ISBN 978-82-326-3463-7 (printed version)
ISBN 978-82-326-3462-0 (electronic version)
ISSN 1503-8181

Gaston Courtade

Doctoral Thesis

Doctoral theses at NTNU, 2018:337

Gaston Courtade

An NMR investigation of lytic polysaccharide monooxygenases

NTNU
Norwegian University of
Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2018:337

NTNU

NTNU
Norwegian University of
Science and Technology

Gaston Courtade

An NMR investigation of lytic polysaccharide monooxygenases

Thesis for the degree of Philosophiae Doctor

Trondheim, October 2018

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences
Department of Biotechnology and Food Science

ISBN 978-82-326-3463-7 (printed version)
ISBN 978-82-326-3462-0 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:337

Printed by Skipnes Kommunikasjon as

Acknowledgements

Biochemistry has allowed humans to do ordinary (albeit challenging) things in the quest to understand the extraordinary and intricate machinery that makes life possible. LPMOs are not short on mysteries and it has been a true adventure to work on elucidating some of their secrets. NMR, being far from an ordinary tool, has been a marvelous companion in this quest – in which this thesis is hopefully just the beginning.

I would like to extend my most sincere gratitude to my supervisors Finn L. Aachmann and Trygve Brautaset. Finn not only dedicated an extraordinary amount of effort to supporting, guiding and encouraging me, but also created a group where comradeship prevails and people flourish. Thank you Aleksander, Annalucia, Edith, Elena, Eva, Georg, Gerd Inger, Gracia, Maya, Margrethe, Olav, Oscar and Yoshi for your companionship in this journey.

This thesis would not have come to be without the support of Zarah Forsberg and Vincent Eijsink at the Norwegian University of Life Sciences (NMBU), to whom I am grateful for their excellent work, help and contributions to this thesis. In addition, many thanks go to Morten Sørlie, Åsmund Røhr Kjendseth, Dejan Petrović and Gustav Vaaje-Kolstad, also at NMBU, for fruitful collaboration and discussions. Likewise, I would like to thank the collaboration with the groups of Roland Ludwig, Mads Sandgren, Marylène Vandevenne, Tamo Fukamizo and Paul Walton. Additionally, I would also like to extend my thanks to SINTEF, Vectron Biosolutions AS and RISE PFI AS for their partnership. I am also grateful to Reinhard Wimmer for kindly welcoming me to stay at his laboratory at Aalborg University.

Friends, far and near, have helped me relax and enjoy my time off from work; for this I am grateful to all of you. Also, thank you Morten and Edith for always having time for a beer, as well as my office mates and lab mates for the great company in and outside of work.

I am grateful to my family who, throughout our longest journey, have always been on my side, nurtured me and cheered me on.

Finally, I wish to thank Mimmi for her endless caring patience and inspiring kindness. This would not have been without you and your love.

Preface

This thesis is submitted as part of the requirements to attain the academic title of Philosophiae Doctor at NTNU Norwegian University of Science and Technology. The work was carried out at NOBIPOL, Department of Biotechnology and Food Science under the supervision of Professor Finn Lillelund Aachmann and co-supervision of Professor Trygve Brautaset. The thesis was financed by the Faculty of Natural Sciences at NTNU.

The thesis is composed of an introduction, aim and highlights of the thesis, methods, additional results and discussion, concluding remarks and future perspectives, appendices, and papers. The appendices include one-page summaries of the following contributions which, though relevant, fall outside the scope of the thesis.

Additional contribution I:

Crasson, O., **Courtade, G.**, Léonard, R. R., Aachmann, F. L., Legrand, F., Parente, R., Baurain, D., Galleni, M., Sørlie, M., Vandevenne, M. (2017). Human chitotriosidase: catalytic domain or carbohydrate binding module, who's leading HCHT's biological function. *Scientific Reports*, 7, 2768–2777

Additional contribution II:

Petrović D. M., Bissaro B., Chylenski P., Skaugen M., Sørlie M., Aachmann F. L., **Courtade G.**, Várnai A., Eijsink V. G. H. (2018) Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation. *Protein Science.* doi: 10.1002/pro.3451

Summary

Advancing towards a bioeconomy requires developing technology that enables the pursuit of a bio-based society to prosper. It is thus crucial to find sustainable biomass sources to use as energy, feed and biomaterials. However, efficient utilization of cellulose and chitin biomass is challenging because of the recalcitrance conferred by their tightly packed crystalline structures. A solution to this challenge came with the discovery of lytic polysaccharide monooxygenases (LPMOs). LPMOs are copper-dependent enzymes that bind to the crystalline surface of a variety of plant cell-wall polysaccharides and chitin, causing cleavage of 1,4-glycosidic bonds by an oxidative mechanism. In addition to requiring copper, LPMOs need a source of electrons and molecular oxygen or hydrogen peroxide. LPMO action on crystalline polysaccharides results in enhanced activity of hydrolytic enzymes, leading to more efficient enzymatic biomass degradation. In spite of the ample industrial and scientific interest for LPMOs, several key aspects about their nature had remained insufficiently understood.

The aim was to use the unique versatility of nuclear magnetic resonance (NMR) spectroscopy to characterize and better understand key aspects of LPMO functionality; including their structural diversity, the effect of copper on LPMO structure and substrate interactions, the molecular basis of their substrate binding, the location of the electron transfer event, their dynamic features, and the role of LPMO-associated carbohydrate-binding modules (CBMs).

The structure of chitin-active *Bl*LPMO10A was elucidated by NMR, verifying the highly conserved structure of LPMOs. The Cu(I)-*Bl*LPMO10A structure was also characterized, showing that copper only has marginal effects on the structure, limited to rearrangements in the copper active site. Mapping interaction of *Nc*LPMO9C with soluble substrates showed that the surface of the LPMO adapts to binding branched substrates such as xyloglucan. Additional substrate interaction studies were carried out on *Sc*LPMO10C and *Bl*LPMO10A, where the substrate binding surfaces were mapped using insoluble cellulose nanofibrils and chitin flakes, respectively. Competition experiments monitored by NMR showed that electron transfer from cellobiose dehydrogenase occurs directly on the copper site of *Nc*LPMO9C, indicating that the electron transfer event during the

LPMO reaction must occur before substrate binding. NMR was used to study the dynamic properties of full-length modular *Sc*LPMO10C consisting of *Sc*CBM2, a linker region and the LPMO domain. Results showed that the linker is disordered and extended, and that most of the cellulose binding affinity resides on *Sc*CBM2. Supplementary cellulose oxidation experiments suggested that these dynamic properties of the linker promote multiple localized oxidation of the substrate by the modular LPMO. LPMO-associated CBMs were studied by elucidating the NMR structures of *Sc*CBM2 (cellulose-binding), *Cj*CBM5 and *Cj*CBM73 (both chitin-binding). The structures depict typical surfaces rich in aromatic residues for binding crystalline polysaccharides.

To obtain these results, experimental strategies were adopted and developed, including the design and construction of a new expression system to produce isotopically enriched bacterial LPMOs, as well as the implementation of various NMR-based approaches supplemented by biophysical (isothermal titration calorimetry, electron paramagnetic resonance), biochemical and computer modelling techniques. Taken together, the outcomes presented here provide comprehensive insights into LPMO structure and function, encompassing key aspects that are now better understood than at the start of the thesis.

Symbols and abbreviations

AA10	Auxiliary Activity Family 10
AA9	Auxiliary Activity Family 9
AFM	Atomic Force Microscopy
AUTOPSY	AUTOmated Peak picking for NMR SpectroscopY (Software)
B_0	Longitudinal magnetic field strength, in units of Tesla (T)
B_1	Amplitude of the transverse magnetic field
*Bl*LPMO10A	First auxiliary activity family 10 lytic polysaccharide monooxygenase from *Bacillus licheniformis*
BMRB	Biological Magnetic Resonance Bank; repository for protein NMR assignments
c^2, CSA	Chemical Shift Anisotripy contribution to relaxation
CARA	Computer Assisted Resonance Assignment (Software)
CBM	Carbohydrate Binding Module
CBM14	Carbohydrate Binding Module 14 of the human macrophage chitotriosidase
CBM33	Carbohydrate Binding Module Family 33
CBP21	Chitin Binding Protein 21, also known as *Sm*LPMO10A
CDH	Cellobiose DeHydrogenase
CelS2	Another name for *Sc*LPMO10C
ChiA, ChiB	Exo-acting chitobiohydrolases
ChiC	Endochitinase
*Cj*CBM5	Carbohydrate Binding Module family 5 in *Cj*LPMO10A
*Cj*CBM73	Carbohydrate Binding Module family 73 in *Cj*LPMO10A
*Cj*LPMO10A	First auxiliary activity family 10 lytic polysaccharide monooxygenase from *Cellvibrio japonicus*
CNF/NFC	Cellulose NanoFibrils/NanoFibrilated Cellulose
COSY	COrrelation SpectroscopY
CPMG	Carr-Purcell-Meiboom-Gill; relaxation dispersion (NMR experiment)
CS-ROSETTA	Chemical Shift- ROSETTA (Software)
CYANA	Combined Assignment and Dynamics Algorithm for NMR Applications (Software)
CYT	Heme *b* cytochrome domain of cellobiose dehydrogenase (CDH)
C^{α}, C^{β}, C^{γ}, C^{δ}, C'	Alpha, beta, gamma, delta or carbonyl carbon atom in an amino acid
d^2	Dipolar contribution to relaxation
D_x, D_y, D_z	Principal components of the diffusion tensor
EPR	Electron Paramagnetic Resonance
f, ν, δ	Frequency, chemical shift
FID	Free Induction Decay
FRET	Förster Resonance Energy Transfer
GH	Glycoside Hydrolase
GH61	Glycoside Hydrolase Family 61
Glc_6	Glucose hexamer
GlcN	Glucosamine, chitosan monomer
GlcNAc	N-acetyl glucosamine, chitin monomer
hetNOE	Heteronuclear $\{^1H\}$-^{15}N NOE
HSQC	Heteronuclear Single Quantum Correlation (NMR experiment)
H^{α}, H^{β}, H^{γ}, H^{δ}, H^N	Alpha, beta, gamma, delta or amide protons in an amino acid
IEX	Ion EXchange chromatography
INEPT	Insensitive Nucleus Enhancement by Polarization Transfer (NMR pulse sequence)
I-PINE	Integrative-Probablistic Interaction Network of Evidence (Software)
ITC	Isothermal Titration Calorimetry
$J(\omega)$	Spectral density function
$^1J_{HN}$	Scalar coupling between H^N and N
$^3J_{HH\alpha}$	Scalar coupling between H^N and Hα
k	Exchange constant

LB	Lysogeny Broth (complex medium)
LPMO	Lytic Polysaccharide MonoOxygenase
M9	Defined medium containing only one source of carbon and only one source of nitrogen
MD	Molecular Dynamics
MOLMOL	Molecule Analysis and Molecule Display (Software)
MUSIC	NMR pulse scheme: Multiplicity Selective In-phase Coherence transfer
Mw	Molecular weight
Mx, My, Mz	Net magnetization in x, y and z axes
N	Amide nitrogen
*Nc*LPMO9C	Third auxiliary activity family 9 lytic polysaccharide monooxygenase from *Neurospora crassa*
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PDB	Protein Data Bank
Pm/XylS	Expression system consisting of the Pm promotor and XylS activator
PRE	Paramagnetic Relaxation Enhancement
r	Internuclear distance
RDCs	Residual Dipolar Couplings
RF	Radio Frequency
ROTDIF	ROTational DIFfusion (Software)
S. D.	Standard Deviation
S^2	order parameter
SAXS	Small Angle X-ray Scattering
*Sc*AA10	Catalytic domain of *Sc*LPMO10C
*Sc*CBM2	Carbohydrate Binding Module family 2 in *Sc*LPMO10C
*Sc*LPMO10C	Third auxiliary activity family 10 lytic polysaccharide monooxygenase from *Streptomyces coelicolor*. Full-length enzyme containing *Sc*AA10, a linker region and *Sc*CBM2
SDS-PAGE	Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SSD	Sum of Squared Differences
SUMO	Small Ubiquitin-like MOdifier
T_1	Spin-lattice/longitudinal relaxation time
$T_1\rho$	T_1 relaxation time measurement in the rotating frame; relaxation dispersion (NMR experiment)
T_2	Spin-spin/transversal relaxatiom time
T7lac / pRSET	Expression system consisting of lac operon and T7 RNA polymerase
TALOS-N	Torsion Angle Likeliness Obtained from Shift and Sequence Similarity (Software)
*Ta*LPMO9A	First auxiliary activity family 9 lytic polysaccharide monooxygenase from *Thermoascus auranticus*
TOCSY	TOtal Correlation Spectroscopy
TSP	3-(TrimethylSilyl)-Propionic acid
W_0, W_1, W_2	Zero, single, double quantum transitions
YASARA	Yet Another Scientific Alternative Reality Application (Software)
γ	Gyromagnetic ratio
δ	Chemical shift
Φ	Phi dihedral angle
Ψ	Psi dihedral angle
Ω	Omega dihedral angle

List of papers

Paper I:

Courtade, G., Wimmer, R., Røhr, Å. K., Preims, M., Felice, A. K. G., Dimarogona, M., Vaaje-Kolstad, G., Sørlie, M., Sandgren, M., Ludwig, R., Eijsink, V. G. H., Aachmann, F. L. (2016) Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences*, 113, 5922–5927

Paper II:

Courtade, G., Le, S. B., Sætrom, G. I., Brautaset, T., Aachmann, F. L. (2017) A novel expression system for lytic polysaccharide monooxygenases. *Carbohydrate Research*, 448, 212–219.

Paper III:

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. *The Journal of Biological Chemistry*, 293, 13006-13015

Paper IV:

Courtade, G.*, Ciano, L.*, Forsberg, Z., Sørlie, M, Wimmer, R., Eijsink, V. G. H., Walton, P. H., Aachmann, F. L. (2018) An integrated NMR/EPR spectroscopic study into the substrate-enzyme interactions of a chitin-active lytic polysaccharide monooxgyenase. *Manuscript in preparation.*

Paper V:

Courtade, G., Balzer, S., Forsberg, Z., Vaaje-Kolstad, G., Eijsink, V. G. H., Aachmann, F. L. (2015) ^{1}H, ^{13}C, ^{15}N resonance assignment of the chitin-active lytic polysaccharide monooxygenase *Bl*LPMO10A from *Bacillus licheniformis*. *Biomolecular NMR Assignments*, 9, 207–210.

Paper VI:

Courtade, G., Wimmer, R., Dimarogona, M., Sandgren, M., Eijsink, V. G. H., Aachmann, F. L. (2016) Backbone and side-chain ^{1}H, ^{13}C, and ^{15}N chemical shift assignments for the *apo*-form of the lytic polysaccharide monooxygenase *Nc*LPMO9C. *Biomolecular NMR Assignments*, 10, 277–280.

Paper VII:

Courtade, G., Forsberg, Z., Vaaje-Kolstad, G., Eijsink, V. G. H., Aachmann, F. L. (2017) Chemical shift assignments for the *apo*-form of the catalytic domain, the linker region, and the carbohydrate-binding domain of the cellulose-active lytic polysaccharide monooxygenase *Sc*LPMO10C. *Biomolecular NMR Assignments*, 11, 257–264

Paper VIII:

Kitaoku, Y.*, **Courtade, G.***, Petrović, D. M., Fukamizo, T., Eijsink, V. G. H., Aachmann, F. L. (2018) Resonance assignments for the *apo*-form of the cellulose-active lytic polysaccharide monooxygenase *Ta*LPMO9A. *Biomolecular NMR Assignments,* 12, 357-361

*These authors contributed equally to this work.

Table of contents

1. Introduction

1.1 Bioeconomy, chitin and lignocellulose

Reducing our dependence on oil and the need for solutions towards climate change[1] are the driving forces for bioeconomy. Bioeconomy is a vision of society where oil-based goods are replaced by biomass-based goods. However, increasing the production of, for example, bioethanol by using foodstocks as biomass sources results in competition for land and raw materials. In turn this leads to negative impacts such as increased food prices, and deforestation for sugar cane, palm oil and soy cultivation. Therefore, there is a critical need to find alternative sources of biomass[2]. Cellulose is the most abundant form of biomass on the planet and it is available as a raw-material from waste from agricultural, forest and sawmill industries[3]. Cellulose is thus the foremost candidate as a non-food raw material to drive the shift towards a bioeconomy[4].

Cellulose (Figure 1.1) is a plant cell-wall polysaccharide composed of chains of β-1,4-linked glucose units, where each unit is rotated approximately 180° with respect to each other, giving rise to the cellobiose disaccharide, which is the repeating unit in a cellulose chain. Approximately 2000-14000 glucose residues make up each cellulose chain[5]. The hydroxyl groups in cellulose enable the formation of various types of inter- and intra-chain hydrogen bonds. This gives rise to networks of hydrogen bonds (Figure 1.1B) that form crystalline cellulose fibrils with different morphologies[6]. Cellulose Iα and Iβ are the natural and most abundant forms of cellulose. Cellulose Iα is the principal allomorph of cellulose in bacteria and algae[7], whereas cellulose Iβ is predominant in plants[6]. Different kinds of treatments involving alkaline solubilization and recrystallization yield cellulose II and III[6].

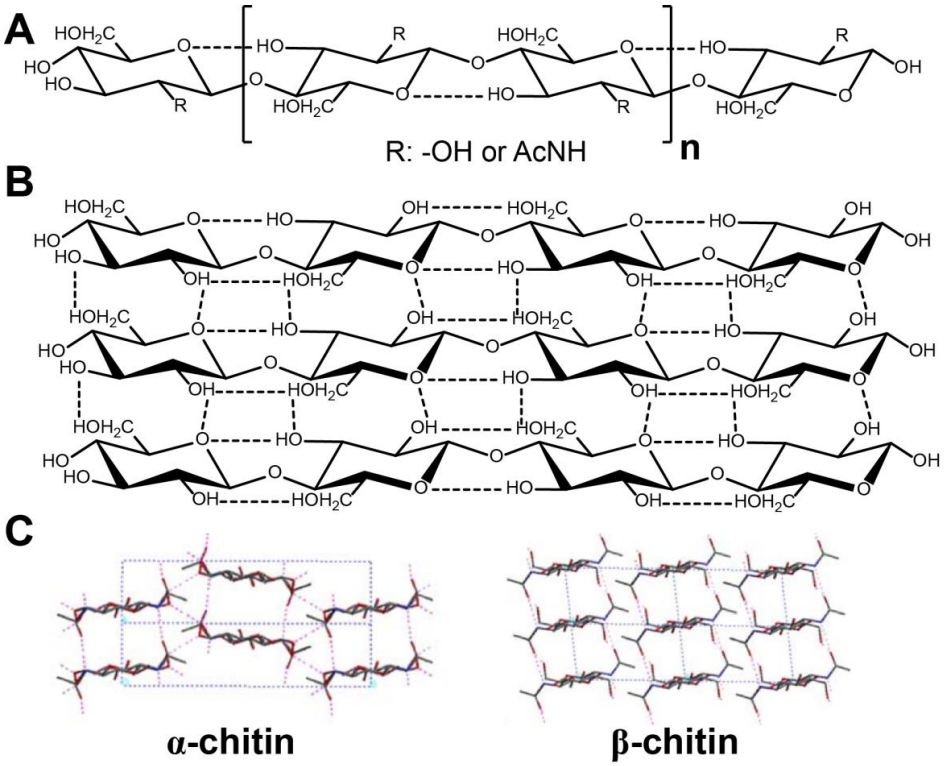

Figure 1.1. Structures of chitin and cellulose. (A) Oligosaccharide chain showing the repeating cellobiose/chitobiose unit and intrachain hydrogen bonding (dashed lines). (B) Intra- and interchain hydrogen bond network in cellulose (dashed lines). (C) Comparison of the crystal packing of α- an β-chitin, resulting in different surface topologies; reprinted with permission from ref. 8.

In plant cell-walls, cellulose is associated with hemicelluloses and lignin (see Figure 1.2A), and the three polymers are collectively referred to as lignocellulose. Hemicelluloses comprise a group of inhomogenous β-(1,4)-linked polysaccharides that include mannans, glucomannans, xylan, xyloglucans, and β-(1,3;1,4)-linked glucans. Flexible hemicelluloses strengthen plant cell-walls through interactions with cellulose fibers and lignin[9]. Lignin is a complex, amorphous polymer of different phenylpropanoid units, which confers strength and structural stability to plants, as well as hydrophobicity to cell-walls to facilitate water transport. Moreover, lignin protects the cellulose core from enzymatic degradation[10].

Analogous to cellulose, chitin is a β-1,4-linked polysaccharide and occurs in the exoskeleton of insects and crustaceans, and in the cell walls of fungi. It is composed of N-acetyl glucosamine (GlcNAc) units that, like in cellulose, are rotated approximately 180° with respect to each other, giving rise to the chitobiose disaccharide, which is the repeating unit in a chitin chain[8]. Individual chitin chains form insoluble crystalline fibrils as a result of inter- and intra-chain stabilization by hydrogen bonds, giving rise to three allomorphs: chitin α, β and γ (Figure 1.1C). In α-chitin, individual chains form antiparallel layers, resulting in dense packing and the most rigid crystalline structure of the three allomorphs[11]. In β-chitin, individual chains are arranged in parallel layers, which results in less dense packing than in α-chitin. γ-chitin is proposed to arise from a mixture of both parallel and antiparallel conformations, but its existence as an independent allomorph remains controversial[8]. Chitin is often found associated with proteins and minerals. Moreover, a percentage of GlcNAc units in chitin may undergo deacetylation to form glucosamine (GlcN). High degrees of deacetylation gives rise to chitosan: a soluble, positively charged polysaccharide at pH values below 6.5[8].

1.2 Degradation of chitin and lignocellulose

Both lignocellulose and chitin are biomass that can be converted into useful biomaterials and platform chemicals through biomass conversion processes. An example of such a process is depicted in Figure 1.2, where the polysaccharide is first harvested from its natural source, mechanically (e.g. milling) and/or chemically pretreated, enzymatically depolymerized, and converted into new products[12]. For cellulose, a common end-product is ethanol: a result of fermentation of glucose, the main product of enzymatic decomposition of cellulose. Ethanol can either be used directly as fuel or as feedstock for the production value-added materials, such as polyethylene[13]. Similarly, while the main commercial product of chitin is N-acetyl glucosamine, it can also been molded into a plastic-like material to manufacture large three-dimensional objects[14].

A bottleneck in the process of cellulose and chitin conversion (Figure 1.2) is the degradation of the dense crystalline polysaccharide matrix into shorter fragments; in the case of cellulose the target degradation products are glucose or cellobiose[10]. This process often requires the use of several types of Carbohydrate-Active enZymes (CAZymes)[15].

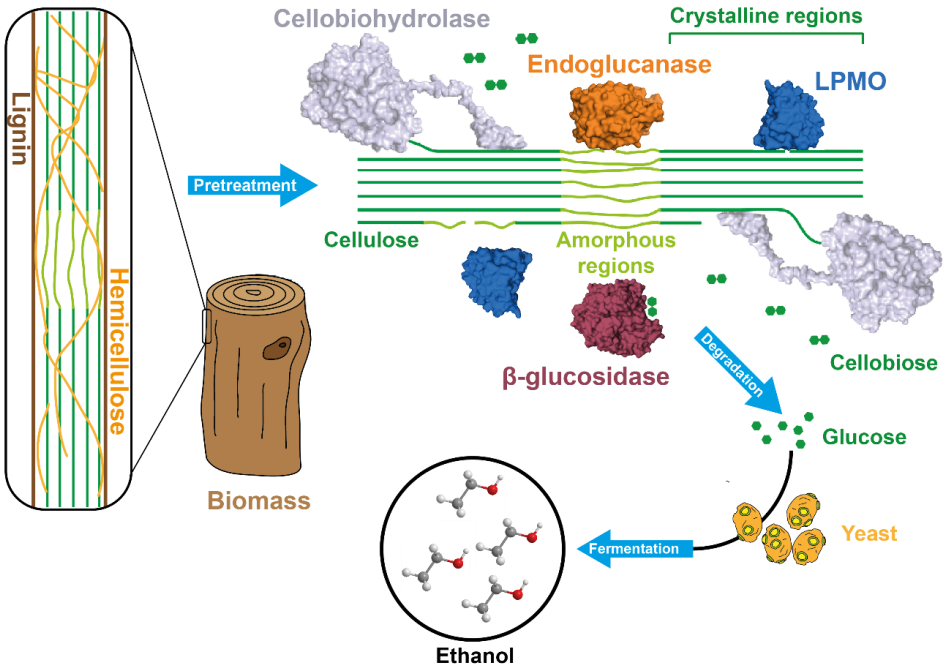

Figure 1.2. Illustrative overview of the lignocellulose degradation process. Biomass is pretreated to remove lignin and hemicelluloses associated with cellulose in the plant cell wall. Crystalline cellulose is shown in green, with a lighter green color to indicate amorphous regions. A combination of cellobiohydrolases (adapted from ref. 16, using PDB IDs: 8CEL[17] and 1CBH[18]), an endoglucanase (PDB ID: 3QR3[19]) and an LPMO (lytic polysaccharide monooxygenase; PDB ID: 4OY7[20]) is used to depolymerize cellulose. Note that the LPMO is the only enzyme that can directly cleave internal glycosidic bonds in crystalline regions. The activity of the enzymes releases cellobiose, which is converted into glucose by a β-glucosidase (PDB ID: 4GXP[21]). Glucose is then fermented by an organism such as *Saccharomyces cerevisiae* to produce ethanol.

Traditionally, enzymatic depolymerization of cellulose relied on the use of enzymes belonging to the glycoside hydrolase (GH) class of CAZymes (www.cazy.org), which hydrolyze glycosidic bonds. Cellulose GHs are collectively referred to as cellulases and include[22] (i) cellobiohydrolases (also known as exoglucanases) that processively depolymerize cellulose starting from chain ends, producing cellobiose or glucose, (ii) endoglucanses that hydrolyze glycosidic bonds at random in amorphous regions on a cellulose surface, and (iii) β-glucosidase that degrades cello-oligosaccharides into glucose. Hydrolysis is however an inherently inefficient method of cellulose depolymerization because water molecules are precluded from penetrating the tightly packed hydrogen bonded network in crystalline cellulose fibrils. Furthermore, open chain ends and amorphous regions seldom occur on a crystalline cellulose fibril. These factors are obstacles to efficient enzymatic degradation[23] of cellulose by GHs and further conversion into fuels and platform chemicals. Due to these inefficiencies, industrial conversion processes of cellulose require high enzyme loads, which, according to a calculation by Johnson[24], account for about 30% of the total costs of producing cellulosic ethanol[24].

Hydrolytic degradation of chitin, like cellulose, requires the use of a set of enzymes with analogous activities, namely chitobiohydrolases ChiA and ChiB, endo-acting ChiC, and GlcNAc-producing chitobiase[25].

1.3 LPMOs – Lytic Polysaccharide MonoOxygenases

In 1950, Reese *et al* proposed the existence of an additional step in the cellulose degradation process, preliminary to cellulose hydrolysis, which they called C_1 and that would have a role in making cellulose more accessible to GHs[26]. Later, in 1974, Eriksson *et al* discovered that oxidative enzymes were involved in cellulose degradation[27]. The discovery of a new class of enzymes known as lytic polysaccharide monooxygenases (LPMOs), about 60 years after the seminal paper by Reese, confirmed the C_1 hypothesis and brought about a paradigm shift regarding the degradation of crystalline polysacharides[28–32]. LPMOs are copper-dependent enzymes that oxidize glycosidic bonds in 1,4-linked polysaccharides; originally their oxidative activity was observed on chitin[29] and cellulose[33–37]. Interestingly, mixtures of cellulases and LPMOs were shown to have much higher cellulose-degradation activity than any of the enzymes by themselves. This synergism can be explained by LPMOs binding and cleaving glycosidic bonds in crystalline regions of polysaccharides. New chain ends generated by LPMO activity lead to decreased crystallinity in the area proximal to the oxidation site, which create new access points for cellulases, overall boosting the activity of the enzyme mixtures[28–30,37–39] (see Figure 1.2). Another mechanism through which LPMOs might enhance the activity of cellulases consists in that cleavage points created by LPMO oxidation could likely promote the dissociation of processive cellobiohydrolases. This would overall enhance their cellulose degradation activity, as low dissociation rates of cellobiohydrolases have been shown to be rate-limiting[40–42].

These initial observations motivated further research on LPMOs and their unique reaction, and led to their inclusion in industrial enzyme cocktails for cellulose degradation[12,43]. To date, LPMOs have been found in bacteria, fungi and insects and overall, LPMO activity has been demonstrated for a variety of 1,4-linked substrates (cellulose[33–37], chitin[29], cellooligosacharides[44,45], xyloglucan[46–48], xylan[49,50], starch[51,52]). LPMOs are classified in the carbohydrate-active enzyme database (CAZy) as auxiliary activity (AA) families AA9[53] (cellulose-oxidizing fungal LPMOs; formerly classified as glycoside hydrolases GH61), AA10[53] (cellulose and chitin oxidizing bacterial LPMOs; formerly classified as carbohydrate-binding modules CBM33), AA11[54] (chitin-oxidizing fungal LPMOs), AA13[51] (starch-oxidizing fungal LPMOs), AA14[49] (xylan-oxidizing

fungal LPMOs), and AA15[55] (cellulose and chitin oxidizing insectile LPMOs). Since no official naming convention for LPMOs exists and different names have been used for the same proteins in literature, it is necessary to specify the LPMO nomenclature used in this thesis. Here, LPMOs are named according to the Latin name of their source organism, their AA family, and a letter corresponding to the LPMO number in that organism. For example, the third cellulose-active LPMO from *Neurospora crassa*, a fungus with 14 LPMOs, is classified in family AA9 and named *Nc*LPMO9C.

1.3.1 Structural features of LPMOs

The first LPMO structures were those of *Sm*LPMO10A[56] (also known as CBP21) and *Hj*LPMO9A[32] (also known as Cel61B). These two LPMOs belong to families AA10 and AA9, respectively, and there is only 16.5% sequence identity between them. From these structures it was clear that LPMOs from different families have overall similar structures, as well as a conserved "histidine brace" active site and conserved surface groups involved in substrate interactions. These similarities have been further confirmed by a comprehensive number of structural studies, which have been reviewed elsewhere[57].

Overall, the catalytic domains of LPMOs are approximately 150-250 residues in length. During translation, LPMOs are produced with an N-terminal signal sequence of approximately 15-30 amino acids (Figure 1.3A), which is proteolytically cleaved when the protein is translocated to the periplasm[29,58]. Mature LPMOs have an N-terminal His. Amino acid numbering in LPMOs, depends on the length of the signal sequence; for example, the N-terminal amino acid in mature *Nc*LPMO9C should be His17. The N-terminus together with the side-chain of the N-terminal His and the side-chain of a more distal His form the "histidine brace" active site of LPMOs. The active site is further shaped by a Tyr or Phe and coordinates one copper ion, which classifies it as a type II copper site[59]. Furthermore, since the active site is highly plastic, it can bind other metal ions such as Zn(II) instead of copper, albeit at lower affinities and producing an inactive form of the enzyme[60]. The N-terminal histidine in LPMOs produced in filamentous fungi carries a τ-nitrogen methylation (see Figure 1.3B)[61].

LPMOs are mostly composed of β-strands, and their core is composed of a β-sandwich fold encompassing two β-sheets, including 8-9 β-strands altogether. The structure is stabilized by hydrophobic residues, as well as by one or two disulfide bridges. The β-sandwich core is decorated with loops and helices, particularly in the first 60-70 amino acids, forming and overall pyramidal shape. The base of the pyramid corresponds to the substrate binding surface and has aromatic residues that mediate interactions and orient the LPMO with the crystalline polysaccharide substrates. The copper active site is located on the center of this substrate binding surface (see Figure 1.3C, D) and when an LPMO binds its substrate, the active site is positioned near the glycosidic bond[57].

While many LPMOs exists solely as a catalytic domain, others are found tethered to carbohydrate binding modules (CBMs) through peptide linkers of various lengths and compositions (see Figure 1.3A and section 1.3.4).

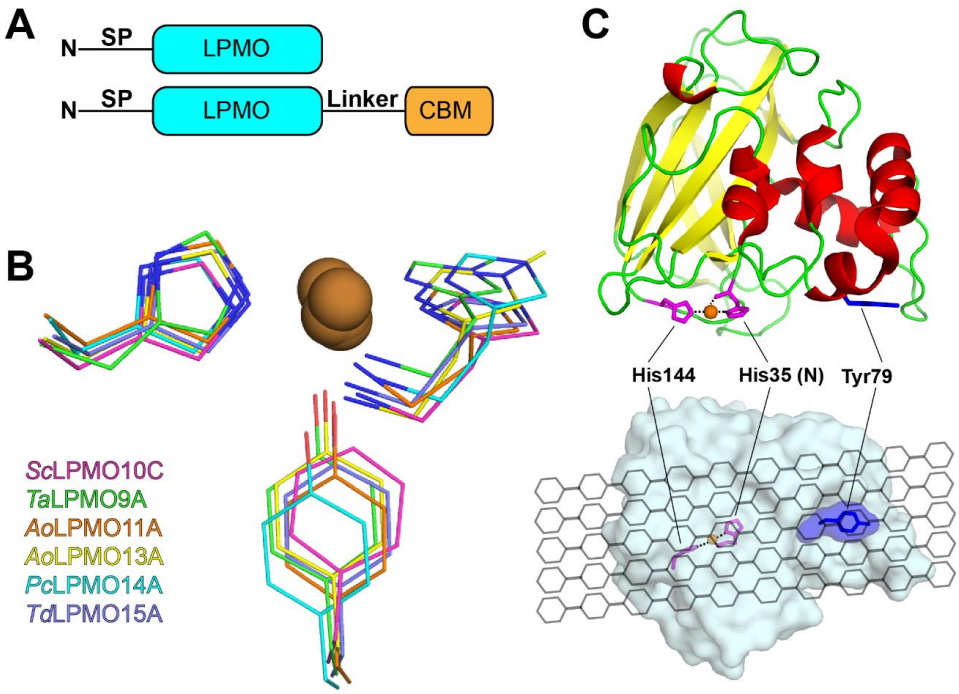

Figure 1.3. Structural features of LPMOs. (A) Overview of a typical LPMO gene including the translocation signal peptide sequence (SP), the LPMO and, in some cases, the presence of a peptide linker and a CBM. (B) Overlay of the histidine brace active site, coordinating copper (brown) of LPMOs from all families: *Sc*LPMO10C (pink; PDB ID: 4OY7[20]), *Ta*LPMO9A (green; PDB ID: 3ZUD[34]) *Ao*LPMO11A (orange; PDB ID: 4MAI[54]), *Ao*LPMO13A (yellow; PDB ID: 4OPB[51]), *Pc*LPMO14A (cyan; PDB ID: 5NO7[49]), *Td*LPMO15A (purple, PDB ID: 5MSZ[55]). Note that *Ta*LPMO9A and *Ao*LPMO13A carry a τ-nitrogen methylation on the N-terminal His. The amino acid in the axial position is Phe for *Sc*LPMO10C and *Td*LPMO15A, and Tyr for the other proteins. (C; top) Overview of the fold of the catalytic domain of *Sc*LPMO10C, showing β-strands (yellow), helices (red) and loops (green), as well as the active site (pink) with a bound copper ion (orange), and Tyr79, which mediates substrate binding. (C; bottom) Surface view of *Sc*LPMO10C overlaid with a model of a cellulose fibril.

1.3.2 Catalytic mechanism of LPMOs

Even though, LPMOs were first thought to be non-catalytic[28], it was promptly shown that the products from LPMO-treated substrates were oxidized on either carbon (C1, C4 or both) in the susceptible glycosidic bond[20,29,62,63]. While, there is currently no consensus on several details of the reaction mechanism, such as the nature of the oxygen species[64–66], there is agreement regarding the overall nature of the reaction (summarized in Figure 1.4). The LPMO reaction involves the reduction of Cu(II) to Cu(I) in the active site by an electron donor that can either be another redox protein such as cellobiose dehydrogenase (CDH)[35,36,67], lignin-derived phenols[68], glucose-methanol-choline oxidases (GMCs)[69] or organic compounds such as ascorbic acid or gallic acid[29,34]. LPMO-bound Cu(I) then binds oxygen (or hydrogen peroxide, see below), which is activated to form a copper-oxygen radical that upon binding of the LPMO to the substrate, abstracts either proton from the scissile bond carbons. This results in the formation of either a lactone or a ketoaldose, depending on whether H1 or H4 is abstracted, respectively. After this initial oxidation, which involves breaking a bond with a binding energy of approximately 95 kcal/mol[27,70,71], an aldonic acid (GlcA) or a geminal diol are formed, respectively, as a consequence of the aqueous environment in which the reaction occurs. The geminal diol has been shown to be at equilibrium with the ketoaldose form[44]. Solubilized products generated by LPMO oxidation have a predominantly even number of sugars. This indicates that LPMOs are active on well-ordered chains in a crystal surface, where only every other glycosidic bond is available to LPMO attack[29,33].

More recently, it has been shown that the real LPMO co-substrate is most likely to be hydrogen peroxide instead of oxygen. Hydrogen peroxide has been shown to not only drive the LPMO reaction in the absence of oxygen, but also promote higher yields of oxidized products. Moreover, product yields have been shown to be in stochiometric ratio with the hydrogen peroxide[65].

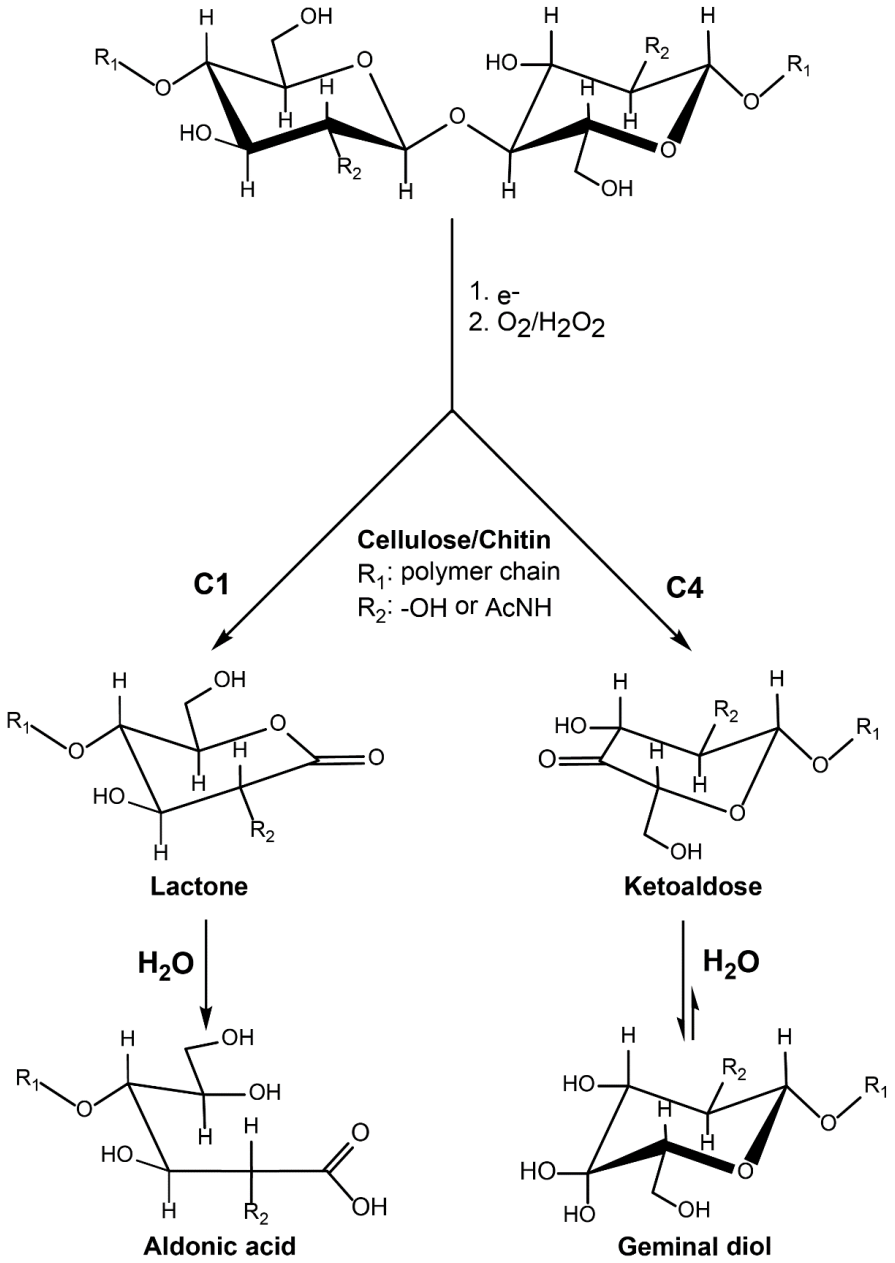

Figure 1.4. LPMO oxidation of chitin/cellulose. The reaction requires the addition of electrons and oxygen/hydrogen peroxide (see text for details). C1 oxidation forms a lactone, and hydration of the carbonyl group yields an aldonic acid as the final product. C4 oxidation forms a ketoaldose, and its subsequent hydration yields a geminal diol. In C4 oxidation, the ketoaldose and its hydrated form are in equilibrium, with preference for the germinal diol. Note that oxidation leads to distortion of the chair conformation of the sugars, which might reduce local crystallinity.

1.3.3 Substrate specificity of LPMOs

Most LPMOs, particularly those belonging to the AA10 family, appear to have narrow substrate specificity, being active on only one kind of substrate such as crystalline chitin or cellulose. Substrate specificity was first demonstrated for *Sm*LPMO10A (also known as CBP21) on β-chitin[29]. Interestingly, even though *Sm*LPMO10A was shown to be less active on the more crystalline α-chitin, its synergy with chitinases ChiA, B and C was observed to increase with the degree of crystallinity of the substrate[72]. The same is also true for LPMOs active on crystalline cellulose, and their synergy with cellulases[20,73]. This is consistent with the idea that LPMOs specifically target crystalline surfaces, where the interaction is primarily mediated by interactions between π-electrons in the aromatic rings on the surface of the LPMO and C-H bonds in the carbohydrate[45,60,74,75].

Some LPMOs, particularly those belonging to the AA9 family (e.g. *Nc*LPMO9C, *Gt*LPMO9A, *Ls*LPMO9A), have been shown to have broad substrate specificity, oxidizing both crystalline cellulose[44,45,47], cellooligosaccharides[21,23], xyloglucan[46] and cellulose-xyloglucan complexes[47]. The molecular basis for LPMO substrate recognition and activity remains, however, insufficiently understood.

1.3.4 LPMO-associated Carbohydrate Binding Modules (CBMs)

Similar to the catalytic domains of LPMOs and other CAZymes, CBMs are divided into families according to similarity of their amino acid sequences. Currently, there are 83 distinct CBM families (www.cazy.org) and these CBMs display significant variation in terms of substrate specificity. CBMs are thought to have three main functions that arise as a consequence of the binding event. They target the "correct" substrate for their attached catalytic domain[76], they create proximity between the catalytic domain and the substrate, and, in certain cases, they have been shown to have a disruptive effect on the polysaccharide, enhancing the degradation ability of the catalytic domain. While extensive reviews of CBMs have been published elsewhere[77], it is of interest to focus on CBMs that are associated with LPMOs, particularly CBM5 and CBM73 in AA10s, and CBM1 and CBM2 in AA9s and AA10s, respectively (see Table 1.1).

Table 1.1. Overview of CBMs commonly associated with LPMOs. Adapted from www.cazy.org[77].

CBM family	Fold type	Approx. sequence length (# of a.a.)	Example of associated LPMO	LPMO substrate
1	Cysteine knot	40	*Nc*LPMO9C	Cellulose, xyloglucan
2	β-sandwich	100	*Sc*LPMO10A	Cellulose
5	Unique	60	*Cj*LPMO10A	Chitin
73	Unknown	65	*Cj*LPMO10A	Chitin

These CBMs belong to "Type A", meaning that they bind to insoluble, crystalline substrates (e.g. chitin and cellulose)[77]. A comparison of the substrate-binding face of the CBMs (see Figure 1.5) reveals that their binding surfaces are flat and that they have a high abundance of aromatic residues. Both of these features are hallmarks of interactions with a crystalline substrate, as aromatic side-chains facilitate interactions to C-H bonds in carbohydrates[74].

For LPMOs, the presence of a CBM significantly enhances LPMO binding to its substrate[78] and results in higher yields of oxidized products, compared to versions of the enzyme comprising only the catalytic LPMO domain[78]. It is however not well understood why some LPMOs have CBMs (in some cases more than one, e.g. *Cj*LPMO10A)[79], while others exist only as a single catalytic domain. Even with the availability of studies on LPMO-associated CBMs, where the presence of CBMs has been correlated with higher yields[79 - 81], the mechanism through which CBMs enhance activity and the interplay between CBMs, linker regions and catalytic domains are not well understood. Moreover, at the start of this thesis, no structures of LPMO-associated CBMs were currently available.

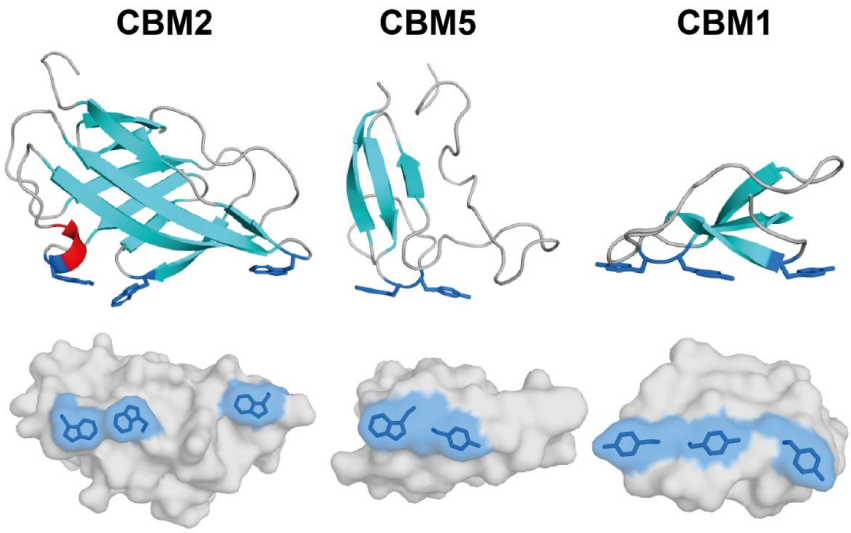

Figure 1.5. Structure comparison of the overall fold and substrate-binding surface of CBMs. Representative structures of CBM2 (PDB ID: 1EXG[82]), CBM5 (PDB ID: 1AIW[83]), and CBM1 (PDB ID: 1CBH[18]). The top panel shows the cartoon representation of the structures (β-strands: cyan, α-helices: red, loops: gray), and the putative substrate binding residues are shown as blue sticks. The bottom panel shows the surface representation of the proteins, viewed from the substrate binding surface, where the putative substrate binding residues are highlighted. These CBMs are associated with other CAZymes, not LPMOs, but are shown here for illustrative purposes. The aromatic amino acids Trp and Tyr are prevalent in this example and are usually conserved, but His, though less conserved, may also occur on the substrate binding surface.

2. Aim and highlights of the thesis

Key aspects of LPMO functionality that remained unexplored and/or poorly understood at the start of the thesis included LPMO structural diversity, the molecular basis of substrate binding, the location of the electron transfer event to the LPMO, the dynamic features of LPMOs, the role of LPMO-associated CBMs, and the effect of copper on LPMO structure and substrate interactions (see Figure 2.1). The main aim was therefore to increase LPMO understanding by providing insights into these key aspects of their functionality by using primarily nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy was the technique of choice because of its unmatched ability to provide both atomic level resolution of protein structure and probe a broad range of time-scales that encompasses substrate binding and protein dynamics, as will be explained in detail in the Experimental strategies section.

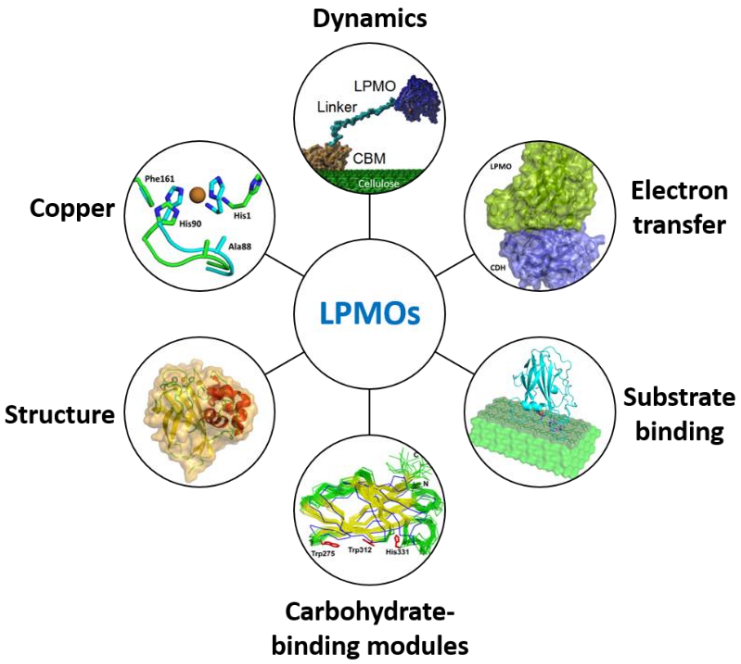

Figure 2.1. Key aspects of LPMO functionality addressed by this thesis.

Structure – At the start of the thesis, a handful of LPMO structures had been determined, primarily by X-ray crystallography, and the structure of one LPMO had been determined by NMR[60]. The contribution from this thesis to understanding structural features of LPMOs has been in the form of elucidating the NMR structure of chitin-active *Bl*LPMO10A, which strengthened the view that LPMO structures are highly conserved (**Paper IV**). The X-ray crystal structures of LPMOs considered in this thesis were available, so efforts were focused on obtaining chemical shift assignments instead. In total, the sequence-specific assignments of four LPMOs, three CBMs and a linker region within a modular LPMO were obtained (**Papers V–VIII**). While the assignments allowed further functional characterization of the proteins, the chemical shifts were also used to estimate the secondary structure content. A comparison of these estimates with the secondary structure elements present in X-ray crystal diffraction structures was used to evaluate whether full structure determination of target LPMOs should be pursued. Overall, there was excellent agreement between the chemical shift based estimates of secondary structure content and the X-ray crystal structures.

Substrate binding – In the beginning of this thesis, only one experimental characterization of LPMO-substrate (chitin) interaction had been performed, by NMR[60]. This was a remarkable feat, since the crystalline nature of LPMO substrates poses a challenge for investigating binding with atomic-resolution. During the course of the thesis, we contributed to broadening the understanding of LPMO-substrate interactions by mapping, with NMR, the substrate binding surface of *Nc*LPMO9C during its interaction with soluble cellohexaose and xyloglucan. Results showed that a larger area on the surface of the LPMO is involved in binding xyloglucan, likely because of the branched architecture of the substrate (**Paper I**). Later in the thesis, we also monitored the interaction of *Sc*LPMO10C and its associated *Sc*CBM2 with insoluble cellulose nanofibrils (**Paper III**), and the interaction of *Bl*LPMO10A with insoluble chitin flakes (**Paper IV**).

Electron transfer – The redox reaction catalyzed by LPMOs requires the transfer of electrons to the copper site. While this is readily observed for small-molecule reductants such as ascorbic acid[29,34], the transfer of electrons from an electron donor protein such as CDH was not well understood at the start of the thesis. It was particularly unclear whether

the electron transfer occurred directly on the copper active site[67], or whether it occurred on a more distal region of the LPMO, followed by subsequent intra-protein electron transfer steps[61]. We clarified this uncertainty in this thesis, by using NMR spectroscopy to probe the interaction between *Nc*LPMO9C, CDH and its isolated heme *b* cytochrome domain (**Paper I**). We observed that addition of full-length CDH or cytochrome perturbed chemical shifts centered around the LPMO's copper-active site, indicating that electron transfer most likely happens directly from the heme group in the cytochrome domain to the copper-site in the LPMO. Since both substrate binding and CDH binding occur on the same place on the LPMO, and substrate binding precludes CDH binding, these observations entail that electron transfer must occur before substrate binding.

Dynamics – Many LPMOs are modular: they consist of catalytic and carbohydrate-binding domains, joined by linker sequences of variable amino acid composition and length. However, despite efforts in understanding structural aspects of individual catalytic domains, there is remarkably little available knowledge about the conformation of modular, linker-containing LPMOs. There is even less knowledge about the role of CBMs and linker regions during LPMO activity. Using NMR, we have studied the conformation and dynamic properties of full-length modular *Sc*LPMO10C (**Paper III**). Results showed that the linker is disordered and extended, creating distance between the CBM and the catalytic domain and allowing these domains to move semi-independently of each other. Accompanying cellulose degradation experiments indicate that these dynamic properties of the linker promote multiple localized oxidation of the substrate by the modular LPMO.

Carbohydrate-binding modules (CBMs) – While the focus of most LPMO investigations has been on the catalytic domain, CBMs associated with LPMOs, and their role in the mode of action of LPMOs, remain largely unexplored. To further understanding on LPMO-tethered CBMs, we characterized the structure and substrate interaction of cellulose-binding domain *Sc*CBM2 from *Sc*LPMO10C (**Paper III**), and the structures of the chitin-binding domains *Cj*CBM5 and *Cj*CBM73 from *Cj*LPMO10A. For the latter, the structure presented in this thesis represents the first structure of a CBM belonging to family 73.

Copper – NMR investigations performed during the course of this thesis involved studying LPMOs in their inactive *apo*-form, in order to avoid the detrimental signal reduction caused by the paramagnetic relaxation enhancement (PRE) effect brought about by the nature of the type II copper site. To verify the extent to which these results and observations could be justified, we investigated the effect of Cu(I) and Cu(II) on the structure of LPMOs by using NMR (**Paper IV**), as well as the effect of Cu(II) on substrate binding by using isothermal titration calorimetry (ITC) (**Paper I**). Results showed that copper only has marginal effects on the structure, limited to rearrangements in the active site histidines, and it does not affect substrate binding significantly.

In addition to these six key aspects, we developed an expression system to reliably produce isotopically enriched LPMOs. This methodological outcome (**Paper II**) is presented in the Experimental strategies section. While the LPMO expression system originally was not an aim in itself, it became a necessary objective that allowed us to use NMR to address the aspects of LPMO functionality outlined in Figure 2.1.

During the courses the thesis, we contributed to two other projects. In the first contribution, we performed an NMR investigation of chitin binding affinity of the CBM14 of human macrophage chitotriosidase, an enzyme involved in innate immunity. Our results contributed to elucidating the substrate recognition and binding mechanisms of CBM14 (Appendix A)[84]. The second contribution consisted in pH titrations of the N-terminal histidine of *Ta*LPMO9A, monitored with NMR, to investigate the way in which His-τ-methylation and Zn-binding affect the pK_a of the N-His. Results showed that neither methylation nor Zn have a significant effect on the N-His pK_a of *Ta*LPMO9A, even though His-τ-methylation likely has a role in protecting the LPMO against self-oxidation (Appendix B)[85].

3. Experimental strategies

Detailed step-by-step methods are provided in each of the **Papers**, therefore this chapter describes the experimental strategies for characterizing the structure and function of LPMOs, and addressing the various key aspects outlined in the previous section. The strategy, presented in Figure 3.1 was applied on several of the target proteins, and can be summarized in five main steps:

1. Production of a pure, ^{13}C and ^{15}N enriched **protein sample**.
2. Recording **NMR spectroscopy** data on the protein sample.
3. Perform **a sequence-specific assignment** of the NMR data to the amino acid sequence.
4. Generate a **structural model** (including iterative NOE assignment, dihedral angle constraint prediction, and structure calculation and refinement).
5. Perform **functional studies** (e.g. interactions with binding partners and dynamics)

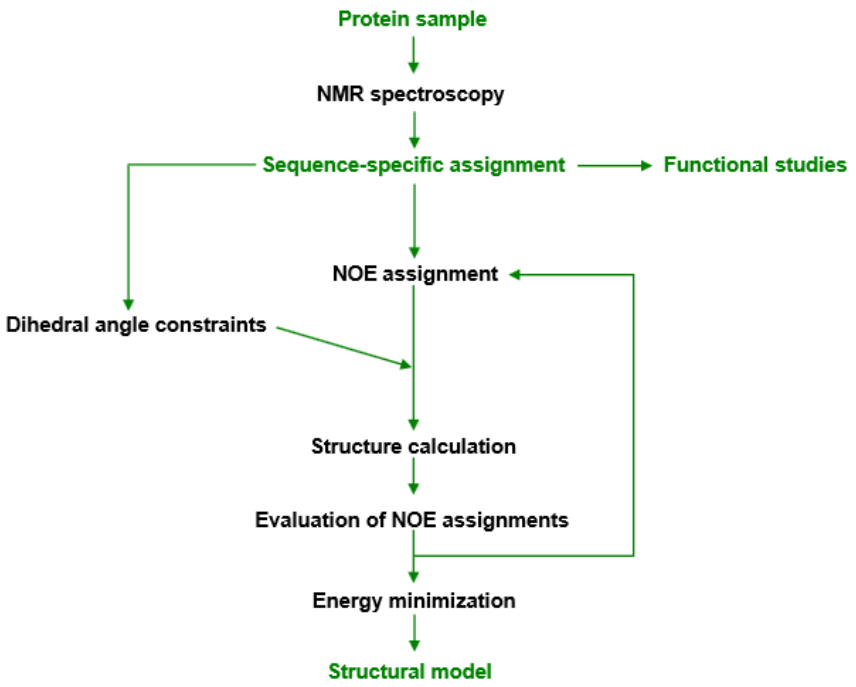

Figure 3.1. Overview of a strategy for structural and functional investigation of a protein by using NMR spectroscopy. Outputs from the strategy are written in green.

3.1 LPMO production and purification

The biochemist Efraim Racker once said: "do not waste pure thoughts on impure proteins"[86]. This statement, when applied to NMR investigations of LPMOs, means that the samples should contain proteins that are pure, correctly folded, isotopically enriched with ^{13}C and ^{15}N, produced in sufficient amounts (>10 mg/L), and correctly processed (i.e. the N-terminal residue should be His and disulfide bridges should be correctly formed).

However, at the time of the start of the thesis, the requirements for LPMO samples for NMR investigations were unmet. Only one type of expression system had been used to produce bacterial LPMOs, namely the T7lac pRSET system relying on "leak" expression in the absence of inducer. This made it difficult to control the level of protein production, leading to low yields and large batch-to-batch variations, especially when using media where ^{13}C-glucose was the sole carbon source. Throughout the thesis, we used different approaches (see Table 2.1), to produce samples in *Escherichia coli* that fulfilled the criteria for NMR investigations. In **Paper II**, we explain why the "leaky" pRSET-based approach was a poor system choice for production of isotopically enriched proteins. We also provide a methodological description of the design and performance evaluation of a new, tunable AA10 LPMO expression vector. This system enabled the stable and reproducible production of isotopically enriched NMR samples for further investigation of *Bl*LPMO10A (**Paper V**) and the catalytic domain of *Sc*LPMO10C (**Paper VII**). Moreover, **Paper VII** describes the production of a modular AA10 LPMO using an N-terminal SUMO-tagged LPMO within a T7lac system, known as the Expresso™ pETite N-His SUMO T7 expression vector (Lucigen).

In addition to developing a new expression system for bacterial LPMOs, investigations of fungal AA9 LPMOs (**Papers I, VI, VIII** and Appendix B) were carried out on isotopically enriched samples produced in *Pichia pastoris* using a protocol described in **Paper VI**. Investigations of CBMs were carried out on isotopically enriched samples produced with His-tags, using a standard T7lac pRSET system, as described in **Paper VII.**

Table 2.1 Comparison of different expression systems used to produce bacterial LPMOs in *E. coli*.

Expression system	General features	Advantages	Disadvantages
T7lac pRSET	• Needs a T7 production strain, such as BL21(DE3), which contains a T7 polymerase in the genomic DNA of the host. • The protein is translocated to the periplasm.	• "Leaky" expression, where there is basal transcription even in the absence of inducer. The whole culture can be grown at the same temperature and no induction is necessary. • Works well in LB media or when using glycerol as the carbon source. • Oxidative conditions in the periplasm facilitate correct disulfide bond formation. • Extraction from the periplasm results in facile purification by affinity chromatography, IEX and SEC.	• If inducer is added, the expression is too high (all-or-none induction), the protein is not correctly translocated, and leads to production of heterogenous mixture of correctly processed and immature protein. • Glucose inhibits the lac operon through the CAP-repressor, therefore media containing ^{13}C-glucose as the sole carbon source result in low (often undetectable) yields.
T7lac SUMO	• Needs a T7 production strain. • The LPMO gene is cloned downstream of a SUMO tag.	• The protein is expressed in the cytoplasm and purified with a His-tag that is included in the SUMO tag. • The system can be induced, overcoming the CAP-repressor problem and thus resulting in higher expression levels.	• SUMO protease must be added to remove the SUMO tag. The resulting, correctly processed LPMO is obtained through another (inverse His-tag) purification step. • Gives lower yields than the other systems. • Maybe not correctly formed disulfide bridges since the protein is produced in the cytoplasm.
Pm/XylS	• The protein is translocated to the periplasm.	• The system is independent of the carbon metabolism of *E. coli*, i.e. glucose has no detrimental effect. • The expression can be fine-controlled, as it is proportional to the amount of inducer. • Homogenous, correctly processed protein. • Oxidative conditions in the periplasm facilitate correct disulfide bond formation. • Extraction from the periplasm results in facile purification by affinity chromatography, IEX and SEC. • Successful protein production at high-cell density cultivations. • Highest LPMO yields reported in literature.	• Unsuccessful production of modular LPMO with a disordered linker region, but only one such target has been tested so far.

3.2 NMR spectroscopy

NMR spectroscopy is a versatile technique that excels at studying the structure, function and dynamics of molecules. NMR measures quantum mechanical properties (e.g. the spin precession) of nuclei within molecules; for proteins the measurable nuclei are typically ^{1}H, ^{13}C and ^{15}N. Reliable measurements of the last two, however, require isotopic enrichment of the protein. The electronic environment, or chemistry, around the nuclei affect their properties and consequently NMR-observable data. NMR spectroscopy can be understood as any other spectroscopy technique, where a system at equilibrium is excited by absorbing energy, and then returns to equilibrium by emitting electromagnetic radiation. The main difference between NMR spectroscopy and for example UV spectroscopy is that the excited state exists for a longer period of time, meaning that one can probe different kinds of phenomena that occur during the life-time of the excited state. Extensive explanations and educational resources about protein NMR are available in the literature[87,88,97–99,89–96] and online[100–104]; here a concise overview of protein NMR is provided based on literature, online resources lectures and personal notes.

In order to observe an NMR signal, nuclei are placed in a strong magnetic field (B_0). The nuclear spin moment ("nuclei" and "spins" are often used interchangeably in NMR terminology) of each individual nucleus in the sample aligns with or against the direction of B_0, resulting in an overall net magnetization in the z-axis, Mz. Nuclei are then excited by means of a radio-frequency (RF) pulse with magnitude B_1 that converts Mz to magnetization in the x,y-plane, Mx,y (see Figure 3.1). For the pulse to be able to excite nuclei and cause this transition, the resonance criterion must be satisfied: the frequency of the RF pulse must match the precession frequency of the nuclei, $\omega_L = \gamma B_0$, where γ is the gyromagnetic ratio of the nucleus. Precessing nuclei in the x,y-plane give rise to a time-domain signal (FID; free induction decay) that decays back to thermal equilibrium in a process known as relaxation. Fourier transformation of the FID gives a frequency-domain depiction of the signal, which is the NMR spectrum.

Integral. At thermal equilibrium, the amount of signal arising from Mz magnetization depends on both the magnetic field strength (which is constant in a modern spectrometer), the magnetic susceptibility of the observed nuclei (manifested through γ), and the concentration of nuclei in the sample. This amount of signal is known as an integral,

which is modulated by several phenomena (e.g. relaxation) and is the reporter of different kinds of NMR experiments, such as relaxation time and NOE measurements (section 3.5.1). Integrals are usually reported as absolute or relative dimensionless values that correspond to the height, area or volume of an NMR signal (see Figure 3.1C)

Chemical shift. Nuclei in a magnetic field naturally precess at frequencies that are determined by the magnetic field strength they experience ($\omega_L = \gamma B_0$), e.g. an isolated ^1H nucleus precesses at 800 MHz in a 18.8 T field. Nuclei in a molecule are located in different chemical (i.e. electronic) environments. Electrons surrounding the nuclei interact with the B_0 field, shielding the nuclei and causing them to experience magnetic fields of slightly different strength. This is the origin of the chemical shift, which is reported as ppm (parts per million) that correct for the relative frequencies in MHz, which vary according to the size of the magnet. For proteins, the secondary and tertiary structure have a large influence on the chemical shift, which means that the chemical shift intrinsically contains structural information (see section 3.4.1).

Couplings. The interaction between nuclei in a molecule through covalent bonds is known as the J-coupling (also called scalar coupling). Interestingly, J-couplings in a protein are different for each type of covalent bond (see Figure 3.4). This feature is exploited when selecting magnetization transfer pathways in multidimensional NMR (see section 3.2.3 for details). These measurable J-coupling interactions result in splitting patterns in spectra (Figure 3.1C) and are also reporters of structural information, as there is a correlation between the size of the J-coupling and dihedral angles in the protein backbone. An example of such correlation is that of Φ-dihedral angles and the magnitude of the ^3J-coupling, described by the Karplus curve in Figure 3.7. Another kind of coupling is the dipolar coupling, which arises from interactions between nuclei through space. Interestingly, the dipolar coupling allows dipolar relaxation to occur, a phenomenon that gives rise to the Nuclear Overhauser Effect (NOE), which enables measurement of distances between nuclei and to probe dynamic processes (see sections 3.4.2 and 3.5.1 for details).

Relaxation. The decay of the FID signal can be described by two parameters: T_1 and T_2 relaxation times. T_1 is the longitudinal (or spin-lattice) relaxation time and it describes the

time it takes for Mz to recover after perturbation by an RF pulse, i.e. for the net magnetization to return to the z-axis (Figure 3.1B). T_2 is the transverse (or spin-spin) relaxation time, describing the time it takes for the Mx,y magnetization to decay, i.e. to lose phase coherence. T_2 decreases with increasing molecular size, and has clear effects on the appearance of spectra such as the linewidth, which is proportional to $1/T_2$. Relaxation times are affected by several dynamic processes (chemical shift anisotropy, spin-rotation, paramagnetic relaxation, quadrupolar relaxation, scalar relaxation and dipolar relaxation). Note that while it is always true that T_1 is larger or equal to T_2, for proteins of 10-25 kDa at room temperature, T_1 is always much larger than T_2.

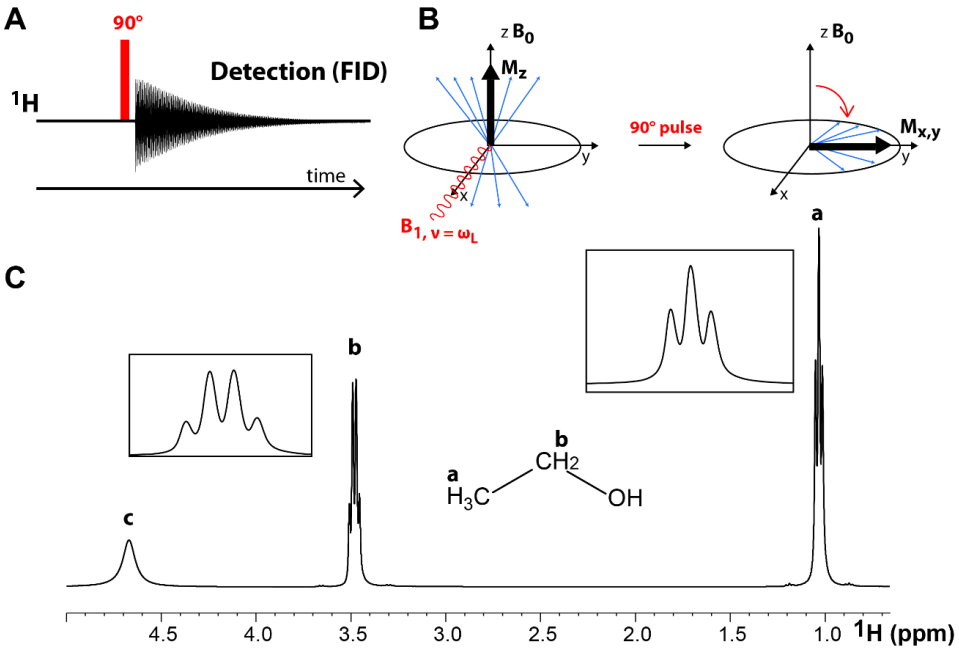

Figure 3.1. Representation of a 1D ^{1}H NMR experiment. (A) Pulse scheme depicting a 90° pulse (red) on ^{1}H resulting in an FID (NMR signal). The same experiment is shown in vector representation in (B), where the net longitudinal magnetization, **Mz** (thick black arrow), composed of individual spins (thin blue arrows) is converted into transversal magnetization, **Mx,y**, by a 90° RF pulse with field B_1 (red) and the resonance criterion $\nu = \omega_L$, where the frequency of the RF pulse (ν) matches the spin precession frequency, ω_L (also called Larmor frequency). (C) ^{1}H spectrum of ethanol, resulting from an experiment like the one shown in (A). The spectrum displays different integrals for (**b**) two and (**a**) three protons, different chemical shifts for (**a**) methyl, (**b**) methylene-alcohol and (**c**) water protons, and different splitting due to J-couplings giving rise to a (**a**) triplet (splitting= 2n +1, n: number of neighboring protons), (**b**) quartet and (**c**) singlet. Noteworthy on this spectrum is that the water protons are in exchange with the -OH group, giving a broad signal at 4.7 ppm. The spectrum was recorded with a 800 MHz Bruker Ascend spectrometer, on a sample of 50% ethanol in 90:10 H_2O:D_2O, at 298.1 K.

3.2.1 Water suppression

The ethanol spectrum in Figure 3.1C is well-resolved and its signals can be easily identified and interpreted because ethanol is a small molecule with few protons. Proteins, on the other hand, give rise to complex 1D spectra with significant overlap of signals (Figure 3.2C) and a predominant water signal (in a 500 µL sample containing 1 mM protein, the water signal is approximately 100000 times stronger), which must be attenuated from NMR spectra by means of water suppression schemes within the pulse sequence. Water suppression can be accomplished by selectively dephasing the water resonance, which is achieved by a combination of selective pulses and gradients.

A pulse sequence such as the one shown in Figure 3.1A uses a broadband rectangular pulse, meaning that the 90° pulse excites all ^{1}H frequencies in the sample. A selective pulse excites a narrow range of frequencies and can thus be used to select specific frequencies (i.e. specific type of nuclei) within the sample, such as water. This can be achieved by using shaped pulses that are defined by a Gaussian or sinc function.

In a perfectly homogenous sample, all nuclei experience the same magnetic field. A gradient introduces a position-dependent magnetic field in one of the sample axes. For example, in Figure 3.2B, a linear gradient along the z-axis after a broadband 90° pulse results in spins located at the top of the NMR sample to experience a slightly stronger magnetic field than the spins located at the bottom of the sample. The difference in magnetic field at each slice causes spins to precess at different frequencies, dephasing them, effectively eliminating the net Mx,y magnetization. Then, a selective 90° pulse is applied on water, followed by a 180° pulse on all spins, and a second selective 90° pulse on water. After a second, identical gradient, resonances that were flipped by the 180° pulse are refocused and give rise to an NMR signal, while the water resonances that experienced a total of 360° (equivalent to no change at all) remain dephased and give no signal. This water suppression scheme has different variants and implementations, two common ones are WATERGATE and excitation sculpting.

Because of water suppression schemes, protein signals can be appreciated in a ^{1}H spectrum like the one shown in Figure 3.2C. Overall, the distribution of signals across a broad frequency range that usually spans -2 to 14 ppm for proteins without a paramagnetic

center (-1 to 11 ppm in Figure 3.2C), as well as overall uniform integrals, indicate that the protein is folded and structured. However, the overlap of ^{1}H signals precludes the extraction of much more information from such a 1D spectrum.

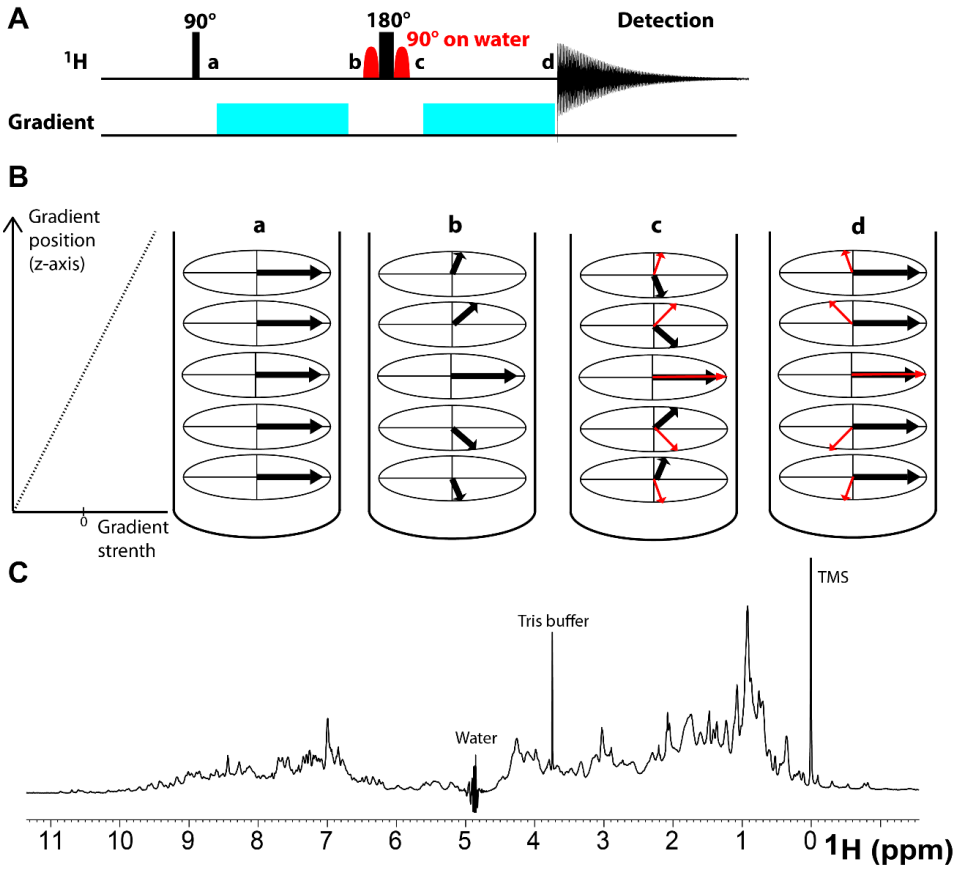

Figure 3.2. Water suppression. (A) Pulse scheme showing the use of selective pulses (red) on water, and gradients (cyan) to achieve water suppression in a protein sample. (B) Shows the effect of a 90° pulse on the spins in five slices in an NMR tube. At time point (**a**), the spin magnetization is converted to Mx,y. After the gradient (**b**), spins in each slice have been shifted proportionally to gradient strength. The combined effect of the selective pulses on water and the 180° pulse on all signals (**c**) rotates all spins, except water (red arrows) 180°. After a second gradient (**d**), the spins are shifted with the same magnitude, but in the opposite direction (because of the 180° flip), giving rise to a signal. Water spins remain dephased since the total 360° pulse angle they experienced means that they are in practice untouched. (C) ^{1}H spectrum of the chitin-active *Bl*LPMO10A, resulting from an experiment like the one shown in (A), where the protein signals can be appreciated, and there is only a minimal residual water signal at ~4.7 ppm. Tris buffer impurities at ~3.8 ppm, and the reference compound trimethylsilyl propionic acid (TSP) used to define 0 ppm are labeled. The spectrum was recorded with a 800 MHz Bruker Ascend spectrometer, on a sample of ~0.2 mM protein in 90:10 H_2O:D_2O, at 298.1 K and pH 5.5.

3.2.2 Two-dimensional NMR

NMR solves the overlap problem by adding more dimensions to the spectrum. Extra dimensions can either be homonuclear (the same nucleus) or heteronuclear (different nuclei). Early 2D NMR spectra relied on adding an extra ^{1}H dimension and observing homonuclear ^{1}H-^{1}H correlations. This strategy can still be applied to small proteins (up to 10 kDa), by using 2D experiments such as 2D COSY, TOCSY and NOESY, but in the case of proteins of the size of LPMOs (15-25 kDa), 2D spectra based on homonuclear correlations are usually too crowded and therefore the strategy is to use heteronuclear correlation to reduce overlap. One obvious correlation that can be exploited is the ^{1}H-^{15}N correlation, present in every amide bond (except that of prolines) in the amino acid backbone of the protein. In order to probe this correlation, the protein must be enriched with ^{15}N-isotopes.

Figure 3.3A describes the overall pulse scheme for this kind of 2D NMR experiment, which is composed of four main segments: preparation, evolution, mixing and detection. The preparation segment (Figure 3.3A; green) involves an interexperiment delay, **d1**, where spins are allowed to relax to thermal equilibrium. During the mixing segment (Figure 3.3A; yellow), like in 1D experiments, Mx,y magnetization is created in ^{1}H by a 90° pulse. This ^{1}H magnetization is mixed with ^{15}N nuclei through a transfer scheme (INEPT; see Appendix C) which is highly selective, enabling magnetization transfer only from ^{1}H that are covalently bonded to ^{15}N. This is achieved by setting the $\boldsymbol{\tau}$ delay proportional to $1/(4J_{HN})$, where J_{HN} is the coupling constant between amide H^{N} and N. Following this transfer, the magnetization is "frequency labeled" with the chemical shift of the ^{15}N nuclei in a step known as **t1** evolution. In this segment, the **t1** delay varies so that the resulting signal is time-dependent with **t1**. This concept applies to all multidimensional NMR experiments and is illustrated in Figure 3.3B, where the intensity of a ^{1}H frequency signal is modulated by the time-varying signal of ^{15}N. Then, the ^{15}N magnetization is mixed back to ^{1}H (Figure 3.3A; orange), by a reverse version of the same INEPT-type transfer. Here, different water suppression schemes may be implemented, before detection of the FID signal. The resulting spectrum, known as a ^{15}N-HSQC (heteronuclear single quantum coherence) spectrum, is shown in Figure 3.3C.

While ^{15}N-HSQC is used as an example here, ^{13}C-HSQC follows the same principle. An HSQC can also be called a fingerprint spectrum, which depicts an "NMR map" of the protein with an altogether small number of signals. HSQC spectra are therefore the starting point for multidimensional NMR spectra and are also the basis for functional NMR investigations; for example, ligand binding can be monitored by signal perturbations in an HSQC upon addition of ligand. However, a prerequisite for using the ^{15}N-HSQC for these purposes requires that the signals are assigned to their respective H^N, N atom pairs in the protein sequence.

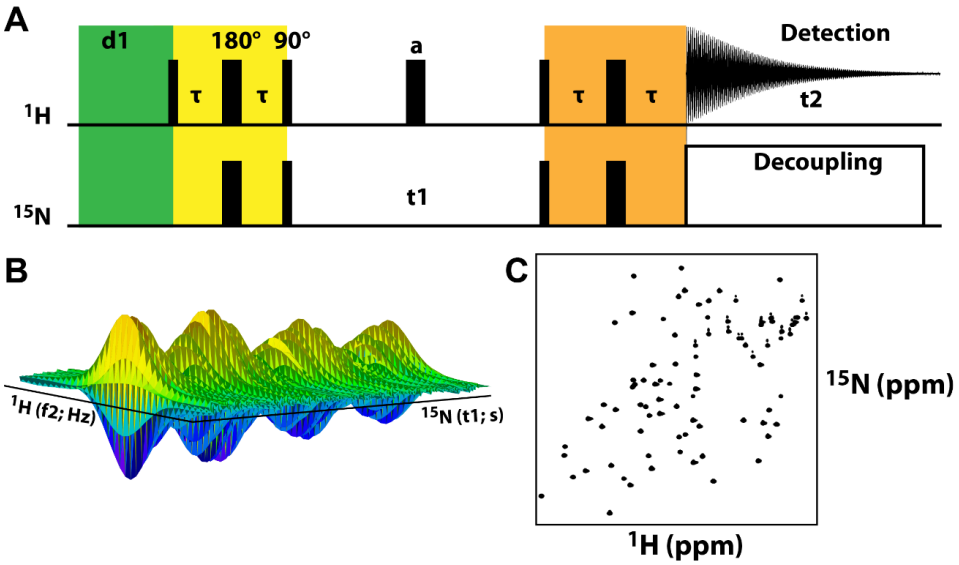

Figure 3.3. Two-dimensional NMR. (A) ^{15}N-HSQC pulse scheme; in summary H^N → N (**t1**) → H^N (**t2**) transferred through scalar coupling evolution during the time delay $\tau = 1/(4^1J_{NH})$. The 180° pulse **a** eliminates scalar couplings between H^N and N. **t2** corresponds to the acquisition time used for detection. Note that while the signal is detected in the ^{1}H channel, there's a decoupling scheme on the ^{15}N channel to reduce the splitting through J-couplings and associated line-broadening. (B) Modulation of ^{1}H frequency by the variable ^{15}N **t1** evolution. (C) Fourier transformation of the time-domain signals in the ^{1}H and ^{15}N (**t2**, **t1**) dimensions results in correlations between both frequency dimensions (**f2**, **f1**), i.e. a ^{15}N-HSQC spectrum. The ^{15}N-HSQC spectrum in (C) was recorded on a sample of ^{13}C and ^{15}N enriched chitin-binding domain *Cj*CBM5, tethered to *Cj*LPMO10A, with a 800 MHz Bruker Ascend spectrometer, on a sample of ~0.5 mM protein in 90:10 H_2O:D_2O, at 298.1 K and pH 5.5. The signal in (B) was extracted from the same ^{15}N-HSQC spectrum.

3.2.3 Multidimensional NMR

Signal overlap in proteins in size of 10-25 kDa precludes sequence-specific assignment through the interpretation of homonuclear 2D spectra. A solution is to add a 3rd (or more) dimensions to the NMR spectrum. This is achieved through NMR spectra that correlate ^{1}H-^{15}N with different, specific ^{13}C atoms. This is possible because pulse sequences select specific J-couplings (see Figure 3.4), which become specific magnetization transfer pathways.

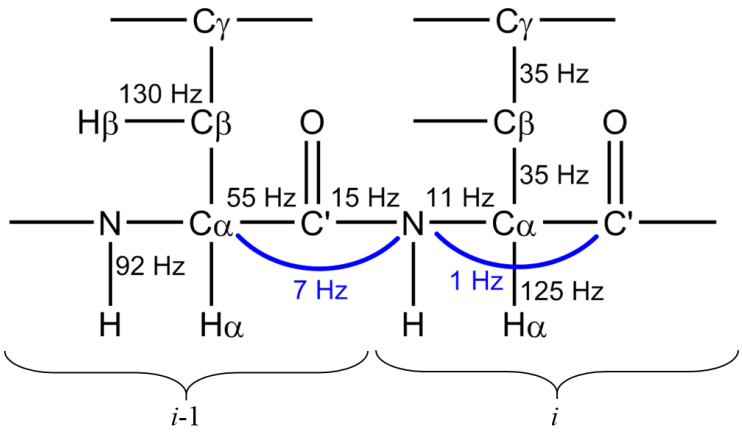

Figure 3.4. J-couplings in proteins. ^{1}J-couplings (black) and ^{2}J-couplings (blue) present in two residues in a protein, where residue *i*-1 is predecessor to residue *i*.

These pathways may correlate H^N, N to other nuclei in the same spin-system (denoted *i*) or to nuclei in the preceding spin-system (denoted *i*-1). A spin-system, in the case of proteins, corresponds to all resonances within the same amino acid residue that are scalar coupled to H^N, N. Typically, J-couplings allow correlation of H^N, N to be made with only one type of ^{13}C atom, such as C^α or C'. Moreover, experiments have been designed so that they contain complementary information, with one experiment showing correlations from H^N, N to only $C^\alpha_{(i-1)}$ (the preceding C^α, in spin-system/residue *i*-1), and the other experiment to both $C^\alpha_{(i)}$ (its own C^α, in spin-system/residue *i*) and $C^\alpha_{(i-1)}$. For example, the HNCA experiment (Figure 3.5) correlates amide H^N, N with $C^\alpha_{(i)}$ and $C^\alpha_{(i-1)}$ by setting the **τ2** delay proportional to 1/9 Hz: the reciprocal of the average between the ^{1}J-coupling between N and C^α (11 Hz) and the ^{2}J-coupling between N and $C^\alpha_{(i-1)}$ (7 Hz). The resulting 3D spectrum has two dimensions that correspond to ^{1}H-^{15}N correlations, and a third

dimension that contains correlations to $C^{\alpha}_{(i)}$ and $C^{\alpha}_{(i-1)}$. The complementary experiment is HN(CO)CA, which correlates H^N, N with $C^{\alpha}_{(i-1)}$ by using two transfer steps, first from N to $C'_{(i-1)}$ (15 Hz) and then from $C'_{(i-1)}$ to $C^{\alpha}_{(i-1)}$. The complementarity of the information in this pair of spectra enables assembly of the sequence-specific assignment.

In addition to these spectra, other pairs that can be used for backbone assignment are HNCO/HN(CA)CO, HN(CO)CACB/HNCACB, and HN(COCACB)HAHB/ HN(CACB)HAHB. All these pairs contain complementary *i*-1/*i*, *i*-1 information for the observed spins. After backbone assignment, side-chains can be assigned through TOCSY-type experiments (e.g. HCCH-TOCSY), which are also 3D experiments that correlate ^1H,^{13}C pairs (two dimensions) in a ^{13}C-HSQC with all side-chain ^1H-^{13}C resonances (the third dimension) in the same spin-system.

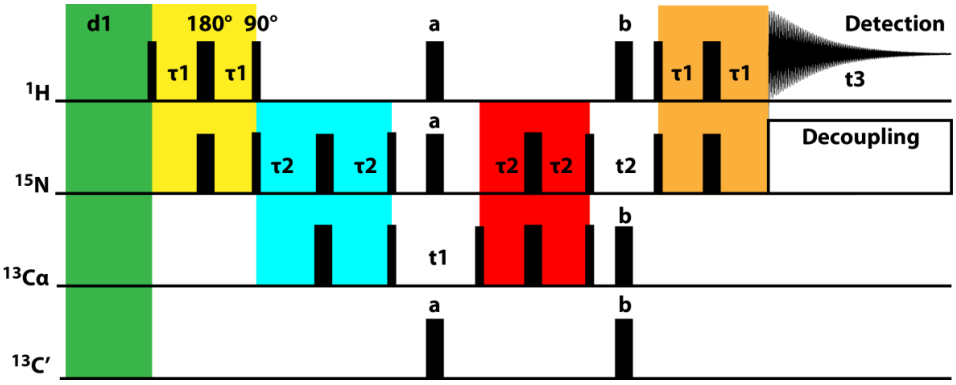

Figure 3.5. HNCA pulse scheme. In summary H^N → N → Cα (**t1**) → N (**t2**) → H^N (**t3**). **t3** corresponds to the acquisition time used for detection. The steps, like in 2D, are preparation (green), INEPT transfer to achieve mixing between ^1H and ^{15}N (yellow; $\boldsymbol{\tau 1} \propto 1/{}^1J_{HN}$), mixing between ^{15}N and ^{13}C (cyan; $\boldsymbol{\tau 2} \propto 1/{}^1J_{NC\alpha}$), **t1** evolution in ^{13}C, mixing between ^{13}C and ^{15}N (red), **t2** evolution in ^{15}N, and mixing between ^{15}N and ^1H (orange) before detection. The 180° pulses **a** and **b** eliminate scalar couplings during chemical shift evolution of C^{α} and N, respectively.

3.3 Sequence-specific assignment

The assignment of ^{1}H, ^{15}N and ^{13}C chemical shifts to specific nuclei and amino acids in the protein is a key step that enables further interpretation of NMR data for structural and functional characterization. The sequence-specific assignment is a bottleneck of any protein NMR investigation, as it requires solving a huge puzzle with several dimensions and thousands of identical-looking pieces. All the proteins in this thesis were assigned in a systematic fashion in a computer-assisted manual approach by using CARA. The assignments have been published in **Papers V–VIII**, which also describe the specific set of experiments used for each protein. Here, an overview of the assignment strategy is explained, exemplified by using the experiment pair HNCA and HN(CO)CA.

Chemical shift patterns, in particular those of ^{13}C, can be readily used to identify specific amino acids in the sequence. Furthermore, the complementary information provided by 3D experiments allows assessment of the connectivity of amino acids in terms of *i*, *i*-1. In the example shown in Figure 3.6, there are five slices of an HN(CO)CA spectrum. Three of these slices show signals $C^{\alpha}_{(i-1)}$ chemical shifts at ~45 ppm. Only glycine residues, due to their unique $C^{\alpha}_{(i)}$ architecture could give rise to these signals. The HNCA experiment confirms that in fact there are four Gly in the segment, connected in a GGXXGG pattern. This is indeed a unique pattern in the amino acid sequence of the LPMO at hand (*Bl*LPMO10A), and after further inspection of all resonances in HNCA, HN(CO)CA and other spectra, the chemical shifts in this segment were assigned to G55-G56-L57-F58-G59-G60.

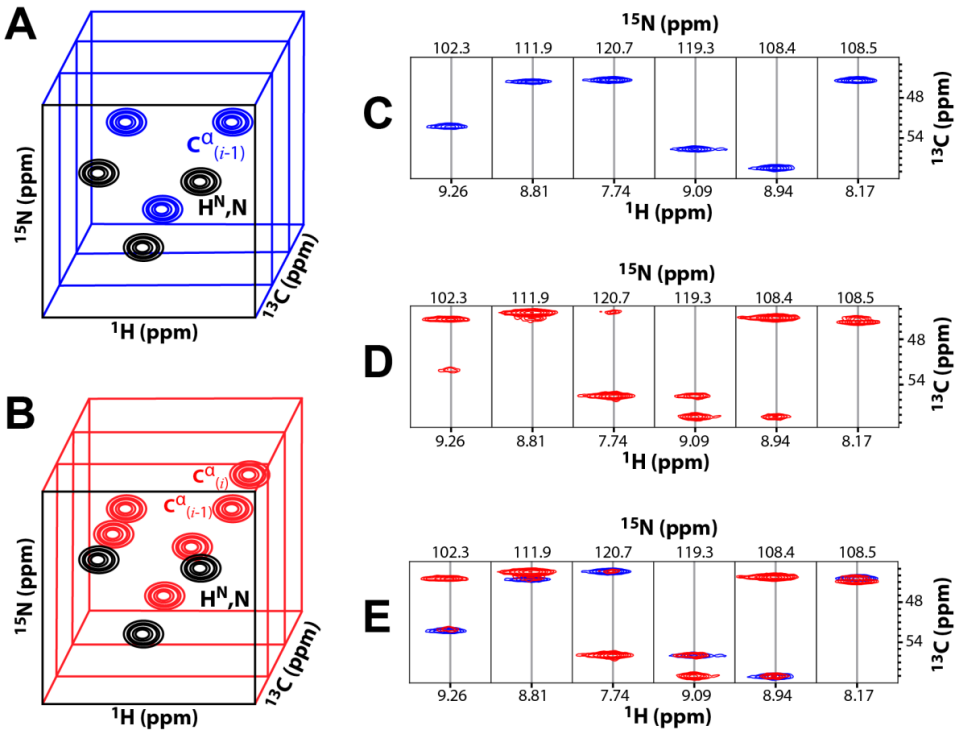

Figure 3.6. Sequence-specific assignment. (A) A representation of a 3D HN(CO)CA spectrum with three signals. The H^N, N resonances are shown in black and the $C^{\alpha}_{(i-1)}$ correlation from each of those resonances are shown in blue. (B) A representation of a 3D HNCA spectrum with three signals. The H^N, N resonances are shown in black and the $C^{\alpha}_{(i)}$/$C^{\alpha}_{(i-1)}$ correlations from each of those resonances are shown in red. The spectra are usually visualized by taking slices through each H^N, N peak, as shown for (C) HN(CO)CA, (D) HNCA and (E), which is an overlap of (C) and (D). Note that in (D), the $C^{\alpha}_{(i-1)}$ peak has a lower intensity than the $C^{\alpha}_{(i)}$. The HN(CO)CA and HNCA spectra were recorded on a sample of ^{13}C and ^{15}N enriched *Bl*LPMO10A with a 600 MHz Bruker Avance III spectrometer, on a sample of ~0.2 mM protein in 90:10 H_2O:D_2O, at 298.1 K and pH 5.5.

3.4 Structural determination

The resonance assignment is the basis for structure determination of proteins by NMR. This is because it can be used to (i) determine or measure dihedral angle constraints, which define the secondary structure of a protein, and (ii) the NOE assignment, which allows distances between ^{1}H nuclei to be estimated from the intensity of cross-peaks in a NOESY-type experiment (see section 3.4.2). Additionally, it is important to highlight that the chemical shift assignment contains intrinsic structural information.

Chemical shifts arise from the electronic environment around nuclear spins. In a small molecule or amino acid residue, the major contributor to chemical shift is the presence of electron-withdrawing (e.g. -OH) and electron-donating groups (e.g. $-CH_3$). In particular, ^{13}C shifts are good reporters of chemical environment, because they are to a lesser extent affected by environmental factors like temperature and pH, than ^{1}H shifts. For example, the C^{β} methyl carbon in Ala has a typical chemical shift of ~20 ppm, in contrast to the OH-bound C^{β} in Ser, with a chemical shift of ~65 ppm.

3.4.1 Dihedral angles

In addition, chemical shift in proteins are affected by secondary and tertiary structure. Secondary structure arises from variations in the backbone dihedral angles (Figure 3.7C) of the amino acid backbone. The relationship between secondary structure and chemical shifts is such that clear patterns have been found. In α-helices, C^{α} shifts are consistently upshifted with respect to random coil shifts, while C^{β} shifts are downshifted (see Figure 3.7D). The pattern is the exact opposite in β-strands. These relationships have been implemented in a number of software, such as TALOS-N and its predecessors, which output per-residue dihedral angle (and secondary structure) estimates, using the chemical shift assignment as an input. Dihedral angles can also be measured experimentally, owing to their effect on J-couplings. This is illustrated in the Karplus curve in Figure 3.7A. Dihedral angles can be used directly as a structural constraint inputs for structure calculation.

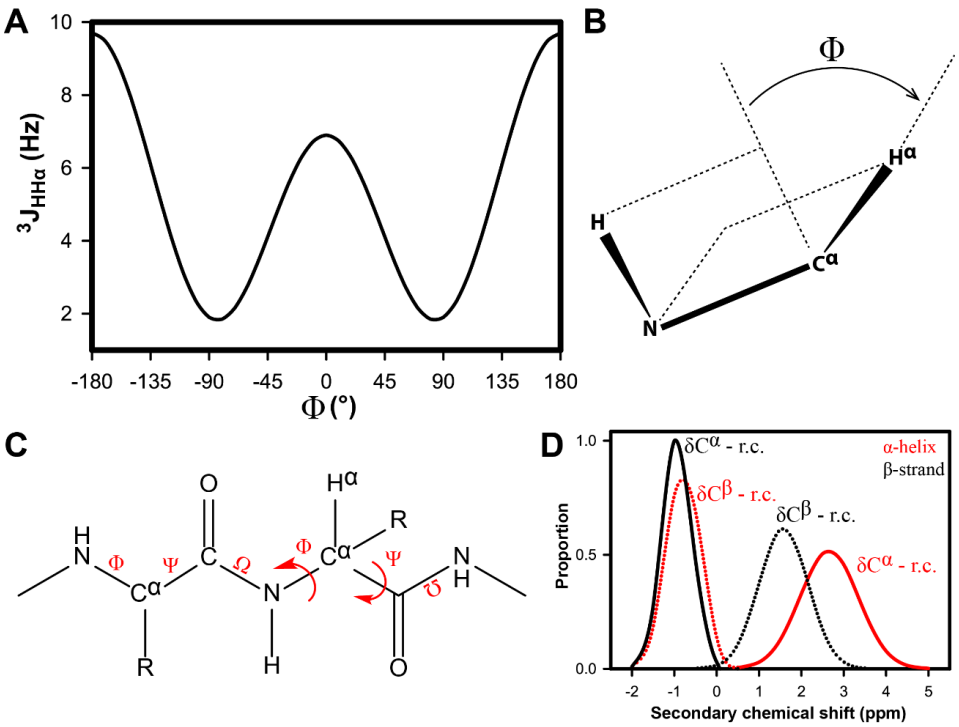

Figure 3.7. Dihedral angles. (A) Karplus curve ($J(\Phi) = A\cos^2\Phi + B\cos\Phi + C$; A=6.4, B=-1.4, C=1.9) showing the relationship between $^3J_{HH\alpha}$ and (B) the Φ-dihedral angle. (C) Dihedral angles in a polypeptide backbone. (D) Secondary chemical shift distributions of C^α (solid lines) and C^β (dotted lines) chemical shifts minus random coil (r.c.) shifts for amino acids in α-helices (red) and β-strands (black). The average and standard deviation shifts are for α-helices (i) δC^α-r.c.=2.66 ± 0.69, (ii) δC^β -r.c.=-0.93 ± 0.35; for β-strands (iii) δC^α-r.c.=-0.78 ± 0.41, (iv) δC^β -r.c.=1.58 ± 0.58. The plot was made with data from RefDB[96] (refdb.wishartlab.com/stat.html).

3.4.2 NOE assignment

Elucidating the tertiary structure of a protein is usually the main goal of any structural determination investigation, as it is the three-dimensional structure that often determines the function of an enzyme. Tertiary structure has a non-negligible effect on chemical shifts, as protein folding and packing of amino acid side-chains leads to interactions between methyl groups and aromatic rings. π-electrons in the rings have significant effects on the chemical shifts of nearby nuclei, by means of the ring current effect, which depends on both distance and orientation to the aromatic ring. A well-known figure depicting this effect is reproduced in Figure 3.8. However, despite software (e.g. CS-ROSETTA) that correlate chemical shifts with tertiary structure, other types of data are required to accurately determine the tertiary structure of a protein by NMR.

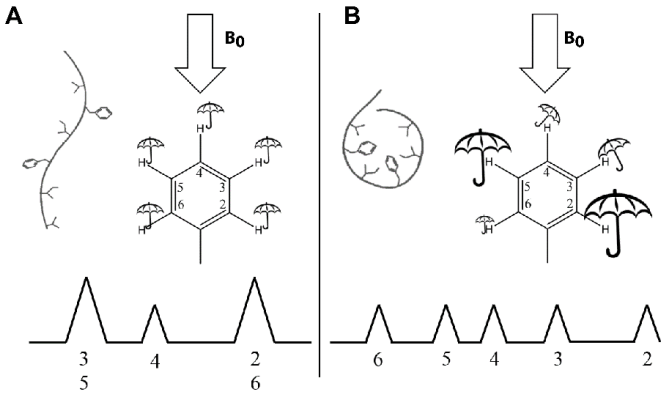

Figure 3.8. Tertiary structure effect on chemical shifts. (A) An unfolded polypeptide chain behaves as a random coil: the side-chains are free to rotate and therefore the protons in e.g. a Phe side-chain experience the same magnetic field strength (B_0; arrow), represented as umbrellas of the same size. The chemical shift pattern (bottom) only depends on the structure of the side-chain (protons 3 and 5 are equivalent, as are 2 and 6). (B) In a folded protein, side-chain rotation is restricted, and the specific electron distribution around each proton arises from tertiary structure. This leads to different shielding of B_0, represented by umbrellas of different sizes. The phenomenon manifests itself in the chemical shift pattern, giving one distinct signal for each proton. The figure is adapted from ref. 88.

The most important type of structural data can be derived from NOE cross-peak intensities (also called NOEs for short). NOEs are the result of the nuclear Overhauser effect (Figure 3.9), which arises when two spins, denoted *S* and *I*, are within ~ 5 Å from each other in space. The interaction between the magnetic dipole moments of each nucleus gives rise to a dipolar coupling between the spins. Since each of the two spins can be in two states (α='spin-up' and β='spin-down'), the dipolar coupling between the spins results in four distinct energy levels (see Figure 3.9A). At equilibrium, a statistically larger number of spins populates low energy levels (α) than high energy levels (β). When spin *S* is excited by a radiofrequency pulse, the equilibrium populations of the energy levels change. Following this perturbation, *S* tries to return to equilibrium through longitudinal relaxation (see section 3.5.1) depicted by the single-quantum (net change of one spin; W_1) transition in Figure 3.9A. However, because *S* has a dipolar coupling with *I*, there are also two alternative relaxation pathways to return to equilibrium: the double quantum (net change of two spins; W_2) and zero quantum (net change of zero spins; W_0) transitions. Relaxation through either of these two transitions involves a magnetization transfer from *S* to *I* through the dipolar coupling. It is the relaxation of S through the W_2 and W_0 transitions that leads to NOE cross-peaks, the intensities of which are proportional to $1/\mathbf{r}^6$, where **r** is the distance between the dipolar coupled spins (Figure 3.9B). In small molecules (Mw < 1000 Da), dipolar relaxation occurs through the W_2 transition, while in proteins (Mw > 1000 Da), dipolar relaxation occurs through the W_0 transition.

NOE assignments for medium-to-large proteins are performed by recording multidimensional spectra, where two dimensions correspond to an HSQC (either ^{13}C or ^{15}N) and the third dimension contains peaks at chemical shifts of ^1H nuclei < 5 Å away from the amide H^N, in the case of the ^{15}N-NOESY-HSQC. These NOE peaks can be identified thanks to the chemical shift assignment, and integration of the peaks allows ^1H-^1H distances to be determined and incorporated as input constraints in the structure calculation.

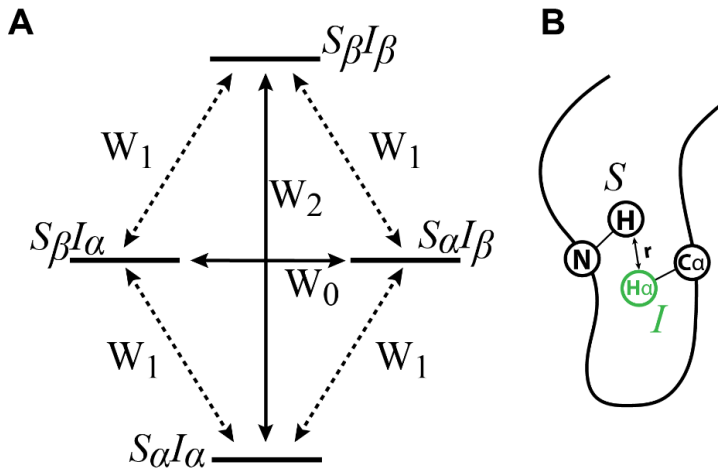

Figure 3.9. Nuclear Overhauser Effect. (A) Energy levels and relaxation pathways (W_1, W_2, W_0) for two dipolar coupled spins *S* and *I*. The W_1 transitions are shown with dashed arrows, and the W_2 and W_0 transitions are shown with solid arrows (see text for details). (B) A representation of a correlation that would give rise to a NOE cross-peak in a NOE-type spectrum (e.g. ^{15}N-NOESY-HSQC). The H^N, N, C^α nuclei are shown in black and H^α (spin *I*) near H^N (spin *S*) is shown in green. The intensity of the H^N, N, H^α cross-peak would depend on $1/\mathbf{r}^6$.

3.4.3 Structure calculation and refinement

The NMR-determined structural model of a protein must satisfy all data input, as well as be chemically correct in terms of bond lengths, bond angles, van der Waals interactions and charge interactions. The CYANA software fulfill the first requirement: it generates random starting models, and through molecular dynamics (MD) it incorporates dihedral angle constraints and NOE-derived distance constraints into the model. The result is an ensemble of three-dimensional conformations of the protein that best satisfy the data. The computational step is remarkably fast, as several, iterative steps of structure calculation, evaluation and manual (re-)interpretation of NOE assignments and their integrals are required to obtain an ensemble of sufficient quality.

CYANA, however, does not describe a protein as a full-atom model, and CYANA-generated models fail at properly packing side-chains and optimizing distances between interacting charged side-chains. This is instead accomplished by an energy minimization approach using software, like YASARA, that operates with a full-atom forcefield and incorporates an explicit solvent and its effects on the protein structure.

3.5 Functional characterization

NMR can be used to survey dynamic processes that occur in a variety of time scales, for example binding between proteins and ligands, and protein mobility. Characterization of these processes is crucial for understanding enzyme function. Whether the time scale of a process is accessible by NMR depends on the experiment used and its conditions. In addition, the magnetic field strength, B_0, has an important role in defining the observable time scale. For example, for a $B_0 = 18.8$ T, the resonance frequency of ^1H is 800 MHz; this results in approximately one precession even every 10^{-9} s. A fascinating coincidence is that many molecular dynamic processes happen on the same time scale as spin precession (see Figure 3.10), therefore these events manifest themselves directly on NMR measurable parameters, such as the appearance of the NMR signal (lineshape), and T_1, T_2 relaxation times.

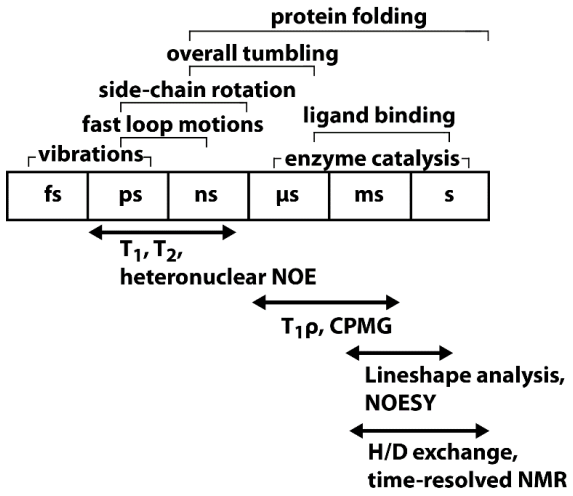

Figure 3.10. Time scales of NMR. Time scales of molecular dynamic processes that occur in a protein and examples of corresponding NMR experiments that can be used to probe such processes.

Many of these phenomena can be described as reversible process where the protein exchanges between two states. For example, during ligand binding the protein can be thought to exchange between the 'free' (F) and 'bound' (B) states, with a rate constant, *k*. Each of the states is assumed to have a unique chemical shift, ν_F and ν_B, respectively. Interestingly, since both the chemical shifts and *k* have units of frequency, the appearance of the peak lineshapes in the NMR spectrum will depend on the relationship between $\Delta\nu = \nu_F - \nu_B$ and *k*.

1. If $k << \Delta\nu$ the process is said to be in slow exchange, and the NMR spectrum will display two sharp signals, one at ν_F and one at ν_B, and their intensities will be proportional to the fraction of 'free' and 'bound' protein.
2. If $k \approx \Delta\nu$ the process is said to be in intermediate exchange, and the NMR spectrum will display one broad signal, often undetectable, at $\bar{\nu} = (\nu_F + \nu_B)/2$
3. If $k >> \Delta\nu$ the process is said to be in fast exchange, and the NMR spectrum will display one sharp signal, at $\bar{\nu} = (\nu_F + \nu_B)/2$.

The approach for extracting functional data from dynamic NMR experiments such as ligand-binding depends on the type of exchange regime. In the case of LPMOs, it is safe to assume that protein-protein binding and protein-carbohydrate binding occur in the fast exchange regime. Evidence supporting this claim is the observation of uniform, unchanging lineshapes of ^{15}N-HSQC peaks during substrate titrations, such as Glc_6 in the case of *Nc*LPMO9C (**Paper I**). While binding of Cu(II) and Zn(II) appears to occur in the slow exchange regime, as in the case of *Sm*LPMO10A[60].

3.5.1 Protein dynamics

Molecular motion is crucial for protein function. Understanding protein dynamics is a prerequisite to fully comprehend protein-ligand interactions, enzyme catalysis, protein folding and stability. Even though, NMR is unrivaled in its potential to characterize protein dynamics in a broad range of time scales, accurate analysis of dynamic data that takes into account the biochemical context of the protein is essential for developing models that describe enzyme function.

As mentioned briefly, NMR is able to measure dynamics because of the unique phenomena that arise from the overlapping time scales of spin precession and molecular motion. A model to describe this relationship is known as the spectral density function, $\boldsymbol{J(\omega)}$. It depends on both the spin-precession frequency, $\boldsymbol{\omega}$, the order parameter, $\boldsymbol{S^2}$, which describes the amplitude of internal motions (1 for fully restricted motions; 0 for completely free motions), and the rotational correlation time ($\boldsymbol{\tau_c}$; i.e. the time it takes a particle to rotate one radian). Since $\boldsymbol{\tau_c}$ describes overall molecular tumbling in solution it can be estimated by the relation $\tau_c = 4\pi\eta r_H^3/3k_BT$, where $\boldsymbol{\eta}$ is the viscosity of the solvent, $\boldsymbol{r_H}$ is the hydrodynamic radius of the particle, $\boldsymbol{k_B}$ is Boltzmann's constant and **T** is temperature.

A

$$J(\omega) = \frac{2}{5}\left(\frac{S^2}{1+(\omega\tau_c)^2}\right)$$

$$1/T_1 = d^2[J(\omega_4) + 3J(\omega_2) + 6J(\omega_3)] + c^2[J(\omega_2)]$$

$$1/T_2 = \frac{1}{2}d^2[4J(\omega_0) + J(\omega_4) + 3J(\omega_2) + 6J(\omega_1) + 6J(\omega_3)] + \frac{1}{6}c^2[3J(\omega_2) + 4(\omega_0)]$$

$$hetNOE = 1 + d^2\gamma[6J(\omega_3) - J(\omega_4)]T_1$$

$$d^2 \propto \frac{1}{r^6}$$

$$c^2 \propto B_0^2$$

B

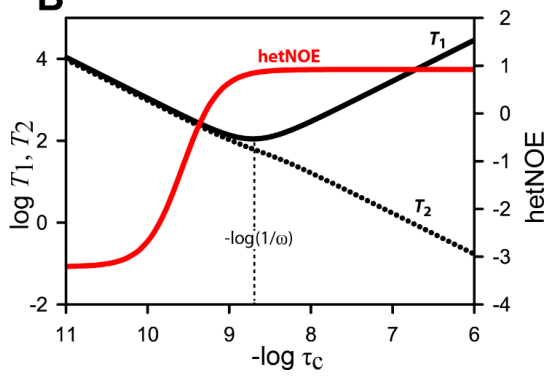

Figure 3.11. Spectral density function. (A) In descending order: Spectral density function $J(\omega)$, ^{15}N T_1, T_2 and heteronuclear $\{^1H\}$-^{15}N NOE expressed as a function of $J(\omega)$, dependence of the dipolar (d^2) and chemical shift anisotropy (CSA; c^2) contributions on $1/r^6$ and B_0^2, respectively. ω_0=0; ω_1= frequency of ^1H; ω_2=frequency of ^{15}N; ω_3=ω_1+ω_2; ω_4=ω_1-ω_2; γ=ratio between the gyromagnetic ratios of ^1H and ^{15}N. See text and ref. 91 for details. (B) Dependence of T_1 (solid black line), T_2 (dotted black line) and heteronuclear NOE (solid red line) on τ_c, calculated at B_0=18.8 T. The minimum value for T_1 is reached when $\tau_c = 1/\omega$.

Interestingly, T_1 and T_2 relaxation times, and the heteronuclear $\{^1H\}$-^{15}N NOE effect can be estimated from the spectral density function (see Figure 3.11). This has two main implications: (i) it demonstrates the connection that exists between the quantum mechanical phenomena of spin precession and relaxation, and the tumbling of particles (ii) it shows that internal and overall motions occurring in the protein can be characterized by measuring these three parameters. A rapid characterization of these motions can be obtained by the heteronuclear NOE $\{^1H\}$-^{15}N NOE (most sensitive to fast, internal motions) and T_2, which is proportional to τ_c (and is thus a good reporter of overall motion). Under the assumption of isotropic tumbling, τ_c can be estimated as a function of T_1/T_2 (see section 4.4.2).

T_1 and T_2 relaxation times are measured by experiments based on the HSQC experiment (see Figure 3.3), with an additional block placed after **t1** evolution that includes a variable time delay, *t*. The length of *t* modulates the intensity of HSQC peaks. Therefore, T_2 and T_1 relaxation times can be determined from the relations $I(t) = I_0\, e^{\frac{-t}{T_2}}$ and $I(t) = I_0(1 - e^{\frac{-t}{T_1}})$, respectively, where *I(t)* is the integral at *t*, and I_0 is the integral at *t*=0.

Heteronuclear $\{^1H\}$-^{15}N NOE measurements follow the same principle as the homonuclear ^1H-^1H NOE described in section 3.4.2, but the measurable quantity arises from dipolar coupling between amide H^N and N. The experiment consists in recording two HSQCs: one with saturation of the W_1 pathway (i.e. "with NOE") and one without. The magnitude of the effect is then determined by the relation hetNOE = 1 + $I_{withoutNOE}/I_{withNOE}$. Note that, for most proteins, the probe for relaxation times and hetNOE measurements is the amide ^{15}N, as it is overall a good reporter of backbone dynamics, but the same principles apply to other nuclei (e.g. ^{13}C and ^1H).

Other noteworthy features of the relationships shown in Figure 3.11A are that the relaxation times are dependent on the dipolar (d^2) and chemical shift anisotropy (CSA; c^2) contributions to relaxation. The dipolar contribution is overall constant, since the distance, **r**, between amide H^N and N does not vary significantly. The CSA contribution arises from the asymmetrical orientation of bond vector (e.g. ^1H–^{15}N) with the magnetic field. This means that the CSA contribution depends on the geometry of the molecule and B_0^2. An implication of this, is that nuclei where CSA is the main contributor to relaxation (e.g. carbonyl ^{13}C) relax at faster rates at higher magnetic fields.

In proteins that contain a paramagnetic center, such as the type II copper site in LPMOs, there is an additional contribution to relaxation through the paramagnetic relaxation enhancement (PRE) effect. The PRE arises from coupling between nuclei and unpaired electrons in paramagnetic centers (known as hyperfine coupling). This creates a new relaxation pathway, enhancing the transversal ($R_2 = 1/T_2$) and, to a lesser extent, longitudinal ($R_1 = 1/T_1$) relaxation rates. This effect leads to severe line broadening, often to undetectable levels, due to the dependence of linewidth on R_2. Interestingly, the hyperfine coupling includes a dipolar interaction component, therefore the PRE effect is also proportional to $1/\mathbf{r}^6$. Because of this feature, paramagnetic labels in proteins can be used to obtain distance constraints.

3.6 Software – format conversion

Analysis of the different types of NMR data described in this chapter requires the use of various software packages, as described in the **Papers**. However, due to the lack of a standard format for handling NMR data, several steps within the experimental strategies required the conversion of data formats. To accomplish this in a timely and robust fashion, a number of scripts were written in Python and LUA. The overall functionality of these scripts consisted in parsing lines from a text file in format X, reordering and reformatting the data and writing a new text file in format Y. Two LUA scripts for CARA produced in this thesis can be found in the CALUA repository http://wiki.cara.nmr.ch/AllCaluaScripts. Python scripts are available from https://bitbucket.org/gcnmr/nmr-format-conversion/.

4. Additional results and discussion

Since the main outcomes of this thesis are already presented in the **Papers**, the purpose of this chapter is to offer an insight into other, unpublished results, as well as inner workings of the project that would not otherwise see the light of day. First, an overview of the sequence-specific assignments, as well as an evaluation of an automatic assignment approach are presented. Then, an excerpt of structural features of *Bl*LPMO10A is described, as well as the structures of *Cj*CBM5 and *Cj*CBM73. An alternative approach for analyzing ligand-binding NMR data of *Nc*LPMO9C is presented. Finally, trends in the overall dynamic features of LPMOs are discussed, with special focus on the domain dynamics of modular *Sc*LPMO10C, as well as the role of the linker region and *Sc*CBM2 domain in directing cellulose oxidation.

4.1 Automated sequence-specific assignments

The sequence-specific assignments form the basis of the LPMO characterization in this thesis. The assignments for all proteins have been attained and are summarized in Table 4.1. These assignments allowed structural and functional characterization of the proteins, including probing enzyme-substrate interactions, monitoring the effects of ligand binding, performing pH titrations, and characterizing protein dynamics.

In addition to the value of the assignments in enabling further experimental characterization of the proteins, the assignment papers mark interesting milestones throughout the course of the thesis. **Paper V** included the proof-of-principle use of the *Pm*/XylS expression system for producing ^{13}C and ^{15}N isotope enriched AA10 LPMOs. This was the basis for the development of the LPMO expression system described in **Paper II**. **Paper VI** describes the first production of a ^{13}C and ^{15}N isotope enriched fungal AA9 LPMO in *P. pastoris*, as well as the first NMR investigation of an AA9 family LPMO. **Paper VII** describes the first NMR investigation of a modular LPMO, including a proposed assignment for the disordered linker region.

Table 4.1. List of assigned proteins. LPMOs for which NMR assignments have been produced during this thesis. The name of the protein, its size in terms of molecular weight (Mw) and number of amino acids (aa) is given, as well as the completion percentage of the assignment and where it is published.

Protein	Size (Mw, aa)	Completion	Published
*Bl*LPMO10A	19.2 kDa, 172 aa	Backbone > 97% H, C side-chains > 90%	**Paper V** BMRB: 19984
*Nc*LPMO9C	23.3 kDa, 227 aa	Backbone > 98% H, C side-chains > 90%	**Paper VI** BMRB: 26717
*Sc*LPMO10C-AA10 domain	21.4 kDa, 196 aa	Backbone > 90 % Aliphatic side-chains > 80%	**Paper VII** BMRB: 27078
*Sc*LPMO10C-CBM2 domain	10.7 kDa, 104 aa	Backbone > 97% All aliphatic and aromatic side-chains > 75%	
*Sc*LPMO10C – full-length	34.6 kDa, 330 aa	Backbone ~ 60% Linker region > 80%	
*Ta*LPMO9A	24.4 kDa, 228 aa	Backbone > 95% Aliphatic side-chains > 79%	**Paper VIII** BMRB: 27411
*Cj*LPMO10A-CBM5 domain	6.6 kDa, 59 aa	Backbone > 95 % Aliphatic side-chains > 90 %	-
*Cj*LPMO10A-CBM73 domain	6.4 kDa, 60 aa	Backbone > 95 % Aliphatic side-chains > 90%	-

The assignments were achieved by using a computer-assisted approach with CARA[97]. Essentially, every signal in the sets of multidimensional spectra was manually peaked and classified in a spin-system, after semi-automatic comparison of the connectivity between spin-systems, resonances were assigned to the amino acid sequence of the protein. This is a reliable but time-consuming approach, as thousands of signals must be manually evaluated and verified, and the difficulty of the process increases with the size of the protein. It is thus enticing to evaluate the extent to which the procedure could be automated.

Peak-picking, the identification of individual peaks in spectra, is a step that still has not been properly automated. Programs such as AUTOPSY[105] and CARA scripts[97] that rely on threshold-based peak-picking, require extensive manual curation of incorrectly picked and omitted peaks. However, the assembly of spin-system connectivity and the assignment of connected fragments to the amino acid sequence can be reliably automated. Here, we have used the I-PINE web server (http://i-pine.nmrfam.wisc.edu/) to automatically assign the *Sc*CBM2 domain of *Sc*LPMO10C.

I-PINE[106] uses Bayesian inference to estimate the likelihood of two spin-systems being connected together. The only input submitted to the server were the unassigned peak lists that had been manually peak-picked. The unsupervised assignment procedure then outputs the sequence-specific assignment. The automatic I-PINE assignment approach, however, only managed to assign 88% of the total number of residues assigned using the manual CARA approach. Moreover, from the assigned residues only 82% of them were correctly assigned (Figure 4.1). In total, this means that even using manually curated peak-lists as input, I-PINE managed to correctly assign only 72% of the chemical shifts to the amino acid backbone. Closer inspection of the misassigned residues shows that they are clustered in misassigned fragments (e.g. 293-301 and 332-335). The misassigned residues contain either no shifts besides H^N, N (e.g. Val335) or aberrant shifts, e.g. Val334 and Val298 were assigned a Cα chemical shift of 50.9 and 48.8 ppm, respectively, far from the expected[107] 62.56 ± 2.84 ppm (± S.D.). This suggests that the while using an automatic method such as I-PINE may help with the bulk of the sequence-specific assignment procedure, the output must be manually verified to ensure that the assignment is of sufficient quality. Therefore, until better unsupervised approaches become available, it is recommended to continue performing sequence-specific assignments in a semi-manual, computer-assisted fashion.

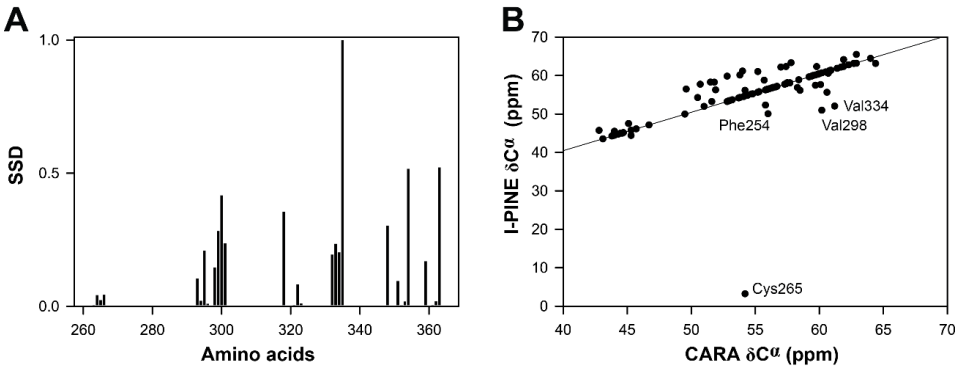

Figure 4.1. Evaluation of automatic I-PINE assignment. (A) Normalized sum of squared difference (SSD) between the CARA- and I-PINE-assigned H^N, N chemical shifts. A higher SSD means larger deviation from the correct (semi-manual CARA) assignment. (B) Plot of Cα chemical shifts (δ) assigned by CARA and I-PINE. The dots aligned with the line indicate that the assignments are identical in both methods. Four selected deviant assignments are labeled. Cys265 was assigned a δC^{α} = 3.22 ppm, while Phe254 as assigned a δC^{α} = 50.1 ppm (expected 58.14 ± 2.57 ppm).

4.2 Structure determination

The assigned chemical shifts served as a basis for further structure elucidation. First, chemical shifts were run through the TALOS-N[95] or SSP servers[108] to obtain estimates of the secondary structure content in the proteins. Generally, the analysis of secondary structure elements showed excellent agreement with the secondary structure elements present in X-ray crystal diffraction structures. However, in some cases, this analysis yielded new insights into the conformation of certain regions of the protein. For example, the secondary structure analysis of full-length *Sc*LPMO10A indicates that its disordered linker region has an overall extended conformation (**Paper VII**). In the case of *Bl*LPMO10A, *Sc*CBM2, *Cj*CBM5 and *Cj*CBM73, where no crystal structures were available, the secondary structure analysis was used as a starting point for the NOE assignment. This is possible because α-helices and β-strands give rise to characteristic NOE patterns, e.g. strong intra-residue H^N-Hα cross-peaks in β-strands, and inter-residue (2, 3 and 4 residues away) H^N-Hα cross-peaks in α-helices (see ref. 88 for details).

4.2.1 Structural features of *Bl*LPMO10A

Currently (July 2018), there are over 60 LPMO-related entries in the Protein Data Bank. Most of these LPMO structures were solved using X-ray crystallography, and two of them by NMR. One of the latter is the structure of *Bl*LPMO10A (PDB ID: 5LW4), presented in **Paper IV.** *Bl*LPMO10A is a chitin-active, C1-oxidizing LPMO from the soil bacterium *Bacillus licheniformis*. The protein depicts the typical LPMO β-sandwich fold, formed by a three-stranded β-sheet and a five-stranded β-sheet. The structure does not show any remarkable differences with other chitin-active family AA10 LPMOs such as *Sm*LPMO10A or *Ba*LPMO10A, and it appears to have an extended L2 loop, similarly to *Sc*LPMO10B. The structure of *Bl*LPMO10A is thus a good example to highlight structural details of LPMOs that manifest themselves in NOESY spectra in distinctive ways.

An example of one of these details is the presence of *cis* prolines in LPMOs. Peptide bonds usually adopt a *trans* configuration (Ω dihedral angle ~180°), but prolines can also form peptide bonds in *cis* configuration (Ω ~ 0°). Proline *cis*/*trans* isomerism often has a significant effect in the way the polypeptide chain folds. The structure of *Bl*LPMO10A starts with the N-terminal histidine (His32), one of the two metal binding site histidines. It is followed by the first β-strand in the three-stranded β-sheet, encompassing Gly33-Lys37. Pro38 is the first *cis* proline (the other one is Pro73); its configuration causes the loop region consisting of residues Pro38-Ser40 to bend (Figure 4.2C), which results in the successive α-helix (Arg41-Glu47) being positioned perpendicular to the β-strand. There is two-fold evidence to assign the Pro38 as a *cis* proline.

First, the *cisprocheck* macro in CYANA analyzes the difference between C^{β} and C^{γ} Pro chemical shifts to test for the presence of *cis* prolines[109]. Prolines in *trans* configuration tend to have C^{β}-C^{γ} values of 0-5 ppm, while for *cis* Pro, C^{β}-C^{γ} values are approximately 7-12 ppm[109]. For Pro38 in *Bl*LPMO10A, the C^{β}-C^{γ} value is ~7 ppm.

Additionally, in the case of *cis* prolines, an NOE cross-peak would be expected between the Pro-H^{α} and the H^{α} of the preceding residue[88]. For *Bl*LPMO10A, an NOE between H^{α}-Lys37 and H^{α}-Pro38 was indeed observed in the ^{13}C-NOESY-HSQC spectrum (Figure 4.2A). This NOE correlation would not be present if Pro38 had been a *trans* proline; the

H^{α}-Lys37 and H^{δ}-Pro38 cross-peak would appear instead[88]. The *cis*Pro38 bend is further defined by the NOE cross-peaks from H^{N}-Gly39 to both H^{α}-Lys37 and H^{α}-Pro38, as seen in the ^{15}N-NOESY-HSQC spectrum (Figure 4.2B).

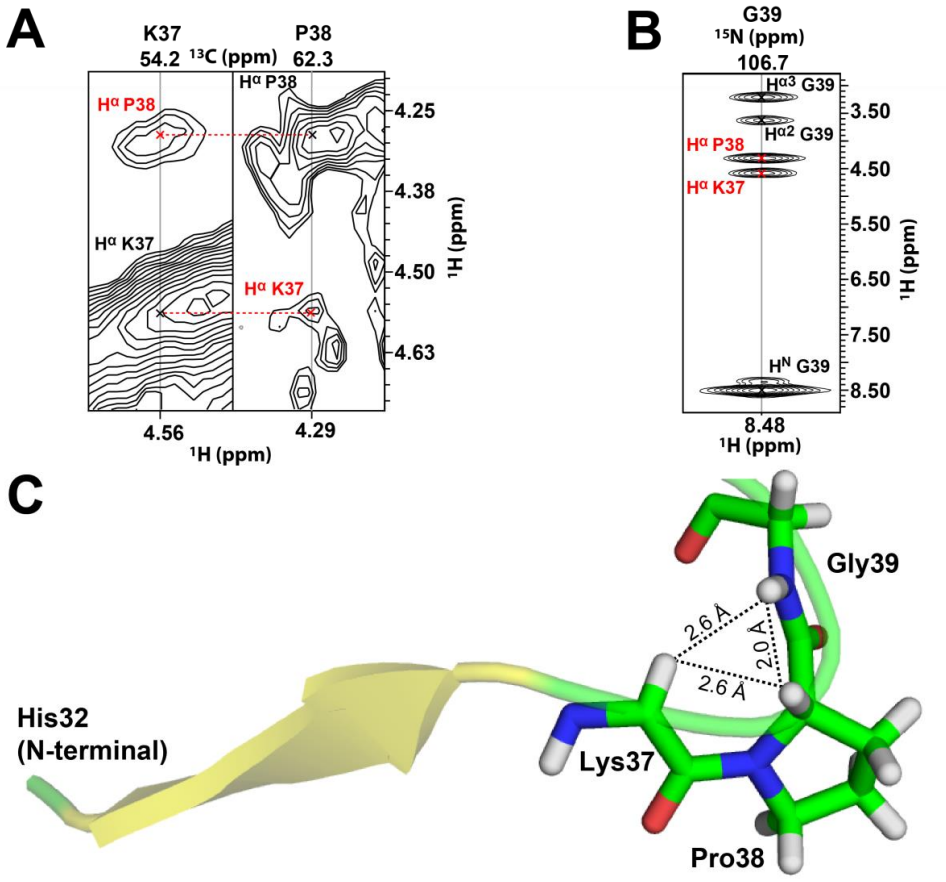

Figure 4.2. NOESY spectra and *cis* proline in *Bl*LPMO10A. (A) Excerpt from the ^{13}C-NOESY-HSQC spectrum of ^{13}C- and ^{15}N-enriched *Bl*LPMO10A showing NOE cross-peak between H^{α}-Lys37 and H^{α}-Pro38. (B) Excerpt from the ^{15}N-NOESY-HSQC spectrum showing NOE cross-peaks from H^{N}-Gly39 to both Gly39 H^{α}'s ($H^{\alpha 2}$ and $H^{\alpha 3}$; black), H^{α}-Lys37 (red) and H^{α}-Pro38 (red). (C) Cartoon representation of the stretch His32-Leu44 (yellow: β-strand; green: loop; red: α-helix), with stick representation of Lys37 (backbone), Pro38 (backbone and side-chains), Gly39 (backbone). The distances between the atoms are indicated in Ångström and with dotted lines.

Right after the first α-helix (Arg41-Glu47), there is a second helix, assigned as a 3_{10}-helix (Gly57-Glu62). These two helical segments are linked together by a conserved disulfide bridge (Cys45-Cys56). This disulfide bridge is crucial because, by "locking" these helices in place, it positions Tyr61 and Phe51 on the same surface plane as the copper-active site. As seen by substrate binding experiments of *Bl*LPMO10A with chitin flakes (**Paper IV**) polar residues on the LPMO surface appear to be key for chitin binding. Tyr61[60] (equivalent to Tyr54 in *SmL*PMO10A[56]) is in the center of this surface and likely promotes LPMO orientation on chitin through interactions of aromatic π-electrons with C-H bonds in GlcNAc rings.

*SmL*PMO10A has an additional disulfide bridge (Cys145-Cys162) that connects the five-stranded β-sheet with the three-stranded β-sheet, stabilizing the β-sandwich core. Interestingly, *Bl*LPMO10A does not have a second disulfide bridge. A structure alignment (Figure 4.3A) shows two isoleucines (Ile149 and Ile167) instead of Cys. The Ile side-chains are close enough in space to give an NOE cross-peak (Figure 4.3B) and it is likely that their methyl groups, through hydrophobic interactions, play a similar structural role as the disulfide bridge in *SmL*PMO10A.

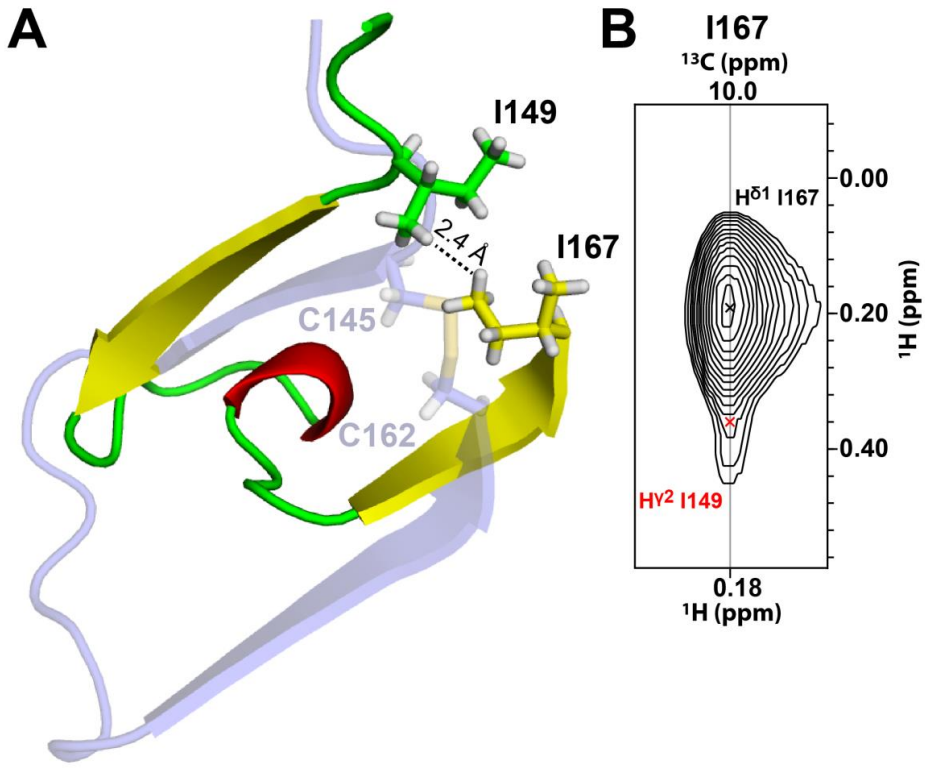

Figure 4.3. Hydrophobic interaction between two Ile in *Bl*LPMO10A. (A) Superposition of a selected region of the structure of *Sm*LPMO10A (blue) and *Bl*LPMO10A (yellow: β-strand; red: helix; green: loop). The disulfide bridge (Cys145-Cys162) in *Sm*LPMO10A is shown, as is the distance in the hydrophobic contact between the Ile167-$H^{\delta1}$ and Ile149-$H^{\gamma2}$. (B) The NOE cross-peak between these atoms is visible in the ^{13}C-NOESY-HSQC. The low ^{1}H chemical shifts (~0.20 ppm) are likely a result of ring-current effects caused by the nearby Phe113 and Tyr128 (not shown).

4.2.2 Structural characterization of carbohydrate-binding modules

Despite the abundance of CBMs in carbohydrate-active enzymes and LPMOs, there is an overall lack of understanding regarding the structure and function of LPMO-associated CBMs. At the start of the thesis, functional studies on CBMs had showed evidence that they enhance the activity of their appended catalytic LPMO domains[78–80]. However, the structures of these CBMs were not known. During the thesis, the structures of *Sc*CBM2 (PDB ID: 6F7E), *Cj*CBM5 and *Cj*CBM73 were characterized by NMR spectroscopy (Table 4.2 and Figure 4.4). Since, the structure of ScCBM2 is presented in **Paper III**, the focus of this section is on *Cj*CBM5 and *Cj*CBM73. Both structures are similar (C^{α} rmsd ~ 4.2 Å) and have a unique fold. The fold, described as a "skiboot shape"[83], consists of a three-stranded antiparallel β-sheet, perpendicular to a loop region. The structures are overall stabilized by a disulfide bridge connecting the N-terminal region of the protein with the C-terminal region. The main difference is that *Cj*CBM73 contains an additional disulfide bridge, connecting the central β-strand with a helical region (residues 333-339). No analog helical region was observed in *Cj*CBM5.

Three aromatic residues are located on the surface of the loop region. For *Cj*CBM5, these residues are Tyr282, Trp283 and Tyr265; for *Cj*CBM73, Tyr315, Trp348 and Trp350. These residues are partially conserved within the CBM5 family, as they can be either Tyr or Trp. Brun *et al*[83] defined these residues as the cellulose-binding surface of *Ec*CBM5, associated with endoglucanase Z from the plant pathogen *Erwinia chrysanthemi*. *Cj*CBM5 has been observed to bind weakly both Avicel and filter paper, but this is not the case for *Cj*CBM73[79]. Both CBMs have been shown to bind chitin, but *Cj*CBM73 does so more strongly than *Cj*CBM5 (Figure 4.4C). This could be explained by the compositions of the chitin-binding surfaces, where tryptophan is expected to be the best carbohydrate-binding residue[74]. Namely, *Cj*CBM73 has two surface tryptophans while *Cj*CBM5 has only one. An interesting feature in both structures is that, opposed to that what would be expected for CBMs (Figure 1.5), the aromatic side-chains on the surface of the protein are not aligned (Figure 4.4A, B). This could imply that the function of *Cj*CBM5 and *Cj*CBM73 may be to bind more than one chitin chain at the same time, thus targeting crystalline chitin surfaces.

As shown in Table 4.2, the overall quality of the structures is consistent with the number of NOE distance restraints included in the calculation. *Cj*CBM73, in particular, will require a larger number of restraints in the loop region in order to reduce the backbone rmsd and obtain a better-defined structure.

Table 4.2. Input data for calculation of the structures of *Cj*CBM5 and *Cj*CBM73 together with the structural assessment statistics. The structures were calculated using the torsion angle dynamics program CYANA 3.97[110]. The structure calculation started by generating 200 conformers with random torsion angles, and the dihedral angles in each conformer were optimized using simulated annealing in 10000 steps, to fit the restraints. The 20 conformers with the lowest CYANA target function values were energy-minimized using YASARA[111] first in vacuo, followed by using water as the explicit solvent and calculating electrostatics by applying the particle mesh Ewald method[112]; both steps used the YASARA force field[113].

	***Cj*CBM5**	***Cj*CBM73**
Total number of NOE distance constraints	581	299
Intraresidue	218	131
Sequential	286	100
Medium-range	23	23
Long-range	54	45
Torsion angle restraints[a]	93	72
Disulfide bridges	253-305	304-360, 330-336
Structure statistics (20 conformers)[b]		
CYANA target function value (Å^2)	2.40 ± 0.27	3.02 ± 0.70
Maximum residual distance constraint violation (Å)	0.42 ± 0.11	0.94 ± 0.23
Maximum torsion angle constraint violation (°)	4.15 ± 1.19	4.27 ± 4.06
PROCHECK-NMR Ramachandran plot analysis		
Residues in favored regions (%)	81.0	80.0
Residues in additionally allowed regions (%)	18.0	19.1
Residues in generously allowed regions (%)	0.9	0.3
Residues in forbidden regions (%)	0.1	0.5
rmsd to the average coordinates (Å)[c]		
N, C$^\alpha$, C’	1.44 ± 0.28	3.69 ± 0.77
Heavy atoms	1.94 ± 0.27	4.32 ± 0.81
N, C$^\alpha$, C’ (Best-defined regions)[d]	0.96 ± 0.25	0.99 ± 0.31
Heavy atoms (Best-defined regions)[d]	1.46 ± 0.27	1.64 ± 0.29

[a] Calculated from secondary chemical shifts using the TALOS-N software[95].
[b] The values are the average and standard deviation over the 20 lowest CYANA target function values before energy minimization.
[c] Average coordinates of the 20 conformers after superposition of the N, C$^\alpha$ and C’ atoms to the lowest target energy conformer.
[d] For *Cj*CBM5: residues 251-289, 297-307. For *Cj*CBM73: residues 304-313, 315-316, 318-330 and 353-360

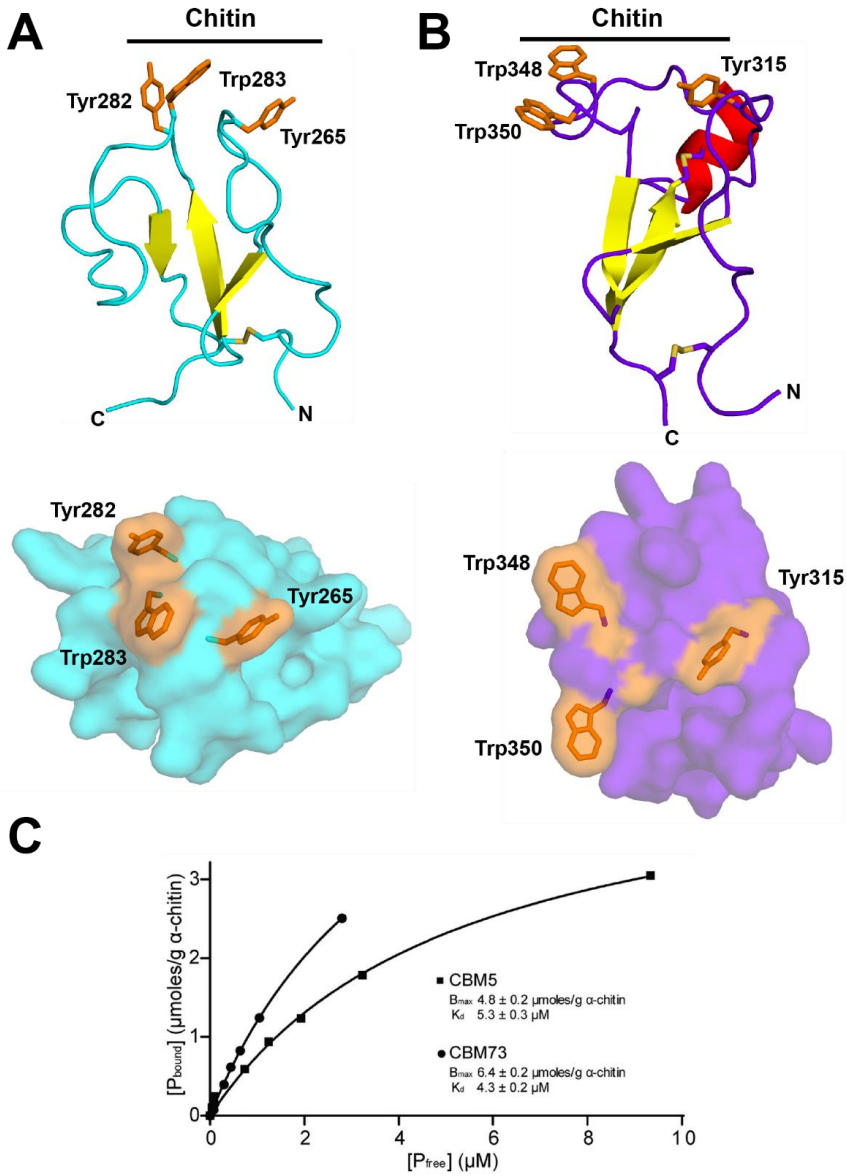

Figure 4.4 Structural characterization of CBMs. (A) *Cj*CBM5 (cyan) and (B) *Cj*CBM73 (purple) shown as cartoon representations (top panels) with sheets (yellow) and helix (red), and as surface representations (bottom panels). The side-chains of the putative substrate binding residues are shown as sticks, labeled, and colored orange. The N- and C-termini are labeled, and disulfide bridges are shown. (C) Binding data for *Cj*CBM5 and *Cj*CBM73 domain incubated with α-chitin. P_{bound} corresponds to bound protein (µmol/g substrate), and P_{free} corresponds to non-bound protein (µM). Panel C is reprinted with permission from ref. 79.

4.3 Binding competition: cellobiose dehydrogenase vs. cellohexaose

The gradual shifting of peaks, unaccompanied by a change in signal intensity, in ^{15}N-HSQC spectra of *Nc*LPMO9C during titrations in **Paper I** suggest that the binding with substrates and the cytochrome domain of cellobiose dehydrogenase (CDH) occur in a fast exchange regime. Even though binding affinity has no effect on the exchange regime, weak binding in the mM range, as is the case for the ligands used in this investigation, tend to be in fast exchange[114]. This reflects the nature of the LPMO reaction, where the protein needs to switch between different binding partners to receive electrons and to oxidate substrate. In a fast exchange regime, interactions with ligands are relatively short lived, allowing the LPMO to swiftly exchange from 'free' to 'CDH-bound' and 'substrate-bound' during its reaction cycle.

One of the main findings from **Paper I** is that CDH binds to the same site as the cellohexaose substrate (Glc_6), on the surface around the active site on *Nc*LPMO9C. Since CDH is an electron donor, this observation entails that electron transfer to the LPMO and LPMO-binding to the substrate cannot happen simultaneously. The bases for this "same-site" binding model are chemical shift perturbations in ^{15}N-HSQC spectra upon adding of either ligand (Glc_6 or CDH; see Figure 3 in **Paper I**). The observation that, upon perturbation, chemical shifts shifted in *different* directions, depending on the ligand, enabled assigning certain H^N, N chemical shifts as ligand-free, Glc_6-bound and CDH-bound, as illustrated in Figure 4.5A. In **Paper I,** we relied on qualitative observation of reporter peaks in ^{15}N-HSQC spectra to assert that the same signals were shifted in different directions by each ligand. These findings led us to conclude that CDH and Glc_6 bound to the same site on *Nc*LPMO9C. However, an alternative, more quantitative approach for analyzing chemical shift perturbations may be applied. Here, we propose the implementation of a vector-projection approach[115] for evaluating whether the two ligands truly elicit changes in chemical shifts in different directions.

The movement of a peak shifting from point $\mathbf{a_1}$ (e.g. ligand-free) to $\mathbf{a_2}$ (e.g. CDH-bound) can be described by a vector $\vec{\boldsymbol{a}}$, and a difference in direction between two vectors $\vec{\boldsymbol{a}}$ (e.g. representing a change in chemical shift from ligand-free to CDH-bound) and $\vec{\boldsymbol{b}}$ (e.g. representing a change in chemical shift from ligand-free to Glc_6-bound) can be described

as the projection of $\vec{b}$ on $\vec{a}$ (see Figure 4.5). A negative value for the projection means that $\vec{a}$ and $\vec{b}$ indeed point in directions opposite to each other; a value of -1 would mean that $\vec{a}$ is the same length as $\vec{b}$, but its direction is the exact opposite. A positive value for the projection means that $\vec{a}$ and $\vec{b}$ point in approximately the same direction; a value of 1 would mean that $\vec{a}$ is the same length and direction as $\vec{b}$. This analysis was carried out for the *Nc*LPMO9C datasets containing chemical shifts perturbations, showing that a number of the H^N, N signals perturbed by CDH are indeed the same that were perturbed by Glc_6, as previously observed. In addition, results shows that H^N, N chemical shifts from amino acids closest to the copper site were shifted in an overall different direction upon addition of CDH, compared with Glc_6, whereas signals from residues farther from the copper site were shifted in overall the same direction (Figure 4.5C, D). Furthermore, the vector projection analysis suggests that an overall larger area, including the L3 loop, might be affected by CDH than what was originally presented in **Paper I.** The "same-site" binding model presented in **Paper I** was also supported by cyclic voltammetry data of CDH and *Nc*LPMO9C[69]. This ultimately led to the conclusion that Glc_6 and CDH compete for the same binding site and that electron transfer occurs directly from the heme *b* cytochrome domain in CDH to the copper active site in LPMOs.

The main difference between this vector-based analysis and the compound chemical shift analysis (see **Paper I**), is that, while the latter gives only information about the magnitude of the chemical shift change, which can be used to estimate binding affinities (e.g. Appendix A), the vector-based analysis contains information about both magnitude and direction of the change in chemical shift. This is interesting because ^1H chemical shifts are primarily perturbed by distant effects (hydrogen bonding, ring currents, pH), whereas ^{15}N shifts are mainly affected by conformational changes in the backbone of the protein[116]. Therefore, it would appear that, in order to characterize ligand binding modes, one needs to account for the multiple effects that the ligands might have on the chemical shifts of the protein upon binding.

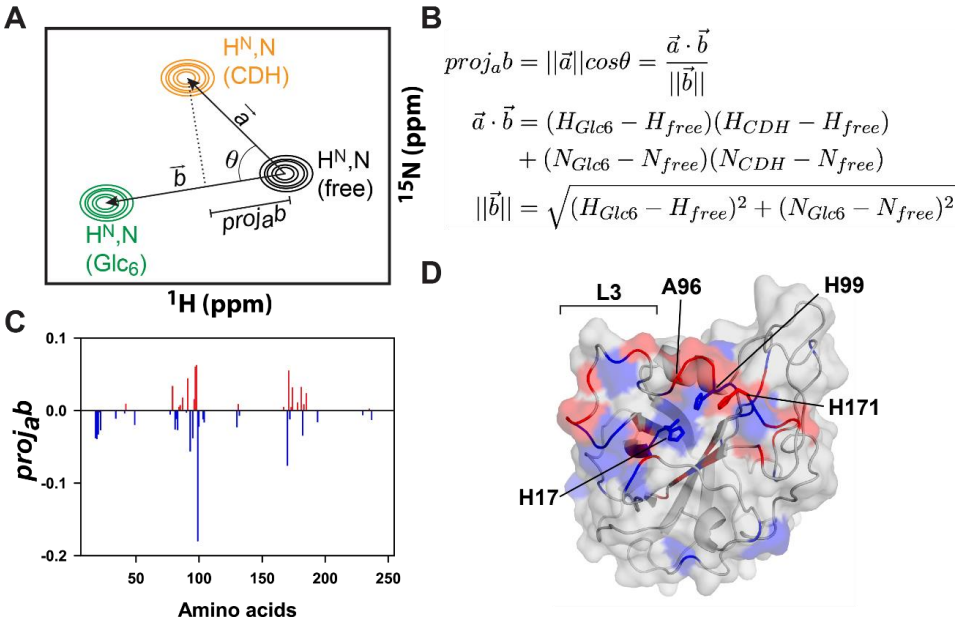

Figure 4.5 Vector projection analysis of chemical shift perturbation by two ligands. (A) ^{15}N-HSQC Representation of titration of free LPMO with Glc_6 (green) and CDH (orange), based on **Paper I** (Fig. 3). $\vec{a}$ is the vector between 'free' and 'CDH-bound', $\vec{b}$ is the vector between 'free' and 'Glc_6-bound'. θ is the angle between the vectors and $proj_a b$ is the projection of $\vec{a}$ on $\vec{b}$. (B) Mathematical description of the vector projection approach using ^{1}H ($H_{Glc6,\ free,\ CDH}$) and ^{15}N ($N_{Glc6,\ free,\ CDH}$) chemical shifts. (C) Plot of the projection value for *Nc*LPMO9C, a positive (red) value means that the CDH vector points approximately in the same direction as the Glc_6 vector; a negative (blue) value means that the CDH vector points approximately in the opposite direction as the Glc_6 vector. The values are highlighted with the same color on the structure of *Nc*LPMO9C in (D), where the protein structure is shown as cartoon and surface representation. The side-chains of H17 (H1 in **Paper I**), A96 (A80 in **Paper I**), H99 (H83 in **Paper I**) and H171 (H155 in **Paper I**) are shown as sticks and highlighted, as is the L3 loop.

4.4 Dynamics of LPMOs

An essential aim of enzymology is to understand the relationships between function, structure and dynamics. Only a holistic approach that considers all three aspects can lead to a complete understanding of how enzymes work. For LPMOs, the challenge is no different: structural studies must be complemented with insights into protein mobility and biochemical function. In this thesis, we have done this by dynamics features of LPMOs, particularly their rotational tumbling in solution. Moreover, the dynamics of the domains and the linker region in full-length *Sc*LPMO10C, together with biochemical assays of cellulose oxidation, represent the first steps towards a comprehensive understanding of modular LPMOs.

4.4.1 Rotational correlation times in LPMOs

When T_1 and T_2 relaxation time data are available, one of the most interesting parameters to calculate using these data is the rotational correlation time (τ_c). For relaxation time data, $\tau_c = 0.5 \cdot (6 \cdot T_1/T_2 - 7)^{1/2}$, under the assumption of isotropic tumbling[92]. As explained earlier, τ_c is a quantity that depends on the molecular weight and hydrodynamic radius of a protein. Therefore, τ_c can also be regarded as an indirect measure of how the mass of the protein is distributed around its center. Normally, globular proteins have a distribution of mass such that the correlation time is approximately 0.6 times their molecular weight (see Figure 4.6). However, for LPMOs the relation is different, namely $\tau_c \approx 0.5 \cdot$Mw, which suggests that LPMOs are overall compact proteins. This observation, together with the finding that plastocyanin and azurin (copper-containing proteins involved in electron transfer reactions) have also been described as compact on the basis of their NMR relaxation properties[117], indicates that compactness might be a hallmark of copper-proteins. This is possibly advantageous for electron tunneling[118] and for structural stability of the copper-active site, as coordination lengths change upon reduction of Cu(II) to Cu(I)[119]. A fortunate consequence of these exceptionally low correlation times is that LPMOs have sharper signals in NMR spectra, facilitating interpretation and analysis.

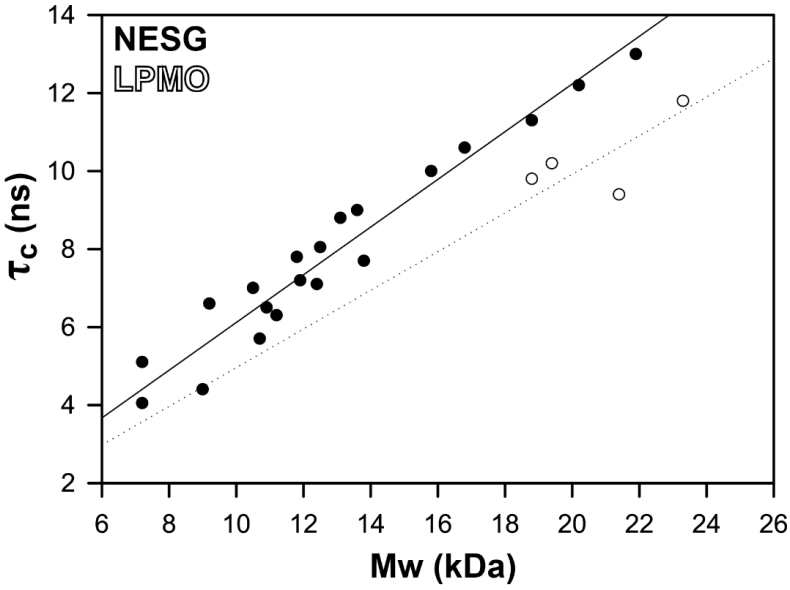

Figure 4.6. Rotational correlation time (τ_c) as a function of molecular weight (Mw). Comparison of proteins from the NESG (Northeast Structural Genomics Consortium) database[100] (black; solid line: $\tau_c \approx 0.6 \cdot Mw$) versus LPMOs (*Nc*LPMO9C, *Sc*LPMO10C-catalytic domain, *Bl*LPMO10A, *Sm*LPMO10A) for which T_1 and T_2 data are available (white; dotted line: $\tau_c \approx 0.5 \cdot Mw$).

4.4.2 Dynamics and cellulose oxidation of modular *Sc*LPMO10C

The main findings in **Paper III** pertain to the relationship between the dynamic nature of the linker region in *Sc*LPMO10A and its effect on the biochemical function of the full-length protein, compared with the isolated catalytic domain. Comparison of the τ_c values for the isolated *Sc*AA10 and *Sc*CBM2 domains, and the domains tethered together in full-length *Sc*LPMO10A showed that τ_c was not significantly affected, indicating that the domains could still move semi-independently from each other when connected together by the linker region. These τ_c values were calculated as a function of T_1/T_2 relaxation times, with the assumption that the rotational tumbling was overall isotropic. While this would be the case for a perfectly spherical protein, it is relevant to evaluate whether the assumption holds for the proteins characterized in **Paper III**.

Rotational anisotropy is a result of the shape of the protein and leads to relaxation properties of the nucleus at hand (e.g. ^{15}N) that are dependent on the H^N–N bond orientation. Rotational anisotropy can be evaluated by calculation of the diffusion tensor, which can be derived from relaxation time data, as well as from the three-dimensional coordinates of the H^N–N atoms, i.e. from the structure of the protein. The principal components of the diffusion tensor (D_x, D_y, D_z) are all equal in the case of isotropic tumbling. If two of the components are equal, the diffusion tensor is said to be axially-symmetric anisotropic, and if they are all different the diffusion tensor is fully anisotropic. Detailed theoretical background for the estimation of the diffusion tensor and dynamic parameters are described elsewhere (e.g. ref. 92). Here, the diffusion tensors for isolated *Sc*AA10 and *Sc*CBM2 domains, and the domains tethered together in full-length *Sc*LPMO10A were calculated using the ROTDIF software[120], which was also used to calculate the new τ_c values. The results, summarized in Table 4.2, show that, whereas the data is best modelled by an axially-symmetric anisotropic model, the overall τ_c values do not deviate largely from the ones estimated under the isotropic tumbling assumption, making them good predictors for the overall rotational tumbling of the domains characterized in **Paper III**. Notably, the diffusion tensor of *Sc*AA10 appears to be most affected by the attached linker and CBM2 domain, as opposed to the diffusion tensor of *Sc*CBM2, which varies to a lesser degree when attached to the linker and AA10 domain. This suggests that the linker and CBM2 could influence the motion of *Sc*AA10, maybe directing the domain towards the cellulose surface.

Table 4.2. *Sc*LPMO10C domain dynamics. Isotropic rotational correlation time (τ_c) calculated as described in section 4.4.1. Diffusion tensor and average anisotrpic τ_c calculated using ROTDIF[120].

Protein	Isotropic τ_c (ns)	Diffusion tensor (D_x, D_y, D_z)	Average anisotropic τ_c (ns)
Isolated *Sc*AA10	9.40	1.36, 2.17, 2.17	8.78
*Sc*AA10 in *Sc*LPMO10C	9.38	1.21, 1.21, 4.54	7.18
Isolated *Sc*CBM2	7.40	1.97, 2.36, 2.36	7.49
*Sc*CBM2 in *Sc*LPMO10C	8.57	1.48, 2.20, 2.20	8.50

The model in **Paper III** proposes that the *Sc*AA10 domain in *Sc*LPMO10C moves around *Sc*CBM2, which is bound to the cellulose surface. The extent of the motion of *Sc*AA10 is determined by the flexibility and length of the linker region. Thus, for modular *Sc*LPMO10C, only cellulose glucose units that can be reached by the copper active site of *Sc*AA10 can be oxidized, leading to a localization of oxidized residues around *Sc*CBM2. For the isolated catalytic domain *Sc*AA10, the motion is random, and no localization on the cellulose surface may be expected. Two oxidation events (i.e. cleavage points) on the same cellulose chain, eight or less glucose residues apart, give rise to the release of soluble products. In **Paper III**, we observed that ~60% of the soluble, oxidized products generated by full-length *Sc*LPMO10C had degrees of polymerization between 2 and 4, compared to ~40% for isolated *Sc*AA10. In order to verify that this distribution of soluble products indeed would be expected for *Sc*LPMO10C if it behaved according to the proposed model, a simulation of the model was scripted in Python (see Figure 4.7). Results from this simulation show that, while the distribution of the degrees of polymerization of soluble products differ from the experimental values, ~60% of the simulated soluble products have degrees of polymerization between 2 and 4, and that ~30% of the products are insoluble. This is in very good agreement with the experimental observations from cellulose degradation experiments in **Paper III**. Despite the obvious need to optimize the simulation and its parameters, such as the size of the cellulose surface, this incipient approach represents a step towards developing computer models to explain correlations between LPMO dynamics and function.

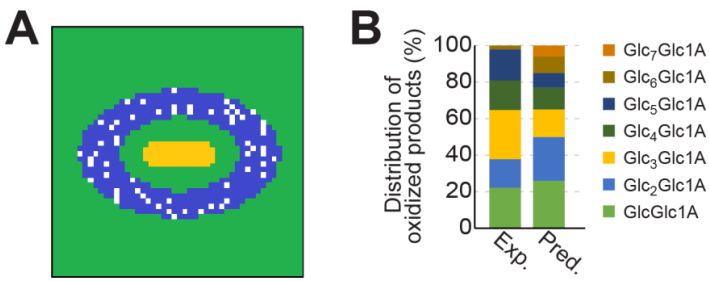

Figure 4.7. Model of cellulose oxidation by *Sc*LPMO10C. (A) Graphical representation of the simulation model generated by the Python script. The simulation was designed as follows. (i) A simplified model of a cellulose surface was generated by placing 48 horizontal cellulose chains 8.2 Å apart from each other. Each cellulose chain contained 48 pseudo-residues placed 5.2 Å apart. This is shown in green, where every pseudo-residue is a square. (ii) The *Sc*CBM2 was placed in the center of the cellulose surface (orange). (iii) The allowed area for *Sc*AA10 (blue) was generated by a uniform distribution of the lower and upper limits of the length of *Sc*LPMO10C, around the anchored *Sc*CBM2. These limits were determined in **Paper III** from the NMR calculated ensemble of and dynamic light scattering of *Sc*LPMO10C. (iv) A 60 minute reaction was started by randomly choosing one of the residues in the allowed area, and 'oxidizing' it. This was repeated 60 times, as the reaction rate for LPMOs[29] is ~ 1 min^{-1}. Oxidation events are shown as white squares. (v) After the reaction, the length of the fragments between two oxidation events was measured to calculate the distributions, shown in (B) together with experimental data from **Paper III**. The Python script to run the simulation is available from http://folk.ntnu.no/courtade/.

5. Concluding remarks and future perspectives

Efforts toward understanding how an enzyme works require the integration of structure, dynamics and functional aspects. Understanding various aspects of LPMOs, including characterization of their structure, the effect of copper, their substrate binding properties, the electron transfer event, their appended carbohydrate-binding modules and their dynamic features have been the focus of this thesis. The main investigation strategy has been sample production followed by chemical shift assignments that enabled studies on these key aspects of LPMO functionality.

The first challenge in the experimental strategy, achieving stable and reliable expression of LPMOs, was solved by constructing a tunable expression system for bacterial LPMOs whose regulation is independent from the carbon metabolism of the cell (**Paper II**). For CBMs, modular *Sc*LPMO10C and fungal LPMOs, alternative and successful approaches were carried out. Altogether, these methods allowed production of high amounts of ^{13}C and ^{15}N isotopically enriched proteins for subsequence characterization with NMR. Even though these efforts were fruitful, future investigations will require further optimization of protein production techniques, including the development of new, general systems for producing isotopically enriched proteins for NMR investigations. For particular research targets, selective labeling of amino acids and methyl groups could help expedite NMR data collection and analysis, especially for dynamic measurements of disordered regions in LPMOs, such as the linker[121]. Similarly, the study of large multidomain LPMOs by NMR could be facilitated by perdeuteration, where the substitution of ^{1}H for ^{2}H would lead to increased magnetization lifetimes of ^{13}C nuclei, by diminishing the effects of scalar and dipolar couplings on relaxation times[122]. Moreover, unnatural amino acids and post-translational modifications could be incorporated by repurposing the seldom used TAG stop codon in *E. coli*[123].

In addition to selective-labeling techniques during protein sample production, alternative NMR methods could facilitate the sequence-specific assignment procedure. MUSIC-type experiments[124], for example, could allow resonances from specific amino acids to be selected, simplifying NMR datasets. Likewise, NMR experiments based on direct detection of ^{13}C-resonances (e.g. MOCCA-XY16[125]) and ^{15}N-resonances[126], could

provide additional possibilities to map Pro-rich linker regions in LPMOs. Lastly, while the current software for fully automatic assignment does not appear to attain the same level of accuracy and quality as a semi-automated, supervised approach, it is important to continue following technological developments. A suitable unsupervised approach may ultimately take advantage of higher-dimensional datasets as input, e.g. 7D HNCO(N)CACONH[127], correlating $H^{N}_{(i)}$, $N_{(i)}$ with all of the backbone atoms in *i*-1, and $C'_{(i-2)}$. These kind of spectra would have to be recorded using fast-acquisition methods such as non-uniform sampling (NUS)[128], and would contain virtually no signal overlap, facilitating automatic peak picking and assembly of the assignment.

Chemical shift assignments (**Papers V–VIII**) were the starting points for the three-dimensional solution structure determination of *Bl*LPMO10A, *Sc*CBM2, *Cj*CBM5 and *Cj*CBM73. Resonance assignments were also the basis for characterizing the secondary structure content of all proteins, and made it possible to combine NMR data with three-dimensional structures of LPMOs that had been determined by X-ray crystallography. This allowed the focus to be on the characterization of functional aspects such as ligand binding and dynamics, instead of full structure elucidation. Despite the contributions of NMR towards obtaining structural insights of LPMOs, X-ray structures are more accurate in the atomic level, especially regarding the coordination geometry of the copper in the LPMO active site. However, for smaller proteins such as CBMs, as well as for disordered regions of LPMOs (i.e. the linker), NMR still is an important technique that will continue contributing to understanding these structural aspects. Moreover, NMR (combined with EPR spectroscopy) has been able to provide insights into interactions of LPMOs with copper in solution, both through the PRE effect of Cu(II) and by studying an oxygen-free Cu(I) sample (**Paper IV)**.

Functional insights regarding substrate interaction have been possible due to the ability of individual LPMOs to bind soluble substrates (**Paper I**) or the availability of custom-made substrates such as nanofibrilated cellulose and chitin (**Papers III and IV**). The latter have allowed interactions between highly dispersed solid substrates and LPMOs to be probed by solution NMR. However, insoluble substrates like cellulose and chitin are interesting targets for solid-state NMR investigations. Until recently, mapping residue-specific interactions between a protein and a solid substrate by measuring ^{1}H-^{15}N

chemical shifts in solid-state NMR spectra would have been an insurmountable challenge due to the broadness of the signals. Fortunately, technological developments have resulted in ultrafast spinning (> 100 kHz) solid-state probes. Here, protein resonances become almost as narrow as in solution NMR, making it possible to record ^{15}N-HSQC-like spectra in solid-state[129], thus enabling potential investigations of interactions between LPMO and their natural substrates in solid-state.

NMR experiments used in this thesis to derive dynamic data for LPMOs probed motions in the ps-ns time scale, and no slower motions were detected for *apo*-LPMOs in the absence of substrate. However, it is conceivable that LPMOs could partake in interactions occurring at slower time-scales (µs-ms time scale), which could be measured by using relaxation dispersion NMR experiments[130,131]. Moreover, it would be interesting to evaluate the extent to which NMR may be used to probe electron transfer kinetics between CDH (or other electron donors) and LPMOs (**Paper I**), as this process is expected to occur within NMR-observable time-scales[132]. For characterizing global LPMO dynamics, NMR techniques such as residual dipolar couplings (RDCs) and PRE labels could be used to analyze the overall conformation and relative orientations of catalytic and carbohydrate-binding domains in modular LPMOs[133]. Alternatives to NMR methods could be fluorescence spectroscopy/FRET (Förster resonance energy transfer) between fluorophores attached to the CBM and LPMO domains[134], small angle X-ray scattering (SAXS) to characterize the overall shape of the modular proteins[135], and atomic force microscopy (AFM) to visualize the movement of LPMOs on crystalline substrate surfaces[136].

The first steps towards understanding how LPMO structure, dynamics and function are related, using *Sc*LPMO10C as a model system, relied on structural information from NMR and X-ray crystallography, supplemented with NMR dynamic data and biochemical assays. Special focus was placed on understanding how the flexibility of the linker region, together with CBM2, determines the product profile of the LPMO domain in *Sc*LPMO10C (**Paper III**). Findings support the idea that CBMs and linkers play a key role in regulating the activity of catalytic domains, as has been suggested previously[84,137]. A model of cellulose oxidation by the modular LPMO that considered both dynamics and biochemistry was constructed. However, obtaining the complete picture of cellulose

oxidation by LPMOs will require integration of several kind of data including NMR, X-ray crystallography, biophysical techniques and computational modelling[138].

Overall, this thesis has provided new insights regarding several key aspects of LPMO functionality. However, paraphrasing Erwin T. Reese[26], from our point of view a number of points regarding the nature of LPMOs require further elucidation. Why do some LPMOs have attached linkers and CBMs while others exist as isolated catalytic domains? What are the dynamic features of linkers in LPMOs and why do they have different compositions and lengths? What are the different biological roles that individual LPMOs play in an organism such as *N. crassa*, which contains 14 different LPMO genes? What is the mode of action of LPMOs, and what are the kinetics of each of the events (substrate binding, electron transfer, oxidation)? These are the questions to which this thesis only begins to give partial answers.

References

1. U. S. Department of Energy. *Biofuels: Bringing biological solutions to energy challenges*. *Biofuels Bulletin by U.S. DoE Office of Science* (2007).

2. Soimakallio, S., Antikainen, R. & Thun, R. *Assessing the sustainability of liquid biofuels from evolving technologies: A Finnish approach*. *VTT Tiedotteita - Research Notes* (2009).

3. Chilvers, A. & Jeswani, H. *Sustainability of liquid biofuels*. *Royal Academy of Engineering* (2017).

4. Sheldon, R. A. Green chemistry, catalysis and valorization of waste biomass. *J. Mol. Catal. A Chem.* **422,** 3–12 (2016).

5. Joseleau, J.-P. & Perez, S. Cellulose, Pectins. *GlycoPedia* (2018). at <https://www.glycopedia.eu/Cellulose-Pectins>

6. Beckham, G. T. *et al.* Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *J. Phys. Chem. B* **115,** 4118–4127 (2011).

7. Mihranyan, A. Cellulose from cladophorales green algae: From environmental problem to high-tech composite materials. *J. Appl. Polym. Sci.* **119,** 2449–2460 (2011).

8. Rinaudo, M. Chitin and chitosan: properties and applications. *Prog. Polym. Sci.* **31,** 603–632 (2006).

9. Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61,** 263–289 (2010).

10. Bomble, Y. J. *et al.* Lignocellulose deconstruction in the biosphere. *Curr. Opin. Chem. Biol.* **41,** 61–70 (2017).

11. Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C. H. & Nah, J.-W. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. *J. Polym. Sci. Part A Polym. Chem.* **42,** 3423–3432 (2004).

12. Harris, P. V, Xu, F., Kreel, N. E., Kang, C. & Fukuyama, S. New enzyme insights drive advances in commercial ethanol production. *Curr. Opin. Chem. Biol.* **19,** 162–170 (2014).

13. Hottle, T. A., Bilec, M. M. & Landis, A. E. Biopolymer production and end of life comparisons using life cycle assessment. *Resour. Conserv. Recycl.* **122,** 295–306 (2017).

14. Fernandez, J. G. & Ingber, D. E. Manufacturing of large-scale functional objects using biodegradable chitosan bioplastic. *Macromol. Mater. Eng.* **299,** 932–938 (2014).

15. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research* **37,** 233–238 (2009).

16. Kont, R., Kari, J., Borch, K., Westh, P. & Väljamäe, P. Inter-domain synergism is required for efficient feeding of cellulose chain into active site of cellobiohydrolase Cel7A. *J. Biol. Chem.* **291,** 26013–26023 (2016).

17. Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **275,** 309–325 (1998).

18. Kraulis, J. *et al.* Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28,** 7241–7257 (1989).

19. Lee, T. M., Mary, F. F., Arnold, F. H. & Mayo, S. L. A structural study of *Hypocrea jecorina* Cel5A. *Protein Sci.* **20,** 1935–1940 (2011).

20. Forsberg, Z. *et al.* Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 8446–8451 (2014).

21. Smith, M. A., Philip, R. A., Wu, T., Brustad, E. M. & Arnold, F. H. Chimeragenesis of distantly- related proteins by noncontiguous recombination. *Protein Sci.* **22,** 231–238 (2012).

22. Montenecourt, B. S. *Trichoderma reesei* cellulases. *Trends Biotechnol.* **1,** 156–161 (1983).

23. Igarashi, K. *et al.* Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333,** 1279–1282 (2011).

24. Johnson, E. Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuels, Bioprod. Biorefining* **10,** 164–174 (2016).

25. Vaaje-Kolstad, G., Horn, S. J., Sørlie, M. & Eijsink, V. G. H. The chitinolytic machinery of *Serratia marcescens*-a model system for enzymatic degradation of recalcitrant polysaccharides. *FEBS J.* **280,** 3028–3049 (2013).

26. Reese, E., Siu, R. & Levinson, H. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J. Bacteriol.* **59,** 485–497 (1950).

27. Eriksson, K. & Pettersson, B. Oxydation: an inportant enzyme reaction in fungal degradation of cellulose. *FEBS Lett* **49,** 282–285 (1974).

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

29. Vaaje-Kolstad, G. *et al.* An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330,** 219–222 (2010).

30. Harris, P. V *et al.* Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic

family. *Biochemistry* **49,** 3305–3316 (2010).

31. Horn, S. J. *et al.* Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 18089–18094 (2006).

32. Karkehabadi, S. *et al.* The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 Å resolution. *J. Mol. Biol.* **383,** 144–154 (2008).

33. Forsberg, Z. *et al.* Cleavage of cellulose by a CBM33 protein. *Protein Sci.* **20,** 1479–1483 (2011).

34. Quinlan, R. J. *et al.* Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 15079–15084 (2011).

35. Phillips, C. M., Beeson, W. T., Cate, J. H. & Marletta, M. A. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem. Biol.* **6,** 1399–1406 (2011).

36. Langston, J. A. *et al.* Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* **77,** 7007–7015 (2011).

37. Eibinger, M. *et al.* Cellulose surface degradation by a lytic polysaccharide monooxygenase and its effect on cellulase hydrolytic efficiency. *J. Biol. Chem.* **289,** 35929–35938 (2014).

38. Vermaas, J. V., Crowley, M. F., Beckham, G. T. & Payne, C. M. Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. *J. Phys. Chem. B* **119,** 6129–6143 (2015).

39. Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. H. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* **5,** 45–56 (2012).

40. Nakamura, A. *et al.* Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. *J. Am. Chem. Soc.* **136,** 4584–4592 (2014).

41. Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K. & Westh, P. Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: Effects of substrate structure and carbohydrate binding domain. *Biochemistry* **52,** 8938–8948 (2013).

42. Kurašin, M. & Väljamäe, P. Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* **286,** 169–177 (2011).

43. Johansen, K. S. Discovery and industrial applications of lytic polysaccharide mono-oxygenases. *Biochem. Soc. Trans.* **44,** 143–149 (2016).

44. Isaksen, T. *et al.* A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *J. Biol. Chem.* **289,** 2632–2642 (2014).

45. Frandsen, K. E. H. *et al.* The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat. Chem. Biol.* **12,** 298–303 (2016).

46. Agger, J. W. *et al.* Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6287–6292 (2014).

47. Kojima, Y. *et al.* A lytic polysaccharide monooxygenase with broad xyloglucan specificity from the brown-rot fungus *Gloeophyllum trabeum* and its action on cellulose-xyloglucan complexes. *Appl. Environ. Microbiol.* **82,** 6557–6572 (2016).

48. Bennati-Granier, C. *et al.* Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. *Biotechnol. Biofuels* **8,** 90–103 (2015).

49. Couturier, M. *et al.* Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nat. Chem. Biol.* **14,** 306–310 (2018).

50. Frommhagen, M. *et al.* Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnol. Biofuels* **8,** 101–113 (2015).

51. Lo Leggio, L. *et al.* Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nat. Commun.* **6,** 5961–5969 (2015).

52. Vu, V. V, Beeson, W. T., Span, E. A., Farquhar, E. R. & Marletta, M. A. A family of starch-active polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 13822–13827 (2014).

53. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6,** 41–64 (2013).

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

55. Sabbadin, F. *et al.* An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat. Commun.* **9,** 756–767 (2018).

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

57. Frandsen, K. E. H. & Lo Leggio, L. Lytic polysaccharide monooxygenases: A crystallographer's view on a new class of biomass-degrading enzymes. *IUCrJ* **3,** 448–467 (2016).

58. Moser, F., Irwin, D., Chen, S. & Wilson, D. B. Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng.* **100,** 1066–1077 (2008).

59. Peisach, J. & Blumberg, W. E. Structural implications derived from the analysis of electron paramagnetic resonance spectra of natural and artificial copper proteins. *Arch. Biochem. Biophys.* **165,** 691–708 (1974).

60. Aachmann, F. L., Sørlie, M., Skjåk-Bræk, G., Eijsink, V. G. H. & Vaaje-Kolstad, G. 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,** 18779–18784 (2012).

61. Li, X., Beeson, W. T., Phillips, C. M., Marletta, M. A. & Cate, J. H. D. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* **20,** 1051–1061 (2012).

62. Beeson, W. T., Phillips, C. M., Cate, J. H. D. & Marletta, M. A. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J. Am. Chem. Soc.* **134,** 890–892 (2012).

63. Borisova, A. S. *et al.* Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity. *J. Biol. Chem.* **290,** 22955–22969 (2015).

64. Walton, P. H. & Davies, G. J. On the catalytic mechanisms of lytic polysaccharide monooxygenases. *Curr. Opin. Chem. Biol.* **31,** 195–207 (2016).

65. Bissaro, B. *et al.* Oxidative cleavage of polysaccharides by monocopper enzymes depends on H_2O_2. *Nat. Chem. Biol.* **13,** 1123–1128 (2017).

66. Hangasky, J. A., Iavarone, A. T. & Marletta, M. A. Reactivity of O_2 versus H_2O_2 with polysaccharide monooxygenases. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1801153115

67. Tan, T.-C. *et al.* Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nat. Commun.* **6,** 7542–7552 (2015).

68. Westereng, B. *et al.* Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Sci. Rep.* **5,** 18561–18577 (2015).

69. Kracher, D. *et al.* Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science* **352,** 1098–1101 (2016).

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

71. Kjærgaard, C. H. *et al.* Spectroscopic and computational insight into the activation of O_2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 8797–8802 (2014).

72. Nakagawa, Y. S., Eijsink, V. G. H., Totani, K. & Vaaje-Kolstad, G. Conversion of α-chitin substrates with varying particle size and crystallinity reveals substrate preferences of the chitinases and lytic polysaccharide monooxygenase of *Serratia marcescens*. *J. Agric. Food Chem.* **61,** 11061–11066 (2013).

73. Hu, J., Arantes, V. & Pribowo, A. Substrate factors that influence the synergistic interaction of AA9 and cellulases during the enzymatic hydrolysis of biomass. *Energy Environ. Sci.* **7,** 2308–2315 (2014).

74. Hudson, K. L. *et al.* Carbohydrate-aromatic interactions in proteins. *J. Am. Chem. Soc.* **137,** 15152–15160 (2015).

75. Beeson, W. T., Vu, V. V, Span, E. A., Phillips, C. M. & Marletta, M. A. Cellulose degradation by polysaccharide monooxygenases. *Annu. Rev. Biochem.* **84,** 923–946 (2015).

76. McLean, B. W. *et al.* Carbohydrate-binding modules recognize fine substructures of cellulose. *J. Biol. Chem.* **277,** 50245–50254 (2002).

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

78. Forsberg, Z. *et al.* Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* **53,** 1647–1656 (2014).

79. Forsberg, Z. *et al.* Structural and functional analysis of a lytic polysaccharide monooxygenase important for efficient utilization of chitin in *Cellvibrio japonicus*. *J. Biol. Chem.* **291,** 7300–7312 (2016).

80. Crouch, L. I., Labourel, A., Walton, P. H., Davies, G. J. & Gilbert, H. J. The contribution of non-catalytic carbohydrate binding modules to the activity lytic polysaccharide monooxygenases. *J. Biol. Chem.* **291,** 7439–7449 (2016).

81. Forsberg, Z. *et al.* Structural determinants of bacterial lytic polysaccharide monooxygenase functionality. *J. Biol. Chem.* **293,** 1397–1412 (2018).

82. Xu, G. Y. *et al.* Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34,** 6993–7009 (1995).

83. Brun, E. *et al.* Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. *Biochemistry* **36,** 16074–16086 (1997).

84. Crasson, O. *et al.* Human chitotriosidase: catalytic domain or carbohydrate binding module, who's leading HCHT's biological function. *Sci. Rep.* **7,** 2768–2777 (2017).

85. Petrović, D. M. *et al.* Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation. *Protein Sci.* (2018). doi:10.1002/pro.3451

86. Kornberg, A. Ten commandments: Lessons from the enzymology of DNA replication. *J. Bacteriol.* **182,** 3613–3618 (2000).

87. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*. *Protein NMR Spectroscopy: Principles and Practice* (Elsevier Academic Press, Amsterdam, 2007).

88. Wüthrich, K. *NMR of Proteins and Nucleic Acids*. (John Wiley & Sons, New York, 1986).

89. Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. (John Wiley & Sons, 2001).

90. Friebolin, H. *Basic one- and two-dimensional NMR spectroscopy*. (WILEY-VCH Verlag GmbH, 2011).

91. Kay, L. E., Torchia, D. & Bax, A. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* **28,** 8972–8979 (1989).

92. Fushman, D. Determination of protein dynamics using ^{15}N relaxation measurements. *BioNMR Drug Res.* **7,** 1–28 (2002).

93. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113,** 5490–5492 (1991).

94. Güntert, P., Mumenthaler, C. & Herrmann, T. *DYANA User's Manual*. (Institut für Molekularbiologie und Biophysik, ETH, Zürich, 1998).

95. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56,** 227–241 (2013).

96. Zhang, H., Neal, S. & Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25,** 173–95 (2003).

97. Keller, R. *The Computer Aided Resonance Assignment Tutorial*. (CANTINA Verlag, Goldau, Switzerland, 2004).

98. Kovacs, H. *Avance 3D / Triple Resonance Manual*. (Bruker Biospin, Fällanden, 2003).

99. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. D. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. **14,** 496–514 (1990).

100. Aramini, J. M. NMR determined Rotational correlation time. *NESG Wiki* (2011). at <http://www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time>

101. Facey, G. University of Ottawa NMR Facility Blog. *University of Ottawa* (2018). at <http://u-of-o-nmr-facility.blogspot.no/>

102. Parella, T. Bruker NMR Guide and Encyclopedia. (2003). at <http://triton.iqfr.csic.es/guide/tutorials/tutorial.html>

103. Reich, H. J. Structure Determination Using Spectroscopic Methods. (2018). at <https://www.chem.wisc.edu/areas/reich/nmr/>

104. Keeler, J. Lectures by James Keeler. (2014). at <http://www-keeler.ch.cam.ac.uk/lectures/>

105. Koradi, R., Billeter, M., Engeli, M., Güntert, P. & Wüthrich, K. Automated Peak Picking and Peak Integration in Macromolecular NMR Spectra Using AUTOPSY. *J. Magn. Reson.* **135,** 288–297 (1998).

106. Lee, W., Tonelli, M., Dashti, H., Eghbalnia, H. & Markley, J. L. I-PINE web server, an integrative probabilistic NMR assignment system. *In preparation* (2018). at <http://i-pine.nmrfam.wisc.edu>

107. Wisconsin-Madison, U. of. Biological Magnetic Resonance Data Bank - Full Chemical Shift Statistics. (2018). at <http://www.bmrb.wisc.edu/ref_info/statsel.htm#20>

108. Marsh, J. a, Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* **15,** 2795–2804 (2006).

109. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. *J. Biomol. NMR* **24,** 149–154 (2002).

110. Güntert, P. Automated NMR Structure Calculation With CYANA. *Methods Mol. Biol.* **278,** 353–378 (2004).

111. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins Struct. Funct. Bioinforma.* **47,** 393–402 (2002).

112. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

113. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinforma.* **77,** 114–122 (2009).

114. Teilum, K., Kunze, M. B. A., Erlendsson, S. & Kragelund, B. B. (S)Pinning down protein interactions by NMR. *Protein Sci.* **26,** 436–451 (2017).

115. East, K., Liptak, C. & Loria, J. P. Template specific fidelity of DNA polymerase β characterized by interactions with the templating bases and a base specific mutator. *Abstr. 59th Exp. Nucl. Magn. Reson. Conf.* 50 (2018).

116. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73,** 1–16 (2013).

117. Ulrich, E. L. & Markley, J. L. Blue-copper proteins: nuclear magnetic resonance investigations. *Coord. Chem. Rev.* **27,** 109–140 (1978).

118. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402,** 47–52 (1999).

119. Gudmundsson, M. *et al.* Structural and electronic snapshots during the transition from a Cu(II) to Cu(I) metal center of a lytic polysaccharide monooxygenase by X-ray photoreduction. *J. Biol. Chem.* **289,** 18782–92 (2014).

120. Berlin, K., Longhini, A., Dayie, T. K. & Fushman, D. Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J. Biomol. NMR* **57,** 333–352 (2013).

121. Tanio, M., Tanaka, R., Tanaka, T. & Kohno, T. Amino acid-selective isotope labeling of proteins for nuclear magnetic resonance study: Proteins secreted by Brevibacillus choshinensis. *Anal. Biochem.* **386,** 156–160 (2009).

122. Venters, R. A., Ii, B. T. F., Fierke, C. A. & Spicer, L. D. Characterizing the use of perdeuteration in NMR studies of large proteins : ^{13}C , ^{15}N and ^{1}H assignments of human carbonic anhydrase II. **264,** 1101–1116 (1996).

123. Wals, K. & Ovaa, H. Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2,** 1–12 (2014).

124. Schubert, M., Oschkinat, H. & Schmieder, P. MUSIC, selective pulses, and tuned delays: Amino acid type-selective ^{1}H-^{15}N correlations. *J. Magn. Reson.* **148,** 61–72 (2001).

125. Felli, I. C., Pierattelli, R., Glaser, S. J. & Luy, B. Relaxation-optimised Hartmann-Hahn transfer using a specifically tailored MOCCA-XY16 mixing sequence for carbonyl-carbonyl correlation spectroscopy in ^{13}C direct detection NMR experiments. *J. Biomol. NMR* **43,** 187–196 (2009).

126. Chhabra, S. *et al.* ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1717560115

127. Zerko, S. & Koźmiński, W. Six- and seven-dimensional experiments by combination of sparse random sampling and projection spectroscopy dedicated for backbone resonance assignment of intrinsically disordered proteins. *J. Biomol. NMR* **63,** 283–290 (2015).

128. Kazimierczuk, K. & Orekhov, V. Non-uniform sampling: Post-Fourier era of NMR data collection and processing. *Magn. Reson. Chem.* **53,** 921–926 (2015).

129. Xue, K. *et al.* Limits of resolution and sensitivity of proton detected MAS solid-state NMR experiments at 111 kHz in deuterated and protonated proteins. *Sci.*

Rep. **7,** 7444–7450 (2017).

130. Kleckner, I. R. & Foster, M. P. An Introduction to NMR based approaches for measuring protein dynamics. *Biochim Biophys Acta* . **1814,** 942–968 (2011).

131. Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A. & Kay, L. E. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.* **124,** 10743–10753 (2002).

132. Zhuravleva, A. V *et al.* Gated electron transfers and electron pathways in azurin: a NMR dynamic study at multiple fields and temperatures. *J. Mol. Biol.* **342,** 1599–611 (2004).

133. Braddock, D. T., Cai, M., Baber, J. L., Huang, Y. & Clore, G. M. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J. Am. Chem. Soc.* **123,** 8634–8635 (2001).

134. Pham, E., Chiang, J., Li, I., Shum, W. & Truong, K. A computational tool for designing FRET protein biosensors by rigid-body sampling of their conformational space. *Structure* **15,** 515–523 (2007).

135. Schmuck, M., Pilz, I., Hayn, M. & Esterbauer, H. Investigation of cellobiohydrolase from trichoderma reesei by small angle X-ray scattering. *Biotechnol. Lett.* **8,** 397–402 (1986).

136. Eibinger, M., Sattelkow, J., Ganner, T., Plank, H. & Nidetzky, B. Single-molecule study of oxidative enzymatic deconstruction of cellulose. *Nat. Commun.* **8,** 894–900 (2017).

137. Papaleo, E. *et al.* The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* **116,** 6391–6423 (2016).

138. Bonomi, M. & Camilloni, C. Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* **33,** 3999–4000 (2017).

Appendix A – Additional contribution I

Contribution to characterization of chitin binding by human macrophage chitotriosidase CBM14 (ref. 84). Here, we assigned the Trp465 and Asn466, using them as reporters for the interaction by measuring chemical shift changes in the H^N and N atoms (^{15}N-HSQC) upon titration of chitin trimer ($GlcNAc_3$). A compound change in chemical shift, $\Delta\delta comp$ (in ppm) was calculated using the following formula: $\Delta\delta comp = [(\Delta\delta_H)^2 + (\Delta\delta_N/x)^2]^{1/2}$. $\Delta\delta_H$ is the change in chemical shift of the amide proton (ppm), $\Delta\delta_N$ is the change in chemical shift of the amide nitrogen (ppm), and x is a constant used to achieve equal contributions from changes in N and H^N shifts. $\Delta\delta comp$ was used to estimate the dissociation constant (K_d), as shown in Figure A1.

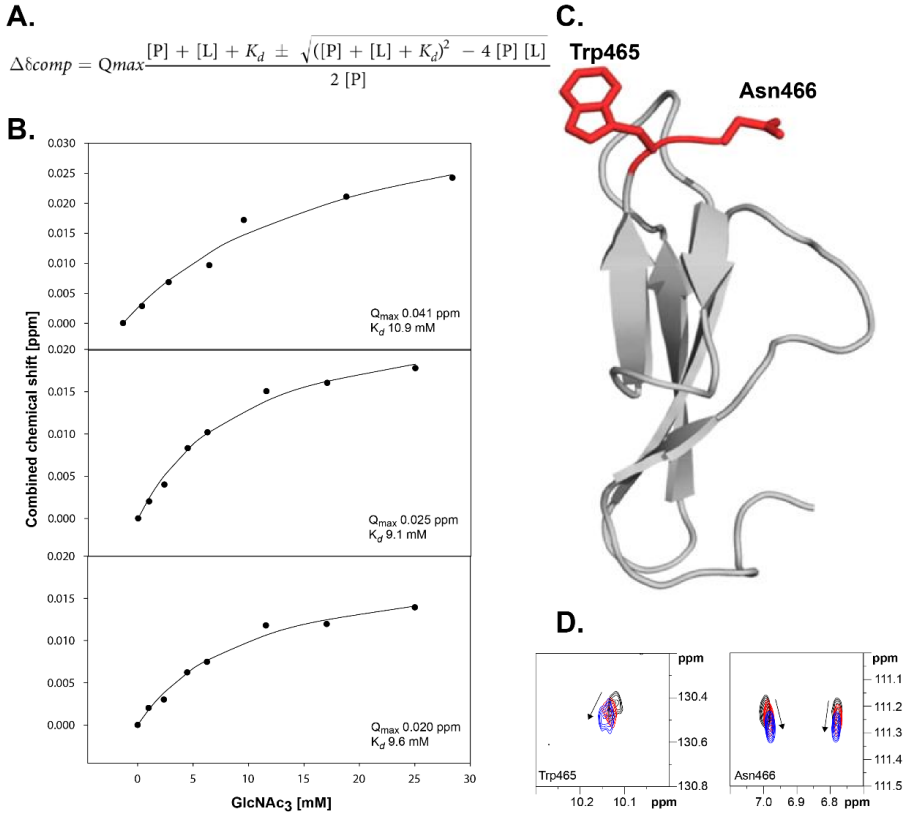

Figure A.1. CBM14 interaction with chitin trimer ($GlcNAc_3$). (A) Relationship between the K_d, $\Delta\delta comp$, protein concentration, [P], and ligand concentration [L]. (B) Combined chemical shift for Trp465 (top panel) and Asn466 (two lower panels) upon titration with GlcNAc3. Kd was estimated for individual atom pairs, as well as Qmax (Δδcomp at full saturation). (C) Binding surface mapped on the structure of CBM14 (PDB ID: 5HBF), where the side chains Trp465 and Asn466 are showed in red. (D) Overlay of an area of interest from the ^{15}N-HSQC spectrum for CBM14 (black) in the presence of 4.5 mM (red) and 25 mM $GlcNAc_3$ (blue). The arrows indicate direction of the change in chemical shift upon titration.

Appendix B – Additional contribution II

In nature, fungal LPMOs undergo several post-translational modifications including signal peptide cleavage, disulfide bond formation and τ-methylation of the N-terminal histidine. Remarkably, the role of the τ-methylation (i.e.methylation of Nε2) of the catalytically crucial N-terminal His remains unclear. In order to investigate the effect methylation and metal-binding (using Zn(II), a non-paramagnetic analogue of Cu(II)) on the pK_a of the N-terminal His, NMR pH titrations of TaLPMO9A were performed in four different forms of the enzyme (see Figure B1). Results showed the pK_a values of the N-terminal His in *Ta*LPMO9A are hardly affected by methylation, regardless of whether a metal ion (Zn(II)) is bound. See ref. 85 for details.

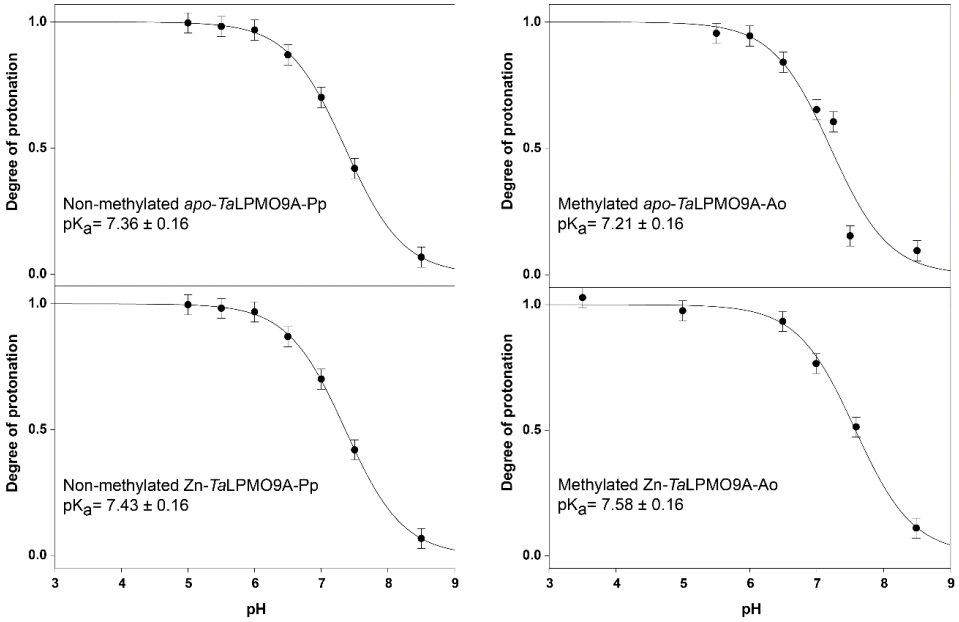

Figure B1. pH titrations of the N-terminal His of *Ta*LPMO9A. The titration points are shown together with the fitted titration curve for each of the four forms of the enzyme. Pp and Ao indicate whether the LPMO was produced in *Pichia pastoris* or *Aspergillus oryzae*, respectively. The titration curves correspond to a rearrangement of the Henderson-Hasselbach equation $\delta = \frac{10^{(pH-pKa)}\delta_B+\delta_{BH}}{1+10^{(pH-pKa)}}$,where δ_{BH} is the chemical shift of the fully protonated histidine, δ_B is the chemical shift of the fully deprotonated histidine, and δ is the measured chemical shift at a certain pH.

Appendix C – INEPT – Product operator formalism

INEPT (Insensitive Nucleus Enhancement by Polarization Transfer; Figure C1) is one of the most essential building blocks in pulse sequences for protein NMR (e.g. HSQC spectra and multidimensional spectra for resonance assignment). It consists on transferring magnetization, through scalar couplings, from a sensitive nucleus (^{1}H) to an insensitive one (^{15}N or ^{13}C). Vector representation of pulse sequences such as the ones used in Figures 3.1 and 3.2 can be useful to explain simple NMR experiments, but a proper explanation of INEPT requires using operators to describe the different spin states. Operators describe the state of a spin, and can be used to analyze the consequences of RF pulses on the spin state, and how chemical shifts and scalar couplings arise (evolve) as a function of time.

- The operators I_x , I_y and I_z represent the x-,y- and z-components of the magnetization of ^{1}H, similarly S_x , S_y and S_z represent the components of the magnetization of ^{15}N. Only x,y-magnetization is observable.
- A pulse with flip angle β in the x-direction is described by the operator βI_x.
- Chemical shift (Ω) evolution of during time τ is described by the operator $\Omega\tau I_z$.
- The scalar coupling between I and S during time τ is described by the operator $2\pi\tau J_{IS} I_z S_z$.

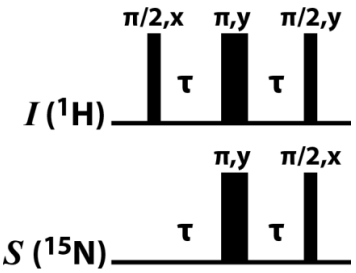

Figure C1. INEPT pulse scheme. Magnetization starts on *I* and is transferred to *S* by a set of π/2 (90°) and π (180°) pulses applied on x- or y-directions.

The state is at equilibrium, with only magnetization the z-axis of *I.* After a π/2 pulse in the x-direction on *I*, followed by J-coupling evolution during τ, the state of the system is:

$$I_z \xrightarrow{\left(\frac{\pi}{2}\right)I_x} -I_y \xrightarrow{2\pi\tau J_{IS} I_z S_z} -I_y \cos(\pi\tau J_{IS}) + 2I_x S_z sin(\pi\tau J_{IS})$$

After the first chemical shift evolution during τ:

$$\xrightarrow{\boldsymbol{\Omega\tau I_z}} -I_y \cos(\pi\tau J_{IS}) \cos(\Omega\tau) + I_x \cos(\pi\tau J_{IS}) \cos(\Omega\tau) + 2I_x S_z \sin(\pi\tau J_{IS}) \cos(\Omega\tau) + 2I_y S_z \sin(\pi\tau J_{IS})$$

Two simultaneous π pulses in the y-direction on both *I* and *S* result in two of the terms changing sign:

$$\xrightarrow{\boldsymbol{\pi I_y + \pi S_y}} -I_y \cos(\pi\tau J_{IS}) \cos(\Omega\tau) - I_x \cos(\pi\tau J_{IS}) \cos(\Omega\tau) + 2I_x S_z \sin(\pi\tau J_{IS}) \cos(\Omega\tau) - 2I_y S_z \sin(\pi\tau J_{IS})$$

After the second chemical shift evolution during τ, the sign inversion caused by the π pulses lead to cancelling of the chemical shift contributions $cos(\Omega\tau I_z)$ and $sin(\Omega\tau I_z)$.

$$\xrightarrow{\boldsymbol{\Omega\tau I_z}} -I_y \cos(\pi\tau J_{IS}) + 2I_x S_z \sin(\pi\tau J_{IS})$$

Following the second J-coupling evolution during τ:

$$\xrightarrow{\boldsymbol{2\pi\tau J_{IS} I_z S_z}} I_y \cos(\pi\tau J_{IS})^2 + 4I_x S_z \sin(\pi\tau J_{IS}) \cos(\pi\tau J_{IS}) + I_y \sin(\pi\tau J_{IS})^2$$

Finally, after the two π/2 pulses in the y-direction (*I*) and the x-direction (*S*), magnetization is transferred to the *S* spin [S_y in the term $4I_z S_y \sin(\pi\tau J_{IS}) \cos(\pi\tau J_{IS})$]:

$$\xrightarrow{\boldsymbol{\frac{\pi}{2} I_y + \frac{\pi}{2} S_x}} -I_y \cos(\pi\tau J_{IS})^2 + 4I_z S_y \sin(\pi\tau J_{IS}) \cos(\pi\tau J_{IS}) + I_y \sin(\pi\tau J_{IS})^2$$

If $\tau = 1/(4J_{IS})$, the *I* magnetization disappears altogether, yielding only the term $2I_z S_y$, which corresponds to observable ^{15}N magnetization.

Paper I

Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase

Gaston Courtade[a], Reinhard Wimmer[b], Åsmund K. Røhr[c], Marita Preims[d], Alfons K. G. Felice[d], Maria Dimarogona[e], Gustav Vaaje-Kolstad[c], Morten Sørlie[c], Mats Sandgren[e], Roland Ludwig[d], Vincent G. H. Eijsink[c,1], and Finn Lillelund Aachmann[a,1]

[a]Norwegian Biopolymer Laboratory (NOBIPOL), Department of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway; [b]Department of Chemistry and Bioscience, Aalborg University, DK-9220 Aalborg Ø, Denmark; [c]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, N-1432 Ås, Norway; [d]Food Biotechnology Laboratory, Department of Food Science and Technology, Vienna Institute of Biotechnology, University of Natural Resources and Life Sciences, Vienna A-1190, Austria; and [e]Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

Edited by Arnold L. Demain, Drew University, Madison, NJ, and approved April 11, 2016 (received for review February 15, 2016)

Lytic polysaccharide monooxygenases (LPMOs) are copper-dependent enzymes that catalyze oxidative cleavage of glycosidic bonds using molecular oxygen and an external electron donor. We have used NMR and isothermal titration calorimetry (ITC) to study the interactions of a broad-specificity fungal LPMO, *Nc*LPMO9C, with various substrates and with cellobiose dehydrogenase (CDH), a known natural supplier of electrons. The NMR studies revealed interactions with cellohexaose that center around the copper site. NMR studies with xyloglucans, i.e., branched β-glucans, showed an extended binding surface compared with cellohexaose, whereas ITC experiments showed slightly higher affinity and a different thermodynamic signature of binding. The ITC data also showed that although the copper ion alone hardly contributes to affinity, substrate binding is enhanced for metal-loaded enzymes that are supplied with cyanide, a mimic of O_2^-. Studies with CDH and its isolated heme *b* cytochrome domain unambiguously showed that the cytochrome domain of CDH interacts with the copper site of the LPMO and that substrate binding precludes interaction with CDH. Apart from providing insights into enzyme–substrate interactions in LPMOs, the present observations shed new light on possible mechanisms for electron supply during LPMO action.

lytic polysaccharide monooxygenase | LPMO | cellulose | xyloglucan | cellobiose dehydrogenase

The polysaccharides in chitinous and lignocellulosic biomass compose large sources of organic carbon and are attractive substrates in biorefineries for the production of biofuels and value-added products. However, the exploitation of these resources is hindered by polysaccharide recalcitrance, which hampers enzymatic depolymerization.

Traditionally, it was thought that hydrolytic enzymes were solely responsible for the degradation of chitin and cellulose. A fundamental change in this model was triggered by the discovery of copper-dependent redox enzymes today known as lytic polysaccharide monooxygenases (LPMOs) (1–8). LPMOs are abundantly present in biomass-degrading microbes and make use of molecular oxygen and an external electron donor to cleave polysaccharides through hydroxylation of one of the carbons in the scissile glycosidic bond (4, 5, 9–13). LPMOs can accept electrons from cellobiose dehydrogenase (CDH) (3, 14, 15) or a variety of small molecule reducing agents such as ascorbate and gallic acid (4, 5) as well as lignin-derived redox mediators (16). Each LPMO reaction cycle is postulated to consume two electrons (3, 5, 6).

Due to the potentially major role of LPMOs in enzymatic biomass conversion and the uniqueness of their catalytic power, there is great interest in unraveling the molecular basis of LPMO activity. This interest has resulted in the discovery and characterization of several LPMOs, currently classified in auxiliary activity (AA) families 9, 10, 11, and 13 in the carbohydrate-active enzymes (CAZy) database (17–19). Bacterial LPMOs occur in family AA10, whereas LPMOs of fungal origin belong to families AA9, AA11, and AA13. LPMOs show large diversity, in terms of domain and sequence composition, as well as in terms of substrate specificity, oxidative regioselectivity, and product profiles. To date, LPMO activity has been demonstrated for β-1,4 glycosidic bonds in chitin (5), cellulose (2), soluble cellulose oligosaccharides (20), hemicelluloses (21), and xylan (22), as well as α-1,4 glycosidic bonds in starch (23). LPMOs acting on β-1,4-linked glucans oxidize either C1 or C4 or show mixed oxidative regioselectivity leading to the formation of both C1 and C4 oxidized products.

*Nc*LPMO9C (also known as NCU02916 or *Nc*GH61-3) is a two-domain C4-oxidizing AA9 LPMO from *Neurospora crassa* that is active on β-1,4 glycosidic bonds in cellulose, cellodextrins, and β-glucan hemicelluloses (20, 21, 24, 25). In addition to its catalytic domain, *Nc*LPMO9C contains a carbohydrate-binding module belonging to the CBM1 family. The recently published crystal structure of the catalytic domain of *Nc*LPMO9C (26) displays a typical core LPMO structure: two β-sheets forming a

Significance

Copper-dependent lytic polysaccharide monooxygenases (LPMOs) are key players in the enzymatic conversion of biomass. LPMOs catalyze oxidative cleavage of glycosidic bonds in a process involving molecular oxygen and an electron donor, such as cellobiose dehydrogenase (CDH). Using protein NMR and isothermal titration calorimetry we have studied the interactions between a fungal LPMO and three soluble substrates and CDH. The results reveal which areas on the LPMO surface interact with the varying substrates and unambiguously show that both the substrate and CDH bind to a region that is centered around the copper site. The data presented here suggest that electron transfer occurs before substrate binding, providing important new leads for understanding the reaction mechanism of LPMOs.

Author contributions: G.C., Å.K.R., G.V.-K., M. Sørlie, M. Sandgren, R.L., V.G.H.E., and F.L.A. designed research; G.C., R.W., Å.K.R., M.P., A.K.G.F., M.D., V.G.H.E., and F.L.A. performed research; G.C., Å.K.R., G.V.-K., M. Sørlie, V.G.H.E., and F.L.A. analyzed data; and G.C., R.W., G.V.-K., M. Sørlie, M. Sandgren, R.L., V.G.H.E., and F.L.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Assigned chemical shifts have been deposited in BioMagResBank (www.bmrb.wisc.edu/) (entry no. 26717).

[1]To whom correspondence may be addressed. Email: finn.l.aachmann@ntnu.no or vincent.eijsink@nmbu.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1602566113/-/DCSupplemental.

β-sandwich fold with several loops protruding from the β-sandwich. The enzyme has a flat surface, which contains the copper coordination site [a common feature for all LPMOs (4, 9, 27)]. The copper ion is coordinated by a histidine brace composed by the N-terminal amino group (His1), the side chain of His-1 ($N^{\delta1}$), and the side chain of His83 ($N^{\varepsilon2}$). The hydroxyl group of a characteristic tyrosine residue, Tyr166, further shapes the copper site by occupying one of the axial coordination positions.

NMR spectroscopy studies have previously shown that the surface surrounding the conserved copper-binding site is responsible for the interaction of the chitin-active AA10 LPMO CBP21 from *Serratia marcescens* to crystalline chitin (9), and a similar substrate binding surface has been suggested for family AA9 LPMOs on the basis of docking and crystallography studies (13, 28). An intriguing question is how the LPMO interacts with CDH during electron transfer, especially when the LPMO is bound to the substrate, which would make the copper site inaccessible to CDH. According to the postulated LPMO mechanisms, CDH has to deliver two electrons to the LPMO during each reaction cycle (6, 10, 29). Docking studies have suggested that surface residues close to the copper site interact with the cytochrome domain of CDH during electron transfer (ET) (14). An alternative CDH docking site that would not be blocked by substrate binding has also been suggested (13). Experimental data that could shed light on the interactions of AA9 LPMOs with their substrates and CDH are lacking.

In the present study, we have used isothermal titration calorimetry (ITC) and NMR spectroscopy techniques to analyze the structure and dynamic features of *Nc*LPMO9C in solution. We have mapped the residues involved in the interaction between the catalytic domain of *Nc*LPMO9C and three soluble substrates: cellulose hexasaccharide (Glc_6), xyloglucan 14-mer (XG_{14} = a cellooctaose backbone with substitutions; *SI Materials and Methods*), and polymeric xyloglucan from tamarind seeds (polyXG). Furthermore, we have used the unique possibilities offered by the NMR assignment to map the interaction of *Nc*LPMO9C with full-length CDH from *N. crassa* and its isolated heme *b*-type cytochrome domain (CYT). Thus, we obtained insights into how fungal LPMOs interact with their substrates and CDH.

Results

Secondary Structure and Mobility in Solution. To verify that the overall NMR structure of *Nc*LPMO9C in solution corresponds to the X-ray crystal structure, the presence of secondary structure elements was analyzed using TALOS-N. This analysis demonstrated that the solution structure and the X-ray crystal structure comprise the same secondary structure elements (Fig. S1), indicating that the two structures are very similar.

*Nc*LPMO9C is able to bind a variety of substrates, which could indicate a flexible binding surface. To gain insight into this issue, ^{15}N-{^1H} NOEs as well as T_1 and T_2 relaxation times (picosecond and nanosecond timescales) were measured. Both the ^{15}N-{^1H} NOEs and the relaxation data are relatively featureless and show the characteristics of a rigid protein (Fig. S2), as also observed previously for an AA10 LPMO (9).

Substrate Binding. The interaction of *apo-Nc*LPMO9C with different ligands was probed by measuring changes in chemical shifts in ^{15}N-HSQC and ^{13}C-aromatic-HSQC spectra upon titration with three substrates (Glc_6, XG_{14}, and polyXG) and $GlcNAc_6$. The larger chemical shift changes are likely to occur at the ligand binding interface. Additional chemical shift changes, which are normally smaller, may be observed for nuclei that are near but not directly involved in the interface or as the result of propagating conformational changes upon binding (30).

All ligands except chitin-derived $GlcNAc_6$ showed binding to an area on the surface of *Nc*LPMO9C clustered around the histidine brace (His1 and His83) (Fig. 1 and Fig. S3). All substrates had a substantial effect on the chemical shifts of residues His1, Ala80, His83 and His155, whereas the effect on the chemical shifts of other amino acids varied according to the substrates used. A large surface loop [also known as the LC loop (31)] showing considerable variation among LPMOs but also containing a highly conserved tyrosine, Tyr204, which has been suggested to contribute to cellulose binding (13, 31), was generally little affected by substrate binding and was more affected by the binding of XG_{14} and polyXG than by the binding of Glc_6 (Fig. 1). Generally, the longer XG substrates (Fig. 1 *C* and *D*) had more extended effects on chemical shifts than Glc_6 (Fig. 1*B*). Differences were observed for a surface loop containing His64, which is part of an insertion [also known as L3 (26)] that only occurs in a subgroup of AA9 LPMOs and that is present in both xyloglucan active LPMOs that have been described so far (26, 32). This region was more affected by the binding of XG_{14} and, particularly, polyXG compared with Glc_6. Compared with other substrates, binding of polyXG had more predominant effects on residues located further away (>15 Å) from the active site surface, such as Tyr166–Cys169 in the β8-strand (Fig. 1*D* and Fig. S3).

Initially, NMR studies were done using the *apo*-enzyme, one reason being the paramagnetic relaxation enhancement effect of Cu^{2+} that would make a larger portion of the substrate binding surface invisible in the NMR experiments (9, 33). To gain further insight into substrate binding, the effect of bivalent metal ions and the effect of CN^-, an analog of superoxide known to inhibit LPMO activity (5), we then carried out a series of ITC experiments (Fig. S4), the results of which are summarized in Table 1. Loading the *apo*-enzyme with copper had little effect on substrate affinity, and both enzyme forms bound XG_{14} slightly better than Glc_6. A clear increase in affinity was observed in the presence of cyanide, for both substrates but more so for Glc_6. This result suggests that the enzyme–substrate interaction is strengthened during the first steps of the LPMO reaction, where a Cu^{2+}–superoxide complex is likely to emerge (29).

The ITC data (Table 1) confirm that Glc_6 and XG_{14} have different binding modes. First, the affinity for XG_{14} is higher compared with Glc_6, except for the Cu^{2+}/CN^- situation, where the two ligands bind with similar affinities. Second, for all LPMO forms, the thermodynamic signatures of Glc_6 and XG_{14} binding differed considerably; the enthalpic effects of binding of XG_{14} were much larger compared with Glc_6. As a prelude to further NMR experiments (where use of Cu^{2+} was to be avoided), an experiment with a Zn^{2+}-loaded enzyme was also conducted, showing slightly weaker binding compared with Cu^{2+}.

Based on the results from the ITC experiments, additional NMR experiments were carried out, using Zn^{2+}-loaded enzyme, in the presence of cyanide, without revealing other or additional substrate interactions compared with the data discussed above.

Docking of Cellulose Hexamer. Docking was used to obtain a conceptual understanding of how binding might occur, using a simple docking algorithm combined with restraints obtained from the NMR data. The results showed that the surface patch surrounding the copper site appears to be a preferred interaction surface for Glc_6 (Fig. 2) and that the interaction energies are favorable. The results suggest that Tyr204 is not involved in binding of this short substrate, whereas residues in the L3 loop (His64), which is characteristic for this group of LPMOs, are more prominent.

Interaction with CDH and CYT. To identify amino acids involved in the interaction of the LPMO with CDH and CYT, changes in chemical shifts in the ^{15}N-HSQC and ^{13}C-aromatic-HSQC spectra were measured upon addition of either protein to *Nc*LPMO9C. Both CDH and CYT showed binding to an area on the surface of *Nc*LPMO9C clustered around the copper coordination site, the effects being somewhat more pronounced for CDH than for CYT (Fig. 3). Thus, the amino acid clusters that

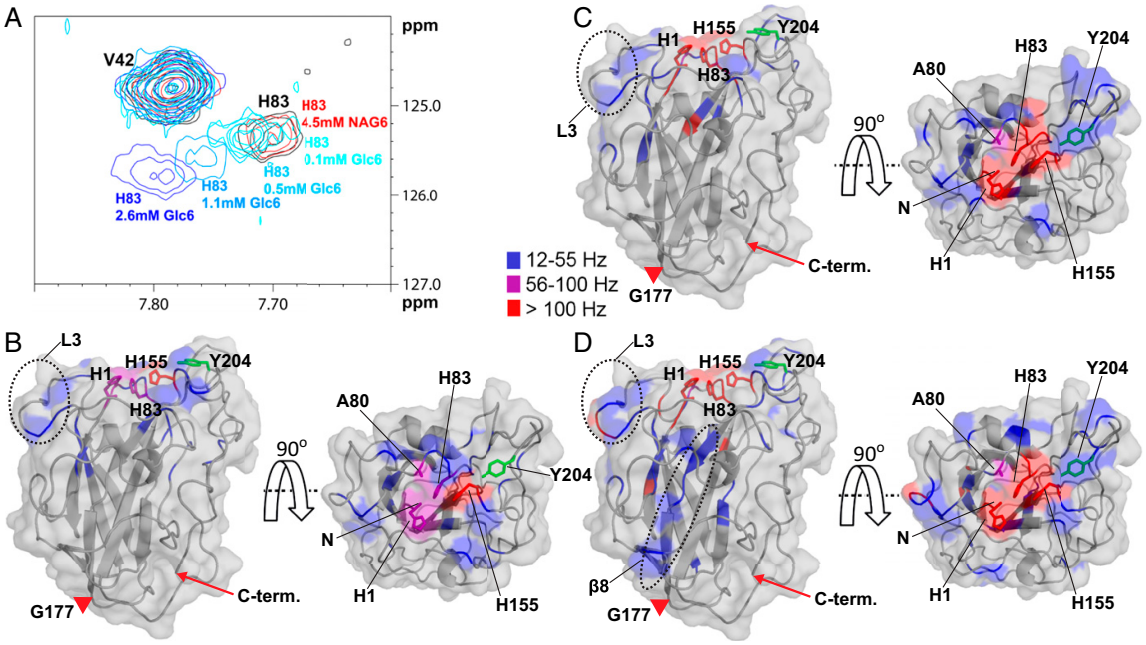

Fig. 1. Interaction of *apo-Nc*LPMO9C with substrates. (*A*) Overlay of an area of interest from the ^{15}N-HSQC spectrum for *Nc*LPMO9C (black) in the presence of 4.5 mM $GlcNAc_6$ (labeled as NAG_6; red) or increasing concentrations of Glc_6 (from lighter to darker blue). The H^N/N chemical shift of Val42 is not affected by the interaction, and therefore, the peak is shown as a reference. (*B–D*) Compound change in chemical shifts larger than 12 Hz (Fig. S3) upon substrate binding mapped on the structure of *Nc*LPMO9C. The backbone of *Nc*LPMO9C (shown in cartoon and surface representation) is colored according to the compound change in chemical shift (^{15}N-HSQC) upon adding 2.6 mM Glc_6 (*B*), 1.3 mM XG_{14} (*C*), or 4.2 μM polyXG (*D*) using the indicated coloring scheme (gray coloring represents no change). The *Nc*LPMO9C structure is shown by a side view (*Left*) and a top view (*Right*). The side chains of residues His1, Ala80, His83, and His155 are shown in stick representation. In addition, the side chain of Tyr204 is shown in green. The positions of the L3 loop and the β8-strand are marked on the structures. The LC loop spans the stretch from Gly177 (marked with a red triangle) to the C terminus (marked with a red arrow). The N-terminal amino group (His1) is not observed in ^{15}N-HSQC spectra because of its fast exchange. The ^{13}C-aromatic HSQC spectra showed clear changes in chemical shift for this residue, with all three substrates, with the strongest effects (a vanished signal) being observed with XG_{14} and polyXG. Based on these observations, for illustrative purposes, His-1 is colored purple (*B*) or red (*C* and *D*) in the figures.

were affected by the binding of these proteins were similar to the regions that were most affected in the substrate titration experiments. However, compared with the substrates, and especially the xyloglucan substrates, the interaction areas on the LPMO are more focused on the catalytic center, i.e., His1, Ala80, His83, and His155, with minimal effects on the LC and L3 loops.

A competition experiment was carried out to verify that the same residues were involved in binding substrates and CYT/CDH. Indeed, upon addition of excess Glc_6 to a mixture of *Nc*LPMO9C and CDH, the chemical shifts of the H^N/N signals in the ^{15}N-HSQC spectrum returned to values similar to those observed for the titration end-point of *Nc*LPMO9C with Glc_6 (Fig. 3*C*). This result unambiguously shows that CDH and Glc_6 bind to the same area on the LPMO.

Despite considerable efforts, we have so far not been able to establish conditions that allow monitoring of the CDH–LPMO interaction by ITC, an interaction that likely is transient in nature.

Discussion

The complex mechanism that LPMOs use to oxidize their substrates involves the interplay between the LPMO, its substrate, the copper ion, molecular oxygen, and an electron donor. Thus, a fundamental understanding of the interaction between the LPMO and each of these factors is necessary for unraveling the molecular basis of the enzymatic mechanism. In the current study, we have used NMR and ITC techniques to provide unprecedented insight into key interactions between *Nc*LPMO9C, its substrates, and a likely natural electron donor, CDH, as they would occur in solution.

It is challenging to obtain information about the interaction of LPMOs with polymeric substrates because these are mostly large, insoluble, and even partially crystalline, which complicates or precludes most experimental approaches. The ability of *Nc*LPMO9C to act on soluble substrates allowed us to probe these interactions by titrating the protein with the substrates and measuring perturbations in the chemical shifts observed by NMR. The results show that the substrate binding surface is centered around the copper site and extends over the relatively flat surface of the LPMO, involving residues in surface loops that are remote from the catalytic center and that show considerable sequence variation. The NMR data show that the L3 and LC loops interact more strongly with XG_{14} and polyXG than with smaller substrates such as Glc_6. The L3 loop only occurs in a subgroup of LPMOs that includes the two LPMOs for which xyloglucan activity has been described so far [*Nc*LPMO9C (26) and *Pa*LPMO9H (32)]. The ITC data showed differences in binding affinity between Glc_6 and XG_{14} and revealed considerable differences in the thermodynamic signatures of binding. Binding of XG_{14} was associated with considerably larger beneficial enthalpic effects, suggesting more extended binding interactions, thus confirming the NMR data.

During the course of this study, Frandsen et al. (28) reported the crystal structure of an AA9 LPMO, *Ls*LPMO9A, in complex with Glc_6 (PDB ID: 5ACI). The docking model of Fig. 2 and the crystal structure show similar binding in the area near the copper-site, spanning subsites −2 to +2. Frandsen et al. showed that the +1 sugar

Table 1. Thermodynamic parameters for binding of XG_{14} and Glc_6 to various forms of *Nc*LPMO9C at t = 25 °C in 20 mM MES (pH 5.5)

Active site ligand	K_d*	$\Delta G^{\circ\dagger}$	$\Delta H^{\circ\dagger}$	$-T\Delta S^{\circ\dagger}$
		XG_{14}		
apo	0.42 ± 0.02	−4.6 ± 0.1	−10.8 ± 1.0	6.2 ± 1.0
Cu^{2+}	0.33 ± 0.04	−4.7 ± 0.1	−10.5 ± 0.5	5.8 ± 0.5
Cu^{2+}/CN^{-}‡	0.14 ± 0.01	−5.3 ± 0.1	−16.7 ± 0.4	11.4 ± 0.4
Zn^{2+}/CN^{-}	0.22 ± 0.02	−5.0 ± 0.1	−9.1 ± 1.0	4.1 ± 1.0
		Glc_6		
apo	1.1 ± 0.1	−4.0 ± 0.1	−4.5 ± 0.5	0.5 ± 0.5
Cu^{2+}	0.81 ± 0.08§	−4.3 ± 0.2§	−2.5 ± 0.5§	−1.8 ± 0.5§
Cu^{2+}/CN^{-}‡	0.13 ± 0.01	−5.3 ± 0.1	−9.2 ± 0.2	3.9 ± 0.2

Metal-loaded enzymes were obtained by adding solutions of $CuCl_2$ or $ZnSO_4$ to the *apo*-enzyme, to reach a final concentration of 30 μM. Cyanide was added to a final concentration of 1 mM. For practical reasons these experiments were done with the full-length enzyme. Thermograms are shown in Fig. S4.

*In mM.

†In kcal·mol^{-1}.

‡Analogous experiments with only the catalytic domain were done, showing essentially similar results.

§Data from ref. 26.

docks onto His1. The NMR data do not provide the same level of atomic accuracy, meaning that defining the +1 sugar in the docking model of Fig. 2 is somewhat arbitrary. Fig. 2 suggests that the binding mode is −3 to +3; it could also be −2 to +4 but certainly not −4 to +2. The main difference between the studies is that the crystal structure shows binding from −4 to +2, whereas the NMR data suggest binding from −3 to +3 or −2 to +4. This difference is accompanied by interesting structural variation. The L3 loop, which is important for interactions in the +3/+4 subsites (Figs. 1 and 2 and Fig. S3), is only 7 residues long in *Ls*LPMO9A, compared with 14 in *Nc*LPMO9C, leading to a quite different interaction surface, albeit with conservation of His64 (His66 in *Ls*LPMO9A). Furthermore, relative to *Ls*LPMO9A, *Nc*LPMO9C carries a six-residue insertion right next to Tyr204 (Tyr203 in *Ls*LPMO9A), which changes the binding surface near subsite −3/−4.

Notably, binding of polyXG to *Nc*LPMO9C perturbed chemical shifts on amino acids that are far from the interaction surface (Fig. 1*D* and Fig. S3). This might be due to indirect effects that result in minor conformational rearrangements in the protein, rather than direct surface binding interactions. Some of these conformational rearrangements occurred on the β8-strand containing the copper site tyrosine Tyr166.

The ITC data show that the presence of Cu^{2+} hardly affects the strength of ligand binding. This observation shows that copper binding does not affect the conformation of the substrate-binding surface, which coincides with the observed rigidity of the protein. Using NMR, it has been shown previously (9) that the only structural effect of metal ion binding is a tightening of the N terminus of the protein (His1). This observation also suggests that the copper ion is too recessed to interact directly with the substrate, as one would indeed conclude from enzyme–substrate complexes that are available in the literature (6, 28, 29, 31). We observed stronger ligand binding in the presence of cyanide. Cyanide is a known inhibitor of LPMOs (5) and a known copper-binding analog of the superoxide ion (34). Cu^{2+}–CN^{-} complexation is analogous to the Cu^{2+}–O_2^{-} complex emerging in the first step of LPMO catalysis initiated by a reduced LPMO (29). Our data thus suggest that ligand binding is enhanced upon single electron transfer from Cu^{1+} in the reduced LPMO to molecular oxygen.

Because of the shared timescales of molecular motions and spin-precession frequencies, we were able to perform NMR measurements that provided dynamic information about the protein. The results show that the backbone of *Nc*LPMO9C is inflexible (Fig. S2). Furthermore, the data were used to calculate the rotational correlation time (τ_c) for *Nc*LPMO9C, which was found to be smaller (τ_c = 11.8 ± 0.5 ns for 23.3 kDa *Nc*LPMO9C) than what would be expected for a globular protein of similar molecular weight (τ_c = 13.0 ns for a 21.9 kDa globular protein; Fig. S2). The same is true for CBP21 (9), confirming that LPMOs are compact, rigid proteins, which is an advantageous feature for proteins involved in ET reactions (35).

Significant progress has been made in understanding the role of CDH in the oxidative degradation of cellulose (3, 14, 15). The CYT domain of CDH contains a heme *b* prosthetic group that, after being reduced by the catalytic dehydrogenase domain of CDH, proceeds to reduce the LPMO through an intermolecular electron transfer (ET) event (3, 14). In early work, CDH was proposed to interact with a conserved region centered around a Pro–Gly–Pro motif situated on a side of the LPMO that is opposite of the copper-binding surface, which implies that long-range electron transfer through the LPMO would take place. Upon the determination of the crystal structure of full-length CDH, Tan et al. (14) concluded from docking studies that CDH is

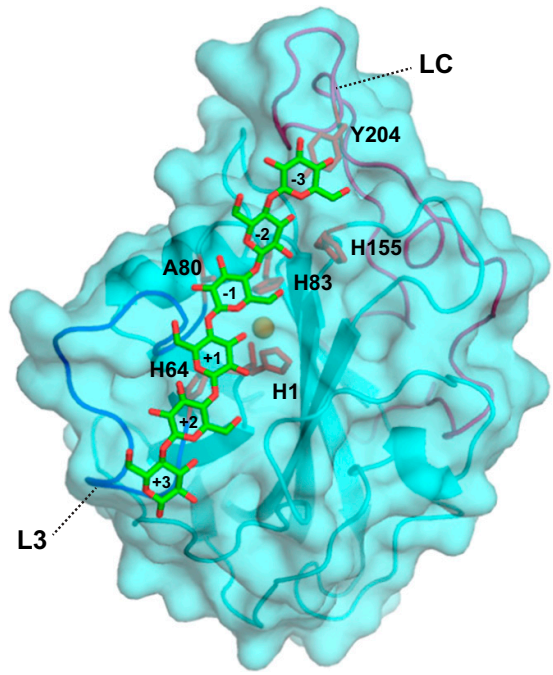

Fig. 2. Interaction model of Glc_6 and *Nc*LPMO9C produced by HADDOCK (39). The backbone is shown as a cartoon and surface, and the side chains of residues known from the NMR experiments to be strongly affected by substrate binding (His1, His64, Ala80, His83, and His155) are shown as sticks. In addition, the picture shows the side chain of a selected surface residue (Tyr204) that shows a high degree of sequence conservation and that may be involved in substrate binding, possibly without an effect of substrate binding on the compound change in chemical shift (^{15}N-HSQC). In the shown complex, the scissile glycosidic bond is located at 5 Å from the copper atom (orange sphere). The L3 loop (containing His64) and the long LC loop (containing Tyr204) are displayed in blue and magenta, respectively. The sugar residues are numbered by subsite, where the sugar that is closest to His1 is sugar +1, in accordance with recent crystallographic data (28) (*Discussion*). The HADDOCK energies were (with SD) as follows: van der Waals energy = −39.6 ± 2.7 kcal·mol^{-1}, electrostatic energy = −27.8 ± 7.7 kcal·mol^{-1}, and desolvation energy = −8.5 ± 3.2 kcal·mol^{-1}.

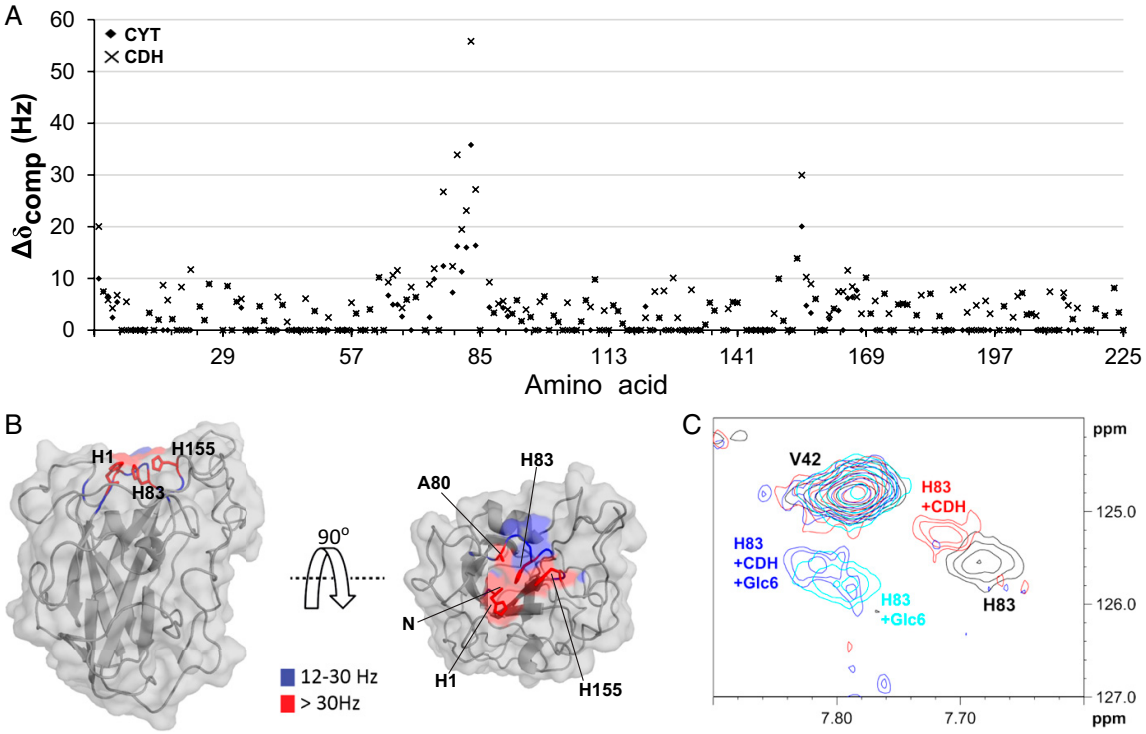

Fig. 3. Interaction of *apo-Nc*LPMO9C with CDH and CYT. (*A*) Compound change in chemical shifts (^{15}N-HSQC) for each amino acid in *Nc*LPMO9C upon addition of CYT (rhombi) or CDH (crosses). (*B*) Compound change in chemical shifts larger than 12 Hz mapped on the *Nc*LPMO9C structure. The backbone of *Nc*LPMO9C (shown in cartoon and surface representation) is colored using the indicated coloring scheme. The side chains of residues His1, Ala80, His83, and His155 are shown in stick representation. The ^{13}C-aromatic HSQC spectra showed clear changes in chemical shift for His1, and this residue was treated as described in the legend of Fig. 1. (*C*) Overlay of an area of interest from the ^{15}N-HSQC spectrum for 0.10 mM *apo-Nc*LPMO9C in the absence of an interaction partner (black) and in the presence of 0.12 mM CDH (red), 2.6 mM Glc_6 (cyan), or both (blue).

likely to interact directly with the copper. The present NMR data show that CDH indeed interacts directly with the copper site and that the interaction involves a relatively narrow surface patch near that site (Fig. 3). Thus, it would seem that the ET event occurs directly between the heme *b* of CYT and the copper ion in *Nc*LPMO9C. This patch contains amino acids that are also involved in the interaction with substrates, and we show that *Nc*LPMO9C will not bind Glc_6 and CDH simultaneously. The changes in chemical shift upon interaction of *Nc*LPMO9C with CYT or CDH were small compared with changes in the chemical shifts caused by the binding of substrates, which is indicative of weak protein–protein interactions (36).

Knowing that CDH by itself is capable of supplying the LPMO with the two electrons it needs per catalytic cycle and the fact that CDH cannot bind to the LPMO in the presence of substrate would imply that these electrons are present in the LPMO before substrate binding. This is a puzzling thought because reduction of the copper transfers only one electron to the LPMO. The presence of two electrons before substrate binding could be taken to imply that LPMOs have the ability to somehow store a second electron. Alternatively, the formation of the copper–superoxo adduct initiated by the first transferred electron could lead to subsequent oxidation of a tyrosine or a tryptophan residue. To complete the reaction cycle, CDH would donate two electrons: one to reduce Cu^{2+} to Cu^{1+} and another one to return the tyrosine or tryptophan radical to its normal form (37). LPMOs are exceptionally rich in aromatic residues and contain a conserved cluster of aromatic residues in vicinity of the copper site, whose π-electron clouds may facilitate intraprotein ET.

Altogether, this study provides insight into LPMO–substrate interactions as well as the first experimental evidence to our knowledge demonstrating that the electron transfer mechanism between CYT and LPMO is based on the direct interaction of the heme *b* and LPMO copper sites. The location of the binding site for CDH, the binding competition between CDH and substrate (Fig. 3), and the increased binding strength observed in the LPMO–Cu^{2+}–CN^{-}–ligand complex suggest that all electrons necessary for the reaction are present in the LPMO before substrate binding.

Materials and Methods

Detailed information for all experimental procedures is provided in *SI Materials and Methods*.

Sample Preparation. The isotope-labeled catalytic domain of *Nc*LPMO9C used in NMR experiments was recombinantly produced in *Pichia pastoris* cultivated in isotope-enriched (^{13}C, ^{15}N) minimal medium and purified by multiple chromatographic steps as described previously (26). Nonlabeled full-length *Nc*LPMO9C and its catalytic domain were produced and purified as previously described. Because the LPMO was produced in *P. pastoris*, His1 was not methylated (26).

Full-length CDH and its isolated CYT, obtained as a proteolytic degradation product, were produced in *P. pastoris* and purified by subsequent hydrophobic interaction and ion exchange chromatography steps, followed by deglycosylation. Homogeneous CDH and CYT preparations were obtained by size exclusion chromatography.

NMR Spectroscopy. NMR data for *apo-Nc*LPMO9C in 25 mM sodium phosphate buffer (pH 5.5) and 10 mM NaCl were obtained using a Bruker Ascend

800-MHz spectrometer and a Bruker Avance III 600-MHz spectrometer. The NMR assignment of *Nc*LPMO9C has been published elsewhere (38).

Detection of residues involved in the interactions between *Nc*LPMO9C and different ligands (Glc_6, XG_{14}, polyXG, CYT, and CDH) was accomplished by measuring chemical shift perturbations in ^{15}N-HSQC spectra upon adding a ligand to 80–100 μM *Nc*LPMO9C.

ITC. Dissociation constants and thermodynamics data for binding of Glc_6 and XG_{14} to various forms of *Nc*LPMO9C were determined with a VP-ITC system from Microcal, Inc., essentially as described previously (26), using 20 mM MES buffer (pH 5.5) at t = 25 °C.

ACKNOWLEDGMENTS. This work was financed by strategic funds from the Norwegian University of Science and Technology and the MARPOL project, a FRINAT project, and the Norwegian NMR Platform (Grants 221576, 214613, and 226244 from the Research Council of Norway, respectively). The NMR laboratory at Aalborg University is supported by the Obel, SparNord, and Carlsberg Foundations. Part of the work was supported by the European Commission through the INDOX project (FP7-KBBE-2013-7-613549).

1. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH (2012) Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* 5(1):45–57.
2. Forsberg Z, et al. (2011) Cleavage of cellulose by a CBM33 protein. *Protein Sci* 20(9): 1479–1483.
3. Phillips CM, Beeson WT, Cate JH, Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol* 6(12):1399–1406.
4. Quinlan RJ, et al. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA* 108(37):15079–15084.
5. Vaaje-Kolstad G, et al. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330(6001):219–222.
6. Kim S, Ståhlberg J, Sandgren M, Paton RS, Beckham GT (2014) Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc Natl Acad Sci USA* 111(1):149–154.
7. Beeson WT, Vu VV, Span EA, Phillips CM, Marletta MA (2015) Cellulose degradation by polysaccharide monooxygenases. *Annu Rev Biochem* 84:923–946.
8. Hemsworth GR, Johnston EM, Davies GJ, Walton PH (2015) Lytic polysaccharide monooxygenases in biomass conversion. *Trends Biotechnol* 33(12):747–761.
9. Aachmann FL, Sørlie M, Skjåk-Bræk G, Eijsink VGH, 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 USA* 109(46): 18779–18784.
10. Beeson WT, Phillips CM, Cate JHD, Marletta MA (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc* 134(2): 890–892.
11. Hemsworth GR, et al. (2013) The copper active site of CBM33 polysaccharide oxygenases. *J Am Chem Soc* 135(16):6069–6077.
12. Forsberg Z, et al. (2014) Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* 53(10):1647–1656.
13. Li X, Beeson WT, 4th, Phillips CM, Marletta MA, Cate JHD (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* 20(6):1051–1061.
14. Tan T-C, et al. (2015) Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nat Commun* 6(May):7542–7553.
15. Langston JA, et al. (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microbiol* 77(19): 7007–7015.
16. Westereng B, et al. (2015) Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Sci Rep* 5:18561–18670.
17. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6(1):41–55.
18. Lo Leggio L, et al. (2015) Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nat Commun* 6:5961.
19. Hemsworth GR, Henrissat B, Davies GJ, Walton PH (2014) Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nat Chem Biol* 10(2): 122–126.
20. Isaksen T, et al. (2014) A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *J Biol Chem* 289(5):2632–2642.
21. Agger JW, et al. (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc Natl Acad Sci USA* 111(17):6287–6292.
22. Frommhagen M, et al. (2015) Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnol Biofuels* 8(1):101–113.
23. Vu VV, Beeson WT, Span EA, Farquhar ER, Marletta MA (2014) A family of starch-active polysaccharide monooxygenases. *Proc Natl Acad Sci USA* 111(38):13822–13827.
24. 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(1):79–92.
25. Vu VV, Beeson WT, Phillips CM, Cate JHD, Marletta MA (2014) Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *J Am Chem Soc* 136(2):562–565.
26. Borisova AS, et al. (2015) Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity. *J Biol Chem* 290(38): 22955–22969.
27. Hemsworth GR, Davies GJ, Walton PH (2013) Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Curr Opin Struct Biol* 23(5):660–668.
28. Frandsen KEH, et al. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat Chem Biol* 12(4):298–303, 10.1038/nchembio.2029.
29. Kjaergaard CH, et al. (2014) Spectroscopic and computational insight into the activation of O_2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc Natl Acad Sci USA* 111(24):8797–8802.
30. Cavanagh J, Fairbrother WJ, Palmer AG, III, Rance M, Skelton NJ (2007) Protein-ligand binding interfaces. *Protein NMR Spectroscopy: Principles and Practice* (Elsevier Academic Press, Amsterdam), 2nd Ed, pp 755–760.
31. Wu M, et al. (2013) Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the Basidiomycota fungus *Phanerochaete chrysosporium*. *J Biol Chem* 288(18):12828–12839.
32. Bennati-Granier C, et al. (2015) Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnol Biofuels* 8:90–104.
33. Bertini I, Pierattelli R (2004) Copper(II) proteins are amenable for NMR investigations. *Pure Appl Chem* 76(2):321–333.
34. Paci M, Desideri A, Rotilio G (1988) Cyanide binding to Cu, Zn superoxide dismutase. An NMR study of the Cu(II), Co(II) derivative. *J Biol Chem* 263(1):162–166.
35. Page CC, Moser CC, Chen X, Dutton PL (1999) Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* 402(6757):47–52.
36. Ubbink M (2009) The courtship of proteins: understanding the encounter complex. *FEBS Lett* 583(7):1060–1066.
37. Warren JJ, Winkler JR, Gray HB (2012) Redox properties of tyrosine and related molecules. *FEBS Lett* 586(5):596–602.
38. Courtade G, et al. (2016) Backbone and side-chain ^{1}H, ^{13}C, and ^{15}N chemical shift assignments for the apo-form of the lytic polysaccharide monooxygenase *Nc*LPMO9C. *Biomol NMR Assign*, 10.1007/s12104-016-9683-x.
39. de Vries SJ, van Dijk M, Bonvin AMJJ (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 5(5):883–897.
40. Sygmund C, et al. (2012) Characterization of the two *Neurospora crassa* cellobiose dehydrogenases and their connection to oxidative cellulose degradation. *Appl Environ Microbiol* 78(17):6161–6171.
41. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handook*, ed Walker JM (Springer, Totowa, NJ), pp 571–607.
42. Keller R (2004) *The Computer Aided Resonance Assignment Tutorial* (CANTINA Verlag, Goldau, Switzerland).
43. Farrow NA, et al. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33(19):5984–6003.
44. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* 28(23):8972–8979.
45. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227–241.
46. Mulder FAA, Schipper D, Bott R, Boelens R (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J Mol Biol* 292(1):111–123.
47. Wiseman T, Williston S, Brandts JF, Lin L-N (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 179(1):131–137.
48. Turnbull WB, Daranas AH (2003) On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc* 125(48):14859–14866.
49. DeLano WL, Lam J (2005) PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc* 230:1371–1372.
50. Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14(1):51–55.
51. Schüttelkopf AW, van Aalten DMF (2004) PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60(Pt 8):1355–1363.
52. Vaaje-Kolstad G, Houston DR, Riemen AHK, Eijsink VGH, van Aalten DMF (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J Biol Chem* 280(12):11313–11319.

Supporting Information

Courtade et al. 10.1073/pnas.1602566113

SI Materials and Methods

Sample Preparation. Cloning, protein expression production, and purification of the uniformly isotope-labeled (^{15}N and ^{13}C) catalytic domain of *Nc*LPMO9C (26) and of unlabeled full-length *Nc*LPMO9C (24), as well as conditions for NMR measurements (30), have been described previously. The catalytic domain of *Nc*LPMO9C is not glycosylated (26), whereas full-length *Nc*LPMO9C carries O-glycosylation in the additional linker and/or carbohydrate-binding module (24). For some experiments (where Zn^{2+} was used), the buffer was exchanged using protein spin concentrators (Vivaspin 6 PES; 5 kDa MWCO; Sartorius) to 20 mM sodium-acetate (pH 5.5), 10 mM NaCl, and 0.5 mM $ZnCl_2$ in 90% (vol/vol) H_2O/10% (vol/vol) D_2O.

Recombinant production of *Neurospora crassa* CDH IIA in *Pichia pastoris* was carried out as previously described (40). Culture supernatant was harvested from the bioreactor 72 h after methanol induction and contained CDH as well as its proteolytic degradation product CYT. After diafiltration with a Microza UF module (SLP-1053; 10 kDa MWCO; Pall Corporation) the proteins, buffered in 50 mM sodium acetate (pH 5.5) containing 20% (wt/vol) (saturated) $(NH_4)_2SO_4$, were loaded onto a Phenyl-Sepharose-FF column (GE Healthcare) and eluted by applying a linear gradient to 50 mM sodium acetate buffer (pH 5.5). Fractions containing CDH and CYT were separately pooled, changed to a 100 mM sodium acetate buffer (pH 5.5), loaded onto a DEAE-Sepharose column (GE Healthcare), and eluted with the same buffer containing 1 M NaCl. Fractions containing CDH or CYT were pooled and deglycosylated with 5,000 U endoglycosidase H_f and 300 U α-1,2/3-mannosidase (New England Biolabs) per mg protein in a 100 mM sodium acetate buffer (pH 5.5) containing 5 mM $ZnCl_2$ for 48 h at 8 °C (14). Size exclusion chromatography with a Superdex 75 column (GE Healthcare) was used to remove glycosidases and low-molecular weight compounds and resulted in homogeneous CDH (6.0 mg·mL^{-1}) and CYT (10 mg·mL^{-1}) preparations in 100 mM sodium acetate buffer (pH 5.5).

Polymeric xyloglucan from tamarind seeds (polyXG, average Mw = 225 kDa), cellulose hexasaccharide (Glc_6), chitin hexasaccharide ($GlcNAc_6$), and the sample referred to as XG_{14} were obtained from Megazyme. XG_{14} is a mixture of longer xyloglucan oligomers that is dominated by species with a cellooctaose backbone with sequence XXXGXXXG (where X stands for glucose with a xylose substitution), where up to three (most commonly one) of the xyloses carry an additional galactose substitution (Megazyme, product number O-XGHDP).

Protein concentration was determined by measuring the A_{280} of the protein solution and deducing the protein concentration based on the theoretical extinction coefficient [calculated using the ProtParam tool; web.expasy.org/tools/protparam/ (41)].

NMR Spectroscopy. NMR spectra of 80–200 μM *Nc*LPMO9C samples were recorded at 25 °C on a Bruker Ascend 800 MHz spectrometer Avance III HD equipped with a 5-mm Z-gradient CP-TCI (H/C/N) cryogenic probe at the NT-NMR-Center/Norwegian NMR Platform (NNP). NMR relaxation measurements were recorded on a Bruker Avance III 600 MHz spectrometer equipped with a 5-mm Z-gradient CPP-TCI (H/C/N) cryogenic probe at the Department of Chemistry and Bioscience, Aalborg University. NMR data were processed using Bruker TopSpin version 3.5. NMR spectral analysis was performed using CARA version 1.5.5 (42). The NMR assignment of the catalytic domain of *Nc*LPMO9C has been published elsewhere (30). The ^{15}N-{^1H} heteronuclear NOEs were derived with Protein Dynamic Center software version 2.3.1 from Bruker BioSpin using two independently measured and integrated ^{15}N-{^1H} heteronuclear correlated spectra with and without ^1H saturation (43). Nuclear magnetic relaxation time measurements of ^{15}N nuclei (T_1 and T_2) were analyzed with Protein Dynamic Center software version 2.3.1 from Bruker BioSpin, using exponential fitting of data from ^{15}N heteronuclear single quantum coherence (HSQC)-type spectra that had been acquired with different relaxation delays (43, 44). Secondary structure elements were analyzed using the web-based version of the TALOS-N software (spin.niddk.nih.gov/bax/nmrserver/talosn/) (45) using the backbone chemical shifts (N, H^N, C', C^α, and H^α) and some side chain chemical shifts (C^β and H^β).

Interaction Studies with NMR. The K_d values for binding to copper-saturated *Nc*LPMO9C at pH 5.5, determined by Borisova et al. (26) using ITC, were used to estimate the amounts of Glc_6 and polyXG added to *apo-Nc*LPMO9C. For Glc_6, the titration points were 0.1, 0.5, 1.1, and 2.6 mM. For polyXG, the titration points were 0.1 μM (20 μg/mL), 0.2 μM (41 μg/mL), 0.7 μM (146 μg/mL), 2.4 μM (538 μg/mL), and 4.2 μM (944 μg/mL). For XG_{14}, the titration points were 0.05 mM (133 μg/mL), 0.1 mM (267 μg/mL), 0.2 mM (400 μg/mL), and 1.3 mM (2822 μg/mL). As a negative control, two titration points were recorded for $GlcNAc_6$ at 0.9 mM and 4.4 mM. The 1D-proton, ^{15}N-HSQC, and ^{13}C-aromatic-HSQC spectra were recorded for each titration point.

To map the interaction of CYT and CDH with *Nc*LPMO9C, the proteins were added to *apo-Nc*LPMO9C in a 1.1:1 ratio. The 1D-proton, ^{15}N-HSQC, and ^{13}C-aromatic-HSQC spectra were recorded before and after the addition of the proteins.

Mapping of the residues involved in the interaction between *Nc*LPMO9C and each of the three substrates, CYT, or CDH was accomplished by measuring chemical shift changes in the N and H^N atoms of the backbone of *Nc*LPMO9C. A compound change in chemical shift was calculated using the following formula: $\Delta\delta_{comp} = \sqrt{(\Delta\delta_H)^2 + x(\Delta\delta_N)^2}$. $\Delta\delta_{comp}$ is the absolute change in chemical shift (Hz). $\Delta\delta_H$ is the change in chemical shift of the amide proton (Hz). $\Delta\delta_N$ is the change in chemical shift of the amide nitrogen atom (Hz). x is the constant used to achieve equal contributions from changes in N and H^N shifts (9, 46). Significant compound changes in chemical shifts were defined as those larger than twice the spectral resolution ($\Delta\delta_{comp} > 12$ Hz).

Affinity Experiments: ITC. Isothermal titration calorimetry (ITC) experiments were performed with a VP-ITC system from Microcal, Inc. Solutions were thoroughly degassed before experiments to avoid air bubbles in the calorimeter. Chelex-treated 20 mM MES buffer (pH 5.5) at t = 25 °C was used for all binding studies. For binding, a 15 μM enzyme solution was placed in the reaction cell with a volume of 1.42 mL, and 4.0 mM of XG_{14} or Glc_6 was placed in the ITC syringe. Aliquots of 4 μL were injected into the reaction cell at 180-s intervals with a stirring speed of 260 rpm. The titrations were complete after 50 injections.

ITC data were collected automatically using the Microcal Origin v.7.0 software accompanying the VP-ITC system. Before further analysis, all data were corrected for heat of dilution by subtracting the heat produced by titrating ligand into buffer alone. The data were fitted using a nonlinear least-squares algorithm using a single-site binding model used by the Origin software that accompanies the VP-ITC system, yielding the stoichiometry (n), the equilibrium binding association constant (K_a), and the enthalpy change (ΔH_r°) of the reaction. Errors in ΔH_r°, K_d, and

$\Delta G_r°$ were obtained as SDs of at least three experiments. Errors in $\Delta S_r°$ and $-T\Delta S_r°$ were obtained as propagation of errors.

The shape of the ITC binding curve may be described by the so-called Wiseman c value, which can be expressed as follows: $c = nK_a[M]_t$, where n is the stoichiometry of the reaction, K_a is the equilibrium binding association constant, and $[M]_t$ is the protein concentration (47). The c values preferably need to be in the range of $10 < c < 1{,}000$ to obtain meaningful calculations of K_a. For the titration of XG_{14} or Glc_6 in all forms of *Nc*LPMO9C, binding isotherms were hyperbolic, indicating weak binding and a c value below 10 (47, 48). In such cases, binding thermodynamics can be obtained even if c is in the range of $0.01 < c < 10$ if a sufficient portion of the binding isotherm is used for analysis (48). This is achieved by ensuring a high molar ratio of ligand versus protein at the end of the titration, accurate knowledge of the concentrations of both ligand and receptor, an adequate level of signal-to-noise in the data, and known stoichiometry. These conditions were all met in the experimental setup used here.

Docking. To gain further insight into the interaction between Glc_6 and *Nc*LPMO9C, high ambiguity-driven biomolecular docking was carried out using the Easy Interface of HADDOCK 2.2 (haddock.science.uu.nl/services/HADDOCK2.2/haddockserver-easy.html) (39). A PDB file containing the X-ray crystal structure of the catalytic domain of *Nc*LPMO9C (26) was obtained from the Protein Data Bank (PDB ID: 4D7U), and protons were added to the residues using PyMol (49). Active residues were defined as those with a compound chemical shift larger than twice the spectral resolution ($\Delta\delta_{comp} > 12$ Hz) and a relative solvent accessibility of either main chain or side chain above 15%, as determined by MolMol (50). Passive residues were determined automatically by HADDOCK as those within a radius of 6.5 Å of any active residue and a relative solvent accessibility above 15% (39). The structure of a fully extended conformation of Glc_6 was constructed using GLYCAM Carbohydrate Builder (glycam.org), and the PDB file was prepared as an input for HADDOCK using PRODRG (davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg) (51).

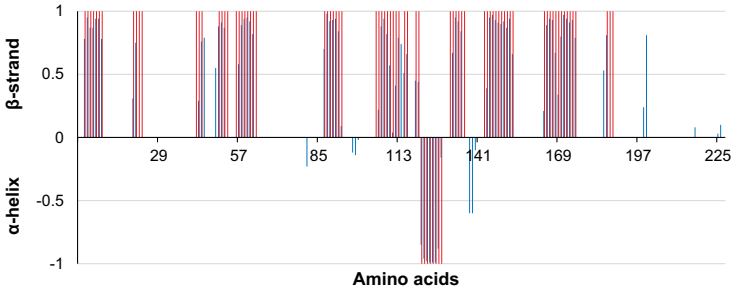

Fig. S1. Secondary structure of *Nc*LPMO9C in solution, as derived from secondary chemical shifts. Secondary chemical shifts allow estimation of dihedral angles, which in turn can be used to predict secondary structure propensities, which are indicated by the blue bars. The pink bars indicate secondary structure assignments based on the crystal structure, and comparison shows that there is excellent agreement between the secondary structure elements observed in solution and those seen in the X-ray crystallographic structure (PDB ID: 4D7U) (26). Such high similarity is not unexpected and has previously been shown to be valid for another LPMO, CBP21, for which both the NMR (PDB ID: 2LHS) (9) and the X-ray (PDB ID: 2BEM) (52) structures are available.

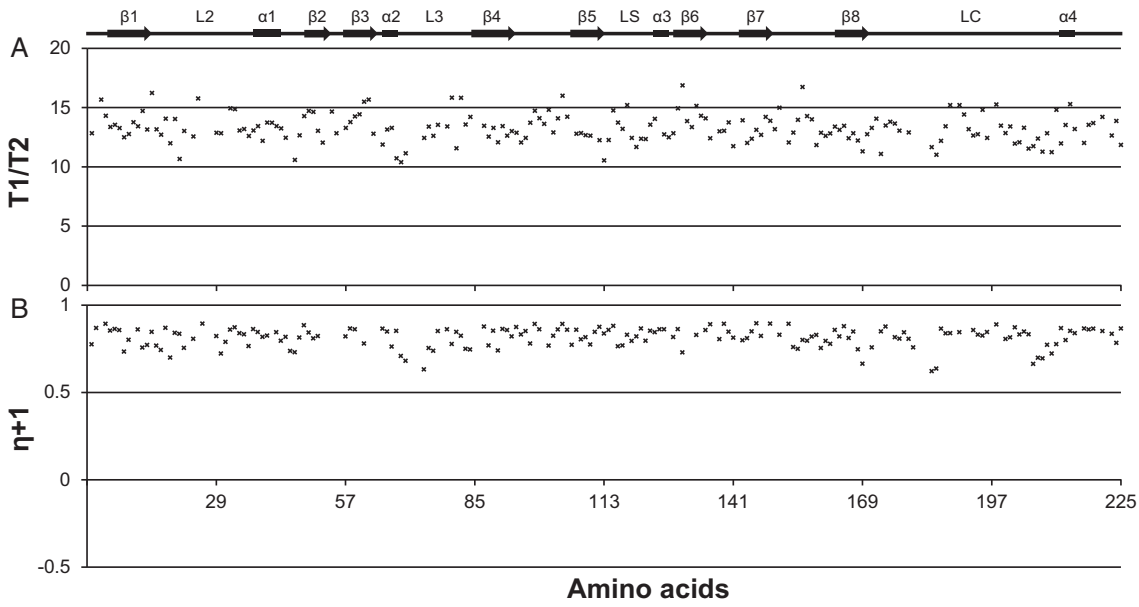

Fig. S2. The ^{15}N relaxation data and ^{15}N-{^{1}H} heteronuclear NOEs for *apo-Nc*LPMO9C. (*A*) Longitudinal relaxation time/transverse relaxation time (T_1/T_2). The T_1:T_2 ratio is a direct measure of the correlation time for overall rotational tumbling of the protein. The average T_1:T_2 ratio for *Nc*LPMO9C was calculated to be 13.8 ± 0.7 (SD), which, assuming a spherical particle, corresponds to an overall rotational correlation time, τ_c, of 11.8 ± 0.5 ns (SD) (44), indicating a tightly packed structure. The Northeast Structural Genomics Consortium has a database with NMR-determined correlation times for globular proteins with different molecular weights. The protein WR73 (which is a 21.9-kDa globular protein) has a longer τ_c = 13.0 ns. The database is available online at www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time (accessed March 22, 2016). (*B*) Steady-state ^{15}N-{^{1}H} NOEs measured for the backbone amide nitrogen atoms.

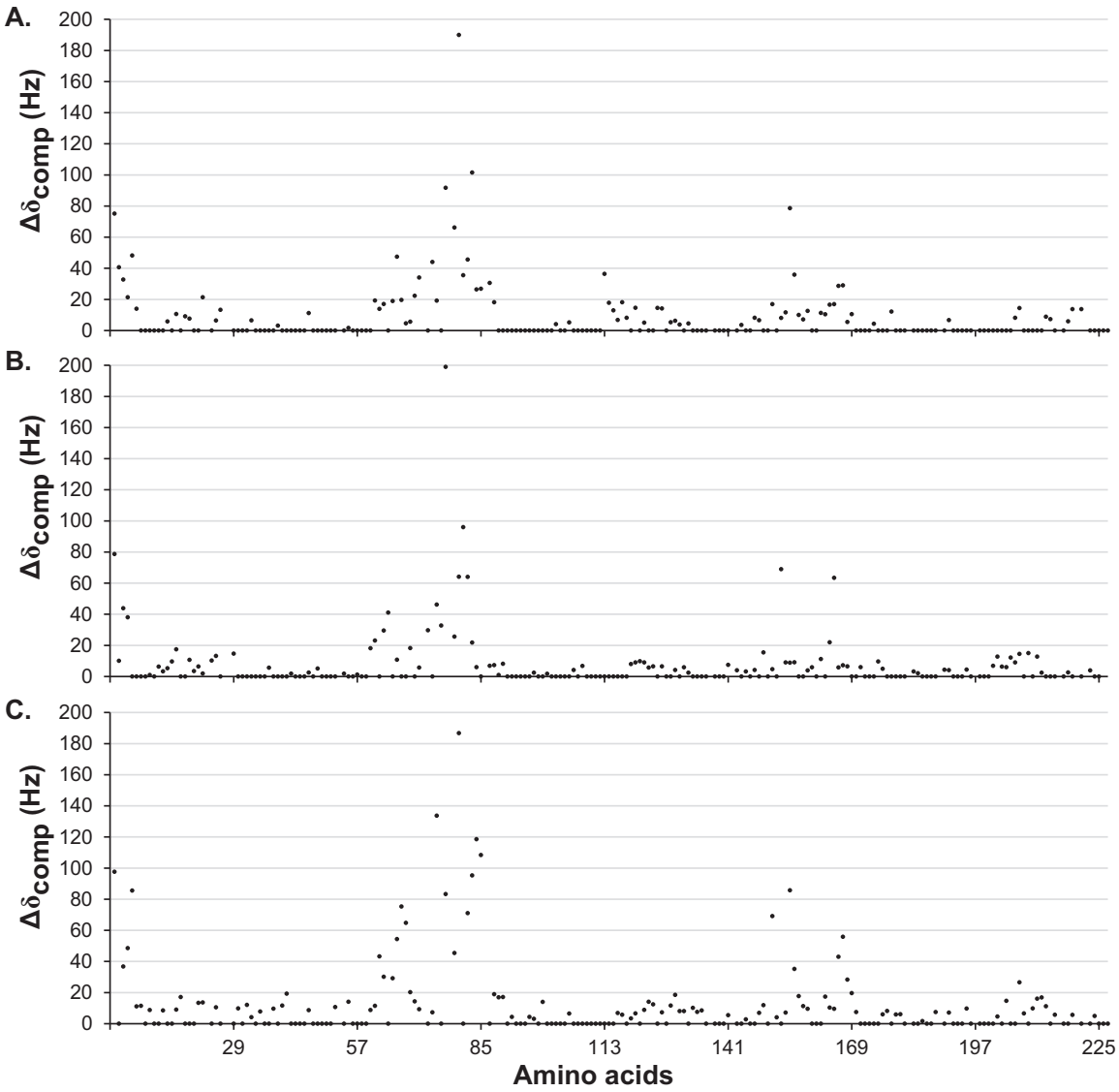

Fig. S3. Compound change in chemical shift (^{15}N-HSQC) upon adding (*A*) 2.6 mM Glc_6, (*B*) 1.3 mM XG_{14}, or (*C*) 4.2 μM polyXG to *apo-Nc*LPMO9C. No change in the chemical shifts was observed upon adding 4.5 mM $GlcNAc_6$. For His1, clear effects on the side chain chemical shifts were observed in a ^{13}C-aromatic HSQC spectrum.

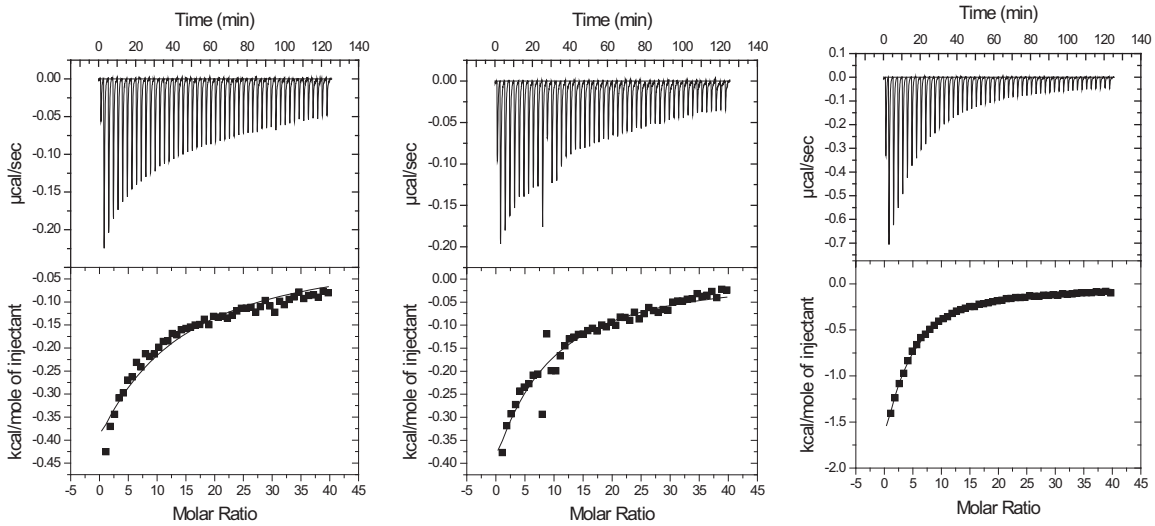

Fig. S4. Thermograms (*Upper*) and binding isotherms with theoretical fits (*Lower*) for the binding of 4.0 mM XG14 to 15 µM of *Nc*LPMO9C *apo* (*Left*), in the presence of Cu^{2+} (*Middle*), and in the presence of Cu^{2+}/CN^{-} (*Right*) at t = 25 °C.

Paper II

Carbohydrate Research 448 (2017) 212–219

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

A novel expression system for lytic polysaccharide monooxygenases

Gaston Courtade [a], Simone Balzer Le [b], Gerd Inger Sætrom [a], Trygve Brautaset [a], Finn L. Aachmann [a, *]

[a] NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491, Trondheim, Norway
[b] Department of Biotechnology and Nanomedicine, SINTEF Materials and Chemistry, N-7034, Trondheim, Norway

ARTICLE INFO

Article history:
Received 18 January 2017
Received in revised form
13 February 2017
Accepted 13 February 2017
Available online 14 February 2017

Keywords:
Heterologous expression system
XylS/*Pm*
Protein production
Lytic polysaccharide monooxygenase (LPMO)
High cell-density cultivation (HCDC)

ABSTRACT

Lytic polysaccharide monooxygenases (LPMOs) are key enzymatic players of lignocellulosic biomass degradation processes. As such, they have been introduced in cellulolytic cocktails for more efficient and less expensive lignocellulose saccharification. The recombinant production of LPMOs in bacteria for scientific investigations using vectors typically based on the T7 and *lac*UV5 promoters has been hampered by low yields. Reasons for this have been catabolite repression when producing the proteins in defined media with glucose as the sole carbon source, as well as the lack of an inducible expression system that allows controlled production of LPMOs that are correctly processed during translocation to the periplasmic space. A cassette vector design containing the XylS/*Pm* system was constructed and evaluated, showing that the expression cassette could easily be used for exchanging LPMO coding genes with or without signal sequences. The cassette was shown to reliably produce mature (translocated) LPMOs under controlled conditions that were achieved by using a low dosage (0.1 mM) of the *Pm* inducer *m*-toluic acid and a low (16 °C) cultivation temperature after induction. Furthermore, the signal sequences of five bacterial LPMOs were tested, and the signal sequence of LPMO10A from *Serratia marcescens* was found to give highest levels of recombinant protein production and translocation. The LPMO expression cassette was also evaluated in cultivations using defined media with glucose as the sole carbon source with a product yield of 7–22 mg per L of culture in shaking flasks. The integrity of the recombinant proteins were analyzed using NMR spectroscopy, showing that the system produced correctly processed and folded LPMOs. Finally, high cell-density cultivations of the recombinant strains were carried out, demonstrating stable protein production levels at similar relative yields (42–1298 mg per L of culture; 3.8–11.6 mg per OD_{600nm} unit) as in shaking flasks, and showing the scale-up potential of the system.

1. Introduction

Cellulose in lignocellulosic biomass (or chitin in fungi and the exocuticle of arthropods) forms crystalline structures that make the polysaccharide resistant to enzymatic hydrolysis. Additional obstacles to enzymatic degradation are introduced by co-polymeric structures with lignin and hemicelluloses such as xylan, glucomannan or xyloglucan. In order to increase the efficiency of biomass utilization, the degradation process normally requires several different enzymes, including hydrolases and the recently discovered lytic polysaccharide monooxygenases (LPMOs) [1]. LPMOs synergize with hydrolytic polysaccharide degrading enzymes to enhance biomass deconstruction, making them a cornerstone of cost-efficient biorefining processes [1,2,10]. They are bacterial and fungal copper-dependent redox enzymes that catalyze oxidative cleavage of β(1 → 4) glycosidic linkages and are classified in the auxiliary-activity families AA9, AA10, AA11 and AA13 [2–9]. The catalytic site of the LPMOs is centered around a copper ion, which is coordinated by the N-terminal His1 (N^H), its side-chain ($N^{\delta 1}$) and the side-chain ($N^{\varepsilon 2}$) of another His. The premature LPMO is synthesized (translated) with an N-terminal signal peptide, which is cleaved off upon translocation to the periplasmic space to produce the native LPMO domain with an N-terminal His that is able to bind copper to form its active site [2,11].

Bacterial LPMOs have been mainly produced recombinantly

* Corresponding author. Sem Sælands vei 6, Kjemi IV 4-140, NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491, Trondheim, Norway.
E-mail address: finn.l.aachmann@ntnu.no (F.L. Aachmann).

http://dx.doi.org/10.1016/j.carres.2017.02.003
0008-6215/

using the pRSET and pET vector systems [2,12,13]. These vectors harbor a T7 or a T7lac promoter recognized by T7 RNA polymerase, which is in turn under transcriptional control of the inducible *lac*UV5 promoter in *Escherichia coli* BL21(DE3) or its derivatives [14,15]. It has been observed that induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) results in a mixture of native and premature (non-cleaved signal peptide) LPMO and cell lysis. This is due to the all-or-none induction in this expression system, which does not allow tuning of protein production [16]. One solution to this issue has been to rely on the leaky expression of the *lac*UV5 promoter [13] in the absence of IPTG, resulting in LPMO yields in the order of 3–9 mg per L of culture [17,18]. While the leaky expression approach has worked well for small-scale batch production of protein in complex media like lysogeny broth (LB), it poses a problem when growing cells in defined media with glucose as the sole carbon source (a requirement to produce isotopically labeled proteins for use in nuclear magnetic resonance (NMR) investigations). A high glucose concentration in the growth medium leads to catabolite repression of the *lac*UV5 promoter, resulting in downregulation of expression of the T7 RNA polymerase encoding gene, and thus low expression of the target LPMO genes [19]. Furthermore, systems based on "leaky expression" lack proper control of protein production, leading to large batch-to-batch variations during in our lab.

A solution to this problem is to use an expression system that is not influenced by the carbohydrate metabolism of the cell. One such expression system is based on the XylS/*Pm* regulator/promoter system [20], which uses benzoic acid derivatives (e.g. *m*-toluic acid) as inducers for recombinant gene expression. XylS is a positive transcriptional regulator that upon binding *m*-toluic acid becomes activated and binds to *Pm*, stimulating transcription. Expression vectors relying on the XylS/*Pm* system have several advantages that include: (1) regulation that is independent of the carbohydrate metabolism of the host [21], (2) tight regulation with minimal leaky expression in absence of inducer [22], (3) dose-dependent induction, allowing tuning of protein production [23], (4) broad host-range compatibility by using the RK2 replicon [20], and (5) uniform expression in the bacterial population as the inducer molecule enters the cell by passive diffusion (no uptake system required) [22].

In addition to a regulatable promoter, the choice of an appropriate translocation signal sequence is of paramount importance due to the need for correct processing of the premature protein. This ensures the presence of His as the first amino acid in the N-terminus (essential for a fully functional catalytic site, as previously stated), as the signal peptide is cleaved during translocation to the periplasmic space [25]. Moreover, signal sequences can influence the total expression levels of recombinant proteins [26].

Recently, the use of the XylS/*Pm* regulator/promotor system for the production of an isotopically labeled LPMO, *Bl*LPMO10A, was demonstrated [24]. In this study, we have built upon this early proof of principle and focused on generalizing it for bacterial LPMO production. The aim of the present study was thus to design a novel standard cassette system that allows LPMOs and translocation signal sequences to be easily exchanged. The performance of the system was evaluated by using five different LPMOs and testing translocation with six signal sequences. Furthermore, high-level production and translocation of bacterial LPMOs in complex and isotope labeled media using the XylS/*Pm* system was demonstrated.

2. Materials and methods

2.1. Biological materials and DNA manipulations

All bacterial strains, vectors, primers and LPMO encoding genes used in this study are described in Table 1. *E. coli* strain DH5α was used as a general cloning host, *E. coli* T7 Express and *E. coli* RV308 were used for recombinant protein production in small-scale shaking flask cultivations, while *E. coli* RV308 was used as the host for LPMO production under high cell-density cultivations (HCDC) in fermentors. Chemically competent *E. coli* cells were transformed using a heat-shock protocol and grown at 37 °C in liquid or solid LB. All recombinant cell cultures were supplemented with 100 μg/mL ampicillin. Standard recombinant DNA techniques were used, unless stated otherwise, according to Sambrook and Russel [27]. Plasmid DNA was isolated using the Wizard® Plus SV Minipreps DNA purification system (Promega) or NucleoBond® Xtra Midi (Macherey-Nagel). Linear DNA fragments were purified from agarose gels using the Zymoclean™ Gel DNA Recovery kit (Zymo Research). PCR were run using standard conditions, with the primers listed in Table 1, and the products were purified using the QIAquick® PCR purification kit (Qiagen).

2.2. Construction of the LPMO expression vectors

In total, six different LPMO encoding genes were investigated and they were expressed with their native signal sequence (nSP) or fused to the signal sequence of the *Sm*LPMO10A gene from *Serratia marcescens* (SP), except *Bl*LPMO10A, which was only expressed with SP. Analysis of patterns in the signal sequences to predict N-, H- and C-domains in the signal peptides was performed using the SignalP 4.1 Server [28].

Plasmid pGM29 [26] without the *gm-csf* insert was used as a vector backbone for the construction of LPMO expression plasmids, and in total ten different LPMO expression vectors were constructed (Table 1). Initially, plasmid pUC57_SP_*Sm* was completely digested with NdeI and HindIII (New England Biolabs) and plasmid pGM29 was partially digested with NdeI and HindIII. The resulting 8.2 kb fragment derived from pGM29 and the 0.6 kb fragment from pUC57_SP_*Sm* were purified, ligated and transformed to *E. coli*. The resulting vector, pJB_SP_*Sm*, was then used for one-step cloning of the LPMO coding regions using the NcoI/NotI restriction sites, and alternatively the LPMO coding region together with a signal sequence by using the NdeI/NotI sites (Fig. 1). As controls, we constructed one vector with the alternative OmpA signal sequence, and the vector pAT64 was included as a negative control. Target LPMOs and signal sequences were cloned using PCR, except for pJB_SP_*Jd*, which was cloned by sequence- and ligation-independent cloning (SLIC) as described by Jeong et al. [29]. All of the plasmids in Table 1 were constructed and expressed to test the performance of the system when combining LPMOs with different signal sequences.

2.3. Production and purification of the recombinant LPMOs from shaking flask cultivations

LPMO production and purification, recombinant cells were grown in ^{15}N-labeled M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl) supplemented with 98% $(^{15}NH_4)_2SO_4$, 4 g/L glucose, 10 mL Bioexpress Cell Growth Media (Cambridge Isotope Laboratories, Tewksbury, MA, USA), 5 mL Gibco™ MEM Vitamin Solution (100x), 2 mM $MgSO_4$ and 10 mL Trace Metal solution (0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, 1 g/L $CaCl_2$) or in 2xLB medium (20 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl).

Pre-cultures were made by inoculating 10 mL LB with recombinant cells followed by incubation at 30 °C and 225 rpm overnight. Main cultures were made in shaking flasks by inoculating 500 mL ^{15}N-labeled M9 medium or 2xLB (100 mL or 500 mL) with 1% pre-culture and incubating at 30 °C and 225 rpm to $OD_{600\ nm}$–0.8. The culture was cooled on ice for 5 min and then induced with 0.1 mM *m*-

Table 1
LPMO genes, vectors, primers and cells used in this study. SP is the signal sequence from *Sm*LPMO10 while nSP is the native and unmodified signal sequence for each protein. All six LPMOs are of bacterial origin and were chosen because of their relevance in recent publications [24,30,31], and to include both chitin-active and cellulose-active enzymes. Restriction sites in the primer sequences are underlined (NdeI: CATATG, NcoI: CCATGGT, NotI: GCGGCCGC, HindIII: AAGCTT).

Source organism and LPMO	Plasmid	Description	Source
Serratia marcescens	pUC57_SP_*Sm*	pUC57 plasmid containing *Sm*LPMO10A from *S. marcescens* with SP	GenScript
*Sm*LPMO10A	pJB_SP_*Sm*	pJB plasmid containing *Sm*LPMO10A from *S. marcescens* with SP	This study
Bacillus licheniformis *Bl*LPMO10A	pJB_SP_*Bl*	pJB containing *Bl*LPMO10A from *B. licheniformis* with SP	Ref. [24]
*Jonesia denitrificans**Jd*LPMO10A	pRSET_nSP_*Jd*	pRSET containing *Jd*LPMO10A from *J. denitrificans* with nSP	Ref. [32]
	pJB_nSP_*Jd*	pJB containing *Jd*LPMO10A from *J. denitrificans* with nSP	This study
	pJB_SP_*Jd*	pJB containing *Jd*LPMO10A from *J. denitrificans* with SP	This study
Streptomyces coelicolor	pJB_nSP_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with nSP	This study
*Sc*LPMO10C	pJB_SP_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with SP	This study
	pJB_OmpA_*Sc*	pJB containing *Sc*LPMO10C from *S. coelicolor* with OmpA	This study
Micromonospora aurantiaca	pUC57_nSP_*Ma*	pUC57 containing *Ma*LPMO10B from *M. aurantiaca* with nSP	GenScript
*Ma*LPMO10B	pJB_nSP_*Ma*	pJB containing *Ma*LPMO10B from *M. aurantiaca* with nSP	This study
	pJB_SP_*Ma*	pJB containing *Ma*LPMO10B from *M. aurantiaca* with SP	This study
Cellvibrio japonicus	pUC57_nSP_*Cj*	pUC57 containing *Cj*LPMO10A from *C. japonicus* with nSP	GenScript
*Cj*LPMO10A	pJB_nSP_*Cj*	pJB containing *Cj*LPMO10A from *C. japonicus* with nSP	This study
	pJB_SP_*Cj*	pJB containing *Cj*LPMO10A from *C. japonicus* with SP	This study
	pGM29	RK2 based plasmid harboring XylS/*Pm* regulator/promoter system, *bla* as reporter gene	Ref. [26]
	pAT64	A vector the XylS/*Pm* system controlling the expression of a interferon-α-coding gene with the PelB signal sequence	Ref. [26]

Primer	Primer sequences (5′–3′)
Jd_fwd_nSP	TACTTCATATGAAGAAGAGAAAGTTGAGAGC
Jd_fwd_SP	CAACAAGCCAACGCCCATGGTTGGGTGACAGATCC
Jd_rev	CTAGCAAGCTTGCGGCCGCTTATGAGACCACAACATCCATAC
Sc_fwd_nSP	ATGAACATATGAACAAAACTTCCC
Sc_fwd_SP	CAATGCCCATGGTAGCGTGGTTGATCCG
Sc_rev	AAGCTTGCGGCCGCTTAGCCGCCGTCGAAG
Ma_fwd_nSP	TACTTCATATGTCAACGCCGTATCG
Ma_fwd_SP	CAACGCCCATGGTAGCGTGGTTGATCCGG
Ma_rev	AAGCTTGCGGCCGCTTAGCCAAAGTCAACATCGCTG
Cj_fwd_nSP	TACTTCATATGTTCAATACCCGTCACC
Cj_fwd_SP	CAACGCCCATGGTTATGTGAGCTCTCCGAAAAG
Cj_rev	AAGCTTGCGGCCGCTTAGCCGAAATCAACGTCAATG

Strains	Description	Source
E. coli T7 Express, a BL21(DE3) derivative	F- λ- *fhuA2* [*lon*] *ompT lacZ*::T7 gene 1 *gal sulA*11 Δ(*mcrC-mrr*)*114*::*IS10 R*(*mcr-*73::*miniTn10-TetS*)*2 R*(*zgb-210*::*Tn10*)(TetS) endA1 [*dcm*]	New England Biolabs
E. coli RV308	*lacI*^{q}-, *su*-, Δ*lacX*74, *gal IS II*::*OP308*, *strA* K12 derivative used for industrial protein production. ATCC strain 31608, deposited by Genentech	ATCC31608
E.coli DH5α	λ^{-} φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17*(r_{K}^{-} m_{K}^{-}) *supE44 thi-1 gyrA relA1*	Bethesda Research Laboratories

toluic acid, followed by incubation at 16 °C and 225 rpm for 20 h. Cells were harvested by centrifugation for 10 min at 4000 × g and 4 °C and periplasmic fractions were prepared by the osmotic shock method as follows: The pellet was resuspended in 30 mL spheroplast buffer (100 mM (0.1 M) Tris-HCl, 0.5 M sucrose, 0.5 mM EDTA, pH 7.5 or 8.5) with half-a-tablet cOmplete™ ULTRA protease inhibitor (Roche). After 5 min incubation on ice followed by a 10 min centrifugation at 5000 × g and 4 °C, the pellet was incubated for 10 min at room temperature and then resuspended on ice in 25 mL ice-cold dd-H_2O supplemented with half-a-tablet cOmplete™ ULTRA protease inhibitor (Roche). 45 s after the pellet was fully suspended, 1.25 mL ice-cold $MgCl_2$ (20 mM) were added to the suspension, which was centrifuged for 10 min at 15000 × g and 4 °C. The supernatant was filtered using a filter (0.22 μm pore size) prior to further protein purification. For total protein extraction experiments, the whole harvested pellet was resuspended in 25 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 1 M $(NH_4)_2SO_4$) and the suspension was sonicated using a Branson Sonifier equipped with a microtip.

*Cj*LPMO10A and *Bl*LPMO10A were purified by loading periplasmic extracts in Buffer A (50 mM Tris-HCl pH 8.5) onto a 5 mL HiTrap® DEAE FF anion exchanger (GE Life Sciences) connected to an ÄKTA Prime Plus FPLC system (GE Life Sciences). Recombinant LPMOs were eluted by using a linear salt gradient (0–500 mM NaCl) over 200 min at a flow-rate of 4.5 mL/min. The fractions containing LPMOs were pooled and concentrated using protein concentrators (10 kDa cut-off, Sartorius). For further purification, the *Cj*LPMO10A sample was loaded onto a HiLoad® 16/600 Superdex® 75 pg size-exclusion column (GE Life Sciences), with a running buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl) and a flow-rate of 1 mL/min.

*Sm*LPMO10A was purified by chitin affinity chromatography as described previously [13], using chitin beads (New England Biolabs) and a buffer consisting of 50 mM Tris-HCl pH 8.0 and 1 M $(NH_4)_2SO_4$. The protein was eluted with 20 mM sodium acetate buffer pH 3.5 using a flow-rate of 2 mL/min. SDS-PAGE gels were run under denaturing conditions using ClearPAGE™ 12% gels and ClearPAGE™ SDS-R Running buffer (C.B.S. Scientific) followed by staining with InstantBlue™ (C.B.S. Scientific). Precision Plus Protein™ standards (Bio-Rad Laboratories) were used for the identification of target proteins.

2.4. NMR spectroscopy

Protein concentrations in the final NMR samples were determined by measuring their A_{280nm} and deducing the protein concentration based on the theoretical extinction coefficient and molecular weight calculated using the ProtParam tool (web.expasy.org/tools/protparam/) [33]. NMR spectra of 0.5–1.2 mM target LPMO samples in 25 mM sodium phosphate buffer pH 5.5 with 10 mM NaCl were recorded at 25 °C on a Bruker Ascend 800 MHz spectrometer Avance III HD equipped with a 5-mm Z-gradient CP-TCI (H/C/N) cryogenic

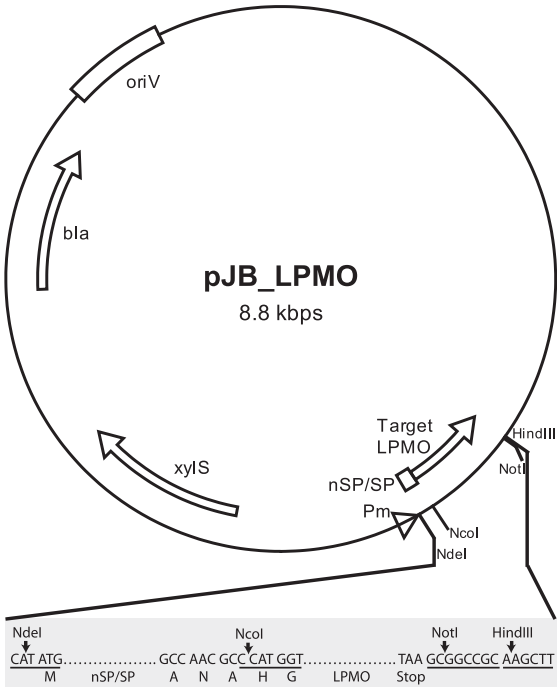

Fig. 1. Vector map of the LPMO expression plasmid pJB_SP_*Sm*. The RK2 replicon (*oriV* and *trfA271C*), the β-lactamase *bla* gene, the *xylS* gene, and the *Pm* promoter region are depicted. The cloning sites for the native signal sequence (nSP) or the *Sm*LPMO10A signal sequence (SP), and the target LPMO gene downstream of the *Pm* promoter, as well as the reading frame are shown in the detail in the grey box. Additional elements relevant for stable maintenance as well as transcriptional terminators are not shown (see Ref. [23] for further information).

probe at the NT-NMR-Center/Norwegian NMR Platform (NNP). NMR data were processed using Bruker TopSpin version 3.5.

2.5. High cell-density cultivation (HCDC)

E. coli RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*) were cultivated in bioreactors to reach high cell-density. All fermentations were performed in 3 L Applikon bioreactors with a 0.75 L working volume. The pH was maintained at pH 6.8 by addition of 30% (w/v) NH_3 (aq). The dissolved oxygen was maintained at 20% (v/v) by adjusting the airflow from 0.35 to 1.5 L/L of medium per min.

Fermentations of both strains were carried out using a 3xLB + glycerol medium (30 g/L tryptone, 15 g/L yeast extract, 10 g/L NaCl and 5 g/L glycerol) and Hf medium (main culture medium described previously [34]) supplemented with 100 μg/mL ampicillin. Fermentations on 3xLB were performed in batch, where the main culture medium was inoculated with pre-culture (in LB medium) to give a starting OD_{600nm}—0.01 and incubated at 30 °C to OD_{600nm}—3. After induction with 0.5 mM *m*-toluic acid, the incubation temperature was reduced to 16 °C.

Fermentations in Hf medium were carried out in fed-batch mode. Pre-culture medium [34] was inoculated with a starter culture (prepared in LB medium To start the HCDC, Hf was inoculated with pre-culture to a give a starting OD_{600nm}—0.05 and the fermentation was divided into three phases as follows: (i) Batch phase. Glucose (25 g/L) was added to the freshly inoculated main culture medium, which was incubated at 30 °C until OD_{600nm}—35. (ii) Exponential feeding phase. Growth was controlled to a specific growth rate (μ) of 0.25 h^{-1} by feeding with 0.6 g/L $MgSO_4$ and 10 g/L glucose. The starting feeding rate was 14 g/h, which was increased exponentially to 35 g/h, and thereafter remained constant. (iii) Induction phase. When OD_{600nm}—90 had been reached, the culture was induced with 0.5 mM *m*-toluic acid, the incubation temperature was reduced to 16 °C, and the feeding rate was reduced to 5 g/h. Cell concentration was measured as OD_{600nm}. The protein concentration in the HCDC was estimated by the following band densitometry analysis. Samples were taken from the cultures at different time points and analyzed on SDS-PAGE. Digital images of the SDS-PAGE gels were captured with the ChemiDoc™ XRS + system (Bio-Rad Laboratories) in transillumination mode and the images were analyzed using the Image Lab™ software (Bio-Rad Laboratories). The protein concentration in the bands at ~20 kDa (i.e. the molecular weight of LPMOs) was estimated by the ratio between their intensity and the intensity of the Bio-Rad Precision Plus Protein™ Dual Color Marker standard band at 20 kDa, which has a known concentration of 15 ng/μl. Cultures were harvested by centrifugation at 4500 × g for 10 min. Periplasmic extraction and purification were carried out using the same methodology as described in the previous section.

3. Results and discussion

The need for a reliable expression system for the production of LPMOs led to the design and construction of a plasmid vector based on the XylS/*Pm* expression cassette. In order to evaluate the performance of the system, five bacterial (AA10) LPMOs were cloned in the newly designed vector together with their native signal sequence. Moreover, six LPMO coding regions were combined with the signal sequence from *Sm*LPMO10A (Table 2) to test the effect of the signal peptide for translocation and overall production of mature LPMOs. In total, eleven vectors were constructed (Table 1), and LPMO production levels were evaluated for all the respective recombinant *E. coli* strains.

3.1. Construction of a new LPMO expression cassette based on the XylS/Pm expression system

A novel, carbon-source independent expression cassette pJB_SP_*Sm* was developed with the aim of producing correctly

Table 2
Amino acid sequence of the native signal peptides for the selected LPMOs and an OmpA-based signal peptide. Gram classification and order of the source organism are included. Text formatting indicates the different features of the sequences as follows. **Bold**: N-domain, regular: H-domain, underlined: C-domain, *italics*: helix-breaking residues Gly and Pro, ***bold italics:*** interspersed positively charged residue.

Source	Amino acid sequence	Gram classification	Order
*Sm*LPMO10A	**MNKTSR**TLLSLGLLSAAMFGVSQQANA	Gram-negative	Enterobacteriales
*Jd*LPMO10A	**MKKRKLR**ASAAIAVLLGAGLVPALSATPAAA	Gram-positive	Actinomycetales
*Sc*LPMO10C	**MVRRTR**LLTLAAVLATLLGSLGVTLLLGQGRAEA	Gram-positive	Actinomycetales
*Cj*LPMO10A	**MFNTRH**LLAGVSQLV***K***PASMMILAMASTLAIHEASA	Gram-negative	Pseudomonales
*Ma*LPMO10A	**MSTPYRR***PLP*LAAAILGVCAVVAALLTTAFSGPASA	Gram-positive	Actinomycetales
OmpA-like (*E. coli*)	**MKK**TAIAIAVALAGFATVAQAAAMA	Gram-negative	Enterobacteriales

processed LPMOs from the *Pm* promoter in *E. coli* using isotopically enriched media (Fig. 1). The result is a cassette design, allowing the LPMO coding region to be easily exchanged by using a combination of restriction enzymes for linearization of the backbone. Insertion of an LPMO gene while keeping the signal sequence of *Sm*LPMO10A is possible with NcoI and NotI. Inserting an LPMO gene with its own signal sequence requires NdeI and NotI, while a combination of NdeI and NcoI exchanges the signal sequence region only. Many bacterial signal sequences like OmpA, PelB, MalE end with the amino acids Ala-X-Ala, and so do the LPMO signal sequences included herein (Table 2) [35]. Using the NcoI site, the transition from Ala in the signal sequence and the His in the start of the LPMO gene (Fig. 1) should ensure that the N-terminus of the LPMO is processed correctly during translocation.

Overall, all signal sequences used have the standard signal peptide architecture consisting of the N-region (composed of 1–5 residues), the hydrophobic H-region (7–15 residues) and the C-region (3–7 residues), which is a determinant of specificity [36]. In particular, the end of the C-region has a characteristic Ala-X-Ala motif [37], where the cleavage site is located. However, they differ in typical features such as the number of positively charged amino acids in the N-region, the number and position of helix-breaking Pro, Gly or Ser residues [38] as well as their length. Signal peptide pattern analysis revealed that two of the signal peptides (*Cj*LPMO10A and OmpA-like) most likely form two consecutive helices instead of a typical single alpha helix in the H-domain, and that the signal sequence from *Sm*LPMO10A seems to be in best agreement with this “consensus” signal sequence.

3.2. The pJB_SP_Sm and pJB_SP_Bl vectors can be used for efficient production and translocation of LPMOs

Controlled protein production is a prerequisite for correct processing of the signal peptide and translocation of LPMOs to the periplasmic space. A hallmark of the XylS/*Pm* expression cassette is the dependency on the inducer dosage [23], allowing control of the recombinant protein production levels. The constructed expression vectors were evaluated by cultivating *E. coli* strains RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*), expressing *Sm*LPMO10A and *Bl*LPMO10A, respectively (see Table 1), in 100 mL shaking flasks, and protein production was controlled by using a low inducer concentration (0.1 mM *m*-toluic acid) and a low cultivation temperature (16 °C). The semiquantitative amount of translocated protein (Table 3) was determined by using SDS-PAGE from the intensity of the bands at 21 kDa and 19.4 kDa (*Sm*LPMO10A and *Bl*LPMO10A, respectively) in the periplasmic extract fraction, and a ½ dilution series (Fig. 2). The absence of visible bands at 21 kDa and 19.4 kDa in the cell pellet after periplasmic extraction (Fig. 2) indicates low accumulation of recombinant protein in the cytoplasm. These results show that, under the chosen cultivation conditions, the XylS/*Pm* expression cassette efficiently produced LPMOs, which were then successfully translocated to the periplasm. Moreover, the expression system showed minimal levels of leaky expression (in absence of inducer), a feature consistent with the XylS/*Pm* system [22].

Table 3
Relative production levels of the different mature LPMOs using three different signal peptides are indicated with plus (+) signs, where +++ corresponds to the highest observed level of expression. N/A means that data for that LPMO/signal peptide combination is not available. In the case of *Sm*LPMO10A, nSP and SP are the same native signal peptide.

LPMO	Source	Native signal peptide (nSP)	*Sm*LPMO10A signal peptide (SP)
*Sm*LPMO10A	*S. marcescens*	+++	
*Jd*LPMO10A	*J. denitrificans*	+	+++
*Sc*LPMO10C	*S. coelicolor*	+	+++
*Cj*LPMO10A	*C. japonicus*	++	+++
*Ma*LPMO10A	*M. aurantiaca*	+	+++
*Bl*LPMO10A	*B. licheniformis*	N/A	+++

3.3. The native signal sequence of SmLPMO10A can be used for efficient translocation of heterologous LPMOs

A signal peptide is vital for the correct translocation of the protein to the periplasmic space. Since the native signal sequences of the target LPMOs (Table 2) showed large translocation variability, it was of interest to test the effects of substituting with alternative signal sequences. Thus, using the XylS/*Pm* cassette design, LPMO genes were inserted with their native signal sequences (nSP) or with the signal sequence of *Sm*LPMO10A (SP), which has been shown to successfully produce mature (correctly processed) LPMOs [24].

Table 3 shows a summary of the test expressions of the six target LPMOs. Where applicable, a comparison of the production level of mature protein for each signal peptide is given. The protein levels were assessed semiquantitatively using serial ½ dilutions of the periplasmic extracts and SDS-PAGE, as described in the previous section.

It is known that combinations of host strain, signal sequences and target proteins can influence translocation, solubility and expression levels [26,39]. The SP from *S. marcescens* was here shown to produce the highest levels of translocated protein compared to the native signal sequences (Table 3). In addition, an *E. coli* OmpA-like signal sequence was tested together with *Sc*LPMO10C, but using this signal sequence did not result in higher

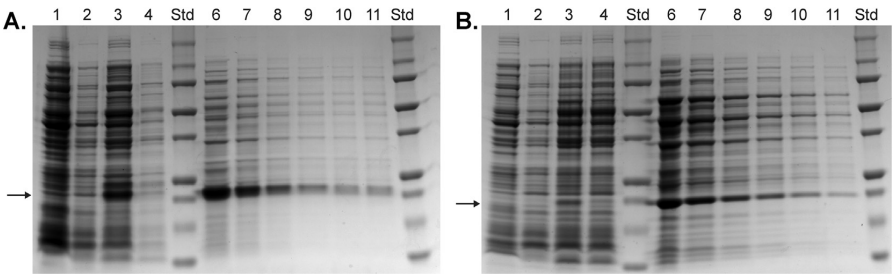

Fig. 2. SDS-PAGE analysis of recombinant strains producing *Sm*LPMO10A and *Bl*LPMO10A. Panel A corresponds to pJB_SP_*Sm*, Panel B corresponds to pJB_SP_*Bl*. For both panels the lanes were as follows: (1) induced negative control vector (pAT64), (2) non-induced cells, (3) induced cells, (4) cell pellet after periplasmic extraction, (6–11) serial ½ dilutions of the extract, and (Std) Bio-Rad Precision Plus Protein™ Dual Color Marker. The arrows mark bands corresponding to the molecular weights of *Sm*LPMO10A (21 kDa, Panel A) and *Bl*LPMO10A (19.4 kDa, Panel B).

protein production levels than SP.

An explanation for the superior performance of SP from *S. marcescens* could be that signal sequences from more closely related bacteria would work better in a certain expression host (Table 2). Indeed, *S. marcescens*, which is a Gram-negative Enterobacterium commonly found in the gastrointestinal tract just like *E. coli*. The signal sequence from LPMO10A in *C. japonicus* (also a Gram-negative bacterium) led to intermediate production levels in *E. coli*. The remaining signal sequences, which gave low production levels, belong to LPMOs from Gram-positive bacteria, which are characterized by marked differences in the signal sequences, such as a longer H-region than signal sequences from Gram-negative organisms [40]. Therefore, it would appear that the signal sequence from an LPMO from an organism that is closely related to the expression host (such as *S. marcescens* and *E. coli*) would perform better than the signal sequence from LPMOs in organisms that are more distantly related to the expression host.

3.4. The XylS/Pm expression cassette is useful for high-level production of isotopically enriched LPMOs

NMR spectroscopy is a versatile and powerful technique for the study of protein structure and function. NMR has provided insights into various aspects of LPMOs including mobility, metal binding, and substrate interactions [41,42]. However, NMR studies require ~0.5–1 mM of LPMO containing the NMR active ^{15}N and ^{13}C isotopes, which have natural abundances of 0.36% and 1.1%, respectively. This means that LPMOs must be produced in isotopically enriched media with defined C and N sources such as ^{13}C-glucose and $(^{15}NH_4)_2SO_4$ [43].

Expression systems traditionally used to produce LPMOs do not perform well when using glucose as the sole carbon source in defined media. For example, as glucose represses the *lac* operon, this impedes the production of LPMOs in defined media using leaky expression of T7 or T7lac promoters-containing vectors (such as pRSET or pET) in *E. coli* strains harboring the T7 polymerase gene (typically under control of the *lac*UV5 promoter). This represents an obstacle in isotope labeling, hence a bottleneck for NMR investigations.

The XylS/*Pm* system is exempt from this problem because its regulation is independent from the carbon source used in *E. coli* [21]. Therefore, the LPMO expression cassette was further tested to evaluate its performance when producing isotopically enriched LPMOs with a defined carbon source (i.e. glucose). Recombinant strains RV308 (pJB_SP_*Sm*), RV308 (pJB_SP_*Bl*) and T7 Express (pJB_SP_*Cj*), were cultivated in defined ^{15}N-enriched M9 medium using 500 mL shaking flasks. These three LPMOs were selected based on the simplicity of their purification protocols (see Materials and Methods) as well as previous knowledge of their ^{15}N HSQC fingerprint spectra [24,44]. Following subsequent purification, the three isotope labeled proteins were analyzed in terms of protein yields and quality using SDS-PAGE, UV/Vis and NMR spectroscopy. The SDS-PAGE results (Fig. 3, top row) show that the elution

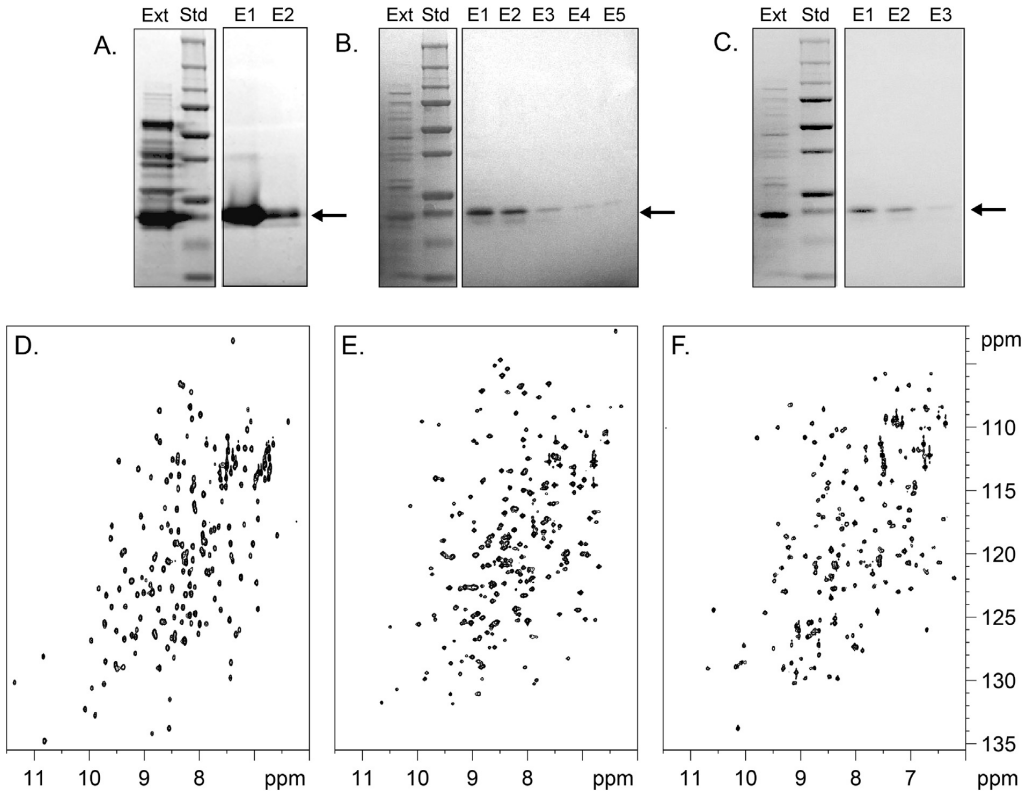

Fig. 3. Top row: Purification of LPMOs from defined ^{15}N-enriched M9 medium. (Std) Bio-Rad Precision Plus Protein™ Dual Color Marker, (Ext) periplasmic extract. Bottom row: ^{15}N HSQC spectra of the LPMO samples, showing the characteristics of pure and correctly processed proteins [43]. A&D: *Sm*LPMO10A, B&E: *Bl*LPMO10A, C&F: *Cj*LPMO10A. . The arrows mark bands corresponding to the molecular weights of each LPMO.

fractions after single step purification (Table 4) contained more than 95% pure protein. Although the data show the presence of weak double bands, the absence of strong, overlapping signals in the center of NMR spectra indicates that this is not a result of incorrect processing of the signal peptide [43,45]. The recombinant protein yields range from 7–26 mg per L of culture (Table 4), which were sufficiently high for NMR analyses, resulting in ^{15}N-HSQC spectra characteristic of pure, correctly processed and folded proteins (Fig. 3, bottom row). Moreover, the product yields were two- to three-fold higher than previously reported for LPMOs produced in LB media using T7 or T7lac containing vectors in *E. coli* containing the T7 polymerase gene [17,18]. This shows that, to date, this XylS/*Pm* system is the best alternative for producing (isotope labeled) bacterial LPMOs in *E. coli*.

3.5. The XylS/Pm expression cassette is applicable to high cell-density cultivation for production of LPMOs

Fermentation in a bioreactor allows precise control of the specific growth rate (μ), ensures that the bacterial population is sampled homogenously, and provides valuable information for further scale-up and optimization of the cultivation process.

The performance of the XylS/*Pm* at HCDC was evaluated by cultivating the recombinant *E. coli* strains RV308 (pJB_SP_*Sm*) and RV308 (pJB_SP_*Bl*) in bioreactors run in batch-mode and 3xLB medium or fed-batch mode and Hf defined medium. The fermentation courses (Fig. 4) show that the production of LPMOs started immediately after XylS/*Pm* induction and, for the fermentations in defined medium, the maximum amount of LPMO was produced within 5 h. Production of LPMOs in 3xLB medium followed cell-growth. Overall, LPMO production remained stable at high cell-density, as observed previously for other XylS/*Pm* containing vectors [23]. Whereas low cell-densities resulted in (protein) yields of 42 and 78 mg per L of culture (3.8 and 11.6 mg per OD unit) for 3xLB in batch-mode (Table 5), which were similar to the yields obtained with shaking flasks (Table 4), cultures in defined medium in fed-batch mode reached higher cell-densities and yields at 514 and 1298 mg per L of culture (3.9 and 9.2 mg per OD_{600nm} unit). No signs of plasmid instability were observed during the fermentation, as the protein yields were maintained or increased, which is expected for the XylS/*Pm* system [23]. The tunable induction of XylS/*Pm* also enabled a controlled and efficient translocation of the target proteins even at high-cell densities, as observed for the shaking flask cultivations. Even at high levels of recombinant LPMO

Table 4
Amount of recombinant LPMO in *E. coli* after incubation in defined ^{15}N-enriched M9 medium and purification.

Plasmid	Purification method	Yield (mg per L of culture)	
pJB_SP_*Sm*	Chitin affinity column	2xLB	26
		M9	22
pJB_SP_*Bl*	Anion exchange DEAE FF	M9	7
pJB_SP_*Cj*	Anion exchange DEAE FF and SEC	M9	10

Table 5
Amount of pure, recombinant LPMO at harvest times in *E. coli* RV308 after HCDC in 3xLB or Hf defined medium. The yields in mg per L of culture were estimated from the SDS-PAGE bands of the elution fractions using the same band densitometry analysis described in Materials and Methods. To facilitate comparison between the different cultures, yields are also given as mg per OD_{600nm} unit.

Plasmid	Media	Yield (mg per L of culture)	Yield (mg per OD_{600nm} unit)
pJB_SP_*Sm*	Defined medium	1298	9.2
	3xLB	78	11.6
pJB_SP_*Bl*	Defined medium	514	3.9
	3xLB	42	3.8

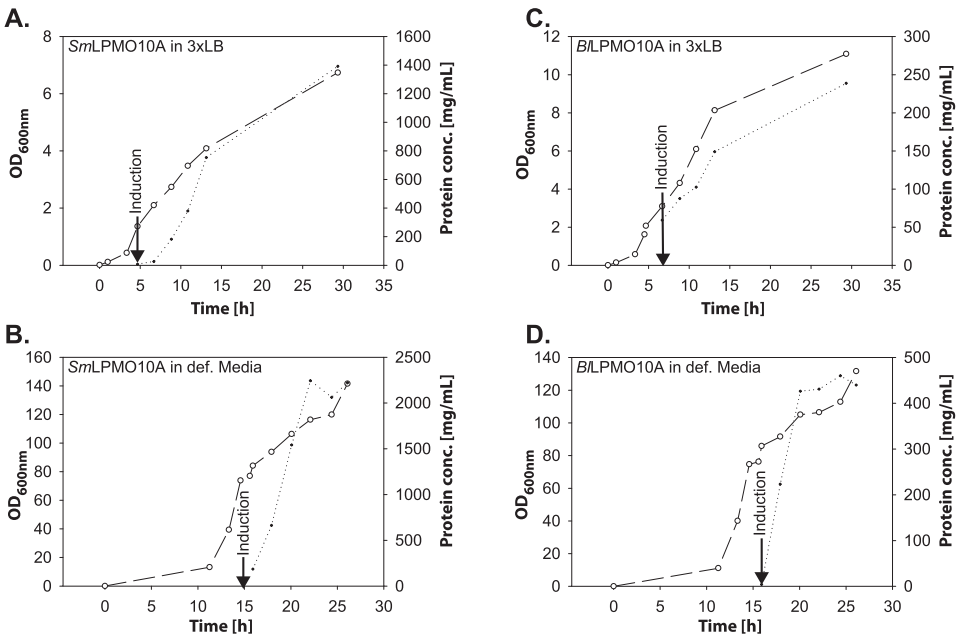

Fig. 4. Fermentation course of *E. coli* RV308 containing pJB_SP_*Sm* (A and B) or pJB_SP_*Bl* (C and D) in 3xLB and Hf defined medium. Time courses of bacterial growth (OD_{600nm}) are shown in dashed lines and LPMO production is shown in dotted lines. An arrow indicates the time of induction with 0.5 mM *m*-toluic acid.

production, no indication of host toxicity was observed during the fermentation. This indicates the scale-up potential of LPMO production using this LPMO expression cassette, even though no further optimization of the cultivation conditions was performed.

4. Conclusions

The need for the development of a controllable expression system for reliable LPMO production, particularly in isotopically enriched media, led to the construction of a novel LPMO expression cassette based on the XylS/*Pm* system. The combination of the XylS/*Pm* regulator/promoter system together with the signal sequence from *Sm*LPMO10A was shown to give the highest production levels of mature LPMOs. The LPMO expression cassette also performed well under HCDC conditions, as well as when producing isotopically enriched LPMOs in defined media with glucose as the sole carbon source. Indeed, the yields reported here using the LPMO expression cassette are the highest to date when producing bacterial LPMOs in *E. coli*.

The XylS/*Pm* system has existed for more than two decades [20] and its relevance has recently been reviewed by Gawin et al. (2017) [46]. While the system is highly flexible and can result in high-level expression of heterologous proteins, it still needs to be fine-tuned to each application. The novel LPMO expression cassette design is an example of such fine-tuning of the XylS/*Pm* system showing its application for the production of LPMOs. Moreover, the cassette system can be further developed for general expression of proteins, especially when isotope labeling is needed (e.g. NMR investigations, MS, etc.). Further refinement of the system could comprise optimization of other features of the vector (e.g. XylS/*Pm* variant, 5′-UTR, plasmid copy number), cultivation parameters (such as composition of the cultivation media, temperature and inducer dosage), and host strains. Altogether, we are confident that the LPMO expression cassette has potential use for scientific and commercial applications.

Acknowledgements

This work was financed by strategic funds from the Norwegian University of Science and Technology, the MARPOL project, and NNP - Norwegian NMR Platform (grants 221576 and 226244 from the Research Council of Norway, respectively). We thank our partners at Vectron Biosolutions AS and the Protein Engineering and Proteomics group at the Norwegian University of Life Sciences who provided insight and expertise that greatly assisted the research. Anne Tøndervik, Per Odin Hansen and Randi Aune from SINTEF are thanked for excellent technical assistance.

References

[1] S.J. Horn, G. Vaaje-Kolstad, B. Westereng, V.G.H. Eijsink, Biotechnol. Biofuels 5 (2012) 45.
[2] G. Vaaje-Kolstad, B. Westereng, S.J. Horn, Z. Liu, H. Zhai, M. Sørlie, et al., Science 330 (2010) 219–222.
[3] R.J. Quinlan, M.D. Sweeney, L. Lo Leggio, H. Otten, J.-C.N. Poulsen, K.S. Johansen, et al., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 15079–15084.
[4] C.M. Phillips, W.T. Beeson, J.H. Cate, M.A. Marletta, ACS Chem. Biol. 6 (2011) 1399–1406.
[5] S. Kim, J. Ståhlberg, M. Sandgren, R.S. Paton, G.T. Beckham, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 149–154.
[6] A. Levasseur, E. Drula, V. Lombard, P.M. Coutinho, B. Henrissat, Biotechnol. Biofuels 6 (2013) 41–54.
[7] G.R. Hemsworth, B. Henrissat, G.J. Davies, P.H. Walton, Nat. Chem. Biol. 10 (2014) 122–126.
[8] V.V. Vu, W.T. Beeson, E.A. Span, E.R. Farquhar, M.A. Marletta, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 13822–13827.
[9] L. Lo Leggio, T.J. Simmons, J.N. Poulsen, K.E.H. Frandsen, G.R. Hemsworth, M.A. Stringer, et al., Nat. Commun. 6 (2015) 1–9.
[10] K.S. Johansen, Biochem. Soc. Trans. 44 (2016) 143–149.
[11] F. Moser, D. Irwin, S. Chen, D.B. Wilson, Biotechnol. Bioeng. 100 (2008) 1066–1077.
[12] Z. Forsberg, A.K. Mackenzie, M. Sørlie, Å.K. Røhr, R. Helland, A.S. Arvai, et al., Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 8446–8451.
[13] G. Vaaje-Kolstad, D.R. Houston, A.H.K. Riemen, V.G.H. Eijsink, D.M.F. van Aalten, J. Biol. Chem. 280 (2005) 11313–11319.
[14] D. Kennell, H. Riezman, J. Mol. Biol. 114 (1977) 1–21.
[15] G.L. Rosano, E.A. Ceccarelli, Front. Microbiol. 5 (2014) 1–17.
[16] T. Afroz, K. Biliouris, Y. Kaznessis, C.L. Beisel, Mol. Microbiol. 93 (2014) 1093–1103.
[17] G.R. Hemsworth, E.J. Taylor, R.Q. Kim, R.C. Gregory, S.J. Lewis, J.P. Turkenburg, et al., J. Am. Chem. Soc. 135 (2013) 6069–6077.
[18] R.C. Gregory, G.R. Hemsworth, J.P. Turkenburg, S. Hart, P.H. Walton, G.J. Davies, Dalton Trans. 45 (2016) 16904–16912.
[19] A.J. Griffiths, J.H. Miller, D.T. Suzuki, R.C. Lewontin, W.M. Gelbart, Catabolite repression of the lac operon: positive control, in: An Introduction to Genetic Analysis, seventh ed., W.H. Freeman, New York, 2000, pp. 342–343.
[20] J.M. Blatny, T. Brautaset, K. Haugan, J.M. Blatny, T. Brautaset, H.C. Winther-Larsen, Appl. Environ. Microbiol. 63 (1997) 370–379.
[21] T. Brautaset, S.B. Petersen, S. Valla, Metab. Eng. 2 (2000) 104–114.
[22] S. Balzer, V. Kucharova, J. Megerle, R. Lale, T. Brautaset, S. Valla, Microb. Cell Fact. 12 (2013) 1–14.
[23] H. Sletta, A. Nedal, T.E.V. Aune, H. Hellebust, S. Hakvåg, R. Aune, et al., Appl. Environ. Microbiol. 70 (2004) 7033–7039.
[24] G. Courtade, S. Balzer, Z. Forsberg, G. Vaaje-Kolstad, V.G.H. Eijsink, F.L. Aachmann, Biomol. NMR Assign. 9 (2015) 207–210.
[25] F.J.M. Mergulhão, D.K. Summers, G.A. Monteiro, Biotechnol. Adv. 23 (2005) 177–202.
[26] H. Sletta, A. Tøndervik, S. Hakvåg, T.E. Vee Aune, A. Nedal, R. Aune, et al., Appl. Environ. Microbiol. 73 (2007) 906–912.
[27] J. Sambrook, D.W. Russell, Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2001.
[28] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, Nat. Methods 8 (2011) 785–786.
[29] J.Y. Jeong, H.S. Yim, J.Y. Ryu, H.S. Lee, J.H. Lee, D.S. Seen, et al., Appl. Environ. Microbiol. 78 (2012) 5440–5443.
[30] Z. Forsberg, A.K. Røhr, S. Mekasha, K.K. Andersson, V.G.H. Eijsink, G. Vaaje-Kolstad, et al., Biochemistry 53 (2014) 1647–1656.
[31] Z. Forsberg, C.E. Nelson, B. Dalhus, S. Mekasha, J.S.M. Loose, L.I. Crouch, et al., J. Biol. Chem. 291 (2016) 7300–7312.
[32] S. Mekasha, Z. Forsberg, B. Dalhus, J.-P. Bacik, S. Choudhary, C. Schmidt-Dannert, et al., FEBS Lett. 590 (2016) 34–42.
[33] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, et al., Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), The Proteomics Protocols Handook, Springer Publishing, New York, 2005, pp. 571–607.
[34] U. Horn, W. Strittmatter, A. Krebber, U. Knüpfer, M. Kujau, R. Wenderoth, et al., Appl. Microbiol. Biotechnol. 46 (1996) 524–532.
[35] P. Pechsrichuang, C. Songsiriritthigul, D. Haltrich, S. Roytrakul, P. Namvijtr, N. Bonaparte, et al., Springerplus 5 (2016) 1–10.
[36] M. Paetzel, N.C.J. Strynadka, CSBMCB/SCBBMC Bull. (2001) 60–65.
[37] G. von Heijne, Eur. J. Biochem. 133 (1983) 17–21.
[38] G. von Heijne, J. Membr. Biol. 115 (1990) 195–201.
[39] D.M. Retallack, J.C. Schneider, J. Mitchell, L. Chew, H. Liu, Biotechnol. Lett. 29 (2007) 1483–1491.
[40] G. von Heijne, L. Abrahmsèn, FEBS Lett. 244 (1989) 439–446.
[41] F.L. Aachmann, M. Sørlie, G. Skjåk-Bræk, V.G.H. Eijsink, G. Vaaje-Kolstad, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 18779–18784.
[42] G. Courtade, R. Wimmer, Å.K. Røhr, M. Preims, A.K.G. Felice, M. Dimarogona, et al., Proc. Natl. Acad. Sci. 113 (2016) 5922–5927.
[43] A.H. Kwan, M. Mobli, P.R. Gooley, G.F. King, J.P. MacKay, FEBS J. 278 (2011) 687–703.
[44] F.L. Aachmann, V.G.H. Eijsink, G. Vaaje-Kolstad, Biomol. NMR Assign. 5 (2011) 117–119.
[45] V. Chupin, J.A. Killian, J. Breg, H.H. de Jongh, R. Boelens, R. Kaptein, et al., Biochemistry 34 (1995) 11617–11624.
[46] A. Gawin, S. Valla, T. Brautaset, Microb. Biotechnol. (2017). In Press.

Paper III

Is not included due to copyright
available in
Journal of Biological Chemistry 2018 ;Volum 293.(34) s. 13006-13015 https://doi.org/10.1074/jbc.RA118.004269

Paper IV

Is not available due to copyright

Paper V

Paper VI

Is not included due to copyright
avalable in
Biomolecular NMR Assignments 2016 ;Volum 10.(2) s. 277-280
https://doi.org/10.1007/s12104-016-9683-x

Paper VII

Paper VIII

Additional contribution I

SCIENTIFIC REPORTS

OPEN

Human Chitotriosidase: Catalytic Domain or Carbohydrate Binding Module, Who's Leading HCHT's Biological Function

Received: 6 October 2016
Accepted: 10 April 2017
Published online: 05 June 2017

Oscar Crasson[1], Gaston Courtade[2], Raphaël R. Léonard[3], Finn Lillelund Aachmann[2], François Legrand[1], Raffaella Parente[1], Denis Baurain[3], Moreno Galleni[1], Morten Sørlie[4] & Marylène Vandevenne[1]

Chitin is an important structural component of numerous fungal pathogens and parasitic nematodes. The human macrophage chitotriosidase (HCHT) is a chitinase that hydrolyses glycosidic bonds between the N-acetyl-D-glucosamine units of this biopolymer. HCHT belongs to the Glycoside Hydrolase (GH) superfamily and contains a well-characterized catalytic domain appended to a chitin-binding domain ($ChBD_{CHIT1}$). Although its precise biological function remains unclear, HCHT has been described to be involved in innate immunity. In this study, the molecular basis for interaction with insoluble chitin as well as with soluble chito-oligosaccharides has been determined. The results suggest a new mechanism as a common binding mode for many Carbohydrate Binding Modules (CBMs). Furthermore, using a phylogenetic approach, we have analysed the modularity of HCHT and investigated the evolutionary paths of its catalytic and chitin binding domains. The phylogenetic analyses indicate that the $ChBD_{CHIT1}$ domain dictates the biological function of HCHT and not its appended catalytic domain. This observation may also be a general feature of GHs. Altogether, our data have led us to postulate and discuss that HCHT acts as an immune catalyser.

Carbohydrate recognition processes are involved in numerous regulatory pathways such as cell signalling and proliferation, fertilization, embryogenesis and in diseases like cancers. Carbohydrates also play a critical role in pathogen recognition, inflammation and innate immune responses through a large array of carbohydrate binding proteins[1]. Chitin, a water-insoluble homopolysaccharide composed of β-1,4-linked N-acetyl-D-glucosamine (GlcNAc) units, is an abundant structural component of arthropods and various infectious organisms like protozoans (e.g. Plasmodium falciparum), nematodes and fungi[2,3]. As mammals do not produce chitin, this polymer is likely a strategic target for innate immune agents. Defense proteins, including lectins, are known to play a crucial role in the initiation of innate immune mechanisms[4]. These carbohydrate binding proteins include numerous members, which are notably synthetized by many organisms including plants and animals, thereby highlighting their ubiquity and necessity for survival[5]. Some of these lectins are able to bind reversibly to chitin and include a conserved structural motif termed "hevein-fold"; which is rich in polar and aromatic residues[6]. Genes encoding these motifs, associated or not with a catalytic domain, are usually expressed after exposure to chitin-containing pathogens. Among other characteristics, lectins have been shown to bear remarkable anti-fungal properties[7,8]. The macrophage chitotriosidase (HCHT) is one of the three active chitinases synthetized by humans, together with acidic mammalian chitinase (AMCase) and the recently discovered exochitinase, chitobiase (CTBS)[9–11]. This protein is synthetized as a 50-kDa soluble monomeric enzyme that is able to hydrolyse colloidal chitin. This modular protein is composed of a catalytic domain that belongs to family 18 of glycoside hydrolases (GH18) and of a carbohydrate-binding module, named $ChBD_{CHIT1}$. The latter domain is stabilized by 3 disulphide bonds

[1]InBioS - Center for Protein Engineering, Department of Life Sciences, Université de Liège, Sart-Tilman, B4000, Liège, Belgium. [2]Department of Biotechnology, Norwegian Biopolymer Laboratory (NOBIPOL), NTNU - Norwegian University of Science and Technology, N-7491, Trondheim, Norway. [3]InBioS - Eukaryotic Phylogenomics, Department of Life Sciences and PhytoSYSTEMS, Université de Liège, Sart-Tilman, B4000, Liège, Belgium. [4]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, N-1432, Ås, Norway. Correspondence and requests for materials should be addressed to M.V. (email: mvandevenne@ulg.ac.be)

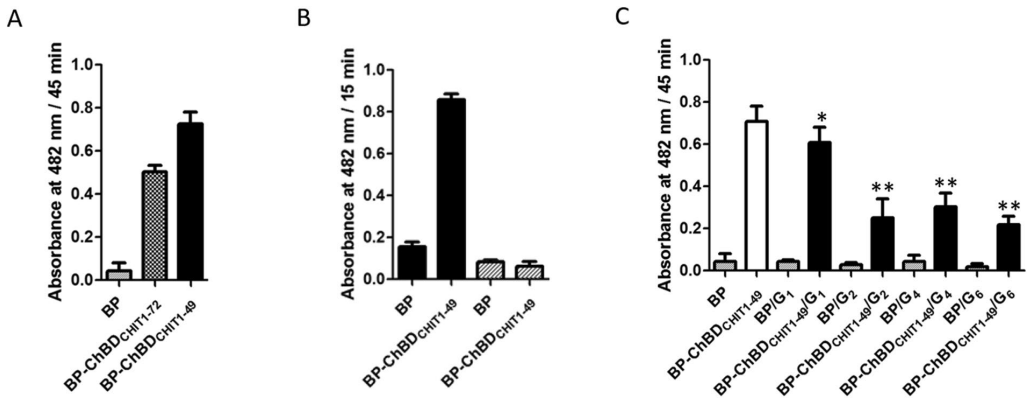

Figure 1. Binding assays of BP-$ChBD_{CHIT1-49}$ towards different ligands. (**A**) Insoluble chitin was first used as a ligand to compare the binding efficiency of the protein used in the present study (BP-$ChBD_{CHIT1-49}$) and the one reported in our previous work[21] (BP-$ChBD_{CHIT1-72}$) using the reporter enzymatic activity of the carrier protein BlaP. The β-lactamase BlaP without any inserted $ChBD_{CHIT1}$ (labelled BP) was used as a negative control (**B**) Hyaluronan (HA; 50% acetylated; black) and peptidoglycan (PG; 50% acetylated; hatched) were also tested using the same procedure. (**C**) $GlcNAc_1$, $GlcNAc_2$, $GlcNAc_4$ and $GlcNAc_6$ (respectively G_1, G_2, G_4 and G_6) were used as competitors. An equimolar amount of competitor was pre-incubated with the protein before incubating the mixture with insoluble chitin. Except for $GlcNAc_1$ (*), all competitors showed a similar inhibition effect (**).

and was classified in the CAZy (Carbohydrate-Active enZYmes) database (http://www.cazy.org) in family 14 (CBM14). This family includes small binding domains like lectin-like proteins and is characterised by the presence of the highly conserved "hevein-fold". Among the three human chitinases, HCHT has drawn most of the attention and is nowadays known to be involved in innate immunity for several reasons. Firstly, this enzyme is mostly synthesized by human macrophages that play a critical role in innate immunity[12]. Secondly, HCHT is overexpressed in several pro-inflammatory diseases and in various human sub-populations more exposed to infectious organisms[13,14]. Finally, the expression of HCHT's gene can be modulated by the action of cytokines and different immune inducers[15]. However, the precise role of HCHT in innate immunity and its associated molecular mechanisms remain unclear. Indeed, most of the work reported on this enzyme focused on the association of its expression with various diseases. Hence, HCHT is used as a biomarker for the diagnostic of Gaucher disease, sarcoidosis, glucose intolerance, β-thalassemia and World Trade Center lung injury[16–18].

In bacteria, insoluble chitin hydrolysis requires the synergic activity of several chitinases. For example, Serratia marcescens expresses a battery of chitinases with complementary activities including exochitinases (ChiA and ChiB), endochitinases (ChiC), a N-acetyl-hexosaminidase (chitobiase) and a lytic polysaccharide monooxygenase (CBP21) that act together to hydrolyse efficiently chitin containing structures in order to generate an energy source[19,20]. In human, in the context of innate immunity, hydrolysis of chitin-containing pathogens would also require complementary chitinase activities. Nevertheless, HCHT appears to be the only chitinase secreted by macrophages. It seems unlikely that HCHT alone could efficiently degrade chitin-containing microorganisms. This intriguing observation explains why the exact function of this enzyme in the context of innate immunity remains unclear.

The aim of this study was to analyse the chitin-binding properties of HCHT chitin binding domain ($ChBD_{CHIT1}$). Our data provide the molecular basis for chitin recognition by $ChBD_{CHIT1}$ and, given the high conservation of the residues involved in chitin binding amongst CBMs 14 as well as other CBM families, we postulated that this binding mode is a hallmark of CBM-carbohydrate interactions.

Moreover, we have interrogated the biological function of HCHT, and more specifically we attempted to understand how such an atypical association between a lectin-like CBM14 ($ChBD_{CHIT1}$) and a glycoside hydrolase (GH18) has been conserved through evolution to generate HCHT homologues involved in defense mechanisms. Our data suggest that $ChBD_{CHIT1}$ has evolved by recruiting a glycoside hydrolase domain, initially used for metabolic purposes in other organisms, to become an important component of innate immunity in humans.

Results

Interaction between $ChBD_{CHIT1-49}$ and carbohydrates. In order to easily monitor the chitin binding activity of $ChBD_{CHIT1}$, we inserted the 49 C-terminal residues of HCHT into the BlaP β-lactamase and the chitin binding activity of the resulting hybrid protein was monitored using the β-lactamase enzymatic activity as a reporter as previously described[21]. Our data indicated that $ChBD_{CHIT1-49}$ displayed chitin-binding activity on chitin-coated magnetic beads (Fig. 1A). Since N-acetyl glucosamine (GlcNAc) is the monomeric subunit of chitin, we have investigated the interaction of $ChBD_{CHIT1-49}$ with different soluble chito-oligosaccharide derivatives ($GlcNAc_1$, $GlcNAc_2$, $GlcNAc_4$ and $GlcNAc_6$). In practice, the protein was pre-incubated with a given

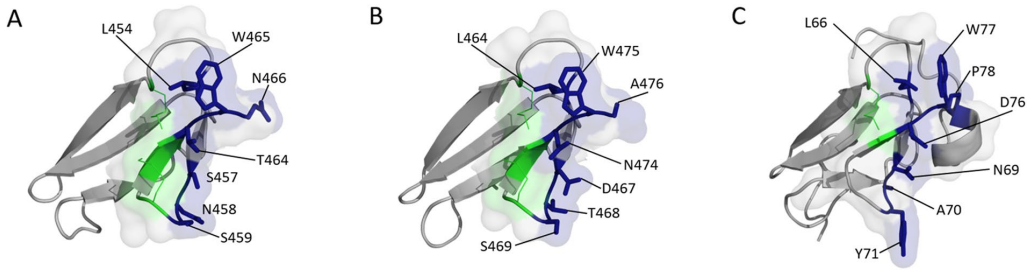

Figure 2. Structures of $ChBD_{CHIT1-49}$, AMCase (3D model) and tachycitin (*"hevein-fold"* motifs are shown in transparent surface representation). (**A**) Crystal structure of $ChBD_{CHIT1-49}$ (PDB ID: 5HBF). The aromatic and polar residues are represented in blue sticks (conserved disulphide bonds are represented as green lines). (**B**) 3D model of AMCase CBM14 (YASARA) built by sequence homology using the $ChBD_{CHIT1-49}$ X-ray 3D structure as template. The corresponding residues shown in Fig. 2A are highlighted in blue sticks. (**C**) NMR structure of tachycitin (PDB ID: 1DQC) where corresponding residues shown in Fig. 2A are highlighted in blue sticks.

chito-oligosaccharide (acting as a competitor) and then the mixture was incubated with insoluble chitin before measuring the β-lactamase activity immobilized on insoluble chitin. The data showed a low competition effect of $GlcNAc_1$, whereas all the other chito-oligomers showed a significant chitin binding inhibition (Fig. 1C). Notably, the inhibition effect of $GlcNAc_2$, $GlcNAc_4$ and $GlcNAc_6$ were in the same order of magnitude. These data suggest that the interaction surface of $ChBD_{CHIT1-49}$ is relatively small and that the residues involved in binding are localised to a limited area.

Other polysaccharides were also tested and $ChBD_{CHIT1-49}$ was found to bind hyaluronan (50% acetylated). Interestingly, no interaction was detected for Escherichia coli peptidoglycan (Fig. 1B), which also includes GlcNAc units.

Molecular basis for chitin recognition. The analysis of the recently solved structure of $ChBD_{CHIT1-49}$ (PDB ID: 5HBF; Fig. 2A) has revealed the presence of a specific structural motif that is also found in tachycitin (PDB ID: 1DQC; Fig. 2C) and hevein (PDB ID: 1T0W)[1,22,23]. This motif known as the "hevein-fold" is well-conserved within CBM14 family and has been shown to be involved in chitin binding[1]. In $ChBD_{CHIT1-49}$, this "hevein" motif is stabilized by two disulphide bonds (Cys450-Cys463, Cys460-Cys462).

Alanine scanning was used to investigate the "hevein" motif. Each of the 12 residues contained within the "hevein" motif of $ChBD_{CHIT1-49}$ was substituted by alanine using directed mutagenesis. In order to preserve the structural integrity of the protein, cysteines and Phe456 were not mutated. All the mutants were expressed and purified as hybrid β-lactamases in order to use the enzymatic activity of the β-lactamase moiety to monitor the chitin-binding activity of $ChBD_{CHIT1-49}$ mutants. The integrity of the fold as well as the functionality of the corresponding mutated domains were probed using Far-UV CD spectra and enzymatic activity assays, respectively (Figure S1, Table S1). The CD measurements were performed on the isolated $ChBD_{CHIT1-49}$ mutants whereas the β-lactamase enzymatic activity assays were recorded on the corresponding hybrid proteins. In general, the chitin binding affinities of the mutants displayed reduced binding affinity. Mutants were classified into three different groups according to the impact of the mutation on the binding affinity: (i) low impact, Thr452, Gly453, Val455, Ser459 and Asn466; (ii) medium impact, Ser457, Asn458, Lys461 and Thr464; (iii) and high impact, Pro451, Leu454 and Trp465 (Fig. 3A,B,C). The deletion of Trp465 has previously been described to have a deleterious effect on binding[24], which is in good agreement with our data, since substitution of this residue showed the strongest impact on $ChBD_{CHIT1-49}$ binding activity (Fig. 3A). The π-electrons on the aromatic residue most likely interact with the C-H bond in the pyranose ring of the GlcNAc unit. Besides Trp465, other apolar residues (Pro451, Gly453, Leu454 and Val455), mostly located in the same loop, appear to be important for binding (Fig. 3A). On the $ChBD_{CHIT1-49}$ structure, these residues form a hydrophobic pocket that stabilizes the loop conformation and consequently the orientation of Trp465 side chain (Fig. 3E). Furthermore, substitution of several polar residues (Ser457, Asn458, Ser459, Thr464 and Asn466) has an impact on chitin binding. These residues likely contribute to binding by providing hydrogen bonds with the ligand (Fig. 3A,D,F). Indeed, polar residues are commonly found in the protein-sugar interfaces[25,26].

Interaction of $ChBD_{CHIT1-49}$ with chito-oligosaccharides. ^{15}N-HSQC spectra of $ChBD_{CHIT1-49}$ were recorded before and after addition of $GlcNAc_3$. The percentages of backbone and side chain assignment covered 92% and 82% of the protein domain, respectively. The side-chain chemical shifts of Asn466 (Hδ1/Nδ1 and Hδ2/Nδ2) and Trp465 (Hε1/Nε1) were the most perturbed in the ^{15}N-HSQC spectra upon titration of the ligands. These side-chains, corresponding to the substrate interaction surface, are shown in Fig. 4B. As no other peaks were significantly affected, chemical shift data for these peaks were used to calculate a K_d for $GlcNAc_3$ of 9.9 ± 0.8 (SD) mM (Figs 4A and S3B). The measured changes in chemical shifts induced by $GlcNAc_2$ were overall too low to calculate a K_d (Figure S3C), however this observation in itself shows that $ChBD_{CHIT1-49}$ binds $GlcNAc_3$ stronger than $GlcNAc_2$.

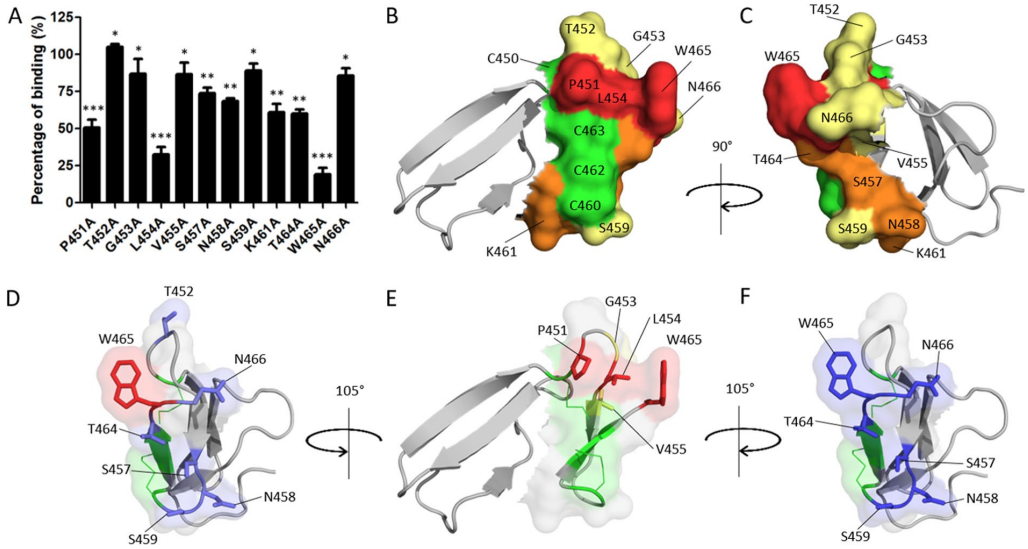

Figure 3. Molecular mechanism of chitin binding and structural features of $ChBD_{CHIT1\text{-}49}$. (**A**) The diagram shows the chitin binding efficiencies of single mutants generated by directed mutagenesis of predicted binding residues. All the protein mutants (except T452A) displayed reduced binding activities and were classified in three different groups depending on the impact of the mutation: (i) low impact *, (ii) medium impact ** and (iii) high impact ***. (**B**) X-ray 3D structure of $ChBD_{CHIT1\text{-}49}$ solved by Fadel and coworkers (PDB ID: 5HBF) where the *"hevein-fold"* motif is shown in surface representation whereas the rest of the structure is represented in cartoon (disulphide bonds are colored in green). Mutated residues were colored in yellow, orange and red depending respectively on the low, medium and high impact of the mutation on the chitin binding function. (**C**) 90° rotated view of the structure. (**D**) Detailed view of polar residues (in blue sticks) and Trp465 (in red) in the *"hevein-fold"* motif of $ChBD_{CHIT1\text{-}49}$ (disulphide bonds in green). (**E**) Representation of key residues (P451, G453, L454 and V455) involved in the hydrophobic pocket which stabilises Trp465 orientation (disulphide bonds in green). (**F**) Surface representation of the binding surface of $ChBD_{CHIT1\text{-}49}$ highlighting the key aromatic residue Trp465 and the key polar residues S457, N458, S459, T464 and N458 (blue sticks) directly involved in chitin binding (disulphide bonds in green).

Phylogenetic Study. $ChBD_{CHIT1\text{-}49}$ is part of the CBM14 family. In HCHT, this CBM is associated by a linker region to a catalytic domain belonging to the GH18 family. Although phylogenetic studies have previously been reported on the GH18 family[27] and CBMs14[28], we have analyzed the taxonomic distribution of CBMs 14, GHs 18 and HCHT in order to trace back the possible origin and evolution of these domains/proteins (Fig. 5A). Hence, we have noticed that GHs 18 proteins are present in genomes of all three domains of Life (Archaea, Bacteria, Eukaryota) and of some viruses. In contrast, CBM14 domains are restricted to specific groups of Eukaryota, suggesting that the appearance of this domain family is more recent. While both domain families (GH18 and CBM14) have coexisted since early eukaryotes [as deduced from their occurrence in both unikont and bikont lineages][29], HCHT-like proteins apparently only assembled in the ancestor of bilaterian animals (i.e., after the divergence of sponges, ctenophores and cnidarians; Fig. 5A). Next we analysed the origin of "hevein-fold" containing proteins, including tachycitin and hevein itself. Tachycitin is also a CBM14 that shares 34% amino acid sequence identity and structure conservation with $ChBD_{CHIT1\text{-}49}$. In contrast, hevein, the first protein in which the "hevein-fold" has been described and characterized, is related to the CBM18 family and is strictly present in Viridiplantae, Fungi and viruses genomes (Fig. 5A).

Therefore, our data support the fact that the structure similarity between the "hevein-folds" of CBMs 14 and CBMs 18 appears to be the result of convergent evolution, which is commonly admitted by the scientific community[23]. Finally, we built domain trees for CBM14 and GH18 families. The CBM14 tree features a subtree that includes all CBMs 14 involved in HCHT-like architectures (Fig. 5B), whereas, in the GH18 tree, most (but not all) of the GHs 18 involved in HCHT-like architectures are located in the same subtree (Fig. 5C). Considering the weak resolution of these two domain trees, it is very possible that the additional HCHT-like GH18 domains that are not part of the main HCHT-like subtree are simply misplaced by phylogenetic inference. This would mean that HCHT-like proteins are all homologous because they have a single origin, tracing back to the original association of a CBM14 domain and of a GH18 domain in a common ancestor of Bilaterians.

Discussion

Our work focuses on the CBM of the HCHT and has highlighted several features of this domain that might be applied to many other CBMs. First, we have identified the chitin binding residues of $ChBD_{CHIT1\text{-}49}$. We have also

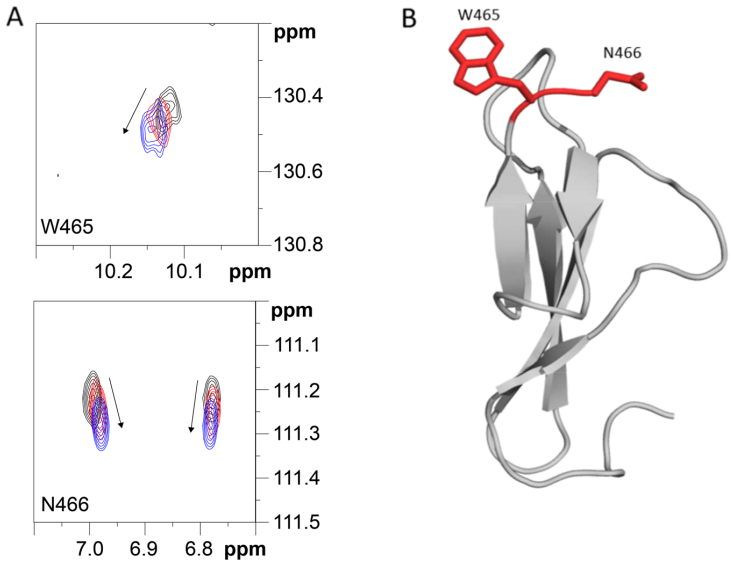

Figure 4. $ChBD_{CHIT1\text{-}49}$ interaction with $GlcNAc_3$ and binding surface. (**A**) Overlay of an area of interest from the ^{15}N-HSQC spectrum for 0.20 mM $ChBD_{CHIT1\text{-}49}$ in 50 mM phosphate buffer pH 7.0 recorded at 25 °C (black) in the presence of 4.5 mM (red) and 25 mM $GlcNAc_3$ (blue). The arrows indicate direction of the change in chemical shift upon titration. (**B**) The binding surface mapped on the structure of $ChBD_{CHIT1\text{-}49}$, where the side chain W465 and N466 are showed in red.

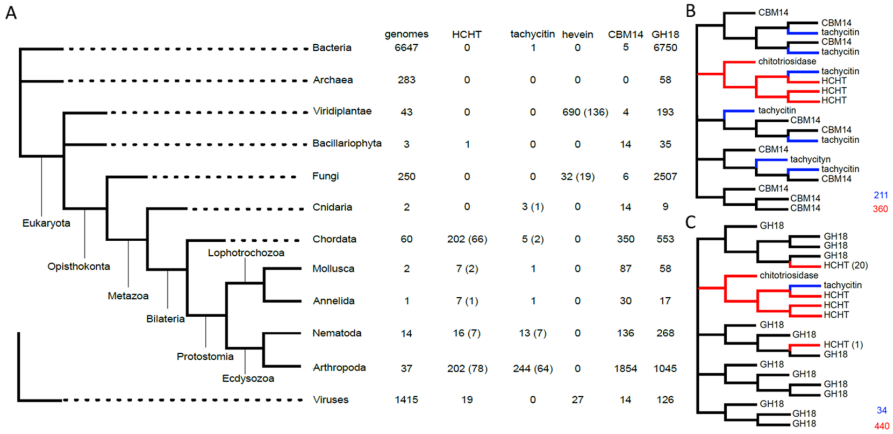

Figure 5. Phylogenetic analyses. (**A**) Reference phylogenetic tree on which are mapped the results of genome mining for chitotriosidase (HCHT-like), tachycitin and hevein proteins, as well as GH18 and CBM14 domains. The number of representative species are written in brackets. (**B**) Schematic tree of CBM14 domains, in which families integrated in HCHT-like architectures are colored in red (families tagged tachycitin are colored in blue). (**C**) Schematic tree of GH18 domains, in which families integrated in HCHT-like architectures are colored in red (families tagged tachycitin shown in blue).

characterized and quantified the interaction between $ChBD_{CHIT1\text{-}49}$ and small chito-oligosaccharides using NMR spectroscopy and competition binding assays. Our data indicate that although $ChBD_{CHIT1\text{-}49}$ interacts tightly to insoluble chitin, the measured affinities for chito-oligosaccharides are weak with dissociation constant values in the mM range. A phylogenetic study was also conducted to analyze the evolutionary paths of both domains included in HCHT (CBM14 and GH18) in order to determine when these domains were combined to give the GH18-CBM14 architecture found in HCHT and how this association is related to the biological role of this chitinase in innate immunity.

A hydrophobic pocket present on $CHBD_{CHIT1-49}$ "hevein" motif important for chitin binding. Our alanine scanning experiment has highlighted the presence of a hydrophobic pocket that includes Leu454 and Val455, both located in a loop induced by Pro451. We hypothesized that this hydrophobic pocket orients the Trp465 indole ring (Fig. 3E) for chitin binding. Molecular dynamic simulations (YASARA) performed on the Leu454Ala mutant; which presented the most impaired chitin binding activity, support this hypothesis. Indeed, the Trp465 indole side chain of the generated mutant exhibits higher flexibility and display more conformational freedom (Figure S2). These observations suggest that Leu454 plays an important role in orientating Trp465 side chain. Also, Pro451, Gly453, Leu454 and Trp465 are relatively well conserved through CBMs 14, which suggests that this binding mechanism can be generalizable to other CBMs 14 (Fig. 2). Our alanine scanning experiment has also highlighted that mutation of five polar residues (Ser457, Asn458, Ser459, Thr464 and Asn466) decreased the chitin binding efficiency. These residues are all located below the Trp465 side chain and well positioned to form hydrogen bonds (Fig. 3D,F). We postulated a chitin binding mechanism for $ChBD_{CHIT1-49}$ where Trp465 oriented by its hydrophobic pocket recognizes and interacts with a first GlcNAc unit of the polysaccharide. Besides this main interaction, polar residues located below Trp465 seem perfectly positioned to interact with a second and probably a third GlcNAc unit. This molecular mechanism reflects very well the binding mechanism of Type C CBMs, where essentially polar residues located in loops and few aromatic residues are directly involved in chitin recognition[6].

The chitin-binding surface of $ChBD_{CHIT1-49}$ can efficiently fit a minimum of two GlcNAc units. Our competition binding assays presented in Fig. 1C has highlighted important information regarding the chitin binding surface of $ChBD_{CHIT1-49}$. These data showed that the smallest chito-oligosaccharide able to compete with insoluble chitin is $GlcNAc_2$. Importantly, $GlcNAc_1$ showed a much lower inhibition effect compared to $GlcNAc_2$. This suggests that at least two GlcNAc units are required to form a stable complex. Moreover, NMR titration showed that a third GlcNAc unit could extend the surface area of the interaction (Figs 4A and S3). Notably, these data contrast with our previous study that didn't show any chitin-binding inhibition of $ChBD_{CHIT1}$ by the chito-oligomers[30]. In the present study, we used another type of chitin (chitin-coated magnetic beads) rather than α-crystalline chitin in our previous work. In addition, the amount of chitin that we used herein was much lower compared to our previous study, which placed us in better condition to see a chitin binding inhibition by chito-oligosccharides given their low binding affinity (K_{ds} in the mM range) for $ChBD_{CHIT1}$.

It is also interesting to comment on the difference in the binding affinities of $ChBD_{CHIT1-49}$ towards soluble chito-oligosaccharides and crystalline chitin. Indeed, although it is difficult to quantify crystalline chitin interaction given the insoluble character of this ligand, we estimated that this interaction is tight given the harsh and denaturing conditions required to remove $ChBD_{CHIT1-49}$ from a crystalline chitin support. In contrast, the measured K_{ds} for chito-oligosaccharides are in the high mM range, which is weak (Figs 4A and S3B). In the context of chitin hydrolysis by HCHT, the weak affinity for small chito-oligosaccharides, which are the hydrolysis products of the enzymatic reaction, is an advantageous feature that avoids $ChBD_{CHIT1-49}$'s binding inhibition, which maintains HCHT efficient towards crystalline chitin.

New insights into the modularity of HCHT using a phylogenetic study approach. HCHT is a modular enzyme composed of a GH18 (catalytic domain) and a CBM14 and is known to be involved in innate immunity by indirect evidences (see Introduction). In this work, we have performed a phylogenetic study (Fig. 5) on the catalytic and chitin binding domains of HCHT. The results showed that GH18-containing genes are found in genomes of all domains of life and that this ancestral family of catalytic domains is present in numerous modular enzymes involved in various biological processes[20,31–33]. It is therefore not surprising that they are associated to a wide variety of domains as it is illustrated by the 377 different architectures classified in Pfam database (http://pfam.xfam.org). In contrast, CBMs 14 are only found in Eukaryota (Fig. 5A), which implies that this protein family is related to strictly eukaryotic biological functions like innate immunity. Interestingly, according to Pfam database CBMs 14 are mostly associated with domains involved in defense/immune functions.

Conservation of the HCHT-like architecture only occurred from Bilateria (Fig. 5A). This lineage includes animals with higher complexity and presumably, more developed immune systems. In these bilaterians, HCHT-like proteins were described as a component of defense and development[34]. All these observations led us to an unexpected conclusion: the leader domain that dictates the biological function and dominates the biological activity of HCHT might not be the catalytic domain but rather the CBM14. If this hypothesis is confirmed, it will be of crucial importance because it would allow us to better understand and predict the biological role of CBM-containing proteins.

Conclusion

Altogether, the data presented in this work bring new insights into the biological function of HCHT. Although there are several evidences that link HCHT to immunity, the precise role of this chitinase in human remains unclear. It is reasonable to postulate that the presence of $ChBD_{CHIT1-49}$ facilitates the recognition of small and hardly accessible motifs in different chitin types. These features probably enhance HCHT's capability to be efficient towards a huge diversity of chitin containing pathogens. Since chito-oligosaccharides were recently shown to display a higher immunogenicity compared to insoluble chitin[35], the lectin-like behaviour of $ChBD_{CHIT1-49}$ might confer to HCHT the ability to enhance the inflammatory response against a wide variety of chitin containing pathogens by releasing immunogenic chito-oligosaccharides and therefore acting as an immune catalyser that can lead to a recruitment of additional immune actors.

Methods

Construction of the Genes Encoding the Hybrid β-lactamases. The genes encoding the chitin-binding domains $ChBD_{CHIT1\text{-}72}$ (residues Pro395 to Asn466) and $ChBD_{CHIT1\text{-}49}$ (residues Thr418 to Asn466) of the HCHT (Uniprot number: Q13231) were amplified by PCR and inserted into the gene coding for the class A β-lactamase exo-small BlaP (BP)[36] previously cloned in the expression vector pET26b(+). The insertion site is located between residues Asp197 and Lys198 of the β-lactamase and the insertions of the gene fragments into the BlaP gene were performed as described in our previous studies[21,30,37,38]. It is important to note that, in our previous work, we used a slightly longer $ChBD_{CHIT1}$ that included the 72 C-terminal residues of HCHT (residues Pro395 to Asn466), however based on sequence alignments; we noticed that only the 49 C-terminal residues of $ChBD_{CHIT1}$ were conserved. In addition previous studies published by Tjoelker and coworkers[24] as well as Fadel and coworkers[22] confirmed that the minimum length chitin-binding domain is $ChBD_{CHIT1\text{-}49}$. This is why we shortened the domain down to 49 residues. Furthermore, the inserted $ChBD_{CHIT1\text{-}49}$ gene was surrounded by two thrombin cleavage sites to release the isolated $ChBD_{CHIT1}$ domain after production and purification when needed as reported previously[21]. The resulting genetic construct, called pET26b(+)-BP-$ChBD_{CHIT1\text{-}49}$, was used to express the hybrid β-lactamases harbouring both the pel B signal peptide for periplasmic secretion and a His6 tag sequence at the N-terminal and C-terminal extremities, respectively.

Alanine Scanning Mutagenesis. We selected a subset of 12 residues present on the predicted chitin-binding surface of $ChBD_{CHIT1\text{-}49}$ for substitution into alanine. The 12 single mutants of the protein BP-$ChBD_{CHIT1\text{-}49}$ were generated using the Quick-Change Multi Site-Directed Mutagenesis kit (Agilent Technologies) following the manufacturer's instructions. Briefly, mixtures of 50% Phusion PCR Master Mix (Thermo Fisher Scientific), 5% DMSO, 100 ng of template DNA and 2 mM of primers were submitted to the following PCR program: 30 s at 98 °C, 10 s at 98 °C, 30 s at 42 °C and 15 s at 72 °C (30 cycles), followed by a final step at 72 °C for 10 min. The PCR products were then digested by DpnI in FastDigest Buffer (Thermo Fisher Scientific) for 10 min at 37 °C to remove template DNA. Finally 10 μL of the resulting mutated plasmids were used to transform E. coli DH5α competent cells for plasmid amplification.

Hybrid Protein Expression and Purification. All hybrid proteins were expressed in E. coli BL21(DE3). Transformed cells were cultured at 37 °C in Terrific Broth medium supplemented with 50 μg/mL kanamycin until OD600nm reached 2.5. Cultures temperature were then decreased to 18 °C for 30 min before proceeding to the induction of protein expression by the addition of isopropyl β-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cultures were grown for 14 h at 18 °C. Cells were harvested by centrifugation (twice) and resuspended successively in two distinct periplasmic extraction buffers: firstly in 1/10th volume of 20 mM Tris 600 mM sucrose 5 mM EDTA pH 8.0 at 37 °C and secondly in 1/4th volume of 5 mM $MgSO_4$ at 4 °C. Each periplasmic extract was supplemented with one tablet of complete Protease Inhibitor Cocktail (Roche).

Periplasmic proteins were loaded on a 5 mL Bio-ScaleTM Mini ProfinityTM IMAC column (BIO-RAD) equilibrated in 300 mM KCl, 50 mM KH_2PO_4 and 5 mM imidazol pH 8.0. The column was successively washed with 300 mM KCl, 50 mM KH_2PO_4, 10 mM imidazol pH 8.0 and the elution was performed with 300 mM KCl, 50 mM KH_2PO_4 and 250 mM imidazol pH 8.0. Imidazol was immediately removed using a 50 ml Bio-ScaleTM Mini Bio-gel® P-6 Desalting column (BIO-RAD). Protein purity level and homogeneity were confirmed by SDS-PAGE and UV-Visible (125–400 nm) spectra. Protein concentrations were determined by UV absorbance at 280 nm.

Isolated native and mutated $ChBD_{CHIT1\text{-}49}$ were released from their carrier protein BlaP by thrombin cleavage as previously described[21]. In this study, isolated $ChBD_{CHIT1\text{-}49}$ domains were purified by molecular exclusion chromatography using XK 26/100 SuperDex 75 PrepGrade column (GE Healthcare) equilibrated in 150 mM NaCl 50 mM phosphate buffer pH 7.5 (PBS).

Isotopic Labelled Protein Expression and Production for NMR studies. ^{13}C, ^{15}N or ^{15}N $ChBD_{CHIT1\text{-}49}$ samples were expressed in *E. coli* BL21(DE3) cells. Pre-culture were grown in LB medium (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl) supplemented with 50 μg/mL kanamycin in a shaking incubator at 225 rpm, 30 °C overnight. A 2L LB main culture with 50 μg/mL kanamycin was inoculated with 1% of the overnight culture and grown in a shaking incubator at 225 rpm, 30 °C to OD600nm reaches ~0.8. Cultures were centrifuged at 4,500 g for 10 min and resuspended on ice in 500 mL M9 media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl) supplemented with 99% $(^{15}NH_4)2SO_4$, 98% $^{13}C_6$-D-glucose, 10 mL Bioexpress Cell Growth Media (Cambridge Isotope Laboratories, Tewksbury, MA, USA), 10 mL Gibco™ MEM Vitamin Solution (100x), 1 mL 1 M $MgSO_4$, 10 mL Trace Metal solution (0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, 1 g/L $CaCl_2$) and 50 μg/mL kanamycin. Expression was induced 15 min after the media change by IPTG to a final concentration of 0.1 mM, and then the culture was incubated at 16 °C, 225 rpm for 20 hours. The cells were harvested by centrifugation at 5,000 g, 10 min, 4 °C and suspended in 30 mL TES buffer pH 8.0 (3.63 g/L TRIS, 1.86 g/L EDTA, 200 g/L sucrose) together with half a tablet Complete Protease Inhibitor (Roche) followed by a centrifugation at 10 min, 4 °C, 6,150 g. The supernatant was removed and the cells incubated at room temperature for 10 min before being resuspended in 25 mL MQ and half a tablet Complete Protease Inhibitor (Roche). The suspension was supplemented with 125 μL 1 M $MgSO_4$ before the final centrifugation at 13,000 g, 45 minutes. The supernatant was filtered through a 0.22 μm Sterile-flip filter unit from Nalgene. Labelled $ChBD_{CHIT1\text{-}49}$ was purified as described above.

Enzymatic Characterization of the Hybrid Proteins. The kinetic parameters of the purified hybrid β-lactamases were determined by measuring the rates of nitrocefin (CalBiochem) hydrolysis at different substrate concentrations. Initial rates were measured so that less than 10% of substrate was hydrolysed. A protein concentration of 25 ng/mL was used in the presence of 0.1 mg/mL BSA (Fermentas) used as a crowding and stabilizing agent. The experiment was performed at 37 °C in PBS (pH 7.5). A spectrophotometer (PowerWave X, TempLab)

was used to monitor the formation of the hydrolysis product at 482 nm. Kinetic parameters (kcat, Km and kcat/Km) were determined for each hybrid protein as described by Matagne and coworkers[39]. Standard deviations were calculated on the basis of the results obtained from 3 technical replicates for each hybrid protein.

NMR Spectroscopy. NMR spectra of 0.2 mM $ChBD_{CHIT1\text{-}49}$ samples in 50 mM Phosphate buffer at pH 5.5 and 7.0 were recorded at 25 °C on a Bruker Ascend 800 MHz spectrometer Avance III HD equipped with a 5-mm Z-gradient CP-TCI (H/C/N) cryogenic probe at the NT-NMR-Center/Norwegian NMR Platform (NNP). NMR data were processed using Bruker TopSpin version 3.5. NMR spectral analysis was performed using CARA version 1.5.5[40]. A partial backbone assignment was accomplished using HNCA, CBCA(CO)NH, HN(CA)CO, HNCO, ^{15}N-HSQC-NOESY and ^{15}N-HSQC spectra.

NMR titration was used to probe the interaction of $ChBD_{CHIT1\text{-}49}$ with its ligands chitotriose ($GlcNAc_3$) and chitobiose ($GlcNAc_2$). For $GlcNAc_3$, the titration points were 1.0 mM, 2.4 mM, 4.5 mM, 6.3 mM, 11.6 mM, 17.1 mM and 25.0 mM. For GlcNAc2, the titration points were 0.5 mM, 1.0 mM, 2.4 mM, 4.5 mM, 8.7 mM, 12.4 mM, 18.7 mM and 23.9 mM. 1D-proton and ^{15}N-HSQC spectra (at 4096 × 1024 point resolution) were recorded for each titration point.

The side-chains of Trp465 and Asn466 were identified based on the partial assignment and their ^{15}N-HSQC peaks were used as reporters for the interaction by measuring chemical shift changes in the N and H^N atoms of the backbone of $ChBD_{CHIT1\text{-}49}$. A compound change in chemical shift, $\Delta\delta comp$ (in ppm) was calculated using the following formula: $\Delta\delta comp = [(\Delta\delta_H)^2 + (\Delta\delta_N/x)^2]^{1/2}$. $\Delta\delta H$ is the change in chemical shift of the amide proton (ppm), $\Delta\delta N$ is the change in chemical shift of the amide nitrogen (ppm), and x is a constant used to achieve equal contributions from changes in N and H^N shifts, which was set to 6.5[41].

Equation (1)[42] was used to estimate the dissociation constant (K_d) of the interaction by using Excel to simultaneously fit K_d and $Qmax$ ($\Delta\delta comp$ at saturation) for $\Delta\delta comp$ at each ligand concentration, [L] (mM), and the protein concentration, [P] (mM), remained constant at 0.2 mM (Fig. S3B).

$$\Delta\delta comp = Qmax\frac{[P] + [L] + K_d \pm \sqrt{([P] + [L] + K_d)^2 - 4\,[P]\,[L]}}{2\,[P]} \quad (1)$$

Alanine Scanning Mutagenesis of the chitin-binding surface of $ChBD_{CHIT1}$. Residues expected to interact with chitin were substituted with alanine. To identify critical amino acids, an excess of purified hybrid proteins expressing a mutated $ChBD_{CHIT1\text{-}49}$ domain were mixed with a final chitin magnetic beads concentration of 2% (v/v; New England BioLabs) and 0.1 mg/mL BSA. A control was conducted with the carrier protein BlaP without any inserted $ChBD_{CHIT1}$. Binding assays were performed at room temperature by orbital mixing of the protein-chitin beads suspension during 30 min. Bound proteins were harvested by magnetic attraction and washed three times with 500 mM NaCl 20 mM Tris-HCl 1 mM EDTA 0.1% Tween (pH 5.0). Immobilized protein levels on chitin beads were determined by incubation of the beads with nitrocefin and monitoring of antibiotic hydrolysis over time at 482 nm (RT).

Phylogenetic Analysis. Starting with the sequences of human chitotriosidase (2201442 A), Tachypleus tridentatus (Arthropoda) tachycitin (1DQC_A), and Hevea brasiliensis (Viridiplantae) hevein (AAO63573.1), and using an E-value threshold of $1e^{-5}$, three separate PHMMER searches were carried out for genome mining on the UniProtKB sequence database, through the HHMI Janelia web portal (http://hmmer.janelia.org/; now available at http://www.ebi.ac.uk/Tools/hmmer/).

For the HCHT search, only the hit sequences simultaneously featuring a CBM14 domain and a GH18 domain were downloaded (both in non-aligned and aligned format). For the two other proteins, all hit sequences were downloaded in non-aligned format. A combination of batch identifier mapping through the UniProt web portal (http://www.uniprot.org/uploadlists/) and custom Perl scripts (Bio-MUST-Core, D. Baurain, R. R. Léonard, unpublished) was then used to recover the complete taxonomic lineage of each sequence. The three non-aligned sequence files were aligned using MAFFT[43] and the resulting alignments cleared of partial sequences, defined as lacking more than 50% positions of the longest sequence in each alignment (Bio-MUST-Core). Final alignments were then submitted to phylogenetic inference using either RAxML[44] and the PROTGAMMALGF model[45,46] or PhyloBayes[47] and the CATGTRG model[48,49] to produce the trees. Both models yielded largely unresolved but broadly similar trees.

Based on the downloaded HMMER-aligned HCHT file, two additional alignments corresponding to each one of the two domains were generated (Bio-MUST-Core) and used to build two HMM profiles with hmmbuild[50]. These profiles were then pasted on the HHMI Janelia web portal to carry out two separate HMM searches on UniProtKB, using an E-value threshold of $1e^{-5}$. All hit sequences featuring at least one copy of the corresponding domain were downloaded in non-aligned format and further processed as above (MAFFT, Bio-MUST-Core, RAxML/PhyloBayes) to produce two phylogenetic trees of the CMB14 and GH18 domain families. To locate chitotriosidase, tachycitin and hevein sequences in the domains trees, a semi-automated annotation pipeline was developed so as to highlight the leaves corresponding to sequences recovered in the three initial PHMMER searches.

References

1. Aboitiz, N. *et al*. NMR and modeling studies of protein-carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *Chembiochem: a European journal of chemical biology* **5**, 1245–1255, doi:10.1002/cbic.200400025 (2004).
2. Gordon-Thomson, C. *et al*. Chitotriosidase and gene therapy for fungal infections. *Cellular and molecular life sciences: CMLS* **66**, 1116–1125, doi:10.1007/s00018-009-8765-7 (2009).
3. Younes, I. & Rinaudo, M. Chitin and chitosan preparation from marine sources. *Structure, properties and applications. Marine drugs* **13**, 1133–1174, doi:10.3390/md13031133 (2015).

4. Zelensky, A. N. & Gready, J. E. The C-type lectin-like domain superfamily. *The FEBS journal* **272**, 6179–6217, doi:10.1111/j.1742-4658.2005.05031.x (2005).
5. Shi, Z. *et al*. Identification of Novel Pathways in Plant Lectin-Induced Cancer Cell Apoptosis. *International journal of molecular sciences* **17**, 228, doi:10.3390/ijms17020228 (2016).
6. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical journal* **382**, 769–781, doi:10.1042/BJ20040892 (2004).
7. Berthelot, K., Peruch, F. & Lecomte, S. Highlights on Hevea brasiliensis (pro)hevein proteins. *Biochimie* **127**, 258–270, doi:10.1016/j.biochi.2016.06.006 (2016).
8. Drummond, R. A. & Lionakis, M. S. Mechanistic Insights into the Role of C-Type Lectin Receptor/CARD9 Signaling in Human Antifungal Immunity. *Frontiers in cellular and infection microbiology* **6**, 39, doi:10.3389/fcimb.2016.00039 (2016).
9. Boot, R. G. *et al*. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *The Journal of biological chemistry* **276**, 6770–6778, doi:10.1074/jbc.M009886200 (2001).
10. Boot, R. G. *et al*. The human chitotriosidase gene. Nature of inherited enzyme deficiency. *The Journal of biological chemistry* **273**, 25680–25685, doi:10.1074/jbc.273.40.25680 (1998).
11. Hussain, M. & Wilson, J. B. New paralogues and revised time line in the expansion of the vertebrate GH18 family. *Journal of molecular evolution* **76**, 240–260, doi:10.1007/s00239-013-9553-4 (2013).
12. Ley, K., Pramod, A. B., Croft, M., Ravichandran, K. S. & Ting, J. P. How Mouse Macrophages Sense What Is Going On. *Frontiers in immunology* **7**, 204, doi:10.3389/fimmu.2016.00204 (2016).
13. Pacheco, N. & Uribe, A. Enzymatic analysis of biomarkers for the monitoring of Gaucher patients in Colombia. *Gene* **521**, 129–135, doi:10.1016/j.gene.2013.03.044 (2013).
14. Barone, R., Simpore, J., Malaguarnera, L., Pignatelli, S. & Musumeci, S. Plasma chitotriosidase activity in acute Plasmodium falciparum malaria. *Journal of tropical pediatrics* **49**, 63–64, doi:10.1093/tropej/49.1.63 (2003).
15. Malaguarnera, L. Chitotriosidase: the yin and yang. *Cellular and molecular life sciences: CMLS* **63**, 3018–3029, doi:10.1007/s00018-006-6269-2 (2006).
16. Bargagli, E. *et al*. Human chitotriosidase: a sensitive biomarker of sarcoidosis. *Journal of clinical immunology* **33**, 264–270, doi:10.1007/s10875-012-9754-4 (2013).
17. Cho, S. J. *et al*. Chitotriosidase is a biomarker for the resistance to World Trade Center lung injury in New York City firefighters. *Journal of clinical immunology* **33**, 1134–1142, doi:10.1007/s10875-013-9913-2 (2013).
18. Musumeci, M. *et al*. Serum YKL-40 levels and chitotriosidase activity in patients with beta-thalassemia major. *Disease markers* **2014**, 965971–6, doi:10.1155/2014/965971 (2014).
19. Horn, S. J. *et al*. Comparative studies of chitinases A, B and C from Serratia marcescens. *Biocatalysis and Biotransformation* **24**, 39–53, doi:10.1080/10242420500518482 (2006).
20. Vaaje-Kolstad, G., Horn, S. J., Sorlie, M. & Eijsink, V. G. The chitinolytic machinery of Serratia marcescens–a model system for enzymatic degradation of recalcitrant polysaccharides. *The FEBS journal* **280**, 3028–3049, doi:10.1111/febs.12181 (2013).
21. Vandevenne, M. *et al*. Rapid and easy development of versatile tools to study protein/ligand interactions. *Protein engineering, design & selection: PEDS* **21**, 443–451, doi:10.1093/protein/gzn021 (2008).
22. Fadel, F. *et al*. X-Ray Crystal Structure of the Full Length Human Chitotriosidase (CHIT1) Reveals Features of Its Chitin Binding Domain. *PloS one* **11**, e0154190, doi:10.1371/journal.pone.0154190 (2016).
23. Suetake, T. *et al*. Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *The Journal of biological chemistry* **275**, 17929–17932, doi:10.1074/jbc.C000184200 (2000).
24. Tjoelker, L. W. *et al*. Structural and functional definition of the human chitinase chitin-binding domain. *The Journal of biological chemistry* **275**, 514–520, doi:10.1074/jbc.275.1.514 (2000).
25. Hamre, A. G., Jana, S., Reppert, N. K., Payne, C. M. & Sorlie, M. Processivity, Substrate Positioning, and Binding: The Role of Polar Residues in a Family 18 Glycoside Hydrolase. *Biochemistry* **54**, 7292–7306, doi:10.1021/acs.biochem.5b00830 (2015).
26. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640–647, doi:10.1038/nature04162 (2005).
27. Funkhouser, J. D. & Aronson, N. N. Jr. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC evolutionary biology* **7**, 96, doi:10.1186/1471-2148-7-96 (2007).
28. Chang, T. C. & Stergiopoulos, I. Inter- and intra-domain horizontal gene transfer, gain-loss asymmetry and positive selection mark the evolutionary history of the CBM14 family. *The FEBS journal* **282**, 2014–2028, doi:10.1111/febs.13256 (2015).
29. Adl, S. M. *et al*. The revised classification of eukaryotes. *The Journal of eukaryotic microbiology* **59**, 429–493, doi:10.1111/j.1550-7408.2012.00644.x (2012).
30. Vandevenne, M. *et al*. The Bacillus licheniformis BlaP beta-lactamase as a model protein scaffold to study the insertion of protein fragments. *Protein science: a publication of the Protein Society* **16**, 2260–2271, doi:10.1110/ps.072912407 (2007).
31. Tetreau, G. *et al*. Overview of chitin metabolism enzymes in Manduca sexta: Identification, domain organization, phylogenetic analysis and gene expression. *Insect biochemistry and molecular biology* **62**, 114–126, doi:10.1016/j.ibmb.2015.01.006 (2015).
32. Tzelepis, G., Dubey, M., Jensen, D. F. & Karlsson, M. Identifying glycoside hydrolase family 18 genes in the mycoparasitic fungal species Clonostachys rosea. *Microbiology* **161**, 1407–1419, doi:10.1099/mic.0.000096 (2015).
33. Nagpure, A., Choudhary, B. & Gupta, R. K. Chitinases: in agriculture and human healthcare. *Critical reviews in biotechnology* **34**, 215–232, doi:10.3109/07388551.2013.790874 (2014).
34. Badariotti, F., Thuau, R., Lelong, C., Dubos, M. P. & Favrel, P. Characterization of an atypical family 18 chitinase from the oyster Crassostrea gigas: evidence for a role in early development and immunity. *Developmental and comparative immunology* **31**, 559–570, doi:10.1016/j.dci.2006.09.002 (2007).
35. Gorzelanny, C., Poppelmann, B., Pappelbaum, K., Moerschbacher, B. M. & Schneider, S. W. Human macrophage activation triggered by chitotriosidase-mediated chitin and chitosan degradation. *Biomaterials* **31**, 8556–8563, doi:10.1016/j.biomaterials.2010.07.100 (2010).
36. Frate, M. C. *et al*. Export and folding of signal-sequenceless Bacillus licheniformis beta-lactamase in Escherichia coli. *European journal of biochemistry/FEBS* **267**, 3836–3847, doi:10.1046/j.1432-1327.2000.01422.x (2000).
37. Vandevenne, M. *et al*. Comparative functional analysis of the human macrophage chitotriosidase. *Protein science: a publication of the Protein Society* **20**, 1451–1463, doi:10.1002/pro.676 (2011).
38. Crasson, O. *et al*. Enzymatic functionalization of a nanobody using protein insertion technology. *Protein engineering, design & selection: PEDS* **28**, 451–460, doi:10.1093/protein/gzv020 (2015).
39. Matagne, A. *et al*. The diversity of the catalytic properties of class A beta-lactamases. *The Biochemical journal* **265**, 131–146, doi:10.1042/bj2650131 (1990).
40. Keller, R. The Computer Aided Resonance Assignment Tutorial. (*CANTINA Ver- lag, Goldau, Switzerland*) (2004).
41. Mulder, F. A., Schipper, D., Bott, R. & Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *Journal of molecular biology* **292**, 111–123, doi:10.1006/jmbi.1999.3034 (1999).
42. Fielding, L. Determination of association constants (K(a)) from solution NMR data. *Tetrahedron* **56**, 6151–6170, doi:10.1016/S0040-4020(00)00492-0 (2000).
43. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772–780, doi:10.1093/molbev/mst010 (2013).

44. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi:10.1093/bioinformatics/btu033 (2014).
45. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Molecular biology and evolution* **25**, 1307–1320, doi:10.1093/molbev/msn067 (2008).
46. Yang, Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular biology and evolution* **10**, 1396–1401 (1993).
47. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288, doi:10.1093/bioinformatics/btp368 (2009).
48. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular biology and evolution* **21**, 1095–1109, doi:10.1093/molbev/msh112 (2004).
49. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC evolutionary biology* **7**(Suppl 1), S4, doi:10.1186/1471-2148-7-S1-S4 (2007).
50. Durbin, R., Eddy, S. R., Krogh, A. & Mitchison, G. In *Biological sequence analysis* (Cambridge University Press, 1998).

Acknowledgements

Most of this work was supported by funds from the National Fund for Scientific Research (F.R.S.-FNRS) and grants from the Walloon Region in Belgium, including a grant from the F.R.S.-FNRS "Crédit de recherche 2014" (CDR J.0080.15) for computational resources. The NMR experiments were financed by SO-funds from the Norwegian University of Science and Technology (NTNU) and by the Norwegian NMR Platform from the Research Council of Norway (grant numbers 226244). Gerd Inger Sætrom is thanked for excellent technical assistance.

Author Contributions

F.L.A., D.B., M.G., M.S. and M.V. designed and supervised the research. O.C., G.C., R.R.L., F.L. and R.P. performed research, and analyzed data. O.C., G.C., F.L.A., D.B. and M.V. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02382-z

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Human Chitotriosidase: Catalytic Domain or Carbohydrate Binding Module, Who's Leading HCHT's Biological Function.

Oscar Crasson[1], Gaston Courtade[2], Raphaël R. Léonard[3], Finn Lillelund Aachmann[2], François Legrand[1], Raffaella Parente[1], Denis Baurain[3], Moreno Galleni[1,] Morten Sørlie[4], and Marylène Vandevenne[1*]

Supplementary Information

FIGURE S1

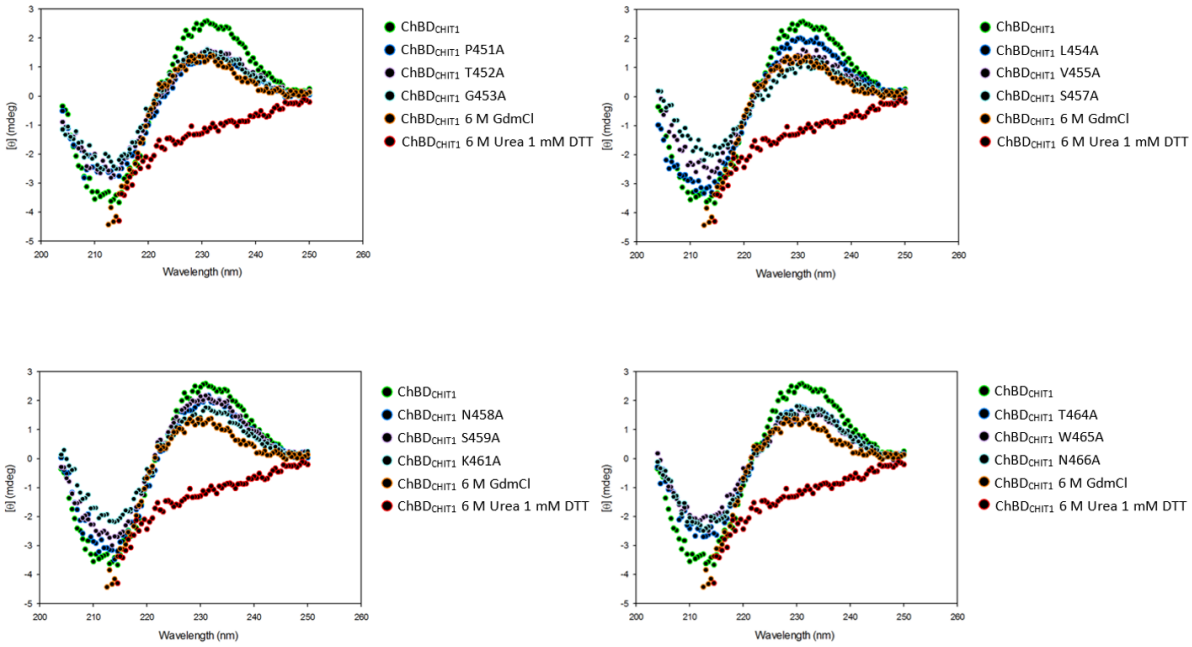

Figure S1. Far-UV CD spectra of ChBDCHIT1-49 and its single mutants. The non mutated $ChBD_{CHIT1-49}$ is represented in green. Unfolded $ChBD_{CHIT1-49}$ samples obtained by incubation with 6 M guanidinium chloride (GdmCl) and 6 M Urea (plus 1 mM DTT) are represented in orange and red dots, respectively. In each spectrum, single mutants are shown in blue, purple and cyan dots. All proteins displayed a β-sheet profile demonstrating that all proteins were well folded and characterized by the expected secondary structure content. Noticeably, the "hevein-fold" motif is a very stable structure that required the addition of reducing agent to abolish the secondary structure content of $ChBD_{CHIT1-49}$ underlining the importance of the disulphide bonds.

FIGURE S2

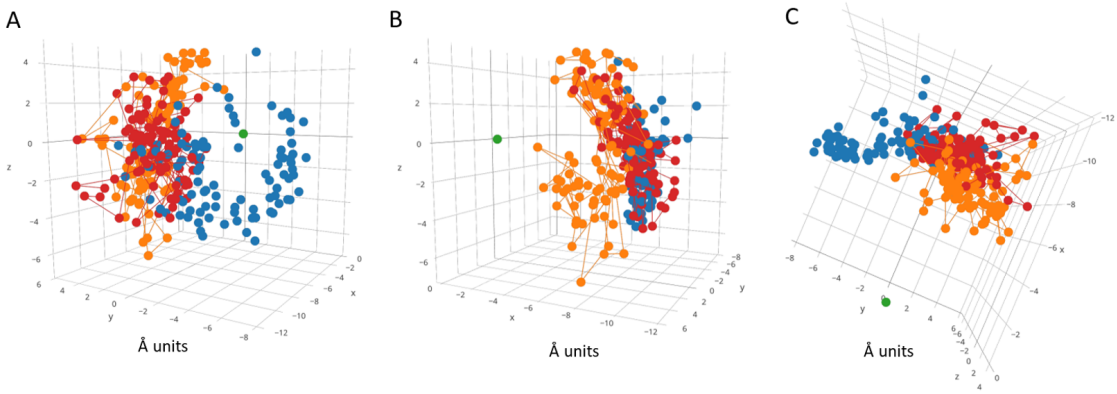

Figure S2. Molecular dynamic simulation (YASARA) of WT $ChBD_{CHIT1-49}$, Leu454Ala and Leu454Val mutants. The center of mass (COM) of the conserved disulphide bond (Cys450-Cys463; green dot) and the COM of Trp465 have been analysed during a molecular dynamic simulation of 25 ns (100 frames) for the WT $ChBD_{CHIT1-49}$ (in blue) and the two mutants Leu454Ala (in orange) and Leu454Val (in dark red). In the WT protein, Trp465 COM displayed a well-defined plane with relatively low flexibility due to the steric hindrances and hydrophobic environment of Leu454. For the two mutants, the Trp465 COM is delocalized in the right inner part of the "hevein-fold" and exhibits more flexibility compared to the WT protein, which results in a loss of binding efficiency. (A) Front view. (B) Right view. (C) Top view.

FIGURE S3

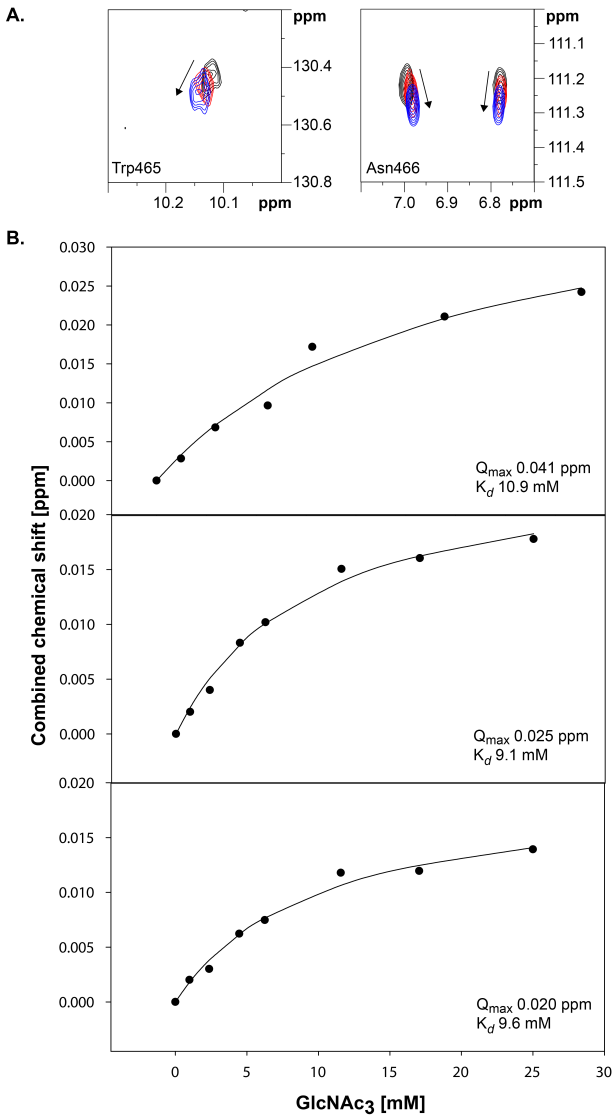

Figure S3. $ChBD_{CHIT1\text{-}49}$ interaction with GlcNAc oligomers. (A) Overlay of an area of interest from the ^{15}N-HSQC spectrum for 0.20 mM $ChBD_{CHIT1\text{-}49}$ in 50 mM phosphate buffer pH 7.0 recorded at 25 °C (black) in the presence of 4.5 mM (red) and 23.9 mM $GlcNAc_2$ (blue). The arrows indicate direction of the change in chemical shift upon titration. (B) Combined chemical shift for three atom pairs [$H^{\varepsilon 1}/N^{\varepsilon 1}$ of Trp465 (top panel) and $H^{\delta 1}/N^{\delta 1}$; $H^{\delta 2}/N^{\delta 2}$ of Asn466 (two lower panels)] upon titration with $GlcNAc_3$. Dissociation constant (K_d) was estimated for individual atom pairs by simultaneously fit K_d and Qmax (Δδcomp at full saturation, and it is assumed equivalent to the bound fraction of the substrate). The average Kd was calculated to 9.9±0.8 (SD) mM for $GlcNAc_3$.

Table S1. Kinetic parameters of BlaP and its chimeric derivatives for nitrocefin

Table S1. *Kinetic parameters of BlaP and its chimeric derivatives for nitrocefin*

Proteins	K_m (μM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($\mu M^{-1}sec^{-1}$)
BP	69 ± 10	432 ± 26	6,3 ± 1
BP-$ChBD_{CHIT1-72}$	98 ± 17	528 ± 49	5,4 ± 1,1
BP-$ChBD_{CHIT1-49}$	91 ± 6	343 ± 13	3,8 ± 0,3
P451A	92 ± 14	494 ± 39	5,4 ± 0,9
T452A	109 ± 16	359 ± 32	3,3 ± 0,6
G453A	93 ± 15	335 ± 28	3,6 ± 0,7
L454A	94 ± 17	293 ± 28	3,1 ± 0,6
V455A	89 ± 16	352 ± 32	4 ± 0,8
S457A	86 ± 19	305 ± 34	3,5 ± 0,9
N458A	120 ± 30	285 ± 39	2,4 ± 0,7
S459A	92 ± 17	386 ± 37	4,2 ± 0,9
K461A	73 ± 12	392 ± 32	5,4 ± 1
T464A	128 ± 33	368 ± 54	2,9 ± 0,9
W465A	78 ± 10	259 ± 16	3,3 ± 0,5
N466A	99 ± 16	348 ± 29	3,5 ± 0,6

Additional contribution II

Protein Science
DOI 10.1002/pro.3451

Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation

Dejan M. Petrović[1], Bastien Bissaro[1], Piotr Chylenski[1], Morten Skaugen[1], Morten Sørlie[1], Marianne S. Jensen[1], Finn L. Aachmann[2], Gaston Courtade[2], Anikó Várnai[1], Vincent G.H. Eijsink[1*]

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

[2]Department of Biotechnology and Food Science, NOBIPOL, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

*Corresponding author:
Vincent G.H. Eijsink,
Faculty of Chemistry, Biotechnology, and Food Science, The Norwegian University of Life Sciences (NMBU), 1432 Ås, Norway. Tel.: +47-67232463; E-mail: vincent.eijsink@nmbu.no

Running title: Role of histidine methylation in fungal LPMOs

Total number of manuscript pages: 42

Total number of figures: 7

Supplementary material:

The supplementary material includes eleven figures. Figure S1: SDS-PAGE of purified *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp. Figure S2: 3D structure of *Ta*LPMO9A (PDB ID: 3ZUD). Figure S3: LC-MS chromatograms of N-terminal peptides. Figure S4: Binding to PASC. Figure S5: Reaction products generated from steam exploded birch. Figure S6: HPAEC-PAD chromatograms for the reactions with cellopentaose. Figure S7: Close-up of MALDI-ToF MS spectra for the reactions with PASC or xyloglucan. Figure S8: NMR pH titration curves. Figure S9: Kinetic information for copper binding obtained by stopped-flow spectrophotometry. Figure S10: Differential scanning calorimetry analysis. Figure S11: Control reactions to check for occurrence of copper-catalyzed Fenton-type chemistry.

Filename:

Petrovic_Role_of_histidine_methylation_in_fungal_LPMOs_Supplementary_material

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as
doi: 10.1002/pro.3451

Received: May 09, 2018; Revised: Jun 01, 2018; Accepted: Jun 04, 2018

Abstract

The catalytically crucial N-terminal histidine (His1) of fungal lytic polysaccharide monooxygenases (LPMOs) is post-translationally modified to carry a methylation. The functional role of this methylation remains unknown. We have carried out an in-depth functional comparison of two variants of a family AA9 LPMO from *Thermoascus aurantiacus* (*Ta*LPMO9A), one with, and one without the methylation on His1. Various activity assays showed that the two enzyme variants are identical in terms of substrate preferences, cleavage specificities and the ability to activate molecular oxygen. During the course of this work, new functional features of *Ta*LPMO9A were discovered, in particular the ability to cleave xyloglucan, and these features were identical for both variants. Using a variety of techniques, we further found that methylation has minimal effects on the pK_a of His1, the affinity for copper and the redox potential of bound copper. The two LPMOs did, however, show clear differences in their resistance against oxidative damage. Studies with added hydrogen peroxide confirmed recent claims that low concentrations of H_2O_2 boost LPMO activity, whereas excess H_2O_2 leads to LPMO inactivation. The methylated variant of *Ta*LPMO9A, produced in *Aspergillus oryzae*, was more resistant to excess H_2O_2 and showed better process performance when using conditions that promote generation of reactive-oxygen species. LPMOs need to protect themselves from reactive oxygen species generated in their active sites and this study shows that methylation of the fully conserved N-terminal histidine provides such protection.

Keywords: lytic polysaccharide monooxygenase, histidine, methylation, hydrogen peroxide, *Thermoascus aurantiacus*

Abbreviations: LPMO, lytic polysaccharide monooxygenase; AA, Auxiliary Activity; PASC, phosphoric acid swollen cellulose;

Lytic polysaccharide monooxygenases (LPMOs) are abundant mono-copper enzymes that are important in industrial biomass processing and the global carbon cycle. We have studied the role of a relatively rare post-translational modification found in fungal LPMOs, namely methylation of the copper coordinating N-terminal histidine. We show that methylation has surprisingly little effect on the properties of His1 and the LPMO as a whole, with one important exception: methylation reduces oxidative self-inactivation, thus improving overall enzyme performance.

Introduction

Copper-dependent lytic polysaccharide monooxygenases (LPMOs) oxidatively cleave polysaccharides[1-5]. They are classified as Auxiliary Activities in the Carbohydrate-Active enzyme (CAZy; http://www.cazy.org[6]) database and belong to the families AA9, AA10, AA11, AA13, AA14 and AA15[6]. The discovery of LPMOs in 2010[1] has been followed by extensive research related to their structure, function and diversity[4,5]. Since LPMO action renders the crystalline structure of recalcitrant substrates such as cellulose and chitin more susceptible to the action of classical glycoside hydrolases (GHs), LPMOs have become an important ingredient of commercial enzyme cocktails for industrial biomass conversion[7,8].

As of today, around 50 LPMOs have been characterized (according to the CAZy database), of which most are fungal LPMOs belonging to the AA9 family[9]. Characterized fungal LPMOs act on various polysaccharides, including cellulose[10,11], xyloglucan and other (1,4)-linked β-glucans[12,13], starch[14,15], and xylan grafted onto cellulose[16]. Regarding their oxidative regioselectivity, cellulose-active LPMOs may act on the C1 or the C4 in the scissile glycosidic bond, or on both[2,17,18]. Structural analyses have shown that the copper ion in the active site is coordinated by a highly conserved "histidine brace" formed by three nitrogen ligands provided by the N-terminal amino group and the side chains of the N-terminal histidine (His1) and the side chain of an additional histidine[2] (Fig. 1).

Fungal LPMOs have been successfully produced in several fungal species such as *Aspergillus oryzae*[2,19], *Neurospora crassa*[17] and *Myceliophthora thermophila*[20], and in the yeasts *Pichia pastoris* (recently renamed to *Komagataella phaffii*) and *Saccharomyces cerevisiae*[18,21-24]. In nature, fungal LPMOs undergo several post-translational modifications (PTMs) including signal peptide cleavage, disulfide bond formation and τ-methylation of the N-terminal histidine[2,14,17]

(Fig. 1), and may also be subject to *O*- and *N*-linked glycosylations[5,25,26]. Accurate cleavage of the signal peptide is important for activity because the N-terminal amino group is part of the catalytic center. Remarkably, the role of the τ-methylation (i.e. methylation of Nε2) of the catalytically crucial His1 (Fig. 1) remains unclear, although some suggestions have been made[4,5,27,28]. It is noteworthy that the N-terminal histidines of bacterial LPMOs are not methylated[4,5,29].

Fungal LPMOs expressed in filamentous fungi or yeasts will display differences in their PTMs. When expressed in fungi, methylation of the His1 occurs[2,14,17,30], whereas yeasts lack the methyltransferase responsible for this PTM, meaning that His1 remains non-methylated. Besides this, different expression hosts may yield proteins with different glycosylation patterns, which could affect enzyme properties, although expression in especially *P. pastoris* is generally considered to result in functional recombinant proteins[25,31]. Expression in *P. pastoris* will lead to LPMOs lacking their natural PTMs, most notably methylation of the N-terminus, but LPMOs produced in this manner tend to be active, as illustrated by their ability to boost the overall cellulolytic activity of Celluclast, a commercial LPMO-poor cellulase cocktail[7,32]. Thus, the functional implications of the methylation remain an enigma.

In this study, we have expressed a cellulose-active LPMO from *Thermoascus aurantiacus* (*Ta*LPMO9A, previously known as *Ta*GH61A) in *P. pastoris* (*Ta*LPMO9A-Pp), and compared its properties with the properties of the same LPMO expressed in the fungus *A. oryzae* (*Ta*LPMO9A-Ao), a host known to carry out PTMs such as methylation[2]. The investigated enzyme properties included activity, substrate specificity, substrate binding, the ability to generate H_2O_2, product profiles, the pK_a of His1, redox potential, copper binding, thermal and operational stability, and the boosting effect on the efficiency of Celluclast, an LPMO-poor

cellulase cocktail. These analyses showed that the only major difference between the two enzyme variants concerns their operational stability, in particular their resistance against oxidative damage, which seems to be a major determinant of LPMO performance during biomass conversion.

Results

Heterologous expression of TaLPMO9A in Pichia pastoris and Aspergillus oryzae

*Ta*LPMO9A expressed in *P. pastoris* (*Ta*LPMO9A-Pp) or in *A. oryzae* (*Ta*LPMO9A-Ao) was purified to homogeneity and the mass of both enzyme variants was between 25 and 30 kDa, based on SDS-PAGE (Fig. S1). These similar masses were slightly higher than the theoretical mass (24.4 kDa) of the protein calculated from its amino acid sequence. This minor discrepancy is likely due to *N*- and/or *O*-glycosylations. The NetNGlyc and NetOGlyc servers (http://www.cbs.dtu.dk/services) predicted one putative *N*-glycosylation site, Asn138, and five putative *O*-glycosylation sites, Ser35, Thr37, Thr39, Thr47 and Thr144. None of these are located on the catalytic surface of the protein (Fig. S2). The glycosylation sites nearest to the histidine brace, Thr39 and Thr47, are located 13.7 and 15.3 Å away from the catalytic copper, respectively.

First, we analyzed the N-terminus of the *Ta*LPMO9A preparations to confirm correct processing of the signal peptide and the presence of methylation on His1. The two *Ta*LPMO9A variants were digested with trypsin, which, in case of correct processing, would yield a non-methylated N-terminal peptide (HGFVQNIVIDGK) or a methylated N-terminal peptide (Met-HGFVQNIVIDGK) with corresponding theoretical masses of 1326.52 or 1340.55 Da, respectively. LC-MS analysis confirmed that the signal peptide was correctly cleaved right

before the His1 in 97 % or >99 % of the cases for *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp, respectively. In addition, methylation of His1 was absent in the *Ta*LPMO9A-Pp, while in the *Ta*LPMO9A-Ao sample, the majority of the protein carried the methylation (Fig. S3).

Substrate binding and specificity

*Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp showed weak and similar binding to PASC and Avicel. Upon mixing with the substrate, the concentration of free enzyme in solution was lowered by 16 % in the presence of PASC (Fig. S4), or ≤5 % in case of Avicel (data not shown). The observation of weak binding is congruent with previously reported binding data for LPMOs lacking a CBM domain[33,34].

Both *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp showed activity towards PASC, tamarind xyloglucan (TXG) and steam-exploded birch (SEB) in the presence of ascorbic acid (Figs. 2 & S5), while both enzymes were inactive towards cellopentaose (Fig. S6), ivory nut mannan, konjac glucomannan, and xylan from birch wood (data not shown).

In the presence of ascorbic acid, *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp released a mixture of C1- and C4-oxidized oligosaccharides (Fig. 2A). Although product quantification is not straightforward, both the HPAEC-PAD chromatograms and MALDI-ToF MS spectra (Figs. 2C & S7) indicated that C4 oxidation dominates. The MALDI-TOF MS data showed signal compatible with the formation of double-oxidized oligosaccharides (Fig. S7).

HPAEC-PAD analysis of products generated from TXG (Fig. 2B), revealed a broad range of products eluting between 22 and 62 min and the product profiles were similar to those previously observed for the xyloglucan-active C1/C4-oxidizing LPMOs *Fg*LPMO9A and *Gt*LPMO9A-2[23,35]. MALDI-ToF MS spectra (Figs. 2D & S7) revealed products with varying combinations of pentoses (Pen) and hexoses (Hex). The appearance of oxidized products with a number of

pentoses not equaling a multiple of 3, suggests that *Ta*LPMO9A is able to oxidize TXG at any position in the backbone (the repeating unit in TXG is the Hex_4Pen_3 heptasaccharide XXXG). Interestingly, a closer look at the MALDI-ToF MS data for the Hex_7Pen_4 cluster (Fig. S7) showed a signal that could reflect a double oxidized species, which would mean that the mixed C1/C4 activity also applies to TXG.

*Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp also liberated oxidized cello-oligosaccharides from SEB (Fig. S5) in the absence of ascorbic acid. SEB contains 36 % of lignin, which has been shown earlier to drive the LPMO reaction[2,7,19,36]. Due to the high background, product identification in the reactions with SEB is not easy, but it is clear form Fig. S5, that the two *Ta*LPMO9A variants give identical results.

Importantly, while revealing novel aspects of the functionality of *Ta*LPMO9A, the activity and binding assays described above did not reveal any functional difference between the methylated and the non-methylated variant of the enzyme.

Variation of the electron donor

Next, we compared product formation by *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp in reactions with PASC using either ascorbic acid or cellobiose dehydrogenase (*Mt*CDH) as electron donor. In the reactions with the enzymatic electron supply, the amount of products increased over 24 h for both LPMOs. After 24 h, product formation was the same for the two versions of *Ta*LPMO9A (114±9 and 126±6 nC×min, for *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp, respectively; Fig. 3A). On the other hand, with the non-enzymatic electron donor, product formation rapidly reached a plateau, after 2 hours for *Ta*LPMO9A-Pp and after 4 hours for *Ta*LPMO9A-Ao (Fig. 3B). Consequently, the methylated *Ta*LPMO9A-Ao version gave higher final product yields than the non-methylated *Ta*LPMO9A-Pp (192 ± 2 and 153 ± 10 nC×min, respectively). These results

indicate that the two enzymes may differ in terms of operational stability, as discussed further, below.

H_2O_2 production

It has been shown that, in the presence of a suitable reductant and in the absence of substrate, LPMOs accumulate H_2O_2[21]. We tested the ability of both *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp to produce H_2O_2 in the absence of substrate with ascorbic acid as reductant. The measured H_2O_2 production rates were the same for both enzymes, 0.93 ± 0.04 min^{-1} and 1.00 ± 0.09 min^{-1}, for *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp, respectively. It is worth noting that, these H_2O_2 production rates are similar to previously reported LPMO catalytic rates[1,12,37]. The similarity of the H_2O_2 production rates of *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp suggests that the methylation on His1 does not affect the LPMO's ability to activate molecular oxygen.

The pK_a value of His1

The pK_a for methylated His1 in apo-*Ta*LPMO9A-Ao was determined to be 7.21 ± 0.16. For non-methylated *apo-Ta*LPMO9A-Pp, the pK_a values of the two active site histidines were determined to be 7.36 ± 0.16 and 7.67 ± 0.16 His1 and His86, respectively (Fig. S8). The pK_a value for His1 in *apo-Ta*LPMO9A-Pp is thus close to the pK_a of 7.21 ± 0.16 for methylated His1 in *apo-Ta*LPMO9A-Ao.

Zn^{2+} is a non-paramagnetic analogue of Cu^{2+} that does not cause paramagnetic relaxation enhancement and can thus be used to study the effect of metal binding on the catalytic histidines. pH titrations of both *Ta*LPMO9A forms were carried out in the presence of Zn^{2+} to investigate the effect of Zn^{2+} on the pK_a of His1. For the Zn^{2+}-loaded forms of the enzymes, the pK_a values for His1 were determined to be 7.43 ± 0.16 and 7.58 ± 0.16 for *Ta*LPMO9A-Pp and

*Ta*LPMO9A-Ao, respectively. (Nb. Since His86 was not observable in the Zn^{2+}-loaded form of *Ta*LPMO9A-Pp, its pKa could not be determined). The pH titration curves are shown in Fig. S8.

Cu^{2+} binding, redox potential and thermal stability

Intense efforts were made to probe binding of Cu(II) to the apo forms of the LPMOs by isothermal titration calorimetry, using a method that has previously been used successfully for various LPMOs[27,38], but it was not possible to obtain solid data. At enzyme concentrations of 10 micromolar and higher and temperatures of 10 °C and higher, the stoichiometry of Cu(II) vs. LPMO was higher than 1. At lower LPMO concentrations (e.g. 5 micromolar) and temperature (e.g. 5 °C), the heats of binding were too small for the determination of any thermodynamic parameters. Therefore, to get an impression of copper binding, we adopted a stopped-flow method described previously by Chaplin et al.[39] to obtain kinetic information for copper binding.

The apo forms of *Ta*LPMO9A-Ao or *Ta*LPMO9A-Pp were mixed with Cu^{2+} solutions of known concentrations and the quenching of Trp fluorescence was monitored as a function of time in a stopped-flow spectrophotometer. For both enzymes, the pseudo first-order rate constants (k_{obs}) were linearly dependent on the Cu^{2+} concentrations (Fig. S9). The slopes of the k_{obs} versus [Cu^{2+}] plots yielded second-order rate constants of 2.19 ± 0.06 · 10^6 $M^{-1}\cdot s^{-1}$ and 1.77 ± 0.04 · 10^6 $M^{-1}\cdot s^{-1}$, for *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp, respectively. Both linear plots cross the axes close to the origin, as previously observed in similar analysis of a family AA10 LPMO from *Streptomyces lividans*, *Sli*LPMO10E, suggesting that association rates for Cu^{2+} are high compared to the dissociation rates[39].

Redox potentials, assessed through equilibrium reactions with N,N,N′,N′-tetramethyl-1,4-phenylenediamine (TMP), were determined to be 195 ± 7 mV for *Ta*LPMO9A-Ao and 186 ± 5mV for *Ta*LPMO9A-Pp.

The thermal stability of the LPMOs was investigated using DSC analysis. The DSC thermograms showed that *Ta*LPMO9A-Ao and the majority *Ta*LPMO9A-Pp are thermostable and unfold irreversibly at an apparent meting temperature ($T_{m,app}$) of ~77.0 °C (Fig. S10). The DSC studies showed that the *Ta*LPMO9A-Pp preparation contained a subfraction with a lower $T_{m,app}$ of ~67.0 °C. It is worth noting that all enzyme assays in this study were done at the much lower temperature of 45 °C.

Boosting effect on cellulase cocktails

Earlier, it has been shown that replacing 15 % of a blend of Celluclast and Novozym 188 (C/N188), an LPMO-poor cellulase preparation, with *Ta*LPMO9A increases saccharification of steam exploded birch[7]. Here, we tested the same blend with either of the *Ta*LPMO9A preparations for the saccharification of sulfite-pulped spruce. As references, we used a mixture of 85 % C/N188 blend and 15 % BSA as well as Cellic® CTec2, a commercial LPMO-containing cellulase cocktail[7,40].

As expected on the basis of previous results[32], Fig. 4 shows that addition of the LPMO had an effect on saccharification efficiency and that this effect was much larger (25 % increase in glucose yield at 96 h) in the presence of ascorbic acid. Importantly, as opposed to reactions carried out in the absence of ascorbic acid (Fig. 4A), differences between the two enzyme preparations became apparent in the reactions with ascorbic acid (Fig. 4B). During the first 24 hours, *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp boosted the C/N188 cellulase blend to the same extent as shown by similar saccharification yields (Fig. 4B). During the next 72 h of incubation, however, saccharification was clearly more efficient in the reactions containing *Ta*LPMO9A-Ao, compared to reactions containing *Ta*LPMO9A-Pp. Indeed, after 96 h, the C/N188 blend with *Ta*LPMO9A-Ao reached 50 % saccharification, which is not significantly different from the

level that was reached with Cellic CTec2 (Fig. 4B), but higher than the saccharification yield reached with *Ta*LPMO9A-Pp. While in the case of *Ta*LPMO9A-Pp, the LPMO effect stopped after 24 h (meaning that after 24 h, the difference between black and blue bars in Fig. 4B is constant), the LPMO effect seemed to continue over the complete incubation period for *Ta*LPMO9A-Ao (shown as an increasing difference between the red and blue bars between 24 h and 96 h). Again, this suggests that the two LPMO variants differ in terms of operational stability.

The effect of H_2O_2

It has been shown that LPMOs not only can produce H_2O_2[21], but that they may also use H_2O_2 as co-substrate during polysaccharide cleavage[41-43]. It has further been shown that a surplus of H_2O_2 leads to oxidative inactivation of LPMOs and that this is due to the action of the LPMO itself (i.e. redox chemistry happening when H_2O_2 enters a reduced active site in the absence of substrate). Finally, it has been shown that LPMOs differ in terms of their susceptibility to H_2O_2 and that binding to substrate has a protective effect[41].

We tested the effect of different H_2O_2 concentrations on the activity of *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp in reactions with PASC and ascorbic acid (Fig. 5). Comparison of apparent initial rates, based on the amounts of oxidized oligosaccharides released after 3 min of incubation, revealed that the two enzyme variants responded differently to varying levels of H_2O_2 (Fig. 6). Consistently with previous observations, a boosting effect of H_2O_2 was observed (Figs. 5 & 6). However, there is a trade-off between the boosting effect of H_2O_2 and H_2O_2–mediated enzyme inactivation, which becomes noticeable for the higher H_2O_2 concentrations and/or after longer incubation times (Figs. 5-7). When it comes to this trade-off, the two enzyme variants clearly differ. For example, whilst at 200 µM H_2O_2 *Ta*LPMO9A-Ao showed a linear

progress curve up to 60 min (Fig. 5C), the progress curve for *Ta*LPMO9A-Pp was not even linear at 100 µM H_2O_2 (Fig. 5A). Consequently, at 200 µM H_2O_2 and after 60 minutes *Ta*LPMO9A-Ao yields much more products than *Ta*LPMO9A-Pp (Fig. 7B). At higher H_2O_2 concentrations, inactivation becomes more prominent for both enzymes, meaning that the yields after 60 minutes are reduced. In some cases inactivation is so fast that the apparent initial rates are reduced and in some cases (for *Ta*LPMO9A-Pp) these (apparent) rates are even lower than those measured in the absence of H_2O_2 (Figs. 5 & 6).

It is noteworthy that *Ta*LPMO9A-Pp was faster than *Ta*LPMO9A-Ao in the initial phase of the reaction (Fig. 7A). This difference was, however, only observed in the very beginning of the reaction and after 30 min the difference was no longer detectable (Fig. 5). Control reactions showed that oxidation products released from PASC did not originate from Fenton-type chemistry as shown by the absence of detectable products when the LPMOs were replaced by $CuSO_4$ (Fig. S11) in reactions that contained ascorbic acid and varying amounts of H_2O_2.

Discussion

Several of the fungal LPMOs characterized so far have been heterologously expressed in fungi or *P. pastoris*, and harbor different post-translational modifications (PTMs) depending on the host organism. While minor variations in glycosylation are generally not considered to be of major functional importance, some LPMOs carry another PTM, namely methylation of the catalytically crucial N-terminal histidine[2]. Comparison of the structures of methylated and non-methylated AA9 LPMOs has not provided obvious clues as to the role of the methylation. We have therefore carried out a functional comparison of *Ta*LPMO9A produced in *P. pastoris* and *Ta*LPMO9A produced in *A. oryzae*. Analysis of the N-terminal peptide confirmed that post-

translational methylation of the N-terminal histidine occurred only in the protein expressed in *A. oryzae*.

*Ta*LPMO9A has a limited number of putative glycosylation sites, which reduces the chance of significant differences in glycosylation between *Ta*LPMO9A-Pp and *Ta*LPMO9A-Ao. Indeed, the two enzyme variants showed similar mobilities in SDS-PAGE analysis. Considering the latter observation and the fact that none of the putative glycosylation sites are located near the catalytic center, it is unlikely that a glycosylation, or its absence, would explain functional differences between *Ta*LPMO9A-Pp and *Ta*LPMO9A-Ao.

The functional characterization of the two *Ta*LPMO9A variants provided novel information on this LPMO. As expected on the basis of previously published data[2,7,44], *Ta*LPMO9A showed a mixed C1/C4-oxidizing activity on PASC, with the C4-oxidizing activity being dominating. Importantly, we show that *Ta*LPMO9A is also able to cleave xyloglucan and that cleavage likely can happen independent of the backbone substitution pattern, similar to *Fg*LPMO9A[23] and *Gt*LPMO9A-2[35]. Interestingly, the MALDI-TOF MS data shown in Fig. S7B may indicate production of a double-oxidized xyloglucan fragment, which has also been observed for *Fg*LPMO9A[23] and which would unambiguously proof that LPMOs can cleave xyloglucan at both C1 and C4 positions. Further analytical work is needed to finally proof the formation of such fragments. Importantly, the studies with various substrates clearly showed that *Ta*LPMO9A-Pp and *Ta*LPMO9A-Ao have identical substrate specificities and yield identical products under the tested conditions.

Reactions with various electron-donating systems (ascorbic acid, CDH or natural compounds in steam-exploded birch) did not reveal differences in terms of the LPMO's ability to recruit electrons from varying sources. While progress curves with CDH showed that product formation

increased during most of the monitored incubation period, the reactions with ascorbic acid, expectedly[45,46], terminated before the monitoring period was over (Fig. 3). Termination of the reaction could be due to depletion of ascorbic acid, which in the presence of transition metals and O_2 will engage in a variety of redox side reactions, and/or to damage done by the reactive oxygen species generated in such side reactions, including H_2O_2 produced by non-substrate bound LPMOs[21,41]. The experiments with ascorbic acid revealed a difference between the two enzyme variants: the activity of *Ta*LPMO9A-Pp terminated more rapidly than the activity of *Ta*LPMO9A-Ao, suggesting that the latter is operationally more stable.

Previous studies have shown that τ-methylation increases the pK_a of histidine in solution by 0.46[47]. Our NMR-based determinations of the pK_a values of His1 in *Ta*LPMO9A showed that methylation hardly affects the pK_a value, regardless of whether a metal ion (Zn^{2+}) is bound. The minimal effect of Zn^{2+} on the pK_a contrasts to earlier observations for *Sm*LPMO10A, showing strong binding of Zn^{2+} leading to an expected reduction of the pK_a of the catalytic histidines[27]. The present observations suggest that LPMOs belonging to the AA9 family (such as *Ta*LPMO9A) interact with Zn^{2+} more weakly than LPMOs belonging to the AA10 family (such as *Sm*LPMO10A), as has indeed been observed by isothermal titration calorimetry[2].

Our measurements of the kinetics of copper binding indicate strong binding and showed that copper binding may be slightly more effective in the methylated enzyme. It is conceivable that methylation locks the histidine side chain in a single tautomeric conformation that is beneficial for copper binding, analogous to what has been suggested for copper binding by amyloid-β peptides[48].

The observed apparent melting temperatures for both enzymes were similar, although the DSC experiments did show differences between the two enzyme variants. While proteins forms

showed a main unfolding peak at 77 °C, the unfolding curve for *Ta*LPMO9A-Pp showed a minor peak at 67 °C. The extra peak observed *Ta*LPMO9A-Pp indicates a change in unfolding mechanism or a heterogeneity in the *Ta*LPMO9A-Pp preparation. A change in unfolding mechanism and sample heterogeneity could both be caused by variation in glycosylation. Notably, such variation cannot have been large because the two enzyme forms have very similar masses (Fig. S1). Importantly, the apparent melting temperatures (67 – 77 °C) are far above the temperature that was used in all performed experiments (45 °C). Therefore, we do not expect experimental artefacts due to differences in the thermal stability of the two enzyme variants.

Looking further into features that affect and/or reflect catalytic efficiency, we found that the two enzyme variants have almost identical redox potentials as well as almost identical abilities to generate H_2O_2. These experimental observations are in accordance with a theoretical study of LPMOs by Kim et al[49], who concluded that the methylation of His1 has only minor effects on the structure and functionality of the catalytic center.

The experiments discussed above all lead to the conclusion that the two LPMO forms are functionally very similar, in accordance with conclusions based on theoretical studies[49]. However, in more application oriented experiments, the two *Ta*LPMO9A preparations showed clear differences. Although both the methylated and non-methylated versions boosted cellulose saccharification by an LPMO-poor cellulase cocktail, the methylated *Ta*LPMO9A-Ao outperformed the non-methylated version, *Ta*LPMO9A-Pp in longer reactions. This difference is likely caused by inactivation of the *Ta*LPMO9A-Pp alone and not by inactivation of other enzymes in the reaction, since after 24 hours, when *Ta*LPMO9A-Pp no longer had a boosting effect, cellulose hydrolysis continued (Fig. 4B, black and blue bars).

Recently, it has been shown that small amounts of H_2O_2 boost LPMO activity, while higher amounts lead to inactivation of the enzyme[41-43,50]. The presence of ascorbic acid in the reactions described above will lead to generation of H_2O_2[51,52], which perhaps could explain the observed apparent enzyme inactivation. It was also shown[41] that inactivation of LPMOs by excess H_2O_2 is accompanied by oxidative damage in the catalytic center only, with the N-terminal histidine being especially vulnerable. Thus, it may seem that H_2O_2–driven inactivation is an autocatalytic process.

The studies with H_2O_2 showed a clear boosting effect of H_2O_2 on LPMO activity and H_2O_2–driven enzyme inactivation. Importantly, these studies revealed that *Ta*LPMO9A-Ao is more H_2O_2 resistant than *Ta*LPMO9A-Pp, suggesting that methylation of His1 protects the enzyme against autoxidation of the LPMO's catalytic center. Thus, methylation of His1 increases the operational stability of the LPMO. In this respect, it is interesting to note the relatively stable kinetics obtained with CDH (Fig. 3A), which is known to form H_2O_2 in a controlled manner at rates comparable to observed catalytic rates of LPMOs[21,45]. In line with the recent claims by Bissaro et al[41], it is tempting to speculate that in these reactions the produced H_2O_2 is consumed immediately by the LPMO, avoiding H_2O_2 accumulation and thus oxidative inactivation.

Oxidation of histidines tends to generate 2-oxo-histidine, which may undergo ring opening and further degradation[41,53,54]. It would be of interest to study the inactivation process of LPMOs in more detail, and to explore how τ-methylation, i.e. methylation of Nε2, reduces the vulnerability of the Cε for oxygenation. Of note, auto-oxidative inactivation is a well known phenomenon in the field of peroxidases[55] and is usually the result of a surplus of H_2O_2 that leads to either heme destruction or protein modification, via molecular mechanisms that remain to be fully resolved[55,56,57]. Interestingly, it has recently been shown that Nδ-methylation of the imidazole ring of a

proximal histidine in a heme-iron peroxidase led to an increase in total turnover numbers[58]. The authors of this study attributed this positive effect to the fixing of the His in a single tautomeric form, which could reduce chances of entering off-pathway processes that lead to enzyme inactivation. Non-methylated LPMOs show variation in their vulnerability for inactivation indicating that other structural factors than methylation co-determine operational stability. The effect of the methyl group in fungal LPMOs may thus be co-determined by interactions that involve specific structural features of these proteins. Another aspect that warrants further studies is that, despite being more prone to oxidative damage, *Ta*LPMO9A-Pp showed higher initial rates than methylated *Ta*LPMO9A-Ao (Figs. 5 & 7A). There is no straightforward explanation for this observation, which is of particular interest in light of the notion that the N-terminal histidines of bacterial LPMOs are not methylated.

In conclusion, we have compared the catalytic properties of *Ta*LPMO9A expressed in two commonly used expression hosts, *P. pastoris* and *A. oryzae*. The most important physical difference between the two enzyme versions concerns the methylation of their N-terminal histidine. We show that this methylation plays a role in protecting the LPMO from oxidative damage, which is most likely autocatalytic. It is interesting to note that our data suggest that the protection against auto-oxidation comes at a cost, since methylation seems to reduce the initial activity. Interestingly, bacterial LPMOs lack the N-terminal methylation and one may wonder whether this has to do with the degree of "oxidative pressure" that the various microbes meet, including, perhaps levels of available hydrogen peroxide. From an applied point of view, it is important to note that the methylation of His1, and thus the selection of the expression host for production of recombinant LPMOs, is important for the robustness of this redox enzyme.

Materials and Methods

Cloning, expression and purification of TaLPMO9A in Pichia pastoris

*Ta*LPMO9A from *Thermoascus aurantiacus* [UniProt: G3XAP7] was cloned, expressed in *P. pastoris*, and purified as described before[32]. The enzyme was saturated with Cu(II) by incubating with an excess of Cu(II)SO_4 (at ~3:1 molar ratio of copper:enzyme) for 90 min at room temperature as described previously[59]. Subsequently, the sample was subjected to size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare, Uppsala, Sweden) equilibrated with 50 mM Bis-Tris/HCl buffer (pH 6.5) containing 150 mM NaCl, using a flow rate of 0.75 $mL \cdot min^{-1}$. Fractions containing the pure protein were identified using SDS-PAGE and subsequently pooled, concentrated and buffer exchanged to 50 mM Bis-Tris/HCl buffer, pH 6.5, using Amicon Ultra centrifugal filters (MWCO 3kDa, Merck Millipore, NJ, USA). The resulting solution with purified protein was filtered through a 0.22 µm Millex®-GV filter (Merck Millipore, NJ, USA) and stored at 4 °C.

Expression and purification of TaLPMO9A in Aspergillus oryzae

*Ta*LPMO9A was expressed and partially purified at Novozymes (Bagsvaerd, Denmark) as described before[60]. The enzyme was saturated with Cu(II) followed by additional purification using the HiLoad 16/60 Superdex 75 size exclusion column (GE Healthcare, Uppsala, Sweden), and subsequent buffer exchange to 50 mM Bis-Tris/HCl buffer, pH 6.5, as described above for *P. pastoris* expressed *Ta*LPMO9A.

The protein concentrations of both final *Ta*LPMO9A preparations were determined by measuring absorbance at 280 nm and using a molar extinction coefficient of 45630 $M^{-1} \cdot cm^{-1}$ determined using ExPASy's ProtParam tool[61].

Apo-enzymes (for the kinetics of Cu^{2+} binding) were generated as described before[62] by 30 min incubation with EDTA in a 1:4 molar ratio (LPMO:EDTA), at room temperature.

Production of ^{13}C, ^{15}N-labeled TaLPMO9A in Pichia pastoris

The production of the isotopically labeled *Ta*LPMO9A-Pp was done as described before, using shake flask cultures[63,64]. Briefly, *P. pastoris* harboring the pPink-GAP plasmid, carrying the *Ta-lpmo9a* gene, was grown in 50 mL of ^{13}C, ^{15}N-labeled buffered minimal glucose medium (^{13}C, ^{15}N-BMD), in a 250-mL shake flask at 29 °C and 200 rpm for 24 hours. This culture was used to inoculate 450 mL ^{13}C, ^{15}N-BMD medium in 2-L shake flasks followed by incubation at 29 °C and 200 rpm for 48 hours. After the first 24 hours, the medium was re-supplemented with 1% (v/v) ^{13}C-labeled glucose. Cells were removed by centrifugation and the supernatant was concentrated to 100 mL using a VivaFlow 50 tangential crossflow concentrator (MWCO 10 kDa, Sartorius Stedim Biotech Gmbh, Goettingen, Germany).

The concentrated supernatant was supplemented with ammonium sulfate to a final concentration of 1.42 M and loaded onto a 5-mL HiTrap Phenyl FF column (GE Healthcare, Uppsala, Sweden), equilibrated with 50 mM Bis-Tris/HCl buffer (pH 6.5) containing 1.42 M ammonium sulfate, using a flow rate of 1 mL·min^{-1}. Proteins bound to the column were eluted using a 25-mL linear gradient from 1.42 M to 0 M ammonium sulfate in 50 mM Bis-Tris/HCl buffer (pH 6.5). Fractions containing the pure protein were identified using SDS-PAGE and subsequently pooled, concentrated and buffer exchanged to 50 mM Bis-Tris/HCl buffer, pH 6.5.

For NMR experiments, the *apo*-forms of *Ta*LPMO9A-Ao and ^{13}C,^{15}N-labeled *Ta*LPMO9A-Pp were obtained by incubating the protein samples with 10 mM Na-EDTA for 45 min at room temperature, followed by buffer exchange to 20 mM MES buffer, pH 5.5.

Analysis of N-terminal methylation

Protein samples were prepared for liquid chromatography – tandem mass spectrometry (LC-MS/MS) analysis as described before[65]. Briefly, dried protein samples were dissolved in 4% SDS, 10 mM dithiothreitol. After incubation for 25 minutes at 95 °C, iodoacetamide was added to a final concentration of 50 mM. The protein solution was then transferred to an STrap tip that was packed with two Empore™ C18 extraction disks (3M, St. Paul, MN, USA) and eleven MK360 quartz filters (Ahlstrom Munktell, Falun, Sweden), and that was prefilled with 90% (v/v) methanol, 100 mM Tris, pH 7.1 (STrapping solution). The samples in the STrap tips were washed first with the STrapping solution and then with 50 mM ammonium bicarbonate. Subsequently, trypsin was added to the STrap tips, which were then incubated for 45 min at 37 °C. The digestion was stopped by washing the STrap tips with 0.5% (v/v) trifluoroacetic acid and peptides were eluted with a solution containing 80 % (v/v) acetonitrile and 0.1 % (v/v) TFA.

The tryptic peptides (~1 µg) were separated on an Acclaim™ PepMap™ 100 C18 column, 50 cm × 75 µm (Thermo Fisher Scientific, Bremen Germany), using a 40 min gradient running from 12 % to 44 % acetonitrile. Chromatography was conducted using an Ultimate3000 RSLCnano/QExactive (Thermo Fisher Scientific, Bremen Germany) system, set up with a Nanospray Flex ion source. MS and MS/MS data were recorded using a standard data dependent acquisition method: *m/z* range 300-1500; automatic gain control targets of 3×10^6 (MS) and 5×10^4 (MS/MS); resolution of 70,000 (MS) and 35,000 (MS/MS), dynamic exclusion set to 20 s, and normalized collision energy set to 28.

Thermo's Xcalibur software (v3.0) was used to evaluate raw data, and to generate extracted ion chromatograms. The raw data were converted to mgf format using the msconvert module of Proteowizard (v 3.0.9016). The mgf files were used in Mascot searches against a database constructed by appending N-terminal variants of the target protein sequence to the reference proteome of *Komagataella phaffii/Pichia pastoris* (UniProt proteomes ID: UP000000314).

Substrates and electron donors

The following substrates were used in this study: phosphoric acid swollen cellulose (PASC), prepared from Avicel as described before[66]; xyloglucan from tamarind seed (TXG), ivory nut mannan, konjac glucomannan and cellopentaose, all purchased from Megazyme (Bray, Ireland); xylan from birch wood obtained from Sigma-Aldrich (St. Louis, MO, USA); steam exploded birch (*Betula pubescens*) wood (SEB) was prepared at 210 °C, with 10 min residence time, as described earlier[67].

Electron donors for the LPMOs were L-ascorbic acid (≥99.0%) purchased from Sigma-Aldrich (St. Louis, MO, USA) and cellobiose dehydrogenase from *Myriococcum thermophilum* (*Mt*CDH) expressed in *P. pastoris* and purified as previously reported[68].

LPMO reactions

Unless otherwise stated, reaction mixtures contained 2 mg·mL^{-1} substrate, 1 µM LPMO and, as electron donor, 1 mM ascorbic acid or 1 µM *Mt*CDH in 40 mM Bis-Tris/HCl buffer (pH 6.5). Reactions were performed in 2 mL Eppendorf tubes containing 100 µL total reaction volume, which were incubated at 45 °C with shaking at 1000 rpm in an Eppendorf Thermomixer (Eppendorf, Hamburg, Germany). Reactions were stopped by filtration using a 96-well plate equipped with a 0.45 µm filter (Merck Millipore, Billerica, MA, USA) that was operated with a vacuum manifold. Control experiments were performed in the absence of reductant.

Substrate specificity of the two *Ta*LPMO9A preparations was characterized using ascorbic acid as electron donor, and the reactions were incubated for 16 h in a total volume of 100 µL.

Product accumulation from PASC over time was followed using ascorbic acid or *Mt*CDH as electron donor. The total reaction volume was 500 µL, from which 100 µL aliquots were taken after 0.5, 1, 2, 4, and 24 h of incubation. Each reaction was done in triplicates.

Detection of oxidized products

Oxidized products were analyzed using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-ToF MS).

HPAEC was performed on a Dionex ICS5000 system, equipped with a CarboPac PA1 analytical column (2×250 mm) and a CarboPac PA1 guard column (2×50 mm), using a 50-minute gradient[69] for cellulosic and a 75-minute gradient[12] for hemicellulosic substrates. Chromatograms were recorded with Chromeleon and analyzed using Origin 9.1 software (OriginLab, Northampton, MA, USA).

MALDI-ToF MS was performed on an Ultraflex MALDI-ToF/ToF instrument (Bruker Daltonik GmbH, Bremen, Germany) equipped with a Nitrogen 337 nm laser, as described earlier [12]. The data were analyzed using mMass software[70]. Baseline correction and Gaussian smoothing (window size 0.3 m/z) were applied to all spectra.

Binding of TaLPMO9A to insoluble substrates

Binding studies with PASC were performed as described before[33]. The reaction mixture contained 2 mg·mL^{-1} substrate and 5 µM of the enzyme in 50 mM Bis-Tris/HCl buffer, pH 6.5, and was carried out at 45 °C with shaking at 1000 rpm in an Eppendorf Thermomixer (Eppendorf, Hamburg, Germany).

Determination of the pK_a of His1

NMR pH titrations of the *apo* forms of *Ta*LPMO9A were performed on ~0.3 mM non-isotopic labelled (natural abundance) methylated *Ta*LPMO9A-Ao and on ~0.6 mM ^{13}C and ^{15}N labeled non-methylated *Ta*LPMO9A-Pp, using 20 mM sodium phosphate buffers with 10 mM NaCl and 10% D_2O in the pH range 5.0 - 8.5. pH titrations were also carried out it in the presence of Zn^{2+} (a non-paramagnetic analogue of Cu^{2+} that has been previously used for NMR investigations of

LPMOs[27]), using 20 mM ammonium acetate buffers with 10 mM NaCl, 2 mM $ZnCl_2$ and 10 % D_2O in the pH range 3.5-8.5. Buffers were exchanged using Amicon Ultra centrifugal filters (MWCO 3 kDa, Merck Millipore, NJ, USA). Equation (1), which is a rearrangement of the Henderson-Hasselbach equation, was used to estimate the pK_a by using a least squares minimization algorithm to fit the pK_a in Excel, where δ_{BH} is the chemical shift of the fully protonated histidine, δ_B is the chemical shift of the fully deprotonated histidine, and δ is the measured chemical shift at a certain pH.

$$\delta = \frac{1 0^{(pH-pKa)} \delta_B + \delta_{BH}}{1 + 10^{(pH-pKa)}} \qquad (1)$$

In the case of methylated *Ta*LPMO9A-Ao, a ^{13}C Heteronuclear Single Quantum Coherence (HSQC) spectrum for the aliphatic region was recorded at each titration point to monitor the changes in chemical shifts of the methyl group of His1. This methyl group has a unique chemical shift[71] usually not observed in proteins, which was assigned at pH 6.5: $\delta^1H/\delta^{13}C$ {3.8 ppm; 35.2 ppm}.

For ^{13}C and ^{15}N labeled non-methylated *Ta*LPMO9A-Pp, a ^{13}C HSQC spectrum for the aromatic region was recorded at each titration point to monitor the changes in chemical shifts of the epsilon atoms ($C^{\varepsilon 1}$/ $H^{\varepsilon 1}$) of His1, which was assigned at pH 6.5: $\delta^1H/\delta^{13}C$ {8.3 ppm; 133.5 ppm}.

Saccharification experiments

Saccharification of sulfite-pulped Norway spruce prepared using the BALI™ process[72] was conducted in 60-mL rubber sealed glass bottles (Wheaton, Millville, NJ, USA) with 10-mL working volume and 50 $g \cdot L^{-1}$ total solids loading in 50 mM sodium acetate buffer, pH 5.0, in the

presence or absence of 1 mM ascorbic acid. The reactions contained 4 mg of enzyme per g of glucan, and the enzyme mixture was composed of 85 % (w/w) Celluclast 1.5L:Novozym 188 blend (in a 5:1 [w/w] ratio), and 15 % (w/w) *Ta*LPMO9A-Pp, *Ta*LPMO9A-Ao or bovine serum albumin (BSA). Earlier, this C/N188 and LPMO ratio had been found optimal for obtaining the highest sugar recovery from steam-exploded birch[7]. Reactions were incubated at 50 °C, with mixing at 38 rpm in a Multi RS-60 programmable rotator (Biosan, Riga, Latvia). 100 µL aliquots were taken after 4, 24, 48, and 96 h of incubation, followed by heat inactivation at 100 °C for 15 min. As a reference, Cellic® CTec2 was used. All three commercial preparations, Celluclast 1.5L, Novozym 188 and Cellic® CTec2, were kindly provided by Novozymes A/S (Bagsværd, Denmark) and their protein concentration was determined with the Bio-Rad Protein Assay (Bio-Rad, Hercules, CA, USA), based on the Bradford method[73], using BSA as a standard.

Glucose and cellobiose released during enzymatic hydrolysis were quantified with High-Performance Liquid Chromatography (HPLC) using a Dionex Ultimate 3000 system (Dionex, Sunnyvale, CA, USA) coupled to a refractive index (RI) detector 101 (Shodex, Tokyo, Japan), as described earlier [32]. Hydrolysis yields were calculated based on detected glucose and cellobiose (typically less than 1 % of the total).

Determination of the Cell Potential (E°)

The cell potentials for the redox couples *Ta*LPMO9A-Ao(Cu^{2+})/*Ta*LPMO9A-Ao(Cu^{+}) and *Ta*LPMO9A-Pp(Cu^{2+})/*Ta*LPMO9A-Pp(Cu^{+}) were determined as described by Aachmann et al[27]. Solutions (50 µL) of oxygen-free N,N,N′,N′-tetramethyl-1,4 phenylenediamine (TMP_{red}) in its reduced form (200 µM) and Cu^{2+}-charged *Ta*LPMO9A-Ao or *Ta*LPMO9A-Pp (70 µM) in 20 mM MES buffer (pH 5.5, t = 25 °C) were mixed and the extent of reaction was determined by measuring absorbance from the formed TMP radical cation (TMP_{ox}) at λ = 610 nm. The

concentrations of TMP_{ox}, which equal concentrations of LPMO-Cu^{+}, were calculated by using an extinction coefficient of 14.0 $mM^{-1}cm^{-1}$ [74]. From the determined concentrations of TMP_{ox} and LPMO-Cu^{+}, the equilibrium constant (K) was calculated. The cell potential for the LPMO-Cu^{2+}/LPMO-Cu^{+} redox couple was determined by adding the known cell potential of 273 mV for TMP_{red}/TMP_{ox} [75] to the cell potential of the equilibrium reaction of TMP_{red} and LPMO-Cu^{2+} [27].

Stopped-flow kinetics

Kinetic experiments to measure changes in fluorescence emission upon binding of Cu^{2+} were carried out using a SFM4000 stopped-flow spectrophotometer (BioLogic Science Instruments, Grenoble, France). Apo-LPMO samples (2.0 µM, final concentration) were prepared in 20 mM MES buffer, pH 5.5 and mixed with samples of Cu^{2+} (in the form of $CuSO_4 \cdot 5H_2O$) in the same buffer with concentrations ranging from 6 - 125 µM after mixing. We used a single mixing set-up, which means that for each experiment a new syringe with the appropriate concentration of $CuSO_4$ was prepared. Tryptophan residues were excited at 295 nm and emitted light was detected after passage through a cut-off filter permitting light with wavelengths above 320 nm to pass, using a photomultiplier tube (670 V) positioned at a 90 degrees angle relative to the exciting beam. Kinetic data were analyzed using the Bio-Kine software (BioLogic Science Instruments, Grenoble, France).

Differential scanning calorimetry

Differential scanning calorimetry (DSC) to analyze the thermal stabilities of *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp was carried out with a Nano-Differential Scanning Calorimeter III instrument (TA Instruments, New Castle, DE, USA). *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp (15 µM and 10 µM, respectively) were saturated with Cu(II) by incubating with an excess of Cu(II)SO_4 (at 3:1 molar ratio of copper:enzyme) for 90 min at room temperature. Subsequently,

the enzymes were dialyzed against 2 × 1000 volumes of 50 mM Na phosphate buffer, pH 6.5, for 24 h at 4 °C. Prior to loading into sample cell, the enzyme solutions were filtered through a 0.22 µm filter (Merck Millipore, Billerica, MA, USA) and degassed. The dialysis buffer was also filtered, degassed and loaded into reference cell. Scanning was performed in the range from 30 – 90 °C, with a scan rate of 1 °C·min^{-1}. The experiments were carried out in duplicate, using freshly dialyzed enzyme for each scan. The data were analyzed using NanoAnalyze software (TA Instruments, New Castle, DE, USA) whereby baseline scans, collected with buffer in both, reference and sample cells, were subtracted from the sample scans. Transition temperatures, T_m (defined as the temperature of maximum apparent heat capacity) were estimated from the baseline-corrected DSC thermograms.

H_2O_2 production

The method for measuring H_2O_2 production was adapted from a previously reported protocol[21]. A reaction mixture (180 µL) containing 1 µM *Ta*LPMO9A, 5 U·mL^{-1} horseradish peroxidase (HRP) and 100 µM Amplex® Red (Thermo Fisher Scientific, Bremen Germany) in 50 mM Bis-Tris/HCl buffer pH 6.5 was incubated for 5 min at 40 °C in a 96-well microtiter plate in a plate reader (Multiskan™ FC Microplate Photometer (Thermo Fisher Scientific, Bremen Germany). The reaction was initiated by the addition of 20 µL of 500 µM ascorbic acid (50 µM final concentration) in each well and the production of resorufin was monitored at 540 nm. Control reactions in the absence of *Ta*LPMO9A were carried out to obtain the LPMO-independent resorufin production rate. This control reaction provided a background signal equal to 0.3 % of the LPMO-catalyzed reaction and was subtracted from the latter. A H_2O_2 standard curve was prepared using the same conditions (without ascorbic acid and LPMO). The reactions

were monitored for 45 min and H_2O_2 production rates were derived from data points in the linear region, between 2 and 20 min.

Sensitivity towards H_2O_2

To assess the impact of H_2O_2 on *Ta*LPMO9A, 500 µL reaction mixtures were prepared as described above with the addition of H_2O_2 corresponding to 0, 50, 100, 200, 300 and 500 µM final concentrations. H_2O_2 was added to the reaction mixture just before initiation of the reaction by addition of ascorbic acid (1 mM final concentration). At regular intervals of 3, 6, 9, 30 and 60 min 55 µL samples were taken from the reaction mixtures, and LPMO activity was immediately stopped by filtration using a 96-well filter plate (Merck Millipore, Billerica, MA, USA) operated with a vacuum manifold. Filtered samples were frozen (-20 °C) prior to further analysis.

Supplementary material

Figure S1: SDS-PAGE of purified TaLPMO9A-Ao and TaLPMO9A-Pp. **Figure S2:** 3D structure of TaLPMO9A (PDB ID: 3ZUD). **Figure S3:** LC-MS chromatograms of N-terminal peptides. **Figure S4:** Binding to PASC. **Figure S5:** Reaction products generated from steam exploded birch. **Figure S6:** HPAEC-PAD chromatograms for the reactions with cellopentaose. **Figure S7:** Close-up of MALDI-ToF MS spectra for the reactions with PASC or xyloglucan. **Figure S8:** NMR pH titration curves. **Figure S9:** Kinetic information for copper binding obtained by stopped-flow spectrophotometry. **Figure S10:** Differential scanning calorimetry analysis. **Figure S11:** Control reactions to check for occurrence of copper-catalyzed Fenton-type chemistry.

Acknowledgments

This work was supported by the Research Council of Norway through grants 226244 and 243663. We thank Novozymes for supply of *Ta*LPMO9A produced in *Aspergillus oryzae*.

Conflicts of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sorlie M, Eijsink VGH (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science 330:219-222.
2. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JCN, Johansen KS, Krogh KBRM, Jorgensen CI, Tovborg M, Anthonsen A and others (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proc Natl Acad Sci USA 108:15079-15084.
3. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH (2012) Novel enzymes for the degradation of cellulose. Biotechnol Biofuels 5:45.
4. Beeson WT, Vu VV, Span EA, Phillips CM, Marletta MA (2015) Cellulose degradation by polysaccharide monooxygenases. Annu Rev Biochem 84:923-946.
5. Hemsworth GR, Johnston EM, Davies GJ, Walton PH (2015) Lytic polysaccharide monooxygenases in biomass conversion. Trends Biotechnol 33:747-61.
6. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels 6:41.

7. Muller G, Varnai A, Johansen KS, Eijsink VGH, Horn SJ (2015) Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. Biotechnol Biofuels 8:187.
8. Hu JG, Chandra R, Arantes V, Gourlay K, van Dyk JS, Saddler JN (2015) The addition of accessory enzymes enhances the hydrolytic performance of cellulase enzymes at high solid loadings. Bioresour Technol 186:149-153.
9. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:490-495.
10. Phillips CM, Beeson WT, Cate JH, Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. ACS Chem Biol 6:1399-1406.
11. Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VGH, Igarashi K, Samejima M, Stahlberg J, Horn SJ, Sandgren M (2011) The putative endoglucanase PcGH61D from Phanerochaete chrysosporium is a metal-dependent oxidative enzyme that cleaves cellulose. PLoS One 6:e27807.
12. Agger JW, Isaksen T, Varnai A, Vidal-Melgosa S, Willats WGT, Ludwig R, Horn SJ, Eijsink VGH, Westereng B (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc Natl Acad Sci USA 111:6287-6292.
13. Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, Fanuel M, Ropartz D, Rogniaux H, Gimbert I and others (2015) Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. Biotechnol Biofuels 8:90.

14. Lo Leggio L, Simmons TJ, Poulsen JCN, Frandsen KEH, Hemsworth GR, Stringer MA, von Freiesleben P, Tovborg M, Johansen KS, De Maria L and others (2015) Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. Nat Commun 6:5961.
15. Vu VV, Beeson WT, Span EA, Farquhar ER, Marletta MA (2014) A family of starch-active polysaccharide monooxygenases. Proc Natl Acad Sci USA 111:13822-13827.
16. Frommhagen M, Sforza S, Westphal AH, Visser J, Hinz SW, Koetsier MJ, van Berkel WJ, Gruppen H, Kabel MA (2015) Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. Biotechnol Biofuels 8:101.
17. Vu VV, Beeson WT, Phillips CM, Cate JH, Marletta MA (2014) Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. J Am Chem Soc 136:562-565.
18. Isaksen T, Westereng B, Aachmann FL, Agger JW, Kracher D, Kittl R, Ludwig R, Haltrich D, Eijsink VGH, Horn SJ (2014) A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. J Biol Chem 289:2632-2642.
19. Harris PV, Welner D, McFarland KC, Re E, Poulsen JCN, Brown K, Salbo R, Ding HS, Vlasenko E, Merino S and others (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. Biochem 49:3305-3316.
20. Frommhagen M, Koetsier MJ, Westphal AH, Visser J, Hinz SWA, Vincken JP, van Berkel WJH, Kabel MA, Gruppen H (2016) Lytic polysaccharide monooxygenases from

Myceliophthora thermophila C1 differ in substrate preference and reducing agent specificity. Biotechnol Biofuels 9:186.

21. 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.
22. Wu M, Beckham GT, Larsson AM, Ishida T, Kim S, Payne CM, Himmel ME, Crowley MF, Horn SJ, Westereng B and others (2013) Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus Phanerochaete chrysosporium. J Biol Chem 288:12828-12839.
23. Nekiunaite L, Petrovic DM, Westereng B, Vaaje-Kolstad G, Hachem MA, Varnai A, Eijsink VGH (2016) FgLPMO9A from Fusarium graminearum cleaves xyloglucan independently of the backbone substitution pattern. FEBS Lett 590:3346-3356.
24. Saloheimo M, Nakari-Setala T, Tenkanen M, Penttila M (1997) cDNA cloning of a Trichoderma reesei cellulase and demonstration of endoglucanase activity by expression in yeast. Eur J Biochem 249:584-91.
25. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27:297-306.
26. Nevalainen H, Peterson R (2014) Making recombinant proteins in filamentous fungi- are we expecting too much? Front Microbiol 5:75.
27. Aachmann FL, Sørlie M, Skjåk-Bræk G, Eijsink VGH, 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 USA 109:18779-18784.

28. Hemsworth GR, Henrissat B, Davies GJ, Walton PH (2014) Discovery and characterization of a new family of lytic polysaccharide rnonooxygenases. Nat Chem Biol 10:122-126.
29. Vaaje-Kolstad G, Forsberg Z, Loose JS, Bissaro B, Eijsink VGH (2017) Structural diversity of lytic polysaccharide monooxygenases. Curr Opin Struct Biol 44:67-76.
30. Li X, Beeson WT, Phillips CM, Marletta MA, Cate JHD (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. Structure 20:1051-1061.
31. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U (2014) Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. Biotechnol Biofuels 7:135.
32. Chylenski P, Petrović DM, Müller G, Dahlström M, Bengtsson O, Lersch M, Siika-aho M, Horn SJ, Eijsink VGH (2017) Enzymatic degradation of sulfite-pulped softwoods and the role of LPMOs. Biotechnol Biofuels 10:177.
33. Forsberg Z, Nelson CE, Dalhus B, Mekasha S, Loose JSM, Crouch LI, Rohr AK, Gardner JG, Eijsink VGH, Vaaje-Kolstad G (2016) Structural and functional analysis of a lytic polysaccharide monooxygenase important for efficient utilization of chitin in Cellvibrio japonicus. J Biol Chem 291:7300-7312.
34. Crouch LI, Labourel A, Walton PH, Davies GJ, Gilbert HJ (2016) The contribution of non-catalytic carbohydrate binding modules to the activity of lytic polysaccharide monooxygenases. J Biol Chem 291:7439-7449.
35. Kojima Y, Varnai A, Ishida T, Sunagawa N, Petrovic DM, Igarashi K, Jellison J, Goodell B, Alfredsen G, Westereng B and others (2016) Characterization of an LPMO from the

brown-rot fungus Gloeophyllum trabeum with broad xyloglucan specificity, and its action on cellulose-xyloglucan complexes. Appl Environ Microbiol 82:6557-6572.

36. Westereng B, Cannella D, Agger JW, Jorgensen H, Andersen ML, Eijsink VGH, Felby C (2015) Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. Sci Rep 5:18561.

37. Frandsen KEH, Simmons TJ, Dupree P, Poulsen JCN, Hemsworth GR, Ciano L, Johnston EM, Tovborg M, Johansen KS, von Freiesleben P and others (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. Nat Chem Biol 12:298-303.

38. Borisova AS, Isaksen T, Dimarogona M, Kognole AA, Mathiesen G, Varnai A, Rohr AK, Payne CM, Sorlie M, Sandgren M and others (2015) Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity. J Biol Chem 290:22955-22969.

39. Chaplin AK, Wilson MT, Hough MA, Svistunenko DA, Hemsworth GR, Walton PH, Vijgenboom E, Worrall JA (2016) Heterogeneity in the histidine-brace copper coordination sphere in auxiliary activity family 10 (AA10) lytic polysaccharide monooxygenases. J Biol Chem 291:12838-12850.

40. Cannella D, Hsieh CWC, Felby C, Jorgensen H (2012) Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. Biotechnol Biofuels 5:26.

41. Bissaro B, Rohr AK, Muller G, Chylenski P, Skaugen M, Forsberg Z, Horn SJ, Vaaje-Kolstad G, Eijsink VGH (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. Nat Chem Biol 13:1123-1128.

42. Bissaro B, Rohr AK, Skaugen M, Forsberg Z, Horn SJ, Vaaje-Kolstad G, Eijsink V (2016) Fenton-type chemistry by a copper enzyme: molecular mechanism of polysaccharide oxidative cleavage. bioRxiv 097022; doi: https://doi.org/10.1101/097022.

43. Hangasky JA, Iavarone AT, Marletta MA (2018) Reactivity of O2 versus H2O2 with polysaccharide monooxygenases. Proc Natl Acad Sci USA.

44. Kim IJ, Seo N, An HJ, Kim JH, Harris PV, Kim KH (2017) Type-dependent action modes of TtAA9E and TaAA9A acting on cellulose and differently pretreated lignocellulosic substrates. Biotechnol Biofuels 10:46.

45. Loose JS, Forsberg Z, Kracher D, Scheiblbrandner S, Ludwig R, Eijsink VGH, Vaaje-Kolstad G (2016) Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. Protein Sci 25:2175-2186.

46. Kracher D, Scheiblbrandner S, Felice AK, Breslmayr E, Preims M, Ludwicka K, Haltrich D, Eijsink VGH, Ludwig R (2016) Extracellular electron transfer systems fuel cellulose oxidative degradation. Science 352:1098-1101.

47. Paiva AC, Juliano L, Boschcov P (1976) Ionization of methyl derivatives of imidazole, histidine, thyreotropin releasing factor, and related compounds. J Am Chem Soc 98:7645-8.

48. Tickler AK, Smith DG, Ciccotosto GD, Tew DJ, Curtain CC, Carrington D, Masters CL, Bush AI, Cherny RA, Cappai R and others (2005) Methylation of the imidazole side chains of the Alzheimer disease amyloid-beta peptide results in abolition of superoxide dismutase-like structures and inhibition of neurotoxicity. J Biol Chem 280:13355-63.

49. Kim S, Stahlberg J, Sandgren M, Paton RS, Beckham GT (2014) Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. Proc Natl Acad Sci USA 111:149-154.
50. Kuusk S, Bissaro B, Kuusk P, Forsberg Z, Eijsink VGH, Sorlie M, Valjamae P (2018) Kinetics of H2O2-driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase. J Biol Chem 293:523-531.
51. Peterson RW, Walton JH (1943) The autoxidation of ascorbic acid. J Am Chem Soc 65:1212-1217.
52. Boatright WL (2016) Oxygen dependency of one-electron reactions generating ascorbate radicals and hydrogen peroxide from ascorbic acid. Food Chem 196:1361-1367.
53. Uchida K, Kawakishi S (1993) 2-Oxo-histidine as a novel biological marker for oxidatively modified proteins. FEBS Lett 332:208-210.
54. Uchida K, Kawakishi S (1989) Ascorbate-mediated specific oxidation of the imidazole ring in a histidine derivative. Bioorg Chem 17:330-343.
55. Valderrama B, Ayala M, Vazquez-Duhalt R (2002) Suicide inactivation of peroxidases and the challenge of engineering more robust enzymes. Chem Biol 9:555-565.
56. Mao L, Luo S, Huang Q, Lu J (2013) Horseradish peroxidase inactivation: heme destruction and influence of polyethylene glycol. Sci Rep 3:3126.
57. Gonzalez-Perez D, Garcia-Ruiz E, Ruiz-Duenas FJ, Martinez AT, Alcalde M (2014) Structural determinants of oxidative stabilization in an evolved versatile peroxidase. ACS Catal 4:3891-3901.

58. Green AP, Hayashi T, Mittl PRE, Hilvert D (2016) A Chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. J Am Chem Soc 138:11344-11352.
59. Loose JSM, Forsberg Z, Fraaije MW, Eijsink VGH, 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.
60. Dotson WD, Greenier J, Ding H (2006) Polypeptides having cellulolytic enhancing activity and polynucleotides encoding same. US20060005279 A1 U.S.A.
61. Gasteiger E. HC, Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The Proteomics Protocols Handbook: Humana Press; 2005. p 571-607.
62. Forsberg Z, Mackenzie AK, Sorlie M, Rohr AK, Helland R, Arvai AS, Vaaje-Kolstad G, Eijsink VGH (2014) Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. Proc Natl Acad Sci USA 111:8446-51.
63. Pickford AR, O'Leary JM, Isotopic labeling of recombinant proteins from the methylotrophic yeast Pichia pastoris. In: Downing AK, editor. Protein NMR Techniques. Totowa, NJ: Humana Press; 2004. p 17-33.
64. Courtade G, Wimmer R, Rohr AK, Preims M, Felice AK, Dimarogona M, Vaaje-Kolstad G, Sorlie M, Sandgren M, Ludwig R and others (2016) Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. Proc Natl Acad Sci USA 113:5922-7.

65. Zougman A, Selby PJ, Banks RE (2014) Suspension trapping (STrap) sample preparation method for bottom- up proteomics analysis. Proteomics 14:1006-1010.

66. Wood TM (1988) Preparation of crystalline, amorphous, and dyed cellulase substrates. Method Enzymol 160:19-25.

67. Vivekanand V, Olsen EF, Eijsink VGH, Horn SJ (2013) Effect of different steam explosion conditions on methane potential and enzymatic saccharification of birch. Bioresour Technol 127:343-349.

68. Zamocky M, Schumann C, Sygmund C, O'Callaghan J, Dobson ADW, Ludwig R, Haltrich D, Peterbauer CK (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.

69. Westereng B, Agger JW, Horn SJ, Vaaje-Kolstad G, Aachmann FL, Stenstrom YH, Eijsink VGH (2013) Efficient separation of oxidized cello-oligosaccharides generated by cellulose degrading lytic polysaccharide monooxygenases. J Chromatogr A 1271:144-152.

70. Strohalm M, Kavan D, Novak P, Volny M, Havlicek V (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal Chem 82:4648-4651.

71. Henderson WW, Shepherd RE, Abola J (1986) Proton and carbon-13 NMR spectra of a series of methyl-substituted imidazole complexes of pentaamminecobalt(III) and crystal structure of the remote isomer of [(4-methylimidazole)pentaamminecobalt(III)] trichloride dihydrate. Inorg Chem 25:3157-3163.

72. Rodsrud G, Lersch M, Sjode A (2012) History and future of world's most advanced biorefinery in operation. Biomass & Bioenergy 46:46-59.

73. Bradford MM (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.

74. Sorlie M, Seefeldt LC, Parker VD (2000) Use of stopped-flow spectrophotometry to establish midpoint potentials for redox proteins. Anal Biochem 287:118-25.

75. Liu Y, Seefeldt LC, Parker VD (1997) Entropies of redox reactions between proteins and mediators: the temperature dependence of reversible electrode potentials in aqueous buffers. Anal Biochem 250:196-202.

Figure legends

Figure 1: The active site of *Ta*LPMO9A. Two histidine side chains (His1 and His86) contribute to copper coordination in a highly conserved structural arrangement[2]. In fungal LPMOs His1 is methylated; the methylgroup is indicated by dotted red lines. Oxidation of histidine, discussed further below, is thought to start at Cε1[41,53,54], which is indicated by a red star and which is adjacent to the methylated nitrogen.

Figure 2: Reaction products generated by *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp from PASC or TXG. (A) HPAEC-PAD profiles showing products released from PASC. Oxidized products are indicated in the Fig. and assignments of these peaks were based on the previous work[18,23]. Native products elute prior to the C1-oxidized products. AA indicates the presence of 1 mM ascorbic acid in the reaction mixture. (B) HPAEC-PAD profiles showing products released

from TXG. (C) MALDI-ToF MS spectra of products released in reactions containing *Ta*LPMO9A-Ao (red line) or *Ta*LPMO9A-Pp (black line), PASC and ascorbic acid, after Na^+ saturation. The spectra show the DP4 – DP8 region (mass values are provided in Fig. S5). (D) MALDI-ToF MS spectra of products obtained in reactions containing *Ta*LPMO9A-Ao (red line) or *Ta*LPMO9A-Pp (black line), TXG and ascorbic acid, after Na^+ saturation. The indicated *m/z* values refer to sodium adducts and the nature of the products is indicated using the following abbreviations: Hex, hexose; Pen, pentose; ox, oxidized. For more details about detected products, see Fig. S7.

Figure 3: Time course for release of oxidized products during incubation of 1 µM *Ta*LPMO9A-Ao (red line) or *Ta*LPMO9A-Pp (black line) with 2 mg/ml PASC and (A) 1 µM *Mt*CDH, or (B) 1 mM ascorbic acid, in 40 mM Bis-Tris, pH 6.5, at 45 °C. The amount of released oxidized products for each sample is expressed as the sum of the integrated peak areas (nC × min) for C1 products only (A; in this case C4 products cannot be detected because they are oxidized by CDH) or C1 and C4 products together (B). Consequently, values on the y-axis in the graph A and B should not be compared; quantification is only valid for comparing the two enzyme variants within the same panel.

Figure 4: Saccharification of sulfite-pulped spruce in the (A) absence and (B) presence of ascorbic acid (1 mM), upon substitution of 15 % of the proteins in a Celluclast and Novozym 188 (C/N188) blend with the same amount of BSA (blue bar), *Ta*LPMO9A-Pp (black bar) or *Ta*LPMO9A-Ao (red bar). For comparison, the results for a reaction with an LPMO-containing commercial cellulase cocktail, Cellic™ CTec2 (green bar), are included. Note that it has been shown previously that in the absence of an externally added reductant (as in panel A) LPMO activity on this substrate is low[32].

Figure 5: The effect of H_2O_2 on product generation by *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp over time. The Fig.s show time-courses for the release of oxidized products in reactions containing 2.5 g/L PASC and 1 µM of *Ta*LPMO9A-Pp (A; zoom-in view in B), or *Ta*LPMO9A-Ao (C; zoom-in view D), in the presence of different initial concentrations of exogenous H_2O_2 (0 – 500 µM) and 1mM ascorbic acid. Note that this Fig. shows that increasing H_2O_2 concentrations lead to higher initial activity (clearly visible in panels B & D) and to higher rates of LPMO inactivation. The trade-off between these two phenomena determines the apparent initial rate (measured at 3 minutes) and the overall shape of the progress curves.

Figure 6: The effect of H_2O_2 on *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp. (A) Relative increase in apparent initial rate, calculated on the basis of product formation after 3 min, for *Ta*LPMO9A-Ao (red line) and *Ta*LPMO9A-Pp (black line) compared to the reference reaction without initial exogenous H_2O_2 (40 mM Bis-Tris, pH 6.5, at 45 °C). See Fig. 7 for the actual product levels and see Fig. 5 for the complete progress curves.

Figure 7: The amount of oxidized products generated during incubation of 1 µM *Ta*LPMO9A-Ao (red line) or *Ta*LPMO9A-Pp (black line) with 2 mg/ml PASC, 1 mM ascorbic acid and different initial concentrations of H_2O_2, after (A) 3 min, and (B) 60 min (40 mM Bis-Tris, pH 6.5, at 45 °C). The amount of released oxidized products is expressed as the sum of the integrated areas (nC × min) of the peaks representing C1 and C4 oxidized products. For complete progress curves, see Fig. 5.

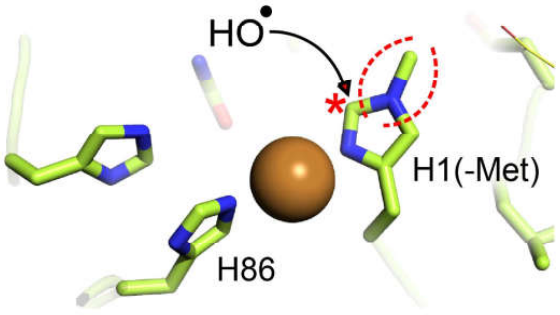

Figure 1: The active site of TaLPMO9A. Two histidine side chains (His1 and His86) contribute to copper coordination in a highly conserved structural arrangement.2 In fungal LPMOs His1 is methylated; the methylgroup is indicated by dotted red lines. Oxidation of histidine, discussed further below, is thought to start at Cε1,41,53,54 which is indicated by a red star and which is adjacent to the methylated nitrogen.

917x617mm (96 x 96 DPI)

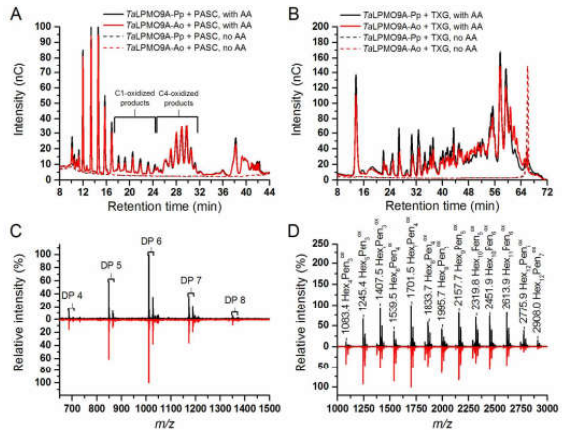

Figure 2. Reaction products generated by TaLPMO9A-Ao and TaLPMO9A-Pp from PASC or TXG. (A) HPAEC-PAD profiles showing products released from PASC. Oxidized products are indicated in the Fig. and assignments of these peaks were based on the previous work18,23. Native products elute prior to the C1-oxidized products. AA indicates the presence of 1 mM ascorbic acid in the reaction mixture. (B) HPAEC-PAD profiles showing products released from TXG. (C) MALDI-ToF MS spectra of products released in reactions containing TaLPMO9A-Ao (red line) or TaLPMO9A-Pp (black line), PASC and ascorbic acid, after Na+ saturation. The spectra show the DP4 – DP8 region (mass values are provided in Fig. S5). (D) MALDI-ToF MS spectra of products obtained in reactions containing TaLPMO9A-Ao (red line) or TaLPMO9A-Pp (black line), TXG and ascorbic acid, after Na+ saturation. The indicated m/z values refer to sodium adducts and the nature of the products is indicated using the following abbreviations: Hex, hexose; Pen, pentose; ox, oxidized. For more details about detected products, see Fig. S7.

160x119mm (300 x 300 DPI)

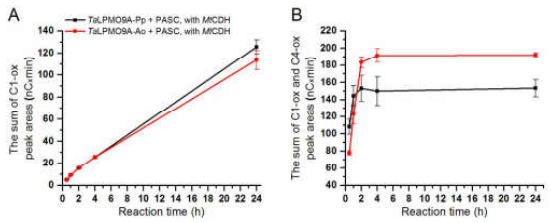

Figure 3: Time course for release of oxidized products during incubation of 1 μM TaLPMO9A-Ao (red line) or TaLPMO9A-Pp (black line) with 2 mg/ml PASC and (A) 1 μM MtCDH, or (B) 1 mM ascorbic acid, in 40 mM Bis-Tris, pH 6.5, at 45 °C. The amount of released oxidized products for each sample is expressed as the sum of the integrated peak areas (nC × min) for C1 products only (A; in this case C4 products cannot be detected because they are oxidized by CDH) or C1 and C4 products together (B). Consequently, values on the y-axis in the graph A and B should not be compared; quantification is only valid for comparing the two enzyme variants within the same panel.

160x59mm (300 x 300 DPI)

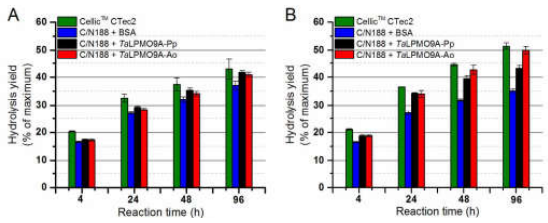

Figure 4: Saccharification of sulfite-pulped spruce in the (A) absence and (B) presence of ascorbic acid (1 mM), upon substitution of 15 % of the proteins in a Celluclast and Novozym 188 (C/N188) blend with the same amount of BSA (blue bar), TaLPMO9A-Pp (black bar) or TaLPMO9A-Ao (red bar). For comparison, the results for a reaction with an LPMO-containing commercial cellulase cocktail, Cellic™ CTec2 (green bar), are included. Note that it has been shown previously that in the absence of an externally added reductant (as in panel A) LPMO activity on this substrate is low.32

160x59mm (300 x 300 DPI)

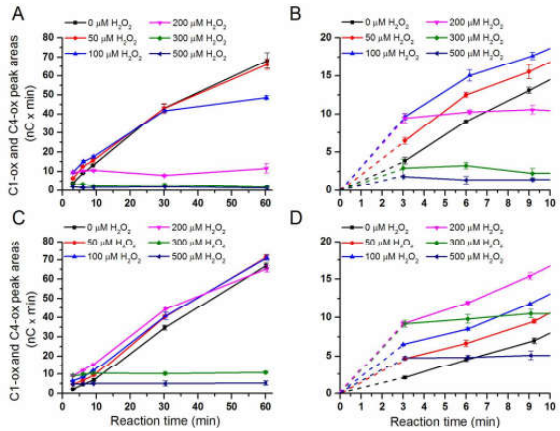

Figure 5: The effect of H2O2 on product generation by TaLPMO9A-Ao and TaLPMO9A-Pp over time. The Fig.s show time-courses for the release of oxidized products in reactions containing 2.5 g/L PASC and 1 μM of TaLPMO9A-Pp (A; zoom-in view in B), or TaLPMO9A-Ao (C; zoom-in view D), in the presence of different initial concentrations of exogenous H2O2 (0 – 500 μM) and 1mM ascorbic acid. Note that this Fig. shows that increasing H2O2 concentrations lead to higher initial activity (clearly visible in panels B & D) and to higher rates of LPMO inactivation. The trade-off between these two phenomena determines the apparent initial rate (measured at 3 minutes) and the overall shape of the progress curves.

160x119mm (300 x 300 DPI)

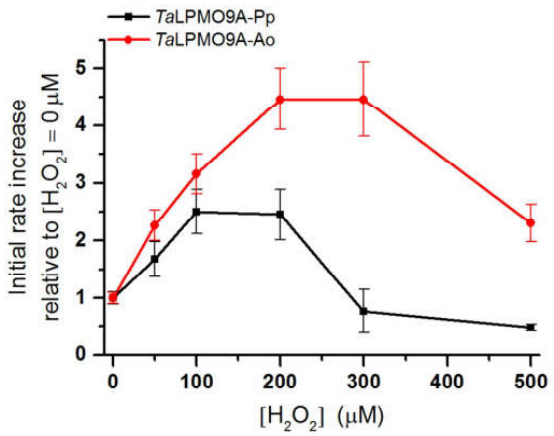

Figure 6: The effect of H2O2 on TaLPMO9A-Ao and TaLPMO9A-Pp. (A) Relative increase in apparent initial rate, calculated on the basis of product formation after 3 min, for TaLPMO9A-Ao (red line) and TaLPMO9A-Pp (black line) compared to the reference reaction without initial exogenous H2O2 (40 mM Bis-Tris, pH 6.5, at 45 °C). See Fig. 7 for the actual product levels and see Fig. 5 for the complete progress curves.

241x190mm (96 x 96 DPI)

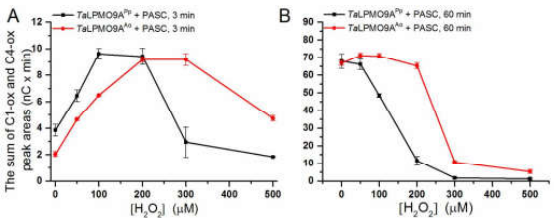

Figure 7: The amount of oxidized products generated during incubation of 1 μM TaLPMO9A-Ao (red line) or TaLPMO9A-Pp (black line) with 2 mg/ml PASC, 1 mM ascorbic acid and different initial concentrations of H2O2, after (A) 3 min, and (B) 60 min (40 mM Bis-Tris, pH 6.5, at 45 °C). The amount of released oxidized products is expressed as the sum of the integrated areas (nC × min) of the peaks representing C1 and C4 oxidized products. For complete progress curves, see Fig. 5.

160x59mm (300 x 300 DPI)

Supplementary Material

Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation

Dejan M. Petrović[1], Bastien Bissaro[1], Piotr Chylenski[1], Morten Skaugen[1], Morten Sørlie[1], Marianne S. Jensen[1], Finn L. Aachmann[2], Gaston Courtade[2], Anikó Várnai[1], Vincent G.H. Eijsink[1]*

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

[2]Department of Biotechnology and Food Science, NOBIPOL, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

*Corresponding author:

Vincent G.H. Eijsink,

Faculty of Chemistry, Biotechnology, and Food Science, The Norwegian University of Life Sciences (NMBU), 1432 Ås, Norway. Tel.: +47-67232463; E-mail: vincent.eijsink@nmbu.no

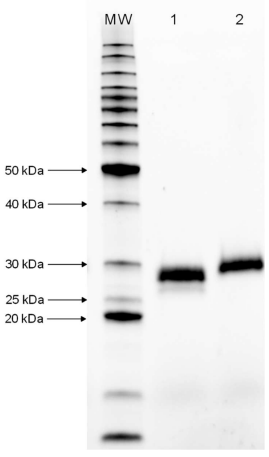

Figure S1. SDS-PAGE of purified *Ta*LPMO9A-Ao and *Ta*LPMO9A-Pp. Lanes: (Mw) Benchmark standard protein molecular weight marker (relevant molecular weight values are indicated); (1) purified *Ta*LPMO9A-Ao after size exclusion chromatography; (2) purified *Ta*LPMO9A-Pp after size exclusion chromatography.

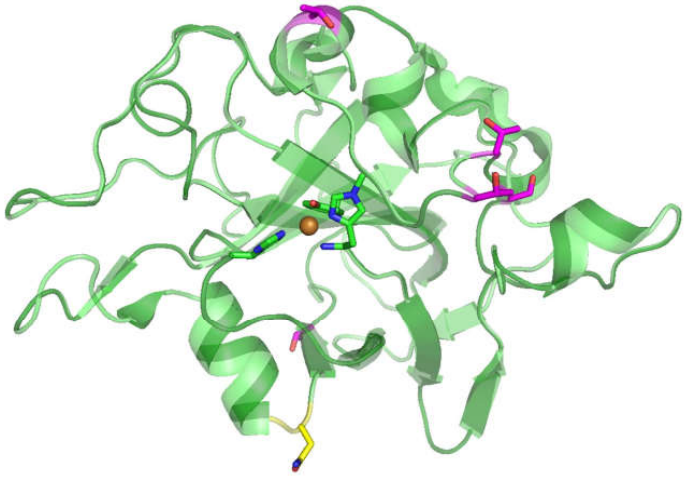

Figure S2. 3D structure of *Ta*LPMO9A (PDB ID: 3ZUD). The side chains of Met-His1, His86 and Tyr175, involved in coordination of the copper (brown sphere) are shown, with green carbons. The side chain of Asn138, which may be *N*-glycosylated, is shown with yellow carbons. The side chains of potentially *O*-glycosylated amino acid residues are shown with purple carbons.

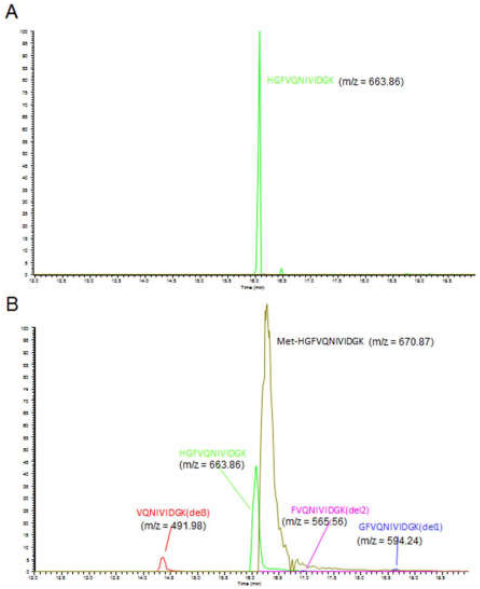

Figure S3. LC-MS extracted ion chromatograms of N-terminal peptides derived from (A) *Ta*LPMO9A-Pp, and (B) *Ta*LPMO9A-Ao. The *Ta*LPMO9A-Pp sample (A) shows a peak with a retention time of 16.10 min and a detected mass of 663.86 Da that can be ascribed to a double charged non-methylated N-terminal peptide (*m/z* for $[HGFVQNIVIDGK]^{2+}$ = 663.26 Da). Other relevant masses, including a mass that corresponds to the methylated N-terminal peptide were not detected. LC-MS analysis of the *Ta*LPMO9A-Ao sample (B) revealed several peaks with retention times between 14.30 and 18.80 min. The most dominant peak appeared at the retention time of 16.28 min and had a mass of 670.87 Da, which corresponds to a double charged methylated N-terminal peptide (*m/z* for $[Met\text{-}HGFVQNIVIDGK]^{2+}$ = 670.29 Da). The second most intense peak

in the sample appeared at 16.08 min and had a mass of 663.86 Da, which corresponds to the double-charged non-methylated N-terminal peptide. Three minor peaks with retention times of 14.36, 16.95 and 18.65 min, and corresponding masses of 491.98, 565.56 and 594.24 Da, respectively, result from enzymes with incorrectly processed signal peptides, as their masses corresponds to double-charged N-terminal peptides that lack one to three N- terminal amino acids. Peak intensity values are presented relative to the intensity of the dominant peak (100 %).

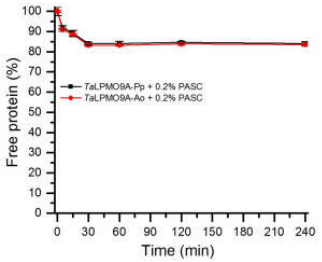

Figure S4. Binding of the *Ta*LPMO9A-Pp and *Ta*LPMO9A-Ao to PASC. The percentage of free LPMO was determined by measuring the reduction in LPMO concentration (A_{280}) over time. The experiments were carried using the same conditions as in LPMO reactions with substrates, but without the addition of a reductant.

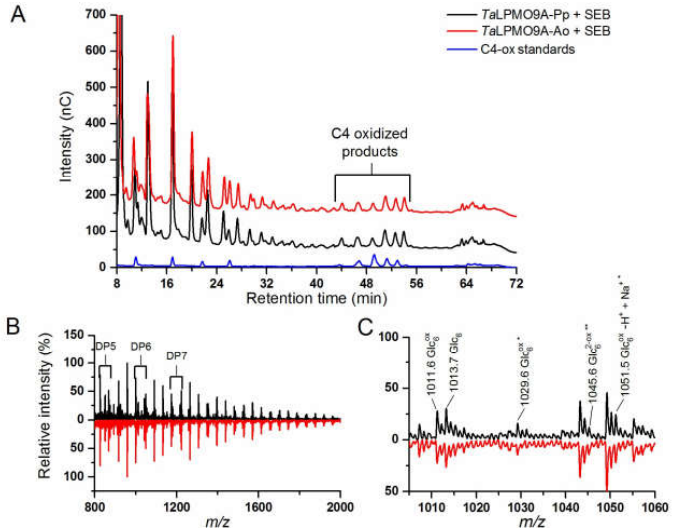

Figure S5. Reaction products generated by *Ta*LPMO9A-Pp and *Ta*LPMO9A-Ao from steam exploded birch (SEB). (A) HPAEC-PAD profiles of various reaction mixtures. Note that ascorbic acid was not added since it is well known that SEB contains sufficient reducing power to drive the LPMO reaction. When using this 75 minute gradient, native cello-oligomers elute between 10 and 26 min, while C4-oxidized products elute between 47 and 56 min as confirmed by analysis of a mixture of C4-oxidized cello-oligomers (solid blue line), obtained by oxidative action of strictly C4-oxidizing *Nc*LPMO9D (NCU01050^{1}) on PASC. It was not possible to accurately identify C1-oxidized products due to their lower abundance and the high complexity of the chromatograms in the region where they elute (30 – 50 min). (B) MALDI-ToF MS spectra of products released from SEB by the action of *Ta*LPMO9A-Pp (black line) or *Ta*LPMO9A-Ao (red line), after Na^+ saturation. Characteristic signal clusters belonging to products in the DP5 to DP7 range are indicated. (C) Close-up of MALDI-ToF MS spectra showing the DP6 product cluster generated by the action of *Ta*LPMO9A-Pp (black line) and *Ta*LPMO9A-Ao (red line). All labeled signals are sodium adducts of native and oxidized products; single oxidation or double oxidation is indicated by "ox" and "2-ox"and single or double hydration is indicated by "*" or "**". The overall signal pattern is characteristic for mixed C1/C4 oxidation.

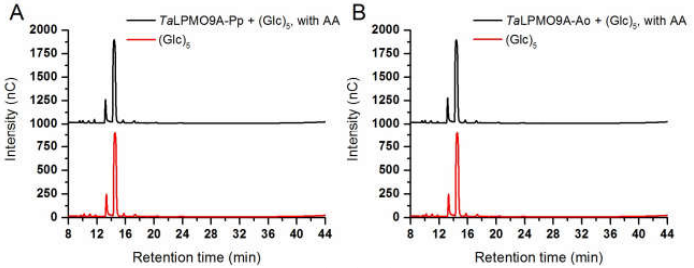

Figure S6. HPAEC-PAD chromatograms showing the outcome of reactions of *Ta*LPMO9A-Pp (A) and *Ta*LPMO9A-Ao (B) with cellopentaose, in the presence of ascorbic acid. Chromatograms for the reaction mixtures are black, while the chromatogram for untreated cellopentaose is shown in red.

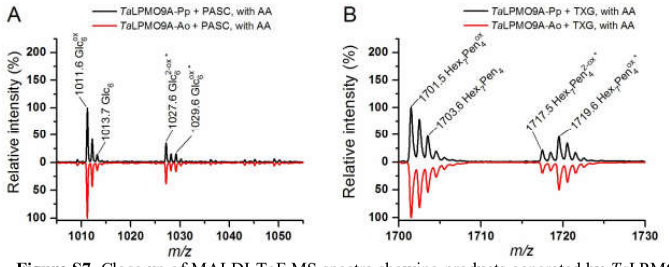

Figure S7. Close-up of MALDI-ToF MS spectra showing products generated by *Ta*LPMO9A-Pp (black line) and *Ta*LPMO9A-Ao (red line). (A) The DP 6 cluster from the reaction with PASC showing products characteristic for mixed C1/C4 oxidation (sodium adducts). The relatively high peak at 1011 (Glc_6^{ox}) indicates C4 oxidation, since the lactone formed by C1 oxidation, which has the same m/z, would be largely in its aldonic acid form (i.e. m/z = 1029 for the single sodium adduct) under these conditions. Furthermore, there is a signal for a hydrated double oxidized species at 1027. (B) The Hex_7Pen_4 cluster from the reaction with TXG, showing oxidized and, possibly, minor amounts of native products. In both spectra, hydration is indicated by "*", single oxidized products are indicated by "ox" and products which *m/z* values corresponding to masses of double-oxidized products are indicated by "2-ox".

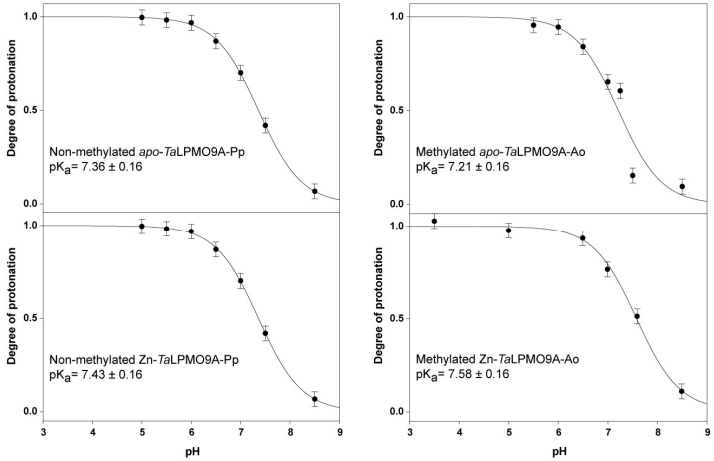

Figure S8. Determination of pK_a values of His1. The graphs show titration curves derived from ^{1}H chemical shift values at each pH point for LPMOs with a methylated (*Ta*LPMO9A-Pp) or a non-methylated (*Ta*LPMO9A-Ao) His1 either in the *apo* form or with Zn^{2+} present. The curves are the result of fitting Equation 1 to the data, and all pK_a values are indicated with their SD.

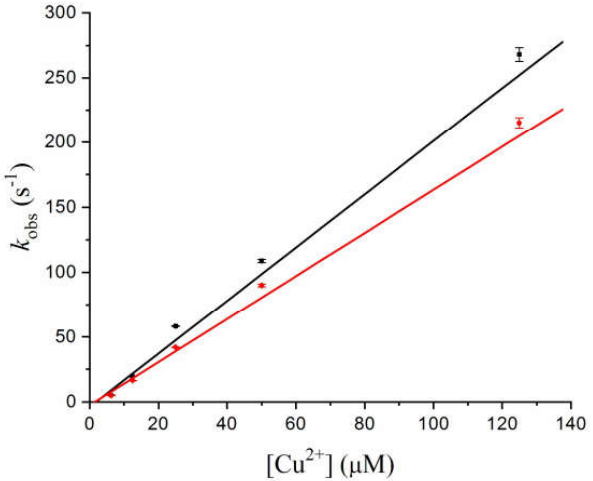

Figure S9. First order rate constants (k_{obs}) for binding of Cu^{2+} as a function of the copper concentration for *Ta*LPMO9A-Pp (●) and *Ta*LPMO9A-Ao (■). The points represent average values derived from four measurements ± the standard deviation.

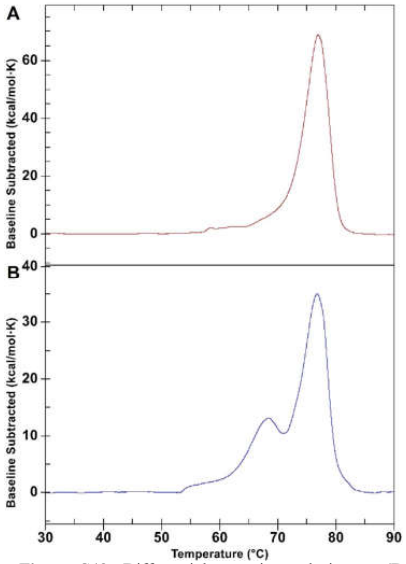

Figure S10. Differential scanning calorimetry (DSC) analysis. The panels show protein unfolding curves for (A) *Ta*LPMO9A-Ao (red), and (B) *Ta*LPMO9A-Pp (blue).

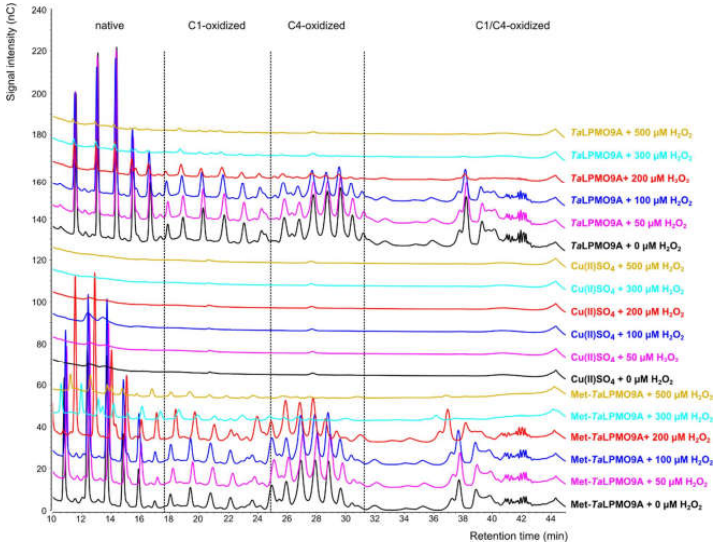

Figure S11. Control reactions to check for occurrence of copper-catalyzed Fenton-type chemistry. The Figure shows HPAEC-PAD product profiles obtained after incubation of 1 µM *Ta*LPMO9A-Ao ("Met-*Ta*LPMO9A"), 1 µM *Ta*LPMO9A-Pp ("*Ta*LPMO9A"), or 1 µM $Cu(II)SO_4$ with PASC in the presence of different initial concentrations of exogenous H_2O_2 (0 – 500 µM). All reactions were carried out in 40 mM BisTris buffer, pH 6.5, with 2.5 g/L PASC and 1 mM ascorbic acid, at 45 °C and 1000 rpm and the reaction time was 60 min.

[1] Vu, V. V., Beeson, W. T., Phillips, C. M., Cate, J. H., and Marletta, M. A. (2014) Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases, *J. Am. Chem. Soc. 136*, 562-565.