Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2023:41
Faculty of Natural Resources and Agricultural Sciences

Structure and function studies of GH45 glycoside hydrolases

Laura Okmane

Structure and function studies of GH45 glycoside hydrolases

Laura Okmane
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Sciences
Uppsala

SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Uppsala 2023

Acta Universitatis Agriculturae Sueciae
2023:41

ISSN 1652-6880
ISBN (print version) 978-91-8046-134-4
ISBN (electronic version) 978-91-8046-135-1
https://doi.org/10.54612/a.3f14k593rr

Swedish University of Agricultural Sciences, Department of Molecular Sciences, Uppsala, Sweden

Print: SLU Grafisk Service, Uppsala 2023

Structure and function studies of GH45 glycoside hydrolases

Abstract

The enzymatic hydrolysis of lignocellulosic material in nature is carried out by a plethora of cellulases. Glycoside hydrolase family 45 (GH45) enzymes are small cellulases most commonly found in fungi that catalyse the hydrolysis of β(1→4) linked glucans. Additionally, GH45 enzymes display a structural resemblance to non-hydrolytic protein groups such as expansins and loosenins.

In this thesis, the distinctness of GH45 enzymes from different subfamilies was explored. The enzymatic activity of GH45 enzymes from *Humicola insolens* (*Hi*Cel45A), *Mytilus edulis* (*Me*Cel45A), *Trichoderma reesei* (*Tr*Cel45A), *Phanerochaete chrysosporium* (*Pc*Cel45A), *Gloeophyllum trabeum* (*Gt*Cel45A) was demonstrated. Among the tested substrates were barley betaglucan, konjac glucomannan, carboxymethyl cellulose, and cellohexaose. Initial hydrolysis rates and hydrolysis yields were determined by reducing sugar assays, product formation was analysed using NMR spectroscopy and HPLC. The subfamily B and C enzymes exhibited mannanase activity, and the subfamily B enzyme *Me*Cel45A appeared cold adapted in comparison to *Tr*Cel45A. GH45 enzymes are known to act using an inverting action mechanism. This action mechanism had not been experimentally demonstrated in subfamilies B and C, however. Here experimental evidence is provided for the inverting nature of GH45 enzymes from all subfamilies.

Comparisons of GH45 enzyme structures were carried out and the first crystal structure of a GH45 subfamily C enzyme from the brown-rot fungus *G. trabeum* was reported at 1.3 Å resolution. Furthermore, structure and function investigations were done on an isotopically labelled Cel45A from the white-rot fungus *P. chrysosporium* using NMR spectroscopy. *Pc*Cel45A was expressed in *Pichia pastoris* with ^{13}C and ^{15}N labelling. A nearly complete assignment of ^{1}H, ^{13}C and ^{15}N backbone resonances was obtained and the interaction of ^{15}N-labelled *Pc*Cel45A with cellobiose was studied.

Keywords: Cel45A, cellulase, DPBB, endoglucanase, GH45, glycoside hydrolase

Struktur och funktion studier av GH45 glykosidhydrolaser

Abstract

Den enzymatiska hydrolysen av lignocellulosa i naturen utförs av en uppsjö av cellulaser. Glykosidhydrolaser från familj 45 (GH45) är små cellulaser, oftast påträffade i svampar, som katalyserar hydrolysen av β(1→4)-bundna glukaner. Dessutom uppvisar GH45-enzymer en strukturell likhet med icke-hydrolytiska proteiner såsom expansiner och looseniner.

I denna avhandling undersöktes särskiljande drag för GH45-enzymer från olika subfamiljer. Den enzymatiska aktiviteten av GH45-enzymer från *Humicola insolens* (*Hi*Cel45A), *Mytilus edulis* (*Me*Cel45A), *Trichoderma reesei* (*Tr*Cel45A), *Phanerochaete chrysosporium* (*Pc*Cel45A) och *Gloeophyllum trabeum* (*Gt*Cel45A) analyserades. Bland de testade substraten var betaglukan från korn, glukomannan, karboxymetylcellulosa och cellohexaos. Initiala hydrolyshastigheter och utbyten bestämdes genom analys av reducerande socker, produktbildning analyserades med hjälp av NMR-spektroskopi och HPLC. Enzymer tillhörande subfamilj B och C uppvisade aktivitet på mannankedjor, och *Me*Cel45A från subfamilj B tycktes vara anpassad för kallare miljö i jämförelse med *Tr*Cel45A från samma subfamilj. GH45-enzymer är kända för att hydrolysera substratet med hjälp av en inverterande reaktionsmekanism. Denna mekanism har tidigare inte påvisats experimentellt i subfamiljerna B och C, men här tillhandahålls experimentella bevis för den inverterande mekanismen hos GH45-enzymer från samtliga subfamiljer.

Strukturen av GH45-enzymer jämfördes och den första kristallstrukturen av ett subfamilj C GH45-enzym från brunrötesvampen *G. trabeum* rapporterades med 1,3 Å upplösning. Vidare gjordes struktur- och funktionsstudier på en isotopinmärkt Cel45A från vitrötesvampen *P. chrysosporium* med hjälp av NMR-spektroskopi. *Pc*Cel45A uttrycktes i *Pichia pastoris* med ^{13}C- och ^{15}N-inmärkning. En nästan fullständig tilldelning av ^{1}H, ^{13}C och ^{15}N från polypeptidkedjan erhölls och interaktionen mellan ^{15}N-inmärkt *Pc*Cel45A och cellobios undersöktes.

Nyckelord: Cel45A, cellulas, DPBB, endoglukanas, GH45, glykosidhydrolas

THE TRUTH IS ~~OUT~~ IN THERE… at the molecular level.

Dedication

To whomever cares about me or this research topic.

Contents

List of publications 9

List of tables 11

List of figures 13

Abbreviations 15

1. Introduction 19

2. Background 21
   - 2.1 An inventory of plant cell wall components 21
     - 2.1.1 The CW polysaccharides 22
     - 2.1.2 The CW proteins 23
     - 2.1.3 Phenolics 24
   - 2.2 Cellulose and hemicellulose 24
     - 2.2.1 Cellulose 24
     - 2.2.2 Hemicellulose 27
   - 2.3 Plant CW degrading organisms 29
     - 2.3.1 CW degrading enzymes: Carbohydrate-active enzymes 32
   - 2.4 Glycoside hydrolase family 45 35
     - 2.4.1 Subfamily A, B, and C 41
     - 2.4.2 Structurally similar proteins 43
   - 2.5 Methods background 46
     - 2.5.1 Heterologous protein expression and purification 46
     - 2.5.2 Biochemical characterization 51
     - 2.5.3 Protein X-ray diffraction crystallography 52
     - 2.5.4 Protein NMR spectroscopy 55

3. Aim of the study 59

4. Results and discussion 61
   - 4.1 Structural studies 62

4.1.1 Structure studies of GH45 enzymes (Paper I, II) 62
4.1.2 The structure of Cel45A from *G. trabeum* (Paper II) 66
4.2 Enzyme activity studies 67
4.2.1 Enzymatic activity divergence across GH45 (Paper I) ... 68
4.2.2 Dissimilarities within a subfamily (Paper I, II) 70
4.3 Additional studies on GH45 enzymes and homologous proteins 71

5. Conclusions and outlook 77

References 81

Popular science summary 95

Populärvetenskaplig sammanfattning 99

Acknowledgements 103

List of publications

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

I. Laura Okmane, Gustav Nestor, Emma Jakobsson, Bingze Xu, Kiyohiko Igarashi, Mats Sandgren, Gerard J. Kleywegt, Jerry Ståhlberg. (2022). Glucomannan and beta-glucan degradation by *Mytilus edulis* Cel45A: crystal structure and activity comparison with GH45 subfamily A, B and C. *Carbohydrate Polymers* 277, 118771.

II. Laura Okmane, Louise Fitkin, Mats Sandgren, Jerry Ståhlberg. The first crystal structure of a Family 45 glycoside hydrolase from a brown-rot fungus, Cel45A from *Gloeophyllum trabeum.* (manuscript)

III. Laura Okmane, Mats Sandgren, Jerry Ståhlberg, Gustav Nestor. ^{1}H, ^{13}C and ^{15}N backbone resonance assignment of Cel45A from *Phanerochaete chrysosporium* and its interaction with cellobiose. (manuscript)

Paper I is reproduced with the permission of the publisher.

The contribution of Laura Okmane to the papers included in this thesis was as follows:

I. Planned the work together with supervisors and performed most of the laboratory work, except for the crystallography part: produced and purified PcCel45A, prepared enzymes and substrates and measured activities by reducing sugar assay, and HPAEC. Did sequence alignment and structure comparison, analysed and interpreted results. Wrote manuscript together with the co-authors, made figures and tables, and submitted to the journal.

II. Planned the work with the co-authors. Performed most of the laboratory work: Expressed, purified and crystallised GtCel45A together with Louise Fitkin (master student under my supervision), and did activity measurements. Collected X-ray data, solved and refined the GtCel45A structure, and deposited in PDB. Analysed and interpreted results and made figures and tables. Wrote the manuscript together with the co-authors.

III. Performed the laboratory work: Produced and purified ^{13}C and ^{15}N isotope labelled PcCel45A, performed titration experiments. Together with G. Nestor, carried out the NMR experiments. Performed backbone resonance assignment. Took part in analysis and result interpretation. Wrote the manuscript together with the co-authors.

List of tables

Table 1. A few examples of one-domain and two-domain GH45 enzymes. *NCBI's conserved domain database..... 38

List of figures

Figure 1. Illustrative image of the plant CW organization. 21

Figure 2. Schematic representation of carbohydrates found in the plant CW. 28

Figure 3. Types of cellulase action on cellulose. 33

Figure 4. Types of active site shapes found in cellulases. 34

Figure 5. Inverting action mechanism in Cel45A from *H. insolens* (Davies et al., 1995). 37

Figure 6. Taxonomic distribution of GH45 enzymes, based on number of genera. 39

Figure 7. Phylogenetic tree of GH45 enzymes and homologous proteins. Depicted in bold are the names of the proteins investigated in this thesis. 44

Figure 8. (A) image of *Gt*Cel45A crystal in a loop directly before data collection; (B) diffraction pattern obtained from the *Gt*Cel45A crystal. 52

Figure 9. Surface views of GH45 endoglucanases from subfamily A, B and C. Arrows point to substrate binding area in *Hi*Cel45A (PDB: 4ENG), *Pc*Cel45A (PDB: 5KJO), *Me*Cel45A (PDB: 1WC2), and *Tr*Cel45A (GenBank: CAA83846.1, homology model). Catalytic acid and base are shown as sticks in the cartoon representation. Letters A, B, C indicate subfamily membership. 63

Figure 10. Surface views of GH45 endoglucanases from subfamily B, active site cleft heights indicated with dashed line. Height measurements shown in Å. (A) *Me*Cel45A (PDB ID 1WC2); (B) *Ac*Cel45A (PDB ID 5XBU); (C) *Tr*Cel45A (predicted structure)..... 64

Figure 11. GH45 subfamily A, B, C consensus sequences with depiction of secondary structure elements..... 65

Figure 12. Cartoon representation of the molecular structure of *Gt*Cel45A. 66

Figure 13. Graphical summary of enzymatic activities determined in Paper I. 68

Figure 14. Schematic representation of hydrolytic activity variance within subfamily B, as determined in Paper I..... 70

Figure 15. $^{1}H,^{15}N$-HSQC spectrum of PcCel45A. 72

Figure 16. Hydrolytic activity of *Pc*Cel45A on bacterial PASC and pulp, with and without subjection to mechanical stress. 74

Figure 17. LOOL12 crystals from crystallization trials..... 74

Figure 18. $^{1}H,^{15}N$-HSQC spectrum of *Tr*LOOL1. 75

Abbreviations

^{13}C	Carbon 13
^{15}N	Nitrogen 15
^{1}H	Proton
2D	Two-dimensional
3D	Three-dimensional
aa	Amino acid
AA	Auxiliary activity
AcCel45A	Cel45A from *A. crossean*
AFM	Atomic force microscopy
AOX	Alcohol oxidase
CAZy	Carbohydrate-active enzyme.
CBM1	Carbohydrate binding module 1
CBM63	Carbohydrate binding module 63
CD	Catalytic domain
CE	Carbohydrate esterase
Cel45A	GH45 cellulase
CMC	Carboxymethylcellulose
CW	Cell wall
D_2O	Deuterium oxide

Da	Dalton
DPBB	Double-psi beta barrel
EC	Enzyme commission number
FID	Free induction decay
GdCel45A	Cel45A from *G. dilepis*
GH	Glycoside hydrolase
GH45	Glycoside hydrolase family 45
GT	Glycosyl transferase
GtCel45A	Cel45A from *G. trabeum*
HGT	Horizontal gene transfer
HiCel45A	Cel45A from *H. insolens*
HPAEC	High performance anion exchange chromatography
Lac	Laccase
LiP	Lignin peroxidase
LOOL	Loosenin-like
MeCel45A	Cel45A from *M. edulis*
MLG	Mixed-linkage glucans
MnP	Manganese peroxidase
MUT	Methanol Utilization Pathway
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide gel electrophoresis
PASC	Phosphoric acid swollen cellulose
PcCel45A	Cel45A from *P. chrysosporium*
PCW	Primary cell wall
PDB	Protein data bank
PHBAH	p-hydroxybenzoic acid hydrazide
PL	Polysaccharide Lyase

RG	Rhamnogalacturonan
RlpA	Rare lipoprotein A
SCW	Secondary cell wall
TMS	Tetramethylsilane
TrCel45A	Cel45A from *T. reesei*
VP	Versatile peroxidase
XG	Xyloglucan
YNB	Yeast nitrogen base

1. Introduction

The advances of biotechnology are behind many of the luxuries we experience in our everyday life. From various food products (cheese, yoghurt, beer) and household products (liquid and powder detergents, paper tissues) to medicine (probiotics and vaccines). The main tools in biotechnology are proteins that catalyse different chemical reactions – *enzymes*. *Enzymes* are used in straightforward applications such as enzymatic hydrolysis of cellulosic material in paper and pulp industry, to more complex methods of molecular engineering, such as production of biological compounds, for example enzymes themselves (Bajpai, 1999; Demuner et al., 2011).

Catalytic proteins are not only used in generation of products, such as foods, but are also used to mitigate certain substances. For example, it is important for foods intended for human consumption to not contain high levels of a neurotoxin *acrylamide* (European Commission regulation (EU) 2017/2158). Unfortunately, acrylamide occurs naturally upon the heat-treatment of many plant-derived foods and therefore is found in processed food products. Food industry utilizes enzyme *asparginase* to reduce the acrylamide content in such foods (Palermo et al., 2016).

Taking advantage of enzymes such as *carbohydrate-active enzymes* (CAZys) is at the foundation of paper and pulp industry, which make paper, carton, tissues and paper/pulp by-products possible (Basak et al., 2021). Utilizing cellulase mixtures is the most commonly chosen strategy when high substrate specificity and hydrolytic potential is of essence (Hu et al., 2013). Not using single enzymes, but instead enzymes mixtures is especially common when it comes to production of household detergents. *Amylases*, *cellulases*, *lipases*, *mannanases*, and *proteases* are the types of enzymes found in household detergents.

Amylases, cellulases and mannanases take care of carbohydrate-related (in other words, sugar) stains, while lipases and proteases deal with fat and protein-derived stains respectively.

What seems counter-intuitive at first, but appears sensible after further reading, is the use of carbohydrate active enzymes in laundry detergents. Considering that clothing is most often made out of cotton and cotton is a carbohydrate, wouldn't we want to *avoid* subjecting textiles to cellulases? Fortunately, the detergent industry recognizes the diversity of enzymes. Not only are enzymes diverse regarding their substrate specificity (for example, proteases catalyse the digestion of other proteins), but also in their molecular structure and enzymatic capabilities or *enzyme activity*. Many of the enzymes used in achieving the colour-protective effects of laundry detergents are cellulases with low enzymatic activity, meaning that the peak catalytic activity is relatively low and is reached under certain conditions. Most of these cellulases are derived from fungi (Arja, 2007). Especially common cellulases in colour-protective washing powders belong to *Glycoside Hydrolase Family 45* (GH45). These enzymes are good at defibrillating textiles, while having rather high resistance to oxidizing agents and anionic surfactants (Shimonaka et al., 2006).

GH45 enzymes catalyse the hydrolysis of cellulosic substrates, and are widely used, yet research on the structure and the molecular action mechanism of these proteins is rather scarce. In nature, GH45 enzymes are extracellularly secreted by wood-degrading fungi and also bivalves. The exact purpose of these low-activity enzymes continues to be under investigation. Interestingly, some members of GH45 family share a remarkable structural resemblance to hydrolytically inactive plant proteins called *expansins* and homologous fungal proteins *loosenins*. While plant expansins are crucial in major plant developmental processes, the current opinion is that wood decaying fungi release GH45 enzymes mainly to degrade substrate. The role of loosenins remains unknown to this date.

The aim of this thesis is to shed some additional light on to GH45 enzymes and the diversity within the family.

2. Background

2.1 An inventory of plant cell wall components

Considering that plants are unable to relocate themselves to more resource rich areas, through evolution plants had to develop mechanisms to survive and thrive in their local environment. To survive predation, harsh climate conditions and competition for resources, plants have adapted at the molecular level. Cells in each tissue are surrounded by cell walls, which are composed of intertwined polysaccharides (Figure 1). This cell architecture strengthens the cell against physical and chemical factors, such as environmental stresses and invading pathogens. It also permits plants to "defy gravity" by growing upwards (negative gravitropism), reach impressive height, take various shapes, in addition to that, it aids in water retention.

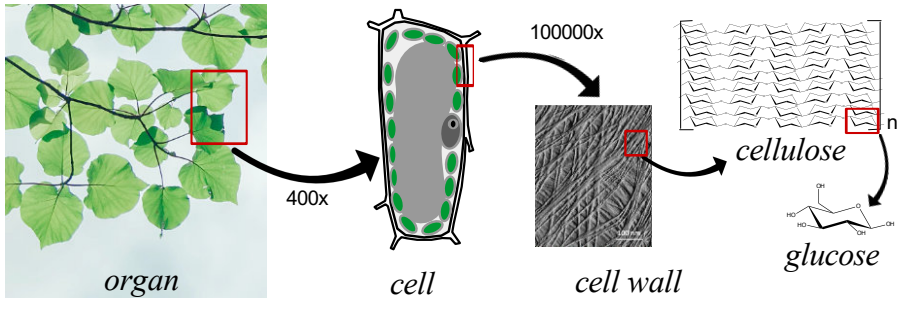


Figure 1. Illustrative image of the plant CW organization.

Plant cell walls (CWs) are traditionally divided into two types, the primary cell wall (PCW) and the secondary cell wall (SCW). The PCW has dynamic properties and forms during division in early plant

development. Its main purpose is to provide structural support and flexibility under cell's development. The PCW is plastic and elastic, it shapes cell's morphology and determines the extension of the cell. Once the cell stops dividing and growing, a thicker, more robust layer develops between the plasma membrane and the PCW. This new, rigid layer is the SCW and compositionally varies from the PCW. The main roles of the SCW are to improve the plants resistance to external stimuli and to allow vertical growth.

The structure of CWs varies within a plant (between tissue types) and between species, but the main components of all plant CWs are: *polysaccharides, proteins, and phenolic compounds.*

Polysaccharides and proteins conduct the functionality of the CW. Proteins modify and customize the CW, while polysaccharides provide CW integrity.

2.1.1 The CW polysaccharides

The architecture of the plant CW has to withstand the tensile forces from within the cell generated by turgor pressure during cell growth. The CW polysaccharides are the basis for CW rigidity.
The CW polysaccharides can be grouped into three major categories: *cellulose, hemicellulose, pectin.*

Cellulose is a homogenous polymer, a chain of D-glucose units linked together with β(1→4) linkages. Examples of *hemicelluloses* are mannans, glucomannans, xylans, xyloglucans, mixed-linkage glucans. Hemicelluloses are heterogenous linear or branched polymers, derived from L-arabinose, D-galactose, D-glucose, D-mannose and D-xylose. *Pectins* are highly heterogenous polymers, which are divided in to four major groups: homogalacturonans, rhamnogalacturonans (RG-I and RG-II), and xylogalacturonans. Most pectins have a chain of α(1→4) linked galacturonic acid (GalA) as the backbone and mixed, highly complex sidechains (Xiao & Anderson, 2013).

The overall structure of plant CWs are cellulose fibres embedded in a non-cellulosic matrix. However, the types and proportion of CW polysaccharides make grasses and related species unique among other land plants.

In most land plants xyloglucan (XG) and pectins dominate the plant CW matrix. However, in grasses, such as barley, cellulose and xylan

are dominant CW constituents. Mixed-linkage glucan (MLG) is absent in the PCWs of most angiosperms, yet it is in significant proportion in the PCWs of grasses. Glucomannans and mannans constitute minimum 5% of the dicotic cell wall, and are of minor abundance in grasses (Vogel, 2008). Dicots appear to have higher proportion of structural proteins in their primary wall as well.

The main structural elements of the SCW in plants are cellulose, xylan and lignin. The proportion of these structural elements slightly varies between monocots and dicots on average.

Different proportions of SCW elements can be observed between angiosperm species (hardwood) and gymnosperm species (softwood) as well. For example, wood from *Acer rubrum* (Red maple tree) consists of 45% cellulose, 24% lignin, 29% hemicellulose (25% glucurono-xylan, 4% glucomannan), 2% pectin/starch/ash, whereas wood from *Thuja occidentalis* (Northern white-cedar tree) contains 41% cellulose, 31% lignin, 26% hemicellulose (14% arabinoglucurono-xylan, 12% galactoglucomannan), 2% pectin/starch/ash (Timell, 1967).

2.1.2 The CW proteins

Despite not constituting majority of CW, there is a vast diversity of proteins in the plant CW. In 2014 the researchers from The University of Toulouse, France (Albenne et al., 2014) summarized 53 proteomics studies on the PCW proteins, pointing to the vast amounts of different proteins that have been identified in the proteomes of various plant species CWs, for example, the CWs of *Arabidopsis thaliana* contain at least 900 different proteins and the CWs of the grass species *Brachypodium distachyon* contain at least 600 different proteins. Main duties of CW proteins are to contribute to the integrity of the CW, modify other CW components and partake in cell signalling (Ferrari et al., 2013; Murphy et al., 2012).

Some types of proteins found in the plant CWs are: *various carbohydrate interacting proteins*, such as glycoside hydrolases, expansins, lectins, carbohydrate esterases, carbohydrate lyases, oxido-reductases including peroxidases, multicopper oxidases; *protein interacting proteins* such as proteases and enzyme inhibitors; *lipid interacting proteins* such as lipid transfer proteins, lipases; *proteins*

associated with signalling such as arabinogalactan proteins; *structural proteins*.

2.1.3 Phenolics

Two forms of phenolic compounds are present in plants – as a part of a phenolic polymer or as smaller phenolic compounds bound to other cell wall components. Phenolic compounds such as ferulic and coumaric acids are found in the PCW, while phenolic polymers such as lignin are part of the SCW (Bonoli et al., 2004; Lygin et al., 2011). The CW bound phenolics are cross-linked to matrix polysaccharides (Mnich et al., 2020) and undergo changes upon subjection to stress (Gupta & De, 2017).

Complex phenolic polymers, such as lignin, are of minor abundance in PCW but of high proportion in the SCW. Lignin is a hydrophobic heteropolymer and serves as dehydrated layer at the SCW compartment. Lignin – cellulose composite (lignocellulose) increases the strength and reduces the flexibility of CWs. Generally, lignocellulose allows plants to create tube-like structures, which are used as water channels and for structural support. The polymer is a mixture of three constituents: p-coumaryl, sinapyl, and coniferyl alcohols. The proportion of these monomers is species dependent.

2.2 Cellulose and hemicellulose

2.2.1 Cellulose

Cellulose is a linear homopolymer, a polysaccharide composed of β(1→4) linked glucose residues. It is present in all plant cell walls in a significant proportion (Timell, 1967). Glucose units in cellulose are organized in cellulose chains, which are then packed in to crystalline microfibrils. In 2020, a group at Department of Plant Biology, Michigan State University, USA used atomic force microscopy (AFM) for direct imaging of elongated PCWs and SCWs from fresh tissues of maize *Zea mays* (B. Song et al., 2020). Cellulose structure was analysed under near-native conditions and the individual microfibrils in elongated PCWs were measured. They affirmed the recent discovery that a single cellulose microfibril is composed of 18

cellulose chains (Kubicki et al., 2018), which are possibly packed in six sheets with 234432 conformation. All individual microfibrils appeared to originate from a microfibril bundle which made the authors suggest that, in PCWs, cellulose fibrils are synthesized in larger bundles, which then split during cell expansion or elongation. A different organization was observed in SCWs. Microfibrils were nearly-parallel twined or individual and appeared to be embedded separately in the matrix. No splitting effect was observed, pointing at differences in the cellulose biosynthesis mechanisms of PCW and SCW.

Despite homogeneity at the unit-level, native cellulose can exist in multiple forms. Multiple solid-state NMR studies carried out towards the end of 20th century, revealed two crystalline forms of native cellulose – Iα and Iβ (Atalla & VanderHart, 1984). In early 2000s, crystallographers at The University of Tokyo managed to solve the structure of Iα and Iβ cellulose isolated from the algae *Glaucocystis nostochinearum* and the animal *Halocynthia roretzi* (sea pineapple) respectively (Nishiyama et al., 2002, 2003). Cellulose isolated from bacteria and algae was mostly Iα, while higher plants such as cotton and ramie mostly contained Iβ cellulose. However, recent re-investigations show that cellulose from PCW of monocots and dicots is polymorphic and often significantly different from crystalline Iα and Iβ allomorphs. By coupling multiple NMR methods, at least five types of different ^{13}C chemical shifts for glucose units have been identified in cellulose microfibrils across the PCWs of variety of plants, such as grasses and Arabidopsis (T. Wang et al., 2016). The Iα and Iβ allomorphs continue to be identified in cotton, suggesting that these structures are mostly present in plants with highly crystalline cellulose (Kirui et al., 2019).

The organisation of cellulose microfibrils within the plant CW is an open investigation, but multiple studies done in the past decade have elucidated polysaccharide interactions within the plant CW using modern methods. Atomic force microscopy (AFM) studies on the PCW of onion epidermal cells have shown that cellulose microfibrils are directly connected and form a laterally bundled network(Tian Zhang et al., 2016), which is not in agreement with the traditional PCW depiction where microfibrils appear mostly separated by the matrix. Solid-state NMR studies reveal that 25 – 50 % of all cellulose

microfibrils in plant CWs are in close contact with pectins (T. Wang et al., 2012), and close contacts between cellulose and hemicellulose chains are in much lesser proportion (Bootten et al., 2004, 2009; Pérez García et al., 2011). These findings play a significant role in understanding the action mechanisms of proteins involved in the loosening and softening of the PCW, such as expansins and loosenins.

Relevance in biotechnology

Cellulose is not only the most abundant naturally available polymer, due to being a major constituent of plant CWs, but also an energy rich material, which makes it an attractive target in biotechnology. In past decades, cellulose as part of lignocellulosic biomass, has been one of the main ingredients in ethanol production.

Despite microbial ethanol production being a rather well-known trade, never has bioethanol as a biofuel been such a hot topic as it is today. Factors such as environmental policies and the skyrocketing gasoline prices has made bioethanol a desirable alternative to fossil fuels in 2023.

To mediate the anthropogenic contribution to the global warming by lowering CO_2 emissions, many countries around the world are issuing environmental policies that aim to increase the dependence on renewable energy, ideally achieving a complete phase-out of fossil fuels. For example, The European Union, through The European Green Deal, aims to make the EU carbon neutral by 2050. This heavily increases the demand on biofuels and brings forward the question about renewable resources for biofuel production. Biofuel production thus far has been divided into three categories, depending on the raw materials. *First-generation biofuels* are produced from starch crops and sucrose, for example cereals and sugarcane. *The second-generation* biofuels are produced from lignocellulosic biomass, such as lignocellulosic agricultural waste. *The third-generation* biofuel production utilises microalgae.

Majority of the ethanol commercially available today is the *first-generation* ethanol produced from sugarcane and corn (Hoang & Nghiem, 2021). Despite lignocellulosic ethanol being around already in 1920s (Eklöf et al., 2012), lignocellulosic biomass continues to remain as only a *potential* feedstock. The bottlenecks of *second-generation* ethanol production are not related to the feedstock supply

(Lin & Tanaka, 2006), but rather the technical and economic aspects. Feedstock and transportation costs compete with the well-established and efficient supply chains developed for the production of *first-generation* biofuels (Cheng & Timilsina, 2011). The major technical challenges involve the recalcitrance of the lignocellulosic biomass (Taherzadeh & Karimi, 2008). The treatment of lignocellulosic biomass involves physical, chemical and enzymatic treatment, the costs of which have to compete with *first-generation* biofuel production.

2.2.2 Hemicellulose

Similarly to cellulose, hemicellulose aids in strengthening the CW to withstand the mechanical and chemical forces induced upon the cell. In 2011, a group from Iowa State University demonstrated that hemicelluloses play a role in stiffening the CW, as cellulose chains in XG deficient CWs exhibited higher mobility both at the unit and saccharide chain level (Pérez García et al., 2011).

The term hemicellulose was introduced in 1891 by Ernst Schulze to distinguish cellulose from other CW polysaccharides. At that time, cellulose was described as a CW polysaccharide that was hydrolysable only by strong acids and upon hydrolysis almost exclusively yielded glucose as product. Hemicellulose was the other cell wall component, which could be hydrolysed by weak mineral acids and gave products such as galactose, mannose, arabinose, xylose (Murneek, 1929). Consequently, the term "hemicellulose" describes a heterogenous polysaccharide group. Hemicellulose polysaccharides have β(1→4) linkages in the backbone, but some hemicelluloses such as mixed-linkage glucans (MLGs), have an additional β(1→3) linkage in the backbone.

Mannan and glucomannan

In mannans and glucomannans the backbone consists predominantly of β(1→4) linked mannose residues and both hemicelluloses are often acetylated. Mannans can be divided in to mannans and galactomannans. In mannans and galactomannans, the backbone consists entirely of mannose residues. In galactomannans, mannose

backbone chain residues are substituted with galactose units by an α(1→6) linkage (Figure 2).

Glucomannans can be divided in to glucomannans and galactoglucomannans. Glucomannans and galactoglucomannans are heteropolymers with both mannose and glucose residues in the backbone. In galactoglucomannans the main chain mannose and occasionally glucose units are substituted by galactose residues through an α(1→6) linkage.

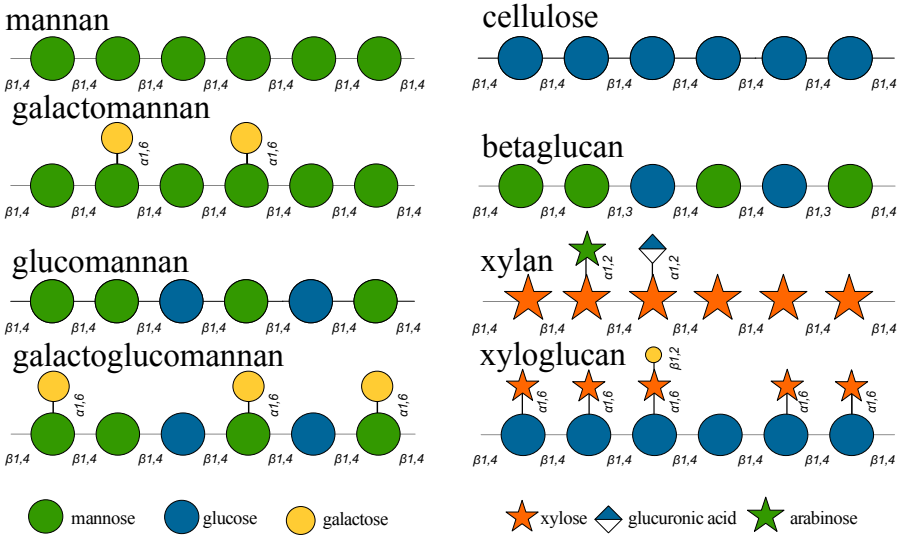


Figure 2. Schematic representation of carbohydrates found in the plant CW.

Xylan

Xylan composition varies between species, but the backbone is a β(1→4) linked branched xylose residue chain. Xylose units are substituted in varying degrees with arabinose and glucuronic acid. Depending on species, xylan may contain galacturonic acid and rhamnose residues (Rennie & Scheller, 2014).

Xyloglucan

Xyloglucan (XG) is a polysaccharide consisting of β(1→4) linked glucose residues, most of which carry xylose substituents through an α(1→6) linkage. A third of xylose residues are furthermore linked through β(1→2) linkage to an acetylated galactose. The galactose

residue is often α(1→2) linked to a fucose unit, thus expanding the level of branching.

Mixed-linkage glucan (MLG)

MLG, often referred to as betaglucan (BG), is an unsubstituted homopolymer consisting entirely of β(1→4) linked glucose oligomers which are then linked through β(1→3) linkages. Glucose oligomer length varies between grass species, but in the majority of MLGs the repeating unit is either cellotriose or cellotetraose (Buckeridge et al., 2004).

2.3 Plant CW degrading organisms

Plant CWs have evolved to resist biotic and abiotic factors. The building blocks of CWs, such as crystalline cellulose, lignin and hemicelluloses, require specialised conditions to be broken down. Despite the robustness of plant CWs, some organisms have evolved to produce enzymes which are capable of degrading plant material. As per the intrinsically egocentric nature of biological organisms, this is to ensure the plant-degrading species survival and propagation. Complex organic matter degradation also leads to finite resource such as carbon recycling, thus ensuring sustainability of life.

In biology, organisms that require the carbon to be uptaken from organic compounds are called *chemoorganoheterotrophs*. These organisms utilize carbohydrates, proteins and lipids to provide themselves with carbon and nutrients. Organisms that break down dead or decaying organic matter are called *decomposers* or *detritivores*. *Detritivores* ingest biomass prior to degradation, while *decomposers* secrete digestive enzymes in the environment to later absorb nutrients from their surroundings.

Invertebrates, such as earthworms (common earthworm *Lumbricus terrestris*), termites and mussels (the blue mussel *Mytilus edulis*), are some examples of detritivores. Fungi, such as mushrooms (*Pleurotus ostreatus*) and molds (*Penicillium sp.*), and soil bacteria (*Streptomyces sp.*) are examples of decomposers. The process that involves nutrient uptake from dead plant matter through extracellular digestion is called *saprotrophic digestion* and organisms employing such feeding mechanism are regarded to as *saprotrophs*. Saprotrophic fungi are the

main decomposers of organic matter in nature (McGuire & Treseder, 2010).

Plant organic matter is found in terrestrial and aquatic environments (Bar-On et al., 2018). Natural waste such as dead foliage, decaying wood, decaying plant matter (soil litter) and anthropogenic waste such as agricultural waste (crop residues, manure, forestry and energy crop residues) are decomposed by a variety of fungal and bacterial species.

Wood-decay fungi

Wood-decay fungi are any fungi which cause the rotting of wood. Both living or dead wood material can be affected by these species. Parasitic wood-decay fungi attack living trees, such as the pathogenic white-rot fungus *Heterobasidion parviporum* which lives on spruce *Picea abies* (Oliva et al., 2011). Saprotrophic species attack dead wood, for example, the ceramic fungus *Xylobolus frustulatus* which grows on decaying oak. Wood-decay fungi are categorized based on the visual consequence of wood decomposition. *Brown-rot* fungi, for example *Gloeophyllum trabeum*, fully degrade cellulose but cannot significantly degrade lignin, therefore the resulting wood residue, containing mostly lignin, is brown in colour (Goodell et al., 2008). *White rot* fungi, such as *Phanerochaete chrysosporium*, are capable of degrading both cellulose and lignin, the visual consequence of which is a bleached woody material (Blanchette, 1991).

White rot and brown rot are commonly caused by *macromycetes* – fungi that form fruiting bodies (sporocarps, mushrooms) visible to the naked eye.

Soft rot, however, is most often caused by *micromycetes* – fungi that produce spores in microscopic structures. Examples of micromycetes are moulds. Soft rot usually occurs in moist conditions on old timber and, generally, visual softening of wood material is apparent but such characteristic is not a requirement for a mould to be considered a soft rot fungus (Levy, 1966). Soft rot in wood visually resembles brown rot as soft rot fungi degrade mainly cellulosic parts of wood.

Soil microorganisms

Dead plant biomass, such as leaf litter, accumulates in soils, where it gets decomposed by soil microorganisms. Most of the forest litter

consist of cellulose, hemicellulose, lignin and pectin, consequently many soil residing microorganisms produce cellulose, hemicellulose, lignin or pectin degrading enzymes (Šnajdr et al., 2011). Soil microbial biomass constitutes mostly bacteria, archaea, and up to 30% soil fungi (B. Wang & Qiu, 2006), which cause mineral weathering and litter degradation, that in turn promotes plant growth by increasing the bioavailability of nutrients. Soil microorganisms have evolutionary become responsible parties for nutrient and mineral recycling (Falkowski et al., 2008; Vieira et al., 2020). Most plant species benefit from microorganisms residing in the *rhizosphere* – zone in the soil with close proximity to plant roots. Plant-soil microorganism interactions are complex and of various kinds. Many species are *symbiotic* – they depend on others to achieve the best growth conditions. Symbiosis can be mutually beneficial (*mutualistic*), beneficial to one organism, but harmless to the other (*commensalistic*) or beneficial to one and detrimental to the other (*parasitic*). Fungi residing the rhizosphere and having symbiotic association with plants are referred to as *mycorrhiza*. Rhizosphere's bacteria species are called *rhizobacteria*.

Plant litter decomposition requires complex degradation mechanisms involving a set of cellulolytic enzymes. Numerous studies have shown that the majority of soil bacteria have cellulase encoding genes (Eichorst & Kuske, 2012; Himmel et al., 2010; Yang et al., 2014), yet studies with Actinobacteria are indicative that the number of genes does not necessarily reflect a cellulolytic potency (Berlemont & Martiny, 2015; Větrovský et al., 2014). However, many soft rot fungi which can be found in the rhizosphere, such as *Aspergillus sp.*, *Trichoderma sp*., *Penicillium sp*. (Workneh & van Bruggen, 1994), are well known to secrete cellulases with high catalytic activity.

Agricultural waste and compost decomposers

Agricultural waste includes livestock manure, post-harvest agricultural waste and agro-industrial residues. A vast number of organisms is involved in the decomposition of agricultural waste. Organisms such as nematodes, micro-arthropods (spiders, mites, springtails, millipedes and more), fungi and bacteria all contribute to the complete degradation of agricultural residues. The degradation process involves aerobic and anaerobic decomposition, diverse

microorganisms with hydrolytic, acidogenic (*Micrococcus sp., Streptococcus sp.,* and others), methanogenic (*Methanobrevibacter sp*.) capabilities, thermophilic compost degraders such as thermophilic bacterium *Geobacillus sp*. and the thermophilic fungus *Humicola insolens*.

The extent to which any species is able to degrade plant CWs is dependent on the enzymatic machinery it employs. Wood and woody agricultural material, such as straw, consists mostly of SCW, where lignocellulose is the main component. Enzymatic degradation of lignocellulosic biomass involves lignin-modifying oxidoreductases such as laccase (Lac), manganese peroxidase (MnP), versatile peroxidase (VP) and lignin peroxidase (LiP), hemicellulose and cellulose degrading enzymes (hemicellulases and cellulases respectively). It is important to note that enzymes work in a synergistic manner. Different enzymes cleave CW components at different locations in a different manner. Cooperation has the potential to achieve complete biomass degradation. All of these enzymes are classified under *carbohydrate active enzymes* (CAZys).

2.3.1 CW degrading enzymes: Carbohydrate-active enzymes

As the building blocks of plant CWs are polysaccharides, plant CW degrading enzymes fall under CAZys.

The CAZy classification is sequence-based and follows the perception that protein sequence defines structure and structure defines function. It currently comprises five families which consist of enzymes that build or cleave carbohydrates, or modify lignin.

GlycosylTransferases (GTs) catalyse the formation of a glycosidic linkage.

Glycoside Hydrolases (GHs) catalyse the hydrolysis of a glycosidic linkage.

Polysaccharide Lyases (PLs) cleave carbohydrates through non-hydrolytic means.

Carbohydrate Esterases (CEs) remove acetylation from carbohydrates.

Auxiliary Activity (AA) enzymes include a variety of lignin-modifying enzymes, oxidases and reductases.

Glycoside Hydrolases

Glycoside Hydrolases (GHs), often referred to as glycosyl hydrolases, GH enzymes or glycosidases, are enzymes that catalyse the hydrolysis of glycosidic linkages. Reaction product profile is dependent on the type of GH enzyme and the interacting saccharide chain (substrate), products can be monosaccharides, oligosaccharides, polysaccharides.

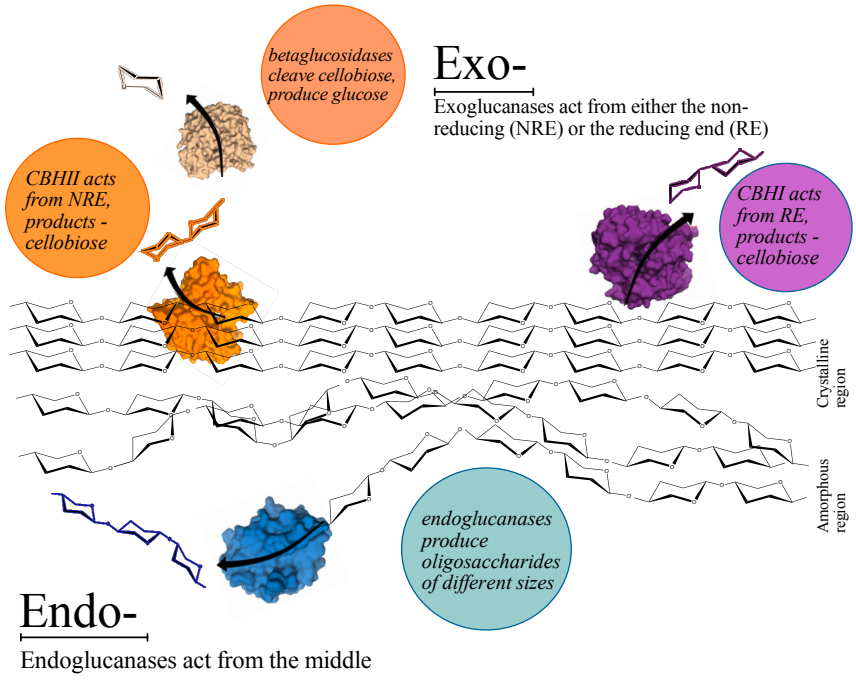


Figure 3. Types of cellulase action on cellulose.

GH enzymes are classified based on multiple characteristics. The cleavage site on the carbohydrate chain determines whether a GH is classified as *exo*-acting or *endo*-acting (Figure 3). *Endo*- glycosidases randomly hydrolyse internal glycosidic linkages of the polysaccharide and yield saccharide chains of various lengths. *Exo*- glycosidases cleave from the terminal ends of the polysaccharide chain, either from the reducing (RE) or non-reducing end (NRE), and mostly yield disaccharides as a product. Certain *exo*-acting GHs named beta glucosidases cleave cellobiose, yielding glucose as product.

The cleavage pattern appears to be protein structure dependant (Payne et al., 2015). *Endo*-acting GHs tend to have a rather open substrate binding groove or cleft, while in *exo*-acting GHs substrates bind in a more closed, tunnel-like structure. Beta glucosidases have a pocket-like crater structure at their active site (Figure 4).

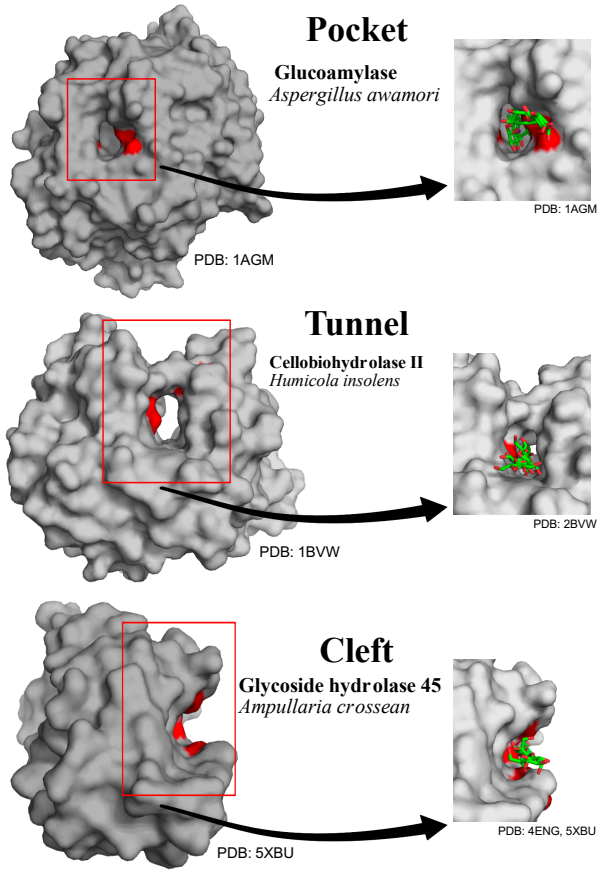


Figure 4. Types of active site shapes found in cellulases.

Another classification system for enzymes is based on the chemical reactions enzymes catalyse. Enzymes are given an enzyme commission (EC) number. GH enzymes are categorized as EC 3.2.1, which entails 50 types of reactions, for example, cellulases are EC 3.2.1.4 and the reaction in enzyme database (enzyme-database.org) is

listed as “Endohydrolysis of (1→4)-β-D-glucosidic linkages in cellulose, lichenin and cereal β-D-glucans”.

The reaction mechanism employed during hydrolysis divides GH enzymes into two groups: *retaining* glycoside hydrolases and *inverting* glycoside hydrolases (Davies & Henrissat, 1995). The name of these mechanisms is in regards to the anomeric configuration of the reaction product. Retaining enzymes retain the configuration by the anomeric carbon, yielding beta anomers from beta-linked substrates. After reaction with inverting enzymes, however, cleavage products have an inverted configuration of the anomeric carbon, yielding alpha anomers from beta-linked substrates.

To this date, sequence-based classification has divided GH enzymes into 180 families (GH1 – GH180). GH family members tend to align in employed reaction mechanism (retaining/inverting), catalytic residues, as well as 3D fold.

2.4 Glycoside hydrolase family 45

GH45 family members are *endo*-acting (endoglucanases) enzymes that are classified as cellulases (EC 3.2.1.4), *endo*-xyloglucanases (EC 3.2.1.151) and endo-β-1,4-mannanases (EC 3.2.1.78). GH45 enzymes are inverting enzymes with an aspartic acid residue acting as a catalytic base at the active site and a double-psi-beta barrel (DPBB) fold. GH45 family is formerly known as cellulase family K and as structurally related to non-hydrolytic protein groups expansins, loosenins, and swollenins. According to the CAZy database, more than 500 sequences are available, 90 GH45 members have been characterized and 9 molecular structures have been published to this date. GH45 enzymes are found in Bacteria, as well as Eukaryota. Amongst Eukaryota, GH45 genes have been identified in kingdoms Animalia, Fungi, and Plantae.

Latest research associates GH45 enzymes with pathogenic pathways, invoking plant immune response, but possible involvement in fungal cell developmental processes cannot be disregarded, as an upregulation of a GH45 enzyme has promoted fungal cell proliferation in *Ustilago esculenta* (X. Guo et al., 2022; Zhang et al., 2022).

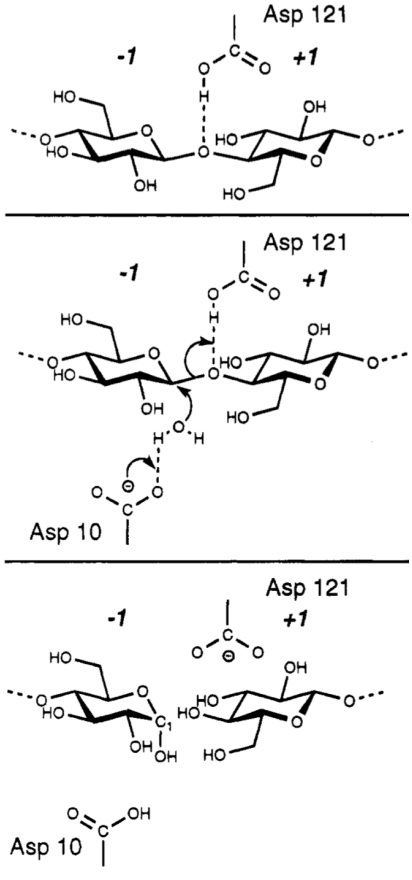
Molecular structure and catalytic mechanism

The building blocks of proteins are amino acids, which then form polypeptides. Polypeptides do not stay as linear chains of amino acids, however. Due to backbone atom interactions, polypeptides are pulled into forming shapes – the secondary structures. Secondary structures are of two conformations: an alpha helix and a beta strand, both of which are made by hydrogen bonding between the backbone carbonyl oxygen (C=O) of one amino acid and the backbone amino hydrogen (N-H) of another amino acid. An alpha helix occurs when carbonyl O forms a hydrogen bond with an amino (N)H which is four residues downstream. This helical structure has 3,6 amino acids in one turn. In a beta strand conformation multiple segments of the polypeptide chain align in a zig-zag pattern. Multiple beta strands can associate by hydrogen bonding forming a beta-sheet. Beta-strands in a beta-sheet can be parallel (peptide N-termini and C-termini align) or antiparallel (a peptides N-terminus aligns with another's C-terminus). Amino acid sidechains determine proteins tertiary structure. Tertiary structures are made by assembling secondary structures through hydrophobic interactions, covalent and a variety of non-covalent bonding in 3D space, for example, sidechains with opposite charges form ionic bonds, hydrophobic interactions group amino acids with non-polar sidechains towards the inside of a protein, disulphide bonds formed by cysteine sidechains fasten the protein structure.

GH45 enzymes have a double-psi beta barrel (DPBB) fold, which contains a six-stranded beta-barrel at its core, often with both types of beta-strand organization (parallel and antiparallel), short alpha helices and long surrounding loops. The substrate binding site is a cleft or an open groove which spans across the surface.

Substrate profile is subfamily-dependent, but GH45 enzyme family is able to hydrolyze β(1→4) linkages in cellulose, hemicellulose, such as MLG, XG, mannan, glucomannan, and, as one study shows, possibly even pectin (Bharadwaj et al., 2020; Davies et al., 1995; Godoy et al., 2018; Karlsson et al., 2002; Zhou et al., 2017). As endoglucanases, GH45 enzymes cleave internal bonds in polysaccharide chains. Cleavage product profile varies between subfamilies, likely due to variance in the size and shape of substrate binding area.

GH45 enzymes catalyse the hydrolysis of β(1→4) linkages in soluble betaglucans via an inverting action mechanism (Figure 5). Two amino acids act as a catalytic acid and a catalytic base in this mechanism. The catalytic acid protonates the glycosidic oxygen and catalytic base activates a water molecule, which then hydrolyses the glycosidic bond.



Cellulose binds in active site. Enzymatic cleavage will take place between subsites -1 and +1. Asp 121 acts as catalytic acid and protonates the glycosidic oxygen.

Asp 10 acts as catalytic base. A solvent water molecule is activated for nucleophilic attack at the C1 position. Presumably , this occurs via a transition-state with an elongated glycosidic bond.

Catalysis facilitated with inversion of configuration at C1. Aspartates 10 and 121 now have protonation states opposite to that before catalysis.

Figure 5. Inverting action mechanism in Cel45A from *H. insolens* (Davies et al., 1995).

The configuration at the anomeric carbon becomes inverted, thus yielding alpha-anomers from β(1 → 4) linked glucans. The catalytic acid residue is an aspartic acid, and it is conserved in all GH45 enzymes. The catalytic base residue is conserved within a subfamily and remains an open question in some GH45 enzymes.

Peculiarities of GH45s

Few characteristics make family 45 enzymes distinctively different from other glycoside hydrolases.

The GHs often exhibit a modular organization consisting of one or more domains. However, majority of GH45 enzymes are single domain proteins, few are bimodular containing an additional carbohydrate binding module (CBM1) attached with a flexible linker to the catalytic domain. Highest proportion of two-domain architecture is found in GH45 subfamily A, lowest in subfamily C (Table 1).

Table 1. A few examples of one-domain and two-domain GH45 enzymes. *NCBI's conserved domain database

Organism name	Molecular size (kDa)	Conserved domains*	GenBank accession number	GH45 subfamily
Cryptopygus antarcticus	23.8	GH45	ACV50414.1	A
Humicola grisea	32.2	GH45; CBM1	BAA74956.1	A
Humicola insolens	22.9	GH45; CBM1	CAB42307.1	A
Melanocarpus albomyces	25.0	GH45	CAD56665.1	A
Neurospora crassa	30.3	GH45; CBM1	CAD70529.1	A
Thermothielavioides terrestris	30.6	GH45; CBM1	AEO64667.1	A
Ampularia crossean	21.3	GH45-like	ABR92638.1	B
Mytilus edulis	19.7	GH45-like	CAC59695.1	B
Trichoderma reesei	22.8	GH45-like; CBM1	CAA83846.1	B
Phanerochaete chrysosporium	20.9	GH45-like	BAG68300.1	C
Gloeophyllum trabeum	20.4	GH45-like	EPQ56593.1	C
Gymnopilus dilepsis	26.0	GH45-like; CBM1	PPQ98991.1	C
Fomitopsis palustris	21.8	GH45-like	AVV62522.1	C

Despite the occasional presence of a second domain, the molecular size of GH45 enzymes usually is relatively small, around 20-30 kDa. These enzymes are amongst the smallest glycoside hydrolases, with

Cel45A from *Trichoderma reesei* (TrCel45A) being one of the smallest cellulases known to date. This unique size, possibly, could have served as an evolutionary advantage in ensuring easier migration between cellulose microfibrils in amorphic cellulose regions of the plant cell wall.

The substrate specificity of GH45 enzymes is broad. In addition to cellulose, these enzymes can degrade other plant CW polysaccharides such as mannans and MLGs. Yet another peculiarity is the relatively low hydrolytic activity of these enzymes. In comparison to other cellulases, GH45 enzyme catalytic capabilities are small, which indicates that the evolutionary role of these proteins has not been to actively cleave every available cellulose fibre.

Taxonomic distribution

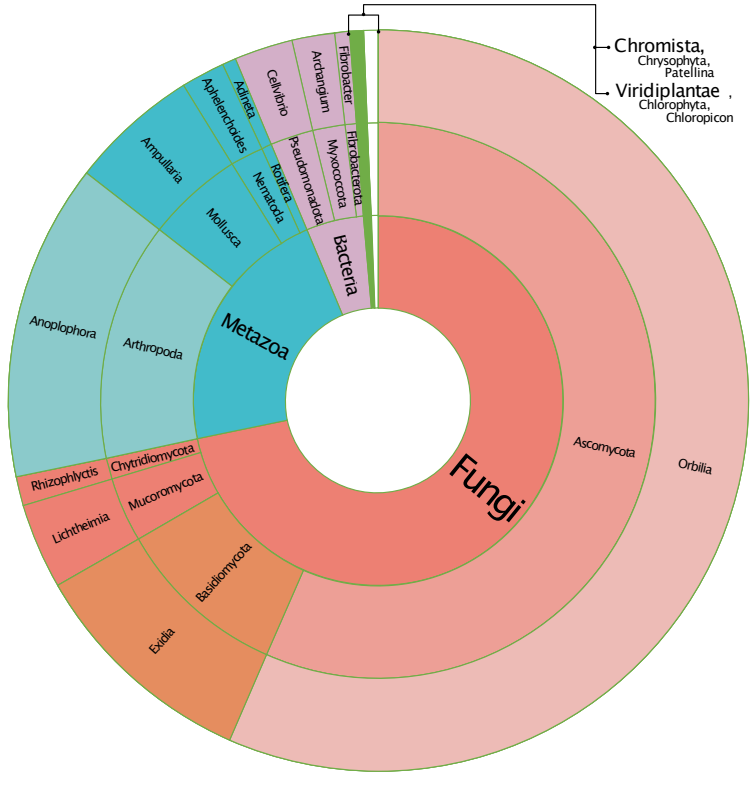


Figure 6. Taxonomic distribution of GH45 enzymes, based on number of genera.

From all GH45 entries in the CAZy.org database, 72% of phyla are from kingdom Fungi, 22% from kingdom Animalia (former Metazoa), 5% are from Bacteria, and 0,6% – from Viridiplantae (or Plantae) (Figure 6).

Fungi

Phyla: Ascomycota, Basidiomycota, Chytridiomycota, Mucoromycota.

The majority of GH45 enzymes have been identified in fungi. These enzymes are found in 114 genera over 4 phyla, including phyla of early-diverging fungi. Some early-diverging species having as many as 20 GH45-encoding genes, for example the anaerobic rumen-residing filamentous chytrid fungus *Neocallimastix californiae* (Lange et al., 2019). GH45-secreting fungi originate from different habitats (anaerobic, aerobic, aquatic, terrestrial), lifestyles (symbiotic, parasitic), and metabolize various kinds of organic matter. Ascomycota comprises largest part of GH45-producing fungi, with minimum of 90 genera containing GH45-encoding genes.

Animalia

Phyla: Mollusca (Class: Bivalvia), Rotifera (Class: Eurotatoria), Arthropoda (Class: Mandibulata).

Up until 1998, a common view was that the cellulolytic capabilities of roaches and lower termites can be attributed to their protozoan symbionts, gut bacteria or fungal cellulases present in their food (Michael, 1991). In 1998, however, a cellulase-encoding gene was discovered in termites (Watanabe et al., 1998), which raised the question of whether cellulases were in fact acquired via horizontal gene transfer (HGT). In 2019, Busch et al. carried out phylogenetic analyses which suggested that GH45 enzyme encoding genes in Phytophaga beetles were acquired by HGT from fungal species. Some bivalves, such as the blue mussel *M. edulis*, also appear to have GH45-encoding genes (Xu et al., 2001), whereas other bivalves, such as shipworms *Teredo navalis*, possess GH45 enzymes due to endosymbiotic bacteria residing in their gills (Sabbadin et al., 2018).

GH45-encoding genes have been discovered also in the microscopic freshwater rotifer *Adineta ricciae* (Szydlowski et al., 2015).

Bacteria

Phyla: Fibrobacterota (Class: Fibrobacteria), Myxococcota (Class: Myxococcia), Pseudomonadota (Class: Gammaproteobacteria).

The genomes of few bacteria contain GH45-encoding genes. This includes 8 genera, which comprise mostly rumen-residing bacteria, soil bacteria, and compost bacteria, but includes also marine gammaproteobacterial, such as the shipworm endosymbiont *Teredinibacter turnerae*.

Viridiplantae

Phylum: Chlorophyta (Class: Chloropicophyceae).

One GH45 enzyme has been identified in green algae *Chloropicon primus* (GenBank: QDZ22954.1). This protein is a two-domain protein with a CBM1 domain at the N-terminus and a GH45 catalytic domain (CD) at the C-terminus. A Protein-BLAST search shows the CD to be more than 40% sequence identical with GH45s from some ascomycetes and rotifers.

2.4.1 Subfamily A, B, and C

Sequence-based phylogenetic analyses group GH45 enzymes into three subfamilies – A, B and C.

The letter assignment indicates chronological discovery of the subfamilies. Subfamily A is the most studied and has incited the discovery of other subfamilies. Subfamily C is the least studied, but is of a research interest due to the peculiar nature of its members, that exhibit relatively low catalytic activities and bear a structural resemblance to non-hydrolytic proteins – expansins, loosenins, swollenins (Cosgrove, 2015).

All three subfamilies have been found in fungi. GH45 members of nematodes, insects, springtail and bacteria belong to subfamily A (Kikuchi et al., 2004; Lee et al., 2004; Mei et al., 2016; Pauchet et al., 2010, 2014; J. M. Song et al., 2017; Valencia et al., 2013), while molluscs only have subfamily B endoglucanases (R. Guo et al., 2008; Sakamoto & Toyohara, 2009; Tsuji et al., 2013; Xu et al., 2000). Subfamily C is only found in basidiomycete fungi, so far (Berto et al., 2019; Cha et al., 2018; Igarashi et al., 2008).

The molecular structure of *Hi*Cel45A has served as the template for catalytic residue assignment in all subsequently discovered GH45 enzymes. Studies on *Hi*Cel45A proposed that Asp121 acts as a general acid and Asp10 most likely is the general base. Introduction of

mutation at these sites, D121N mutant and D10N, led to complete inactivation of *Hi*Cel45A (Davies et al., 1995). Interestingly, a D114N retained no more than 1% of its activity comparatively to the wildtype, indicating a significant role of Asp114 in the catalytic mechanism of *Hi*Cel45A. This residue was assigned the role of an assisting residue.

The role of an aspartic acid positioned at the location of Asp121 as the catalytic acid has been experimentally proven also in the other subfamilies. Mutations of the corresponding residue in the subfamily B member *Ac*Cel45A (D137A) and subfamily C members *Fp*Cel45A (D117N), from the brown-rot fungus *Fomitopsis pinicola*, and *Pc*Cel45A (D114N) lead to complete a removal of the enzymatic activity of these enzymes(Godoy et al., 2018; Nakamura et al., 2015; Nomura et al., 2019).

Mutation of the proposed catalytic base, however, does not lead to complete inactivation of subfamily B members, although only few percent of the activity remains (*Ac*Cel45A D27A) (Nomura et al., 2019). An example of further divergence among GH45 enzymes perhaps is the existence of subfamily C, where the catalytic base is not found in the aforementioned location. Instead, an asparagine residue residing on a loop above the active site cleft, at the position of the assisting residue in *Hi*Cel45A (D114), has been proposed to function as an alternate base (Asn92 in *Pc*Cel45A). Furthermore, a drastic reduction of the enzymatic activity occurs upon the mutation of Asn92 in *Pc*Cel45A and the corresponding residues in all GH45 subfamilies were tested (D114N in *Hi*Cel45A, N112A in *Ac*Cel45A, N95D in *Fp*Cel45, N92D in *Pc*Cel45A)(Cha et al., 2018; Davies et al., 1995; Nakamura et al., 2015; Nomura et al., 2019).

The uniqueness of GH45 subfamily C enzymes lies in their ability to catalyse hydrolysis despite the absence of the traditional catalytic base residue at the active site. In 2015 Nakamura et al. (Nakamura et al., 2015) proposed a novel catalytic mechanism for the subfamily C member *Pc*Cel45A. They carried out high-resolution X-ray and neutron diffraction analyses and suggested a proton relay-network to underlie the hydrolytic action of *Pc*Cel45A, this network links Asn92 to the catalytic acid (Asp114 in *Pc*Cel45A) by a chain of hydrogen bonding.

2.4.2 Structurally similar proteins

The DPBB domain found in GH45 enzymes is bearing the name GH45 or GH45-like domain. Interestingly, this domain is found also in proteins which do not exhibit any hydrolytic activity, such as expansins, loosenins, swollenins, cerato-platanins (Figure 7). GH45 domains are also found in wound-induced plant protein barwin (Huet et al., 2013), which belongs to pathogenesis-related protein PR-4 proteins, and the virulence-associated protease inhibitor rare lipoprotein A (RlpA) from *Rhizoctonia solani* (Charova et al., 2020), as well as in bacterial murein lytic transglycosylases such as *E. coli* MltA, involved in remodelling of bacterial cell wall peptidoglycan (van Straaten et al., 2007). Some of the homologous proteins contain only the GH45 domain, for example loosenins, cerato-platanins, barwin and the protease inhibitor. Others contain a beta-sandwich domain (CBM63) attached to the DPBB (plant expansins). Other domains are also found in combination with the GH45-like domain and the CBM63 domain, which have been found in microbial expansins, swollenins.

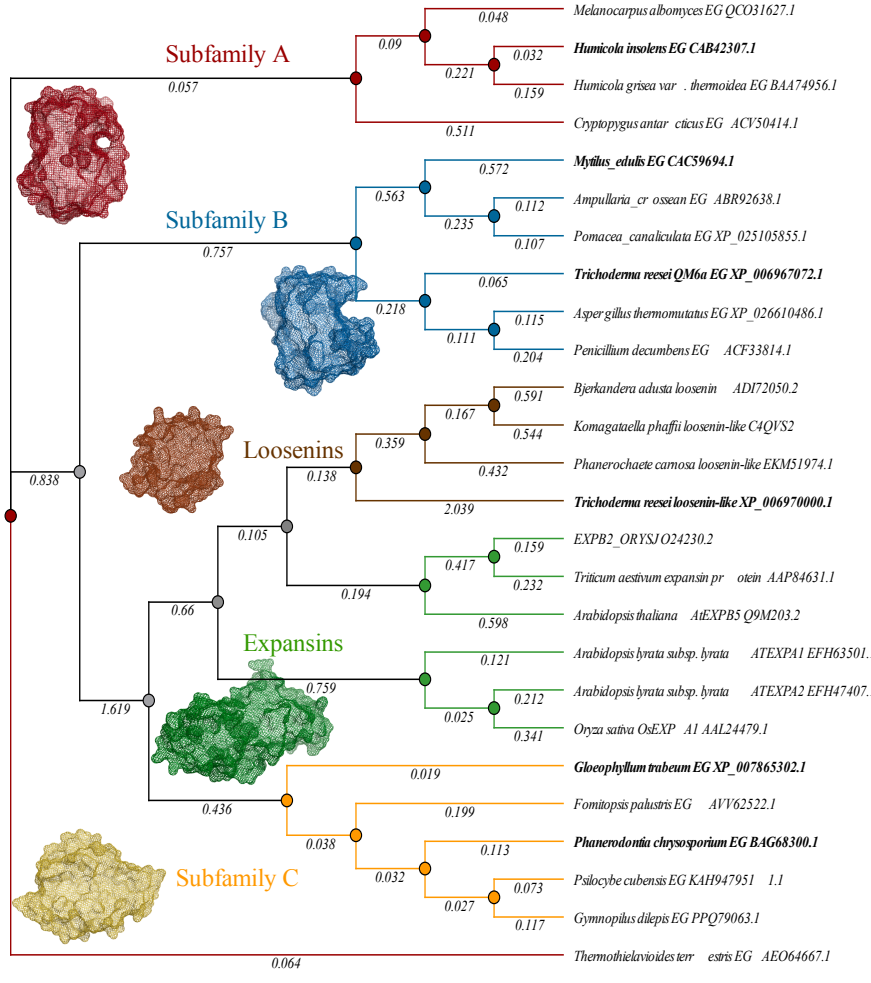


Figure 7. Phylogenetic tree of GH45 enzymes and homologous proteins. Depicted in bold are the names of the proteins investigated in this thesis.

Expansins

Expansins are plant cell wall loosening proteins associated with cell enlargement. They are involved in a variety of developmental processes in which cell wall modification occurs. For example, seed germination, leaf development and root hair development. Suppression of expansin genes in plants has been associated with delayed seed germination, underdeveloped root system and plants generally grow to be up to 80% smaller if expansin gene suppression

occurs (Lü et al., 2013; Weitbrecht et al., 2011; ZhiMing et al., 2011). Expansins and homologous proteins are abundant in plant kingdom but can also be found in some bacteria, fungi and molluscs. Expansin roles in these organisms are associated with the degradation of cell-wall material rather than the organism's cellular development. For example, knock-out of expansin genes in Bacillus and Clavibacter reduces their ability to colonize plant tissues (Jahr et al., 2000; Kerff et al., 2008). Some phytopathogenic organisms express expansins, such as the potato plant pathogen nematode *Globodera rostochiensis* (Kudla et al., 2005).

The action mechanism of cell wall loosening by expansins remains unknown, but is speculated to be highly selective towards load-bearing structures in the PCW. Expansins are believed to unravel cellulose fibrils through a non-hydrolytic action mechanism (Cosgrove, 2000, 2015).

Organisms which express expansin-like proteins can be found outside of plant kingdom, but the terms *expansin* and *expansin-like* are reserved for specific proteins. Only proteins bearing the two domains found in plant expansins are classified under this category, therefore homologous single-domain proteins have been assigned other names.

Fungal homologues to plant expansins

Swollenins are proteins homologous to expansins that were first discovered in *T. reesei* (Saloheimo et al., 2002). Experiments with swollenin gene overexpression or silencing showed that these proteins significantly increase *T. asperellum* ability to colonize cucumber *Cucumis sativus* roots (Brotman et al., 2008). Swollenins are ~75 kDa in size and consist of three domains – CBM1 at the N-terminu, GH45 domain at the core, and a CBM63 domain at the C-terminus. The N-terminal CBM1 has been proved to be indispensable for the action of this protein (Brotman et al., 2008). In addition to the substrates which GH45 enzymes bind to, swollenins bind to xylan and lignin. Generally, swollenin activity does not release reducing sugar, with some exceptions of low initial sugar release on β(1→4) glucans (Eibinger et al., 2016).

Cerato-platanins are one-domain proteins, ~12 kDa in molecular size, which were first discovered in the plant pathogenic fungus *Ceratocystis platani* and referred to as phytotoxic proteins (Pazzagli

et al., 1999). These proteins serve as virulence factors of pathogenic fungi and as signalling molecules in ectomycorrhizal communities to elicit plant immune response (Gaderer et al., 2014).

Loosenin genes are wide-spread among cellulolytic fungi. These proteins were first discovered in *Bjerkandera adusta* (Quiroz-Castañeda et al., 2011). Loosenins exhibit cellulose fibril loosening behaviour, while it is unclear whether fungi utilize these proteins mainly in pathogenic mechanisms or they serve an analogous function to expansins in plants (Monschein et al., 2023; Quiroz-Castañeda et al., 2011). No structure of loosenins has been experimentally obtained to this date, therefore all structure comparisons rely on homology modelling. Loosenins are one-domain proteins, ~11 kDa in size and consisting solely of GH45-like domain surrounded by loops. Sequence-based analyses reveal these proteins to be remarkably similar to subfamily C enzymes of GH45 family, minus the hydrolytic activity (Payne et al., 2015). Similar to GH45 subfamily C enzymes, loosenins have conserved an aspartic acid residue at the same location as the catalytic acid. Moreover, the alternate base of subfamily C members is also present in loosenins, yet loosenins exhibit no hydrolytic activity on cellulosic substrates (Monschein et al., 2023).

2.5 Methods background

2.5.1 Heterologous protein expression and purification

In order to produce proteins of interest in a laboratory setting, different microorganism platform strains have been created. Protein of interest encoding genes become expressed in a host organism which naturally does not possess the gene. Gene insertion is done with molecular cloning methods – an assembly of a recombinant DNA and its replication within host cells.

The methylotrophic yeast *Pichia pastoris* or *Komagataella phaffii* is a well-known platform strain for heterologous protein expression. *P. pastoris* can be grown to high densities, has an efficient secretory system and is a popular choice over bacterial expression systems when eukaryotic post-translational modification is essential for the protein production. This yeast is capable of metabolizing methanol as a carbon source through Methanol Utilization Pathway (MUT Pathway), which

takes place in the peroxisome and the cytosol. In MUT, alcohol oxidase (AOX) oxidizes methanol to formaldehyde and H_2O_2. *P. pastoris* genome contains two AOX encoding genes – AOX1 and AOX2. AOX1 is the stronger promoter. Next major step in the pathway is the breakdown of H_2O_2 into water and oxygen by catalase. In cytoplasm, a portion of the formaldehyde gets oxidized to formate and CO_2 releasing energy in the form of NADH. The other portion of formaldehyde is used to produce dihydroxyacetone (DHA) and glyceraldehyde-3-phosphate (GAP). The exposure of yeast cells to methanol as carbon source leads to production of two enzymes – AOX and peroxisomal enzyme dihydroxyacetone synthase (DAS).

Three phenotypes of *P. pastoris* strains are currently commercially available:

1) methanol utilization plus (Mut+), AOX1 and AOX2 genes are intact and active, MUT pathway repressed by glucose and glycerol;
2) methanol utilization slow (MutS), AOX1 gene is knocked-out, MUT pathway repressed by glucose and glycerol;
3) methanol utilization minus (Mut-), AOX1 and AOX2 genes knocked-out, strain unable to grow on methanol as sole carbon source.

Protein expression is affected by the type of promoter present in the expression system – unregulated (constitutive) or regulated (inducible). If an expression system has the recombinant protein under an inducible promoter, then protein expression can be manipulated with addition or removal of oppressor molecules. MUT pathway in *P. pastoris* Mut+ and MutS strains is oppressed by high levels of glucose and glycerol, removal of these molecules is crucial for protein expression.

Cell cultivation

Proteins described in this study were produced using zeocin-resistant *P. pastoris* X-33 (Mut+, constitutive expression) and KM71H (MutS, inducible promoter) expression systems. The protein-expressing strains were kindly provided by prof. K. Igarashi at Tokyo University, Japan.

P. pastoris culture cultivation was carried out both in shake-flasks and in bioreactors, according to Invitrogen Pichia Fermentation

Process Guidelines and in-house methods. *P. pastoris* is a facultative anaerobe which, despite recombinant expression significantly increasing upon oxygen deprivation (Baumann et al., 2011), requires some oxygen dispersion in growth media. Therefore, aeration was done in all cultivation instances. To acquire high cell densities, cultures were incubated at 30°C. Methanol-induced protein expression was carried out at 20-25°C.

GtCel45A expression in A. nidulans

For production of *Gt*Cel45A of subfamily C, from the brown-rot fungus *Gloeophyllum trabeum*, the filamentous fungus *Aspergillus nidulans* was instead used as expression system. An *A. nidulans* strain expressing *Gt*Cel45A had previously been constructed by the group of Prof Fernando Segato, University of Sao Paulo, Brazil (Berto et al., 2019) and was kindly provided to us.

Isotope-labelled protein production

Cultivation in Multifors bioreactors was chosen for isotope-labelled protein production. The advantage of bioreactors is the highly controllable environment, contrary to shake-flasks where condition such as aeration, feed and pH control is fairly limited.

The production of isotopically labelled *Pc*Cel45A was done applying a two-phase fermentation scheme: 1) biomass accumulation; 2) induction phase. Cells were grown as a batch-culture in a minimal medium (YNB w/o aa and $(NH_4)_2SO_4$) with isotope ^{15}N-labelled ammonium sulphate and ^{13}C-labelled glucose. The induction was started by lowering the temperature and adding ^{13}C-labelled methanol at the lowest feed rate. Lower temperatures during protein expression in MUT strains have shown to lead to higher protein yields (Anasontzis et al., 2014). Low methanol feed rates are advised by *P. pastoris* fermentation manual, as the organism is sensitive to methanol concentration (Trinh et al., 2003).

P. pastoris cells grow to highest densities with glycerol as the carbon source, but, when a ^{13}C-labelled carbon source was of essence, ^{13}C labelled glucose instead was chosen for cell cultivation to limit the cost.

Protein purification

To get to protein structure and function studies, the protein of interest must be isolated from its culture medium, which is done through series of purification steps. The following methods were utilized for purification of proteins used in this study.

Gel filtration chromatography is used to fractionate molecules in a sample by size. This characteristic can be utilized to: fractionate molecules and complexes within a specific size range; determine the size of a molecule; do a buffer exchange; desalt the sample; remove unwanted large or small molecules, such as nucleotides and contaminants from the sample; and more.

The stationary phase of a gel filtration chromatography column is composed of a porous matrix with beads that have a defined pore size. The mobile phase is a buffer flowing between the matrix beads. All molecules too large to enter the pores remain in the mobile phase and are moved through the column along with the buffer.

Due to the longer path taken to migrate through the bead pores, molecules able to enter the pores of the stationary phase, move through the gel filtration column much slower. This leads to fractionation of molecules according to their sizes in time. Smallest molecules will be hindered by beads the most, thus eluting last, while largest molecules will be the least hindered and will become eluted from the column first.

Desalting columns, such as Biogel P6DG (BioRad), were used for buffer exchange – to place protein solution into the equilibration buffer of next purification method.

Ion exchange chromatography (IEX) is used to fractionate molecules by charge. Protein molecules are composed of amino acids with different charge, which gives each protein a unique overall charge. Protein charge is pH dependent, it can be positive, negative or no charge. Isoelectric point (pI), is used to determine the charge of a protein molecule in a solution. It represents the pH at which the protein has no net charge.

The stationary phase is an immobile matrix with ionic functional groups on the surface that interact with ions of the opposite charge. When loaded on to IEX column, protein molecules with appropriate charge bind to the oppositely charged resin and then become eluted upon addition of exchangeable counterions, as they get out-competed

for binding to the stationary phase. Initially, molecules unable to bind or with weak binding to the stationary phase are first to be exit the chromatography column.

Based on type of net charge on the protein of interest, two types of ion exchange chromatography are available: 1) anion-exchange chromatography; 2) cation-exchange chromatography.

For cation-exchange chromatography, the molecule of interest has to be positively charged. To achieve a positive charge, the protein must be put in a solution with a pH that is between 0.5 to 1.5 units less than the protein's pI. In cation-exchange chromatography, the stationary phase is negatively charged and positively charged molecules bind the matrix.

For anion-exchange chromatography, the molecule of interest has to be negatively charged to bind the column resin, as the stationary phase matrix is positively charged. A negative protein charge can be achieved by placing the protein into a solution with a pH 0.5 to 1.5 units greater than the pI.

The most frequently used IEX method in our studies was anion exchange chromatography with DEAE Sepharose as stationary phase.

Immobilized metal ion affinity chromatography (IMAC) is used to separate proteins which have affinity towards certain metals (protein molecules containing specific residues, molecular tags) from protein mixtures.

The stationary phase matrix contains immobilized metal ions (for example, Ni2+, Cu2+, Co2+) that have high selectivity towards specific amino acids, such as histidine and cysteine. Amino acid residues in peptides and proteins form complexes with the stationary phase at pH 6 – 8, thus capturing the molecules on the IMAC column. The binding reaction is pH dependent and the sample can be eluted in a number of ways: by reducing the pH, increasing buffer's ionic strength, or by adding other chemical agents, such as EDTA or imidazole, to the buffer.

HisTrap HP (Cytiva) with immobilized nickel ions were the IMAC columns used in our studies. Ni-Affinity Chromatography utilizes the interaction of nickel and histidine for purifying the Histidine Tagged proteins (6X His Tag).

2.5.2 Biochemical characterization

GH45 enzymes act on carbohydrate polymers, which makes the activity quantification through traditional means cumbersome. Classically, quantification of catalytic activity in enzymatic assays is measured by quantification of product formation, which is ideal for small, soluble and homogenic substrates. However, polysaccharides are more or less heterogenous with respect to the accessibility of different cleavage sites, meaning that their properties will change during the course of degradation and the rate typically decrease and often level off by time. Therefore, great caution is needed for proper activity comparisons.

As the hydrolytic activity of GHs often leads to the formation of monosaccharides or oligosaccharides which contain a reducing end, cellulase activity is usually quantified by assays that require the presence of a reducing end on a sugar. DNSA and PHBAH are examples of commonly used reducing sugar assays for cellulase activity quantification (Mellitzer et al., 2012). Glucose in a solution exists mainly in a pyran ring form, with an open chain making up less than 0,02%. Most of glucopyranose in a solution is in a beta pyranose (62%), the rest is in an alpha pyranose form. At 25°C in pure water, these forms exist in a 62:38 equilibrium which is reached through mutarotation (opening and closing of the ring), via a linear chain form (Alexandersson & Nestor, 2022; Ouellette & Rawn, 2015). Any monosaccharide which contains a hemiacetal group is able to act as a reducing agent and therefore is called a *reducing sugar*. When glucopyranoses are joined together they are linked by an acetal group. Therefore, only upon cleavage of glucans, new free reducing ends are formed.

Principles behind PHBAH assay relay on the formation of osazones – compounds which form when reducing sugars react with hydrazines (Lever, 1972; Lever et al., 1984). PHBAH reagent contains para-hydroxybenzoic acid hydrazide that, when subjected to boiling temperatures at alkaline pH, reacts with the aldehyde group of an open-chain glucose (Lever et al., 1984). Upon heating, the reaction solution gains colour, which is proportional to the number of reducing ends in the sample. The colour intensities can then be quantified with a spectrophotometer.

2.5.3 Protein X-ray diffraction crystallography

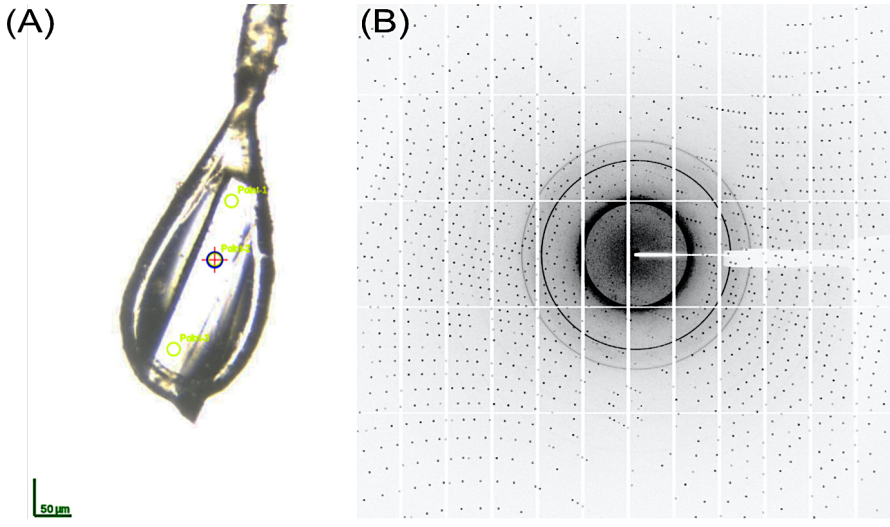


Figure 8. (A) image of *Gt*Cel45A crystal in a loop directly before data collection; (B) diffraction pattern obtained from the *Gt*Cel45A crystal.

Nowadays, X-ray crystallography is a popular technique for structure determination of proteins, but during the 19th century, crystallization was mainly used as a purification tool. Determining the structure of biological macromolecules is an essential step in understanding the molecular interactions which often are at the essence of research in most biological sciences. X-ray crystallography has provided insights into molecular mechanisms involved in many research areas, for example, virus structure studies, structure-based drug design, elucidation of enzyme mechanisms. This method has been used to determine the size of atoms, atomic-scale differences in materials, the lengths and types of chemical bonds. X-ray crystallography has led to vast amount of significant scientific discoveries, perhaps most famous of which is the structure of DNA (McPherson & Gavira, 2013; Watson & Crick, 1953). From 1901 to 2013 total of 29 Nobel Prizes have been awarded to research conducted through crystallography.

The path towards 3D molecular structure determination starts with a crystal. A protein crystal is obtained from a pure protein. The importance of purity is high, as contaminations might interfere with the crystal formation process. The principles behind protein

crystallization are similar to crystallization of non-biological compounds. Crystals form when atoms or molecules become highly ordered in a structure. To achieve the high organization, macromolecules in a solution must be supersaturated.

Supersaturation is obtained through the addition of precipitating agents or precipitants. Examples of precipitants are salts or hydrophilic polymers. Manipulation of parameters such as temperature, ionic strength and pH also play a crucial role in producing supersaturation. Often added are substances which can affect the structural state of the macromolecule, such as metal ions, ligands, inhibitors or other small molecules.

Macromolecular crystallization does not have a strict comprehensive theory to be used as a guide to predict a sure (or even semi definite) outcome. Crystallographer success is based in various principles, experience, and inquisitiveness. To quote Alexander McPherson and Jose A. Gavira (McPherson & Gavira, 2013):

> […] macromolecular crystal growth is largely empirical in nature, and demands patience, perseverance and intuition.

Crystallization process involves a vast number of variables, if altered slightly all of which could lead to failure. Naturally, experiments begin with trial and error. Initially, the aim is to cover as wide range of variables as practically possible. The basic variables include: the protein solubility and concentration, the precipitant and its concentration, the buffer and its pH, the temperature, the crystallization technique, presence or absence of additives, handling errors.

To alleviate initial condition screening process, commercially available “crystal screens” are used to find possible conditions in which the protein of interest forms crystals. These screens contain solutions with varying precipitants, buffers, salts, additives and more.

The crystallization can be carried out in multiple ways, using different crystallization plates, depending on the type of the set up – sitting drop vapor diffusion technique, hanging drop vapor diffusion technique, or dialysis.

Some projects can be deemed impossible at this step, because of the failure to crystallize the protein. After finding the crystallization conditions, crystallographers proceed with condition optimization. Condition optimization can involve a vast variety of methods, including a process called seeding, which involves introduction of micro-/nanocrystals into the condition to possibly serve as a nucleation point and lead to the growth of larger crystals. Before moving forwards with structure determination, crystal verification should be carried out in some way, for example, subjecting a crystal to a polyacrylamide gel electrophoresis analysis (PAGE), staining of the crystal, or, if available, a test X-ray exposure.

Despite being the fundamental base of X-ray crystallography and a lifeline of a project, obtained crystals serve simply as a part of the raw input data which later is used to visualize the protein molecules. After obtaining a macrocrystal, it is analysed using X-rays. The X-ray beam is commonly generated at a synchrotron source, as these use high-quality optics and are capable of producing intense X-rays. Data collection at a synchrotron allows shorter exposure times and a higher signal to noise ratio in regards to the diffraction image. During a crystal X-ray diffraction measurement, a finely focused monochromatic beam of X-rays is directed towards a crystal, which then causes the X-rays to diffract. The X-ray diffraction creates a pattern of regularly positioned spots or reflections. Two-dimensional (2D) images of the reflection pattern are collected from multiple angles of the crystal. Afterwards, the patterns from 2D images are processed and eventually converted into an electron density map – a three-dimensional (3D) model created using the mathematical method of Fourier transformation.

From the spacing of the spots, a unit cell can be determined – the larger the cell, the more spots per unit area. A unit cell is the smallest repeating unit within the crystal and its dimensions are described with three lengths (a, b, c) and three angles (α, β, γ). The shape of the unit cell determines the crystal system (7 systems in total). The symmetry of the diffraction pattern provides with the space group. Total of 65 space groups are possible in proteins, space groups describe the symmetries of crystal patterns – the packing of the molecules in the crystal lattice (Hahn, 2002).

Many software packages and program suites are available for the processing of the diffraction data. Data processing itself is mathematically complex, but to date we have well established algorithms which enables novice structure biologists to process data and obtain an electron density map. After data processing, follow numerous structure refinement cycles, which eventually lead to model building.

In this thesis, XDS software package was used for data processing, structure was solved and refined using PHENIX Suite, visual inspections and real space refinement were carried out using COOT, and final structure visualization was carried out in PyMOL.

2.5.4 Protein NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is used to identify substances and to provide detailed information about the molecule's structure, conformation, dynamics and interactions with other species. NMR-based methods for protein analysis (or protein NMR) are well-established for analysing protein structures and interactions at atomic resolution. Analyses can be carried out in multiple sample states – solution state, solid state, and even membranous environment. A major advantage of NMR is its non-destructive nature and thus the number of times the same sample can be reused is dependent on the sample rather than the NMR spectrometer.

Protein interaction with ligands can be of different strength but this does not represent the biological significance of the interaction as significant low-affinity reactions exist. Often proteins bind several ligands with equal affinity. However, non-specific interactions tend to be of low affinity. Determination of affinity when the reaction is either very strong (K_d < nM) or very weak (K_d > μM) can become challenging as the quantification methods are heavily impacted by the sample solubility and assay sensitivity. NMR spectroscopy stands out in this regard, as it can provide quantitative measurements for very low affinities. In NMR experiments the chemical exchange is described as low, intermediate or fast on the NMR time scale (Teilum et al., 2017).

NMR spectroscopy is based on the principles of the nuclei spin – an intrinsic property of the atomic nuclei that gives rise to a magnetic moment when the spin quantum number (I) is not zero. Only if nuclei

have a magnetic spin, can nuclei be observed by NMR. If the numbers of protons and neutrons are both even, the nuclear spin quantum number I is 0, which means they will not produce an NMR signal. For example, the number of protons (6) and the number of neutrons (6) of carbon-12 are both even therefore, ^{12}C won't be visible in an NMR spectrum. However, odd mass nuclei have half-integer spins, for example, I = 1/2 for ^{13}C, which have a spherical charge distribution, and they are easier to utilize for NMR spectroscopic studies.

Hydrogen nuclei, when placed in a magnetic field, will split into two different spin states (Zeeman effect), which are described as vectors with quantized size and direction. The nuclei will align along with the applied magnetic field but, if excited out of that state, the nuclei can also be positioned against the magnetic field. There is a slight difference in energy between these two states, which can be measured at a specific frequency – the resonance frequency. Protons at different positions will have a slightly different resonance frequency, because, the exact resonance frequency is also affected by the surrounding electronic environment, including neighbouring atoms. The variation of the resonance frequency with the local environment is measured by the *chemical shift*, δ, and is expressed as the difference in resonance frequency compared to a reference standard. To avoid that the chemical shift is dependent on the magnetic field strength, it is divided by the resonance frequency of the reference and expressed in ppm. In ^{1}H and ^{13}C NMR spectroscopy, tetramethylsilane (TMS) is used as a reference.

As the name suggests, spectroscopy methods generate spectra for data visualization. Essentially, what is recorded by an NMR spectrometer is a combination of radiofrequencies, known as the free induction decay (FID), which is converted to an NMR spectrum by a so-called Fourier transformation. The ^{1}H NMR spectrum shows signals corresponding to differently positioned hydrogen atoms with defined chemical shifts. The information that can be extracted from a ^{1}H NMR spectrum is not restricted to the chemical shift, but also involves coupling constants, the linewidth, and the relative intensities of the signals.

NMR experiments

Protein structure determination by NMR spectroscopy begins with sample preparation and NMR experiments, followed by processing and interpretation of the spectra, structure calculation and validation.

Most commonly, proteins intended for NMR spectroscopy analyses are recombinantly expressed using *E. coli* expression systems (Mondal et al., 2013), but the popularity of *P. pastoris* as an expression system for this purpose is increasing (Clark et al., 2018; Courtade et al., 2016; Rodriguez & Krishna, 2001). The simplest labelling method for protein NMR is to uniformly label all carbon and nitrogen atoms with the stable isotopes ^{13}C and/or ^{15}N. The protein concentration should be around 500 μM or higher and for a successful resonance assignment, the degree of labelling must be high. In addition, the protein should be dissolved in a buffer and the sample should contain 5-10% D_2O. In case of high purity and successful labelling, a single protein sample can provide a lot of the structural information about the protein and the same sample can be used for various purposes, including protein-ligand interaction studies.

The most common procedure for protein NMR relies heavily on the backbone amide groups for defining the different amino acid residues. By using a ^{13}C,^{15}N-double labelled protein, triple resonance experiments utilizing both ^{1}H, ^{13}C and ^{15}N can be carried out. These are most often 3D NMR experiments, where the extra dimensions allow enough spectral resolution to distinguish between the different residues.

Some protein NMR experiments show correlations not only within the same amino acid residue (*i*), but also from the previous amino acid which shares a peptide bond (*i-1*). A set of standard NMR experiments are used in conjunction to produce protein backbone assignments.

The following NMR experiments were carried out:

1) **^{1}H,^{15}N-HSQC** (2D): One-bond ^{1}H-^{15}N correlations are shown; cross-peaks represent mainly the backbone amide groups, but also sidechain NH are present. Used as a quality check for the protein NMR sample, the spectrum defining the different amino acid residues during resonance assignment and for obtaining chemical shift changes from titration experiments with ligands.

2) **^{1}H,^{13}C-HSQC** (2D): One-bond ^{1}H-^{13}C correlations are shown similarly to the ^{1}H,^{15}N-HSQC. Carried out as quality check for ^{13}C labelling.
3) **HN(CA)CO** (3D): shows correlations between amides and carbonyl carbons; produces two cross-peaks with CO from *i* and *i-1*, where *i-1* is weaker.
4) **HNCO** (3D) – shows correlations between amides and carbonyl carbons; generates one cross-peak per aa, the CO of *i-1*.
5) **CBCA(CO)NH** (3D) – shows correlations between amides and CA and CB of the preceding residue (*i-1*).
6) **HNCA** (3D) – shows correlations between amides and CA from *i* and *i-1*
7) **HNCACB** (3D) – shows correlations between amides and CA and CB from *i* and *i-1*, where *i-1* is weaker.

3. Aim of the study

The aim of this study was to investigate the structural and enzymatic differences between members of different GH45 subfamilies.

To be able to conduct these studies, the following objectives were set:

- Compare the molecular structures and the enzymatic activities of GH45 enzymes from subfamily A, B, C – Cel45A from *H. insolens, T. reesei, M. edulis, P. chrysosporium*. (Paper I)
- Solve the molecular structure of a GH45 enzyme – Cel45A from *G. trabeum*. (Paper II)
- Investigate the protein-substrate interactions in Cel45A from *P. chrysosporium* (Paper III)

To fulfil the set objectives, the following methods were applied:

- heterologous protein expression in shake-flasks and fermenters (non-labelled, single-labelled, double-labelled protein expression),
- protein purification,
- protein X-ray crystallography,
- NMR spectroscopy,
- Biochemical characterization – reducing sugar assay, HPAEC,
- Bioinformatics – homology modelling, multiple sequence alignments.

4. Results and discussion

This thesis describes the structural and functional variance within carbohydrate-active enzymes belonging to the GH45 family.

The majority of proteins used in this study were obtained via heterologous protein expression in *P. pastoris*, except *Gt*Cel45A that was produced in the filamentous fungus *Aspergillus nidulans.* Some proteins were kindly provided by collaborators or available as already produced and purified protein samples in-house. During this project, the catalytic activities of a total five GH45 enzymes were investigated (*Pc*Cel45A, *Me*Cel45A, *Hi*Cel45A, *Tr*Cel45A, *Gt*Cel45A), two of the GH45 proteins were produced in-house (*Pc*Cel45A, *Gt*Cel45A), one (*Gt*Cel45A) was crystallized, structure determined and deposited in PDB (PDB ID: 8BZQ).

Attempts were also made to crystallize six loosenin proteins: LOOL12 (GenBank: EKM51974.1), LOOL2 (GenBank: EKM55357.1), LOOL7 (GenBank: EKM53490.1), LOOL9 (GenBank: EKM52742.1), *Ba*LOOS1 (GenBank: ADI72050.2), TrLOOL1 (GenBank: XP_006970000.1). Crystals were obtained of two loosenin proteins (LOOL12, *Ba*LOOS1), but none of the tested crystals provided sufficient X-ray diffraction datasets for structure determination. Loosenins were also included in a few activity studies, but no hydrolytic activity was detected.

Method development was carried out for the production of an isotope labelled *Pc*Cel45A (^{15}N-, and ^{13}C,^{15}N- respectively), as well as one loosenin, ^{13}C,^{15}N-*Tr*LOOL1. Furthermore, the backbone resonance assignment for ^{13}C,^{15}N -*Pc*Cel45A and ligand interaction studies with cello-oligosaccharides were done using NMR spectroscopy.

Amino acid sequence comparisons were performed in conjunction with structure studies to elucidate the distinctness of individual GH45 subfamilies and investigate the similarities to loosenins.

4.1 Structural studies

GH45 enzymes share a highly conserved DPBB domain, some contain an additional domain – a carbohydrate binding module CBM1. Currently, there are nine structures of GH45 enzymes deposited in the PDB. In our studies we compared the molecular structure of Cel45A enzymes from: *H. insolens* (*Hi*Cel45A, PDB ID: 4ENG), *T. reesei* (*Tr*Cel45A, predicted), *M. edulis* (*Me*Cel45A, PDB ID: 1WC2), *A. crossean* (*Ac*Cel45A, PDB ID: 5XBU), *G. trabeum* (*Gt*Cel45A, PDB ID: 8BZQ), *P. chrysosporium* (*Pc*Cel45A, PDB ID: 5KJO), *G. dilepis* (*Gd*Cel45A, predicted). To investigate the Cel45A structures, protein structure models were downloaded from PDB and homology modelling was carried out for Cel45A enzymes without publicly available 3D structures. Structural comparisons were made across GH45 subfamilies and within a subfamily.

4.1.1 Structure studies of GH45 enzymes (Paper I, II)

The most obvious structural variance is between members from different subfamilies. While the core of the catalytic domain superposes well, variations in the lengths of flanking loops cause structural dissimilarities. All GH45s have a globular structure, but the shape of substrate binding area differs.

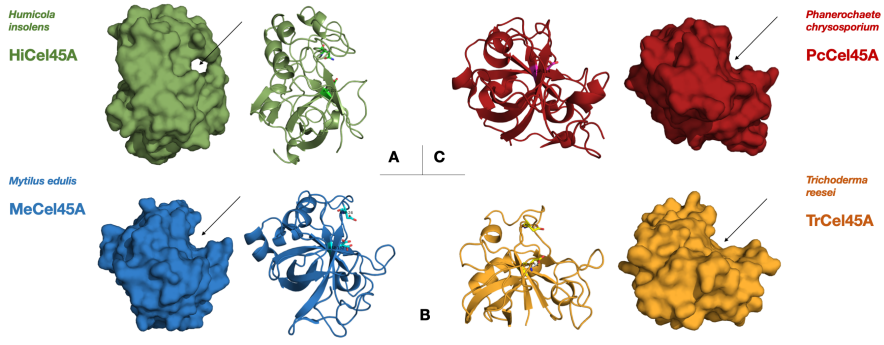


Figure 9. Surface views of GH45 endoglucanases from subfamily A, B and C. Arrows point to substrate binding area in *Hi*Cel45A (PDB: 4ENG), *Pc*Cel45A (PDB: 5KJO), *Me*Cel45A (PDB: 1WC2), and *Tr*Cel45A (GenBank: CAA83846.1, homology model). Catalytic acid and base are shown as sticks in the cartoon representation. Letters A, B, C indicate subfamily membership.

The substrate binding and active site area in subfamily A member *Hi*Cel45A resembles a tunnel (Figure 9). Studies on *Hi*Cel45A revealed this tunnel to be the result of a conformational change upon binding of the substrate, where a loop carrying residues 111-118 encloses into the active site cleft (Davies et al., 1995). This structure was observed only in a crystal complex with cellobiose (PDB ID: 3ENG) and cellohexaose (PDB ID: 4ENG), whereas the apo structure did not have a sufficient electron density in the 111-118 area for this loop to be modelled. The complex structures indicate a substrate induced "lid closure", which positions another aspartic acid residue, Asp114, near the catalytic centre (Davies et al., 1995).

The substrate binding and active site areas of subfamily B and C members are more alike, resembling an open cleft (Figure 9). Among GH45 subfamily B members, enzymes from molluscs, *Me*Cel45A and *Ac*Cel45A, have narrower active site clefts, whereas the central part of the binding area in fungal *Tr*Cel45A seems to be more open and shallower (Figure 10). None of the two mollusc enzymes contain an additional domain, contrary to *Tr*Cel45A which has a linker and CBM1 attached at the C-terminus.

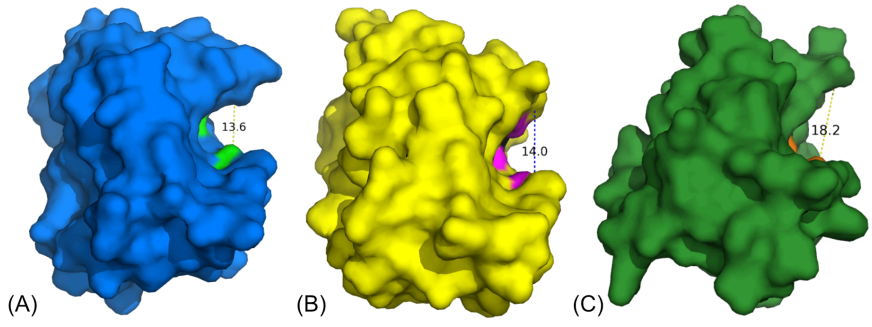


Figure 10. Surface views of GH45 endoglucanases from subfamily B, active site cleft heights indicated with dashed line. Height measurements shown in Å. (A) *Me*Cel45A (PDB ID 1WC2); (B) *Ac*Cel45A (PDB ID 5XBU); (C) *Tr*Cel45A (predicted structure).

In the subfamily C members *Pc*Cel45A and *Gt*Cel45A, long loops elongate the binding surface. The central part of the substrate binding area is noticeably more open than in subfamily B. The subfamily C member *Gd*Cel45A exhibits nearly identical structure to the catalytic domain of *Gt*Cel45A but has a linker and CBM1 domain attached at the N-terminus (Paper II).

The location of catalytic residues in subfamily A and B members is well conserved. Subfamily C, however, deviates regarding catalytic base. The corresponding location is occupied by a glycine residue instead. The proposed alternate base residue (Asn92 in *Pc*Cel45A, Asn95 in *Gt*Cel45A) is positioned on the opposite side of the cleft. Subfamily B members have conserved the positioning of the asparagine as well (Asn 109 in *Me*Cel45A), contrary to *Hi*Cel45A where this location is occupied by an aspartic acid (Asp114) that is regarded as an assisting residue in *Hi*Cel45A.

In regard to the positioning of catalytic residues in *Pc*Cel45A, Paper I showed that nearly all of the residues proposed to be involved in the proton-relay chain in *Pc*Cel45A share a similar location in *Me*Cel45A and *Ac*Cel45A, while majority of these residues are not conserved in *Hi*Cel45A.

Resemblances were investigated also at the amino acid sequence level. The Basic Local Alignment Search Tool (BLAST) blastp suite (protein-protein BLAST) was used to search for similar sequences. The following proteins were used as query sequences: *Hi*Cel45A (GenBank: CAB42307.1), *Me*Cel45A (GenBank: CAC59694.1), *Ac*Cel45A (GenBank: ABR92638.1), *Tr*Cel45A (GenBank:

CAA83846.1), *Pc*Cel45A (GenBank: BAG68300.1). The top 100 protein sequences with a minimum of 50% identity were collected and later revised to remove incomplete or identical sequences.

For each subfamily, a multiple sequence alignment was generated by Robert C. Edgar's MUSCLE (version 5) tool for multiple alignment. Consensus sequences were created using Ugene (version 41.0), with consensus type "strict" and 90% threshold. Minimum of 50 sequences were used for the generation of each consensus sequence.

Sequence analyses revealed total of 14 residues to be strictly conserved across all GH45 subfamilies.

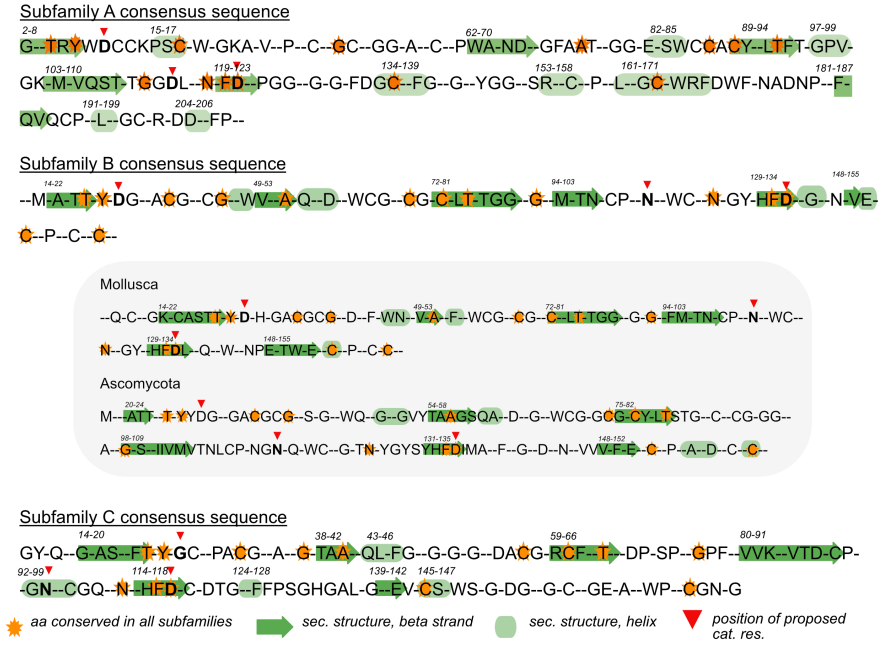


Figure 11. GH45 subfamily A, B, C consensus sequences with depiction of secondary structure elements.

A tryptophan residue coming from different locations of sequence, is conserved at the -4 subsite in all of the GH45 enzymes utilized in this study, likely exhibiting a significant function, such as serving as a sugar binding platform. Mutation of this tryptophan residue in

*Pc*Cel45A (W154A) leads to 50% decrease in catalytic activity (Godoy et al., 2018).

4.1.2 The structure of Cel45A from *G. trabeum* (Paper II)

During the thesis project, a subfamily C member *Gt*Cel45A was crystallized and its molecular structure solved at 1.3Å resolution. *Gt*Cel45A has a globular structure, with approximating dimensions of 26 Å x 37 Å x 48 Å. It contains a six-stranded beta barrel with a double-psi beta barrel (DPBB), which affirms the belonging to GH45 subfamily. The DPBB core is surrounded by short helices and long loops. An open substrate binding groove spans across the surface and is 48 Å in length, with the widest part of around 14 Å near the centre and depth around 12 Å.

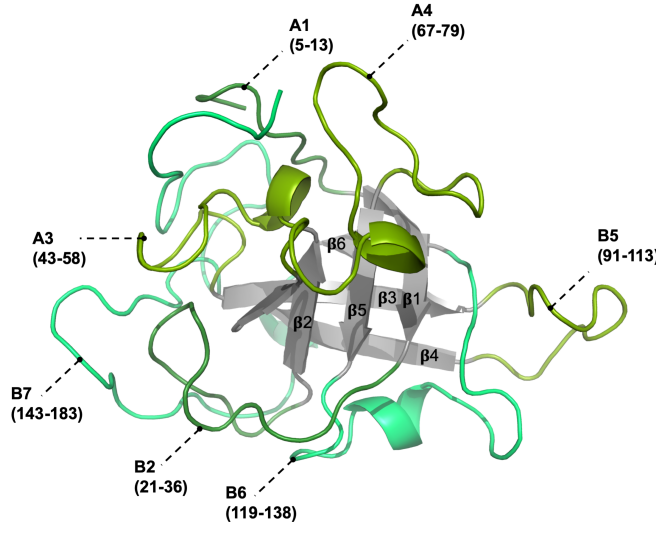


Figure 12. Cartoon representation of the molecular structure of *Gt*Cel45A.

The first 4 amino acids that are present in the mature protein sequence (LEER) were not visible in the crystal structure. The structure model begins with loop A1, which contains a three amino acid long alpha-helix. Five amino acids connect the helix to the first beta strand (beta1), which is connected to the next beta strand (beta2) via 15 amino acid loop (B2). The secondary structure pattern of a helix – beta-strand – beta-strand is repeated once more, where at least 12

amino acid long loops connect the secondary structure elements. This pattern is followed by two repetitions of helix – beta-strands. Two helices and long loops make up the C-terminus of the structure (Figure 12).

Structure comparison to members of subfamily C

*Gt*Cel45A is 82% sequence identical to *Pc*Cel45A. *Gt*Cel45A is one amino acid shorter than *Pc*Cel45A – beta strand B4 has one histidine residue (His80) where *Pc*Cel45A has two residues (Gly76 and Gln77). The unique residues in *Gt*Cel45A are not positioned in the substrate binding groove or the catalytic site. The overall structure of *Gt*Cel45A and *Pc*Cel45A is nearly identical.

Subfamily C enzymes are usually one domain proteins, but in our sequence comparisons we noted at least one subfamily C member to have a CBM. *Gymnopilus dilepis* Cel45A (GenBank: PPQ98991.1) (Reynolds et al., 2018) is 55% sequence identical to *Gt*Cel45A and additionally has a CBM1 domain attached with a linker to the GH45 domain at the N-terminus. The catalytic GH45 domain of *G. dilepis* Cel45A (*Gd*Cel45A) has a very high (75%) sequence identity to *Gt*Cel45A.

4.2 Enzyme activity studies

All GH45 enzymes have been regarded as inverting hydrolases, but before our studies, to our knowledge, it had been experimentally proven only for subfamily A member Cel45A from *H. insolens* (Schou et al., 1993). The NMR analyses described in Paper I provided an experimental confirmation of the inverting nature of Cel45A from *M. edulis, T. reesei* and *P. chrysosporium*, thus proving that this mechanism indeed is most likely common among all GH45 subfamilies known to date.

The enzymatic activity of GH45 members has been investigated on various cellulosic and hemicellulosic substrates (PASC, Avicel, bacterial cellulose, CMC, xyloglucan, xylan, arabinoxylan, mannan, glucomannan, betaglucan, lichenan) but the exact activity profile remains unclear. Qualitative and quantitative assays have been carried out under conditions varying in substrate concentration, buffer pH, incubation time (from 1 min to 24 h) and temperature (4-70°C),

therefore reaching an activity consensus becomes problematic. It is clear, however, that all GH45 enzymes tested so far, are active on CMC and barley betaglucan (Berto et al., 2019; Cha et al., 2018; Karlsson et al., 2002; Okmane et al., 2022; Vlasenko et al., 2010).
Mainly three methods were utilized in our enzyme activity studies: reducing sugar assay with PHBAH reagent, solution state NMR spectroscopy, and HPAEC.

4.2.1 Enzymatic activity divergence across GH45 (Paper I)

The innate division of GH45 enzymes into three subfamilies (A, B and C) displays itself not only in sequence-based phylogenetic analyses but also the respective hydrolytic activities (Figure 13).

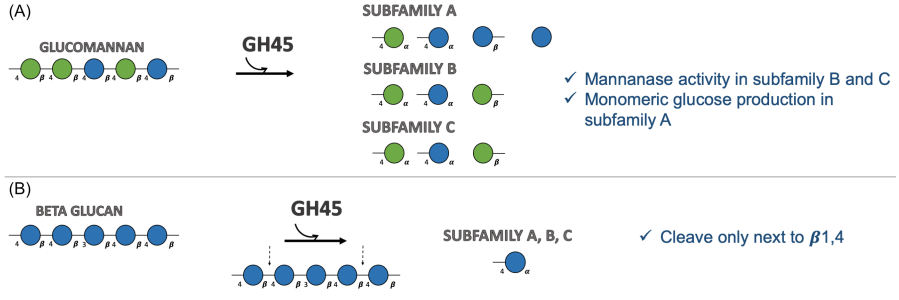


Figure 13. Graphical summary of enzymatic activities determined in Paper I.

Barley betaglucan

Reducing sugar assays revealed higher initial activity rates on barley betaglucan than CMC. The initial rates were 1.6 to 7.5 times higher in comparison. Initial rates (µM/µM)min^{-1} ±SD on 0.1% CMC and 0.1% BG

for the subfamily A member *Hi*Cel45A were 235 ±9.9 and 1139 ±58.5 (µM/µM)min-^{1}, respectively. In comparison to *Hi*Cel45A, the subfamily B members exhibited significantly lower initial rates of 183 ±9.7 and 296 ±43.5 (µM/µM)min^{-1} (*Me*Cel45A), 12 ±4.1 and 90±37.3 (µM/µM)min^{-1} (*Tr*Cel45A), respectively. The subfamily C member *Pc*Cel45A was more alike subfamily B in regard to activity rate, yet exhibited the same dynamic of higher activities on barley betaglucan, 56 ±5.0 and 109±58.6 (µM/µM)min^{-1} respectively.

NMR experiments on barley betaglucan pointed to certain cleavage site restrictions – Cel45A enzymes required a β(1→4) linkage between subsites -2 and -1 to hydrolyze barley betaglucan. However, further studies are necessary to draw conclusions about restrictions for other subsites.

Konjac glucomannan

NMR spectroscopy revealed the distinctiveness of GH45 subfamilies by expanding on the multiplicity of cleavage preference regarding glycosidic linkage type.

The subfamily A member, *Hi*Cel45A, was able to cleave β(1→4) linkage between mannose and glucose (Man-Glc), while exhibiting a preference towards cleavage of linkages joining two glucose residues (Glc-Glc). *Hi*Cel45A was also unique in its ability to produce monomeric glucose from glucomannan. The product profile of *Hi*Cel45A on glucomannan did not contain non-reducing end mannose, contrary to *Tr*Cel45A, *Me*Cel45A and *Pc*Cel45A that displayed a preference towards cleavage between glucose and mannose residues (Glc-Man). Furthermore, all GH45 members with the exception of *Hi*Cel45A facilitated slow formation of reducing-end mannose accompanied by an increase of non-reducing end mannose on glucomannan, which pointed to cleavage between two mannose residues (Man-Man), therefore a mannanase activity in subfamily B and C members. All of the Cel45A enzymes were active against glucomannan, but Cel45A from *T. reesei* was distinguished in its ability to proceed with glucomannan degradation even after 24-hour incubation mark. This confirmed the previously suggested mannanase activity in *T. reesei* (Karlsson et al., 2002).

CMC

All GH45 enzymes have been proven to be catalytically active on CMC. Our studies showed comparatively slower hydrolysis rates and lower reducing end yields than on betaglucan or glucomannan. Interestingly, this pattern is not observed across studies, particularly in reference to subfamily C. Multiple studies show higher product yields on CMC compared to betaglucan in subfamily C (Amengual et al., 2022; Berto et al., 2019; Cha et al., 2018). Subfamily C members have been studied under multiple incubation settings with aeration ranging

from 100 to 1200 rpm, different incubation times with discoveries of optimal pH values of 4.0 – 6.5. Interestingly, subfamily C enzymes prove to be rather thermostable, retaining high activities at temperatures as high as 70°C (Cha et al., 2018). In addition to varying experimental conditions across studies, the variance in hydrolytic activity on CMC could presumably be attributed to a discrepancy in the degree of carboxymethyl substitutions on CMC across studies, which can range from 0.60 to 0.95. One may hypothesize that the presence of substituents on glucose residues could hinder the docking of the carbohydrate chain in the optimal position for hydrolysis. However, in our studies *Hi*Cel45A gave comparatively higher reducing end yield on CMC, suggesting greater tolerance for substitutions regardless of possessing a more enclosed active site than the other enzymes (Paper I).

4.2.2 Dissimilarities within a subfamily (Paper I, II)

To gain further insight in to enzymatic activity variation at a subfamily level, two subfamily B enzymes, *Me*Cel45A and *Tr*Cel45A, were incubated with cellohexaose.

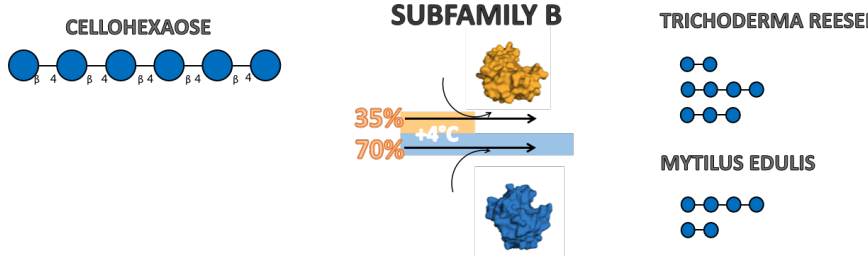


Figure 14. Schematic representation of hydrolytic activity variance within subfamily B, as determined in Paper I.

*Me*Cel45A and *Tr*Cel45A were subjected to low temperature (+4°C) with the aim to observe any variance in enzymatic activity. Considering that *Me*Cel45A is an enzyme from a blue mussel – a mollusc which resides in a boreal habitat – and *Tr*Cel45A is isolated from a fungus which resides in a tropical climate and in its natural habitat grows on rotting material, we set out to test for cold adaptivity in *Me*Cel45A.

Perhaps not surprisingly, the enzyme from blue mussel retained over 70% of its activity at +4°C relative to +30°C, while Cel45A from *T. reesei* retained 35%, which could indicate GH45 enzyme evolution towards cold adaptation in the mollusc (Figure 14). Xu et al. (Xu et al., 2000) had previously demonstrated a broad optimum temperature range for the *Me*Cel45A, which points to the necessity of further investigation of how the structure of this enzyme has evolved to achieve thermostability.

To investigate substrate binding groove characteristics, we looked at the product profile after cellohexaose cleavage. The cellohexaose cleavage product profile, as analysed with HPAEC, differed between the two enzymes. Predominantly cellobiose and cellotetraose were produced by *Me*Cel45A, while *Tr*Cel45A produced cellotriose in addition to cellobiose and cellotetraose. These results imply a more distinct substrate organization in the active site cleft in *Me*Cel45A in comparison to *Tr*Cel45A.

Non-uniform catalytic activities were noted also in subfamily C. In Paper II we compared *Gt*Cel45A and *Pc*Cel45A for activity on barley beta glucan. Betaglucan degradation was nearly 4 times more rapid with *Gt*Cel45A than *Pc*Cel45A. This comparatively higher activity was previously reported in another study which, in addition to higher hydrolysis rate, observed the formation of glucose as a product at 50°C. Such activities seem unlikely for proteins with wide substrate binding grooves, but conformation change upon subjection to higher temperatures cannot be dismissed.

4.3 Additional studies on GH45 enzymes and homologous proteins

NMR spectroscopy studies on $^{13}C,^{15}N$-labelled PcCel45A (Paper III)

To study *Pc*Cel45A interaction with cello-oligosaccharides, a single labelled ^{15}N-*Pc*Cel45A and a double labelled $^{13}C,^{15}N$-*Pc*Cel45A was produced and an amino acid backbone resonance assignment was made (Paper III) (Figure 15). $^{1}H,^{15}N$-HSQC experiments were carried out where ^{15}N-*Pc*Cel45A was titrated with 0-100 and 0-10 molar ratios of cellobiose and cellohexaose,

respectively. Backbone amide chemical shift differences were considered as an indication for substrate interactions.

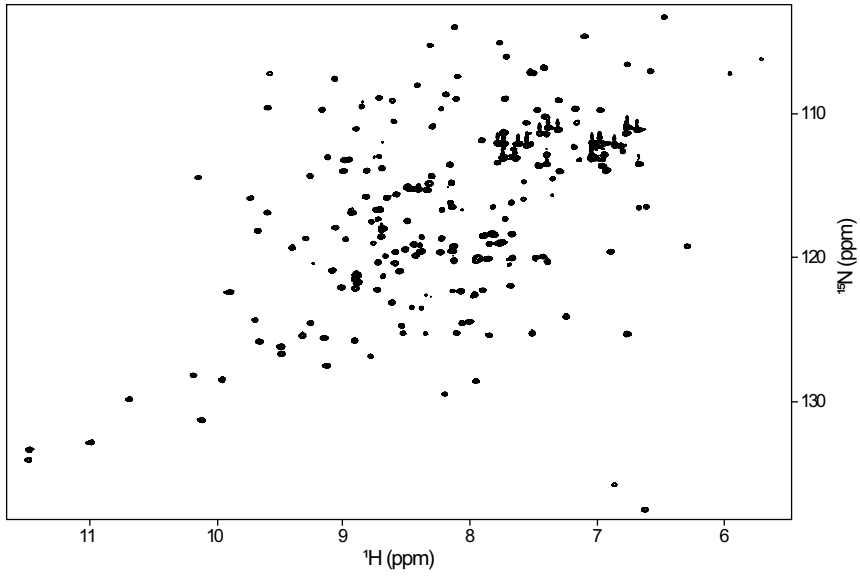


Figure 15. $^{1}H,^{15}N$-HSQC spectrum of PcCel45A.

The following amino acids underwent a significant change upon addition of cellobiose: T16-S19, A36, Q39, Y67, C87, Q90-G97, H107, D114-I115, and G131. The cellobiose interaction was in fast exchange on the NMR time-scale and the overall affinity towards cellobiose was low. The chemical shift differences were visualized in PyMOL. Comparison of the binding area to the available crystal structure of a *Pc*Cel45A-cellobiose complex indicated that the backbone amides which had a significant change in chemical shift correspond to the same amino acids that interact with the cellobiose complex in the crystal structure.

Analysis of the ^{15}N-*Pc*Cel45A interaction with cellohexaose was a lot more complex. The involved amino acids appeared as slow, intermediate and fast chemical exchange with the cellohexaose. Many more chemical shift changes were observed upon addition of cellohexaose compared to cellobiose. The majority of the substrate binding groove was affected, as well as the backbone amides of amino acids 180° from the substrate binding area. The affinity to

cellohexaose was determined to be high, but due to amino acids exhibiting all types of chemical exchange, the dissociation constant (K_d) could not easily be calculated. In order to assign NH cross-peaks of the cellohexaose complex in slow exchange, backbone assignment of ^{13}C,^{15}N-*Pc*Cel45A in a complex with cellohexaose will be required.

Hydrolytic activity of PcCel45A under mechanical stress (unpublished)

The action mechanism of expansins is assumed to be directed towards load-bearing structures in the plant CW, structures which undergo mechanical stress. *Pc*Cel45A exhibits relatively low hydrolytic activities and we explored the possibility of activity increase upon mechanic impact, such as the stretching of cellulose microfibrils. We utilized a system for nanocellulose generation developed in house at The Royal Institute of Technology (KTH), which involves processing of pre-treated pulp (TEMPO pulp) through a capillary tube. Bacterial cellulose (PASC) and TEMPO pulp were loaded in to a homogenizer with and without *Pc*Cel45A. The formation of reducing ends was measured in samples processed in the homogenizer (with/without *Pc*Cel45A), un-treated samples with/without *Pc*Cel45A and a sample subjected to combined conditions – post-homogenization treatment with *Pc*Cel45A. The samples were processed through homogenizer for up to 15 minutes.

Not surprisingly, mechanical treatment of cellulosic substrate subjected to additional enzymatic digestion lead to higher concentration of reducing ends, however simultaneous homogenization process (*Pc*Cel45A present in the sample undergoing homogenization) caused the highest concentration of reducing ends, increasing with an increase in incubation time (Figure 16). These results served as an experimental indication that mechanical stretching of microfibrils indeed aids in hydrolytic disruption.

The value distribution in the non-homogenized samples was quite high due to heterogeneity of the sample suspensions, therefore more repetitions were necessary. Unfortunately, these experiments were not practically feasible, as the minimum volumes of sample solutions subjected to homogenization were in hundreds of millilitres and the require enzyme amounts were too high.

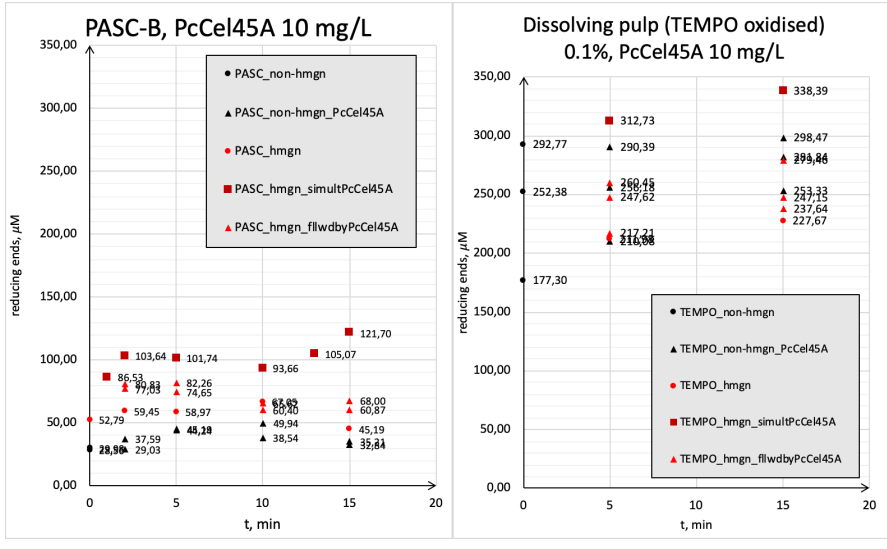


Figure 16. Hydrolytic activity of *Pc*Cel45A on bacterial PASC and pulp, with and without subjection to mechanical stress.

Studies on loosenins and loosenin-like proteins (unpublished)

Production and purification methods were developed for small (~11 kDa) non-hydrolytic proteins – a loosenin from *B. adusta* (*Ba*LOOS1) and a loosenin-like protein from *T. reesei* (*Tr*LOOL1).

Crystallization trials were carried out for *Ba*LOOS1, *Tr*LOOL1 and four LOOL proteins received from collaborators at Aalto University, Finland. Crystals were obtained of *Ba*LOOS1 and one of the LOOL proteins (Figure 17), but none of the obtained crystals diffracted. Homology models were made for structure studies.

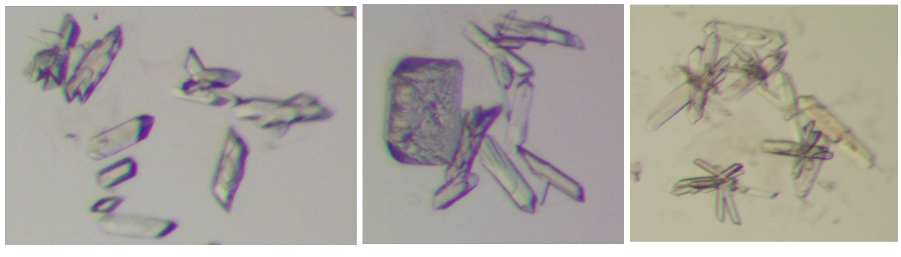


Figure 17. LOOL12 crystals from crystallization trials.

Possible enzymatic activity was evaluated for *Ba*LOOS1 and *Tr*LOOL1. *Ba*LOOS1 was incubated with cellohexaose and product

profile was analysed with HPAEC. Cellohexaose concentration in the sample remained steady, *Ba*LOOS1 exhibited no catalytic activity towards the degradation of cellohexaose.

No TrLOOL1 activity was detected on any of the tested substrates (2-HE-cellulose, barley betaglucan, yeast betaglucan, curdlan, pachyman, pullan, dextran, chitosan, xyloglucan, xylan, amylopectin, amylose, pectic galactan, arabinan, galactomannan, rhamnogalacturonan).

A trial run was carried out for the production and purification of a double labelled $^{13}C,^{15}N$-*Tr*LOOL1, a $^{1}H,^{15}N$-HSQC spectrum was generated as a result, but the protein concentration was too low for a successful backbone assignment from the trial run (Figure 18).

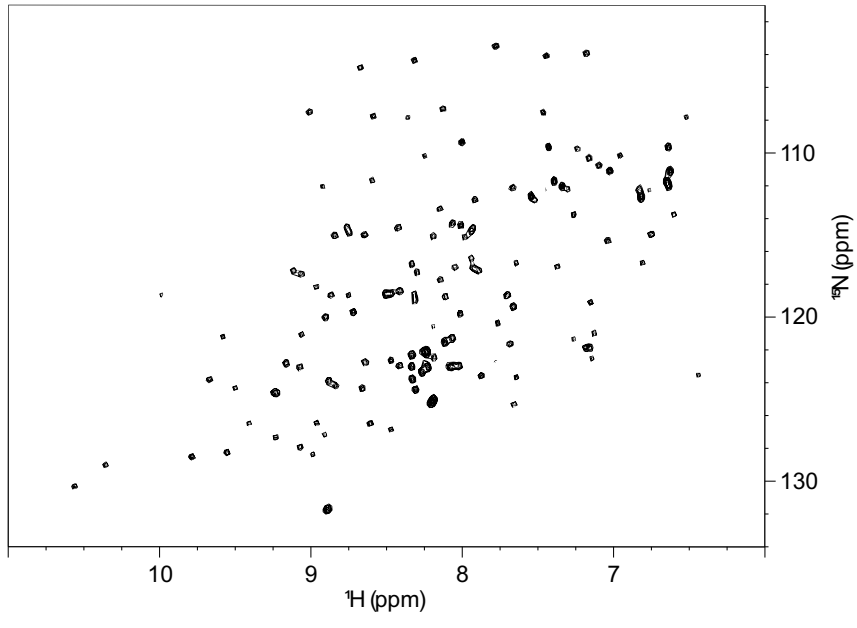


Figure 18. $^{1}H,^{15}N$-HSQC spectrum of *Tr*LOOL1.

5. Conclusions and outlook

This study showed that GH45 enzymes differ significantly regarding enzymatic activity rates and the architecture of substrate binding grooves, while exhibiting remarkable overall structural resemblance. Certain GH45 subfamilies can be distinguished based on amino acid sequence, hydrolytic activity rates, substrate and product profile or the surface structure of substrate binding area. Information about the catalytic site residues can facilitate subfamily distinction at the sequence level – the absence of catalytic base residue distinguishes subfamily C, the presence of aspartic acid residue at the assisting residue position distinguishes subfamily A. Our studies showed subfamily A to be uniquely able to produce glucose monomers from glucomannan. Therefore, glucose production on glucomannan can be an indication of belonging to subfamily A. Inability to cleave cello-oligomers with less than six glucose units indicates subfamily C. An open and wide substrate binding groove points to subfamily C.

More studies on the GH45 enzymatic activities are required to reach a clearer consensus. In my opinion, the previous enzyme activity investigations have mostly been carried out in too different experimental conditions for the results to be directly comparable. The future activity studies would benefit from additional NMR spectroscopy experiments, as here we showed this method to be an impressive tool in analysing the hydrolysis products, even leading to the affirmation of the inverting nature of GH45 enzymes.

We showed that GH45 can undergo evolutionary adaptations, such as cold adaptation in the Cel45A from blue mussel, but further research is required to discover the intricacies of cold adaptation in GH45 enzymes.

The structural resemblance to loosenins and the role of loosenins, in my opinion, should be investigated further. If loosenins have become fungal homologues to plant expansins, do they partake in cell wall loosening in fungal cells as well or is this plant cell wall loosening mechanism hijacked by pathogens and turned in to a weapon against the evolutionary host? Unfortunately, researching loosenins is not an easy task, which can be intuitively discerned from the near absence of research in this area. Currently, there are 5 published studies which mention loosenins, but only 2 have loosenins as the main focus. Study by researchers at the Autonomous University of the State of Morelos, Mexico (Quiroz-Castañeda et al., 2011), which coined the term loosenin, describes a loosenin isolated from the white-rot fungus *Bjerkandera adusta* with a capability to loosen cotton fibres. A recent study on the substrate binding of loosenins, demonstrated loosenin binding preference for chitin, suggesting a function in fungal cell wall morphogenesis (Monschein et al., 2023). Development of new methods for action mechanism determination would be beneficial in order to study non-hydrolytic proteins, as an absence of activity is a result difficult to publish. Also, the absolute lack of published loosenin molecular structures points to difficulties involved in achieving this task. Based on my experience with crystallization of these proteins, I believe it requires daring, dedication and a creative approach. However, since loosenins are very small, typically only 110-130 amino acids, i.e. even smaller than GH45s, perhaps structure determination by NMR may instead be the road forward. The success in double labelling of a loosenin from *T. reesei* has taken us one step closer in discovering the structure of these proteins and currently is a work in progress in our group.

As mentioned in the subchapter about cellulose, the current barriers towards a cost-efficient system for second-generation biofuel production are related to the recalcitrance of lignocellulose. The costs for overcoming the recalcitrance must be lowered which might be achieved by introducing cellulose loosening agents that are non-disruptive towards the cellulases utilized in feedstock degradation. Non-hydrolytic proteins, such as expansins or loosenins, could be utilized to alleviate cellulase access to cellulose fibrils. Synergy experiments with cellulases and expansins have shown the mixtures to be effective enough to increase cellulose degradation rates 2.4 times

(Seki et al., 2015). Further steps towards achieving this goal would involve searching for the middle ground between efficacy and efficiency, where cost-effective additives are mixed with cellulase enzymes in ideal proportions to achieve a fast and complete degradation.

References

Albenne, C., Canut, H., Hoffmann, L., & Jamet, E. (2014). Plant Cell Wall Proteins: A Large Body of Data, but What about Runaways? *Proteomes*, *2*(2), 224–242. https://doi.org/10.3390/proteomes2020224

Alexandersson, E., & Nestor, G. (2022). Complete 1H and 13C NMR spectral assignment of d-glucofuranose. *Carbohydrate Research*, *511*, 108477. https://doi.org/10.1016/j.carres.2021.108477

Amengual, N. G., Csarman, F., Wohlschlager, L., & Ludwig, R. (2022). Expression and characterization of a family 45 glycosyl hydrolase from Fomitopsis pinicola and comparison to Phanerochaete chrysosporium Cel45A. *Enzyme and Microbial Technology*, *156*, 110000. https://doi.org/10.1016/j.enzmictec.2022.110000

Anasontzis, G. E., Salazar Penã, M., Spadiut, O., Brumer, H., & Olsson, L. (2014). Effects of temperature and glycerol and methanol-feeding profiles on the production of recombinant galactose oxidase in Pichia pastoris. *Biotechnology Progress*, *30*(3), 728–735. https://doi.org/10.1002/btpr.1878

Arja, M.-O. (2007). Cellulases in the Textile Industry. In J. Polaina & A. P. MacCabe (Eds.), *Industrial Enzymes: Structure, Function and Applications* (pp. 51–63). Springer Netherlands. https://doi.org/10.1007/1-4020-5377-0_4

Atalla, R. H., & VanderHart, D. L. (1984). Native Cellulose: A Composite of Two Distinct Crystalline Forms. *Science*, *223*(4633), 283–285.

Bajpai, P. (1999). Application of Enzymes in the Pulp and Paper Industry. *Biotechnology Progress*, *15*(2), 147–157. https://doi.org/10.1021/bp990013k

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Basak, P., Adhikary, T., Das, P., Shee, M., Dutta, T., Biswas, S., Paul, S., & Manna, S. (2021). Chapter 13 - Cellulases in paper and pulp, brewing and food industries: Principles associated with its diverse applications. In D. K. Tuli & A. Kuila (Eds.), *Current Status and Future Scope of Microbial Cellulases* (pp. 275–293). Elsevier. https://doi.org/10.1016/B978-0-12-821882-2.00003-X

Baumann, K., Dato, L., Graf, A. B., Frascotti, G., Dragosits, M., Porro, D., Mattanovich, D., Ferrer, P., & Branduardi, P. (2011). The impact of oxygen on the transcriptome of recombinant S. cerevisiae and P. pastoris—A

comparative analysis. *BMC Genomics*, *12*(1), 218. https://doi.org/10.1186/1471-2164-12-218

Berlemont, R., & Martiny, A. C. (2015). Genomic Potential for Polysaccharide Deconstruction in Bacteria. *Applied and Environmental Microbiology*, *81*(4), 1513–1519. https://doi.org/10.1128/AEM.03718-14

Berto, G. L., Velasco, J., Tasso Cabos Ribeiro, C., Zanphorlin, L. M., Noronha Domingues, M., Tyago Murakami, M., Polikarpov, I., de Oliveira, L. C., Ferraz, A., & Segato, F. (2019). Functional characterization and comparative analysis of two heterologous endoglucanases from diverging subfamilies of glycosyl hydrolase family 45. *Enzyme and Microbial Technology*, *120*, 23–35. https://doi.org/10.1016/j.enzmictec.2018.09.005

Bharadwaj, V. S., Knott, B. C., Ståhlberg, J., Beckham, G. T., & Crowley, M. F. (2020). The hydrolysis mechanism of a GH45 cellulase and its potential relation to lytic transglycosylase and expansin function. *Journal of Biological Chemistry*, jbc.RA119.011406. https://doi.org/10.1074/jbc.RA119.011406

Blanchette, R. A. (1991). *Delignification by Wood-Decay Fungi*. *September*(29), 381–398. https://doi.org/10.1146/annurev.py.29.090191.002121

Bonoli, M., Marconi, E., & Caboni, M. F. (2004). Free and bound phenolic compounds in barley (Hordeum vulgare L.) flours. Evaluation of the extraction capability of different solvent mixtures and pressurized liquid methods by micellar electrokinetic chromatography and spectrophotometry. *Journal of Chromatography. A*, *1057*(1–2), 1–12. https://doi.org/10.1016/j.chroma.2004.09.024

Bootten, T. J., Harris, P. J., Melton, L. D., & Newman, R. H. (2004). Solid-state 13C-NMR spectroscopy shows that the xyloglucans in the primary cell walls of mung bean (Vigna radiata L.) occur in different domains: A new model for xyloglucan–cellulose interactions in the cell wall. *Journal of Experimental Botany*, *55*(397), 571–583. https://doi.org/10.1093/jxb/erh065

Bootten, T. J., Harris, P. J., Melton, L. D., & Newman, R. H. (2009). Solid-State 13C NMR Study of a Composite of Tobacco Xyloglucan and Gluconacetobacter xylinus Cellulose: Molecular Interactions between the Component Polysaccharides. *Biomacromolecules*, *10*(11), 2961–2967. https://doi.org/10.1021/bm900762m

Brotman, Y., Briff, E., Viterbo, A., & Chet, I. (2008). Role of Swollenin, an Expansin-Like Protein from Trichoderma, in Plant Root Colonization. *Plant Physiology*, *147*(2), 779–789. https://doi.org/10.1104/pp.108.116293

Buckeridge, M., Rayon, C., Urbanowicz, B., Tiné, M., & Carpita, N. (2004). Mixed Linkage (1→3),(1→4)-β-D-Glucans of Grasses. *Cereal Chemistry - CEREAL CHEM*, *81*, 115–127. https://doi.org/10.1094/CCHEM.2004.81.1.115

Cha, J.-H., Yoon, J.-J., & Cha, C.-J. (2018). Functional characterization of a thermostable endoglucanase belonging to glycoside hydrolase family 45 from Fomitopsis palustris. *Applied Microbiology and Biotechnology*, *102*(15), 6515–6523. https://doi.org/10.1007/s00253-018-9075-5

Charova, S. N., Dölfors, F., Holmquist, L., Moschou, P. N., Dixelius, C., & Tzelepis, G. (2020). The RsRlpA Effector Is a Protease Inhibitor Promoting Rhizoctonia solani Virulence through Suppression of the Hypersensitive Response. *International Journal of Molecular Sciences*, *21*(21), Article 21. https://doi.org/10.3390/ijms21218070

Cheng, J. J., & Timilsina, G. R. (2011). Status and barriers of advanced biofuel technologies: A review. *Renewable Energy*, *36*(12), 3541–3549. https://doi.org/10.1016/j.renene.2011.04.031

Clark, L., Dikiy, I., Rosenbaum, D. M., & Gardner, K. H. (2018). On the use of Pichia pastoris for isotopic labeling of human GPCRs for NMR studies. *Journal of Biomolecular NMR*, *71*(4), 203–211. https://doi.org/10.1007/s10858-018-0204-3

Cosgrove, D. J. (2000). Loosening of plant cell walls by expansins. *Nature*, *407*(6802), 321–326. https://doi.org/10.1038/35030000

Cosgrove, D. J. (2015). Plant expansins: Diversity and interactions with plant cell walls. *Current Opinion in Plant Biology*, *25*, 162–172. https://doi.org/10.1016/j.pbi.2015.05.014

Courtade, G., Wimmer, R., Dimarogona, M., Sandgren, M., Eijsink, V. G. H., & Aachmann, F. L. (2016). Backbone and side-chain 1H, 13C, and 15N chemical shift assignments for the apo-form of the lytic polysaccharide monooxygenase NcLPMO9C. *Biomolecular NMR Assignments*, *10*(2), 277–280. https://doi.org/10.1007/s12104-016-9683-x

Davies, G. J., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C., & Schülein, M. (1995). Structures of oligosaccharide-bound forms of the endoglucanase V from Humicola insolens at 1.9 A resolution. *Biochemistry*, *34*(49), 16210–16220. https://doi.org/10.1021/bi00049a037

Demuner, B. J., Pereira Junior, N., & Antunes, A. M. S. (2011). Technology Prospecting on Enzymes for the Pulp and Paper Industry. *Journal of Technology Management & Innovation*, *6*(3), 148–158. https://doi.org/10.4067/S0718-27242011000300011

Eibinger, M., Sigl, K., Sattelkow, J., Ganner, T., Ramoni, J., Seiboth, B., Plank, H., & Nidetzky, B. (2016). Functional characterization of the native swollenin from Trichoderma reesei: Study of its possible role as C1 factor of enzymatic lignocellulose conversion. *Biotechnology for Biofuels*, *9*(1), 178. https://doi.org/10.1186/s13068-016-0590-2

Eichorst, S. A., & Kuske, C. R. (2012). Identification of Cellulose-Responsive Bacterial and Fungal Communities in Geographically and Edaphically Different Soils by Using Stable Isotope Probing. *Applied and Environmental Microbiology*, *78*(7), 2316–2327. https://doi.org/10.1128/AEM.07313-11

Eklöf, J. K., Ekerholm, H., & Mårald, E. (2012). Promoting Ethanol in the Shadow of Oil Dependence: 100 years of arguments and frictions in Swedish politics. *Scandinavian Journal of History*, *37*(5), 621–645. https://doi.org/10.1080/03468755.2012.725677

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*, 1034. https://doi.org/10.1126/science.1153213

Ferrari, S., Savatin, D. V., Sicilia, F., Gramegna, G., Cervone, F., & Lorenzo, G. D. (2013). Oligogalacturonides: Plant damage-associated molecular patterns and regulators of growth and development. *Frontiers in Plant Science*, *4*, 49. https://doi.org/10.3389/fpls.2013.00049

Gaderer, R., Bonazza, K., & Seidl-Seiboth, V. (2014). Cerato-platanins: A fungal protein family with intriguing properties and application potential. *Applied Microbiology and Biotechnology*, *98*(11), 4795–4803. https://doi.org/10.1007/s00253-014-5690-y

Godoy, A. S., Pereira, C. S., Ramia, M. P., Silveira, R. L., Camilo, C. M., Kadowaki, M. A., Lange, L., Busk, P. K., Nascimento, A. S., Skaf, M. S., & Polikarpov, I. (2018). Structure, computational and biochemical analysis of Pc Cel45A endoglucanase from Phanerochaete chrysosporium and catalytic mechanisms of GH45 subfamily C members. *Scientific Reports*, *8*(1), 3678. https://doi.org/10.1038/s41598-018-21798-9

Goodell, B., Qian, Y., & Jellison, J. (2008). Fungal Decay of Wood: Soft Rot—Brown Rot—White Rot. In *Development of Commercial Wood Preservatives* (Vol. 982, pp. 9–31). American Chemical Society. https://doi.org/10.1021/bk-2008-0982.ch002

Guo, R., Ding, M., Zhang, S.-L., Xu, G., & Zhao, F. (2008). Molecular cloning and characterization of two novel cellulase genes from the mollusc Ampullaria crossean. *Journal of Comparative Physiology B*, *178*(2), 209–215. https://doi.org/10.1007/s00360-007-0214-z

Guo, X., Liu, N., Zhang, Y., & Chen, J. (2022). Pathogen-Associated Molecular Pattern Active Sites of GH45 Endoglucanohydrolase from Rhizoctonia solani. *Phytopathology*, *112*(2), 355–363. https://doi.org/10.1094/PHYTO-04-21-0164-R

Gupta, P., & De, B. (2017). Differential responses of cell wall bound phenolic compounds in sensitive and tolerant varieties of rice in response to salinity. *Plant Signaling & Behavior*, *12*(10), e1379643. https://doi.org/10.1080/15592324.2017.1379643

Hahn, Th. (2002). Graphical symbols for symmetry elements in one, two and three dimensions. In Th. Hahn (Ed.), *International Tables for Crystallography Volume A: Space-group symmetry* (pp. 7–11). Springer Netherlands. https://doi.org/10.1107/97809553602060000503

Himmel, M. E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R., & Bayer, E. A. (2010). Microbial enzyme systems for biomass conversion: Emerging paradigms. *Biofuels*, *1*(2), 323–341. https://doi.org/10.4155/bfs.09.25

Hoang, T.-D., & Nghiem, N. (2021). Recent Developments and Current Status of Commercial Production of Fuel Ethanol. *Fermentation*, *7*(4), Article 4. https://doi.org/10.3390/fermentation7040314

Hu, J., Arantes, V., Pribowo, A., & Saddler, J. N. (2013). The synergistic action of accessory enzymes enhances the hydrolytic potential of a "cellulase mixture" but is highly substrate specific. *Biotechnology for Biofuels*, *6*(1), 112. https://doi.org/10.1186/1754-6834-6-112

Huet, J., Teinkela Mbosso, E. J., Soror, S., Meyer, F., Looze, Y., Wintjens, R., & Wohlkönig, A. (2013). High-resolution structure of a papaya plant-defence barwin-like protein solved by in-house sulfur-SAD phasing. *Acta Crystallographica Section D: Biological Crystallography*, *69*(10), Article 10. https://doi.org/10.1107/S0907444913018015

Igarashi, K., Ishida, T., Hori, C., & Samejima, M. (2008). Characterization of an Endoglucanase Belonging to a New Subfamily of Glycoside Hydrolase Family 45 of the Basidiomycete Phanerochaete chrysosporium. *Appl. Environ. Microbiol.*, *74*(18), 5628–5634. https://doi.org/10.1128/AEM.00812-08

Jahr, H., Dreier, J., Meletzus, D., Bahro, R., & Eichenlaub, R. (2000). The Endo-β-1,4-glucanase CelA of Clavibacter michiganensis subsp. Michiganensis Is a Pathogenicity Determinant Required for Induction of Bacterial Wilt of Tomato. *Molecular Plant-Microbe Interactions®*, *13*(7), 703–714. https://doi.org/10.1094/MPMI.2000.13.7.703

Karlsson, J., Siika-aho, M., Tenkanen, M., & Tjerneld, F. (2002). Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of Trichoderma reesei. *Journal of Biotechnology*, *99*(1), 63–78. https://doi.org/10.1016/S0168-1656(02)00156-6

Kerff, F., Amoroso, A., Herman, R., Sauvage, E., Petrella, S., Filée, P., Charlier, P., Joris, B., Tabuchi, A., Nikolaidis, N., & Cosgrove, D. J. (2008). Crystal structure and activity of Bacillus subtilis YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(44), 16876–16881. https://doi.org/10.1073/pnas.0809382105

Kikuchi, T., Jones, J. T., Aikawa, T., Kosaka, H., & Ogura, N. (2004). A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode

Bursaphelenchus xylophilus. *FEBS Letters*, *572*(1–3), 201–205. https://doi.org/10.1016/j.febslet.2004.07.039

Kirui, A., Ling, Z., Kang, X., Dickwella Widanage, M. C., Mentink-Vigier, F., French, A. D., & Wang, T. (2019). Atomic resolution of cotton cellulose structure enabled by dynamic nuclear polarization solid-state NMR. *Cellulose*, *26*(1), 329–339. https://doi.org/10.1007/s10570-018-2095-6

Kubicki, J. D., Yang, H., Sawada, D., O'Neill, H., Oehme, D., & Cosgrove, D. (2018). The Shape of Native Plant Cellulose Microfibrils. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-32211-w

Kudla, U., Qin, L., Milac, A., Kielak, A., Maissen, C., Overmars, H., Popeijus, H., Roze, E., Petrescu, A., Smant, G., Bakker, J., & Helder, J. (2005). Origin, distribution and 3D-modeling of Gr-EXPB1, an expansin from the potato cyst nematode Globodera rostochiensis. *FEBS Letters*, *579*(11), 2451–2457. https://doi.org/10.1016/j.febslet.2005.03.047

Lange, L., Barrett, K., Pilgaard, B., Gleason, F., & Tsang, A. (2019). Enzymes of early-diverging, zoosporic fungi. *Applied Microbiology and Biotechnology*, *103*(17), 6885. https://doi.org/10.1007/s00253-019-09983-w

Lee, S. J., Kim, S. R., Yoon, H. J., Kim, I., Lee, K. S., Je, Y. H., Lee, S. M., Seo, S. J., Dae Sohn, H., & Jin, B. R. (2004). CDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, Apriona germari. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *139*(1), 107–116. https://doi.org/10.1016/j.cbpc.2004.06.015

Lever, M. (1972). A new reaction for colorimetric determination of carbohydrates. *Analytical Biochemistry*, *47*(1), 273–279. https://doi.org/10.1016/0003-2697(72)90301-6

Lever, M., Walmsley, T. A., Visser, R. S., & Ryde, S. J. (1984). Optimal conditions for 4-hydroxybenzoyl- and 2-furoylhydrazine as reagents for the determination of carbohydrates, including ketosamines. *Analytical Biochemistry*, *139*(1), 205–211. https://doi.org/10.1016/0003-2697(84)90406-8

Levy, J. (1966). The Soft Rot Fungi: Their Mode of Action and Significance in the Degradation of Wood. In R. D. Preston (Ed.), *Advances in Botanical Research* (Vol. 2, pp. 323–357). Academic Press. https://doi.org/10.1016/S0065-2296(08)60253-0

Lin, Y., & Tanaka, S. (2006). Ethanol fermentation from biomass resources: Current state and prospects. *Applied Microbiology and Biotechnology*, *69*(6), 627–642. https://doi.org/10.1007/s00253-005-0229-x

Lü, P., Kang, M., Jiang, X., Dai, F., Gao, J., & Zhang, C. (2013). RhEXPA4, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to Arabidopsis. *Planta*, *237*(6), 1547–1559. https://doi.org/10.1007/s00425-013-1867-3

Lygin, A. V., Upton, J., Dohleman, F. G., Juvik, J., Zabotina, O. A., Widholm, J. M., & Lozovaya, V. V. (2011). Composition of cell wall phenolics and polysaccharides of the potential bioenergy crop –Miscanthus. *GCB Bioenergy*, *3*(4), 333–345. https://doi.org/10.1111/j.1757-1707.2011.01091.x

McGuire, K. L., & Treseder, K. K. (2010). Microbial communities and their relevance for ecosystem models: Decomposition as a case study. *Soil Biology and Biochemistry*, *42*(4), 529–535. https://doi.org/10.1016/j.soilbio.2009.11.016

McPherson, A., & Gavira, J. A. (2013). Introduction to protein crystallization. *Acta Crystallographica. Section F, Structural Biology Communications*, *70*(Pt 1), 2–20. https://doi.org/10.1107/S2053230X13033141

Mei, H.-Z., Xia, D.-G., Zhao, Q.-L., Zhang, G.-Z., Qiu, Z.-Y., Qian, P., & Lu, C. (2016). Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGaseI) in the beetle Batocera horsfieldi. *Gene*, *576*(1, Part 1), 45–51. https://doi.org/10.1016/j.gene.2015.09.057

Mellitzer, A., Glieder, A., Weis, R., Reisinger, C., & Flicker, K. (2012). Sensitive high-throughput screening for the detection of reducing sugars. *Biotechnology Journal*, *7*(1), 155–162. https://doi.org/10.1002/biot.201100001

Michael, M. M. (1991). *The Evolution of Cellulose Digestion in Insects*. *333*(1267), 218–288. https://doi.org/10.1098/rstb.1991.0078

Mnich, E., Bjarnholt, N., Eudes, A., Harholt, J., Holland, C., Jørgensen, B., Larsen, F. H., Liu, M., Manat, R., Meyer, A. S., Mikkelsen, J. D., Motawia, M. S., Muschiol, J., Møller, B. L., Møller, S. R., Perzon, A., Petersen, B. L., Ravn, J. L., & Ulvskov, P. (2020). Phenolic cross-links: Building and de-constructing the plant cell wall. *Natural Product Reports*, *37*(7), 919–961. https://doi.org/10.1039/C9NP00028C

Mondal, S., Shet, D., Prasanna, C., & Atreya, H. S. (2013). High yield expression of proteins in E. coli for NMR studies. *Advances in Bioscience and Biotechnology*, *4*(6), Article 6. https://doi.org/10.4236/abb.2013.46099

Monschein, M., Ioannou, E., Koitto, T., Al Amin, L. A. K. M., Varis, J. J., Wagner, E. R., Mikkonen, K. S., Cosgrove, D. J., & Master, E. R. (2023). Loosenin-Like Proteins from Phanerochaete carnosa Impact Both Cellulose and Chitin Fiber Networks. *Applied and Environmental Microbiology*, *89*(1), e01863-22. https://doi.org/10.1128/aem.01863-22

Murneek, A. E. (1929). Hemicellulose as a Storage Carbohydrate in Woody Plants, with Special Reference to the Apple. *Plant Physiology*, *4*(2), 251–264. https://doi.org/10.1104/pp.4.2.251

Murphy, E., Smith, S., & De Smet, I. (2012). Small Signaling Peptides in Arabidopsis Development: How Cells Communicate Over a Short Distance. *The Plant Cell*, *24*(8), 3198–3217. https://doi.org/10.1105/tpc.112.099010

Nakamura, A., Ishida, T., Kusaka, K., Yamada, T., Fushinobu, S., Tanaka, I., Kaneko, S., Ohta, K., Tanaka, H., Inaka, K., Higuchi, Y., Niimura, N., Samejima, M., & Igarashi, K. (2015). "Newton's cradle" proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Science Advances*, *1*(7), e1500263. https://doi.org/10.1126/sciadv.1500263

Nishiyama, Y., Langan, P., & Chanzy, H. (2002). Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society*, *124*(31), 9074–9082. https://doi.org/10.1021/ja0257319

Nishiyama, Y., Sugiyama, J., Chanzy, H., & Langan, P. (2003). Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society*, *125*(47), 14300–14306. https://doi.org/10.1021/ja037055w

Nomura, T., Iwase, H., Saka, N., Takahashi, N., Mikami, B., & Mizutani, K. (2019). High-resolution crystal structures of the glycoside hydrolase family 45 endoglucanase EG27II from the snail Ampullaria crossean. *Acta Crystallographica Section D: Structural Biology*, *75*(4), 426–436. https://doi.org/10.1107/S2059798319003000

Okmane, L., Nestor, G., Jakobsson, E., Xu, B., Igarashi, K., Sandgren, M., Kleywegt, G. J., & Ståhlberg, J. (2022). Glucomannan and beta-glucan degradation by Mytilus edulis Cel45A: Crystal structure and activity comparison with GH45 subfamily A, B and C. *Carbohydrate Polymers*, 118771. https://doi.org/10.1016/j.carbpol.2021.118771

Oliva, J., Bendz-Hellgren, M., & Stenlid, J. (2011). Spread of Heterobasidion annosum s.s. And Heterobasidion parviporum in Picea abies 15 years after stump inoculation. *FEMS Microbiology Ecology*, *75*(3), 414–429. https://doi.org/10.1111/j.1574-6941.2010.01020.x

Ouellette, R. J., & Rawn, J. D. (2015). 13—Carbohydrates. In R. J. Ouellette & J. D. Rawn (Eds.), *Principles of Organic Chemistry* (pp. 343–370). Elsevier. https://doi.org/10.1016/B978-0-12-802444-7.00013-6

Palermo, M., Gökmen, V., Meulenaer, B. D., Ciesarová, Z., Zhang, Y., Pedreschi, F., & Fogliano, V. (2016). Acrylamide mitigation strategies: Critical appraisal of the FoodDrinkEurope toolbox. *Food & Function*, *7*(6), 2516–2525. https://doi.org/10.1039/C5FO00655D

Pauchet, Y., Kirsch, R., Giraud, S., Vogel, H., & Heckel, D. G. (2014). Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica. *Insect Biochemistry and Molecular Biology*, *49*, 1–13. https://doi.org/10.1016/j.ibmb.2014.03.004

Pauchet, Y., Wilkinson, P., Chauhan, R., & Ffrench-Constant, R. H. (2010). Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PloS One*, *5*(12), e15635. https://doi.org/10.1371/journal.pone.0015635

Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., & Beckham, G. T. (2015). Fungal Cellulases. *Chemical Reviews*, *115*(3), 1308–1448. https://doi.org/10.1021/cr500351c

Pazzagli, L., Cappugi, G., Manao, G., Camici, G., Santini, A., & Scala, A. (1999). Purification, characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from Ceratocystis fimbriata f. Sp. Platani. *The Journal of Biological Chemistry*, *274*(35), 24959–24964. https://doi.org/10.1074/jbc.274.35.24959

Pérez García, M., Zhang, Y., Hayes, J., Salazar, A., Zabotina, O. A., & Hong, M. (2011). Structure and Interactions of Plant Cell-Wall Polysaccharides by Two- and Three-Dimensional Magic-Angle-Spinning Solid-State NMR. *Biochemistry*, *50*(6), 989–1000. https://doi.org/10.1021/bi101795q

Quiroz-Castañeda, R. E., Martínez-Anaya, C., Cuervo-Soto, L. I., Segovia, L., & Folch-Mallol, J. L. (2011). Loosenin, a novel protein with cellulose-disrupting activity from Bjerkandera adusta. *Microbial Cell Factories*, *10*(1), 8. https://doi.org/10.1186/1475-2859-10-8

Rennie, E. A., & Scheller, H. V. (2014). Xylan biosynthesis. *Current Opinion in Biotechnology*, *26*, 100–107. https://doi.org/10.1016/j.copbio.2013.11.013

Reynolds, H. T., Vijayakumar, V., Gluck-Thaler, E., Korotkin, H. B., Matheny, P. B., & Slot, J. C. (2018). Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. *Evolution Letters*, *2*(2), 88–101. https://doi.org/10.1002/evl3.42

Rodriguez, E., & Krishna, N. R. (2001). An Economical Method for 15N/13C Isotopic Labeling of Proteins Expressed in Pichia pastoris1. *The Journal of Biochemistry*, *130*(1), 19–22. https://doi.org/10.1093/oxfordjournals.jbchem.a002957

Sabbadin, F., Pesante, G., Elias, L., Besser, K., Li, Y., Steele-King, C., Stark, M., Rathbone, D. A., Dowle, A. A., Bates, R., Shipway, J. R., Cragg, S. M., Bruce, N. C., & McQueen-Mason, S. J. (2018). Uncovering the molecular mechanisms of lignocellulose digestion in shipworms. *Biotechnology for Biofuels*, *11*(1), 59. https://doi.org/10.1186/s13068-018-1058-3

Sakamoto, K., & Toyohara, H. (2009). Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam Corbicula japonica. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, *152*(4), 390–396. https://doi.org/10.1016/j.cbpb.2009.01.010

Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssönen, E., Bhatia, A., Ward, M., & Penttilä, M. (2002). Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits

disruption activity on cellulosic materials. *European Journal of Biochemistry*, *269*(17), 4202–4211. https://doi.org/10.1046/j.1432-1033.2002.03095.x

Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B., & Schülein, M. (1993). Stereochemistry, specificity and kinetics of the hydrolysis of reduced cellodextrins by nine cellulases. *European Journal of Biochemistry*, *217*(3), 947–953. https://doi.org/10.1111/j.1432-1033.1993.tb18325.x

Seki, Y., Kikuchi, Y., Yoshimoto, R., Aburai, K., Kanai, Y., Ruike, T., Iwabata, K., Goitsuka, R., Sugawara, F., Abe, M., & Sakaguchi, K. (2015). Promotion of crystalline cellulose degradation by expansins from Oryza sativa. *Planta*, *241*(1), 83–93.

Shimonaka, A., Murashima, K., Koga, J., Baba, Y., Nishimura, T., Kubota, H., & Kono, T. (2006). Amino Acid Regions of Family 45 Endoglucanases Involved in Cotton Defibrillation and in Resistance to Anionic Surfactants and Oxidizing Agents. *Bioscience, Biotechnology, and Biochemistry*, *70*(10), 2460–2466. https://doi.org/10.1271/bbb.60200

Šnajdr, J., Cajthaml, T., Valášková, V., Merhautová, V., Petránková, M., Spetz, P., Leppänen, K., & Baldrian, P. (2011). Transformation of Quercus petraea litter: Successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. *FEMS Microbiology Ecology*, *75*(2), 291–303. https://doi.org/10.1111/j.1574-6941.2010.00999.x

Song, B., Zhao, S., Shen, W., Collings, C., & Ding, S.-Y. (2020). Direct Measurement of Plant Cellulose Microfibril and Bundles in Native Cell Walls. *Frontiers in Plant Science*, *11*. https://www.frontiersin.org/articles/10.3389/fpls.2020.00479

Song, J. M., Hong, S. K., An, Y. J., Kang, M. H., Hong, K. H., Lee, Y.-H., & Cha, S.-S. (2017). Genetic and Structural Characterization of a Thermo-Tolerant, Cold-Active, and Acidic Endo-β-1,4-glucanase from Antarctic Springtail, Cryptopygus antarcticus. *Journal of Agricultural and Food Chemistry*, *65*(8), 1630–1640. https://doi.org/10.1021/acs.jafc.6b05037

Szydlowski, L., Boschetti, C., Crisp, A., Barbosa, E. G. G., & Tunnacliffe, A. (2015). Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer Adineta ricciae. *Gene*, *566*(2), 125–137. https://doi.org/10.1016/j.gene.2015.04.007

Taherzadeh, M. J., & Karimi, K. (2008). Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. *International Journal of Molecular Sciences*, *9*(9), 1621–1651. https://doi.org/10.3390/ijms9091621

Teilum, K., Kunze, M. B. A., Erlendsson, S., & Kragelund, B. B. (2017). (S)Pinning down protein interactions by NMR. *Protein Science : A Publication of the Protein Society*, *26*(3), 436–451. https://doi.org/10.1002/pro.3105

Tian Zhang, Yunzhen Zheng, & Daniel J. Cosgrove. (2016). Spatial organization of cellulose microfibrils and matrix polysaccharides in primary plant cell walls as imaged by multichannel atomic force microscopy. *The Plant Journal*, *85*, 179–192. https://doi.org/10.1111/tpj.13102

Timell, T. E. (1967). Recent progress in the chemistry of wood hemicelluloses. *Wood Science and Technology*, *1*(1), 45–70. https://doi.org/10.1007/BF00592255

Trinh, L. B., Phue, J. N., & Shiloach, J. (2003). Effect of methanol feeding strategies on production and yield of recombinant mouse endostatin from Pichia pastoris. *Biotechnology and Bioengineering*, *82*(4), 438–444. https://doi.org/10.1002/bit.10587

Tsuji, A., Tominaga, K., Nishiyama, N., & Yuasa, K. (2013). Comprehensive enzymatic analysis of the cellulolytic system in digestive fluid of the Sea Hare Aplysia kurodai. Efficient glucose release from sea lettuce by synergistic action of 45 kDa endoglucanase and 210 kDa ß-glucosidase. *PloS One*, *8*(6), e65418. https://doi.org/10.1371/journal.pone.0065418

Valencia, A., Alves, A. P., & Siegfried, B. D. (2013). Molecular cloning and functional characterization of an endogenous endoglucanase belonging to GHF45 from the western corn rootworm, Diabrotica virgifera virgifera. *Gene*, *513*(2), 260–267. https://doi.org/10.1016/j.gene.2012.10.046

van Straaten, K. E., Barends, T. R. M., Dijkstra, B. W., & Thunnissen, A.-M. W. H. (2007). Structure of Escherichia coli Lytic transglycosylase MltA with bound chitohexaose: Implications for peptidoglycan binding and cleavage. *The Journal of Biological Chemistry*, *282*(29), 21197–21205. https://doi.org/10.1074/jbc.M701818200

Větrovský, T., Steffen, K. T., & Baldrian, P. (2014). Potential of Cometabolic Transformation of Polysaccharides and Lignin in Lignocellulose by Soil Actinobacteria. *PLOS ONE*, *9*(2), e89108. https://doi.org/10.1371/journal.pone.0089108

Vieira, S., Sikorski, J., Gebala, A., Boeddinghaus, R. S., Marhan, S., Rennert, T., Kandeler, E., & Overmann, J. (2020). Bacterial colonization of minerals in grassland soils is selective and highly dynamic. *Environmental Microbiology*, *22*(3), 917–933. https://doi.org/10.1111/1462-2920.14751

Vlasenko, E., Schülein, M., Cherry, J., & Xu, F. (2010). Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology*, *101*(7), 2405–2411. https://doi.org/10.1016/j.biortech.2009.11.057

Vogel, J. (2008). Unique aspects of the grass cell wall. *Current Opinion in Plant Biology*, *11*(3), 301–307. https://doi.org/10.1016/j.pbi.2008.03.002

Wang, B., & Qiu, Y.-L. (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza*, *16*(5), 299–363. https://doi.org/10.1007/s00572-005-0033-6

Wang, T., Yang, H., Kubicki, J. D., & Hong, M. (2016). Cellulose Structural Polymorphism in Plant Primary Cell Walls Investigated by High-Field 2D Solid-State NMR Spectroscopy and Density Functional Theory Calculations. *Biomacromolecules*, *17*(6), 2210–2222. https://doi.org/10.1021/acs.biomac.6b00441

Wang, T., Zabotina, O., & Hong, M. (2012). Pectin–Cellulose Interactions in the *Arabidopsis* Primary Cell Wall from Two-Dimensional Magic-Angle-Spinning Solid-State Nuclear Magnetic Resonance. *Biochemistry*, *51*(49), 9846–9856. https://doi.org/10.1021/bi3015532

Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. *Nature*, *394*(6691), Article 6691. https://doi.org/10.1038/28527

Watson, J. D., & Crick, F. H. C. (1953). *A structure for deoxyribose nucleic acid. 177*, 737–738.

Weitbrecht, K., Müller, K., & Leubner-Metzger, G. (2011). First off the mark: Early seed germination. *Journal of Experimental Botany*, *62*(10), 3289–3309. https://doi.org/10.1093/jxb/err030

Workneh, F., & van Bruggen, A. H. C. (1994). Microbial density, composition, and diversity in organically and conventionally managed rhizosphere soil in relation to suppression of corky root of tomatoes. *Applied Soil Ecology*, *1*(3), 219–230. https://doi.org/10.1016/0929-1393(94)90013-2

Xiao, C., & Anderson, C. (2013). Roles of pectin in biomass yield and processing for biofuels. *Frontiers in Plant Science*, *4*. https://www.frontiersin.org/articles/10.3389/fpls.2013.00067

Xu, B., Hellman, U., Ersson, B., & Janson, J.-C. (2000). Purification, characterization and amino-acid sequence analysis of a thermostable, low molecular mass endo-β-1,4-glucanase from blue mussel, Mytilus edulis. *European Journal of Biochemistry*, *267*(16), 4970–4977. https://doi.org/10.1046/j.1432-1327.2000.01533.x

Xu, B., Janson, J.-C., & Sellos, D. (2001). Cloning and sequencing of a molluscan endo-β-1,4-glucanase gene from the blue mussel, Mytilus edulis. *European Journal of Biochemistry*, *268*(13), 3718–3727. https://doi.org/10.1046/j.1432-1327.2001.02280.x

Yang, J.-K., Zhang, J.-J., Yu, H.-Y., Cheng, J.-W., & Miao, L.-H. (2014). Community composition and cellulase activity of cellulolytic bacteria from forest soils planted with broad-leaved deciduous and evergreen trees. *Applied Microbiology and Biotechnology*, *98*(3), 1449–1458. https://doi.org/10.1007/s00253-013-5130-4

Zhang, Z., Bian, J., Zhang, Y., Xia, W., Li, S., & Ye, Z. (2022). An Endoglucanase Secreted by Ustilago esculenta Promotes Fungal Proliferation. *Journal of Fungi*, *8*(10), Article 10. https://doi.org/10.3390/jof8101050

ZhiMing, Y., Bo, K., XiaoWei, H., ShaoLei, L., YouHuang, B., WoNa, D., Ming, C., Hyung-Taeg, C., & Ping, W. (2011). Root hair-specific expansins

modulate root hair elongation in rice. *The Plant Journal*, *66*(5), 725–734. https://doi.org/10.1111/j.1365-313X.2011.04533.x

Zhou, Q., Ji, P., Zhang, J., Li, X., & Han, C. (2017). Characterization of a novel thermostable GH45 endoglucanase from Chaetomium thermophilum and its biodegradation of pectin. *Journal of Bioscience and Bioengineering*, *124*(3), 271–276. https://doi.org/10.1016/j.jbiosc.2017.03.017

Popular science summary

Cellulose degradation occupies a significant role in our everyday lives, irrespective to whether we acknowledge that or not. Land plants annually fix about 15% of the carbon dioxide in the atmosphere and produce over 40 billion tons of plant biomass, approximately half of which consists of cellulose. If not for cellulose degrading organisms, none of the woody biomass would become thoroughly degraded, which would require us to manually dispose of it, in order to avoid being surrounded by large deposits of plant matter. Or, perhaps, human life on Earth would not even be possible, as plants are the primary producers of food for animals and fungi and all other non-photosynthesizing organisms.

Decomposition of plant material and recirculation of carbon is mainly done by microorganisms, mainly fungi. Cellulose degrading organisms have evolved to produce a highly potent cellulose degrading machinery – enzymes called *cellulases*. The evidence for the efficacy of *cellulases* is the obvious absence of yesteryear foliage.

Cellulases interact with a number of other enzymes which are crucial to break down other components of the plant material, for example, other polysaccharides called hemicellulose, and a phenolic polymer lignin. Lignin binds cellulose and hemicellulose together, makes wood hard and gives wood its signature light brown colour.

We have learned to utilize these naturally abundant enzymes – from food and material production, to medicines. Production of highly complex compounds is possible due to the development of enzyme production and purification systems. The advances in protein structure and function studies are behind the majority of medicines we rely on throughout our life. We rely also on the direct application of industrially obtained cellulases, for example, the GH45 enzymes

found in detergents such as textile washing powders, keep our clothes visually pleasing.

The vast majority of plant-degrading enzymes are glycoside hydrolases (GH enzymes). GH enzymes cleave polysaccharides by means of hydrolysis, ultimately down to their original building blocks, monosaccharides, or, in other words, simple sugars (e.g. glucose, mannose, xylose). Enzymes that are evolutionarily related and similar in amino acid sequence, structure and properties have been classified into different GH families in a database named CAZy (Carbohydrate Active enZymes). Enzymes in the GH45 family are among the smallest known GH enzymes, which is believed to be an advantage in being able to enter tight pores and spaces in plant material that larger enzymes cannot access. What make these enzymes more interesting is that they are similar and related to expansins, a type of protein that all plants have, and must have in order to grow. Expansins soften the cell wall around plant cells so they can expand in size after cell division, perhaps 10-100 times in volume, but how this happens at the molecular level is still unclear. Related proteins also exist in fungi and are named loosenins (which are even smaller) and swollenins, which also appear to be able to soften plant cell walls, but their role and function are largely unknown so far. GH45 enzymes are mainly found in plant-decomposing fungi (ascomycetes and basidiomycetes), but also in molluscs (mussels, bivalves, snails), insects, crustaceans, and a few bacteria. Closer comparison shows that they fall into three subfamilies, A, B and C.

This thesis focuses on the dissimilarities within the GH45 enzyme family. We have studied the differences across GH45 subfamilies and between members of the same subfamily. We have noted the variation in enzymatic activity, such as different cleavage patterns, activity rates and substrate specificities, and we have investigated the structural differences between these enzymes. We carried out a side-by-side comparison of the enzymatic activities and were the first to affirm that members from all GH45 subfamilies act via the inverting action mechanism. Additionally, we demonstrated beta-1,4-mannanase activity in GH45. We were also the first to crystallize and determine the 3D structure of a GH45 enzyme from a brown-rot fungus, belonging to subfamily C where there was only one known structure previously, from a white-rot fungus. Surprisingly, the enzyme from

the brown rot fungus was four times faster than that from the white rot fungus, when comparing the initial rate of beta-glucan degradation, even though the enzymes are very similar (82% sequence identity).

Finally, an enzyme in GH45 subfamily C has been produced that is isotope-labelled, i.e. which have ^{13}C (carbon-13) and ^{15}N (nitrogen-15) instead of the usual isotopes of carbon and nitrogen, to be used in various analyses with nuclear magnetic resonance spectroscopy (NMR). There we have, among other things, succeeded in assigning NMR signals for the main chain of the protein, which paves the way for structural studies of the enzyme in solution. Experiments with isotopically labelled proteins can reveal how the protein binds to polysaccharides, how it breaks them down, and give additional information on the dynamics of the structure.

Populärvetenskaplig sammanfattning

Cellulosanedbrytning spelar en viktig roll i vårt dagliga liv, oavsett om vi är medvetna om det eller inte, och är en process av gigantiska proportioner. Landlevande växter fixerar årligen ca 15% av den koldioxid som finns i atmosfären och producerar över 40 miljarder ton växtbiomassa, som består till ungefär hälften av cellulosa. Om det inte var för cellulosanedbrytande organismer, så skulle växtmaterialet inte brytas ned ordentligt, koldioxiden skulle ta slut och vi skulle vada i stora avlagringar av växtbiomassa. Eller egentligen funnes inte möjlighet för mänskligt liv på jorden, eftersom växter är primärproducenter av föda för djur och svampar och alla andra icke-fotosyntetiserande organismer.

Nedbrytning av växtmaterial och återcirkulation av kol görs i huvudsak av mikroorganismer, främst svampar. De producerar ett mycket potent cellulosanedbrytande maskineri med enzymer, som aktiva proteiner kallas. De enzymer som görs i störst mängd bryter ned cellulosa och kallas *cellulaser*. De samverkar med en mängd andra enzymer som behövs för att bryta ned andra komponenter i växtmaterialet, framförallt andra polysackarider som kallas hemicellulosa, och lignin som binder ihop cellulosa och hemicellulosa och gör trä hårt (och ger gul-brun färg).

Vi har lärt oss att använda en del sådana naturligt rikligt förekommande enzymer – för alltifrån mat och material till mediciner. Produktion av mycket komplexa föreningar är möjlig tack vare utvecklingen av enzymproduktion och reningssystem. Framstegen i studier av proteinstruktur och funktion ligger bakom de flesta mediciner vi är beroende av under våra liv. Vi förlitar oss också på direkt applicering av industriellt erhållna cellulaser, till exempel

GH45-enzymer som används i s.k. ”colour”-tvättmedel för att färgade textilier ska behålla färgen och inte bli urblekta och tråkiga.

De allra flesta växtnedbrytande enzymerna är glykosidhydrolaser (GH-enzymer), vilket betyder att de klyver polysackarider med hjälp av hydrolys, slutligen ner till sina ursprungliga byggstenar, monosackarider, eller enkla sockerarter som de också kallas (t.ex. glukos, mannnos, xylos). Enzymer som är evolutionärt besläktade och liknar varandra i aminosyrasekvens, struktur och egenskaper har klassificerats i olika GH-familjer i en databas vid namn CAZy (Carbohydrate Active enZymes).

Enzymer i familj GH45 är bland de minsta GH-enzymer som finns, vilket tros var en fördel för att kunna komma in i trånga porer och utrymmen i växtmaterial som större enzymer inte kommer åt. De är också intressanta därför att de är besläktade med och liknar expansiner, en typ av proteiner som alla växter har, och måste ha för att kunna växa. Expansiner mjukar upp cellväggen runt växtceller så att de kan expandera i storlek efter celldelning, kanske 10-100 gånger i volym, men hur det går till på molekylär nivå är fortfarande oklart. Det finns också besläktade proteiner hos svampar, looseniner, som är ännu mindre, och swolleniner som har flera funktionella delar, vilka också verkar kunna mjuka upp växtcellväggar, men deras roll och funktion är i stort sett okänd än så länge. GH45-enzymer finns framförallt hos växtnedbrytande svampar (ascomyceter och basidiomyceter), men även hos mollusker (musslor, snäckor, sniglar), insekter, kräftdjur, och ett fåtal bakterier. Närmare jämförelse visar att de fördelar sig i tre underfamiljer, A, B och C.

Denna avhandling fokuserar på olikheterna inom GH45-enzymfamiljen. Vi har studerat skillnaderna mellan GH45-underfamiljer och mellan medlemmar av samma underfamilj. Vi har noterat variationen i enzymatisk aktivitet på cellulosa och hemicellulosa (beta-glukan, glukomannan), såsom olika klyvningsmönster, aktivitetshastigheter och substratspecificiteter, och vi har undersökt de strukturella skillnaderna mellan dessa enzymer. Vi utförde jämförelse sida-vid-sida av de enzymatiska aktiviteterna och var de första att bekräfta att medlemmar från alla GH45-underfamiljer verkar via den inverterande mekanismen, och påvisade även beta-1,4-mannanas-aktivitet hos GH45.

Vi var också först med att kristallisera och bestämma 3D-strukturen av ett GH45-enzym från en brunrötesvamp, som tillhör underfamilj C där det bara fanns en känd struktur tidigare, från en vitrötesvamp. Överraskande nog så var enzymet från brunrötesvampen fyra gånger snabbare än det från vitrötesvampen, vid jämförelse av initial hastighet vid nedbrytning av beta-glukan, trots att enzymerna är väldigt lika (82% sekvensidentitet).

Slutligen så har ett enzym i GH45 subfamilj C producerats som är isotop-märkt, dvs. som har ^{13}C (kol-13) och ^{15}N (kväve-15) istället för de vanliga isotoperna av kol och kväve, för att användas i olika analyser med NMR-spektroskopi (nuclear magnetic resonance spectrosopy). Där har vi bl.a. lyckats assignera NMR-signaler för huvudkedjan i proteinet, vilket banar väg för strukturstudier av enzymet i lösning, hur det binder till och bryter ned olika polysackarider, och eventuell dynamik i strukturen när enzymet arbetar.

Acknowledgements

Hi there! Congratulations *to me* that you made it this far! Not that I believe you read the whole thesis, but it simply means I play important enough role in your life for you to remember my name and check whether I remember you or not. Could be an indication that you care. Possibly.

Fortunately, I don't intend to make this section overly long or emotional, let's see how that goes. (spoiler alert: I failed.)

I would like to acknowledge the people who have aided in the creation of this thesis, whether it was by giving direct advice or through emotional support during this journey.

First and foremost, I would like to thank my main supervisor *Jerry Ståhlberg* who gave me the opportunity to enter the GH45 project. Jerry, you have my utmost gratitude for giving me this chance. Thank you for your guidance, encouragement and knowledge. Thank you for trusting me with this project and for your patience. I admire your wisdom and enthusiasm about science. You've been an inspiration.

Equally as responsible for my enrollment in this project is *Mats Sandgren*, my co-supervisor. *Mats*, I would like to thank you for all of the support and advice I have received from you over the years, both science- and life- related.

Gustav Nestor, you joined my supervisor team at the other half of my PhD project, but this last year was so NMR-heavy that you felt like the main supervisor at times. Thank you for your time and support, your knowledge and essentially sparking my interest in protein NMR.

I would like to thank *Saeid Karkehabadi, Henrik Hansson,* and *Miao Wu* who at the beginning of my PhD taught me a lot about protein purification and protein crystallography.

Some labs are fortunate to have people who never refuse help or an advice. Thank you, *Igor, Topi, Mikolaj* and *Johanna* for sharing your knowledge and giving a helping hand when needed.

Thank you, *Nils*, for all the work you do to manage our corridor, and also for the interesting conversations during morning fika. Often it is much more fun to start a day with a debate about politics.

Giselle, thank you for being *you*. I appreciate our conversations, especially when we disagree on something. Thank you for lending a friendly ear and your honest opinion every time I was going through some emotional ~~er~~ irregularities.

Everyday life at SLU would be boring without *Yasha*. Thank you for making days more cheerful and fun!

Sanjana, Florentine, Pernilla, and *Ani* thank you for all the after-works, fun non-scientific activities and also deeper conversations.

Additional thanks to *Alyona* for infecting others with your relentless enthusiasm about science and to *Jonas* for practical advice regarding lab work and paper writing.

Once in a while I felt like complaining and venting about life. Thank you, *Mathilde*, for being there equipped with a ton of patience and being generous with support.

I was very fortunate to greet the latest two PhD students in our group, *Naike* and *Piera*. *Naike*, your relentlessness is inspiring, don't ever lose that! *Piera*, your tendency to *not avoid* social responsibilities is remarkable and refreshing. Our group needs people like the both of you!

I would like to thank the people from C3 corridor, *Volkmar, Simon, Leslie, Jule, Thomas, Andrea, Adrian, Sumitha, Sabine, Bettina, Kerstin, Micke, Hasi*, who made the days more enjoyable by contributing to the discussions during *fikas*, lunch or after-works.

Also, huge thanks to *Monika* for organizing PhD student activities and keeping everyone up-to-date with the responsibilities and happenings of PhD student life.

Thank you to *Robert* and *Magnus* for keeping the building standing!

There are quite a few people outside of SLU who have significantly contributed to this thesis by being part of my emotional support system.

I would like to thank *my family* for providing me with the access to education and encouraging me to choose my own paths in life. Paldies par visu!

I would like to thank my dear friend *Irina*, whom I first met at SLU, but who has been a significant part of my life ever since. Jag är tacksam över att ha fått lära känna dig och att du fortsätter att vara en del av mitt liv!

Then there are the Latvian bunch – *Elīna A., Toms un Karīna,* who, with their encouraging words, have played a much bigger role than I suspect they realize. Paldies, draugi, par to "*Ai, bet Tu tak to vari!*" attieksmi!

Last, but definitely not least, I would like to thank the love of my life, *Tom*, without whom I would have given up a very long time ago. Paldies Tev, *mīļum*, par itin visu – Tavu pacietību, mīlestību, rūpēm, uzmundrinājumu un neatlaidīgo ticību man! Vārdi nespēj noraksturot to, ko Tu man nozīmē, bet es ceru, ka Tu apzinies manas mīlestības apmērus. *Negribējās jau šo pataisīt par drāmas teātri, bet kaut kā tā sanāca... Nu, ja jau, tad jau… laikam jānobeidz ar citātu:*

"Šausmīgi daudz taisnību. Nojukt var, ja nav savējās."
— Imants Ziedonis, Epifānijas

I

Carbohydrate Polymers 277 (2022) 118771

Contents lists available at ScienceDirect

Carbohydrate Polymers

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

Research Paper

Glucomannan and beta-glucan degradation by *Mytilus edulis* Cel45A: Crystal structure and activity comparison with GH45 subfamily A, B and C☆

Laura Okmane [a], Gustav Nestor [a], Emma Jakobsson [b,1], Bingze Xu [c,2], Kiyohiko Igarashi [d], Mats Sandgren [a], Gerard J. Kleywegt [b,3], Jerry Ståhlberg [a,*]

[a] Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden
[b] Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden
[c] Center for Surface Biotechnology, Uppsala University, Uppsala, Sweden
[d] Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, University of Tokyo, 1–1–1 Yayoi, Bunkyo-ku, Tokyo 113–8657, Japan

ARTICLE INFO

Keywords:
Endoglucanase
Blue mussel
Cel45A
GH45
Beta-glucan
Glucomannan

ABSTRACT

The enzymatic hydrolysis of barley beta-glucan, konjac glucomannan and carboxymethyl cellulose by a β-1,4-D-endoglucanase MeCel45A from blue mussel, *Mytilus edulis*, which belongs to subfamily B of glycoside hydrolase family 45 (GH45), was compared with GH45 members of subfamilies A (*Humicola insolens* HiCel45A), B (*Trichoderma reesei* TrCel45A) and C (*Phanerochaete chrysosporium* PcCel45A). Furthermore, the crystal structure of MeCel45A is reported.

Initial rates and hydrolysis yields were determined by reducing sugar assays and product formation was characterized using NMR spectroscopy. The subfamily B and C enzymes exhibited mannanase activity, whereas the subfamily A member was uniquely able to produce monomeric glucose. All enzymes were confirmed to be inverting glycoside hydrolases. MeCel45A appears to be cold adapted by evolution, as it maintained 70% activity on cellohexaose at 4 °C relative to 30 °C, compared to 35% for TrCel45A. Both enzymes produced cellobiose and cellotetraose from cellohexaose, but TrCel45A additionally produced cellotriose.

1. Introduction

In aquatic ecosystems, cellulose is produced in large quantities by algal plankton, which is an important energy source for filter feeding organisms such as mussels (Newell et al., 1989). Not surprisingly, cellulase activity has been demonstrated in the digestive tract of several bivalves (Kaur, 1997; Onishi et al., 1985; Purchon, 1977). The highest activity was found in so called crystalline styles, which most bivalves and some gastropods (snails and slugs) use for digestion. The crystalline style is a jelly-like translucent rod protruding into the stomach of the mussel. It aids in digestion by being rotated and pushed against the gastric shield, thus dragging the food from the gills into the stomach and grinding the food like a pestle and mortar. The style dissolves gradually and releases various enzymes that initiate extracellular digestion in the stomach (Purchon, 1977).

In the blue mussel, *Mytilus edulis*, three enzymes with activity against carboxymethyl cellulose (CMC) have been detected, the smallest of which (around 20 kDa) has been purified from blue mussel collected off the Swedish west coast. It has been characterized and the protein and gene sequences have been determined (Xu et al., 2000). The mature

Abbreviations: AcCel45A, *Ampullaria crossean* Cel45A; BG, betaglucan; CBM, carbohydrate binding module; CMC, carboxymethyl cellulose; DPBB, double psi beta barrel; GH, glycoside hydrolase; GH45, glycoside hydrolase family 45; GM, glucomannan; HiCel45A, *Humicola insolens* Cel45A; MeCel45A, *Mytilus edulis* Cel45A; PASC, phosphoric acid swollen cellulose; PcCel45A, *Phanerochaete chrysosporium* Cel45A; PHBAH, p-Hydroxybenzoic acid hydrazide; RMSD, root mean square deviation; TrCel45A, *Trichoderma reesei* Cel45A.

☆ Enzymes: EC3.2.1.4.

* Corresponding author at: Department of Molecular Sciences, Swedish University of Agricultural Sciences, POB 7015, SE-750 07 Uppsala, Sweden.
E-mail address: Jerry.Stahlberg@slu.se (J. Ståhlberg).
[1] Present addresses: Emma Jakobsson, CIC biomaGUNE, Basque Research and Technology Alliance (BRTA), San Sebastián, 20014 Guipúzcoa, Spain.
[2] Present addresses: Bingze Xu, Medical Inflammation Research, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Solna, Sweden.
[3] Present addresses: Gerard J. Kleywegt, EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

https://doi.org/10.1016/j.carbpol.2021.118771
Received 10 August 2021; Received in revised form 24 September 2021; Accepted 11 October 2021
Available online 21 October 2021

protein has 181 amino acid residues and consists of a single glycoside hydrolase family 45 (GH45) catalytic domain without any carbohydrate binding module or other accessory modules and was designated MeCel45A. It has a broad temperature optimum between 30 and 50 °C. Interestingly, it retains over 50% of the maximum activity at 0 °C. This may be related to observations that in Sweden, mussels actively ingest seston (suspended particles) at temperatures below 0 °C, suggesting that they can utilize spring phytoplankton blooms in boreal waters even at low temperatures (Loo, 1992). Although the enzyme did not show any activity at very high temperatures, it could withstand shorter periods (10 min) at +100 °C without irreversible loss of enzymatic activity (Xu et al., 2000).

Today there are over 460 GH45 entries in the CAZy database (cazy.org), which have been divided into three subfamilies, A, B and C (Couturier et al., 2011; Igarashi et al., 2008; Nomura et al., 2019). All three have been found in fungi. GH45 members of nematodes, insects, springtail and bacteria belong to subfamily A (Kikuchi et al., 2004; Lee et al., 2004; Mei et al., 2016; Pauchet et al., 2010, 2014; Song et al., 2017; Valencia et al., 2013), while molluscs only have subfamily B endoglucanases (Guo et al., 2008; Sakamoto & Toyohara, 2009; Tsuji et al., 2013; Xu et al., 2000). Subfamily C is only found in basidiomycete fungi, so far.

Three-dimensional (3D) structures of nine GH45 enzymes are available in the PDB (rcsb.org), one of which is the structure of MeCel45A described in this paper. They all share a conserved six-stranded double-ψ β-barrel (DPBB) also known as GH45-like domain in their structure. The DPBB domain is evolutionarily and structurally related to that of expansins and their homologues (loosenins, swollenins), where the same catalytic acid (aspartic acid) has been conserved in the catalytic center motif (Payne et al., 2015). In GH45 subfamily A and B, an additional aspartic acid at the catalytic center is proposed to act as catalytic base in the inverting hydrolytic mechanism, but the corresponding residue is not conserved in subfamily C. In this regard, subfamily C appears to be more similar to expansin-like one-domain proteins named loosenins, which also lack the putative catalytic base residue. However, no hydrolytic activity of loosenins has been documented yet, as opposed to subfamily C members which have shown β(1 → 4)-endoglucanase activity (Igarashi et al., 2008).

Thus far, the most studied GH45 enzyme has been HiCel45A from the ascomycete fungus *Humicola insolens*. As such, it often serves as a GH45 reference in structure and activity comparisons. HiCel45A is a subfamily A member that is widely used in treating textiles, for example as part of washing powders in the form of the product Carezyme from Novozymes. In subfamily B, the first published structure was that of AcCel45A from the snail *Ampullaria crossean* (Nomura et al., 2019), also known as EG27II, which appears to be the most similar structure to MeCel45A. In subfamily C there is only one enzyme with structures available, namely PcCel45A from the white-rot basidiomycete fungus *Phanerochaete chrysosporium*. Due to the exceptional nature of the catalytic center among subfamily C members, a distinctive catalytic "Proton-relay" mechanism has been proposed for PcCel45A (Nakamura et al., 2015).

GH45 enzymes hydrolyze β(1 → 4) linkages in soluble beta-glucans via an inverting action mechanism, where one amino acid residue acts as a general acid that protonates the glycosidic oxygen and another residue acts as a general base that activates a water molecule to hydrolyze the glycosidic bond. Such a catalytic mechanism leads to inversion of position at the anomeric carbon, thus producing alpha-anomers from β(1 → 4) linked glucans. The highest catalytic activities of GH45 enzymes have been demonstrated on barley beta-glucan and lichenan, lower activities on CMC, phosphoric-acid-swollen cellulose (PASC) and hydroxyethyl cellulose (HEC), and minute activities on crystalline cellulose substrates such as Avicel (Gilbert et al., 1990; Saloheimo et al., 1994; Schou et al., 1993). HiCel45A and other subfamily A members are able to hydrolyze various cellulosic substrates (PASC, CMC, Avicel and bacterial cellulose), as well as xylan and xyloglucan among others (Vlasenko et al., 2010). Subfamily B members have been shown to hydrolyze CMC (Karlsson et al., 2002; Liu et al., 2010; Nomura et al., 2019), PASC, Avicel, glucomannan (Karlsson et al., 2002; Liu et al., 2010), and xylan (Liu et al., 2010), whereas subfamily C member PcCel45A is unable to hydrolyze xyloglucan and Avicel (Godoy et al., 2018).

The roles of three residues proposed to be involved in the catalytic mechanism of GH45 enzymes have been investigated experimentally by mutations at those sites. Mutation of the catalytic acid leads to complete inactivation of the enzyme in all GH45 subfamilies (D121N in HiCel45A, D137A in AcCel45A, D117N in *Fomitopsis palustris* FpCel45, D114A and D114N in PcCel45A) (Cha et al., 2018; Davies et al., 1995; Godoy et al., 2018; Nakamura et al., 2015; Nomura et al., 2019). Mutation of the catalytic base proposed for subfamily A and B, leads to inactivation in subfamily A (HiCel45A D10N), but the activity is not completely lost in subfamily B (AcCel45A D27A) (Davies et al., 1995; Nomura et al., 2019). Subfamily C members lack an acidic residue at the corresponding position. Instead, an asparagine residue at another position, Asn92 in PcCel45A, has been proposed to act as catalytic base. Mutation of this residue, or the corresponding residue in other GH45 enzymes, drastically reduced the enzymatic activity in members from all GH45 subfamilies (D114N in HiCel45A, N112A in AcCel45A, N95D in FpCel45, N92D in PcCel45A) (Cha et al., 2018; Davies et al., 1995; Nakamura et al., 2015; Nomura et al., 2019).

There are very few studies where enzymes from different GH45 subfamilies have been compared side-by-side (Berto et al., 2019; Vlasenko et al., 2010). Here we describe the crystal structure of MeCel45A and compare its enzymatic activity with representatives of GH45 subfamilies A, B, and C. We hypothesise that i) the reaction mechanism is inverting also in subfamily B and C, previously proven only for a subfamily A enzyme; ii) GH45 enzymes of subfamilies A, B and C show differences in substrate specificity and bond cleavage preference on beta-glucan and glucomannan; iii) MeCel45A is more similar in structure and activity properties to the other GH45 subfamily B enzyme used in this study, TrCel45A, than to the members of subfamily A and C; iv) MeCel45A from blue mussel that lives in cold waters is more cold-adapted and retains higher activity at low temperatures than the homologous enzyme from a tropical fungus, TrCel45A.

2. Results

2.1. Isolation, purification and structure determination of MeCel45A

The MeCel45A enzyme was isolated from the digestive gland of the common blue mussel, *M. edulis*, from waters off the Swedish west coast. From 29 kg of blue mussel, 8 mg of pure enzyme was obtained, by using a three-step purification procedure with immobilized metal affinity (IMAC), size exclusion, and cation-exchange chromatography.

During the screening for crystallisation conditions it turned out that microcrystals appeared already in the concentrated protein solution. Crystals for X-ray analysis could be grown by equilibration against 0.6 M sodium acetate, pH 5.5, and 0.1–0.5 M NaCl without addition of other precipitants. The crystals were orthorhombic and the space group was $P2_12_12_1$ with one protein molecule per asymmetric unit. The structure was solved by SIRAS (single isomorphous replacement with anomalous scattering) using a heavy-atom derivative with Baker's dimercurial ($C_{10}H_{16}Hg_2O_6$; 1,4-diacetoxymercuri-2,3-dimethoxybutane) for phasing, and was refined against a high-resolution dataset at 1.2 Å.

The refined structure model contains the complete polypeptide chain (residues 1 to 181), 305 water molecules, one acetate molecule and two polyethylene glycol (PEG) molecules. Proline residues 8 and 108 are involved in cis-peptide bonds and all the 12 cysteine residues are involved in disulfide bridges with the pairings 4/16, 30/69, 32/176, 65/178, 72/157, and 103/113. Electron density maps indicated distinct alternate conformations for the side chains of Thr20, Met55, Gln67, Lys74, Gln85, Ser94, Asn105, His122, His130 and Asp132, which were included in the refinement. Statistics relating to the quality of the X-ray

diffraction data and the refined protein model are summarized in Tables S1 and S2.

2.2. Structure of MeCel45A

M. edulis Cel45A has a compact and globular structure with approximate dimensions of 30 × 40 × 50 Å (Fig. 1A) built around a six-stranded β-barrel that has the characteristic DPBB fold (Fig. 1B). Loops that connect the beta-strands combine to extend one of the faces of the barrel into a shallow substrate-binding cleft (upwards in Fig. 1A). The protein contains a few secondary structure elements in addition to the canonical DPBB fold. At the N-terminus the first 14 residues form a very short strand-turn-strand anti-parallel β-sheet directly before strand β1. A short α-helix is present in the long loop between strands β1 and β2. Finally, two α-helices are present in the stretch of 26 residues after strand β6 at the C-terminus. The latter two helices form a protrusion from the β-barrel on the side opposite to the substrate-binding surface (downwards in Fig. 1A). On this protrusion three surface histidines are located.

Among the GH45s with known structure, MeCel45A is most similar to the other subfamily B enzyme from a mollusc, AcCel45A from *Ampullaria crossean*. As expected from the high sequence similarity (48% identity) the structures are very similar with a low root-mean-square deviation (RMSD) of 1.7 Å over 168 aligned Cα atoms. In the overall fold, MeCel45A has a three-residue deletion at the tip of a loop near the N-terminus, and an eight-residue insertion near the C-terminus that forms an extra alpha helix and extends the size of the protrusion from the β-barrel. However, both these regions are distant from the active site and not likely to influence the catalytic properties. The active site of MeCel45A is nearly identical to that of AcCel45A, including the positions of Asp24, Asn109 and Asp132 that correspond to the proposed catalytic residues of AcCel45A (Asp27, Asn112, Asp137; Fig. 2), suggesting that these residues have the same function in MeCel45A. Asp132 at the bottom of the cleft is the catalytic acid that protonates the glycosidic oxygen, Asp24 corresponds to the proposed catalytic base in subfamily A enzymes (Asp10 in HiCel45A) and Asn109 is in the same position as the proposed catalytic base of the subfamily C enzyme PcCel45A (Asn92). Minor differences relative to AcCel45A include the lengthening of the loop where Asn109 is located by one residue (Tyr107), as well as the substitution of two residues on either side of subsite +2, where Arg24 and Lys89 in AcCel45A are replaced by Asn21 and Gln85, respectively, in MeCel45A.

2.3. Structure comparison with other GH45 enzymes

The structure of MeCel45A was further compared with the other GH45 enzymes used in the activity measurements, HiCel45A (subfamily A), TrCel45A (subfamily B) and PcCel45A (subfamily C). For TrCel45A no experimentally determined structure is yet available. Therefore, a structure model of the catalytic domain of TrCel45A was built by homology modelling using SWISS-MODEL and the structure of MeCel45A as template. Percentage sequence identities and structural deviations (RMSD) relative to MeCel45A are listed in Table 1, and a multiple sequence alignment is shown in Fig. 3. The MeCel45A and PcCel45A enzymes consist of a single catalytic domain alone, whereas HiCel45A and TrCel45A are bimodular with a carbohydrate binding module (CBM1) attached by a Ser/Thr-rich linker peptide to the catalytic domain at the C-terminus.

The β-strands of the DPBB core superpose closely, but the surface structures differ because of variations in the lengths of loop regions flanking the β-barrel (Fig. 4). The subfamily B and C enzymes (MeCel45A, TrCel45A, PcCel45A) are more similar to each other than to HiCel45A of subfamily A. HiCel45A has longer loops surrounding the catalytic center, forming a closed structure that resembles a tunnel, while the others have an open cleft. In PcCel45A, loops extend the cleft at both ends making the cellulose binding surface >5 Å longer (Fig. 5B). The central part of the cleft is noticeably narrower in MeCel45A than in TrCel45A and PcCel45A.

The location of catalytic residues is well conserved, except for the catalytic base corresponding to Asp24 in MeCel45A, which is not present in PcCel45A (Fig. 5). In PcCel45A a glycine residue occupies this position instead. Furthermore, HiCel45A has an aspartic acid (Asp114) instead of asparagine at the location of the alternate base (Asn109 in MeCel45A). Apart from the catalytic residues, a few additional amino acids are conserved near the catalytic acid. These are Thr20, Tyr22 and His130 in MeCel45A. His130 is on the same beta-strand as the catalytic

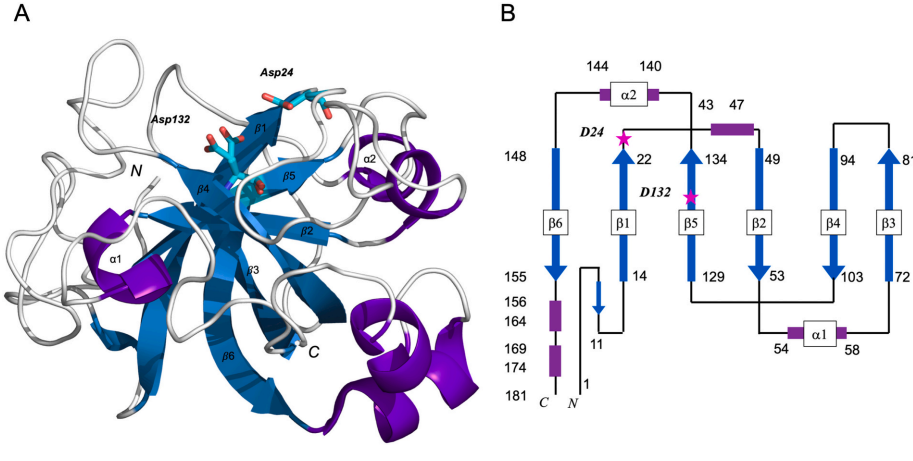


Fig. 1. Overall structure of *M. edulis* Cel45A. (A) Ribbon drawing showing the location of a shallow cleft on one face of the central six-stranded β-barrel with the putative catalytic aspartate residues 24 and 132 sitting on either side of the cleft. On the other side, two α-helices at the C-terminus protrude from the β-barrel. (B) Folding topology diagram with β-strands and α-helices numbered according to the generalized double-psi fold (Castillo et al., 1999). Cel45A contains an extra α-helix at β1/β2, one short β-strand at the *N*-terminus and two *C*-terminal α-helices. The residue numbers of Cel45A at each end of the secondary structure elements are indicated.

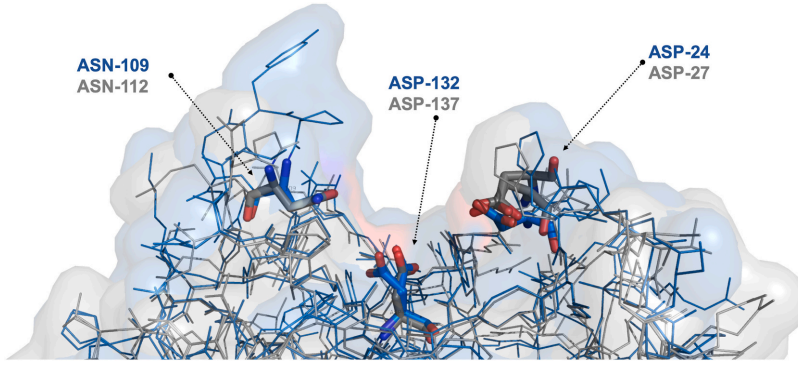


Fig. 2. *M. edulis* Cel45A (blue) structure superimposed on *A. crossean* Cel45A (gray). Assisting residues (Asn109 in MeCel45A; Asn112 in AcCel45A) and catalytic center residues (Asp132, Asp24 in MeCel45A; Asp137, Asp27 in AcCel45A) are represented as sticks (from left to right respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
GH45 endoglucanase structure and sequence similarities with MeCel45A.

Name	Organism	GenBank *accession ID*	PDB ID	*UniProt accession ID*	Percent sequence identity[a]	Structure similarity	
						RMSD (Å)	Cα
HiCel45A	*H. insolens*	CAB42307.1	2eng	P43316	24	4.1	80
TrCel45A	*T. reesei*	CAA83846.1	N/A	P43317	44	N/A	N/A
AcCel45A	*A. crossean*	ABR92638.1	5xbu	A7KMF0	48	1.7	168
PcCel45A	*P. chrysosporium*	BAG68300.1	5kjo	B3Y002	30	4.3	144

[a] CBMs removed.

acid Asp132. The sidechain of Asp132 is positioned between Thr20 and Tyr22 from the adjacent beta-strand, and conserved hydrogen bonds connect the sidechains in the order Asp132-Thr20-His130.

In order to anticipate possible interactions with substrates, the MeCel45A structure was superposed with available GH45 ligand complexes, and protein-ligand interactions were analyzed using LIGPLOT (Figs. S1, S2). The structures chosen for comparison were: i) AcCel45A with two cellobiose molecules bound in subsites −3/−2 and +1/+2, respectively (PDB code 5XBX); ii) HiCel45A D10N mutant in complex with cellohexaose where two cellotriosyl units are seen in subsites −4/−3/−2 and +1/+2/+3, respectively (PDB code 4ENG; Fig. 5A); and iii) PcCel45A with two cellopentaose molecules bound in subsites −5 to −1 and +1 to +5, respectively (PDB code 3X2M; Fig. 5B). In the following, residue numbers refer to MeCel45A unless indicated otherwise.

The position of sugar residues in subsites +1/+2 is very similar in all the structures with several interactions in common. The glucose unit at +1 is bound by hydrogen bonds between O4 and the catalytic acid (Asp132) and between O6 and the alternate catalytic base (Asn109). At subsite +2 the 6-hydroxyl is held in place by hydrogen bonds to the backbone N and O atoms of Asn21 and to the sidechain of an asparagine (Asn147), except in PcCel45A where the latter interaction is instead with a backbone O atom (Gly131 in PcCel45A; Fig. 5). The subfamily B enzymes also have a hydrogen bond between Trp112 NE1 and O3 that is not present in HiCel45A or PcCel45A. There are several additional interactions in HiCel45A formed by the tunnel-enclosing loops that cover the +1 subsite and partially subsite +2.

While the position of sugar units is similar at +1/+2, and presumably at −1, cellulose binding deviates towards both ends of the active site. At subsite +3 there is a small shift in the position of the glucose residue between HiCel45A and PcCel45A. However, both positions would clash with a protein loop in MeCel45A (at Gly84-Gln85) as well as in AcCel45A and TrCel45A, suggesting that either the cellulose chain takes on a different orientation in subfamily B enzymes from subsite +3 and onwards, or the loop assumes a different conformation when accommodating a substrate.

Towards the other end of the active site the −1 subsite is only occupied in PcCel45A but the mode of binding is likely similar in all enzymes due to the high degree of conservation of the structures here. The 6-hydroxymethyl arm of the sugar unit is deeply buried at the bottom of the cleft and is used as a handle for positioning by hydrogen bonding to the catalytic acid (Asp132) and by hydrophobic binding to the tyrosine conserved at this site (Tyr22). On the other side of the sugar ring, O3 is H-bonded to the alternate base (Asn109). At subsites −2 and −3 the sugar positions are very similar in AcCel45A and PcCel45A. The cellotrioside in HiCel45A is slightly shifted at subsite −2 and displays increasing deviation over subsites −3 and −4 relative to the cellopentaose in PcCel45A, showing that the orientation of the cellulose chain differs between these enzymes. The substrate binding in MeCel45A at −2/−3 is likely similar to that seen in AcCel45A but may deviate from PcCel45A at subsite −4 due to the difference in position of the tryptophan residue that forms a sugar-binding platform at this subsite. All the enzymes have a tryptophan sidechain exposed at subsite −4, but this residue occupies different positions in the sequence in the respective subfamilies and are oriented differently in the structures (Fig. 5). The sidechain indole of Trp64 in MeCel45A (and Trp68 in AcCel45A) is shifted around 4.5 Å and is tilted roughly 30 degrees relative to Trp154 in PcCel45A, suggesting that a sugar residue at subsite −4 would likely be tilted to a similar extent. In HiCel45A it is Trp18 that acts as the sugar-binding platform at this site.

2.4. Enzymatic activity

The hydrolytic activity of family GH45 endoglucanases HiCel45A, MeCel45A, TrCel45A and PcCel45A were evaluated on soluble fractions

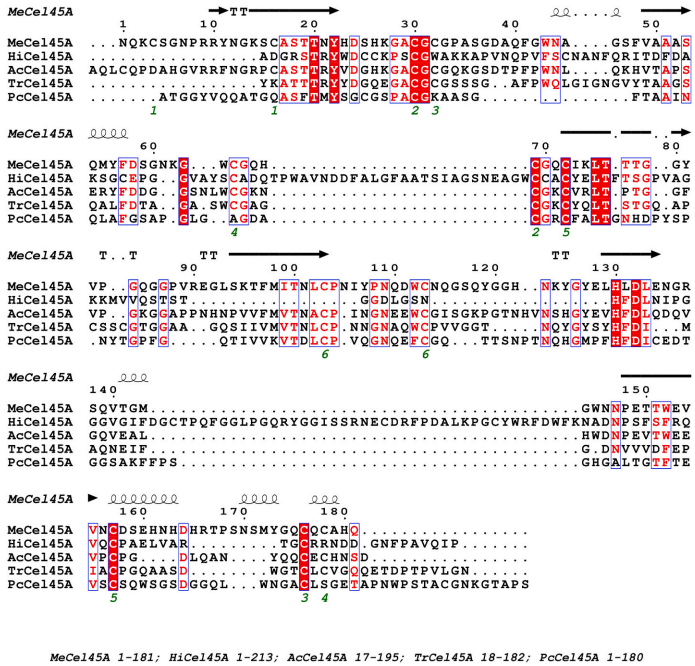


Fig. 3. Sequence alignment of *M. edulis* Cel45A, *H. insolens* Cel45A, *A. crossean* Cel45A *T. reesei* Cel45A and *P. chrysosporium* Cel45A catalytic modules. Alignment visualized in ESPript 3.0. Secondary structure elements of MeCel45A are represented as springs (α-helices) and arrows (β-strands). Character coloration according to ESPript 3.0: green numbers indicate cysteine pairings; filled red box and a white character indicate strict identity; red character – similarity within a group; blue frame – similarity across groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

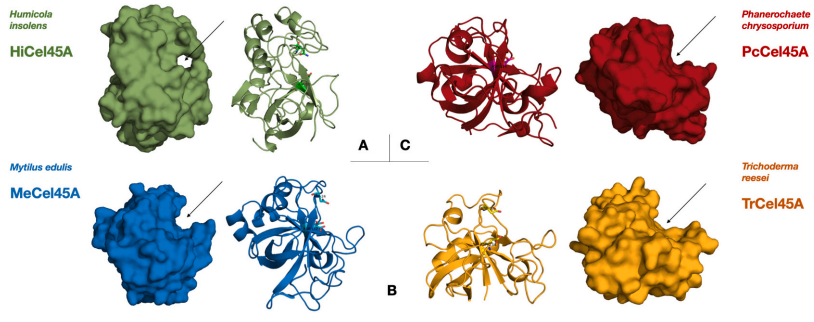


Fig. 4. Surface views of GH45 endoglucanases from subfamily A, B and C. Arrows point to substrate binding area in HiCel45A (PDB: 4ENG), PcCel45A (PDB: 5KJO), MeCel45A (PDB: 1WC2), TrCel45A (GenBank: CAA83846.1, homology model). Catalytic acid and base are shown as sticks in the cartoon representation. Letters A, B, C indicate subfamily membership.

of barley beta-glucan (BG), konjac glucomannan (GM) and carboxymethyl cellulose (CMC) (Fig. 6).

Activity was expressed as formation of reducing ends measured by PHBAH assay. For all proteins the highest initial hydrolysis rates were observed on beta-glucan, for which the initial production of reducing ends was 2–7 times more rapid than on CMC (Table 2). On both CMC and BG, MeCel45A showed the second highest initial rate, preceded by HiCel45A and followed by PcCel45A and TrCel45A in that order.

With GM as a substrate the enzymes did not show any linear phase at the start of the reaction, probably due to its heteropolymer nature, and reliable initial rates could not be determined for GM. Therefore, in order to gain a general understanding of the enzyme relative initial rates on the different substrates we chose an initial product concentration that was covered by all experiments (70 μM reducing ends), and then compared the time needed to reach this concentration among the enzymes (Table 2). For all enzymes GM was hydrolyzed faster than CMC, but much more slowly than BG. The highest activity on GM was exhibited by HiCel45A and TrCel45A, followed by MeCel45A and PcCel45A in that order.

With BG as a substrate the reaction rapidly leveled off and appeared to reach an end point within less than 1 h. Therefore, a 60 min time point was used to calculate the yield of reducing ends from BG. The activity on

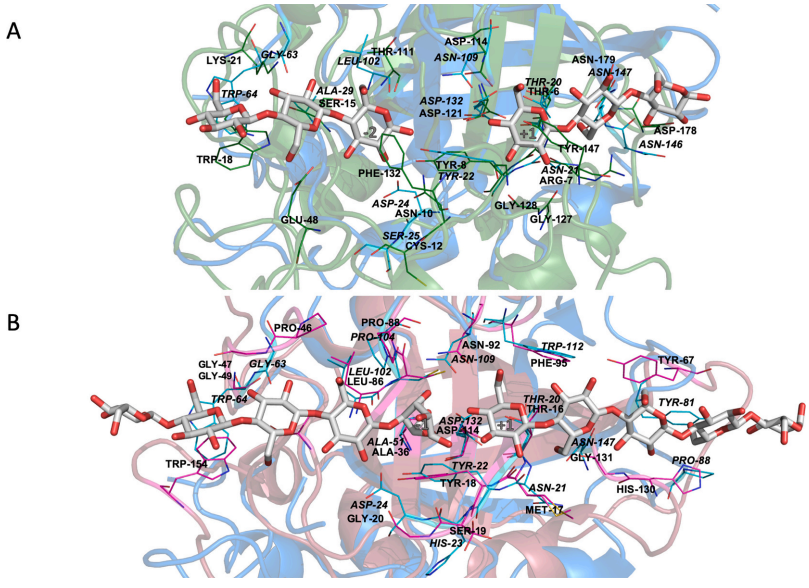


Fig. 5. Residues likely to interact with substrate in the active-site cleft of MeCel45A. (A) The active-site cleft of MeCel45A (PDB: 1WC2) aligned with the structure of HiCel45A (PDB: 4ENG) with two cellotriose molecules bound in the substrate binding groove; (B) The active-site cleft of MeCel45A (PDB: 1WC2) aligned with the structure of PcCel45A (PDB: 3X2M) with two cellopentaose molecules bound in the substrate binding groove. Substrate interacting residues are displayed as lines, and residue labels are in italics for MeCel45A.

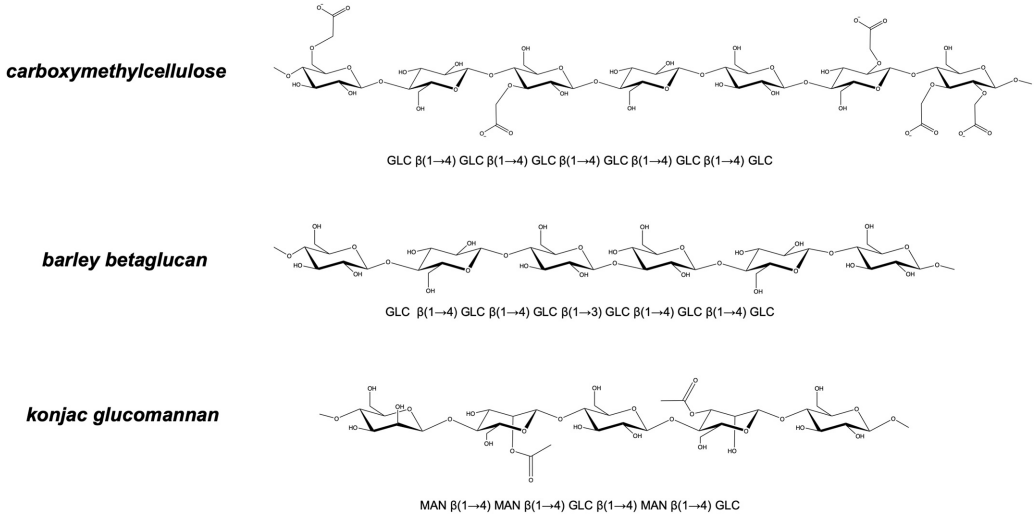


Fig. 6. General representations of the chemical structures of carboxymethyl cellulose (CMC), barley betaglucan (BG), and konjac glucomannan (GM).

CMC and GM did also level off, but no clear end point was reached, and the yield was instead calculated from the amount of reducing ends after 24 h of hydrolysis. The yield is a measure of how many bonds in the polymers the respective enzymes are able to hydrolyze.

The HiCel45A enzyme gave significantly higher yields on BG and on CMC than the other enzymes, especially on BG with more than twice as high concentration of reducing ends (Table 2). However, on GM the pattern was quite different. The highest yield of reducing ends from GM

Table 2
Family GH45 endoglucanase substrate specificity. Initial rate of formation and yield of reducing ends on 0.1% carboxymethyl cellulose (CMC), barley beta-glucan (BG) or konjac glucomannan (GM) as substrate. ±SD is standard deviation of triplicate determinations.

Name	Subfamily	Initial rate on 0.1% substrate		Yield on 0.1% substrate from 0.1 μM enzyme			Required time for 0.1 μM Cel45A to produce 70 μM of reducing sugar on 0.1% substrate		
		CMC	BG	CMC	BG	GM	CMC	BG	GM
		(μM/μM)min^{-1} ± SD	(μM/μM)min^{-1} ± SD	μM ± SD, 24 h (*22 h)	μM ± SD, 60 min	μM ± SD, 24 h	t, min	t, min	t, min
HiCel45A	A	235 ± 9,9	1139 ± 58,5	188 ± 4,8	547 ± 23,5	192 ± 4,4	>60	<1	10
MeCel45A	B	183 ± 9,7	296 ± 43,5	128 ± 3,3	229 ± 6,6	308 ± 42,0	>60	<5	60
TrCel45A	B	12 ± 4,1	90 ± 37,3	92 ± 4,0*	225 ± 4,0	1087 ± 43,2	>100	<30	30
PcCel45A	C	56 ± 5,0	109 ± 58,6	104 ± 14,7	195 ± 20,2	240 ± 22,5	>60	<10	>60

was observed with TrCel45A and the lowest with HiCel45A. TrCel45A deviated from the other enzymes in its activity on GM. The rate of hydrolysis did not decline to the same extent over time as with the other enzymes. The yield of reducing ends for TrCel45A on GM after 24 h was nearly 10 times higher than it was on CMC after a similar period of time.

Enzymatic activities of MeCel45A and TrCel45A were additionally evaluated on cellohexaose at +4 °C and +30 °C. Samples were analyzed every 65 min for 6 h using high performance anion exchange chromatography (HPAEC). The main cleavage products of MeCel45A on cellohexaose were cellobiose and cellotetraose, while the main products of TrCel45A were cellobiose, cellotriose and cellotetraose (Fig. 7A–D). At 30 °C MeCel45A and TrCel45A showed similar cellohexaose degradation rates (1.38 ± 0.05 min^{-1} and 1.26 ± 0.20 min^{-1}, respectively), but at 4 °C these enzymes performed significantly differently. For MeCel45A the activity at 4 °C was 72% of that at 30 °C, whereas it was only 35% for TrCel45A (Fig. 7E).

2.5. NMR spectroscopy

The products from enzyme degradation of GM and BG were investigated by NMR spectroscopy to observe differences in specificity on the two substrates. Degradation of GM at 0.3 μM enzyme concentration was observed by following the formation of reducing end α-Glc and yielded initial build-up curves similar to those observed from PHBAH assays (Fig. S3A). It confirmed that all four enzymes are acting through an inverting mechanism, because α-Glc was formed rapidly before equilibrium with the β anomer was reached.

In order to obtain higher levels of degradation products from GM, the enzyme concentration was increased to 35 μM. This caused reducing end α-Glc to be formed very quickly before reducing end β-Glc was formed by mutarotation (Fig. 8 and S3B), which further confirmed the inverting mechanism of the enzymes. More interestingly, the formation of mannose reducing ends was also observed (Fig. 8). Among the four tested enzymes, TrCel45A was the most efficient in mannose cleavage, whereas HiCel45A, MeCel45A and PcCel45A were similar in efficiency after 23 h. However, HiCel45A and TrCel45A gave rise to the steepest increase of Man reducing ends during the first 6 h, whereas MeCel45A was the slowest.

Degradation of GM at high enzyme concentration (35 μM) also allowed the detection of other degradation products. The ^{1}H NMR spectra after 23 h of degradation clearly shows a difference between HiCel45A and the other enzymes (Fig. S4). Degradation by HiCel45A produced non-reducing end glucose, whereas the other three enzymes produced non-reducing end mannose (Fig. 8E and F). The non-reducing end residues formed instantly, similar to the reducing-end glucose residues. Non-reducing end mannose was not detected from HiCel45A degradation and only small amounts of non-reducing end glucose was detected from TrCel45A, MeCel45A and PcCel45A degradation. In addition, non-reducing end 2-acetylated mannose (2Ac-Man) was formed from degradation by TrCel45A, MeCel45A and PcCel45A with the highest amount from TrCel45A (Fig. S3C). Furthermore, a small amount of monomeric glucose was observed from HiCel45A degradation (Fig. S4) in contrast to the other enzymes.

Enzyme degradation of beta-glucan was also monitored by NMR with a low enzyme concentration (0.3–0.5 μM). Formation of reducing-end α-glucose was followed (Fig. S5) and corresponded well with the results from PHBAH assays. The reducing-end residues formed were linked at position 4 rather than position 3 (Fig. S6), showing a preference for cleavage next to a β(1 → 4) linkage. This preference was the same for all the four enzymes.

3. Discussion

All of the Cel45A enzymes tested in this study were able to produce reducing ends on barley beta-glucan, konjac glucomannan and carboxymethyl cellulose. Barley beta-glucan was degraded most rapidly among the three selected substrates. It is a linear glucose homopolymer with β(1 → 3) and β(1 → 4) linkages, where three or four successive β(1 → 4) linked glucose residues are followed by one pair of β(1 → 3) linked residues (Fig. 6). During NMR experiments involving barley beta-glucan, we found that the reducing-end residues were β(1 → 4) and not β(1 → 3) linked. This shows that the Cel45A enzymes require a β(1 → 4) linkage between subsites −2 and −1. However, further studies are required to draw conclusions about restrictions for other subsites. The NMR analyses confirm that Cel45A from *M. edulis*, *T. reesei* and *P. chrysosporium* hydrolyze glycosidic bonds with an inverting action mechanism, thus proving that this mechanism is indeed common among all GH45 subfamilies known to date. An inverting catalytic mechanism has long been proposed for GH45 enzymes, but to our knowledge had only been experimentally proven for subfamily A member Cel45A from *H. insolens* (Schou et al., 1993).

NMR spectroscopy on glucomannan revealed that HiCel45A can cleave β(1 → 4) linkage between mannose and glucose (Man-Glc) as pointed out by the cyan arrow in Fig. 8A, but the rate and preference for such cleavage is low. However, HiCel45A was much faster at producing non-reducing end glucose alongside reducing-end glucose, indicating cleavage between two glucose residues (Glc-Glc) and a preference for such linkages. Cleavage by HiCel45A also led to formation of monomeric glucose, which was not observed from the other enzymes. Another major difference was the absence of non-reducing end mannose in the HiCel45A product profile, whereas in the case of TrCel45A, MeCel45A and PcCel45A the formation of non-reducing end mannose and reducing-end glucose was fast and simultaneous, suggesting a cleavage between glucose and mannose (Glc-Man, magenta arrow in Fig. 8A) and a preference for such linkages. Furthermore, a slow formation of reducing-end mannose was accompanied by an increase of non-reducing end mannose in TrCel45A, MeCel45A and PcCel45A, which indicates cleavage between two mannose residues (Man-Man), thus mannanase activity. While all of the Cel45A enzymes showed an activity against glucomannan, Cel45A from *T. reesei* was outstanding in its ability to continue glucomannan degradation even after 24-hour incubation. We attribute this to the aforementioned mannanase activity, which was also demonstrated in a previous study by Karlsson et al. (2002). An interesting observation was made regarding the levels of non-reducing end

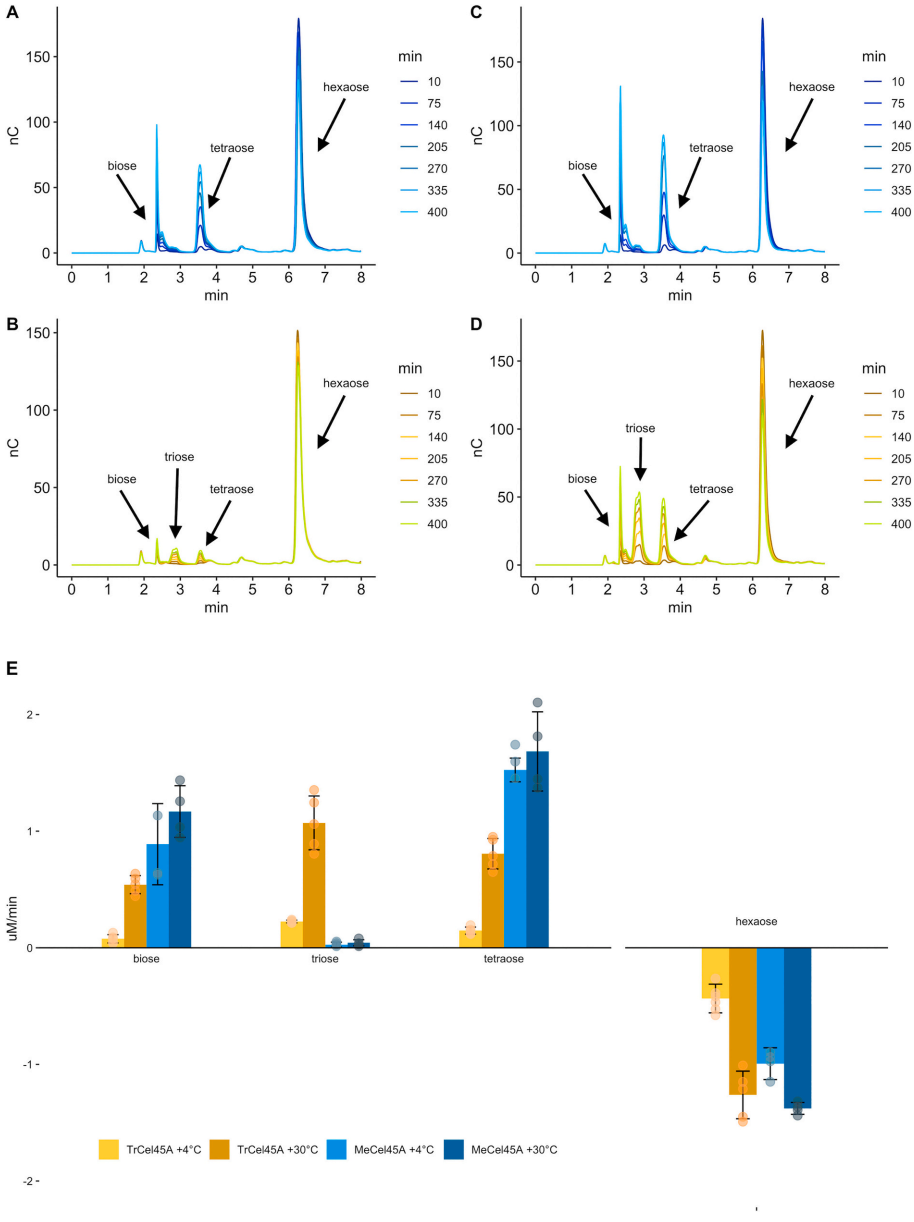


Fig. 7. Enzymatic activity of *M. edulis* Cel45A and *T. reesei* Cel45A on cellohexaose at +4 °C and +30 °C. HPAEC chromatogram time-lapse of product formation from cellohexaose degradation by MeCel45A at (A) +4 °C and (C) +30 °C, by TrCel45A at (B) +4 °C and (D) +30 °C. Chromatogram legends represent timepoints of degradation. (E) Hydrolytic activity is expressed as μM/min per 1 μM of enzyme of MeCel45A and TrCel45A on cellohexaose. The product formation rate is shown with positive values and substrate degradation rate with negative values. Average enzymatic activity rates are depicted as bars and individual datapoints as filled circles. Error bars represent standard deviation (n ≥ 3).

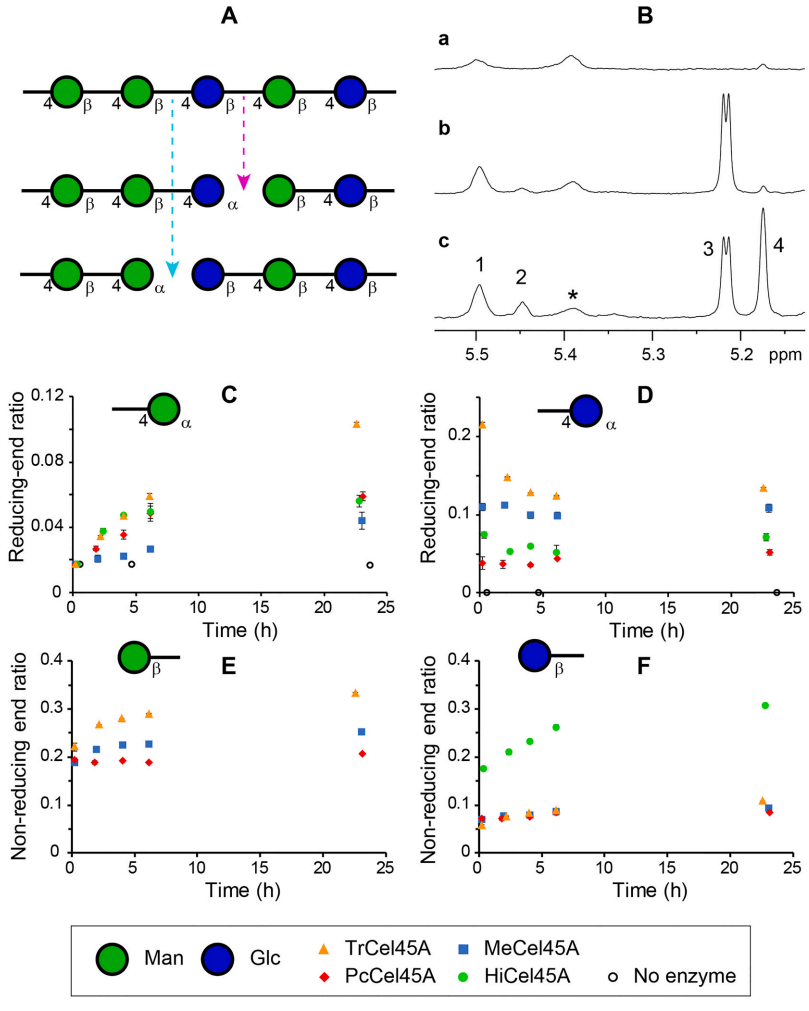


Fig. 8. (A) Simplified formula of a glucomannan chemical structure and exemplified degradation pathways. (B) ^{1}H NMR spectra from a) GM control, b) GM after 12 min TrCel45A degradation, and c) GM after 24 h TrCel45A degradation. Signal assignments are 1) 2Ac-Man H2, 2) non-reducing end 2Ac-Man H2, 3) reducing end Glc H1-α, and 4) reducing end Man H1-α. The asterisk highlights a starch-like impurity. NMR-derived progress curves are plotted to show the formation of (C) reducing-end α-Man, (D) reducing-end α-Glc, (E) non-reducing end Man, and (F) non-reducing end Glc from degradation of glucomannan with 35 μM enzyme. Non-reducing end Man was not detected in samples with HiCel45A. Error bars correspond to ±1 standard deviation.

acetylated mannose residues. Cel45A from *H. insolens* did not facilitate formation of products with an acetylated mannose residue at the non-reducing end. As expected, the cleavage product reducing-end glucose residues were alpha-anomers for all enzymes, which once again confirms the inverting nature of these endoglucanases.

All of the evaluated Cel45A enzymes were able to hydrolyze CMC, a synthesized cellulose derivative with carboxymethyl group substitutions (Fig. 6). Hydrolysis rates were comparatively slower and with lower yields of reducing ends, presumably due to the presence of carboxymethyl substituents on some glucose residues. Interestingly, HiCel45A gave a higher reducing end yield than the other enzymes, suggesting more tolerance for substitutions despite having a more enclosed active site. It would be reasonable to assume that such bulky substituents could restrict the binding or cleavage in narrow active sites due to steric hindrance.

Of the two subfamily B enzymes, MeCel45A retained over 70% of its activity at +4 °C relative to +30 °C, compared to 35% for TrCel45A (Fig. 7E). Considering the tropical habitat of *T. reesei* and the boreal habitat of *M. edulis*, this could indicate GH45 enzyme evolution towards cold adaptation in the mollusc. A broad optimum temperature range has previously been demonstrated by Xu et al. (2000). Further investigation of how the structure of this enzyme has evolved to retain high activity at

low temperatures could be an interesting study. The cellohexaose cleavage product profile differed between the enzymes (Fig. 7A–D). Cel45A from *M. edulis* produced predominantly cellobiose and cellotetraose, which implies a distinct organization of substrate in the active site cleft. Cel45A from *T. reesei* produced cellotriose in addition to cellobiose and cellotetraose, indicating a less definitive cleavage preference.

Structure and sequence comparisons reveal several similarities and differences between Cel45A from *M. edulis* and HiCel45A, AcCel45A, TrCel45A and PcCel45A. While the DPBB core superposes closely, there are variations in surface structures between the subfamilies due to differences in the lengths of flanking loop regions. In this regard, subfamily B and C enzymes are more similar to each other than to the subfamily A enzyme, where longer loops enclose the catalytic center, thus creating a tunnel-like structure. In the subfamily C member PcCel45A, loops at both ends of the substrate-binding cleft make the cleft longer than in the other enzymes. MeCel45A appears to have the narrowest substrate binding cleft among the subfamily B members. Substrate binding and the catalytic mechanism of MeCel45A are most likely very similar to that in AcCel45A, since their active sites are nearly identical. In all of the enzymes a tryptophan residue is found at subsite −4 (Trp64 in MeCel45A), which most likely serves as a sugar binding platform. Interestingly, this residue comes from different locations in the GH45 sequences, possibly underlining the functional significance of this residue. Godoy et al. (2018) mutated the corresponding tryptophan residue in PcCel45A (W154A), which led to 50% decrease in catalytic activity on lichenan.

There is no GH45 structure available yet with substrate spanning over the active site with an un-cleaved bond connecting the −1 and +1 subsites. There are only a few structures where the −1 subsite is occupied by the reducing-end residue of an oligosaccharide, but the conformations differ between all of them. A distorted $^{4}H_{5}$ half-chair is refined in a cellotetraose complex with *Thielavia terrestris* Cel45A of subfamily A (PDB code 5GLY) (Gao et al., 2017), whereas three variants of PcCel45A of subfamily C in complex with cellopentaose display $^{3}S_{1}$ skew, distorted $^{2}H_{1}$ half chair, and $^{4}C_{1}$ chair conformations, respectively, of the glucose residue at subsite −1 (wildtype, N92D, N105D variants; PDB codes 3X2M, 3X2H, 3X2K)(Nakamura et al., 2015).

In the absence of clear information from X-ray crystallography of how a sugar residue will bind at the −1 subsite prior to, during, and after cleavage, the catalytic reaction has been examined by computational methods. QM/MM calculations and transition path sampling (TPS) molecular dynamics (MD) simulations of celloheptaose hydrolysis by HiCel45A showed a conformational itinerary for the glucose unit at subsite −1, from $^{4}C_{1}$ chair to $^{2}S_{O}$ skew in the Michaelis complex, over a $^{2,5}B$ boat transition state, via $^{5}S_{1}$ skew to a $^{1}C_{4}$ chair for the α-anomer product (Bharadwaj et al., 2020). This type of catalytic itinerary, $^{2}S_{O} \rightarrow {}^{2,5}B^{\ddagger} \rightarrow {}^{5}S_{1}$, has been proposed earlier for other GH families, including inverting beta-glucoside active GH6 and GH8, and both retaining (GH11, GH120) and inverting (GH43) beta-xylosidases (Ardèvol & Rovira, 2015).

While the role of the catalytic acid seems to be well established in all three GH45 subfamilies, the assignment of catalytic base function is more uncertain. The first assignment of catalytic residues was based on structures of HiCel45A of subfamily A, where Asp121 was proposed to act as general acid, implied by its hydrogen bonding to the glycosidic oxygen (4OH) of a glucose residue at subsite +1, and Asp10 was proposed as the most likely general base. This was supported by complete inactivation of HiCel45A upon mutations at these sites (D121N and D10N mutants) (Davies et al., 1995). Consequently, these residues are considered to be indispensable for hydrolysis by GH subfamily A members. However, a role was also implicated for Asp114, since a HiCel45A D114N mutant only retained less than 1% activity compared to wildtype. It is interesting that in the subfamily A enzyme an Asp at this location is much more active than Asn, whereas it seems to be the opposite for subfamily C. PcCel45A has an Asn at this location and the activity was drastically reduced upon mutation to Asp (N92D mutant) (Nakamura et al., 2015).

Members of GH45 subfamily C lack the Asp residue corresponding to the base in subfamily A. Instead, Nakamura et al. (2015) proposed Asn92 as the catalytic base, on the basis of structure studies of PcCel45A where high-resolution X-ray and neutron diffraction analyses revealed amide/imidic acid tautomerization of Asn92 and a proton relay-network that links Asn92 to the catalytic acid (Asp114 in PcCel45A) through a chain of hydrogen bonds. Nearly all of the residues in the proton-relay chain in PcCel45A (Asn92, Cys96, Phe95, Asn105, Ser14, His112, Thr16, Asp114) are also present in MeCel45A (Asn109, Cys113, Trp112, Asn123, Ser18, His130, Thr20, Asp132), but most are not conserved in HiCel45A.

In the subfamily B enzymes, MeCel45A, AcCel45A and TrCel45A, both of the candidate base residues are present. Mutation of either of them in AcCel45A drastically reduced the activity but did not lead to complete inactivation (Nomura et al., 2019). The activity reduction was rather similar for the D27A and N112A mutants, about 30-fold and 60-fold, respectively. Based on these results and the position of the residues relative to a glucoside unit modelled at subsite −1 of AcCel45A, Nomura et al. (Nomura et al., 2019) proposed that Asn112 in its imidic acid form acts as the catalytic base that activates a water molecule for nucleophilic attack at the anomeric carbon, whereas Asp27 is of primary importance for productive positioning of the glucose residue at subsite −2. However, the orientation of the modelled glucoside is quite different from previous observations and models. The glucose residue modelled at subsite −1 of AcCel45A is flipped so that the 2OH and 3OH groups point towards the bottom of the cleft, while the 6OH arm is pointing out from the active site. The anomeric carbon is exposed on the side of the glucose ring that is facing Asn112 and not Asp27 (Nomura et al., 2019). This is in contrast to the crystal structures with sugar bound at subsite −1 as well as the QM/MM MD study by Bharadwaj et al. (2020). There, the glucose residue is instead oriented with its 6OH arm bound at the bottom of the cleft, while the 2OH and 3OH groups point out from the active site, which exposes the anomeric carbon on the side of the ring that is facing the Asp residue and not on the side where the Asn residue is located.

Considering the high structural similarity between AcCel45A and MeCel45A, the corresponding residues in *M. edulis* Cel45A (Asp24, Asn109 and Asp132) should be of similar importance. Asp132 is the catalytic acid, but which residue is acting as the catalytic base, Asp24 or Asn109, remains an open question. Further research is obviously needed to fully elucidate the catalytic mechanism of GH45 and its subfamilies at the molecular level.

4. Conclusions

Our results show that the GH45 enzymes studied here share several common properties. All can hydrolyze barley beta-glucan, konjac glucomannan and carboxymethylcellulose with an apparent preference for barley beta-glucan. Hydrolysis by these enzymes leads to inversion of configuration at the anomeric carbon, thus indicating an inverting action mechanism. We also demonstrated a few key differences such as mannanase activity among subfamily B and C members, and the ability of subfamily A member to produce monomeric glucose. We pointed out a variation in product profile within subfamily B and a possible evolutionary cold adaptation of the enzyme in blue mussel.

5. Methods

5.1. Extraction and purification of MeCel45A

MeCel45A from Blue Mussel, *Mytilus edulis*, (UniProt entry P82186) was prepared as described (Xu et al., 2000) with minor modifications – only the digestive gland of the mussel was used and the acid precipitation and heat precipitation steps were omitted. The blue mussels were collected from waters off the Swedish west coast, frozen and their digestive glands excised (total 1.42 kg glands from 28.7 kg of whole

frozen mussels). The digestive glands were homogenized in a meat blender (cooled with ice), stirred for 30 min and insoluble materials were removed by centrifugation. The target protein was captured by immobilized metal affinity chromatography (IMAC) on a STREAMLINE Chelating gel (Amersham Biosciences) saturated with zinc ions and elution with 20 mM sodium phosphate, pH 7.0, 1 M NaCl, 50 mM EDTA, followed by size-exclusion chromatography on Superdex 75 pg (Amersham Biosciences) and finally cation exchange chromatography on a Mono S column (Amersham Biosciences) eluted with a gradient of 0–300 mM NaCl in 20 mM sodium acetate, pH 5.5. MeCel45A eluted at ~200 mM NaCl. Activity against CMC during MeCel45A purification was assayed using dinitrosalicylic acid as reducing sugar reagent according to Bhat and Wood (1998). The purified enzyme was concentrated by ultra-filtration in 50 mM sodium acetate, 0.2 M NaCl, pH 5.5, and was stored at +4 °C prior to crystallisation screening, and at −20 °C for further use. The final yield was 8 mg of pure MeCel45A enzyme (extinction coefficient of 41,660 M^{-1} cm^{-1}). Further details are provided in the Supplementary Information file.

5.2. Preparation of HiCel45A, PcCel45A, TrCel45A

All protein stock solutions were kept in 10 mM sodium acetate buffer pH 5 with protein concentration of less than 1.0 mg/mL, stored at room temperature directly before use, at +4 °C the day before use and −20 °C for further use.

Purified HiCel45A (GenBank: QCH00668, ext. coefficient 0.1% = 1.956) from *Humicola insolens* (including enzymatic deglycosylation with endo H, followed by size exclusion chromatography on Superdex 75) was kindly provided by Kristian Bertel Rømer Mørkeberg Krogh at Novozymes A/S, Denmark.

Methanol-induced heterologous expression of PcCel45A (GenBank: BAG68300, ext. coefficient 0.1% = 1.270) was done in *Komagataella pastoris* KM71H expression system according to Igarashi and colleagues (Igarashi et al., 2008). Culture filtrate, supplemented with 1 M $(NH_4)_2SO_4$, was applied on a hydrophobic interaction chromatography (HIC) column (Phenyl Sepharose; elution buffer 10 mM NaAc pH 5.0). The collected protein solution was desalted on Biogel P6 column into a 20 mM TrisHCl pH 8.0 buffer. Desalting was followed by anion exchange chromatography (DEAE Sepharose) with an elution buffer 0.5 M NaCl 20 mM TrisHCl pH 8.0 using the following elution gradient: 0–50% 400 mL; 50–100% 100 mL; 100% 200 mL. Size exclusion chromatography (Superdex 75, buffer: 10 mM NaAc pH 5.0) was carried out as a final purification step.

TrCel45A from *Trichoderma reesei* (GenBank: CAA83846, ext. coefficient 0.1% = 2.003) was kindly provided by Matti Siika-aho, VTT Biotechnology, Finland. The full-length TrCel45A enzyme (with CD-linker-CBM1) was obtained from culture filtrate of a *T. reesei* strain lacking the genes expressing the two major endoglucanases Cel7B and Cel5A and was purified as described in Karlsson et al. (2002).

5.3. Crystallisation and structure determination of MeCel45A

Crystallisation was done at room temperature by hanging drop vapour diffusion (McPherson, 1982). Crystals were grown with 12 mg/mL protein solution in 50 mM sodium acetate, pH 5.5, 0.2 M NaCl, mixed with and equilibrated against 0.6 M sodium acetate, pH 5.5, and 0.125 or 0.375 M NaCl. The heavy-atom derivative was obtained by adding a grain of solid of Baker's dimercurial ($C_{10}H_{16}Hg_2O_6$; 1,4-diacetoxymercuri-2,3-dimethoxybutane; Anatrace) to drops with crystals. Crystals to be used for X-ray analysis were kept for 1–5 min in a cryoprotectant solution (30% monomethyl PEG 5000, 12.5% glycerol, 0.1 M sodium morpholine-ethane-sulfonic acid, pH 6.0, and 10 mM $CoCl_2$) and then flash-cooled using liquid nitrogen. X-ray diffraction data were recorded at 100 K, in-house on a CuKα RIGAKU/R-AXIS IIC system (native, 1.85 Å; dimercurial, 2.0 Å), and at beamline ID14-3, ESRF, France (native, 1.2 Å resolution). Diffraction data were indexed, processed and scaled with the HKL program package (Otwinowski & Minor, 1997). The structure was solved with SIRAS (single isomorphous replacement with anomalous scattering) using a derivative with Baker's dimercurial. The final structure model was refined with Refmac5 (Murshudov et al., 1997) at 1.2 Å resolution and final R and R_{free}-values of 0.147 and 0.162. Further details are provided in the Supplementary Information file. Statistics of X-ray diffraction data sets and structure refinement are summarized in Tables S1 and S2, respectively. Atomic coordinates of the structure model and the structure factors have been deposited with the Protein Data Bank (wwpdb.org) (wwPDB consortium, 2019) as PDB entry 1WC2.

5.4. Structure comparison

To include Cel45A from *T. reesei* in structure comparisons, a homology model was created for TrCel45A, built in SWISS-MODEL (Waterhouse et al., 2018) using MeCel45A (PDB code 1WC2; this work) as a template. The PyMOL Molecular Graphics System, Version 1.8.6.0 (Schrödinger, LLC) was used to align structures and investigate structural similarities. The structures were superposed using CEAlign command and structure deviation was expressed as root mean square deviation (RMSD, Å) over aligned Ca atoms, as generated by the CE algorithm. LIGPLOT v.4.5.3 (Wallace et al., 1995) was used for plotting protein-ligand interactions. The multiple sequence alignment was created using MUSCLE (Madeira et al., 2019) online at the EMBL-EBI server (ebi.ac.uk), later adjusted and visualized in ESPript (espript.ibcp.fr) (Robert & Gouet, 2014). The percentage sequence-identity matrix was created with Clustal 2.1 online using the EMBL-EBI server (ebi.ac.uk) (Madeira et al., 2019).

5.5. Reducing sugar assay

MeCel45A, TrCel45A, PcCel45A and HiCel45A were each incubated at 30 °C, 400 rpm, in 0.1 M sodium acetate buffer pH 5.0 on CMC (β-1,4 linkage, degree of carboxymethyl substitutions 0.60–0.95, dynamic viscosity 0.7–1.5 Pa.s, purity >99%, BioChemika, Fluka), konjac glucomannan (β-1,4 linkage, mannose: glucose = 60: 40, kinematic viscosity ~1.0×10^{-5} m^2/s, purity >98%, Megazyme) or barley betaglucan (mixed β-1,4 and β-1,3 linkages, kinematic viscosity 2.0–3.0×10^{-5} m^2/s, purity ~95%, Megazyme) dissolved in water. Reaction was stopped by adding 1 M sodium hydroxide solution in a 1:1 ratio and cooling samples on ice. Hydrolytic activity was determined colorimetrically by quantifying the reducing sugar formation with p-Hydroxybenzoic acid hydrazide (PHBAH, Sigma) solution (Lever, 1972): 0.1 M PHBAH; 0.2 M sodium potassium tartrate; 0.5 M sodium hydroxide solution. PHBAH solution was added in 1:1 ratio to the samples, followed by incubation at 95 °C for 15 min and cooling on ice for 10 min. Samples were kept at room temperature for 5 min before absorption was read at 410 nm.

5.6. Hydrolysis product analysis by high-performance anion-exchange chromatography (HPAEC)

0.075 μM MeCel45A and 0.075 μM TrCel45A each were incubated at +30 °C and +4 °C with 85 μM cellohexaose (Seikagaku Corporation, dissolved in water) in 0.1 M sodium acetate buffer pH 5.0.

Cellohexaose hydrolysis products were analyzed periodically (every 65 min for 6 h) using DIONEX ICS3000 equipped with Dionex CarboPac PA200 column, HPAE-PAD detector. Column was equilibrated in 150 mM NaOH/50 mM Na acetate. Peak separation was achieved at a flowrate 0.450 mL/min by three gradients: 1) from 150 mM NaOH/50 mM Na acetate to 150 mM NaOH/125 mM Na acetate in 20 min; 2) from 150 mM NaOH/125 mM Na acetate to 150 mM NaOH/250 mM Na acetate in 15 min; 3) 150 mM NaOH/250 mM Na acetate for 15 min. Separation was followed by a cleaning step gradient from 150 mM NaOH/250 mM Na acetate to 150 mM NaOH/50 mM Na acetate in 15 min. Cellobiose, cellotriose and cellotetraose (Seikagaku Corporation)

standards were used for peak identification and product quantification.

5.7. NMR spectroscopy

Degradation of beta-glucan and glucomannan with MeCel45A, TrCel45A, PcCel45A, and HiCel45A was investigated by NMR spectroscopy. The substrates were dissolved in D_2O to yield 1% stock solutions. Substrates and enzymes were mixed in 3 mm NMR tubes to a total volume of 160 μL, a final substrate concentration of 5–7 mg/mL, and 50 mM phosphate buffer (pH 5.0). The enzyme concentration was either 0.3–0.5 μM (for glucomannan and beta-glucan degradation) or 35 μM (for glucomannan degradation).

NMR spectra were recorded on a Bruker Avance III 600 MHz spectrometer using a 5 mm inverse detection cryoprobe equipped with z axis gradient. The spectra were acquired at 30 °C with the HDO signal as internal reference (δ 4.71 ppm). After the enzyme was added, the tube was quickly placed into the probe and hydrolysis was monitored as a function of time. The first spectrum was acquired within 10 min after the reaction started, and from then on spectra were acquired every 5–10 min. 1D spectra were obtained from the 1D NOESY experiment (noesygppr1d) for water suppression, with water presaturation during the relaxation delay (1 s) and the mixing time (50 ms). A sweep width of 6000 Hz was used, and 32 scans of 64 K data points were acquired. The enzyme and substrate concentrations were chosen to ensure fast enough reaction to allow determination of the glucose anomers before mutarotation occurred. Samples for glucomannan degradation with high enzyme concentration (35 μM) were run in duplicate for each enzyme to yield an error estimation.

Signal assignments were carried out with the help of various 2D NMR experiments, including $^1H,^1H$-TOCSY, $^1H,^1H$-NOESY, and $^1H,^{13}C$-HSQC, as well as comparison with previous assignments of beta-glucan (Petersen et al., 2013) and glucomannan (Mikkelson et al., 2013; Teleman et al., 2003) and predicted chemical shifts from the CASPER tool (Jansson et al., 2006; Lundborg & Widmalm, 2011).

Proton signals were integrated and normalized to yield ratios of reducing-end or non-reducing end residues to interior residues. Since H1 of interior Man and Glc residues of glucomannan were hidden by the water signal or affected by the water suppression, H2 of interior residues was used for normalization. Reducing-end Man was calculated from the ratio between reducing-end Man H1-α (5.17 ppm) and interior Man H2 (4.11 ppm). Reducing-end Glc was calculated from the ratio between reducing-end Glc H1-α (5.22 ppm) and interior Glc H2 (3.35 ppm) and from the ratio between reducing-end Glc H2-β (3.28 ppm) and interior Glc H2 (3.35 ppm). For partly overlapping signals and signals with very low abundance, peak heights were used rather than integrals. By this means, non-reducing end 2Ac-Man was calculated from the ratio between non-reducing end 2Ac-Man H2 (5.45 ppm) and interior 2Ac-Man H2 (5.50 ppm). Similarly, non-reducing end Man was calculated from the ratio between non-reducing end Man H2 (4.05 ppm) and interior Man H2 (4.11 ppm), and non-reducing end Glc was calculated from the ratio between non-reducing end Glc H2 (3.30 ppm) and interior Glc H2 (3.35 ppm). For beta-glucan, reducing-end Glc was calculated from the ratio between the integrated signal of reducing-end Glc H1-α (5.23 ppm) and β(1 → 4)-linked Glc H1 (4.54 ppm).

Funding sources

This work was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), Grant Number 2017-01130. GJK was supported by the Swedish Structural Biology Network (SBNet), the Swedish Natural Science Research Council (NFR) and the Royal Swedish Academy of Sciences (KVA). EJ was supported by the Swedish Structural Biology Network (SBNet). KI acknowledges Grant-in-Aid for Innovative Areas (No. 18H05494 to K.I.) from the Japanese Ministry of Education, Culture, Sports and Technology (MEXT).

Database

Structural data have been deposited with the Protein Data Bank under accession code 1WC2.

CRediT authorship contribution statement

Laura Okmane: Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Gustav Nestor:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Emma Jakobsson:** Validation, Formal analysis, Investigation, Data curation, Writing – original draft. **Bingze Xu:** Methodology, Formal analysis, Investigation, Resources, Writing – original draft. **Kiyohiko Igarashi:** Investigation, Resources, Supervision. **Mats Sandgren:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Gerard J. Kleywegt:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Supervision, Funding acquisition. **Jerry Ståhlberg:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

No real or perceived conflicts.

Acknowledgements

We thank Jan-Christer Janson for initiation of the investigation of enzymes from Blue mussel, Sabah Mahdi for help with protein crystallisation, Christina Divne for help with initial phasing, Matti Siika-aho and Inés G. Muñoz for preparation and purification of TrCel45A enzyme, Miao Wu for help with cultivation of *K. pastoris*, and Kristian Bertel Rømer Mørkeberg Krogh at Novozymes A/S, Denmark, for kindly providing the HiCel45A enzyme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carbpol.2021.118771.

References

Ardèvol, A., & Rovira, C. (2015). Reaction mechanisms in carbohydrate-active enzymes: Glycoside hydrolases and glycosyltransferases. Insights from ab initio quantum Mechanics/Molecular mechanics dynamic simulations. *Journal of the American Chemical Society, 137*(24), 7528–7547. https://doi.org/10.1021/jacs.5b01156

Berto, G. L., Velasco, J., Tasso Cabos Ribeiro, C., Zanphorlin, L. M., Noronha Domingues, M., Tyago Murakami, M., Polikarpov, I., de Oliveira, L. C., Ferraz, A., & Segato, F. (2019). Functional characterization and comparative analysis of two heterologous endoglucanases from diverging subfamilies of glycosyl hydrolase family 45. *Enzyme and Microbial Technology, 120*, 23–35. https://doi.org/10.1016/j.enzmictec.2018.09.005

Bharadwaj, V. S., Knott, B. C., Ståhlberg, J., Beckham, G. T., & Crowley, M. F. (2020). The hydrolysis mechanism of a GH45 cellulase and its potential relation to lytic transglycosylase and expansin function. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.RA119.011406. jbc.RA119.011406.

Bhat, K. M., & Wood, T. M. (1998). Methods for measuring cellulase activities. *Methods Enzymol., 160*, 87–112.

Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L., & Murzin, A. G. (1999). A six-stranded double-psi β barrel is shared by several protein superfamilies. *Structure, 7*(2), 227–236. https://doi.org/10.1016/S0969-2126(99)80028-8

Cha, J.-H., Yoon, J.-J., & Cha, C.-J. (2018). Functional characterization of a thermostable endoglucanase belonging to glycoside hydrolase family 45 from Fomitopsis palustris. *Applied Microbiology and Biotechnology, 102*(15), 6515–6523. https://doi.org/10.1007/s00253-018-9075-5

Couturier, M., Feliu, J., Haon, M., Navarro, D., Lesage-Meessen, L., Coutinho, P. M., & Berrin, J.-G. (2011). A thermostable GH45 endoglucanase from yeast: Impact of its

atypical multimodularity on activity. *Microbial Cell Factories, 10*, 103. https://doi.org/10.1186/1475-2859-10-103

Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C., & Schülein, M. (1995). Structures of oligosaccharide-bound forms of the endoglucanase V from Humicola insolens at 1.9 a resolution. *Biochemistry, 34*(49), 16210–16220. https://doi.org/10.1021/bi00049a037

Gao, J., Huang, J.-W., Li, Q., Liu, W., Ko, T.-P., Zheng, Y., Xiao, X., Kuo, C.-J., Chen, C.-C., & Guo, R.-T. (2017). Characterization and crystal structure of a thermostable glycoside hydrolase family 45 1,4-β-endoglucanase from Thielavia terrestris. *Enzyme and Microbial Technology, 99*, 32–37. https://doi.org/10.1016/j.enzmictec.2017.01.005

Gilbert, H. J., Hall, J., Hazlewood, G. P., & Ferreira, L. M. A. (1990). The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspecies cellulosa constitutes a cellulose-binding domain that is distinct from the catalytic centre. *Molecular Microbiology, 4*(5), 759–767. https://doi.org/10.1111/j.1365-2958.1990.tb00646.x

Godoy, A. S., Pereira, C. S., Ramia, M. P., Silveira, R. L., Camilo, C. M., Kadowaki, M. A., Lange, L., Busk, P. K., Nascimento, A. S., Skaf, M. S., & Polikarpov, I. (2018). Structure, computational and biochemical analysis of pc Cel45A endoglucanase from phanerochaete chrysosporium and catalytic mechanisms of GH45 subfamily C members. *Scientific Reports, 8*(1), 3678. https://doi.org/10.1038/s41598-018-21798-9

Guo, R., Ding, M., Zhang, S.-L., Xu, G., & Zhao, F. (2008). Molecular cloning and characterization of two novel cellulase genes from the mollusc Ampullaria crossean. *Journal of Comparative Physiology B, 178*(2), 209–215. https://doi.org/10.1007/s00360-007-0214-z

Igarashi, K., Ishida, T., Hori, C., & Samejima, M. (2008). Characterization of an endoglucanase belonging to a new subfamily of glycoside hydrolase family 45 of the basidiomycete Phanerochaete chrysosporium. *Applied and Environmental Microbiology, 74*(18), 5628–5634. https://doi.org/10.1128/AEM.00812-08

Jansson, P.-E., Stenutz, R., & Widmalm, G. (2006). Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel web-based version of the computer program Casper. *Carbohydrate Research, 341*(8), 1003–1010. https://doi.org/10.1016/j.carres.2006.02.034

Karlsson, J., Siika-aho, M., Tenkanen, M., & Tjerneld, F. (2002). Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of Trichoderma reesei. *Journal of Biotechnology, 99*(1), 63–78. https://doi.org/10.1016/S0168-1656(02)00156-6

Kaur, H. (1997). In , *1. Hydrolases in the crystalline style of a common fresh water mussel, Lamellidens corrianus (Lea)* (pp. 67–72) (2).

Kikuchi, T., Jones, J. T., Aikawa, T., Kosaka, H., & Ogura, N. (2004). A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode Bursaphelenchus xylophilus. *FEBS Letters, 572*(1–3), 201–205. https://doi.org/10.1016/j.febslet.2004.07.039

Lee, S. J., Kim, S. R., Yoon, H. J., Kim, I., Lee, K. S., Je, Y. H., Lee, S. M., Seo, S. J., Dae Sohn, H., & Jin, B. R. (2004). CDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry Longicorn beetle, Apriona germari. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 139*(1), 107–116. https://doi.org/10.1016/j.cbpc.2004.06.015

Lever, M. (1972). A new reaction for colorimetric determination of carbohydrates. *Analytical Biochemistry, 47*(1), 273–279. https://doi.org/10.1016/0003-2697(72)90301-6

Liu, G., Wei, X., Qin, Y., & Qu, Y. (2010). Characterization of the endoglucanase and glucomannanase activities of a glycoside hydrolase family 45 protein from Penicillium decumbens 114–2. *The Journal of General and Applied Microbiology, 56*(3), 223–229. https://doi.org/10.2323/jgam.56.223

Loo, L. (1992). Filtration, assimilation, respiration and growth of Mytilus edulis L. at low temperatures. *Ophelia, 35*(2), 123–131. https://doi.org/10.1080/00785326.1992.10429974

Lundborg, M., & Widmalm, G. (2011). Structural analysis of glycans by NMR chemical shift prediction. *Analytical Chemistry, 83*(5), 1514–1517. https://doi.org/10.1021/ac1032534

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., … Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research, 47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

McPherson, A. (1982). *Preparation and analysis of protein crystals*. John Wiley & Sons.

Mei, H.-Z., Xia, D.-G., Zhao, Q.-L., Zhang, G.-Z., Qiu, Z.-Y., Qian, P., & Lu, C. (2016). Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGaseI) in the beetle Batocera horsfieldi. *Gene, 576*(1, Part 1), 45–51. https://doi.org/10.1016/j.gene.2015.09.057

Mikkelson, A., Maaheimo, H., & Hakala, T. K. (2013). Hydrolysis of konjac glucomannan by Trichoderma reesei mannanase and endoglucanases Cel7B and Cel5A for the production of glucomannooligosaccharides. *Carbohydrate Research, 372*, 60–68. https://doi.org/10.1016/j.carres.2013.02.012

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

Nakamura, A., Ishida, T., Kusaka, K., Yamada, T., Fushinobu, S., Tanaka, I., Kaneko, S., Ohta, K., Tanaka, H., Inaka, K., Higuchi, Y., Niimura, N., Samejima, M., & Igarashi, K. (2015). "Newton's cradle" proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Science Advances, 1*(7), Article e1500263. https://doi.org/10.1126/sciadv.1500263

Newell, C., Shumway, S., Cucci, T. L., & Selvin, R. (1989). The effects of natural seston particle size and type on feeding rates, feeding selectivity and food resource availability for the mussel Mytilus edulis linnaeus, 1758 at bottom culture sites in Maine. *Journal of Shellfish Research, 8*, 187–196.

Nomura, T., Iwase, H., Saka, N., Takahashi, N., Mikami, B., & Mizutani, K. (2019). High-resolution crystal structures of the glycoside hydrolase family 45 endoglucanase EG27II from the snail Ampullaria crossean. *Acta Crystallographica Section D: Structural Biology, 75*(4), 426–436. https://doi.org/10.1107/S2059798319003000

Onishi, T., Suzuki, M., & Kikuchi, R. (1985). In , *51. The distribution of polysaccharide hydrolase activity in gastropods and bivalves* (pp. 301–308) (2).

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology, 276*, 307–326.

Pauchet, Y., Kirsch, R., Giraud, S., Vogel, H., & Heckel, D. G. (2014). Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica. *Insect Biochemistry and Molecular Biology, 49*, 1–13. https://doi.org/10.1016/j.ibmb.2014.03.004

Pauchet, Y., Wilkinson, P., Chauhan, R., & Ffrench-Constant, R. H. (2010). Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PloS One, 5*(12), Article e15635. https://doi.org/10.1371/journal.pone.0015635

Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., & Beckham, G. T. (2015). Fungal cellulases. *Chemical Reviews, 115*(3), 1308–1448. https://doi.org/10.1021/cr500351c

Petersen, B. O., Olsen, O., Beeren, S. R., Hindsgaul, O., & Meier, S. (2013). Monitoring pathways of β-glucan degradation by enzyme mixtures in situ. *Carbohydrate Research, 368*, 47–51. https://doi.org/10.1016/j.carres.2012.12.006

Purchon, R. D. (1977). *The biology of the mollosca* (2nd ed.). Pergamon Press.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research, 42*(W1), W320–W324. https://doi.org/10.1093/nar/gku316

Sakamoto, K., & Toyohara, H. (2009). Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam Corbicula japonica. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology, 152*(4), 390–396. https://doi.org/10.1016/j.cbpb.2009.01.010

Saloheimo, A., Henrissat, B., Hoffrén, A. M., Teleman, O., & Penttilä, M. (1994). A novel, small endoglucanase gene, egl5, from Trichoderma reesei isolated by expression in yeast. *Molecular Microbiology, 13*(2), 219–228. https://doi.org/10.1111/j.1365-2958.1994.tb00417.x

Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B., & Schülein, M. (1993). Stereochemistry, specificity and kinetics of the hydrolysis of reduced cellodextrins by nine cellulases. *European Journal of Biochemistry, 217*(3), 947–953. https://doi.org/10.1111/j.1432-1033.1993.tb18325.x

Song, J. M., Hong, S. K., An, Y. J., Kang, M. H., Hong, K. H., Lee, Y.-H., & Cha, S.-S. (2017). Genetic and structural characterization of a thermo-tolerant, cold-active, and acidic Endo-β-1,4-glucanase from Antarctic springtail, Cryptopygus antarcticus. *Journal of Agricultural and Food Chemistry, 65*(8), 1630–1640. https://doi.org/10.1021/acs.jafc.6b05037

Teleman, A., Nordström, M., Tenkanen, M., Jacobs, A., & Dahlman, O. (2003). Isolation and characterization of O-acetylated glucomannans from aspen and birch wood. *Carbohydrate Research, 338*(6), 525–534. https://doi.org/10.1016/S0008-6215(02)00491-3

Tsuji, A., Tominaga, K., Nishiyama, N., & Yuasa, K. (2013). Comprehensive enzymatic analysis of the cellulolytic system in digestive fluid of the sea hare Aplysia kurodai. Efficient glucose release from sea lettuce by synergistic action of 45 kDa endoglucanase and 210 kDa ß-glucosidase. *PLoS One, 8*(6), Article e65418. https://doi.org/10.1371/journal.pone.0065418

Valencia, A., Alves, A. P., & Siegfried, B. D. (2013). Molecular cloning and functional characterization of an endogenous endoglucanase belonging to GHF45 from the western corn rootworm,Diabrotica virgifera virgifera. *513*(2), 260–267. https://doi.org/10.1016/j.gene.2012.10.046

Vlasenko, E., Schülein, M., Cherry, J., & Xu, F. (2010). Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology, 101*(7), 2405–2411. https://doi.org/10.1016/j.biortech.2009.11.057

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, 8*(2), 127–134. https://doi.org/10.1093/protein/8.2.127

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

wwPDB consortium. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research, 47*(D1), D520–D528. https://doi.org/10.1093/nar/gky949

Xu, B., Hellman, U., Ersson, B., & Janson, J.-C. (2000). Purification, characterization and amino-acid sequence analysis of a thermostable, low molecular mass endo-β-1,4-glucanase from blue mussel,Mytilus edulis. *267*(16), 4970–4977. https://doi.org/10.1046/j.1432-1327.2000.01533.x

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2023:41

Glycoside hydrolase family 45 enzymes (GH45) are small inverting cellulases most commonly found in fungi, which bear a structural resemblance to non-hydrolytic protein groups such as plant expansins and fungal loosenins. This thesis investigated the differences of GH45 enzymes across subfamilies and between members of the same subfamily. Enzymatic activity studies revealed dissimilarities in hydrolysis product profile, suggesting varying substrate specificities. Structural studies revealed differences in catalytic site residues and the shape of substrate binding area.

Laura Okmane received her graduate education at the Department of Molecular Sciences, SLU, Uppsala. She received her Master's degree in Biology from the University of Latvia.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-134-4
ISBN (electronic version) 978-91-8046-135-1