°42.257′W). The incubator was deployed and collected using a remote operating vehicle on board the research vessel *G.O. Sars*. After incubating the chamber *in situ* for 1 year, DNA was extracted from 6.9 g of the material inside the chamber using the FastDNA spin kit for soil (MP Biomedicals, Santa Ana, CA), according to the manufacturer's protocol. Metagenomic sequencing of total DNA was carried out using Illumina MiSeq 300 paired-end chemistry at the Norwegian Sequencing Centre (www.sequencing.uio.no). In total, 1.1 μg of DNA was submitted for sequencing.

Filtering, assembly, and ORF prediction. Filtering of raw Illumina MiSeq 300 paired-end reads and assembly were performed using the CLC genomics workbench (Qiagen, v.10). Filtering was done using the CLC default parameters (quality 0.05; length, minimum 40 and maximum 1,000 nucleotides). In addition, one nucleotide was removed from terminal read ends. Assembly was done using default parameters with an automatic k-mer size and bubble size. A minimum contig length was set to 1,000 bases with scaffolding enabled. ORF prediction was performed using the *-p meta* option in Prodigal (51, 52) for metagenomic data sets. Full details of the metagenomic data set will be published elsewhere.

Bioinformatic analysis. The metagenomic data set was mined for putative glycosyl hydrolases belonging to family GH10 using dbCAN (csbl.bmb.uga.edu/dbCAN [53]). This analysis resulted in the identification of a 1,767-bp gene encoding a putative GH10 xylanase, *amor_gh10a*. The AMOR_GH10A amino acid sequence was analyzed for the presence of a signal peptide using SignalP (54), and a search for homologous GH10 domains was performed against a customized UniProt GH10 database. Bacterial and archeal GH10 sequences were downloaded from the UniProt knowledge base and subjected to cd-hit (55, 56) to reduce the sequence redundancy with a 95% sequence identity cutoff. The AMOR_GH10A sequence similarities were assessed using a stand-alone version of the BLASTp algorithm, and a multiple sequence alignment was performed in MEGA 7.0.26 (57). Sequences not containing the conserved catalytic residues were removed from the alignment (58). In total, 42 sequences were used for construction of a phylogenetic tree using RAxML v.8.2.9 (59). The confidence for tree topologies was estimated by bootstrap values based on 1,000 replicates. Homologs of the GH10 domain with known three-dimensional structures were aligned using Clustal Omega (60). SWISS-MODEL (https://swissmodel.expasy.org [31]) was used to construct a structure model based on the closest PDB homolog (PDB 1UQY [22]), and the predicted structure was visualized in PyMOL (Schrödinger). Putative homologs of the N-terminal AMOR_GH10A domain of unknown function (residues 19 to 180) were identified and aligned using the SWISS-MODEL server. All sequence alignments were visualized using ESPript 3 (61).

Cloning, expression, and purification of AMOR_GH10A. The *amor_gh10a* gene (codon optimized for *E. coli* expression) was synthesized by GenScript (Piscataway, NJ) and bp 55 to 1767 (omitting the predicted 18 amino acid signal peptide) was amplified by PCR using Q5 DNA polymerase (New England Biolabs, Ipswich, MA) and forward primer 5′-TTAAGAAGGAGATATACTATGGCTGAACCACAAGTTCCGGC-3′, where the nucleotides utilized as a complementary sequence for ligation-independent cloning are underlined. Two different reverse primers were used, one to amplify the full-length gene (*amor_gh10a*) with sequence 5′-AATGGTGGTGATGATGGTGCGCACGCAGCGGCAGGGTCAGT-3′ and one to amplify the N-terminal gene fragment putatively encoding a CBM (*amor_gh10acbm*) with sequence 5′-AATGGTGGTGATGATGGTGCGCACGTTTGTCGTAGCTGATCAGAC-3′. The resulting PCR products were separated on a 1% agarose gel and purified using the Nucleospin gel and PCR Clean-Up kit from Macherey-Nagel (Düren, Germany). Purified DNA fragments were cloned into the pNIC-CH expression vector (AddGene, Cambridge, MA) by ligation-independent cloning (62), followed by heat shock transformation into chemically competent OneShot *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA). As a result of this cloning strategy, the N terminus of the (signal peptide-free) protein is extended with a methionine, while a seven-residue His tag (AHHHHHH) is added at the C terminus. The cells were propagated in SOC medium for 60 min prior to plating on lysogenic broth (LB) agar supplemented with 5% sucrose and 50 μg/ml kanamycin. The plated cells were incubated overnight (ON) at 37°C. Colonies were picked and inoculated in liquid LB containing 50 μg/ml kanamycin and incubated overnight at 37°C with shaking at 200 rpm. The plasmids were isolated from the culture using the NucleoSpin plasmid kit from Macherey-Nagel, and the sequence of the cloned DNA fragments were confirmed by Sanger sequencing (GATC, Constance, Germany). The

pNIC-CH plasmids harboring *amor_gh10a* and *amor_gh10acbm* were further transformed into chemically competent OneShot BL-21 Star (DE3) *E. coli* cells, as described by the supplier (Invitrogen). The cells were propagated in SOC medium, plated on LB agar supplemented with 50 μg/ml kanamycin, and subsequently incubated at 37°C overnight.

Correct transformants were inoculated and grown in Terrific broth (TB) supplemented with 50 μg/ml kanamycin in a Harbinger system (Harbinger Biotechnology & Engineering, Markham, Canada) at 23°C overnight. Induction of protein expression was achieved by addition of 0.2 mM IPTG to the cells, followed by an additional incubation for 24 h. Then, 500 ml of cell culture was harvested by centrifugation at 5,000 $\times$ *g* for 15 min at 4°C using a Beckman coulter centrifuge (Brea, CA). To promote cell lysis, the cell pellet was frozen for at least 1 h at minus 80°C. The cells were subsequently thawed and resuspended in 60 ml of 50 mM Tris-HCl (pH 8.0) supplemented with 500 mM NaCl and 5 mM imidazole, followed by sonication on ice using a Vibracell sonicator (Sonics & Materials, Inc., Newtown, CT) with 5-s on/off pulses for 3 min at 30% amplitude. Cell debris was removed by centrifugation at 15,000 $\times$ *g* for 15 min at 4°C, and the supernatants containing the soluble proteins were filtered using a sterile 0.45-μm syringe filter (Sarstedt, Nümbrecht, Germany).

Cell-free protein fractions were subjected to immobilized metal affinity chromatography using an Äkta pure chromatography system and a Ni^{2+} affinity HisTrap HP 5-ml column (GE HealthCare, Chicago, IL). The His-tagged proteins were eluted using a linear gradient of 5 to 500 mM imidazole in 50 mM Tris-HCl (pH 8.0) supplemented with 500 mM NaCl. Protein fractions were analyzed by SDS-PAGE (Bio-Rad, Hercules, CA), and fractions containing the overexpressed protein were combined and concentrated using a 10,000 MWCO Vivaspin ultrafiltration tube (Sartorius, Göttingen, Germany), with concomitant buffer exchange to 50 mM sodium acetate (NaOAc) buffer (pH 5.6) supplemented with 500 mM NaCl. The protein concentration was measured by recording absorbance at 280 nm with a Biophotometer (Eppendorf, Hamburg, Germany) and by using the theoretical extinction coefficient (web.expasy.org/protparam) to calculate the concentration. The purified His-tagged protein was stored at 4°C and was used as is in subsequent studies.

Apparent melting temperature. A Nano-Differential Scanning Calorimeter III (Calorimetry Sciences Corporation, Lindon, UT) was used to determine the apparent melting temperature [$T_{m(app)}$] of AMOR_GH10A. Approximately 1 mg/ml of AMOR_GH10A was dialyzed overnight at 4°C in 25 mM NaOAc buffer (pH 5.6) with or without 0.5 M NaCl. The protein concentration was adjusted to 0.5 mg/ml, and both the protein solution and the dialysis buffer were subsequently filtered (0.22-μm pore size) and degassed for 10 min. The dialysis buffer was used to record the buffer baselines, which were subtracted from the protein scans. The scan rate was set to 1°C/min from 25 to 110°C. The obtained data were analyzed with NanoAnalyze software (TA Instruments, New Castle, DE).

Substrates. Enzyme activity and binding were assessed on a wide range of carbohydrate substrates. Birchwood glucuronoxylan (BGX) was purchased from Carl Roth GmbH (Karlsruhe, Germany). Avicel PH-101 microcrystalline cellulose was purchased from Sigma-Aldrich (St. Louis, MO). Phosphoric acid-swollen cellulose (PASC) was prepared from Avicel essentially as described previously (63). Wheat flour arabinoxylan (WAX), tamarind xyloglucan (TXG), konjac glucomannan (KGM), ivory nut mannan (INM), barley β-glucan (BG), xylooligosaccharides (DP2 to DP6), and cellooligosaccharides (DP2 to DP6) were all purchased from Megazyme (Wicklow, Ireland). Commercial chitin prepared from shrimp (*Pandalus borealis*) shells was purchased from Chitinor AS (Senjahopen, Norway).

To study the effect of acetylation, freeze-dried Birchwood acetyl glucuronoxylan was prepared in house from steam-exploded birch as previously described (30) (note that all other experiments involving BGX were done with the commercially available BGX described above). A deacetylated version of this substrate was prepared by incubation in 200 mM NaOH at 4°C for 24 h. The deacetylated polysaccharides were subsequently precipitated by adjusting the liquid to 80% (vol/vol) ethanol with ice-cold ethanol, followed by incubation at 4°C for 10 min and subsequent centrifugation at 2,000 $\times$ *g* for 15 min at 4°C. The precipitate was washed four times with ice-cold 80% ethanol to remove the NaOH, and the material was dissolved in 50 mM NaOAc buffer (pH 5.0).

Activity assays. To determine the optimum temperature and pH for AMOR_GH10A activity, reactions were performed using citrate-phosphate buffer (pH 3.7 to 7.3) or glycine-NaOH buffer (pH 8.5), both supplemented with 0.5 M NaCl. In setting up the reactions buffer solutions were diluted 2-fold, leading to a final NaCl concentration of 0.25 M. The final enzyme and substrate (BGX) concentrations were 2 μM and 0.5% (wt/vol), respectively. Reaction mixtures were incubated in a Thermomixer (Eppendorf) for 10 min at 600 rpm using various temperatures (25 to 90°C) at pH 5.6 or at different pH values at 80°C. The effect of the NaCl concentration on enzyme activity was investigated by running reactions at 80°C in 25 mM NaOAc buffer (pH 5.6) supplemented with various concentrations of NaCl (0 to 3.0 M). The xylanase activity was determined by measuring the release of reducing sugars using the 3,5-dinitrosalicylic acid (DNS) reagent (64) and xylose as a standard.

Kinetic parameters were determined using the oligosaccharides xylopentaose (X_5) and cellopentaose (G_5) as the substrates. AMOR_GH10A was incubated for 60 min with different substrate concentrations ranging from 0.05 to 3.2 mM. The reactions were carried out at 60°C in 25 mM NaOAc (pH 5.6) supplemented with 0.5 M NaCl. For G_5, the enzyme loading was 0.5 μM, while for X_5 the enzyme dosage was increased 100-fold to 50 μM. After 15, 30, 45, and 60 min, aliquots were withdrawn and mixed with an identical volume of stopping reagent (0.1 M NaOH). All reactions were performed in triplicates. The samples were analyzed by HPAEC-PAD (see below), and initial reaction rates were calculated based on the depletion of the substrate peak area. GraphPad was used to construct Michaelis-Menten plots and to calculate k_{cat}, V_{max}, and K_m values for each substrate using nonlinear regression.

Substrate specificity. Substrate specificity was determined by incubating 2 μM enzyme with 0.5% (wt/vol) BGX, WAX, TXG, KGM, or PASC, in 25 mM NaOAc buffer (pH 5.6) supplemented with 0.5 M NaCl at 60°C and 600 rpm. The reactions were stopped at various time points by boiling for 15 min, followed by centrifugation at 10,000 $\times$ *g* for 5 min. The concentrations of reducing ends in the resulting supernatants were determined using the DNS method and xylose as the standard.

In an alternative experiment set up to determine the initial rate of AMOR_GH10A on various substrates, 1 μM enzyme was incubated with 0.5% (wt/vol) BGX, WAX, or PASC, in 25 mM NaOAc buffer (pH 5.6) supplemented with 0.5 M NaCl at 80°C and 600 rpm for 10 min. Samples were withdrawn at 2-min intervals and immediately mixed with the DNS reagent.

Product analysis. AMOR_GH10A was mixed with either 0.5% (wt/vol) BXG, 1 mM cellopentaose (G_5), or 1 mM xylopentaose (X_5) in 25 mM NaOAc buffer (pH 5.6), supplemented with 0.5 M NaCl, and incubated at 60°C at 600 rpm. The enzyme concentrations were 2 and 1 μM in the BXG and the oligosaccharide reactions, respectively. Reactions were stopped by boiling for 15 min at various time points, and the boiled reaction mixtures were filtered through 0.45-μm filters before further analysis.

For analysis of products generated from in-house-prepared acetylated and deacetylated BGX, 6 μM AMOR_GH10A was incubated with 5% (wt/vol) of either acetylated or deacetylated BGX in 50 mM NaOAc buffer (pH 5.0) at 37°C for 24 h at 800 rpm. Reaction mixtures containing deacetylated BGX were stopped by boiling for 15 min. Reaction mixtures containing acetylated substrate were subjected to deacetylation (and thereby stopped) using the deacetylation procedure described above, followed by boiling for 15 min. All reactions were filtered prior to analysis.

Product analysis by HPAEC-PAD. Xylose, glucose, and soluble xylo- and cello-oligomers (DP2 to DP6) were analyzed by HPAEC using a Dionex ICS3000 system (Thermo Scientific, San Jose, CA), equipped with PAD and a CarboPac PA1 analytical column (2 by 250 mm). Products were separated by applying a stepwise gradient with an increasing amount of eluent B (0.1 M NaOH and 1 M NaOAc; eluent A is 0.1 M NaOH), as follows: 0 to 10% B over 10 min, 10 to 30% B over 25 min, 30 to 100% B over 5 min, 100 to 0% B over 1 min, and 0% B over 9 min. Data collection and analysis were carried out with the Chromeleon 7.0 software, and xylo- and cellooligosaccharide standards in the DP1 to DP6 range were used to quantify the products. Reaction products generated from acetylated and deacetylated birchwood glucuronoxylan were analyzed with a Dionex ICS 5000 system using the same column type and elution gradient as described above.

Product analysis by MALDI-TOF MS. Qualitative analysis of product formation from BGX was achieved by using a MALDI-TOF UltrafleXtreme mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) equipped with a nitrogen 337-nm laser. Portions (1 μl) of each sample were mixed with 2 μl of 9 mg/ml 2,5-dihydroxybenzoic acid (DHB) dissolved in 30% acetonitrile and applied to an MTP 384 ground steel target plate TF (Bruker Daltonics). Spectra were collected using Bruker FlexControl software and analyzed with Bruker flexAnalysis software.

Binding assays. Protein binding to various insoluble substrates was investigated using a slightly modified version of a previously published method (65). Briefly, reaction mixtures containing 0.1% (wt/vol) of carbohydrate and 0.1 mg/ml protein in 25 mM NaOAc buffer (pH 5.6) supplemented with 0.5 M NaCl were incubated at 22°C and 1,000 rpm. At various time points samples were taken and filtered using a 0.45-μm-pore-size filter, and protein binding was determined by measuring the A_{280} in the filtrate (unbound protein). Protein concentrations were calculated using theoretical extinction coefficients. Protein concentrations in reaction mixtures lacking substrate were used as negative controls (0% binding).

Affinity PAGE was employed to assess binding of the CBM to soluble substrates. After incorporating different soluble polysaccharides in the gels, the effect of the polysaccharides on the migration rate of the CBM was investigated under nondenaturing conditions. The CBM did not migrate under standard (basic) native gel electrophoresis conditions (for which established protocols exist for CBM studies [66, 67]), presumably due to a high isoelectric point (the theoretical pI is 8.4). Therefore, an acidic native gel protocol called RENAGE, developed for prion protein oligomers, was adapted and used instead (68). The experiment was set up as described previously by Ladner and Wishart, except that the separating gel contained 12% polyacrylamide and was supplemented with 0.5 M NaCl. The gels were polymerized in the presence of various polysaccharides (KGM, TXG, WAX, or barley β-glucan [BG]), which were added at a final concentration of 0.1% (wt/vol). The gels were run for 4 h at 80 V in a Mini Protean II system (Bio-Rad) before staining with Coomassie blue. A control gel without added substrate was also run. RNase A (Thermo Scientific) was used as a reference protein in all gels since this enzyme has a high pI (calculated to 9.6) and migrated with a rate similar to the CBM in the acidic gel. The reference protein was included in the gels to account for possible variation in gel pore size caused by the presence of polysaccharides in the gel matrix. To determine whether the CBM interacted with the polysaccharides, the ratio of the migration distance of the CBM in the presence of a polysaccharide to that in the control gel (without polysaccharide) was divided by the corresponding ratio for the reference protein, RNase A. In the case of little to no binding, this would result in a value of approximately 1, while interaction of the CBM with the polysaccharide would lead to ratios lower than 1. Values of $<$0.85 were considered as positive binding (67).

Data availability. The sequence of AMOR_GH10A has been submitted to GenBank under accession number MH727997 and archived under BioProject PRJNA266758 and BioSample SAMN09768207.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02970-18.

SUPPLEMENTAL FILE 1, PDF file, 1.4 MB.

ACKNOWLEDGMENTS

This study was funded by the Research Council of Norway through NorZymeD project 221568. The infrastructure was in part supported by NorBioLab grants 226247 and 270038 provided by the Research Council of Norway. B.W. was funded by Research Council project 244259.

We thank Rolf-Birger Pedersen at the University of Bergen for leading the Centre for Geobiology research cruise to the Arctic Mid-Ocean Ridge vent fields and Gustav Vaaje-Kolstad for advice regarding some of the experiments.

REFERENCES

1. Rennie EA, Scheller HV. 2014. Xylan biosynthesis. Curr Opin Biotechnol 26:100–107. https://doi.org/10.1016/j.copbio.2013.11.013.
2. Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annu Rev Plant Biol 61:263–289. https://doi.org/10.1146/annurev-arplant-042809-112315.
3. Nguyen STC, Freund HL, Kasanjian J, Berlemont R. 2018. Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. Appl Microbiol Biotechnol 102:1629–1637. https://doi.org/10.1007/s00253-018-8778-y.
4. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:490–495.
5. Biely P, Singh S, Puchart V. 2016. Towards enzymatic breakdown of complex plant xylan structures: state of the art. Biotechnol Adv 34:1260–1274. https://doi.org/10.1016/j.biotechadv.2016.09.001.
6. Collins T, Gerday C, Feller G. 2005. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev 29:3–23. https://doi.org/10.1016/j.femsre.2004.06.005.
7. Chakdar H, Kumar M, Pandiyan K, Singh A, Nanjappan K, Kashyap PL, Srivastava AK. 2016. Bacterial xylanases: biology to biotechnology. 3 Biotech 6:150. https://doi.org/10.1007/s13205-016-0457-z.
8. Chu YD, Tu T, Penttinen L, Xue XL, Wang XY, Yi ZL, Gong L, Rouvinen J, Luo HY, Hakulinen N, Yao B, Su XY. 2017. Insights into the roles of non-catalytic residues in the active site of a GH10 xylanase with activity on cellulose. J Biol Chem 292:19315–19327. https://doi.org/10.1074/jbc.M117.807768.
9. Andrews SR, Charnock SJ, Lakey JH, Davies GJ, Claeyssens M, Nerinckx W, Underwood M, Sinnott ML, Warren RA, Gilbert HJ. 2000. Substrate specificity in glycoside hydrolase family 10: tyrosine 87 and leucine 314 play a pivotal role in discriminating between glucose and xylose binding in the proximal active site of *Pseudomonas cellulosa* xylanase 10A. J Biol Chem 275:23027–23033. https://doi.org/10.1074/jbc.M000128200.
10. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781.
11. Herve C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. 2010. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proc Natl Acad Sci U S A 107:15293–15298. https://doi.org/10.1073/pnas.1005732107.
12. Beg QK, Kapoor M, Mahajan L, Hoondal GS. 2001. Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol 56:326–338.
13. Hu J, Arantes V, Pribowo A, Saddler JN. 2013. The synergistic action of accessory enzymes enhances the hydrolytic potential of a "cellulase mixture" but is highly substrate specific. Biotechnol Biofuels 6:112. https://doi.org/10.1186/1754-6834-6-112.
14. Verma D, Kawarabayasi Y, Miyazaki K, Satyanarayana T. 2013. Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. PLoS One 8:e52459. https://doi.org/10.1371/journal.pone.0052459.
15. Sinma K, Khucharoenphaisan K, Kitpreechavanich V, Tokuyama S. 2011. Purification and characterization of a thermostable xylanase from *Saccharopolyspora pathumthaniensis* S582 isolated from the gut of a termite. Biosci Biotechnol Biochem 75:1957–1963. https://doi.org/10.1271/bbb.110353.
16. Shi H, Li X, Gu H, Zhang Y, Huang Y, Wang L, Wang F. 2013. Biochemical properties of a novel thermostable and highly xylose-tolerant β-xylosidase/α-arabinosidase from *Thermotoga thermarum*. Biotechnol Biofuels 6:27. https://doi.org/10.1186/1754-6834-6-27.
17. Wu S, Liu B, Zhang X. 2006. Characterization of a recombinant thermostable xylanase from deep-sea thermophilic *Geobacillus* sp. MT-1 in East Pacific. Appl Microbiol Biotechnol 72:1210–1216. https://doi.org/10.1007/s00253-006-0416-4.
18. Miroshnichenko ML, Bonch-Osmolovskaya EA. 2006. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. Extremophiles 10:85–96. https://doi.org/10.1007/s00792-005-0489-5.
19. Xie H, Gilbert HJ, Charnock SJ, Davies GJ, Williamson MP, Simpson PJ, Raghothama S, Fontes CM, Dias FM, Ferreira LM, Bolam DN. 2001. *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. Biochemistry 40:9167–9176. https://doi.org/10.1021/bi0106742.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
21. Zverlov VV, Schantz N, Schmitt-Kopplin P, Schwarz WH. 2005. Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. Microbiology 151:3395–3401. https://doi.org/10.1099/mic.0.28206-0.
22. Pell G, Taylor EJ, Gloster TM, Turkenburg JP, Fontes CMGA, Ferreira LMA, Nagy T, Clark SJ, Davies GJ, Gilbert HJ. 2004. The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. J Biol Chem 279:9597–9605. https://doi.org/10.1074/jbc.M312278200.
23. Ducros V, Charnock SJ, Derewenda U, Derewenda ZS, Dauter Z, Dupont C, Shareck F, Morosoli R, Kluepfel D, Davies GJ. 2000. Substrate specificity in glycoside hydrolase family 10: structural and kinetic analysis of the *Streptomyces lividans* xylanase 10A. J Biol Chem 275:23020–23026.
24. Ihsanawati, Kumasaka T, Kaneko T, Morokuma C, Yatsunami R, Sato T, Nakamura S, Tanaka N. 2005. Structural basis of the substrate subsite and the highly thermal stability of xylanase 10B from *Thermotoga maritima* MSB8. Proteins 61:999–1009.
25. Derewenda U, Swenson L, Green R, Wei Y, Morosoli R, Shareck F, Kluepfel D, Derewenda ZS. 1994. Crystal structure, at 2.6-Å resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of β-1,4-D-glycanases. J Biol Chem 269:20811–20814.
26. Fontes CM, Gilbert HJ, Hazlewood GP, Clarke JH, Prates JA, McKie VA, Nagy T, Fernandes TH, Ferreira LM. 2000. A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. Microbiology 146:1959–1967. https://doi.org/10.1099/00221287-146-8-1959.
27. Xie H, Flint J, Vardakou M, Lakey JH, Lewis RJ, Gilbert HJ, Dumon C. 2006. Probing the structural basis for the difference in thermostability displayed by family 10 xylanases. J Mol Biol 360:157–167. https://doi.org/10.1016/j.jmb.2006.05.002.
28. Moreau A, Roberge M, Manin C, Shareck F, Kluepfel D, Morosoli R. 1994. Identification of two acidic residues involved in the catalysis of xylanase A from *Streptomyces lividans*. Biochem J 302:291–295. https://doi.org/10.1042/bj3020291.
29. Kastritis PL, Papandreou NC, Hamodrakas SJ. 2007. Haloadaptation: insights from comparative modeling studies of halophilic archaeal DHFRs. Int J Biol Macromol 41:447–453. https://doi.org/10.1016/j.ijbiomac.2007.06.005.
30. Biely P, Cziszarova M, Uhliarikova I, Agger JW, Li XL, Eijsink VG, West-

ereng B. 2013. Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. Biochim Biophys Acta 1830:5075–5086. https://doi.org/10.1016/j.bbagen.2013.07.018.
31. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340.
32. Jiao YL, Wang SJ, Lv MS, Xu JL, Fang YW, Liu S. 2011. A GH57 family amylopullulanase from deep-sea *Thermococcus siculi*: expression of the gene and characterization of the recombinant enzyme. Curr Microbiol 62:222–228. https://doi.org/10.1007/s00284-010-9690-6.
33. Inoue A, Anraku M, Nakagawa S, Ojima T. 2016. Discovery of a novel alginate lyase from *Nitratiruptor* sp. SB155-2 thriving at deep-sea hydrothermal vents and identification of the residues responsible for Its heat stability. J Biol Chem 291:15551–15563. https://doi.org/10.1074/jbc.M115.713230.
34. Paul S, Bag SK, Das S, Harvill ET, Dutta C. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Biol 9:R70. https://doi.org/10.1186/gb-2008-9-4-r70.
35. Linares-Pasten JA, Aronsson A, Karlsson EN. 2018. Structural considerations on the use of endo-xylanases for the production of prebiotic xylooligosaccharides from biomass. Curr Protein Peptide Sci 19:48–67.
36. Meng X, Shao Z, Hong Y, Lin L, Li C, Liu Z. 2009. A novel pH-stable, bifunctional xylanase isolated from a deep-sea microorganism, *Demequina* sp. JK4. J Microbiol Biotechnol 19:1077–1084.
37. Notenboom V, Birsan C, Warren RA, Withers SG, Rose DR. 1998. Exploring the cellulose/xylan specificity of the β-1,4-glycanase Cex from *Cellulomonas fimi* through crystallography and mutation. Biochemistry 37:4751–4758. https://doi.org/10.1021/bi9729211.
38. Xue X, Wang R, Tu T, Shi P, Ma R, Luo H, Yao B, Su X. 2015. The N-terminal GH10 domain of a multimodular protein from *Caldicellulosiruptor bescii* is a versatile xylanase/β-glucanase that can degrade crystalline cellulose. Appl Environ Microbiol 81:3823–3833. https://doi.org/10.1128/AEM.00432-15.
39. Zhang J, Siika-Aho M, Puranen T, Tang M, Tenkanen M, Viikari L. 2011. Thermostable recombinant xylanases from *Nonomuraea flexuosa* and *Thermoascus aurantiacus* show distinct properties in the hydrolysis of xylans and pretreated wheat straw. Biotechnol Biofuels 4:12. https://doi.org/10.1186/1754-6834-4-12.
40. Wang K, Luo H, Tian J, Turunen O, Huang H, Shi P, Hua H, Wang C, Wang S, Yao B. 2014. Thermostability improvement of a streptomyces xylanase by introducing proline and glutamic acid residues. Appl Environ Microbiol 80:2158–2165. https://doi.org/10.1128/AEM.03458-13.
41. Christakopoulos P, Kekos D, Macris BJ, Claeyssens M, Bhat MK. 1996. Purification and characterization of a major xylanase with cellulase and transferase activities from *Fusarium oxysporum*. Carbohydr Res 289:91–104. https://doi.org/10.1016/0008-6215(96)00146-2.
42. Alahuhta M, Xu Q, Bomble YJ, Brunecky R, Adney WS, Ding SY, Himmel ME, Lunin VV. 2010. The unique binding mode of cellulosomal CBM4 from *Clostridium thermocellum* cellobiohydrolase A. J Mol Biol 402:374–387. https://doi.org/10.1016/j.jmb.2010.07.028.
43. Henshaw JL, Bolam DN, Pires VM, Czjzek M, Henrissat B, Ferreira LM, Fontes CM, Gilbert HJ. 2004. The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem 279:21552–21559. https://doi.org/10.1074/jbc.M401620200.
44. Sermsathanaswadi J, Baramee S, Tachaapaikoon C, Pason P, Ratanakhanokchai K, Kosugi A. 2017. The family 22 carbohydrate-binding module of bifunctional xylanase/β-glucanase Xyn10E from *Paenibacillus curdlanolyticus* B-6 has an important role in lignocellulose degradation. Enzyme Microb Technol 96:75–84. https://doi.org/10.1016/j.enzmictec.2016.09.015.
45. Sainz-Polo MA, Gonzalez B, Menendez M, Pastor FI, Sanz-Aparicio J. 2015. Exploring multimodularity in plant cell wall deconstruction: structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem. J Biol Chem 290:17116–17130. https://doi.org/10.1074/jbc.M115.659300.
46. Matsushima N, Yoshida H, Kumaki Y, Kamiya M, Tanaka T, Izumi Y, Kretsinger RH. 2008. Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. Curr Protein Pept Sci 9:591–610.
47. Sjöde A, Frölander A, Lersch M, Rødsrud G. December 2013. Lignocellulosic biomass conversion by sulfite pretreatment. Global patent EP2376642 B1.
48. Rødsrud G, Lersch M, Sjöde A. 2012. History and future of world's most advanced biorefinery in operation. Biomass Bioener 46:46–59. https://doi.org/10.1016/j.biombioe.2012.03.028.
49. Pedersen RB, Thorseth IH, Hellevang B, Schultz A, Taylor P, Knudsen H, Steinsbu BO. 2005. Two vent fields discovered at the ultraslow spreading Arctic ridge system. Eos Trans AGU, Fall Meet Suppl 86.
50. Schander C, Rapp HT, Kongsrud JA, Bakken T, Berge J, Cochrane S, Oug E, Byrkjedal I, Todt C, Cedhagen T, Fosshagen A, Gebruk A, Larsen K, Levin L, Obst M, Pleijel F, Stöhr S, Warén A, Mikkelsen NT, Hadler-Jacobsen S, Keuning R, Petersen KH, Thorseth IH, Pedersen RB. 2010. The fauna of hydrothermal vents on the Mohn Ridge (North Atlantic). Marine Biol Res 6:155–171. https://doi.org/10.1080/17451000903147450.
51. Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
52. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 28:2223–2230. https://doi.org/10.1093/bioinformatics/bts429.
53. 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:W445–W451. https://doi.org/10.1093/nar/gks479.
54. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
55. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.
56. UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092.
57. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
58. Chen CC, Luo H, Han X, Lv P, Ko TP, Peng W, Huang CH, Wang K, Gao J, Zheng Y, Yang Y, Zhang J, Yao B, Guo RT. 2014. Structural perspectives of an engineered β-1,4-xylanase with enhanced thermostability. J Biotechnol 189:175–182. https://doi.org/10.1016/j.jbiotec.2014.08.030.
59. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
60. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.
61. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324. https://doi.org/10.1093/nar/gku316.
62. Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074.
63. Wood TM. 1988. Preparation of crystalline, amorphous, and dyed cellulase substrates. Methods Enzymol 160:19–25.
64. Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428. https://doi.org/10.1021/ac60147a030.
65. Vaaje-Kolstad G, Houston DR, Riemen AH, Eijsink VG, van Aalten DM. 2005. Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. J Biol Chem 280:11313–11319. https://doi.org/10.1074/jbc.M407175200.
66. Tomme P, Boraston A, Kormos JM, Warren RA, Kilburn DG. 2000. Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. Enzyme Microb Technol 27:453–458.
67. Cockburn D, Wilkens C, Dilokpimol A, Nakai H, Lewińska A, Abou Hachem M, Svensson B. 2016. Using carbohydrate interaction assays to reveal novel binding sites in carbohydrate active enzymes. PLoS One 11:e0160112. https://doi.org/10.1371/journal.pone.0160112.
68. Ladner CL, Wishart DS. 2012. Resolution-enhanced native acidic gel electrophoresis: a method for resolving, sizing, and quantifying prion protein oligomers. Anal Biochem 426:54–62. https://doi.org/10.1016/j.ab.2012.04.005.

Supplementary Material

A

ATGAACGCCATGCTCACACTGGCAGCCGCCCTGGCTGTGACGGCAATGTGCGCCGCCGAACCGCAAGTGCCG
GCAGGGGGACAGGTGCTGGTGGCCGGCGAGCCGTCGCCGGACTGGACCGGCCGCTACGAACAGGGCTACAG
CGAGGTGGCACAGGTGCAGGGCCCCGGCTTCTCGCGGGCCGTCAGGGTCACCACCGCCGAGCGCACGGCGC
CCTGGAACGTCCAGTTCCGCGTGCCCGTCAACGCCCCGGTCGCCAAAGGGGACGTGCTGCTTCTGGAGTTCTG
GGCCCGGTGCCTGCGCACCGAGGACGAGTCGGGACAGGGCTACATACGCGCCTGCGTGGAGAAACGCGACC
GACCCTGGACCAAGTCCCTCGAGCGCACCGTGACCGTGGGCTACAGGTGGCAACGCTTCCTGTTGCCCTGCAC
CTGCCGCGAGGACTACGGGCCGGGCGGACTCTACCTGGCCTTCGGCACCGGCCAGGTGCCGCAGGTCATCGA
AATCGGAGGCATCCGGCTCATCTCCTACGACAAACGGCTCTCGGCCGATGACCTGCCGCGCACGCGCGTGACC
TACGCCGGCCGCGAACCAGACGCCGCGTGGCGCAAAGAGGCATTCGAGCGAATCGAACGGATACGCACCGCC
CCCCTGGTGGTCCGGGTGGTGGACGAGGCCGGACGGGCCGTGGCCGGGGCGAAGGTCCAGGTAACGATGGTC
CAGCCCGCCTTCCAGTTCGGCTGCGCCGTGCCGGCATGGACCATCGCCGACCAGCAGGACCCCTCAAACGCCA
GGCGCCGCCAGGCCTTCCTCGAACTGTTCAACGCCGGCTCGTTCGTCAACGCGCTCAAATGGCCACCCTGGGT
GGGCGACTGGGGCAAGCGTTTCTCGCGGGAGGTGGTCTTCACGGCCCTCAACTGGTTCAAAGAGCACGACCT
GCACTTCCGCGGACACGTGCTCATCTGGCCCGGCTGGCGGCACCTGCCCCGCTTCATGCGGGCCTACCGCGGG
CACCCCGACCCCCAGGTCATCCGCCGCGAGGTGCTCAAGCACATCGACGATATCACCACCGCCACCCGCGGCTA
CGTGGAGGAGTGGGACGTCATAAACGAGCCCTACCACAACCACGACCTGATGGACATCTGCGGGCGGGACCT
GCTGGTCGAATGCTTCAAGCGGGCCCGCCACAACCTGCCCGACGCGCGCCTTGCCCTCAACGACTTCGGCATA
CTCACAGCCCTGACCGACGACGCCCACCAGGCCCACTACGAGCAGACCATCCAGTATCTGCTTGACAACGGCG
CGCCCCTGACCGTGCTGGGCATGCAGGGCCACTTCGGCGGCACCGTGCCGCCGCCTGCCCGGATGCTCAAAGT
CCTGGACCGCTACGCCCGGTTCGGCCTGCCCATCCGCATCACCGAGTTCACCGTGGGCACCGACGACGAGCAA
CTGCGGGCCGACTTCCTGCGTGACGCGCTCATCGTGGCCTACAGCCATCCGGCGGTCATCGGGTTCCAGTTCTG
GGGGCTCGGCACGCTGTTCGACCGCGAGATGCGTCCCCGCCCCGCCCTCGCGGCATACAAGGACCTGGTCCTG
GGGCGCTGGCGGACGCGTCTGGAAGGCCGCACCGACCCGATGGGCCGCTTCGCCGGCCGCGGCCACCTTGGC
CGCTACAACGTCAAGGCCACGCTGAACGGTCGGCTTGCCGAGGGCACGCTTGAACTGACCCGGGCCGGCCAG
GCAGCCGAACTGACCCTGCCCCTGCGC

B

MNAMLTLAAALAVTAMCAAEPQVPAGGQVLVAGEPSPDWTGRYEQGYSEVAQVQGPGFSR
AVRVTTAERTAPWNVQFRVPVNAPVAKGDVLLLEFWARCLRTEDESGQGYIRACVEKRDR
PWTKSLERTVTVGYRWQRFLLPCTCREDYGPGGLYLAFGTGQVPQVIEIGGIRLISYDKR
LSADDLPRTRVTYAGREPDAAWRKEAFERIERIRTAPLVVRVVDEAGRAVAGAKVQVTMV
QPAFQFGCAVPAWTIADQQDPSNARRRQAFLELFNAGSFVNALKWPPWVGDWGKRFSREV
VFTALNWFKEHDLHFRGHVLIWPGWRHLPRFMRAYRGHPDPQVIRREVLKHIDDITTATR
GYVEEWDVINEPYHNHDLMDICGRDLLVECFKRARHNLPDARLALNDFGILTALTDDAHQ
AHYEQTIQYLLDNGAPLTVLGMQGHFGGTVPPPARMLKVLDRYARFGLPIRITEFTVGTD
DEQLRADFLRDALIVAYSHPAVIGFQFWGLGTLFDREMRPRPALAAYKDLVLGRWRTRLE
GRTDPMGRFAGRGHLGRYNVKATLNGRLAEGTLELTRAGQAAAELTLPLR

Figure S1. AMOR_GH10A DNA and amino acid sequences. Panel A shows the native DNA sequence of the *AMOR_GH10A* gene. Panel B shows the full-length AMOR_GH10A amino acid sequence, including the predicted signal peptide (green; 1-18), the CBM85 domain (blue, 19-180), and the GH10 domain (yellow; 239-529).

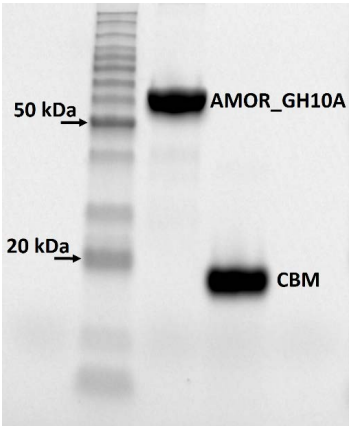

Figure S2. SDS-PAGE analysis of purified enzymes. Next to a lane with size marker proteins the gel shows lanes with 5 μg of the purified full-length enzyme (AMOR_GH10A; predicted size 65.7 kDa) and the purified carbohydrate binding module (CBM; predicted size 19.2 kDa).

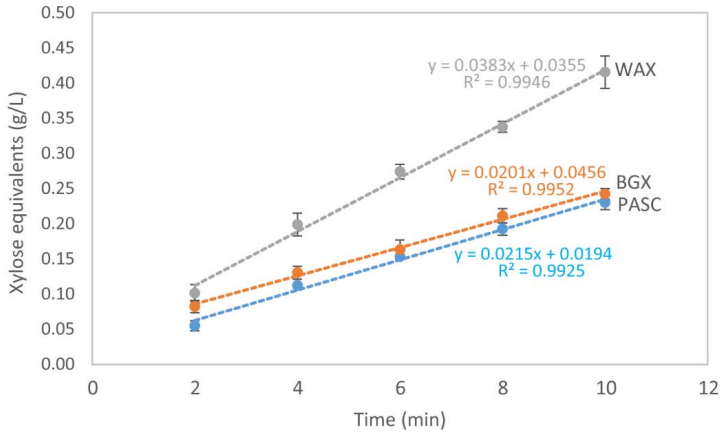

Figure S3. Determination of initial rates. The graphs show progress curves for the reaction of AMOR_GH10A (1 µM) with g/L of the various indicated substrates at pH 5.6 and 80 °C. These graphs yield estimated initial rates of 4.3 s^{-1} for WAX, wheat arabinoxylan, 2.2 s^{-1}, for BGX, birchwood glucuronoxylan and 2.4 s^{-1} for PASC, phosphoric acid-swollen cellulose. Quantification of the products was done according to the DNS method and with xylose as standard.

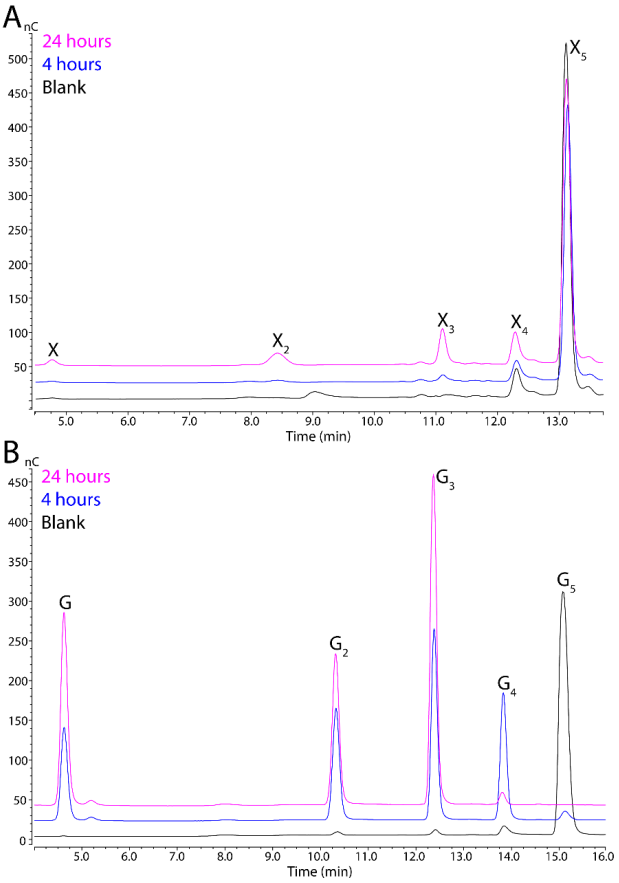

Figure S4. Product formation from xylopentaose and cellopentaose. The panels show chromatographic analysis of reaction products arising upon incubation of 1 µM AMOR_GH10A with 1 mM of xylopentaose (X_5, A) or cellopentaose (G_5, B) for 4 h (blue chromatograms) or 24 h (pink chromatograms), at 60°C, pH 5.6 (25 mM NaOAc, 0.5 M NaCl). Products originating from X_5 and G_5 are annotated as X_1-X_4 and G_1-G_4, respectively. The black chromatograms show background signals obtained by analyzing reaction mixtures with substrate that had been incubated for 24 h without enzyme.

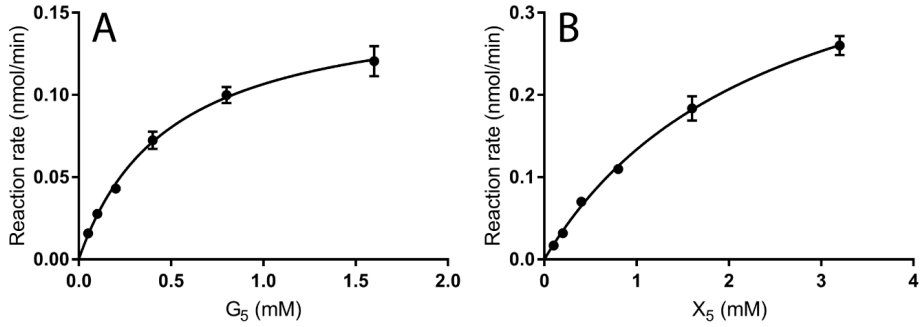

Figure S5. Michaelis-Menten plot for AMOR_GH10A acting on cellopentaose (G_5) and xylopentaose (X_5). The curves show the nonlinear least-squares fit of the experimental data to the Michaelis–Menten equation for the reaction with cellopentaose (G_5; panel A; R^2=0.9873) and xylopentaose (X_5; panel B; R^2=0.9928). The enzyme concentrations were 0.5 µM and 50 µM, respectively. Fitting and calculation of K_m and k_{cat} values were done using GraphPad.

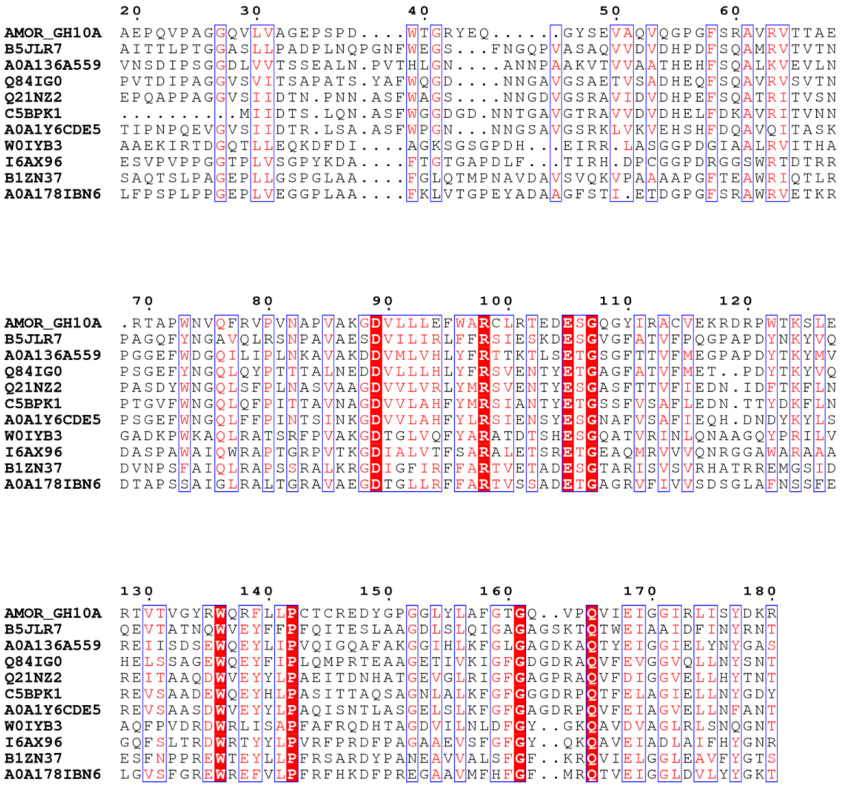

Figure S6. Alignment of hypothetical CBM domains in phylogenetically related GH10 enzymes. 10 of the 12 phylogenetically related GH10 enzymes shown in Fig. 1B have domains that resemble the CBM85 of AMOR_GH10A, with sequence identities ranging from 28%-34%. These sequences were aligned to the CBM85 using ClustalOmega, and the resulting sequence alignment was visualised using ESPRipt 3.0. With the exception of AMOR_GH10A, proteins are identified by their Uniprot id. The sequence numbering for AMOR_GH10A is shown above the alignment. Conserved residues are printed in white on a red background, and conservatively substituted residues are printed in red.

Paper III

A thermostable bacterial AA10 LPMO with high operational stability in a wide temperature range

Marianne Slang Jensen[1], Zarah Forsberg[1], Tina Rise Tuveng[1], Lasse Fredriksen[1], Gustav Vaaje-Kolstad[1], Vincent G. H. Eijsink[1]

[1]Faculty of Chemistry, Biotechnology and Food Science, NMBU - Norwegian University of Life Sciences, Ås, Norway.

For correspondence; E-mail: vincent.eijsink@nmbu.no

Manuscript in preparation

Abstract

LPMOs are oxidative, copper-dependent enzymes that function as powerful tools in the turnover of various biomasses, including cellulosic plant biomass. We describe a thermostable, cellulose-active LPMO (mgLPMO10A) from a high-temperature compost metagenome with an apparent melting temperature of 83 °C and with high activity up to at least 80 °C. Interestingly, mgLPMO10A showed good activity and operational stability over a wide temperature range, which is of both fundamental and applied interest. The LPMO had a clear boosting effect on the efficiency of a GH48 and a GH6 cellobiohydrolase, and on the efficiency of a commercial cellulase cocktail. ICP-MS analyses showed that the enzyme obtained after the final purification step was largely devoid of copper. Accordingly, this enzyme preparation was much less active than the same preparation after being subjected to a copper saturation procedure. Strikingly, the untreated, less active enzyme preparation displayed slow but steady kinetics and worked well in synergy experiments with cellulases.

Introduction

Lignocellulosic biomass (i.e., plant-based biomass) is the most abundant source of renewable carbon on Earth (Pauly and Keegstra, 2008) and its major component is cellulose, a linear polymer of glucose units joined by β-1,4-glycosidic bonds (Klemm et al., 2005). Individual cellulose chains commonly arrange into crystalline fibrils consisting of multiple chains that are stabilized by extensive hydrogen bonding, and this assembly makes cellulose highly resistant to degradation (Duchesne and Larson, 1989, Somerville et al., 2004). Microorganisms have evolved to overcome this recalcitrance by producing specialized enzymes, and these may be employed in industrial bioprocessing for the sake of generating renewable energy and commodities (Merino and Cherry, 2007, Harris et al., 2014).

Cellulose-degrading enzymes include cellulases that cleave glycosidic bonds using a hydrolytic mechanism, and that are further classified as endoglucanases or cellobiohydrolases. Endoglucanases attack randomly in amorphous/loosely packed cellulose by engulfing the cellulose chain in a catalytic cleft, while cellobiohydrolases attack from accessible chain ends using a tunnel-shaped active site (Davies and Henrissat, 1995). The turnover of cellulose by cellulases is hampered by the fact that the enzymes cannot easily access the crystalline surfaces of the cellulose fibrils (Horn et al., 2012).

Enzymes called lytic polysaccharide monooxygenases (LPMOs) employ a powerful oxidative mechanism to cleave glycosidic bonds within crystalline regions of densely packed polysaccharides such as cellulose (Vaaje-Kolstad et al., 2010). This mode of action is enabled by a flat substrate-binding surface with a surface-exposed active site. The ability of LPMOs to disrupt crystalline cellulose fibrils, thereby granting the hydrolytic enzymes access to binding sites in parts of the substrate that they would otherwise struggle to reach (Vaaje-Kolstad et al., 2010, Eibinger et al., 2014, Vermaas et al., 2015, Eibinger et al., 2017), makes LPMO activity crucial in the development of industrial bioprocessing technologies (Johansen, 2016, Müller et al., 2018).

The catalytic function of LPMOs was discovered in 2010 (Vaaje-Kolstad et al., 2010) and they have since been classified as auxiliary activities (AA) in the carbohydrate-active enzymes database (cazy.org; Levasseur et al., 2013). LPMOs are categorized in CAZy families AA9-11 and AA13-16 on the basis of sequence similarity, and may be active on cellulose, hemicellulose, chitin, starch and/or oligosaccharides (Tandrup et al., 2018, Filiatrault-Chastel et al., 2019). Thus, LPMOs hold an important role in the Earth's carbon cycle, but might also have alternative functions, including a role as virulence factors in microbial pathogenicity (Loose et al., 2014, Paspaliari et al., 2015, Agostoni et al., 2017).

LPMOs hydroxylate either the C1 or C4 carbon of the scissile glycosidic bond in cellulose (Vaaje-Kolstad et al., 2010, Phillips et al., 2011, Quinlan et al., 2011), whereas some are less regioselective and produce a mixture of C1- and C4-oxidized products. Oxidation of the C1 carbon results in the formation of 1,5-δ-lactones that are spontaneously converted to aldonic acids, while oxidation of the C4 carbon produces 4-ketoaldoses that are hydrated to their corresponding gemdiol form (Isaksen et al., 2014). Like cellulases, LPMOs may be associated with a carbohydrate-binding module (CBM), often through a flexible linker region (Courtade et al., 2018).

Two highly conserved histidines located on the flat substrate-binding surface constitute the catalytic centre of LPMOs. The histidines coordinate a single copper ion, which must be reduced from Cu(II) to Cu(I) state before the LPMO can initiate oxidative cleavage (Quinlan et al., 2011). To ensure that all LPMO molecules coordinate a copper ion, it is common to include a copper saturation step prior to running LPMO reactions. Notably, reduced LPMOs are known to be prone to oxidative damage when exposed to O_2 or H_2O_2 in the absence of a proper substrate (Bissaro et al., 2017, Loose et al., 2018, Eijsink et al., 2019).

LPMO catalysis has generally been believed to require molecular oxygen and a reductant that delivers two electrons for each catalytic cycle (Vaaje-Kolstad et al., 2010, Beeson et al., 2012, Kjaergaard et al., 2014). However, a recent study has indicated that H_2O_2 may be the true co-substrate of LPMOs (Bissaro et al., 2017), a scenario that would only require reductant for a priming reduction of the LPMO, after which the enzyme can catalyze multiple reactions. Notably, under the conditions normally used in LPMO reactions, H_2O_2 will be formed either by O_2 reacting with reduced LPMO molecules that are not bound to the substrate (Kittl et al., 2012, Isaksen et al., 2014) or by reactions involving O_2 and the reductant in the presence of trace amounts of transition metals.

Although LPMOs have been intensely studied since their discovery in 2010 (Tandrup et al., 2018, Chylenski et al., 2019, Eijsink et al., 2019, Forsberg et al., 2019), and although they are used in industry today (Johansen, 2016, Müller et al., 2018), little is known about industrially relevant properties such as stability. Thermostability is a commonly desired trait for enzymes employed in biorefineries because stable enzymes last longer and because it is often considered favourable to run processes at higher temperatures, for example to avoid microbial contamination issues in the industrial facilities (Haki and Rakshit, 2003). Thermostable enzymes can be discovered by bioprospecting of high temperature environments where biomass turnover occurs in Nature.

The present study describes the characterization of a thermostable bacterial AA10 LPMO with a C-terminal CBM2 domain. The gene encoding the enzyme has previously been identified as overexpressed in a metatranscriptome originating from rice straw that was inoculated with compost and incubated at high temperature (Simmons et al., 2014)). We have compared the properties of this LPMO, termed mgLPMO10A (mg, for metagenome), with the properties of a similar two-domain mesophilic LPMO from *Streptomyces coelicolor*, *Sc*LPMO10C (Forsberg et al., 2014a), and we have assessed its potential to boost the activity of various cellulases. The effect of copper saturation on the stability and catalytic ability of this novel LPMO was also explored.

Materials and methods

LPMO expression and purification

The mgLPMO10A-encoding gene originates from publicly available metagenome data accessible in the Joint Genome Institute IMG/M database (https://img.jgi.doe.gov/cgi-bin/m/main.cgi; IMG genome ID 2199352008). The microbial community was sampled from rice straw that had been inoculated with compost and incubated at high temperature (55 °C), and has been studied previously based on metagenomic and metatranscriptomic analyses

(Reddy et al., 2013, Simmons et al., 2014). The 1092 bp gene encoding mgLPMO10A (IMG/M gene ID 2200500718) was detected as overexpressed in the community ((Simmons et al., 2014); note that the LPMO is referred to as a CBM33 in this paper).

The mgLPMO10A encoding gene (codon-optimized for *Escherichia coli* expression) was synthesized (GenScript) and bp 94-1089 (omitting the predicted 31 amino acid signal peptide sequence and the stop codon) were cloned in the pRSET B expression vector (Invitrogen, Carlsbad, California, USA), already containing the well-functioning signal sequence of *Sm*LPMO10A (also known as CBP21), following the protocol described by (Forsberg et al., 2014b). The forward primer employed for insertion in the pRESET B vector was 5'-CGCAACAGGCGAATGCCCACGGTGCGGCGATGGT-3' and the reverse primer was 5'-CAGCCGGATCAAGCTTTTACGCGCTGGTGCAGGTC-3'. For production of the catalytic domain alone (mgLPMO10Acd; bp 94-669) the reverse primer 5'-CAGCCGGATCAAGCTTTTAATCAAAAACCACGTCGCT-3' was used instead. The underlined nucleotides represent an over-hang sequence with the pRSET B vector (for fusing the gene in the pRSET B vector using the In-Fusion HD cloning kit (Clontech, Mountain View, California, USA)), and the "TTA" sequence of the reverse primers (in bold) was implemented to introduce a stop codon at the end of the LPMO-encoding sequence. The gene sequences were verified by Sanger sequencing (Eurofins GATC, Cologne, Germany).

Chemically competent OneShot BL21 Star™ (DE3) *E. coli* cells (Invitrogen) harbouring the pRSET B vector with the LPMO encoding genes were used to inoculate 500 mL Terrific Broth (TB) supplemented with 100 µg/mL ampicillin, followed by incubation at 37 °C for 18 hours, 30 °C for 24 hours or 22 °C for 48 hours in a Harbinger system (Harbinger Biotechnology & Engineering, Markham, Canada). Note that expression was driven by leakiness of the T7 promoter and that no inducer molecule was added. Cells were harvested by centrifugation at 5000 × g for 15 minutes at 4 °C (Beckman Coulter Brea, California, USA), after which a periplasmic extract was prepared using an osmotic shock method (Manoil and Beckwith, 1986).

Periplasmic extracts were filtered using 0.45 µm syringe filters (Sarstedt, Nümbrecht, Germany) and adjusted to 50 mM Tris-HCl (pH 8.0) after which the proteins were purified by anion exchange chromatography using an ÄKTA pure chromatography system (GE Healthcare, Chicago, USA) equipped with a 5-ml HiTrap DEAE FF column (GE Healthcare). 50 mM Tris-HCl (pH 8.0) was used as running buffer. mgLPMO10A and mgLPMO10Acd displayed poor binding to the DEAE column, and thus mainly ended up in the flow through. Protein fractions were examined by SDS-PAGE (Bio-Rad, Hercules, California, USA) and relevant fractions (i.e. the flow through in this case) were concentrated to 1 mL using 10,000 MWCO (Molecular Weight Cut-Off) Vivaspin ultrafiltration tubes (Sartorius, Göttingen, Germany).

The concentrated protein samples were purified in a second step by size exclusion chromatography using a HiLoad 16/60 Superdex 75 column with a running buffer consisting of 50 mM Tris-HCl (pH 8.0) and 200 mM NaCl. Fractions containing the LPMO, as determined by SDS-PAGE, were concentrated using 10,000 MWCO Vivaspin ultrafiltration tubes (Sartorius) with simultaneous buffer exchange to 20 mM sodium phosphate buffer (pH 6.0).

Prior to use, purified enzymes were incubated for 30 minutes at room temperature with $Cu(II)SO_4$ in a 1:3 molar ratio (LPMO:Cu^{2+}) or EDTA in a 1:10 molar ratio (LPMO:EDTA) to

generate copper-saturated LPMO or apo-enzyme, respectively. PD Midi-Trap G-25 columns (GE Healthcare) were subsequently used to desalt the samples and remove excess Cu^{2+}/EDTA. The columns had been equilibrated with 15 mL 5 mM EDTA in Chelex (Bio-Rad) treated 20 mM sodium phosphate buffer (pH 6.0), followed by 15 mL Chelex (Bio-Rad) treated 20 mM sodium phosphate buffer (pH 6.0) to avoid unintended metals in the enzyme samples. Protein concentrations were determined by measuring A_{280} (absorbance at 280 nm) in a spectrophotometer (Eppendorf Biophotometer, Eppendorf, Hamburg, Germany) and absorbances were converted to protein concentrations using theoretical extinction coefficients calculated with the ExPASy ProtParam tool (web.expasy.org/protparam). The proteins were stored at 4 °C.

Production of additional enzymes

Full-length *Sc*LPMO10C (also in the pRSET B vector) was produced using the same two chromatographic steps as for mgLPMO10A, and as described previously by (Forsberg et al 2014). For ion exchange purification, the periplasmic extract was adjusted to 50 mM Tris-HCl (pH 7.5) which was also used as running buffer, and the protein was eluted by applying a linear salt gradient (0-500 mM NaCl) over 60 column volumes (CV). Fractions containing the LPMO were pooled and subjected to size exclusion chromatography, as described above.

Two putative cellobiohydrolases (CBHs), namely a 101.6 kDa GH48 (mgCel48A; IMG/M gene ID 2200387045; overexpressed together with mgLPMO10A in the thermophilic compost/rice straw community (Simmons et al., 2014) and a 61.4 kDa GH6 (mgCel6B; IMG/M gene ID 2200705178) from the same metagenome (Reddy et al., 2013), were produced and included in synergy experiments with mgLPMO10A. Like crLPMO, both cellulases contain a CBM2. The cellulases were produced with a C-terminal His-tag according to the procedure described for a GH6 endoglucanase (mgCel6A) from the same metagenome (Jensen et al., 2018). To enable cloning in the pNIC-CH expression vector, the forward primer for mgCel48A was 5'TTAAGAAGGAGATATACTATGGCACCGGCATGTGAAGTTACCTAT3' and the reverse primer was 5'AATGGTGGTGATGATGGTGCGCACCAATCAGACGATCATAATCGGC3'. For mgCel6B, the forward primer was 5'TTAAGAAGGAGATATACTATGGCATTTGCAGCACCGGGTTGTAGC3' and the reverse primer was 5'AATGGTGGTGATGATGGTGCGCCAGAGGCGGATATGCATTATCCAT3'. The underlined nucleotides represent an over-hang sequence for ligase-independent cloning in the pNIC expression vector.

Commercial cellulase cocktail

Celluclast® 1.5L, a commercial cellulose cocktail with little LPMO activity (Müller et al., 2015) was kindly provided by Novozymes (Bagsværd, Denmark). The Celluclast cocktail was first mixed with a β-glucosidase kindly supplied by Novozymes in a 10:1 ratio (on a protein basis) to facilitate degradation of solubilized sugars to the monomer (glucose).

Substrates

Evaluation of LPMO oxidative activity and saccharification experiments were carried out with the microcrystalline cellulosic substrate Avicel PH-101 (Sigma-Aldrich). PASC (phosphoric acid-swollen cellulose) was prepared from Avicel PH-101 as described by (Wood, 1988).

Apparent melting temperature

A Nano-Differential Scanning Calorimeter III (Calorimetry Sciences Corporation, Lindon, USA) was used to determine the apparent melting temperatures [$T_{m(app)}$] of mgLPMO10A variants. Approximately 0.4 mg/ml LPMO in 50 mM sodium phosphate buffer pH 6.0 (filtered and degassed) was heated from 25 °C to 100 °C at 1 °C/min. Buffer baselines were recorded and subtracted from the protein scans. The data were analyzed with NanoAnalyze software (www.tainstruments.com).

Activity assays

Unless stated otherwise, reactions were performed in 50 mM sodium phosphate buffer (pH 6.0) at 800 rpm in an Eppendorf thermomixer. All reactions were performed in triplicates.

For the initial verification of oxidative cellulolytic activity, mgLPMO10A was incubated at 1 µM with 5 g/L PASC in the presence of 1 mM ascorbic acid (reductant) for 24 hours, at 60 °C. The supernatant was analyzed by MALDI-ToF (see below) for detection of oxidized cello-oligomers and determination of LPMO regioselectivity.

Progress curves for the degradation of 10 g/L Avicel by 1 µM LPMO in the presence of 1 mM ascorbic acid were determined at various temperatures ranging from 40-80 °C. Aliquots were withdrawn at selected time points and filtered (0.45 µm) before being merged with an identical volume of 2 µM endoglucanase (mgCel6A; (Jensen et al., 2018)) followed by overnight incubation at 37 °C. As a result of this procedure, all oxidized products were converted to oxidized cellobiose or oxidized cellotriose, which simplified product quantification. The products were analyzed using HPAEC-PAD (see below).

Synergy experiments

Investigation of the synergistic relationship between mgLPMO10A and two putative cellobiohydrolases (mgCel48A and mgCel6B) was done using 10 g/L Avicel as substrate, which was subjected to degradation by 1 µM mgLPMO10A and/or 1 µM of mgCel48A or mgCel6B. The reactions contained 50 mM sodium phosphate buffer (pH 6.0), and 1 mM ascorbic acid. Reactions involving mgCel48A were supplemented with 1 mM $CaCl_2$ as this is beneficial to the enzyme (data not shown). Samples were withdrawn after 2, 24 and 48 hours incubation at 60 °C, and mixed with an equal volume of 0.1 M NaOH to stop the enzyme activity. Products were analyzed using HPAEC-PAD (see below).

The effect of mgLPMO10A on the ability of a mixture of Celluclast® 1.5L and a β-glucosidases (see above) to degrade cellulose was investigated using 50 g/L Avicel as substrate. The total protein load was 4 mg enzyme per gram glucan (the glucan content of Avicel is 92.2 %; Müller et al., 2015). The reactions contained 50 mM sodium acetate, pH 5.0, and 1 mM ascorbic acid, and were incubated at 50 °C (optimal conditions for Celluclast). Samples were withdrawn after 4, 24 and 96 hours, the supernatants were separated from the insoluble substrate by filtration over a 0.45 µM filter using microtiter filter plates and kept at -20 °C until product analysis. Products were analyzed using UHPLC (see below).

HPAEC-PAD product analysis

Oxidized products, and native products generated in the synergy experiments with the CBHs, were detected by high-performance anion-exchange chromatography (HPAEC) using a

Dionex™ ICS5000 system (Thermo Scientific, Sunnyvale, CA) equipped with a disposable electrochemical gold electrode and a CarboPac PA1 column (Dionex) operated with 0.1 M NaOH (eluent A) at a column temperature of 30 °C (Westereng et al., 2013). A multistep linear gradient with increasing amounts of eluent B (0.1 M NaOH + 1 M NaOAc) was used to elute the products, going from 0–10 % B over 10 min; 10–14 % B over 5 min; 14–30 % B over 1 min; 30–100 % B over 2 min; 100–0 % B over 0.1 min; and 0 % B (reconditioning) for 10.9 min. Data collection and analysis were carried out with the Chromeleon 7.0 software. Cellobiose (Sigma Aldrich) and cellotriose (Megazyme) were used as substrates for production of C1-oxidized standards of cellobionic acid (GlcGlc1A) and cellotrionic acid (Glc_2Glc1A), respectively, by incubation with cellobiose dehydrogenase from *Myriococcum thermophilum* (MtCDH) (Zamocky et al., 2008).

UHPLC product analysis

Quantification of glucose and cellobiose resulting from reactions with the Celluclast enzyme cocktail was achieved using a Dionex Ultimate 3000 UHPLC system (DionexCorp., Sunnyvale, CA, USA) equipped with a Rezex ROA-Organic Acid H+ (8 %), 300 × 7.8 mm analytical column equipped and a SecureGuard Carbo-H+ 4 × 3.0 mm guard column (Phenomenex, Torrance, CA, USA) operated at 65 °C. Sample components were eluted isocratically using 5 mM sulfuric acid as mobile phase with a flow rate of 0.6 mL/min, and were detected using a refractive index (RI) detector 101 (Shodex, Tokyo, Japan). Data collection and analysis were carried out with the Chromeleon 7.0 software.

MALDI-ToF product analysis

Reaction supernatants were assayed qualitatively using a matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) UltrafleXtreme mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) equipped with a Nitrogen 337-nm laser. Reaction products (1 µl) were applied to an MTP 384 ground steel target plate TF (Bruker Daltonics) and merged with 2 µl of 9 mg/ml of 2,5-dihydroxybenzoic acid (DHB) dissolved in 30% acetonitrile, followed by air-drying. Data collection and analysis were carried out using the Bruker FlexAnalysis software.

Metal content analysis by inductively coupled plasma mass spectrometry (ICP-MS)

Protein samples (untreated, EDTA-treated or copper-saturated mgLPMO10A) were analysed by inductively coupled plasma mass spectrometry (ICP-MS) to determine the amount of copper in the enzyme preparations. Protein and buffer samples (0.5 mL) were mixed with one droplet of concentrated HNO_3 and diluted to 1.5 mL before analysis on an Agilent 8800 ICP-MS (Agilent technologies, Santa Clara, California, USA) with NH_3 as reaction gas.

To estimate the fraction of metals bound to the protein relative to free metals in solution, the protein samples were concentrated on 3K centrifugal filters (VWR) that were first rinsed with Chelex-treated 20 mM sodium phosphate buffer (pH 6.0). The flow through of the protein samples were analysed by ICP-MS to estimate the concentration of free metals. The concentrated protein samples were analysed to estimate the total amount of metals in the samples (protein concentrations were calculated by A_{280} prior to the ICP-MS analysis). The amount of protein-bound metals was estimated by subtracting the concentration of free metals in the flow though from the total metal concentration in the protein samples.

Results and discussion

LPMO production and verification of activity

The full-length LPMO (mgLPMO10A; residues 32-363) and the catalytic domain (mgLPMO10Acd; residues 32-223) were expressed and purified to electrophoretic homogeneity (Fig. S1). The same was done for the well-studied *Sc*LPMO10C, which was included in the experiments to enable comparison of mgLPMO10A to a mesophilic homologue. These two LPMOs have a very similar domain structure, including a proline- and threonine-rich linker that separates the LPMO domain and the CBM2.

The oxidative activity of mgLPMO10A on cellulosic substrates was initially verified using PASC under standard conditions (1 mM ascorbic acid, aerobic conditions), and this showed that the LPMO oxidizes the C1-carbon adjacent to the scissile bond (Figure 1).

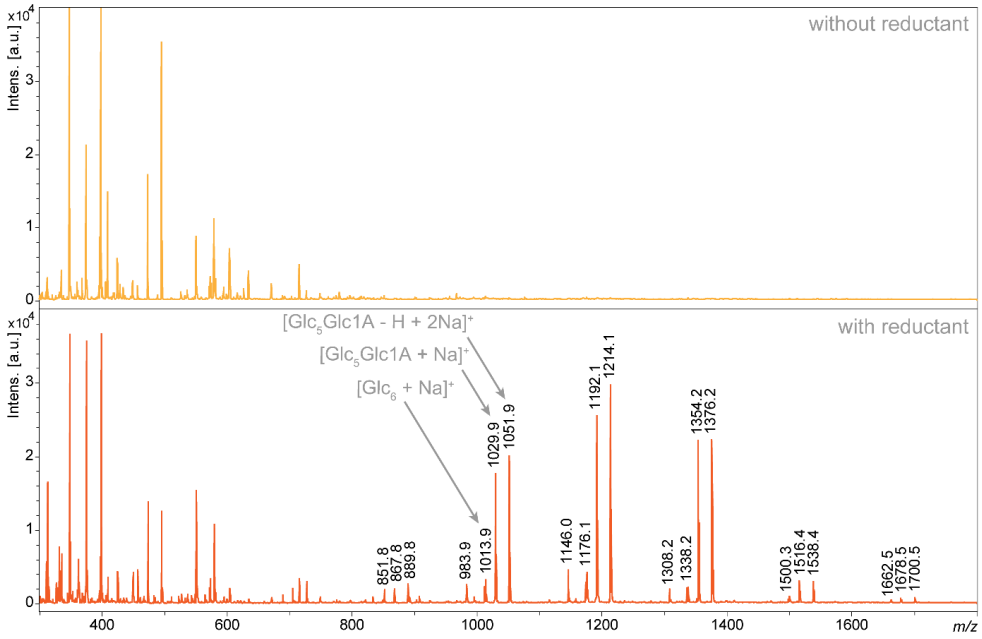

Figure 1. Initial verification of LPMO activity. Full-length mgLPMO10A was incubated with 0.5 % PASC in 50 mM sodium phosphate buffer (pH 6.0) supplemented with 1 mM ascorbic acid (AscA) at 60 °C overnight. The supernatant was subjected to MALDI-ToF analysis. The product profile shows peaks characteristic for oxidized products, including a large peak representing the sodium salt of the sodium adduct of the aldonic acid, which is diagnostic for C1-oxidation. The peaks of the hexamer cluster are denoted by grey arrows and show the sodium adduct of native cellohexaose $[Glc_6 + Na]^+$ and two larger peaks that represent sodium adducts of C1-oxidized cellohexaose (aldonic acid), namely $[Glc_5Glc1A + Na]^+$ and $[Glc_5Glc1A - H + 2Na]^+$. Note that the lactone is also visible (at 1011; not labeled). Of note, the spectrum shows a series of minor signals (984, 1146, 1308) differing by one glucose (162 *m/z*) that represent products of unknown nature. None of the labelled peaks were observed for a reaction with only PASC and AscA, whereas the upper panel shows that only a few, and very minor signals were observed for a reaction with PASC and the LPMO but lacking AscA.

Thermal stability

Differential scanning calorimetry (DSC; Figure 2) showed that the [$T_{m(app)}$] of copper saturated mgLPMO10A, is approximately 83 °C, and that both truncation ([$T_{m(app)}$] 76 °C) and removal of the copper ([$T_{m(app)}$] 73 °C) have negative effects on the thermal stability. All unfolding events were irreversible. These results suggests that both copper binding and the CBM have a protective effect that confers improved stability to the LPMO catalytic domain. The positive effect of copper binding on LPMO stability is consistent with previous studies (Hemsworth et al., 2013) and can be explained by the stabilizing effect of copper binding on the histidine brace, which includes the N-terminus that is known to be relatively flexible in the apo-enzyme (Aachmann et al., 2012). To our knowledge, the [$T_{m(app)}$] of 83 °C observed for mgLPMO10A-Cu^{2+} is the highest melting temperature ever recorded for an LPMO.

The effect of the CBM on the apparent melting temperature of mgLPMO10A is not easily explained. Studies of *Sc*LPMO10C have shown that the linker between the CBM2 and the LPMO domain is flexible and the solution structure of full-length *Sc*LPMO10C showed this linker to be quite extended (Courtade et al., 2018). Thus, one might expect the catalytic domain and CBM to act as separated units that fold and unfold independently. The unfolding curve of mgLPMO10A in Fig. 2 shows only a single transition, which indicates that the two domains do not unfold independently or that they unfold independently and have approximately the same melting temperature. Based on the present data it cannot be excluded that the two domains interact with each other in a manner that has a protective effect. However, the decrease in stability may also be due to the actual truncation, which may have created an unfavourable configuration at the C-terminus of the truncated protein.

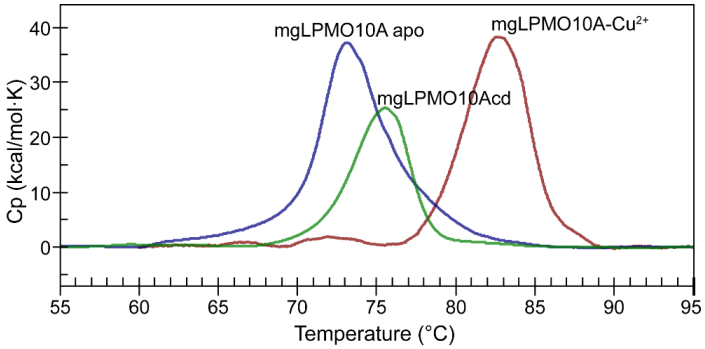

Figure 2. Thermal unfolding of mgLPMO10A. The apparent melting temperatures [$T_{m(app)}$] were determined by differential scanning calorimetry (DSC) by heating the copper saturated enzyme (approximately 0.4 g/L) at 1 °C/min from 25 °C to 100 °C in 50 mM sodium phosphate buffer (pH 6.0). The experiment was repeated for the apo-form of mgLPMO10A (mgLPMO10A apo, in Chelex-treated buffer), and the copper saturated but truncated version of mgLPMO10A (mgLPMO10Acd).

The effect of temperature on activity

Monitoring of product formation over time at varying incubation temperatures (Fig. 3) showed a trade-off between activity and enzyme inactivation and revealed a clear difference between mgLPMO10A and *Sc*LPMO10C. For mgLPMO10A, the initial activity increased with temperature all the way up to 80 °C (the highest tested temperature). Only at 80 °C the product formation curve showed clear signs of enzyme inactivation, with product levels stabilizing already after approximately 20 minutes with 350 μM product. In the other reactions, product levels reached almost 500 μM, which , apparently, is the maximum level that can be reached under these conditions.

In contrast, *Sc*LPMO10C showed clear signs of enzyme inactivation in the reaction carried out at 70 °C (Fig. 3). Considering its apparent T_m of 64 °C (Jensen et al., 2019 submitted for publication), it may seem surprising that the enzyme works at all at 70 °C, but this can be explained by the substrate having a protective effect (the melting temperature was determined in the absence of substrate). It is remarkable that the maximum product levels reached by *Sc*LPMO10C are lower (approx. 200 μM) compared to mgLPMO10A (approx. 500 μM). More experiments would be needed to explain this difference and to verify why product formation stops in the first place. One possible explanation for the difference between the two LPMOs is that *Sc*LPMO10C is more prone to autocatalytic oxidative inactivation than mgLPMO10A, which would, obviously, be very interesting.

As to the effect of temperature on LPMO activity and stability, it is very important to note that we do not know the effects of elevated temperature on the reductant, the co-substrate, and the H_2O_2-generating oxidase activity of the LPMO. Product yields and apparent enzyme inactivation do not necessarily reflect only thermal stability of the LPMOs or the general effect of temperature on enzyme catalysis. For example, increased LPMO activity at higher temperatures could reflect that H_2O_2 is generated at a faster rate as the temperature increases, and may not necessarily reflect the catalytic rate of the LPMOs. Likewise, temperature effects on LPMO reduction and substrate binding, where the latter would affect both the oxidase activity of the LPMO and the sensitivity to autocatalytic inactivation, may partly underlie the temperature effects shown in Fig. 3. Of note, the concentrations of dissolved oxygen become less as temperature increases, but this does not seem to negatively affect initial LPMO activity.

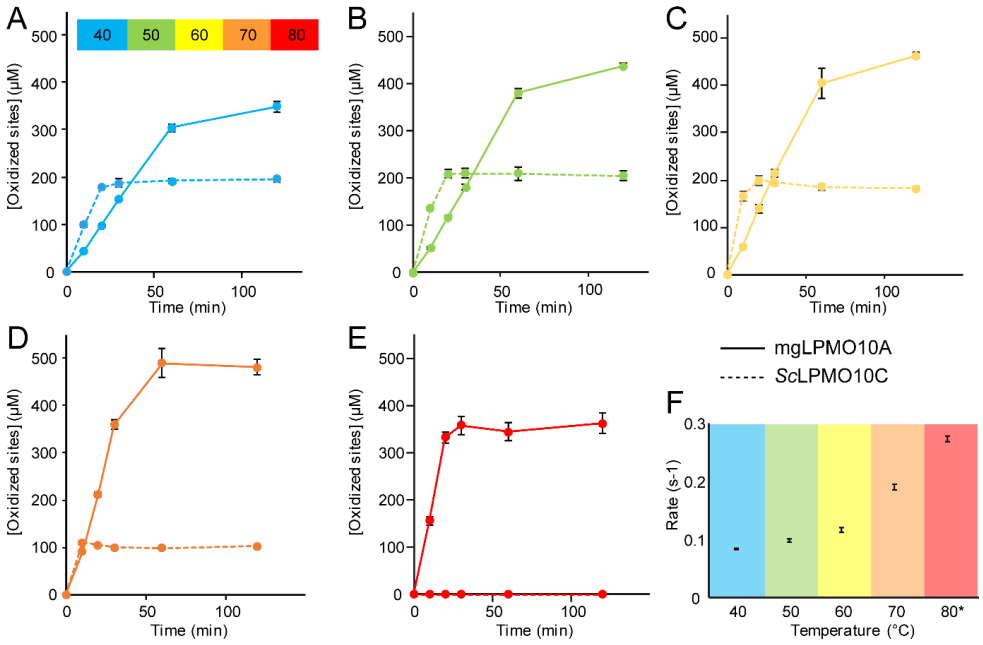

Figure 3. Activity of mgLPMO10A and *Sc*LPMO10C at various temperatures. Panels A-E show progress curves for copper saturated mgLPMO10A (solid curves) and *Sc*LPMO10C (dashed curves) at 40-80 °C. The reactions were carried out with 1 µM LPMO and 10 g/L Avicel in 50 mM sodium phosphate buffer pH 6.0. The reactions were pre-incubated for 10 min at various temperatures before 1 mM ascorbic acid was added to start the LPMO reaction. The reactions were carried out in Eppendorf Thermomixers set to 40-80 °C and 800 rpm and samples were taken and filtered at 10, 20, 30, 60 and 120 min. The solubilized products were degraded to oxidized dimers and trimers by incubation with an endoglucanase prior to product quantification. Panel F shows the initial rate of mgLPMO10A as a function of temperature, ranging from 0.08-0.27 s^{-1} and with the highest rate measured at 80 °C. The rates were calculated from the linear range of the progress curves, that is 0-30 min, with the exception of the 80 °C curve, of which only the 0-20 min points were used. The error bars show ± S.D. (n=3).

Synergy experiments

We then studied synergy between the mgLPMO10A and two putative CBHs derived from the same metagenome, namely a GH6 (mgCel6B) and a GH48 (mgCel48A), where the latter was known to be co-expressed with mgLPMO10A (Simmons et al., 2014). Since both the mgLPMO10A (Fig. 2) and the two CBHs (unpublished data) were thermostable, the reactions were carried out at elevated temperature (60 °C). GH48 CBHs and GH6 CBHs generally attack cellulose chains from the reducing and non-reducing ends, respectively (Barr et al., 1996). Both CBH-LPMO combinations showed clear synergistic effects, and, perhaps not surprisingly considering the metatranscriptomel data, this effect was particularly prominent for the GH48 (Fig. 4). Synergy experiments were conducted using both copper-saturated mgLPMO10A and mgLPMO10A "as is", i.e., as purified, without copper saturation. The difference between these two experiments is discussed below.

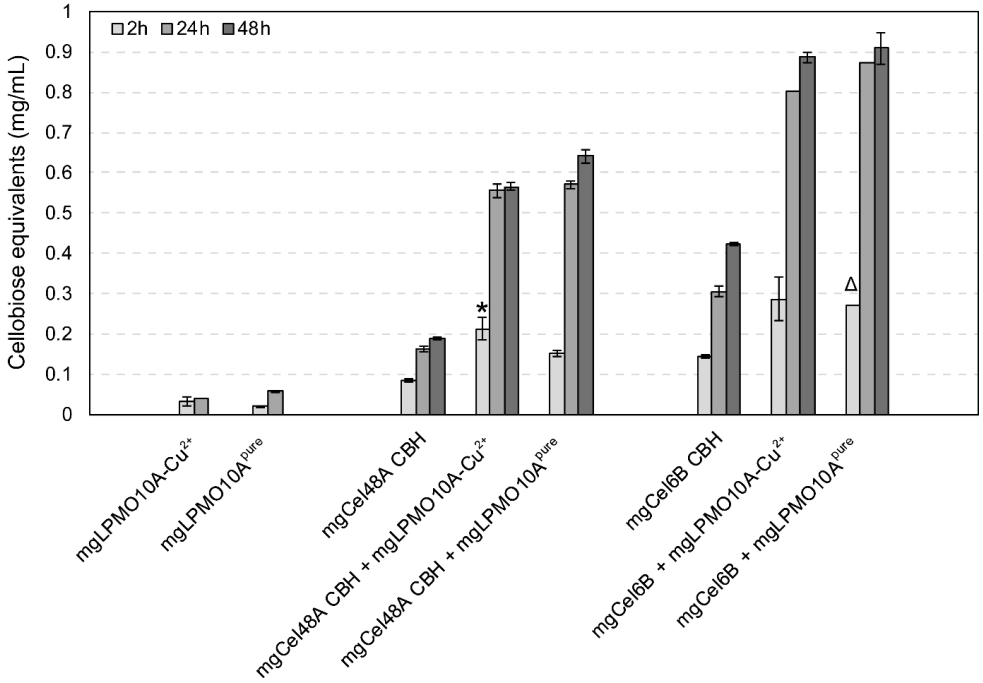

Figure 4. Synergy with mgLPMO10A and cellulases. Reactions mixtures were incubated at 60 °C and contained 10 g/L Avicel in 50 mM sodium phosphate buffer pH 6.0, and 1 µM of each enzyme in the reaction. mgLPMO10A^{pure} denotes untreated purified enzyme "as is", while mgLPMO10A-Cu^{2+} denotes copper saturated enzyme. Error bars represent standard deviations with n=3, except for reactions marked with a star and a triangle (indicating n=2 and n=1, respectively). No samples were taken for the reactions with only LPMO after 48 hours.

The mgLPMO10A was also tested in a spiking experiment with the commercial cellulase cocktail Celluclast, which is assumed to be largely devoid of LPMO activity (Müller et al., 2015). In the experiment, 15 % of the standard enzyme cocktail, consisting of Celluclast and β-glucosidase (in a 10:1 ratio on a protein basis), was replaced by mgLPMO10A or, in a control experiment, BSA (bovine serum albumin; used to test the effect of just adding protein). This setup generated clearly more products than when using 100 % of the cocktail, meaning that the presence of mgLPMO10A enables the cellulases to release more sugars (Fig. 5).

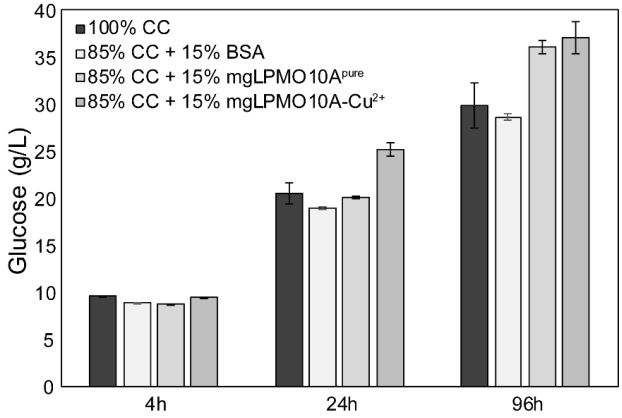

Figure 5. Effect of mgLPMO10A on saccharification of Avicel by Celluclast®. Reactions mixtures were incubated at 50 °C and contained 50 g/L Avicel in 50 mM sodium acetate buffer pH 5.0. The total protein dosage was 4 mg/g glucan, with 100% of CC (= mixture of Celluclast® and β-glucosidase), or 85 % CC and 15 % mgLPMO10A or 15 % BSA. mgLPMO10A^{pur} denotes untreated purified enzyme, while mgLPMO10A-Cu^{2+} denotes copper-saturated enzyme. All reactions were run in triplicates and standard deviations are indicated.

The effect of copper saturation

The synergy experiments were carried out with copper saturated mgLPMO10A, but also with an untreated mgLPMO10A sample taken directly from the purification (referred to as mgLPMO10A-Cu^{2+} and mgLPMO10A^{pure}, respectively). Interestingly, both synergy experiments showed that after longer processing times, the untreated LPMO was as beneficial to the reaction as the copper-saturated LPMO. However, reactions with the untreated LPMO seemed slower at some of the earlier time points, suggesting that the untreated enzyme works more slowly than the copper-saturated enzyme.

The effect of copper saturation was further investigated in additional experiments with both the copper saturated enzyme and the untreated enzyme, “as is”, i.e., used directly after the final purification step. Time course experiments with the untreated LPMO showed that product generation was slow but steady at 50 °C and also at 60 °C, albeit only for 24 h (Fig. 6). At 60 °C the untreated LPMO was about 20 times slower compared to the copper-saturated enzyme (Fig. 3). Additional time courses showed a high degree of inactivation at 70 °C, and, especially 80 °C (Fig. 6). Note that the seemingly lower stability at 70 °C and 80 °C, compared to Figure 3 may be due to both much longer incubation times and the lack of the stabilizing effect provided by copper binding.

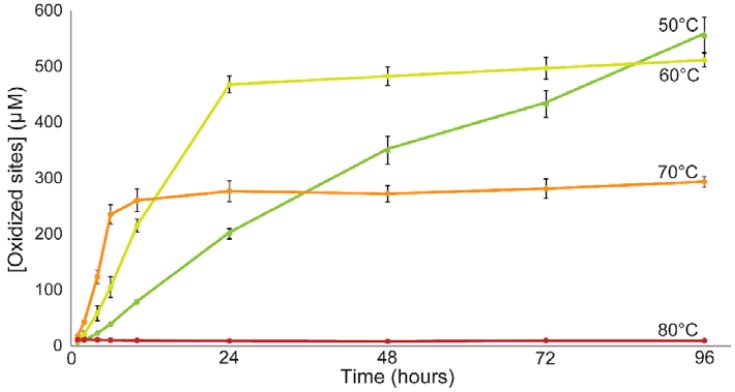

Figure 6. Progress curves using mgLPMO10A without copper saturation. The upper panel shows oxidized products generated by untreated mgLPMO10A or copper saturated *Sc*LPMO10C from 10 g/L Avicel at 40 °C and 60 °C. The Figure shows that using an LPMO that had not been copper saturated yielded stable and slow reaction kinetics. The lower panel shows only untreated mgLPMO10A at different temperatures, indicating that the reaction rate increases with the temperature (although still extremely slow compared to LPMO reactions with copper saturated enzyme).

The contrast between the progress curves for non-copper saturated mgLPMO10A (Fig. 6) and copper saturated mgLPMO10A (Fig. 3) is striking: The copper-saturated enzymes work much faster, reaching reaction end points within a few hours. These end points likely reflect a combination of depletion of ascorbic acid and enzyme inactivation. One explanation for the slow action of the untreated enzyme would be that the purified mgLPMO10A contains very little copper, meaning that most of the enzyme is in an inactive state that neither catalyzes glycosidic bond cleavage nor is vulnerable to inactivation. It is conceivable that this leads to low production of H_2O_2 in the system. Since reduced LPMOs bind the substrate more strongly (Kracher et al., 2018) and since the LPMO-substrate complex has high affinity for H_2O_2 (Bissaro et al., 2016), it is likely that the amount of reduced copper-containing LPMOs in solution is low and that the slowly produced H_2O_2 is used productively, i.e., in substrate oxidation. All in all, this would yield slow and steady kinetics.

The presence of copper in mgLPMO10A was investigated using ICP-MS, which showed that a solution of copper saturated LPMO contained approximately 1.35 copper per LPMO molecule, whereas an ultrafiltrate of this solution contained approximately 0.35 copper per LPMO molecule. Thus, one copper was bound to the LPMO, whereas the solution also contained a significant amount of free copper (approximately 0.35 per LPMO molecule).

Importantly, the ICP-MS data showed that solutions of the untreated LPMO contained only 0.07 copper per LPMO and that all this copper was bound to the enzyme. It is thus clear that most of the recombinantly produced enzyme molecules indeed were in an inactive state, as they lack copper. It remains to be studied whether the difference in kinetics between the copper saturated and the non copper saturated enzyme only relates to the lack of LPMO-bound copper or whether the presence of free copper in the saturated enzyme also plays a role (since free copper may affect the levels of H_2O_2).

Concluding remarks

The LPMO described here is, to the best of our knowledge, the most thermally stable LPMO ever described, showing appreciable activity at temperatures as high as 80 °C. Interestingly, the enzyme showed good performance at all temperatures tested and differed greatly from mesophilic *Sc*LPMO10C tested under the same conditions. It will be of great interest to further explore the mechanistic and structural basis of the special properties of mgLPMO10A. The enzyme is active and operationally stable over a wide temperature range and may thus be used in future studies of the effect of temperature of the complex LPMO reaction.

Synergy experiments showed good effects of the LPMO, comparable to what has been observed for the well-known fungal LPMO *Ta*LPMO9A (also known as *Ta*GH61A; (Harris et al., 2010, Müller et al., 2015)). It seems worthwhile to further explore how the stability of mgLMPO10A can be harnessed in biomass processing, especially since there are indications that LPMO inactivation is a limiting factor during industrial saccharification of lignocellulosic biomass (Müller et al., 2018, Chylenski et al., 2019). Of note, the low and steady activity of non copper saturated mgLPMO10A and the good results obtained when combining this LPMO with Celluclast suggest that it could be beneficial to use a lesser amount of active LPMOs in bioprocessing settings. Large amounts of reduced copper-saturated LPMOs may generate all the products within a few hours of the reaction, while industrial processing with cellulases is usually conducted over several days. At the same time, the LPMOs may experience a higher degree of autocatalytic inactivation because the ratio between reduced (and thus potentially self-destructive) LPMOs in solution and available substrate may become unfavourable. Thus, there is a risk that the LPMOs will not be active towards the end of the process, and are thus unable to assist the cellulases in attacking the remaining cellulose.

As a spin-off of this study, we revealed the enormous effect of copper saturation on LPMO activity and we showed that purified mgLPMO10A, as is, bound very little copper. The degree of copper saturation of a purified LPMO likely varies between production batches and between LPMOs, since we do not have control of the potential scavenging of copper during expression in *E. coli* and because LPMOs may have different copper affinities. Of note, EDTA is present during one of the steps used for preparation of the periplasmic extracts that were the starting samples for LPMO purification, which may explain the lack of copper in the purified enzyme. Copper saturation may secure that every LPMO has a copper coordinated in the histidine brace and is thus capable of oxidative cleavage of polysaccharides. However, copper saturation, at least when done as in this study, also adds some free copper to the reaction system and the effects of this free copper still need to be established. In any case, we show here that the effects of copper saturation of LPMOs, and, thus, the amount of active LPMO in a reaction mixture, may be considerable. Since different research groups active in the LPMO field use different procedures for protein production, purification and, if applied, copper saturation, quantitative comparison of LPMO activities in literature is challenging.

Acknowledgements

This study was funded by the Research Council of Norway through the NorZymeD project, project number 221568. We thank Phillip B. Pope and Alasdair Mackenzie for assistance in metagenome mining.

References

AACHMANN, F. L., SORLIE, M., SKJAK-BRAEK, G., EIJSINK, V. G. & VAAJE-KOLSTAD, G. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proc Natl Acad Sci U S A,* 109**,** 18779-84.

AGOSTONI, M., HANGASKY, J. A. & MARLETTA, M. A. 2017. Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases. *Microbiol Mol Biol Rev,* 81.

BARR, B. K., HSIEH, Y. L., GANEM, B. & WILSON, D. B. 1996. Identification of two functionally different classes of exocellulases. *Biochemistry,* 35**,** 586-92.

BEESON, W. T., PHILLIPS, C. M., CATE, J. H. & MARLETTA, M. A. 2012. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc,* 134**,** 890-2.

BISSARO, B., RØHR, Å. K., MÜLLER, G., CHYLENSKI, P., SKAUGEN, M., FORSBERG, Z., HORN, S. J., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2017. Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. *Nat Chem Biol,* 13**,** 1123-1128.

BISSARO, B., RØHR, Å. K., SKAUGEN, M., FORSBERG, Z., HORN, S. J., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2016. Fenton-type chemistry by a copper enzyme: molecular mechanism of polysaccharide oxidative cleavage. *bioRxiv***,** 097022.

CHYLENSKI, P., BISSARO, B., SØRLIE, M., RØHR, Å. K., VÁRNAI, A., HORN, S. J. & EIJSINK, V. G. H. 2019. Lytic Polysaccharide Monooxygenases in Enzymatic Processing of Lignocellulosic Biomass. *ACS Catalysis,* 9**,** 4970-4991.

COURTADE, G., FORSBERG, Z., HEGGSET, E. B., EIJSINK, V. G. H. & AACHMANN, F. L. 2018. The carbohydrate-binding module and linker of a modular lytic polysaccharide monooxygenase promote localized cellulose oxidation. *J Biol Chem,* 293**,** 13006-13015.

DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure,* 3**,** 853-859.

DUCHESNE, L. C. & LARSON, D. W. 1989. Cellulose and the Evolution of Plant Life. *BioScience,* 39**,** 238-241.

EIBINGER, M., GANNER, T., BUBNER, P., ROSKER, S., KRACHER, D., HALTRICH, D., LUDWIG, R., PLANK, H. & NIDETZKY, B. 2014. Cellulose surface degradation by a lytic polysaccharide monooxygenase and its effect on cellulase hydrolytic efficiency. *J Biol Chem,* 289**,** 35929-38.

EIBINGER, M., SATTELKOW, J., GANNER, T., PLANK, H. & NIDETZKY, B. 2017. Single-molecule study of oxidative enzymatic deconstruction of cellulose. *Nat Commun,* 8**,** 894.

EIJSINK, V. G. H., PETROVIC, D., FORSBERG, Z., MEKASHA, S., RØHR, Å. K., VARNAI, A., BISSARO, B. & VAAJE-KOLSTAD, G. 2019. On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). *Biotechnol Biofuels,* 12**,** 58.

FILIATRAULT-CHASTEL, C., NAVARRO, D., HAON, M., GRISEL, S., HERPOËL-GIMBERT, I., CHEVRET, D., FANUEL, M., HENRISSAT, B., HEISS-BLANQUET, S., MARGEOT, A. & BERRIN, J.-G. 2019. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnology for Biofuels,* 12**,** 55.

FORSBERG, Z., MACKENZIE, A. K., SORLIE, M., RØHR, Å. K., HELLAND, R., ARVAI, A. S., VAAJE-KOLSTAD, G. & EIJSINK, V. G. 2014a. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc Natl Acad Sci U S A,* 111**,** 8446-51.

FORSBERG, Z., MACKENZIE, A. K., SØRLIE, M., RØHR, Å. K., HELLAND, R., ARVAI, A. S., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2014b. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 8446-8451.

FORSBERG, Z., SORLIE, M., PETROVIC, D., COURTADE, G., AACHMANN, F. L., VAAJE-KOLSTAD, G., BISSARO, B., RØHR, Å. K. & EIJSINK, V. G. 2019. Polysaccharide degradation by lytic polysaccharide monooxygenases. *Curr Opin Struct Biol,* 59**,** 54-64.

HAKI, G. D. & RAKSHIT, S. K. 2003. Developments in industrially important thermostable enzymes: a review. *Bioresource Technology,* 89**,** 17-34.

HARRIS, P. V., WELNER, D., MCFARLAND, K. C., RE, E., NAVARRO POULSEN, J. C., BROWN, K., SALBO, R., DING, H., VLASENKO, E., MERINO, S., XU, F., CHERRY, J., LARSEN, S. & LO LEGGIO, L. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry,* 49**,** 3305-16.

HARRIS, P. V., XU, F., KREEL, N. E., KANG, C. & FUKUYAMA, S. 2014. New enzyme insights drive advances in commercial ethanol production. *Curr Opin Chem Biol,* 19**,** 162-70.

HEMSWORTH, G. R., TAYLOR, E. J., KIM, R. Q., GREGORY, R. C., LEWIS, S. J., TURKENBURG, J. P., PARKIN, A., DAVIES, G. J. & WALTON, P. H. 2013. The Copper Active Site of CBM33 Polysaccharide Oxygenases. *Journal of the American Chemical Society,* 135**,** 6069-6077.

HORN, S. J., VAAJE-KOLSTAD, G., WESTERENG, B. & EIJSINK, V. G. 2012. Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels,* 5**,** 45-45.

ISAKSEN, T., WESTERENG, B., AACHMANN, F. L., AGGER, J. W., KRACHER, D., KITTL, R., LUDWIG, R., HALTRICH, D., EIJSINK, V. G. & HORN, S. J. 2014. A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *J Biol Chem,* 289**,** 2632-42.

JENSEN, M. S., FREDRIKSEN, L., MACKENZIE, A. K., POPE, P. B., LEIROS, I., CHYLENSKI, P., WILLIAMSON, A. K., CHRISTOPEIT, T., OSTBY, H., VAAJE-KOLSTAD, G. & EIJSINK, V. G. H. 2018. Discovery and characterization of a thermostable two-domain GH6 endoglucanase from a compost metagenome. *PLoS One,* 13**,** e0197862.

JOHANSEN, K. S. 2016. Discovery and industrial applications of lytic polysaccharide mono-oxygenases. *Biochem Soc Trans,* 44**,** 143-9.

KITTL, R., KRACHER, D., BURGSTALLER, D., HALTRICH, D. & LUDWIG, R. 2012. Production of four Neurospora crassa lytic polysaccharide monooxygenases in Pichia pastoris monitored by a fluorimetric assay. *Biotechnol Biofuels,* 5**,** 79.

KJAERGAARD, C. H., QAYYUM, M. F., WONG, S. D., XU, F., HEMSWORTH, G. R., WALTON, D. J., YOUNG, N. A., DAVIES, G. J., WALTON, P. H., JOHANSEN, K. S., HODGSON, K. O., HEDMAN, B. & SOLOMON, E. I. 2014. Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc Natl Acad Sci U S A,* 111**,** 8797-802.

KLEMM, D., HEUBLEIN, B., FINK, H. P. & BOHN, A. 2005. Cellulose: fascinating biopolymer and sustainable raw material. *Angew Chem Int Ed Engl,* 44**,** 3358-93.

KRACHER, D., ANDLAR, M., FURTMÜLLER, P. G. & LUDWIG, R. 2018. Active-site copper reduction promotes substrate binding of fungal lytic polysaccharide monooxygenase and reduces stability. *The Journal of biological chemistry,* 293**,** 1676-1687.

LEVASSEUR, A., DRULA, E., LOMBARD, V., COUTINHO, P. M. & HENRISSAT, B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels,* 6**,** 41.

LOOSE, J. S., FORSBERG, Z., FRAAIJE, M. W., EIJSINK, V. G. & VAAJE-KOLSTAD, G. 2014. A rapid quantitative activity assay shows that the Vibrio cholerae colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS Lett,* 588**,** 3435-40.

LOOSE, J. S. M., ARNTZEN, M. O., BISSARO, B., LUDWIG, R., EIJSINK, V. G. H. & VAAJE-KOLSTAD, G. 2018. Multipoint Precision Binding of Substrate Protects Lytic Polysaccharide Monooxygenases from Self-Destructive Off-Pathway Processes. *Biochemistry,* 57**,** 4114-4124.

MANOIL, C. & BECKWITH, J. 1986. A genetic approach to analyzing membrane protein topology. *Science,* 233**,** 1403-8.

MERINO, S. T. & CHERRY, J. 2007. Progress and challenges in enzyme development for biomass utilization. *Adv Biochem Eng Biotechnol,* 108**,** 95-120.

MÜLLER, G., CHYLENSKI, P., BISSARO, B., EIJSINK, V. G. H. & HORN, S. J. 2018. The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail. *Biotechnol Biofuels,* 11**,** 209.

MÜLLER, G., VARNAI, A., JOHANSEN, K. S., EIJSINK, V. G. & HORN, S. J. 2015. Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. *Biotechnol Biofuels,* 8**,** 187.

PASPALIARI, D. K., LOOSE, J. S., LARSEN, M. H. & VAAJE-KOLSTAD, G. 2015. Listeria monocytogenes has a functional chitinolytic system and an active lytic polysaccharide monooxygenase. *FEBS J,* 282**,** 921-36.

PAULY, M. & KEEGSTRA, K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J,* 54**,** 559-68.

PHILLIPS, C. M., BEESON, W. T., CATE, J. H. & MARLETTA, M. A. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. *ACS Chem Biol,* 6**,** 1399-406.

QUINLAN, R. J., SWEENEY, M. D., LO LEGGIO, L., OTTEN, H., POULSEN, J. C., JOHANSEN, K. S., KROGH, K. B., JORGENSEN, C. I., TOVBORG, M., ANTHONSEN, A., TRYFONA, T., WALTER, C. P., DUPREE, P., XU, F., DAVIES, G. J. & WALTON, P. H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A,* 108**,** 15079-84.

REDDY, A. P., SIMMONS, C. W., D'HAESELEER, P., KHUDYAKOV, J., BURD, H., HADI, M., SIMMONS, B. A., SINGER, S. W., THELEN, M. P. & VANDERGHEYNST, J. S. 2013. Discovery of Microorganisms and Enzymes Involved in High-Solids Decomposition of Rice Straw Using Metagenomic Analyses. *PLOS ONE,* 8**,** e77985.

SIMMONS, C. W., REDDY, A. P., D'HAESELEER, P., KHUDYAKOV, J., BILLIS, K., PATI, A., SIMMONS, B. A., SINGER, S. W., THELEN, M. P. & VANDERGHEYNST, J. S. 2014. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnology for Biofuels,* 7**,** 495.

SOMERVILLE, C., BAUER, S., BRININSTOOL, G., FACETTE, M., HAMANN, T., MILNE, J., OSBORNE, E., PAREDEZ, A., PERSSON, S., RAAB, T., VORWERK, S. & YOUNGS, H. 2004. Toward a systems approach to understanding plant cell walls. *Science,* 306**,** 2206-11.

TANDRUP, T., FRANDSEN, K. E. H., JOHANSEN, K. S., BERRIN, J. G. & LO LEGGIO, L. 2018. Recent insights into lytic polysaccharide monooxygenases (LPMOs). *Biochem Soc Trans,* 46**,** 1431-1447.

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**,** 219-22.

VERMAAS, J. V., CROWLEY, M. F., BECKHAM, G. T. & PAYNE, C. M. 2015. Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. *J Phys Chem B,* 119**,** 6129-43.

WESTERENG, B., AGGER, J. W., HORN, S. J., VAAJE-KOLSTAD, G., AACHMANN, F. L., STENSTROM, Y. H. & EIJSINK, V. G. 2013. Efficient separation of oxidized cello-oligosaccharides generated by cellulose degrading lytic polysaccharide monooxygenases. *J Chromatogr A,* 1271**,** 144-52.

WOOD, T. M. 1988. Preparation of crystalline, amorphous, and dyed cellulase substrates. *Methods in Enzymology.* Academic Press.

ZAMOCKY, M., SCHUMANN, C., SYGMUND, C., O'CALLAGHAN, J., DOBSON, A. D., LUDWIG, R., HALTRICH, D. & PETERBAUER, C. K. 2008. Cloning, sequence analysis and heterologous expression in Pichia pastoris of a gene encoding a thermostable cellobiose dehydrogenase from Myriococcum thermophilum. *Protein Expr Purif,* 59**,** 258-65.

Paper IV

Tuning the substrate specificity of an AA10 LPMO from cellulose to chitin

Marianne S. Jensen[1], Geir Klinkenberg[2], Bastien Bissaro[1], Piotr Chylenski[1], Gustav Vaaje-Kolstad[1], Hans Fredrik Kvitvang[2], Guro Kruge Nærdal[2], Zarah Forsberg[1,*], Vincent G.H. Eijsink[1,*]

[1]Faculty of Chemistry, Biotechnology and Food Science, NMBU - Norwegian University of Life Sciences, Ås, Norway
[2]SINTEF Industry, Department of Biotechnology and Nanomedicine, Trondheim, Norway

Running title: *Tuning LPMO substrate specificity*

*For correspondence; E-mail: zarah.forsberg@nmbu.no & vincent.eijsink@nmbu.no

Keywords: cellulose, chitin, LPMO, directed evolution, substrate specificity

ABSTRACT

Lytic polysaccharide monooxygenases (LPMOs) are oxidative enzymes that play an important role in the turnover of polysaccharides in Nature, including cellulose and chitin. Oxidative cleavage of the polymers is initiated at the active site copper that is coordinated in a conserved histidine brace, but the structural determinants of LPMO substrate specificity remain largely unknown. In this study, we have used a directed evolution approach to investigate structural determinants of LPMO substrate specificity, with the aim of tuning the substrate preference of an AA10 LPMO from *Streptomyces coelicolor* A3(2) (ScLPMO10C, also known as CelS2) from cellulose to chitin. A rationally designed mutant library with multiple mutations at five positions on the substrate-binding surface was screened using a mass spectrometry-based high-throughput assay for detection of chitinolytic activity. A subset of the mutants that passed the screening process, containing four mutations, were characterized in detail and showed significant chitinolytic activity and a concomitant decrease in cellulose-activity. The chitin-active mutants became more rapidly inactivated during catalysis compared to a natural chitin-active AA10 LPMO, which is likely indicative of sub-optimal substrate binding leading to autocatalytic oxidative damage of the enzymes. Thus, while showing that it is possible to convert the substrate specificity of an LPMO by engineering, the present results also underpin the notion that productive substrate binding by these enzymes is a complex phenomenon that depends on a multitude of amino acids on the substrate-binding surface.

Lytic polysaccharide monooxygenases (LPMOs) constitute a class of enzymes that employ a powerful oxidative mechanism to cleave glycosidic bonds within crystalline regions of recalcitrant biomasses, such as cellulose and chitin. After the discovery of their catalytic function in 2010 (1), LPMOs have been classified as auxiliary activities (AA) in the carbohydrate active enzymes database (cazy.org; (2)) where they are categorized in families AA9-11 and AA13-16 on the basis of sequence similarity. LPMOs are abundant in Nature, where they are frequently produced by fungi and bacteria involved in the degradation of structural polysaccharides in plants, crustaceans, fungi, insects and molluscs. Thus, LPMOs hold an important role in the Earth's carbon cycle, whereas some studies suggest alternative functions, including

a role as virulence factors in microbial pathogenicity, that are not associated with biomass turnover (3-5).

LPMOs engage in a synergistic interplay with hydrolytic enzymes (e.g. cellulases and chitinases) during the conversion of recalcitrant polysaccharides, an important aspect of LPMO functionality that was observed prior to unravelling their mode of action (6,7). The crystalline surface of densely packed polysaccharides can be disrupted by the oxidative action of LPMOs, thereby generating access points for hydrolytic enzymes in areas that otherwise would be less accessible (1,8-10). The resulting boosting effect on biomass turnover makes LPMOs key components in the development of industrial bioprocessing technology (11,12).

While some LPMOs are expressed as single catalytic domains, others are associated with a carbohydrate-binding module (CBM) through a flexible linker region (13). CBMs are thought to increase the effective enzyme concentration on the substrate surface by anchoring the catalytic domain in proximity of substrate binding sites (14,15). Two conserved histidines located on the flat substrate-binding surface of the catalytic domain constitute the catalytic centre of the LPMO. These histidines coordinate a single copper ion, the reduction of which from Cu(II) to Cu(I) is crucial for catalysis (16).

LPMO catalysis has generally been believed to require molecular oxygen and a reductant that delivers two electrons for each catalytic cycle (1,17,18). However, a recent study has indicated that LPMOs may employ H_2O_2 as co-substrate (19), which implies that the reductant is only needed for a priming reduction of the LPMO, after which the enzyme can catalyze multiple reactions. While reduced LPMOs are prone to oxidative damage when exposed to high levels of O_2 or H_2O_2 in the absence of a proper substrate (19-21), well controlled H_2O_2-mediated reactions result in drastically improved catalytic rates (22) and stable reaction kinetics (12,23). Notably, under the conditions normally used in LPMO reactions, H_2O_2 will be formed either by O_2 reacting with reduced LPMO molecules that are not bound to the substrate (24,25) or by reactions involving O_2 and the reductant in the presence of trace amounts of transition metals.

LPMOs may be active on cellulose, hemicellulose, chitin, starch and/or oligosaccharides (26). Most known LPMOs are active on cellulose or chitin, which are among the most abundant polysaccharides on Earth, both consisting of β-1,4-linked monomers (glucose in cellulose and *N*-acetylglucosamine (GlcNAc) in chitin) rotated 180 ° relative to each other. LPMO-mediated cleavage of cellulose and chitin occurs by hydroxylation of the C1 or C4 carbon of the sugar monomers adjacent to the scissile glycosidic bond (1,16,27), and different LPMOs can be limited to one of these mechanisms or be capable of both. Oxidation of the C1 carbon results in the formation of 1,5-δ-lactones that are spontaneously converted to their more stable aldonic acid forms, while oxidation of the C4 carbon produces 4-ketoaldoses that are hydrated to their corresponding gemdiol form (25). Of note, for chitin, only C1 oxidation has been observed.

NMR-based studies of LPMO-substrate interactions (28,29), modelling (30), and mutational studies (20,30-33) suggest that precise arrangements of multiple amino acid side chains covering a major fraction of the substrate-binding surface dictate enzyme properties such as oxidative regioselectivity and substrate preferences. Simultaneously, these multi-side chain arrangements ensure operational stability for the LPMO by promoting precise substrate binding that protects the LPMO from oxidative self-inactivation (20). To create more insight into the structural determinants of LPMO substrate specificity, we have adopted a directed

evolution approach to convert the cellulose-active LPMO10C from *Streptomyces coelicolor* A3(2) (hereinafter called *Sc*LPMO10C, previously known as CelS2 (34)) to a chitin-specific enzyme. This was done by design of a mutant library with multiple mutations at five positions on the substrate-binding surface, which subsequently was screened using a mass spectrometry based high throughput assay for LPMO activity.

Results

Library design and mutant generation

Figure 1 shows the location of residues Tyr79, Asn80, Phe82, Tyr111, and Trp141 (position 1-5) that were targeted for mutation. The five residues are located in the L2 loop, covering a relatively large area on the flat substrate-binding surface of *Sc*LPMO10C. In terms of subsites, the mutations affect subsites -4 to -2, which are known to play an important role in substrate-binding, both in cellulose- (35) and chitin-active (30) LPMOs. These residues were postulated to dictate substrate specificity based on conservation patterns identified when comparing the sequences of cellulose- and chitin- active AA10 LPMOs, as outlined in Figure 1C. Furthermore, a $^{2}H/^{1}H$ exchange experiment had previously shown that Tyr54, Glu55, Gln57 and Thr111 in *Sm*LPMO10A (position 1-3 and 5, see below) interact with crystalline chitin (29).

Tyr79 (position 1) is the only aromatic residue on the AA10 substrate-binding surface that stacks with the carbohydrate substrate (-4 subsite in *Sm*LPMO10A (30)). This position has been targeted in recent site-directed mutagenesis studies which revealed its importance for both substrate binding (20,30) and enzyme activity and operational stability (20,31). The most common natural residues at this position are Tyr and Trp, but non-aromatic residues also occur. Thus, the Tyr at this position was mutated to Trp/Thr/Asp/Glu and Ala. Alanine substitutions were included at all positions.

The second position, Asn80 in *Sc*LPMO10C, was selected on the basis of being highly conserved in C1 and mixed C1/C4 oxidizing cellulose-active LPMO10s, and its different and strictly conserved nature in chitin-active LPMO10s, where it is a Glu. Mutation of this residue in chitin-oxidizing *Sm*LPMO10A resulted in strongly reduced substrate binding and abolished enzyme activity (6,20). In the library, Asn80 in *Sc*LPMO10C was mutated to residues with similar side chain size and polarity (Gln/Asp/Glu), as well as to Ser and Ala, which are also, but rarely, found at this position.

The third position, Phe82 in *Sc*LPMO10C, was recently shown to be important for the regioselectivity of cellulose oxidation in an LPMO10 with mixed C1/C4 activity (31). LPMO10s with mixed C1/C4-cellulose oxidation activity tend to have an Asn at this position (see Figure 1C), whereas those that are strict C1-oxidizers tends to have Phe (cellulose-oxidizers) or Gln (chitin-oxidizers) at this position. Of note, C1/C4 oxidizers active on cellulose tend to show activity on chitin, such as *Ma*LPMO10B from *Micromonospora aurantiaca*. Mutation of Asn85 in *Ma*LPMO10B to Phe almost abolished the C4-activity, but also the chitin-oxidizing activity (31), indicating an important role in substrate interactions. Position 3 was mutated to Ala/Gln/Asn/His/Tyr based on the natural variation in all types of LPMO10s.

Position 4, Tyr111 in *Sc*LPMO10C, is located in a less conserved loop in *Sc*LPMO10C and other cellulose C1-oxidizing LPMO10s. This extra loop (*Sc*LPMO10C_P110_Y111_D112) always comes with the upstream motif Asn-Trp-Phe (*Sc*LPMO10C-N80-W81-F82), which includes position 2 and 3. Position 4 was mutated to Ala/Phe/His to vary the side chain properties at this position.

The fifth position, Trp141 in *Sc*LPMO10C, is highly conserved within each of the

LPMO10 substrate-specificity groups. While the typical chitin oxidizing LPMOs (e.g. *Sm*LPMO10A-like) tend to have a polar Thr at this position, the majority of LPMO10s with mixed activities (C1/C4 on cellulose and C1 on chitin) have a Gln and the strict C1-oxidizing cellulose-active LPMO10s (*Sc*LPMO10C-like) have a Trp at this position. Here, Trp141 was mutated to Ala/Gln/Thr/Phe. The final library contained 4320 variants.

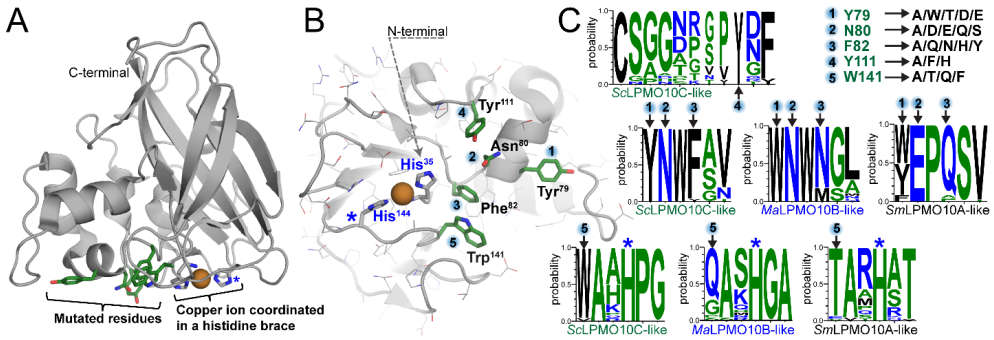

Figure 1. Structural overview of targeted residues in *Sc*LPMO10C. Panel A displays a side-view of the crystal structure of the catalytic domain of wild-type *Sc*LPMO10C (PDB 4OY7; (34)) with the substrate binding surface facing downwards. The side chains of residues targeted for mutation are shown with green-colored carbons whereas carbons in the side chains of the two histidines that coordinate the copper ion (orange sphere) are colored grey. Panel B shows the substrate-binding surface with residue numbering according to the PDB structure, which is used throughout this study. Note that His35 is the N-terminal residue of the mature protein and that the mutational work was carried out on the full-length enzyme that includes a family 2 CBM connected by a flexible Pro/Thr-rich linker (13). Panel C shows amino acid frequencies encountered at each mutated position (1-5) in multiple sequence alignments of sequences belonging to various LPMO10 subgroups: *Sc*LPMO10C-like (26 sequences; C1-oxidizing, cellulose), *Ma*LPMO10B-like (28 sequences; C1/C4 oxidation of cellulose and C1 oxidation of chitin) and *Sm*LPMO10A-like (49 sequences; C1-oxidizing, chitin). Position 4 (*Sc*LPMO10C_Y111) is exclusively found in some *Sc*LPMO10C-like sequences and is always present with the upstream motif Asn-Trp-Phe in which the first and the last make up positions 2 and 3 in the library. Mutations included in the library are shown in the top right corner of the Figure. Panels A and B were made using PyMOL and the graphs in panel C were generated using WebLogo (36). A blue star in panels A-C marks the location of the second histidine (His144) of the histidine brace. See Figure S1 for a structural comparison of natural LPMO10s with experimentally verified different substrate specificities.

High-throughput library screening

Chitin-activity was measured using a high-throughput mass spectrometry-based assay for detection of oxidized chito-oligomers. An Agilent RapidFire 365 system was applied and the solid phase extraction and triple-quadropole mass spectrometry detection methods used were established based on reaction products generated by the positive control enzyme *Sm*LPMO10A (CBP21) acting on beta-chitin as substrate. Negative ESI mode was chosen for detection because this mode led to less noise in the mass spectra. Precursor ions used for fragmentation and subsequent quantification of selected daughter fragments in the triple-

quadropole instrument corresponded to the aldonic acid forms of oxidized chitin oligomers consisting of 2 to 6 monomers (Fig. S2). Protocols for LPMO enzyme reactions in 384-well microplates were established using robotic liquid handling methods optimized to evenly distribute suspensions of solid beta-chitin.

The designed LPMO library was expressed in *E. coli* XL10 Gold and screened for production of LPMOs with activity towards β-chitin as described in the Materials and methods section. A primary screen of 11520 transformants led to selection of 42 transformants. After a second screening of these mutants for verification of activity and after discarding two mutants that contained unintended mutations discovered during sequencing, 37 transformants remained, representing 27 unique mutants, as detailed below.

Sequence analysis of putative chitin-active mutants

Table 1 provides sequence information for the 37 mutants that remained after the screening process. The collection contained 27 unique mutants, including two that were detected four times, one that was detected three times and two that were detected twice. The combined frequencies of residues among all the 37 variants (Table 1A) and among the variants that were detected multiple times (Table 1B), suggested a consensus sequence for chitinolytic activity of Tyr, Asp, Xaa, Phe and Gln for positions 79, 80, 82, 111 and 141, respectively. At position 79, 95 % of the selected variants had kept the wild-type amino acid. At the other four positions, mutation frequencies were higher and all were mutated, relative to wild-type *Sc*LPMO10C, in the deduced consensus sequence. Both N80 and Y111 were substituted in approximately 75 % of the 37 mutants, with the most common substitutions being N80D and Y111F. Residue F82 was replaced by a variety of different amino acids in 95 % of the putative chitin-active variants, and almost 50 % of these carried the mutation F82Q. W141 was the only residue that was mutated in all selected variants, with W141Q being the most frequent substitution.

Based on these observations, a subset of the mutants that were detected multiple times and carried the mutation pattern YDxFQ (M2, M5 and M18; Table 1C) were selected for production and further investigation. Of note, this mutation pattern was present in 11 of the 37 selected mutants. The contribution of individual mutations to chitinolytic activity was investigated by studying the effect of mutating single residues in M18, the most active variant (see below), back to the wild-type sequence. This required construction of two variants of M18, whereas the effect of the other reverse mutations could be assessed by characterization of variants that were among the 37 selected mutants (Table 1C).

Production of putative chitin-active mutants

The selected *Sc*LPMO10C variants and *Sc*LPMO10C wild-type were expressed and purified to electrophoretic homogeneity (Figure S3) using anion exchange purification followed by size exclusion chromatography. The yields of purified protein were typically 10-20 mg per litre culture.

Table 1. Sequence analysis and selection of mutants. Table 1A shows the frequency of different substitutions of the targeted residues (Tyr79, Asn80, Phe82, Tyr111, and Trp141) in the 37 selected putatively chitin-active mutants. Mutations included in the library are shown below the boxes. Note that for most variants chitin-activity was not verified experimentally beyond the screening procedure, so the presence of false positives cannot be excluded. Table 1B shows the frequency of different substitutions in variants that occurred multiple times. Table 1C shows the three mutants (M2, M5 and M18) that were selected for further analysis, as well as four additional mutants that were generated by site-directed mutagenesis or selected from the mutant collection to verify the contribution of individual residues to the chitin-activity of M18.

A

	Residue 79	80	82	111	141
*Sc*LPMO10C wt	Y	N	F	Y	W
substitutions in chitin-active mutants	95% Y 5% W	70% D 27% N 3% E	49% Q 24% H 14% N 8% A 5% F	62% F 27% Y 5% A 5% H	51% Q 38% A 11% T
	A/W/T/D/E	A/D/E/Q/S	A/Q/N/H/Y	A/F/H	A/T/Q/F

B

	79	80	82	111	141
variants occurring multiple times	100% Y	87% D 13% N	60% Q 27% H 13% A	80% F 20% Y	87% Q 13% A
	A/W/T/D/E	A/D/E/Q/S	A/Q/N/H/Y	A/F/H	A/T/Q/F

C

	79	80	82	111	141
M2	Y	D	Q	F	Q
M5	Y	D	H	F	Q
M18	Y	D	A	F	Q
M18_D80N	Y	N	A	F	Q
M18_A82F	Y	D	F	F	Q
M18_F111Y	Y	D	A	Y	Q
M18_Q141W	Y	D	A	F	W
*Sc*LPMO10C-wt	Y	N	F	Y	W

Verification of chitinolytic activity

Analysis of reaction products after incubation of β-chitin with purified *Sc*LPMO10C variants containing the mutation pattern YDxFQ showed that the selected enzymes indeed were active towards chitin (Fig. 2). MALDI-ToF mass spectra of reaction products generated by the three mutants (Fig. 2D-F) were very similar to spectra of products generated by the well-known chitin-active LPMO from *Serratia marcescens* (*Sm*LPMO10A or CBP21) after degradation of β-chitin (Figure 2A,B). Comparison of the zoom-in views of the hexamer cluster ($GlcNAc_5GlcNAc1A$) for *Sm*LPMO10A (Figure 2B) and the three mutants (Figure 2D-F), show identical dominating signals representing the aldonic acid (*m/z* = 1276) and the sodium salt of the aldonic acid (*m/z* = 1298). There are, however, some minor differences: next to the peak for the lactone form (*m/z* 1258), the spectra for the mutants show signals at *m/z* = 1256 and *m/z* = 1260 (see arrows in Figures 2D-F), which could indicate LPMO side activities, as discussed below. Using the same reaction set-up and analytical procedures, no oxidized chito-oligomers were detected for reactions with wild-type *Sc*LPMO10C (Figure 2C).

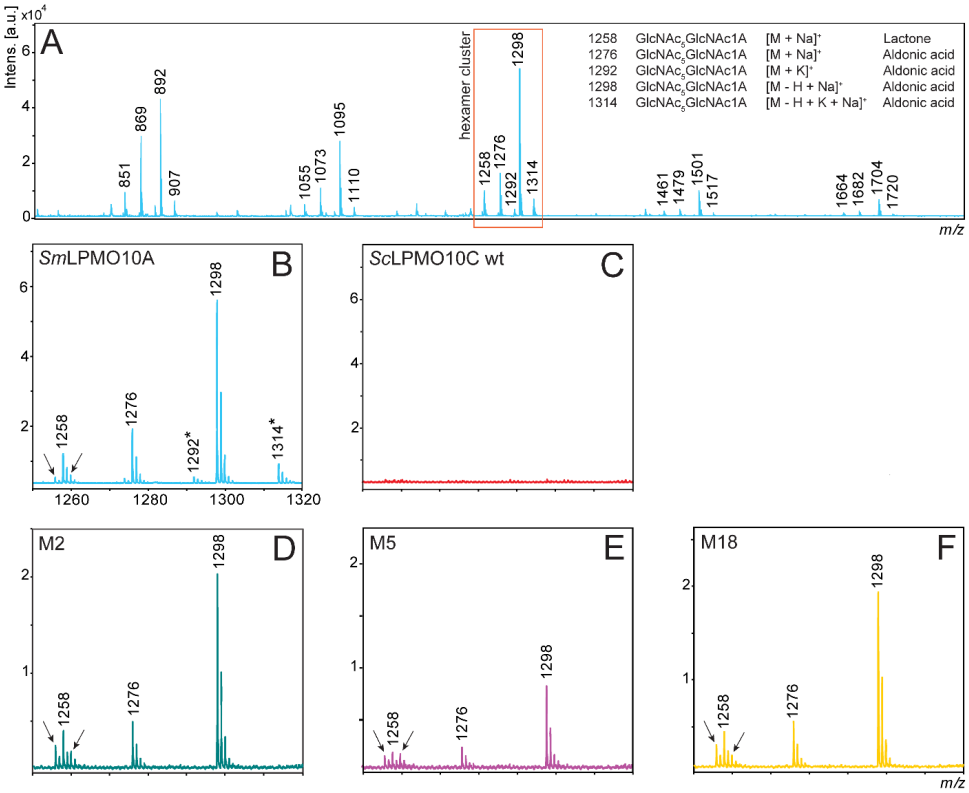

Figure 2. Oxidized products generated from chitin. Panel A shows the results of MALDI-ToF MS analysis after a 24 hours incubation of the positive control LPMO (*Sm*LPMO10A) with 10 g/L β-chitin at 40 °C in sodium phosphate pH 6.0 in the presence of 1 mM ascorbic acid (reducing agent) and displays the C1-oxidized products (aldonic acids and lactones) that can be expected for a chitin-active LPMO. Peaks representing oxidized chitooligomers are marked with their *m/z* value (see panel A for details). Panel B shows a zoom-in view of the hexamer cluster displaying *Sm*LPMO10A-generated products within the 1250 to 1320 *m/z* range. The peaks that are marked with an asterisk represent potassium adducts, which are absent in the spectra for the *Sc*LPMO10C mutants due to different storage buffers (*Sm*LPMO10A was stored in potassium phosphate pH 6.0, and the *Sc*LPMO10C mutants in sodium phosphate pH 6.0). Panel C shows a zoom-in view of the hexamer cluster area for wild-type *Sc*LPMO10C and shows no oxidized products. Panels D-F display zoom-in views of the hexamer cluster area for mutants M2, M5, and M18, respectively. Arrows in panels D-F indicate minor signals at *m/z* = 1256 and *m/z* = 1260 that are more prominent in the mutants (relative to the signals for regular oxidized products) compared to *Sm*LPMO10A.

Quantitative analysis of chitin degradation

Monitoring of chitin degradation over time showed that the mutants generated oxidized products at an initial rate close to that of *Sm*LPMO10A and that the product yield obtained in reactions with the best mutant, M18, after four hours was about half of the yield obtained with *Sm*LPMO10A (Figure 3).

Interestingly, a hardly detectable, but still significant amount of oxidized product was detected in the reaction with *Sc*LPMO10C wild-type, suggesting that the *Sc*LPMO10C wild-type has a hitherto undetected very weak intrinsic capability of cleaving chitin. Figure 3 shows that, while the mutants had good initial rates, product formation stopped earlier, compared to *Sm*LPMO10A. M18 was able to generate products over a longer period compared to M2 and M5.

The stagnation of the product formation could be caused by reductant-, substrate- or enzyme-related issues, and was investigated by adding fresh reductant, fresh substrate, or fresh enzyme to the reaction at a point where the product generation had stopped (Figure S4A). This experiment showed that enzyme inactivation is the primary reason for the cessation of product formation, as only addition of fresh enzyme led to recovery of product generation (Figure S4B).

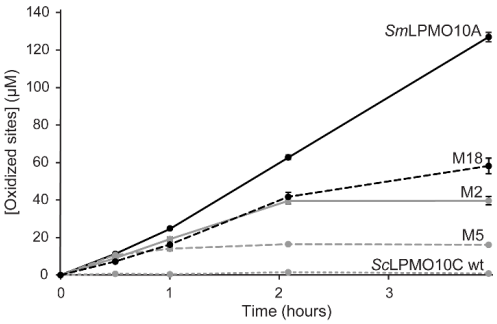

Figure 3. Time course of chitin degradation. *Sm*LPMO10A, *Sc*LPMO10C wild-type, and the three chitin-active *Sc*LPMO10C variants M2, M5 and M18 were incubated at 1 µM with 10 g/L β-chitin at 40 °C and pH 6.0 in the presence of 1 mM ascorbic acid (reducing agent) for 4 hours. A chitinase (*Sm*ChiC) was also added at 0.25 µM to obtain more rapid solubilization of LPMO generated chain ends from the insoluble substrate. The reaction was stopped by vacuum filtering and the soluble oxidized products were converted to oxidized dimers by treatment with chitobiase prior to analysis. All reactions were performed in triplicates. The error bars show ± S.D. (n = 3).

Thermostability of the mutants

The apparent melting temperatures ($T_{m,app}$) of the mutants was determined using a real-time PCR-based approach where increased binding of a fluorescent dye (SYPRO orange) indicates exposure of the hydrophobic core of the protein upon unfolding. The melting curves in Figure 4 show that *Sm*LPMO10A ($T_{m,app}$ = 71.2 °C) and *Sc*LPMO10C wild-type (64.1 °C) are more stable than M2 (52.5 °C), M5 (53.5 °C), and M18 (54.8 °C). The decreased thermal stability of the mutants is somewhat surprising as the mutants contain only four mutations relative to the wild-type that are all located on the protein surface. Still, the apparent melting temperatures of the mutants are well above the reaction temperature used in the experiments (i.e. 40 °C).

Oxidative stability in H_2O_2-driven reactions

Based on the recent finding that H_2O_2 is a relevant co-substrate of LPMOs (19), it is conceivable that LPMO products observed in reactions without added H_2O_2 are in fact generated through H_2O_2-mediated catalysis after H_2O_2 has been formed from O_2 (24,25). This would mean that the apparent initial reaction rates reflect the rate-limiting step of H_2O_2 generation from O_2.

H_2O_2 mediated catalysis was investigated in reactions with M18 by adding 15 µM H_2O_2 at 15 minute intervals (Figure 5A). Reaction samples taken in between addition of fresh H_2O_2 (after 7.5, 22.5, 37.5 and 52.5 minutes) indicated that the reactions must be fast as both M18 and the positive control, *Sm*LPMO10A, had depleted the H_2O_2 in less than 7.5 minutes, meaning that the product yields did not increase any further until fresh H_2O_2 was added. Overall, the progress curves of Fig. 5A indicate that, initially, the *Sm*LPMO10A and M18 are equally fast, but that their stabilities

differ, similar to what was concluded from Figure 3. Thus, *Sm*LPMO10A and M18 differ in terms of how much of the added H_2O_2 is consumed in productive LPMO reactions rather than in non-productive reactions with either ascorbic acid or reduced LPMOs, where the latter may generate oxidative damage on the enzyme leading to enzyme inactivation (Bissaro et al., 2017; Loose et al., 2018). This difference was even more prominent in reactions where the amount of added H_2O_2 was increased to 100 µM: in this case *Sm*LPMO10A still performed well, whereas M18 was almost immediately inactivated (Figure S5A).

Fig. 5A shows that *Sm*LPMO10A stays active for the full hour of the reaction. In fact, product formation seems to increase over time, which may be due to the accumulation of ascorbic acid, leading to in situ H_2O_2 production and/or may reflect that soluble products are more easily released as the reaction progresses (37). To gain further insight in the utilization of H_2O_2 we determined the total amount of oxidized sites in the reaction mixtures. Figure 5B shows that *Sm*LPMO10A indeed generated slightly more than 60 µM oxidized products (of which 34.6 % was soluble), in line with the total amount of added H_2O_2 being 60 µM. In contrast, M18 only generated approximately 35 µM product (of which 41.7 % was soluble), indicating that almost half of the added H_2O_2 was not converted to products. Of note, a control experiment showed that product formation after one hour by M18 was similar in H_2O_2-driven reactions with 10 times less ascorbic acid (Fig. 5C), showing that both product formation and enzyme inactivation are driven by added H_2O_2 and that reduction of the LPMO is not rate-limiting.

It is conceivable that the higher frequency of enzyme-inactivation in M18 relative to *Sm*LPMO10A could be due to weaker substrate binding of the former, since non substrate-bound reduced LPMOs are susceptible to damaging reactions with H_2O_2. However, due to the presence of the CBM2, wild-type ScLPMO10C binds well to chitin (Fig. S6; (38)), and so does M18 (which also has an intact CBM2). It would of course be interesting to assess binding of the catalytic domain of M18 to chitin and compare this with binding of the catalytic domain of the wild-type enzyme. However, despite many attempts, we have not been able to produce the catalytic domain of M18.

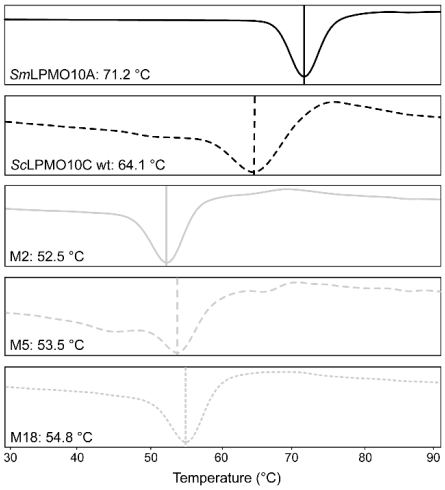

Figure 4. Mutational effects on thermostability. The plots show the apparent melting temperature ($T_{m,app}$) of copper-saturated *Sm*LPMO10A, *Sc*LPMO10C wild-type, M2, M5 and M18. The derivative of the fluorescence signal is plotted as a function of the temperature (39). The reactions contained 0.1 g/L LPMO and were heated from 25 °C to 99 °C, at a rate of 1.5 °C /min, in the presence of a fluorescent dye (SYPRO orange). The scans were performed four times for each protein and the Figures show a typical scan. All apparent melting temperatures ($T_{m,app}$) had standard deviations below ±0.1 °C.

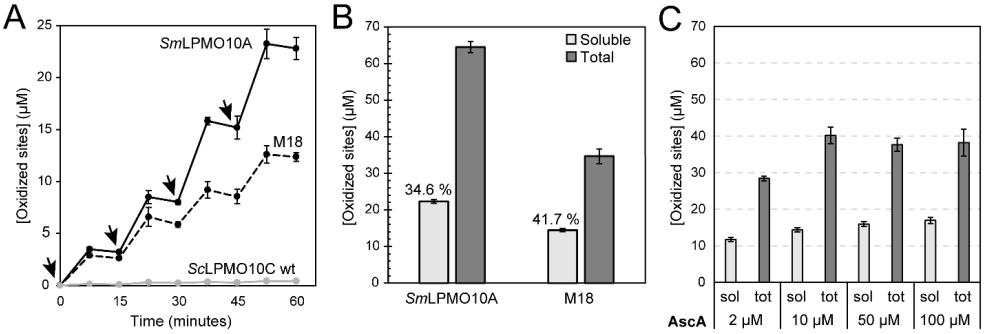

Figure 5. H_2O_2 as co-substrate. Panel A shows the amount of solubilized oxidized sites produced by *Sm*LPMO10A (orange), M18 (blue) and *Sc*LPMO10C wild-type (grey) over 60 minutes in a reactions with 10 g/L β-chitin in 50 mM sodium phosphate pH 6.0, that where supplied with 100 µM ascorbic acid (AscA) and 15 µM H_2O_2 every 15 minutes. The enzyme concentration was 1 µM in all cases. Aliquots were withdrawn both in between H_2O_2 additions (at 7.5 min, 22.5 min, 37.5 min and 52.5 min) and immediately prior to addition of fresh H_2O_2 (at 0 min, 15 min, 30 min, and 45 min). The arrows denote time points where fresh H_2O_2 was added to the reactions. Control reactions without added H_2O_2, shown in a zoom-in view in Fig. S5, yielded hardly any products). Panel B shows the total amounts of oxidized sites at the end of the reactions shown in panel A (i.e. after 60 min). Panel C shows the effect of varying the amount of added ascorbic acid in the reactions displayed in panel A (addition of 15 µM H_2O_2 + varying amounts of AscA every 15 minutes) for M18, and shows that essentially similar product yields were obtained when using a 10 times lower AscA concentration compared to the reactions shown in panel A.

Effect of individual mutations on chitinolytic activity

To investigate the contribution of each mutated residue to chitinolytic activity, M18 (YDAFQ) was used as template to reverse single mutations back to the wild-type sequence. Two of these combinations (M18_A82F and M18_F111Y) were already present amongst the mutants that had passed the screening process, and could be produced directly from the library. The two other residue combinations (M18_D80N and M18_Q141W) were generated by site-directed mutagenesis. Progress curves (Figure 6) showed that none of these single mutations reduced the chitinolytic activity of M18 to the (very low) level of wild-type *Sc*LPMO10C, but all variants performed significantly worse than M18. This suggests a cumulative effect of the mutations. Similarly to the observations made when comparing M2 and M5 to M18 (Figure 3), the mutational effects seem to primarily relate to operational stability. Three of the mutants exhibited an initial rate very similar to M18, but became inactivated at earlier time points compared to M18, whereas M18_Q141W showed a slightly reduced initial rate and slightly retarded inactivation

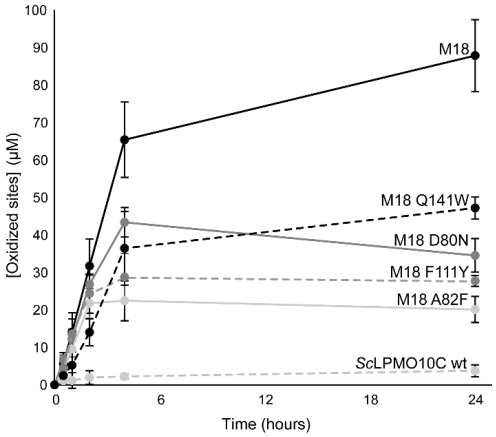

Figure 6. Effect of single mutations on chitin degradation by M18. *Sc*LPMO10C wild-type, M18 and the four M18 mutants were incubated at 1 µM with 10 g/L β-chitin at 40 °C in 50 mM sodium phosphate buffer (pH 6.0) in the presence of 1 mM ascorbic acid for 24 hours. *Sm*ChiC was also added, at 0.25 µM, to obtain more rapid solubilisation of LPMO-generated oxidized chain ends from the insoluble substrate. The reaction was stopped by filtration and soluble oxidized products were degraded to oxidized dimers with chitobiase prior to analysis. All reactions were performed in triplicates. The error bars show ± S.D. (n = 3).

Residual cellulolytic activity

All variants generated oxidized cello-oligomers upon degradation of phosphoric acid-swollen cellulose (PASC) (Figure 7), meaning that cellulose-activity was not completely lost. However, the mutants exhibited a drastic decrease in cellulolytic activity compared to wild-type *Sc*LPMO10C (Figure 7B), with product yields after four hours amounting to 3 % to 7 % of the yields obtained with the wild-type enzyme (set to 100 % product yield). M18_Q141W stood out in that product yields were high (52.5 % of wild-type levels) and because this variant showed stable kinetics, in contrast to all the other variants, which did not generate products after the first hour of incubation. It is worth noting that low activity on cellulose, resulting from the mutations needed to convert *Sc*LPMO10C into a chitin-active LPMO, is associated with reduced operational stability of the enzyme in reactions with cellulosic substrates.

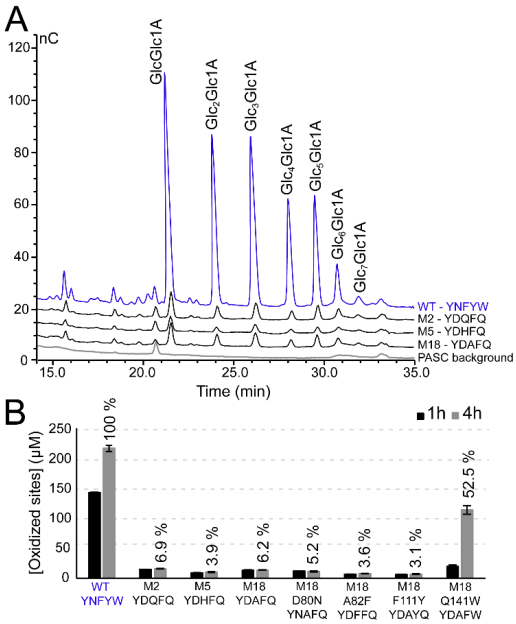

Figure 7. Residual cellulolytic activity. *Sc*LPMO10C wild-type and mutants were incubated at 1 µM with 5 g/L PASC and 1 mM ascorbic acid at 40 °C and in 50 mM sodium phosphate (pH 6.0). Panel A shows oxidized products after approximately 24 hours of incubation, varying from the dimer (GlcGlc1A) to the octamer (Glc_7Glc1A) generated by *Sc*LPMO10C wild-type, M2, M5 and M18. The background signal of the substrate (PASC) incubated with LPMO in the absence of reductant is also shown. Note that the size of the peaks of the wild-type chromatogram was reduced by 25 % before making the plot. Panel B shows the amount of

soluble oxidized sites in aliquots that were withdrawn after 1 hour and 4 hours of incubation; in this case oxidized products were quantified after the solubilized oxidized products had been degraded to dimers (GlcGlc1A) using the endoglucanase *Tf*Cel5A. Relative activities were calculated relative to the amount of product generated by the wild-type enzyme after four hours (100 %).

When comparing the activity of M18 and *Sc*LPMO10C wild type on β-chitin and PASC, it is clear that the substrate specificity has been dramatically shifted towards a preference for chitin. While M18 only gave a relative yield of 6.2 % compared to the wild-type enzyme (100 %) on PASC, M18 produced a relative yield of 100 % compared to the wild-type enzyme (1.4%) on β-chitin (Figure 8). This represents approximately a 16-fold decrease in cellulosic activity and approximately a 70-fold increase in chitinolytic activity (note that the estimated increase in chitinolytic activity is inaccurate as the product generation by the wild-type on chitin is close to the lower detection limit).

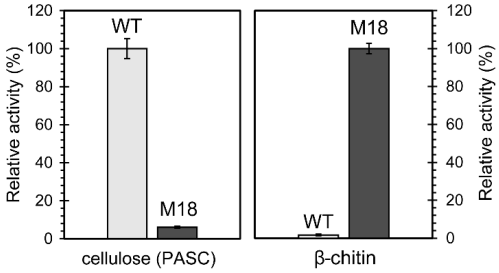

Figure 8. Comparison of chitinolytic and cellulolytic activity. Comparison of solubilized oxidized sites generated in 4 hours by 1 µM M18 or 1 µM *Sc*LPMO10C wild-type from PASC (5 g/L) or β-chitin (10 g/L). 100% activity for *Sc*LPMO10C wild-type on PASC is equal to 218.6 µM (GlcGlc1A and Glc_2Glc1A). 100% activity for M18 on β-chitin is equal to 20.6 µM (GlcNAcGlcNAc1A). The solubilized oxidized products were converted to oxidized dimers before quantification. Note that the wild-type product generation on chitin is inaccurate as the levels are close to the lower detection limit.

Discussion

Using a directed evolution approach, we were able to convert *Sc*LPMO10C from an almost exclusively cellulose-active LPMO into a chitin-active LPMO with only weak remaining cellulose-activity. None of the chitin-active variants showed C4-oxidizing activity on cellulose, as is the case for naturally occurring AA10s that act on both cellulose and chitin. As such, M18 and some of the M18 derivatives are AA10 LPMOs that have never been encountered in Nature so far: exclusively C1-oxidizing and with activity on chitin and cellulose. In this respect, M18_Q141W stands out, since it retains much of the chitin-degrading ability of M18, while regaining much of the cellulose-degrading ability of *Sc*LPMO10C. Of note, a recently described LPMO of eukaryotic origin that is the founding member of the AA15 family (40) has similar properties.

Despite the success of our approach, the selected mutants are not perfect chitin-active LPMOs in the sense that they become inactivated faster during catalysis, compared to e.g. natural chitin-active *Sm*LPMO10A. Recent mutagenesis studies on a chitin-active (20) and cellulose-active (13,31) AA10s have shown that a considerable fraction of mutations with a negative effect on LPMO performance primarily affect LPMO stability under turnover conditions. These studies showed clear correlations between decreased substrate-binding and increased inactivation. (19) have shown that reduced LPMOs are prone to oxidative damage as a consequence of the formation of reactive oxygen species through reaction of the reduced copper with O_2 or H_2O_2 in the absence of substrate. Accordingly, in their mutational study, (20) not only showed a correlation between

decreased substrate binding and an increased rate of inactivation, but also that increased inactivation is correlated to increased auto-catalytic damage (i.e., chemical modification) of residues in the catalytic center.

Accumulating data indicate that the operational stability of LPMOs likely depends on the ability of the enzyme to precisely bind to the substrate in a way that confines the emerging highly powerful reactive oxygen species to only act productively, i.e. abstracting a proton from the C1 in the scissile glycosidic bond. Modelling studies have shown that, indeed, binding of the correct substrate strongly confines the spatial orientation of the reactive oxygen species (30,41), whereas structural studies of the binding of oligomeric substrates to a fungal LPMO (42) have shown that binding of a preferred substrate leads to a more confined catalytic center compared to binding of a less preferred substrate. Thus, we propose, that binding of chitin to M18 is not yet "perfect". This will lead to imprecise coordination of reactive oxygen species in the space between the active site copper and the chemical bond that is to be cleaved, and may lead to oxidation of the active site residues, which results in irreversible inactivation of the LPMO.

On the latter note, the MALDI-ToF analyses of Figure 2 show that the three tested mutants, and in particular the two less stable mutants, M2 and M5, generated minor amounts of unexpected products (at *m/z* 1256 and *m/z* 1260). Both these species, which could be a double oxidized and a native product, respectively, are almost or completely absent in the reaction with *Sm*LPMO10A. Although the origin and nature of these species remains unknown, it is conceivable that imprecise binding of the substrate, leading to less confinement of the reactive oxygen species, may lead to non-specific reactions giving unusual products. It is also possible that auto-catalytic damage to the catalytic center reduces enzyme specificity as the reaction progresses.

Accumulating data (20,31,32,43); this study) suggest that productive interactions between LPMOs and their polymeric substrates are governed by a multitude of interactions involving many of the residues in the substrate-binding surface. The four mutations in M18, relative to *Sc*LPMO10C wild-type, seem to have a cumulative effect on the chitinolytic activity as none of the single mutations back to wild-type abolished chitinolytic activity (Fig. 6). Additionally, all the tested *Sc*LPMO10C variants, except M18_Q141W, showed low relative cellulose-activities of 3-7 % compared to the wild-type (Fig. 7). Residue 141 stood out in that it, alone, seemed to be a major determinant of cellulolytic activity, as shown by the high cellulose-activity of M18_Q141W. This is discussed further below.

Of the five residues targeted in the library, Tyr 111 (position 4), stands out because it does not have an obvious analogue in several other AA10 LPMOs, including the well-studied chitin-active *Sm*LPMO10A. This tyrosine occurs in an extra loop region of *Sc*LPMO10C and was selected because the Tyr side chain protrudes from the surface in what seems to be a critical region for substrate-binding (Fig. 1). Indeed, the sequence data (Table 1) show that almost 60 % of the selected mutants and 80 % of the mutants occurring multiple times had a phenylalanine (F) in this position, indicating that this is a favourable modification for improving productive binding of *Sc*LPMO10C to chitin. Substitution of a tyrosine with a phenylalanine may not seem as a drastic modification, but the data clearly indicate that the extra, protruding hydroxyl group of the tyrosine is unfavourable for chitinolytic activity. The fact that a smaller residue at this position is favourable for chitinolytic activity (Table 1) corresponds well with this loop being absent in chitin-oxidizing *Sm*LPMO10-like LPMOs.

The other four targeted residues all have similarly positioned analogues in natural

chitin-active LPMOs. Both MD simulations (30) and NMR studies (29) have shown that residues Tyr54, Glu55, Gln57 and Thr111 in *Sm*LPMO10A (equivalent to Tyr79, Asn80, Phe82, and Trp141 in *Sc*LPMO10C) interact with chitin during catalysis (mainly in the -5 to -2 subsites; Fig. 1). For each of the four positions, there is mutational data from at least one other LPMO, as outlined below.

Sequencing of the putatively chitin-active mutants revealed that position 1 (Tyr79) needs to be an aromatic residue, and preferably a tyrosine, since the latter was unchanged in 95 % of the variants, while the other 5 % carried a Trp. This result is consistent with the observation that mutation of this tyrosine (Tyr54) to alanine in *Sm*LPMO10A drastically reduces chitinolytic activity (20,43). While the importance of an aromatic residue at this position seems clear, the preference for Tyr or Trp seems context dependent. The Y79W mutation had hardly any effect on the cellulolytic activity of *Sc*LPMO10C, whereas mutating the Trp present at this position in the chitinolytic and cellulolytic C1/C4 oxidizing *Ma*LPMO10B to Tyr (Y82W) reduced activity on both chitin and cellulose (31).

The targeted asparagine residue (Asn80) of *Sc*LPMO10C was mutated to aspartate in 70 % of the chitin-active mutants and in almost 90 % of the mutants selected multiple times. *Ma*LPMO10B-like LPMOs (C1/C4-specific cellulose-activity and C1-specific chitinolytic activity) also have an asparagine in this position, while a glutamate occurs in C1-specific and strictly chitin-active LPMOs (*Sm*LPMO10A-like). Mutational studies of *Sm*LPMO10A have shown that this glutamate (Glu55) is crucial for activity of *Sm*LPMO10A (20,43). Glutamate and aspartate are very similar, which could explain why an aspartate was identified in the majority of the selected variants. It is worth noting the Glu was present in the library but was only found in one (3 %) of the 37 selected variants.

The mutants did not show any clear consensus for position 3 (Phe82), but among the variants that were detected multiple times the most common mutation was F82Q (M2), followed by F82H (M5) and F82A (M18). M2 and M18 generated the best product yields from chitin, and while the F82Q mutation that is present in M2 fits well with glutamine being present in the corresponding position of most *Sm*LPMO10A-like LPMOs, the F82A mutation identified in M18 is more difficult to explain. It seems clear however, that reduction of the size of the side chain at this position is beneficial for chitinolytic activity, as the original Phe was replaced by a smaller residue in 95 % of the selected mutants. This corresponds well with data obtained by (31), who showed that replacing the corresponding asparagine residue (Asn85) in cellulose- and chitin-active *Ma*LPMO10B with a phenylalanine drastically reduced chitinolytic activity.

Trp141 was the only residue that was mutated in all selected chitin-active variants, and nearly 50 % of the mutants and 90 % of the mutants occurring multiple times had the W141Q mutation in this position. While this substitution does not correspond well with threonine being present in the corresponding position of strictly chitin-active *Sm*LPMO10A-like LPMOs (Fig. 1C), it fits well with glutamine frequently being present in *Ma*LPMO10B-like LPMOs. Of note, (20) showed that mutating the corresponding residue in *Sm*LPMO10A (Thr111Ala) did not have a major impact on the activity, which could indicate that the primary reason for Trp not being favourable is its large size. The W141Q mutation alone drastically reduced the cellulolytic activity of *Sc*LPMO10C (31), but did not yield chitinolytic activity. Likewise, introduction of the Q141W mutation in M18 led to a minor reduction in chitinolytic activity, but a large increase in cellulolytic activity.

The discovery of the peroxygenase activity of LPMOs (19) has created discussions in the

field as to the nature of the natural co-substrate of the LPMOs (44-46). It has been claimed that H_2O_2-driven reactions are less specific and lead to more enzyme damage than O_2-driven reactions (46), but this is disputed (21,45). While an in-depth discussion or experimental assessment of these issues is beyond the scope of the present study, it is important to note that Fig. 5 shows that both *Sm*LPMO10A and M18 exert peroxygenase activity. Importantly, also in this case, the primary difference between M18 and the natural chitin-active *Sm*LPMO10A lies in the lower operational stability of the former.

In conclusion, the results presented in this study reveal structural determinants of the substrate specificity of LPMOs. However, first and foremost, these results underpin the complexity of substrate specificity in these fascinating enzymes. Apparently, substrate specificity depends on a multitude of correctly positioned collaborating residues and, despite progress such as described here, rational design of substrate specificity seems not yet feasible. Intriguingly, the multitude of mutations needed to confer chitinolytic activity upon *Sc*LPMO10C and the rather poor properties of intermediate stages of evolution, raises questions as to how natural evolution of LPMOs takes place. Most importantly, the present results add to the notion that operational stability, provided by an optimized LPMO-substrate combination that protects the LPMO from entering off-pathway processes, really is a key determinant of LPMO performance.

Experimental procedures

Library design and generation of ScLPMO10C mutants

Based on conservation patterns detected in a sequence alignment of 130 AA10 sequences belonging to LPMOs acting on cellulose (C1-specific and mixed C1/C4-oxidizing), chitin (C1- specific), and to LPMOs with as yet unknown substrate-specificity, residues Tyr79, Asn80, Phe82, Tyr111 and Trp141 in the N-terminal AA10 domain of *Sc*LPMO10C (Uniprot ID; Q9RJY2) were selected as targets in the design of a rational mutant library (see results section and Figure 1). In order to limit the library size, amino acid substitutions in these positions were restricted on the basis of naturally observed sequence variation. Mutations to alanine were included. The substitutions were Ala/Trp/Thr/Asp/Glu for Tyr79, Ala/Asp/Glu/Gln/Ser for Asn80, Ala/Gln/Asn/His/Tyr for Phe82, Ala/Phe/His for Tyr111, and Ala/Thr/Gln/Phe for Trp141 (see Figure 1C). This implies a library size of 4320 variants. The library was synthesized by GeneArt® Combinatorial Libraries (Life Technologies), and the amplified library was cloned into plasmid pWW0 for controlled expression with the XylS/Pm system (47,48). The plasmids harbouring the *Sc*LPMO10C encoding mutant genes were subsequently transformed into the *E. coli* XL10 gold expression host.

High throughput screening of the mutant library

Cultivation of the mutant library

The library, in *E. coli* XL10 gold, was spread on LA agar supplemented with 100 µg/ml Ampicillin to ensure single colonies. The agar plates were incubated at 37 °C overnight. Colonies were picked and inoculated to clear, flat, sterile polystyrene 384 microwell plates (Nunc 242757) containing 30 µl LB medium supplemented with 100 µg/ml Ampicillin using a Genetix QPixII robot. 11 520 colonies were picked and stored as glycerol stocks at -80 °C.

The 11 520 transformants were inoculated from the glycerol stocks using a Tecan Freedom EVO-2 200 robot system equipped with an MCA384 well pipetting tool, and cultivated in 96-well flat, polystyrene microwell plates (Greiner 655163) with Costar

lids (Corning 3931). Each well contained 80 µl of a defined basic medium (Red Hi+YE basic medium; see Supplementary Information) with 100 µg/ml Ampicillin and 2 µl inoculum. The microwell plates with cultures were incubated at 37 °C, at 90 % relative humidity, 900 rpm, and 3 mm amplitude in an Infors HT Multitron Pro shaker incubator for 20 hours. To minimise evaporation, the microwell plates were incubated in stacks of 5 and wrapped in autoclave paper.

After incubation, 40 µl of defined induction medium (Red Hi+YE induction medium; see Supplementary Information) supplemented with 1.5 mM m-toluate and 100 µg/ml ampicillin was added to the cultures using a Beckman Coulter Biomek NXP and SCARA robotic system. The cultures were then incubated at 16 °C, 900 rpm, for 24 hours, wrapped in autoclave paper. Upon this incubation, the cells were pelleted by centrifugation of the microwell plates at 4700 rpm (3828 x g) for 20 minutes in a Thermo Scientific SL 40R laboratory centrifuge and supernatants were discarded. The cells were then lysed by addition of 120 µl ion-free water followed vortex-treatment using a Quantifoil Instruments GmbH BioShake 3000 elm, at 1200 rpm for 60 sec. The goal of preparing cell lysates was to liberate correctly processed LPMO molecules from the periplasmic space. Cell lysates were stored at -20 °C.

Primary screen for enzyme activity

96-well microwell plates containing cell lysates were thawed in room temperature. Enzyme assays were performed in 384-well deep microwell plates (Corning™ 3965 Costar™ 384-Well Storage Block). The Tecan robot system equipped with an MCA384 well pipetting tool was used to add 20 µl of substrate solution (4 mg/ml β-chitin in 100 mM ammonium acetate buffer supplemented with 2 mM ascorbic acid and 15.8 µM $CuCl_2$, pH 6.0) to each well of the 384-well plates. The Beckman and SCARA robot system was then used to add 20 µl cell lysate to each well containing substrate solution (one well per transformant). The microwell plates were then sealed and given a short spin in the centrifuge to collect all sample and substrate solution in the bottom of the wells. Next, the plates were vortexed on a BioShake at 2000 rpm for 20 seconds to ensure proper mixing. The plates were then incubated over night at 40 °C, 1000 rpm, 3 mm amplitude in an Infors HT Multitron Pro shaker incubator. The enzyme reaction was stopped by addition of 160 µl 100 % acetonitrile. Plates were re-sealed, vortexed at 2000 rpm for 20 seconds and centrifuged at 4700 rpm (3828 x g) for 15 minutes to pellet insoluble material. 80 µl of precipitate-free supernatant was transferred to new 384 deep microwell plates and stored at -20 °C awaiting analysis.

The presence of soluble oxidized chitin-oligomers was assessed using an Agilent RapidFire 365 instrument (Agilent G9530A) coupled to an Agilent triple quadrupole mass spectrometry detector (Agilent G6490A). An Agilent Hilic Rapidfire cartridge (HILIC Type H1) was used as stationary phase in an SPE-fashion, whilst the mobile phases used to retain and elute the oxidized chitin oligomers were a 90-10 (v/v) mixture of acetonitrile-water and a 10-90 (v/v) mixture of acetonitrile-water, each with 0.1% (v/v) formic acid and applied using a flow rate of 1 ml/min, respectively. The mass spectrometer was equipped with an ESI ion source run in negative mode (Gas temperature: 220 °C, Gas flow: 20 L/min, Nebulizer:40 psi, Sheath Gas Temperature: 400 °C, Sheath Gas Flow: 11 L/min, Capillary: 3300 V, Nozzle Voltage: 500 V, iFunnel High: 90; iFunnel Low 60V), and the chitin oligomers were quantified in MRM mode using the following mass transitions: $DP2^{ox}$, *m/z* 439.1-> *m/z* 116.0; $DP3^{ox}$, *m/z* 642-> *m/z* 625.3; $DP4^{ox}$, 845.4 *m/z* -> *m/z* 668.2; $DP5^{ox}$, *m/z* 1048.5-> *m/z* 871.2; $DP6^{ox}$, *m/z* 1251.1-> *m/z* 871.3; where DPn^{ox}, stands for an oligomer of *N*-acetylglucosamine (GlcNAc)

comprising n sugar residues of which one is oxidized to GlcNAc1A [e.g., $DP3^{ox}$ stands for $(GlcNAc)_{n-1}GlcNAc1A$], and where the first *m/z* value given for each oligomer represents $[M-H^+]$, where M is the mass of the aldonic acid.

Cultivation and screening of candidates selected for validation screening

42 out of the 11520 transformants were selected for further screening. The location of each selected candidate in the library was identified and selected clones were spread on petri dishes containing LA agar supplemented with 100 µg/ml Ampicillin to obtain single colonies. The agar plates were incubated at 37 °C overnight. Four single colonies were selected for each candidate and inoculated to 120 µl medium in 96-well flat, polystyrene microwell plates (Greiner 655163) with Costar lids (Corning 3931). Two parallels were made: one microwell plate with LB medium supplemented with 100 µg/ml Ampicillin for use as a glycerol stock and one microwell plate with defined medium (Red Hi+YE basic medium) supplied with 100 µg/ml Ampicillin for further screening. The validation screening was performed as described above and included eight wells with *E. coli* producing the wild type *Sc*LPMO10C-enzyme and four wells with *E. coli* without plasmid as controls. After the validation screening, three of the 42 initially selected transformants were discarded as they did not show convincing chitinolytic activity.

Sequence analysis of selected variants

Single colonies of the 39 selected variants were inoculated in LB medium supplemented with 100 µg/ml Ampicillin and incubated at 37 °C and 200 rpm overnight. Glycerol stocks were prepared and stored at -80 °C, and plasmids were isolated using a NucleoSpin Plasmid kit (Macherey-Nagel, Düren, Germany) and sent for Sanger sequencing (Eurofins GATC, Cologne, Germany). The sequence data was analysed in BioEdit (www.mbio.ncsu.edu/BioEdit/bioedit.html). Two variants contained unintended mutations, leaving 37 variants for further studies. After removing redundant sequences, 27 unique variants remained (see Results section for details).

Generation of additional ScLPMO10C mutants

The gene encoding the M18 mutant (Y79/D80/A82/F111/Q141) was transferred from the XylS/Pm controlled plasmid pWW0 to the pRSETB expression vector using the InFusion HD cloning kit (Clontech) as previously described for wild-type *Sc*LPMO10C (34). Transfer to this smaller plasmid was necessary to facilitate cloning of additional variants based on the M18 construct using the QuikChange II XL site-directed mutagenesis kit (Agilent Technologies). Then, two variants of M18 were produced to partially revert M18 to *Sc*LPMO10C-WT, M18_D80N and M18_Q141W. After verification of the sequences of the mutated plasmids by Sanger sequencing (see above) the plasmids were transformed into One Shot® BL21 Star™ (DE3) chemically competent *E. coli* cells (Invitrogen) for protein expression (see below).

Enzyme production

Expression with the XylS/Pm system (pMM0 vector)

A subset of the mutants were selected for production and were produced essentially as described in (48). Briefly, *E. coli* XL10 gold carrying a mutant plasmid was inoculated from a glycerol stock in 333 mL of a defined basic medium (Red Hi+YE basic medium) supplied with 100 µg/mL ampicillin and incubated at 37 °C and 200 rpm in a shake flask, overnight. After cooling the cultures on ice for five min, 167 mL of a defined induction medium (Red Hi+YE induction medium) with 100 µg/mL

ampicillin and 1.5 mM m-tolutate (the inducer molecule of the XylS/Pm expression system) were added and the incubation was continued at 16 °C and 200 rpm for another 24 hours. Cell pellets were harvested by centrifugation at 5000 x g for 15 minutes at 4 °C (Beckman Coulter Brea, California, USA), after which a periplasmic extract was prepared using an osmotic shock method (49).

Expression with the pRSETB vector

For production of mutants M18_D80N and M18_Q141W generated in the pRSETB vector, cells from a glycerol stock were used to inoculate 500 mL Terrific Broth (TB) supplemented with 100 µg/mL ampicillin, followed by incubation at 37 °C for approximately 20 hours in a Harbinger system (Harbinger Biotechnology & Engineering, Markham, Canada). Note that expression was driven by leakiness of the T7 promoter and that no inducer molecule was added. Cells were harvested and a periplasmic extract was prepared as described above.

Protein purification

Periplasmic extracts containing the mature (i.e. signal peptide-free) proteins were filtered using 0.45 µm syringe filters (Sarstedt, Nümbrecht, Germany) and adjusted to 50 mM Tris-HCl (pH 7.5) after which the proteins were purified by anion exchange chromatography using an ÄKTA pure chromatography system (GE HealthCare, Chicago, USA) equipped with a 5-ml HiTrap DEAE FF column (GE Healthcare), with 50 mM Tris-HCl (pH 7.5) as running buffer. The proteins were eluted by applying a linear salt gradient (0-500 mM NaCl) over 60 column volumes (CV). Protein fractions were examined by SDS-PAGE (Bio-Rad, Hercules, California, USA) and relevant fractions were pooled and concentrated to 1 mL using 10,000 MWCO (Molecular Weight Cut-Off) Vivaspin ultrafiltration tubes (Sartorius, Göttingen, Germany). The proteins were further purified by size exclusion chromatography using a HiLoad 16/60 Superdex 75 column with a running buffer consisting of 50 mM Tris-HCl (pH 8.0) and 200 mM NaCl. Fractions containing the LPMO were concentrated using 10,000 MWCO Vivaspin ultrafiltration tubes (Sartorius) with simultaneous buffer exchange to 20 mM sodium phosphate buffer (pH 6.0). Protein purity was assessed by SDS-PAGE. Before being employed in enzymatic reactions, the purified LPMOs were saturated with Cu(II) by a 30 minute incubation with $CuSO_4$ in a 1:3 molar ratio (LPMO:Cu^{2+}), followed by desalting using PD Midi-Trap G-25 columns (GE Healthcare) equilibrated with 20 mM sodium phosphate (pH 6.0) to remove excess Cu^{2+}, as described previously (5). Protein concentrations were determined by measuring A_{280} (absorbance at 280 nm) in a spectrophotometer (Eppendorf Biophotometer, Eppendorf, Hamburg) and absorbances were converted to protein concentrations using theoretical extinction coefficients calculated with the ExPASy ProtParam tool. The purified and copper saturated LPMOs were stored at 4 °C.

Additional enzymes

The *Serratia marcescens* GH18 exo-chitinase *Sm*Chi18A and endo-chitinase *Sm*Chi18C, its GH20 chitobiase (*Sm*GH20) as well as its chitin-oxidizing AA10 LPMO (*Sm*LPMO10A, also known as CBP21) were expressed and purified as previously described (5,43,50). Full-length LPMO10C from *S. coelicolor* A3(2) (*Sc*LPMO10C) and a truncated variant lacking the CBM2 (*Sc*LPMO10C^{cd}) were produced and purified as previously described (34). Purified *Thermobifida fusca* GH5 endoglucanase *Tf*Cel5A (51) was provided as a kind gift from the late Prof. David Wilson (Cornell University, NY, USA).

Determination of apparent melting temperature (T_m)

The apparent melting temperature (T_m) of the proteins was determined according to a protein thermal shift assay (ThermoFisher Scientific) based on using the fluorescent dye SYPRO orange to monitor protein unfolding (39). The quantum yield of the dye is significantly increased upon binding to hydrophobic regions of the protein that become accessible as the protein unfolds. The fluorescence emission was monitored using a StepOnePlus real-time PCR machine (ThermoFisher Scientific). 0.1 g/L LPMO in 50 mM sodium phosphate buffer (pH 6.0) was heated in the presence of the dye in a 96-well plate from 25 °C to 99 °C over 50 minutes. For each protein, the experiment was carried out in quadruplicates (i.e. n = 4).

Substrates

Activity was assessed on α- and β-chitin and phosphoric acid swollen cellulose (PASC). β-chitin (extracted from squid pen and deproteinized, batch 20140101, France Chitin, Orange, France) was ball-milled and sieved in order to produce particles with a size < 0.85 mm. Shrimp (*Pandalus borealis*) shell α-chitin was purchased from Chitinor AS (Senjahopen, Norway; demineralized by hydrochloric acid treatment and subsequent deproteinized by alkaline (NaOH) treatment). PASC was prepared from Avicel PH-101 (microcrystalline cellulose) as described by (52).

Substrate binding assay

Protein binding to β-chitin was evaluated using a slightly modified variant of the A_{280} method described by (43). The binding mixtures contained 10 mg/mL β-chitin and 0.1 mg/mL protein in 50 mM sodium phosphate pH 6.0 and were incubated at 22 °C and 1000 rpm in an Eppendorf thermomixer (Eppendorf, Hamburg, Germany). Substrate binding was monitored by measuring the A280 of the supernatant at various time points, using aliquots that were taken from the binding reactions and immediately vacuum filtered over a 0.45 µm filter to separate unbound protein from protein bound to substrate. Theoretical extinction coefficients calculated with the ExPASy ProtParam tool were used to determine the protein concentration.

Chitin degradation experiments

Activity assays

Unless stated otherwise, reactions were performed in 50 mM sodium phosphate buffer (pH 6.0) in the presence of 1 mM ascorbic acid (reductant) at 40 °C and 800 rpm in an Eppendorf thermomixer. All reactions were performed in triplicates.

Qualitative analysis of chitinolytic activity

For the initial verification of chitinolytic activity, the mutants were incubated at 10 µM with 10 g/L β-chitin. Wild-type *Sc*LPMO10C and chitin-active *Sm*LPMO10A (CBP21) were included in the experiment as controls. After separation of the solid material by vacuum filtering through a 0.45 µm filter, supernatants were analysed by MALDI-ToF (see below) for detection of oxidized chito-oligomers.

Quantitative analysis of β-chitin activity

For quantitative analysis of the activity towards β-chitin, the LPMOs were incubated at 1 µM with 10 g/L β-chitin for up to 24 hours. A purified chitinase (ChiC from *S. marcescens*) was added to all reactions at 0.25 µM to facilitate solubilization of oxidized chain ends. Aliquots were withdrawn at selected time points and filtered before the soluble fraction was merged with an identical volume of a 1 µM solution of chitobiase in 5 mM Tris-HCl (pH 8.0), followed by overnight incubation at 37 °C. As a result of this procedure, all oxidized products appear as oxidized chitobiose, which facilitates

quantification. The products were analyzed using an UHPLC system (see below).

Probing the cause of enzyme inactivation

Reactions were set up as above, and samples were collected and filtered after four hours and six hours. At six hours, the remaining reaction mixture was divided into three identical fractions, where fresh reductant (1 mM) was added to one fraction, fresh reductant and fresh substrate (10 g/L) to another, and fresh reductant and fresh LPMO (1 µM) to the third fraction. The incubation was continued overnight before stopping the reactions by filtration. All supernatants were subjected to a chitobiase treatment, as above, before product quantification using an UHPLC system (see below).

LPMO reactions with H_2O_2 feeding

The LPMOs were pre-incubated at 1 µM in 50 mM sodium phosphate (pH 6.0) containing 10 g/L β-chitin and at 1000 rpm for 10 minutes to reach 40 °C. This pre-incubation was followed by addition of ascorbic acid to a final concentration of 100 µM to prime the LPMOs by reduction of the catalytic copper from the Cu(II) to the Cu(I) state. Addition of ascorbic acid was immediately followed by addition of H_2O_2 to a final concentration of 15 µM. Control reactions without added H_2O_2 were also included. Aliquots were withdrawn and immediately vacuum filtered every 7.5 minutes, and every 15 minutes a fresh batch of ascorbic acid and H_2O_2 was added subsequently to withdrawing an aliquot from the reaction. The H_2O_2-driven reactions were also carried out using different ascorbic acid concentrations (2-100 µM) and a higher H_2O_2 concentration (100 µM added every 20 minutes for 60 minutes). Reaction supernatants were subjected to chitobiase treatment, as above. At the final time point (60 minutes), next to preparing a product-containing supernatant by the usual filtration procedure, part of the reaction mixture (containing soluble and insoluble products) was boiled for 10 minutes at 98 °C to stop the LPMO reaction and merged with an identical volume of a chitinase cocktail (10 µM ChiA, 6 µM ChiC, 1 µM chitobiase), followed by overnight incubation at 37 °C and 600 rpm. The resulting hydrolysates were used to assess the total amount of oxidized products using UHPLC (see below).

Cellulose degradation experiments

The LPMOs were incubated at 1 µM with 5 g/L PASC, prepared from Avicel essentially as described in (52), and aliquots were taken and filtered after 1 and 4 hours. Oxidized products were quantified by HPAEC-PAD (see below) after conversion of the solubilized oxidized products to a mixture of oxidized dimers (GlcGlc1A) and trimers (Glc_2Glc1A) by overnight incubation at 37 °C with 0.5 µM of an endoglucanase (*Tf*Cel5A from *Thermobifida fusca*). An aliquot not degraded with endoglucanase was also analysed.

Product analysis

MALDI-ToF analysis

Reaction supernatants were assayed qualitatively using a matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) UltrafleXtreme mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) equipped with a Nitrogen 337-nm laser. Reaction products (1 µl) were applied to an MTP 384 ground steel target plate TF (Bruker Daltonics) and merged with 2 µl of 9 mg/ml of 2,5-dihydroxybenzoic acid (DHB) dissolved in 30% acetonitrile, followed by air-drying. Data collection and analysis were carried out using the Bruker FlexAnalysis software.

UHPLC analysis

Quantification of oxidized chitobiose was achieved using a Dionex Ultimate 3000 UHPLC system (DionexCorp., Sunnyvale, CA, USA) equipped with a Rezex RFQ-Fast

Acid H^+ (8%) 7.8×100 mm column (Phenomenex, Torrance, CA) operated at 85 °C. Sample components were eluted isocratically using 5 mM sulfuric acid as mobile phase, and were detected using UV absorption at 194 nm. Data collection and analysis were carried out with the Chromeleon 7.0 software. Standards were generated in-house by complete oxidation of N-acetyl-chitobiose (Megazyme; 95% purity) with a chitooligosaccharide oxidase from *Fusarium graminearum* (ChitO; (53) as described previously (5).

HPAEC-PAD analysis

Oxidized products generated from cellulosic substrates were analysed by high-performance anion-exchange chromatography (HPAEC) using a Dionex™ ICS5000 system (Thermo Scientific, Sunnyvale, CA) equipped with a disposable electrochemical gold electrode and a CarboPac PA1 column (Dionex) operated with 0.1 M NaOH (eluent A) at a column temperature of 30 °C. A multistep linear gradient with increasing amounts of eluent B (0.1 M NaOH + 1 M NaOAc) was used to elute the products, going from 0–10 % B over 10 min; 10–30 % B over 25 min; 30–100 % B over 5 min; 100–0% B over 1 min; and 0% B (reconditioning) for 9 min. Data collection and analysis were carried out with the Chromeleon 7.0 software. For analysis of endoglucanase-treated samples containing only two oxidized products, oxidized cellobiose and cellotriose, a steeper gradient of eluent B was used, as follows: 0–10 % B over 10 min; 10–14 % B over 5 min; 14–30 % B over 1 min; 30–100 % B over 2 min; 100–0 % B over 0.1 min; and 0% B (reconditioning) for 10.9 min. Cellobiose (Sigma Aldrich) and cellotriose (Megazyme) were used as substrates for production of the C1-oxidized standards cellobionic acid (GlcGlc1A) and cellotrionic acid (Glc_2Glc1A), respectively, by incubation with cellobiose dehydrogenase from *Myriococcum thermophilum* (MtCDH) (54).

Acknowledgements

This study was supported by the Research Council of Norway through the NorZymeD project, project number 221568. Infrastructure was in part supported by NorBioLab grant 226247 provided by the Research Council of Norway. We thank Lasse Fredriksen for helpful discussions.

Author contribution

MSJ carried out and interpreted most of the experiments and wrote the manuscript; BB planned the work and designed the library; ZF planned the work, designed the library, carried out experiments and supervised the project; PC designed experiments and carried out experimental work; GK & GKN developed and carried out the library screening. HFK developed the high throughput mass spectrometry method. GVK planned and supervised the project and interpreted results; VGHE planned and supervised the project, interpreted results and managed writing of the manuscript; all authors contributed to writing of the manuscript and have approved the final version for submission.

Reference list

1. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**, 219-222
2. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* **6**, 41
3. Agostoni, M., Hangasky, J. A., and Marletta, M. A. (2017) Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases. *Microbiol Mol Biol Rev* **81**
4. Paspaliari, D. K., Loose, J. S., Larsen, M. H., and Vaaje-Kolstad, G. (2015) *Listeria monocytogenes* has a functional chitinolytic system and an active lytic polysaccharide monooxygenase. *FEBS J* **282**, 921-936
5. Loose, J. S., Forsberg, Z., Fraaije, M. W., Eijsink, V. G., and Vaaje-Kolstad, G. (2014) A rapid quantitative activity assay shows that the *Vibrio cholerae* colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS Lett* **588**, 3435-3440
6. Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M., Synstad, B., and Eijsink, V. G. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J Biol Chem* **280**, 28492-28497
7. Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* **49**, 3305-3316
8. Vermaas, J. V., Crowley, M. F., Beckham, G. T., and Payne, C. M. (2015) Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. *J Phys Chem B* **119**, 6129-6143
9. Eibinger, M., Ganner, T., Bubner, P., Rosker, S., Kracher, D., Haltrich, D., Ludwig, R., Plank, H., and Nidetzky, B. (2014) Cellulose surface degradation by a lytic polysaccharide monooxygenase and its effect on cellulase hydrolytic efficiency. *J Biol Chem* **289**, 35929-35938
10. Eibinger, M., Sattelkow, J., Ganner, T., Plank, H., and Nidetzky, B. (2017) Single-molecule study of oxidative enzymatic deconstruction of cellulose. *Nat Commun* **8**, 894
11. Johansen, K. S. (2016) Discovery and industrial applications of lytic polysaccharide mono-oxygenases. *Biochem Soc Trans* **44**, 143-149
12. Müller, G., Chylenski, P., Bissaro, B., Eijsink, V. G. H., and Horn, S. J. (2018) The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail. *Biotechnol Biofuels* **11**, 209
13. Courtade, G., Forsberg, Z., Heggset, E. B., Eijsink, V. G. H., and Aachmann, F. L. (2018) The carbohydrate-binding module and linker of a modular lytic polysaccharide monooxygenase promote localized cellulose oxidation. *J Biol Chem* **293**, 13006-13015
14. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781
15. Herve, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15293-15298
16. Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J. C., Johansen, K. S., Krogh, K. B., Jorgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J., and Walton, P. H. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15079-15084
17. Kjaergaard, C. H., Qayyum, M. F., Wong, S. D., Xu, F., Hemsworth, G. R., Walton, D. J., Young, N. A., Davies, G. J., Walton, P. H., Johansen, K. S., Hodgson, K. O., Hedman, B., and Solomon,

E. I. (2014) Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc Natl Acad Sci U S A* **111**, 8797-8802

18. Beeson, W. T., Phillips, C. M., Cate, J. H., and Marletta, M. A. (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc* **134**, 890-892
19. Bissaro, B., Røhr, Å. K., Müller, G., Chylenski, P., Skaugen, M., Forsberg, Z., Horn, S. J., Vaaje-Kolstad, G., and Eijsink, V. G. H. (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2. *Nat Chem Biol* **13**, 1123-1128
20. Loose, J. S. M., Arntzen, M. O., Bissaro, B., Ludwig, R., Eijsink, V. G. H., and Vaaje-Kolstad, G. (2018) Multipoint Precision Binding of Substrate Protects Lytic Polysaccharide Monooxygenases from Self-Destructive Off-Pathway Processes. *Biochemistry* **57**, 4114-4124
21. Eijsink, V. G. H., Petrovic, D., Forsberg, Z., Mekasha, S., Røhr, Å. K., Varnai, A., Bissaro, B., and Vaaje-Kolstad, G. (2019) On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). *Biotechnol Biofuels* **12**, 58
22. Kuusk, S., Bissaro, B., Kuusk, P., Forsberg, Z., Eijsink, V. G. H., Sørlie, M., and Valjamae, P. (2018) Kinetics of H_2O_2-driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase. *J Biol Chem* **293**, 12284
23. Hegnar, O. A., Petrovic, D. M., Bissaro, B., Alfredsen, G., Varnai, A., and Eijsink, V. G. H. (2019) pH-Dependent Relationship between Catalytic Activity and Hydrogen Peroxide Production Shown via Characterization of a Lytic Polysaccharide Monooxygenase from *Gloeophyllum trabeum*. *Appl Environ Microbiol* **85**
24. Kittl, R., Kracher, D., Burgstaller, D., Haltrich, D., and Ludwig, R. (2012) Production of four *Neurospora crassa* lytic polysaccharide monooxygenases in *Pichia pastoris* monitored by a fluorimetric assay. *Biotechnol Biofuels* **5**, 79
25. Isaksen, T., Westereng, B., Aachmann, F. L., Agger, J. W., Kracher, D., Kittl, R., Ludwig, R., Haltrich, D., Eijsink, V. G., and Horn, S. J. (2014) A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *J Biol Chem* **289**, 2632-2642
26. Tandrup, T., Frandsen, K. E. H., Johansen, K. S., Berrin, J. G., and Lo Leggio, L. (2018) Recent insights into lytic polysaccharide monooxygenases (LPMOs). *Biochem Soc Trans* **46**, 1431-1447
27. Phillips, C. M., Beeson, W. T., Cate, J. H., and Marletta, M. A. (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol* **6**, 1399-1406
28. Courtade, G., Wimmer, R., Røhr, Å. K., Preims, M., Felice, A. K., Dimarogona, M., Vaaje-Kolstad, G., Sørlie, M., Sandgren, M., Ludwig, R., Eijsink, V. G., and Aachmann, F. L. (2016) Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5922-5927
29. Aachmann, F. L., Sørlie, M., Skjak-Braek, G., Eijsink, V. G., and Vaaje-Kolstad, G. (2012) NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 18779-18784
30. Bissaro, B., Isaksen, I., Vaaje-Kolstad, G., Eijsink, V. G. H., and Røhr, Å. K. (2018) How a Lytic Polysaccharide Monooxygenase Binds Crystalline Chitin. *Biochemistry* **57**, 1893-1906
31. Forsberg, Z., Bissaro, B., Gullesen, J., Dalhus, B., Vaaje-Kolstad, G., and Eijsink, V. G. H. (2018) Structural determinants of bacterial lytic polysaccharide monooxygenase functionality. *J Biol Chem* **293**, 1397-1412
32. Danneels, B., Tanghe, M., and Desmet, T. (2019) Structural Features on the Substrate-Binding Surface of Fungal Lytic Polysaccharide Monooxygenases Determine Their Oxidative Regioselectivity. *Biotechnol J* **14**, e1800211

33. Span, E. A., Suess, D. L. M., Deller, M. C., Britt, R. D., and Marletta, M. A. (2017) The Role of the Secondary Coordination Sphere in a Fungal Polysaccharide Monooxygenase. *ACS Chem Biol* **12**, 1095-1103
34. Forsberg, Z., Mackenzie, A. K., Sørlie, M., Røhr, Å. K., Helland, R., Arvai, A. S., Vaaje-Kolstad, G., and Eijsink, V. G. (2014) Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 8446-8451
35. Frandsen, K. E., Simmons, T. J., Dupree, P., Poulsen, J. C., Hemsworth, G. R., Ciano, L., Johnston, E. M., Tovborg, M., Johansen, K. S., von Freiesleben, P., Marmuse, L., Fort, S., Cottaz, S., Driguez, H., Henrissat, B., Lenfant, N., Tuna, F., Baldansuren, A., Davies, G. J., Lo Leggio, L., and Walton, P. H. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat Chem Biol* **12**, 298-303
36. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190
37. Frommhagen, M., Westphal, A. H., Hilgers, R., Koetsier, M. J., Hinz, S. W. A., Visser, J., Gruppen, H., van Berkel, W. J. H., and Kabel, M. A. (2018) Quantification of the catalytic performance of C1-cellulose-specific lytic polysaccharide monooxygenases. *Appl Microbiol Biotechnol* **102**, 1281-1295
38. Forsberg, Z., Røhr, Å. K., Mekasha, S., Andersson, K. K., Eijsink, V. G., Vaaje-Kolstad, G., and Sørlie, M. (2014) Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* **53**, 1647-1656
39. Huynh, K., and Partch, C. L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* **79**, 28 29 21-14
40. Sabbadin, F., Hemsworth, G. R., Ciano, L., Henrissat, B., Dupree, P., Tryfona, T., Marques, R. D. S., Sweeney, S. T., Besser, K., Elias, L., Pesante, G., Li, Y., Dowle, A. A., Bates, R., Gomez, L. D., Simister, R., Davies, G. J., Walton, P. H., Bruce, N. C., and McQueen-Mason, S. J. (2018) An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat Commun* **9**, 756
41. Wang, B., Johnston, E. M., Li, P., Shaik, S., Davies, G. J., Walton, P. H., and Rovira, C. (2018) QM/MM Studies into the H_2O_2-Dependent Activity of Lytic Polysaccharide Monooxygenases: Evidence for the Formation of a Caged Hydroxyl Radical Intermediate. *ACS Catalysis* **8**, 1346-1351
42. Simmons, T. J., Frandsen, K. E. H., Ciano, L., Tryfona, T., Lenfant, N., Poulsen, J. C., Wilson, L. F. L., Tandrup, T., Tovborg, M., Schnorr, K., Johansen, K. S., Henrissat, B., Walton, P. H., Lo Leggio, L., and Dupree, P. (2017) Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates. *Nature Communications* **8**, 1064
43. Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G., and van Aalten, D. M. (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J Biol Chem* **280**, 11313-11319
44. Bissaro, B., Varnai, A., Røhr, Å. K., and Eijsink, V. G. H. (2018) Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass. *Microbiol Mol Biol Rev* **82**
45. Forsberg, Z., Sørlie, M., Petrovic, D., Courtade, G., Aachmann, F. L., Vaaje-Kolstad, G., Bissaro, B., Røhr, Å. K., and Eijsink, V. G. (2019) Polysaccharide degradation by lytic polysaccharide monooxygenases. *Curr Opin Struct Biol* **59**, 54-64
46. Hangasky, J. A., Iavarone, A. T., and Marletta, M. A. (2018) Reactivity of O_2 versus H_2O_2 with polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4915-4920
47. Gawin, A., Valla, S., and Brautaset, T. (2017) The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering. *Microb Biotechnol* **10**, 702-718

48. Courtade, G., Le, S. B., Sætrom, G. I., Brautaset, T., and Aachmann, F. L. (2017) A novel expression system for lytic polysaccharide monooxygenases. *Carbohydr Res* **448**, 212-219
49. Manoil, C., and Beckwith, J. (1986) A genetic approach to analyzing membrane protein topology. *Science* **233**, 1403-1408
50. Mekasha, S., Byman, I. R., Lynch, C., Toupalová, H., Anděra, L., Næs, T., Vaaje-Kolstad, G., and Eijsink, V. G. H. (2017) Development of enzyme cocktails for complete saccharification of chitin using mono-component enzymes from Serratia marcescens. *Process Biochemistry* **56**, 132-138
51. Irwin, D. C., Spezio, M., Walker, L. P., and Wilson, D. B. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnol Bioeng* **42**, 1002-1013
52. Wood, T. M. (1988) Preparation of crystalline, amorphous, and dyed cellulase substrates. in *Methods in Enzymology*, Academic Press. pp 19-25
53. Heuts, D. P., Winter, R. T., Damsma, G. E., Janssen, D. B., and Fraaije, M. W. (2008) The role of double covalent flavin binding in chito-oligosaccharide oxidase from *Fusarium graminearum*. *Biochem J* **413**, 175-183
54. Zamocky, M., Schumann, C., Sygmund, C., O'Callaghan, J., Dobson, A. D., Ludwig, R., Haltrich, D., and Peterbauer, C. K. (2008) Cloning, sequence analysis and heterologous expression in *Pichia pastoris* of a gene encoding a thermostable cellobiose dehydrogenase from *Myriococcum thermophilum*. *Protein Expr Purif* **59**, 258-265

Tuning the substrate specificity of an AA10 LPMO from cellulose to chitin.

Marianne S. Jensen[1], Geir Klinkenberg[2], Bastien Bissaro[1], Piotr Chylenski[1], Gustav Vaaje-Kolstad[1], Hans Fredrik Kvitvang[2], Guro Kruge Nærdal[2], Zarah Forsberg[1,*], Vincent G.H. Eijsink[1,*]

Running title: *Tuning LPMO substrate specificity*

[1]Faculty of Chemistry, Biotechnology and Food Science, NMBU - Norwegian University of Life Sciences, Ås, Norway
[2]SINTEF Industry, Department of Biotechnology and Nanomedicine, Trondheim, Norway

*For correspondence; E-mail: zarah.forsberg@nmbu.no & vincent.eijsink@nmbu.no

Supporting Information

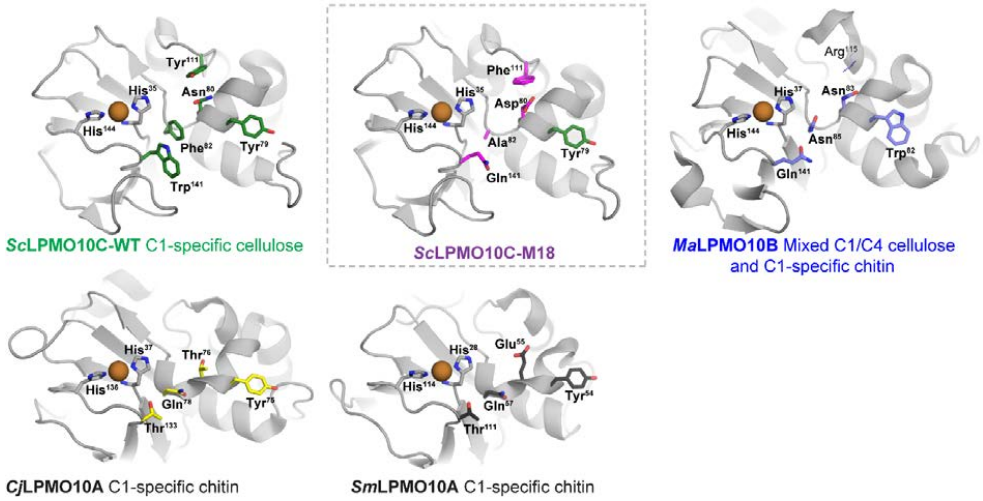

Figure S1. Substrate binding surfaces of various LPMO10s, highlighting the residues targeted in this study. The pictures show the catalytic domains of *Sc*LPMO10C (PDB 4OY7; (Forsberg et al. 2014a)) representing a C1-specific cellulose oxidizer, *Ma*LPMO10B (PDB 5OPF; (Forsberg et al. 2018)) representing LPMO10s with mixed C1/C4 cellulose oxidizing activity and C1-specific chitin-oxidizing activity, M18 with C1-specific chitin oxidizing and some remaining cellulose-oxidizing activity (this study), and *Cj*LPMO10A (PBD 5FJQ; (Forsberg et al. 2016)) and *Sm*LPMO10A (PDB 2BEM; (Vaaje-Kolstad et al. 2005b)), which are both C1-specific chitin oxidizers. The structural model of M18 was made using the mutagenesis tool in PyMOL. Position 4 is only found in *Sc*LPMO10C-like enzymes, although a conserved arginine in *Ma*LPMO10B (R115, shown as lines) has its side chain in the same position as *Sc*LPMO10C_Y111.

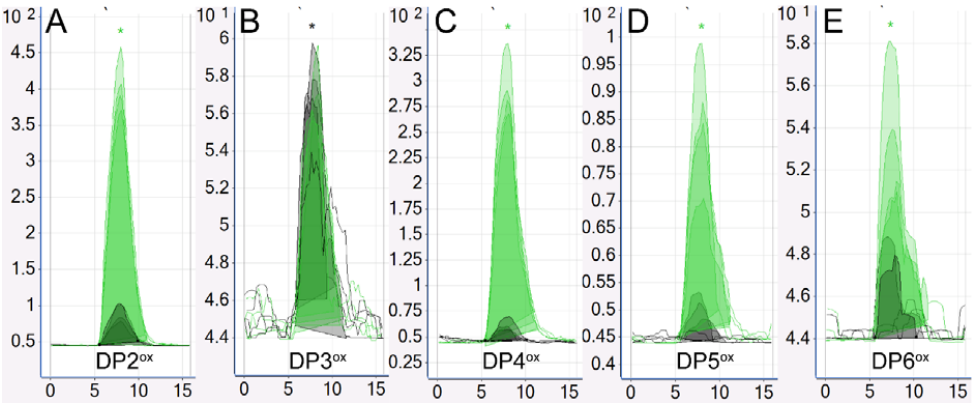

Figure S2. Screening for chitinolytic activity. RapidFire ESI triple quadrupole MRM chromatograms for M2 (four independent wells, light green) and wild-type *Sc*LPMO10C (four independent wells, grey). The ESI MRM intensity is given on the y-axis while time (seconds) is on the x-axis Chromatograms A-E: A) DP2^{ox}: ESI MRM m/z 439.1-> m/z 116.0, B) DP3 ^{ox}: m/z 642-> m/z 625.3, C) DP4 ^{ox}: 845.4 m/z -> m/z 668.2, D) DP5 ^{ox}: m/z 1048.5-> m/z 871.2, E) DP6 ^{ox}: m/z 1251.1-> m/z 871.3. Note the differences in intensity on the y-axis.

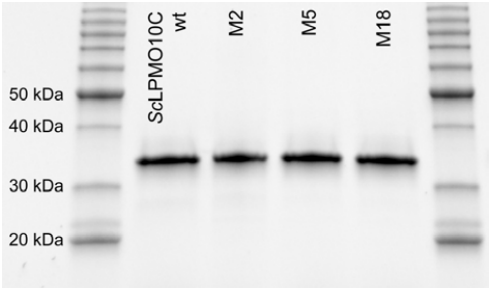

Figure S3. Enzyme purity. SDS-PAGE of *Sc*LPMO10C wild-type and mutants M2, M5 and M18 after expression using the Pm/XylS system and purification of the enzyme from periplasmic extracts by anion exchange followed by size exclusion chromatography. The samples loaded onto the gel contained 50 pmol of protein (approx. 1.7 µg). A protein benchmark ladder was included on either side to verify the molecular size of *Sc*LPMO10C (calculated to be 34.6 kDa).

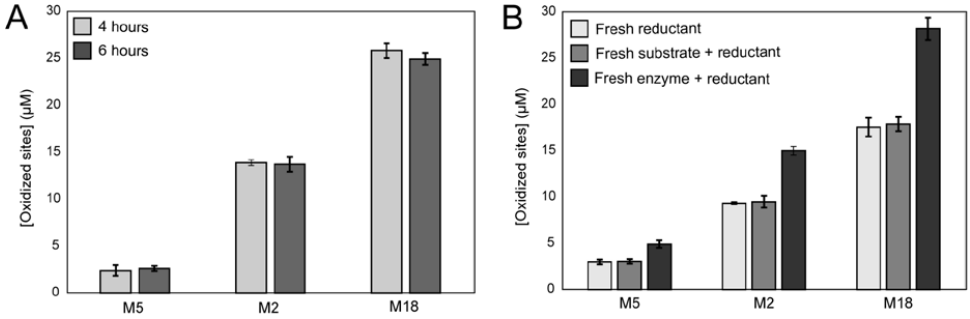

Figure S4. Probing the cause of early enzyme inactivation. The mutants M5, M2 and M18 were incubated at 1 µM with 10 g/L β-chitin at 40 °C and pH 6.0 in the presence of 1 mM ascorbic acid. The oxidized dimer was quantified after degradation of the solubilized oxidized products by chitobiase. Panel A demonstrates that for the three mutants, product generation had stopped prior to four hours of incubation, as no increase in product yield was detected between four and six hours. Panel B shows product formation upon addition of fresh reactants at 6 hours followed by continued incubation for another 18 hours. The data show that neither addition of fresh reductant (1 mM ascorbic acid) nor fresh substrate (10 g/L) together with fresh reductant restored the reaction. When fresh LPMO (1 µM) and fresh reductant were added, the product yield increased for all three mutants, indicating that the stagnation in product formation is primarily due to inactivation of the LPMO.

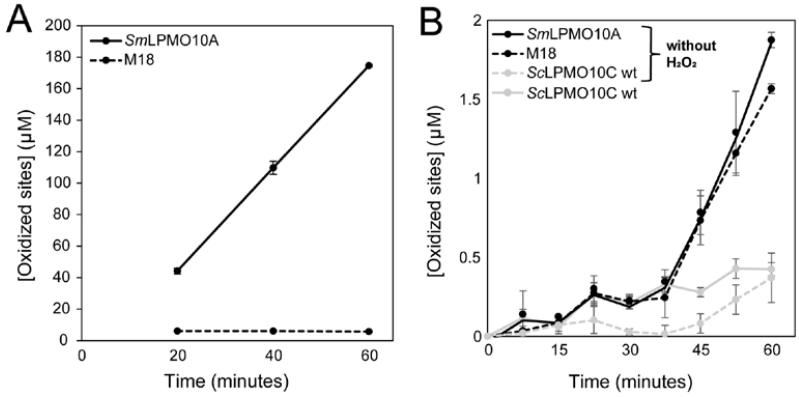

Figure S5. Oxidative instability of M18 and control reactions without addition of H_2O_2. Panel A shows the same reaction as displayed in Figure 5, except that H_2O_2 was added every 20 minutes at 100 µM (together with 100 µM ascorbic acid). Sample aliquots were only withdrawn immediately prior to adding fresh H_2O_2. Panel B shows a zoom-in view of control reactions related to Figure 5A. In the reactions labeled "without H_2O_2", water was added to the reactions instead of H_2O_2. The graph also shows H_2O_2–driven product formation by wild-type *Sc*LPMO10C (same curve as in Fig. 5A). Note that these are very low product levels, making quantification difficult.

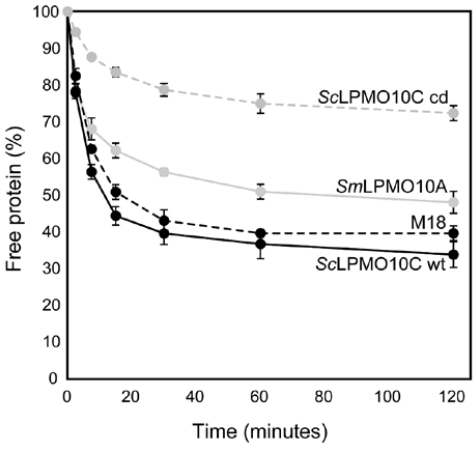

Figure S6. Chitin-binding by various LPMOs. The curves show binding to β-chitin for *Sm*LPMO10A (which does not contain a CBM), wild-type *Sc*LPMO10C (with a CBM2), M18 (with a CBM2) and the catalytic domain (cd) of *Sc*LPMO10C.

Supplementary methods

Recipe for Red Hi+Ye medium.
The initial phase of the incubation was carried out in the basic medium, while the induction medium is added later. See details in the experimental procedures.

1. Basic medium

Stock solutions:

Component	g/L
Fe(III) citrate hydrate	6
H_3BO_3	30
$MnCl_2*4H_2O$	10
$EDTA*2H_2O$	84
$CuCl_2*2H_2O$	15
$Na_2Mo_4O_4*2H_2O$	25
$CoCl_2*6H_2O$	25
$Zn(CH_3COO)_2*2H_20$	4

Solution A:

Component	g/L	ml stock/L
$Na_2HPO_4*2H_2O$	8.6	
KH_2PO_4	3	
NH_4Cl	0.3	
NaCl	0.5	
Fe(III) citrate hydrate		3.3
H_3BO_3		0.03
$MnCl_2*4H_2O$		0.5
$EDTA*2H_2O$		0.03
$CuCl_2*2H_2O$		0.03
$Na_2Mo_4O_4*2H_2O$		0.03
$CoCl_2*6H_2O$		0.03
$Zn(CH_3COO)_2*2H_20$		0.67
Yeast extract (Oxoid)	2	
Glucose*H2O	11	

Solution B:

Component	g/L
Citrate monohydrate	20

Adjust pH to 7.0 with NaOH

Solution C:

Component	Molar
$MgSO_4 * 7H_2O$	1

650 mL Solution A is mixed with 50 mL Solution B and 0.9 mL Solution C.
Ampicillin is added to a final concentration of 100 μg/mL before use.

2. Induction medium

Component	g/L
Glycerol (87%)	15
Yeast extract	24

1.5 M m-toluate stock in ethanol is added to a final concentration of 1.5 mM before use.

Ampicillin is added to a final concentration of 100 μg/mL before use.

ISBN: 978-82-575-1622-2
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

104083 / ANDVORDGRAFISK.NO