THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Structural and functional investigation of underexplored carbohydrate-active enzyme families

ANDREA SEVESO

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

Structural and functional investigation of underexplored carbohydrate-active enzyme families
Andrea Seveso
ISBN 978-91-7905-905-7

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5371
ISSN 0346-718X

Division of Industrial Biotechnology
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
The experimentally determined tridimensional structure of *Pv*CE15 (**Paper I,** lower) compared to the model predicted by AlphaFold2 (upper). On the right, superimposed with one another. The two structures are almost exactly the same. For more considerations see section 3.4.2.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

"An army of other explorers, the explorers of science, operates in laboratories, and it will be the sum of their efforts that will give life to great feats."

- Walter Bonatti

Preface

This dissertation serves as partial fulfillment of the requirements for obtaining the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The work was supported by grant awarded to Johan Larsbrink by the Knut and Alice Wallenberg Foundation within the Wallenberg Wood Science Center (WWSC). The PhD studies were carried out between April 2019 and September 2023 under the supervision of Assoc. Prof. Johan Larsbrink and co-supervision of Dr. Lauren McKee (KTH). The thesis was examined by Prof. Pernilla Wittung-Stafshede.

The majority of the work in this thesis was carried out at the Division of Industrial Biotechnology (IndBio) at Chalmers University of Technology. Crystallization experiments were performed at the University of Gothenburg or at Copenhagen University and X-ray diffraction data were collected at the MAX IV Laboratory, Lund and at the European Synchrotron Radiation Facility (ESRF) by Scott Mazurkewich or Sanchari Banerjee.

Andrea Seveso
September 2023

Structural and functional investigation of underexplored carbohydrate-active enzyme families

Andrea Seveso

Division of Industrial Biotechnology
Department of Life Sciences
Chalmers University of Technology

Abstract

The known consequences of the current fossil-based economy require a transition towards a bio-based economy. Development of biorefineries in which plant biomass can be utilized as a renewable source of energy and building blocks to produce both commodities and high-value products, is a key step in this transition. Lignocellulosic biomass has, however, evolved a highly complex architecture to be recalcitrant to degradation, and this represents a major challenge in its utilization. In nature, a wide variety of microorganisms has evolved to exploit lignocellulose as carbon source. They produce carbohydrate-active enzymes (CAZymes) to degrade lignocellulose polymers into components that can be utilized for their growth. CAZymes therefore represent powerful tools that could be utilized in industrial settings for the degradation of plant biomass.

In this thesis, I investigated different CAZymes belonging to relatively unexplored families. The aim was to expand our yet limited knowledge and to gain further insights into their physiological roles. Bacterial enzymes belonging to the carbohydrate esterase family 15 (CE15) were identified in putative pectin-targeting polysaccharide utilization loci (PULs) – clusters of co-regulated genes coding for proteins involved in the degradation of specific polysaccharide motifs. These CE15 enzymes showed comparable activities on model substrates mimicking pectin-esters and on canonical model substrates. This result led to study their activity also on extracted pectins and pectin-rich biomass, although no new activities were revealed. X-ray protein crystallography was used to obtain structures of *Pv*CE15, also in complex with the sugar moiety of the model substrates, to gain insight into its likely specificity. A broader selection of CE15 enzymes of both fungal and bacterial origin was characterized on an additional, non-conventional, model substrate to define their substrate specificity in regards of the position of the ester substituents in the targeted bond. Furthermore, one of the first bacterial copper radical oxidases, belonging to an unexplored clade of the Auxiliary Activity family 5 (AA5), was heterologously produced and characterized on a wide range of alcohol substrates. Finally, I determined the structure of a previously characterized AA9 lytic polysaccharide monooxygenase with broad substrate specificity, indicating certain structural features as possible determinants of the described specificity.

Keywords: carbohydrate-active enzymes, carbohydrate esterase, copper radical oxidases, lytic polysaccharide monooxygenase, plant biomass degradation, protein structure determination, protein structure prediction

List of publications

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Seveso A.**, Mazurkewich S., Banerjee S., Lo Leggio L., Larsbrink J. Polysaccharide utilization loci from Bacteroidota encode CE15 enzymes with possible role in cleaving pectin-lignin bonds. *manuscript*

II. **Seveso A**.*, Coleman T.*, Carbonaro M., Krogh K., Lo Leggio L., Larsbrink J. The substrate specificity of CE15 glucuronoyl esterases suggest distinct roles in processing different esters in lignin-carbohydrate complexes. *manuscript*

III. Mazurkewich S*, **Seveso A**.*, Larsbrink J. A unique AA5 alcohol oxidase fused with a catalytically inactive CE3 domain from the bacterium *Burkholderia pseudomallei.* FEBS Lett. 2023 May 4. doi: 10.1002/1873-3468.14632. Print Aug 2023. PMID: 37143387

IV. Mazurkewich S.*, **Seveso A**.*, Hüttner S, Brändén G, Larsbrink J. Structure of a C1/C4-oxidizing AA9 lytic polysaccharide monooxygenase from the thermophilic fungus *Malbranchea cinnamomea*. Acta Crystallogr D Struct Biol. 2021 Aug 1;77(Pt 8):1019-1026. doi: 10.1107/S2059798321006628. Epub 2021 Jul 29. PMID: 34342275; PMCID: PMC8329866

*These authors contributed equally.

Contribution summary

I. First author. I planned and performed most of the experiments, analyzed the data, interpreted the results, and drafted the manuscript. Shared in structural work (where I solved the structure of the apo-enzyme) and in editing of the manuscript.

II. First author (shared). Shared the experimental work, analyzed the data and interpreted the results together with my co-authors. I performed the structural work and shared the writing of the manuscript.

III. First author (shared). Shared the experimental work, where I focused more on the production of all the different constructs and divided the workload for the biochemical characterization. Analyzed the data and interpreted the results together with my co-authors. Shared the writing of the manuscript.

IV. First author (shared). Shared in the structural work. I drafted the manuscript and shared in editing of the manuscript.

Table of contents

Abbreviations

CAZymes – carbohydrate-active enzymes
CE15 – carbohydrate esterase family 15
LCC – lignin carbohydrate complex
PULs – polysaccharide utilization loci
AA – auxiliary activity
4-*O*-MeGlcA – 4-*O*-methylglucuronic acid
GlcA – glucuronic acid
GalA – Galacturonic acid
HG – homogalacturonan
RGI – rhamnogalacturonan I
RGII – rhamnogalacturonan II
ORF – open reading frame
PULDB – polysaccharide utilization loci database
UDH – uronate dehydrogenase
PL – polysaccharide lyase family
GT – glycosyltransfer
FPLC – Fast Protein Liquid Chromatography
IMAC– Immobilized Metal Ion Affinity Chromatography
PME – pectin methylesterase
DNSA – dinitrosalicylic acid
pNP – para-nitrophenol
BnzGlcA – glucuronate benzyl ester
AllylGlcA – glucuronate allyl ester
MeGlcA – glucuronate methyl ester
MeGalA – galacturonate methyl ester
BnzGalA – galacturonate benzyl ester
3ppGlcA – 3-phenylpropyl glucuronate
IC – ion chromatography
MS – mass spectrometry
NMR – nuclear magnetic resonance

Chapter I: Introduction

The world's current linear fossil-based economy poses a sword of Damocles on our society in the form of ruinous consequences for the environment and crisis of resources [1]. With the evidence of these perils inevitably becoming apparent to most people, there has been an increasing drive for the adoption of more sustainable and climate-friendly alternatives with the goal of mitigating climate change. If most noticeable efforts have been made in the energy production and in private transportation [2], an important challenge is also represented by the need to replace plastics, their derivates and life cycles and the petrochemical industry. In this context the need of transition towards a circular bioeconomy takes shape, where renewable plant biomass (lignocellulose) is used as a resource to replace fossil-derived alternatives. Biorefineries are a key infrastructure for this transition, facilities where lignocellulose is converted into materials, chemicals, fuels, heat and energy carriers, from commodities like bioethanol, to high-value products [3].

1.1 Enzymes in biorefineries

Woody biomass has evolved to the complex structure we see today, to provide protection for the plant against the external environment, so it is not surprising that it is highly recalcitrant to degradation and difficult to process in industrial settings. Depending on the final product, it could be of interest to extract and purify the single polymers constituting lignocellulose, or to break it down to oligomers, for use as high-value building blocks for several applications. The complete saccharification is also viable, to then utilize single sugars as feed for fermentative processes [4-6]. The development of efficient pre-treatment and separation techniques is thus crucial in a biorefinery setting, and there is a strong drive to replace the established harsh chemical treatments that have been predominantly utilized so far, with enzymatic catalysis [7]. Enzymes are not only interesting for the sake of making the processes more sustainable - by reducing environmentally harmful waste products generated by traditional methods - but also for the several benefits that their utilization could provide. Enzymes are typically very specific, which allows to use them as catalysts without the risk of unwanted side reactions. They can target specific substrates and are also regio- and stereoselective [8, 9], minimizing the risks of generating unwanted isomers and of the desired product(s). Enzymes can also work under mild conditions such as ambient temperature and moderate pH values, thus reducing the cost associated in maintaining specific harsh conditions that are often needed in chemical processes such as high temperatures and pressures, or very acidic or basic conditions [10-12]. On the other hand, enzymes may need to be developed to meet pre-existing or "mandatory" conditions of industrial processes. This is of course an important challenge in the development of biorefineries, but it can be addressed. For instance, biodiversity is so vast that in principle it is possible to find enzymes in nature that would work in a specific industrial setting. In fact, microorganisms have evolved to colonize virtually every environment on earth, surviving in very extreme conditions. And these microorganisms produce enzymes that are able to work in these environmental conditions, so a good strategy would be to look for an enzyme of interest by a niche of microorganisms in a specific environment. A possibility

is also that of *engineering* already known enzymes to generate variants of interest, but this topic will not be object of this thesis. What it is worth highlighting here though is the important role of scientific research in finding new enzymes that could be suitable for industrial application. However, determining the industrial use of newly found enzymes is another big topic, and will not be discussed in this thesis, which represents a collection of studies focused on the discovery and fundamental characterization of novel enzymes.

1.2 Aims of this thesis

The overall aim of this thesis was to expand the knowledge on relatively understudied families of carbohydrate-active enzymes (CAZymes), by discovering new members of the families and conducting both their biochemical and structural characterization. This is a very broad and open-ended goal, and more specific ones related to individual projects/papers were developed throughout the thesis work.

I started my PhD studies with one main project regarding a series of putative enzymes from carbohydrate esterase family 15 (CE15). Enzymes from this family are known to cleave ester bonds between lignin and xylan in lignin carbohydrate complexes (LCCs). The targets I started to study were identified in polysaccharide utilization loci (PULs): large clusters of genes encoding the proteins needed to deconstruct specific polysaccharides. The research question here was to understand why these enzymes were located in PULs, some of which predicted to be targeted towards pectin rather than the expected glucuronoxylan. Considering the proposed biological role of this family of enzymes, the fact that the genes were located in these clusters, was rather interesting. *Did this mean that the family includes activities yet not described*? The goal of this work then, reported in **Paper I**, was to produce and characterize these enzymes to gain further insights on their activities and reasons that they were found in in the selected putative PULs. Another project focused on CE15 enzymes aimed to investigate the substrate specificity of these enzymes towards two different esters species found in the LC bond they target, α- and γ-ester. The aim was to assess whether the enzymes could exhibit a certain preference towards a specific type of the two ester bonds. Selected members were thus characterized on model substrate mimicking the different esters species allowing to assess their substrate specificities. This work is described in **Paper II**.

One way to get a broader understanding of an enzyme family is to explore its diversity. This was our approach adopted in **Paper III,** where we characterized a bacterial member of auxiliary activity family 5 (AA5), a family in which almost only fungal members have previously been studied. Here my goals were to assess why a putative galactose oxidase domain was fused together in a predicted multicatalytic enzyme to a putative acetylxylan esterase domain in the genome of a pathogenic bacterium. . Multi-modularity is common in bacterial CAZymes and, in many cases, the domains work together to achieve a synergistic effect on a common substrate, but the target and function of this particular modular enzyme were not easy to predict.

To get a better understanding of an enzyme's mode of action, often a structural analysis is pursued, and it is something that was done in each of the papers included in this thesis, whether using x-ray crystallography-determined structures or models computationally

predicted by new powerful tools that became available during the course of my doctoral studies. In **Paper IV** specifically we experimentally determined the structure of a fungal lytic polysaccharide monooxygenase from auxiliary activity family 9 (AA9) that had been previously biochemically characterized, with the aim of gaining insights into the structural determinants of its peculiar substrate specificity.

Chapter II: Lignocellulose

Lignocellulose is the most abundant renewable bio-material in the world. The term lignocellulose refers to the matrix formed by various carbon-based polymers that constitutes the plant cell wall, comprising polysaccharides such as cellulose (40-50 %), and hemicellulose (20-40 %), and the aromatic polymer lignin (20-30 %) [13]. The relative abundances vary significantly depending on the type of plant [14].

2.1 The plant cell wall

In lignocellulose, plant cells are enclosed in a strong, protective cell wall, a complex hierarchical structure composed of different layers or *strata*: primary cell wall, secondary cell wall and middle lamella [15]. The outermost layer is represented by the middle lamella, a substance rich in pectins that cements together the primary walls of adjacent cells. The primary cell wall is formed during cell division and is a thin and malleable layer protecting the plasma membrane of every plant cell, while accommodating for their expansion during development. The secondary cell wall on the other hand is formed only in some cell types in certain plants, and is deposited between the primary wall and the plasma membrane only once the cell has stopped growing, making the cell wall much ticker, more rigid and responsible for the stiffness and strength of the larger plant, that allow it to grow upright [16, 17]. The composition of all these layers is determined by a wide array of factors, from the plant species and cell type, together with several environmental factors [18].

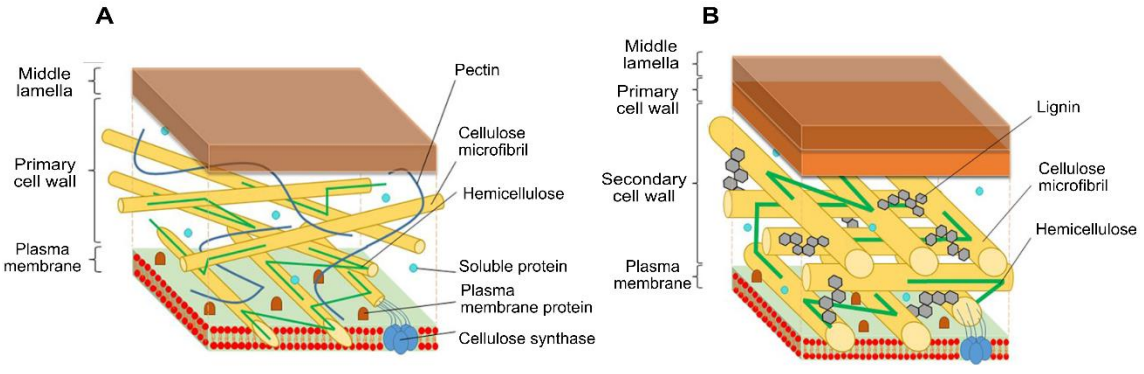

Figure 1. The structure and composition of the primary and secondary cell wall of plants. **A)** The primary cell wall is located outside of the plasma membrane and consists of cellulose microfibrils, hemicellulose, lignin and soluble proteins. **B)** The secondary cell wall is found between the primary cell wall and the plasma membrane. Lignin molecules are impregnated between cellulose microfibrils which appear more orderly arrange. (Figure adapted from Loix et al., 2017 [19] under the CC BY 4.0 license).

The primary cell wall generally consists of cellulose microfibrils interwoven with hemicelluloses in a network embedded in pectin with sparse glycoproteins. When present, the secondary cell wall is generally the thickest layer of the cell wall (Figure 1), comprising the bulk of the lignocellulose biomass and thus the most relevant in an industrial/biorefinery context [20]. It is a complex architecture (Figure 2) in which the "skeleton" cellulose chains are arranged together in microfibrils that can be entangled into macrofibrils, all covered by hemicelluloses of different nature depending on the plant species, which in turn are embedded in a matrix of lignin polymers. Minor levels of pectins are also present in comparison to the primary wall.

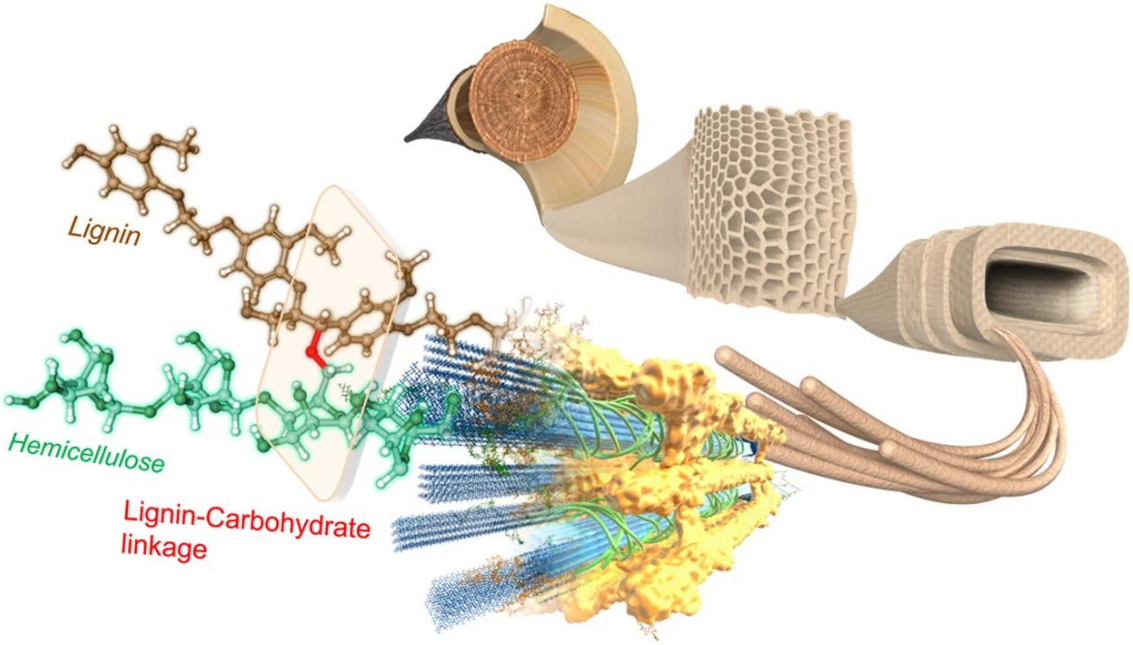

Figure 2. The hierarchical structure of wood and ultrastructure of secondary cell walls. showing the stem, wood cells, the cell wall layers and the arrangement of lignin and hemicelluloses around the organized cellulose microfibrils. (Figure adapted from Nishimura et al., 2018 [21], under the CC BY 4.0 license).

2.2 Polysaccharides

As described in the previous section, polysaccharides constitute the main components of the plant cell wall. These are comprised of covalently linked simple sugar molecules (monosaccharides) that form long chains, which can be thousands of monosaccharides long [22]. Shorter molecules are referred to as oligosaccharides, from two monosaccharides joined together that are named disaccharides, to trisaccharides, tetrasaccharides and so forth. The covalent bond between two sugars is called glycosidic bond and, the most common one, is formed between the hydroxyl group of the anomeric carbon of one monosaccharide and a hydroxyl group of the other. The glycosidic bond is classified as α-

or β- on the basis of the relative stereochemistry, *cis* (α-) or *trans* (β-) of the anomeric position and the stereocenter furthest from the anomeric carbon. The ends of the main chain of a polysaccharide can be referred to as reducing or non-reducing depending if the anomeric carbon is involved in a glycosidic bond (non-reducing) or is free to switch between an open chain and a closed ring forms, and available to be oxidized (reducing) [23]. From the main chain of a polysaccharide, side chains can branch off from the linear backbone, and can be composed of sugars different from the one in the main chain, in this case the macromolecule is referred to as heteropolysaccharide. On the other hand, a homopolysaccharide is constituted by identical monomers. There are many different kinds of polysaccharides that can be found in lignocellulosic biomass, but only the most relevant to this thesis are described in the following subsections.

2.2.1 Cellulose

The major component in wood is cellulose, is also the most abundant bio-polymer on earth [24], and it is linear homopolysaccharide of β-1,4-linked glucose monomers. Each monomer is rotated 180 degrees with respect to the adjacent ones, so the actual smallest repeating unit is the disaccharide cellobiose (Figure 3). The linear nature of the cellulose chains allows them to organize in a parallel fashion, holding together by hydrogen bonds in crystalline microfibrils that are insoluble in water and difficult to hydrolyze. Regions where these interactions are less regular are defined as amorphous cellulose, which results in more disordered, less compact and consequently more accessible polymer sections [25].

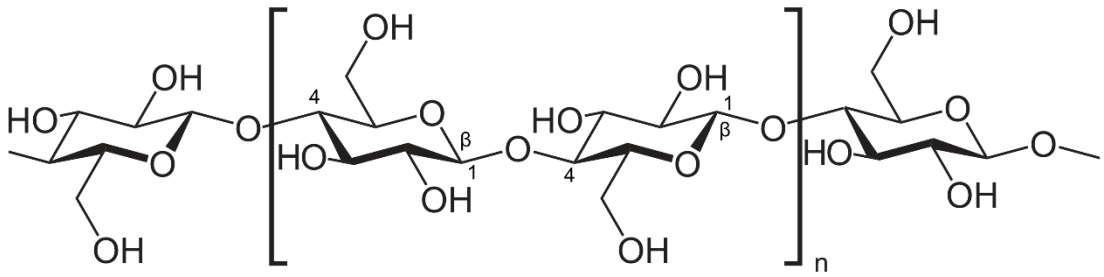

Figure 3. The chemical structure of cellulose, made of monomers of β-1,4-D-glucan. In brackets is the smallest repeating unit, cellobiose.

2.2.2 Hemicelluloses

Hemicellulose is the name that was originally given to a heterogeneous group of polysaccharides with large diversity of chemical composition, thought to be structurally similar to cellulose and defined by their extractability with alkaline solution and neutral charge [26]. Regardless of the ambiguous definition, hemicellulose are abundant in lignocellulose, contribute to the strength, porosity, and hydration of the cell wall [25], and represents an important feedstock in the biorefinery context [27]. They are typically heteropolymers (with exceptions such as mixed linkage β-glucans) with the backbone made

of one or more types of monosaccharides, and they are further classified and named on the basis of the major sugar component of the main chain. The most common hemicellulose polymers in lignocellulose are different variants of xylans and mannans, while others like mixed-linked glucans or xyloglucan are less abundant. As mentioned earlier they can be found coating the cellulose microfibrils and are embedded into the lignin matrix to which they can be also covalently linked, forming so called lignin carbohydrate complexes (LCCs) (Figure x)[28], a feature that contributes to the recalcitrance of lignocellulosic biomass to enzymatic hydrolysis [29].

Xylan is one of the most abundant hemicelluloses. It consists of a main chain of β-1,4-linked xylosyl residues, with side chains that can be of different nature depending on the source of the biomass. On the basis of their substituents, they can be further classified in homoxylan, heteroxylan, arabinoxylan, glucuronoxylan, and arabinoglucuronoxylan (figure 4) [30]. The latter two are characterized by α-1,2-linked 4-*O*-methylglucuronic acid (4-*O*-MeGlcA) substitutions. The various branching and substitutions patterns in xylans are extremely important, as are known to xylan properties and behavior [31], but are too numerous to be described individually in this thesis. Generally, other commonly found xylan substitutions include acetyl groups linked at the C2 or C3 position, or α-1,2 and α-1,3 l-arabinofuranoside units.

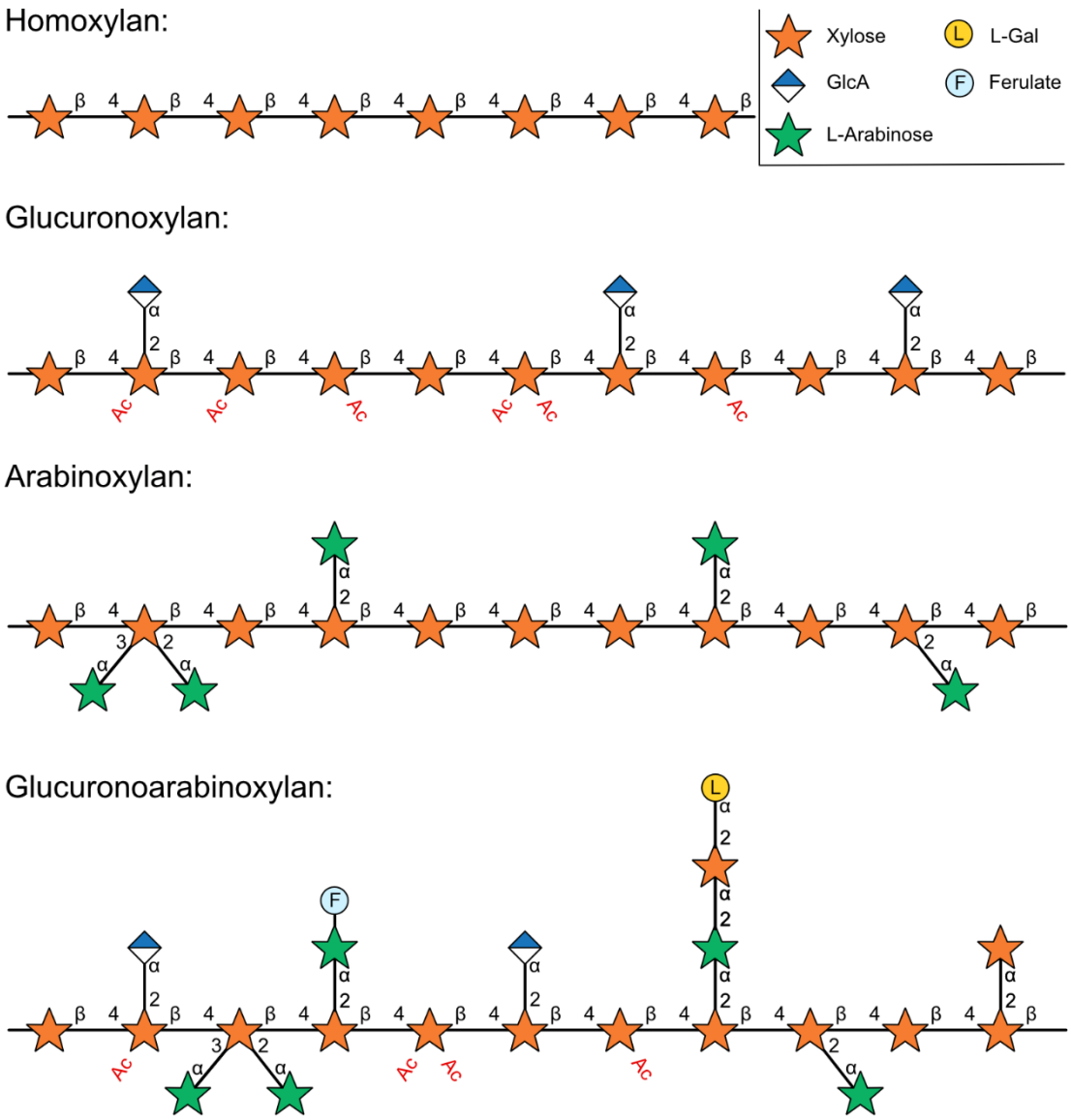

Figure 4. The main classes of xylan. Homoxylan, glucuronoxylan, arabinoxylan, and glucuronoarabinoxylan have the β-1,4-linked xylose backbone in common, but carbohydrate and non-carbohydrate decorations differ. Acetyl groups are common on the C2 and C3 positions of xylose units in glucuronoxylan and glucuronoarabinoxylan and are symbolized by "Ac" (red).

2.2.3 Pectin

Pectin is a class of heterogeneous polysaccharides rich in galacturonic acid (GalA) that are charged and not classified as hemicelluloses. Pectin can be found in many plant species, fruits and vegetables, where it contributes to structural strength and integrity, but it only represents a minor component in lignocellulose [32]. The most important pectic polysaccharides are regarded as homogalacturonan (HG) and rhamnogalacturonan (RG) I

and II (figure 5). The general structure consists predominantly of a backbone of α-1,4-linked galacturonic acid units (HG), that can be alternated with rhamnose residues (RG I). Rhamnose can also be the main component of the side chains (RG II). Additionally, the galacturonic acid residues can be methylated at the C6 position and *O*-acetylated at the C2 and/or C3 positions [33]. Pectin side branches can be very complex, being composed of several different types of monosaccharides and covalent linkages and their complete structure has likely not been fully elucidated. Some evidence have suggested that pectin, like hemicelluloses, could be covalently attached to lignin [34, 35]. Different pectic polysaccharides and pectin-rich biomass from different sources was used in a work described in this thesis to assess a putative pectin-targeting activity of glucuronoyl esterases (**Paper I**)

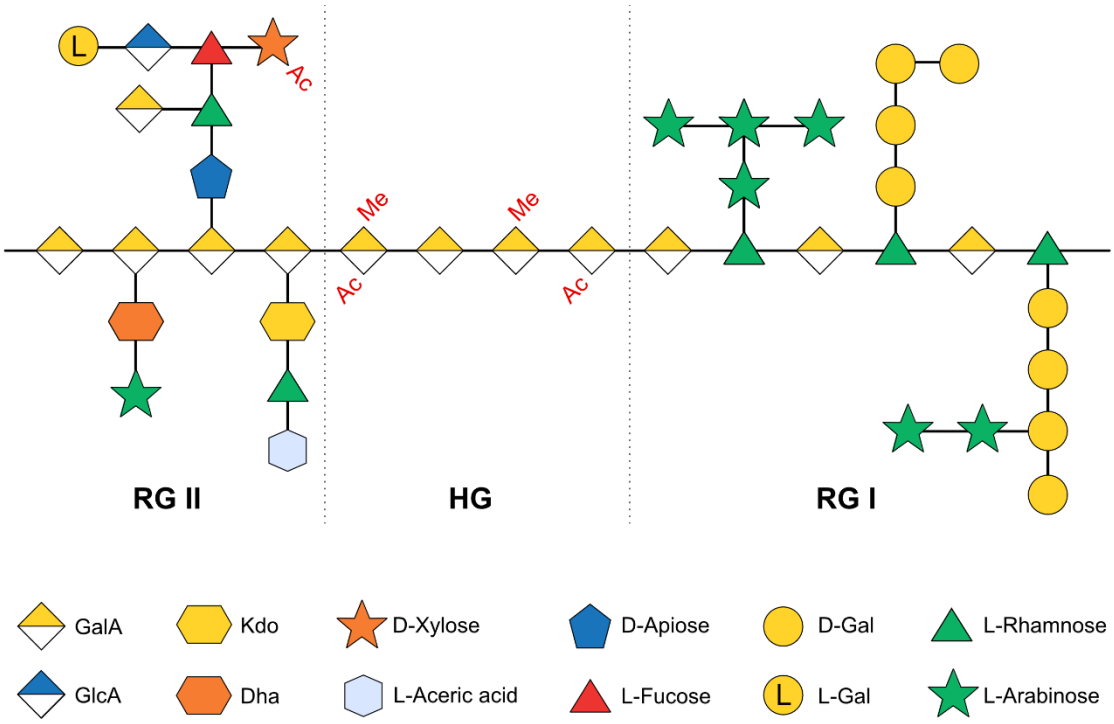

Figure 5. Structure of the main pectic polymers. Rhamnogalacturonan II (RG II), homogalacturonan (HG), and rhamnogalacturonan I (RG I). Methyl groups and acetyl groups are symbolized by “Me” and “Ac” respectively (red).

2.3 Lignin

Lignin is the third major component of the plant cell wall. It is an aromatic polymer consisting of a network of disorderly distributed monomers of monolignols, that are polymerized by radical reactions and covalently connected by carbon-carbon and/or ether bonds. It is embedded around hemicelluloses and cellulose to contribute to cell wall

stifness, water-resistance and providing protection against microbial degradation [36]. The three most abundant monolignols present in lignin are *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, but their relative amounts can vary significantly depending on the source. A multitude of species of covalent bonds are present in lignin, but the most important is regarded to be the β-*O*-4 ether bond [37]. In the biorefinery context, lignin is still vastly underutilized as nearly all of it is just burned for energy, which is sadly ironic considering how well regarded it is a source of value-added chemicals [38, 39]. The challenges to its utilization are represented by its complex structure and the difficulties that it poses for its decoupling in the building blocks of interest through green chemistry methods (eg. Enzyme based) [40-42] (REF). Challenges that are emphasized by the difficulty of purifying lignin from the rest of the polymers of the plant cell wall, to which it can be also cross-linked [43, 44].

2.4 Lignin-carbohydrate complexes

Lignin has been reported to be connected to the polysaccharides of the plant cell wall through different types of covalent bonds forming so-called lignin carbohydrate complexes. These LCCs are believed to be crucial in providing stability for the cell wall, providing recalcitrance to degradation which is a fantastic feature in nature, that though poses several challenges from a biorefinery perspective, increasing the difficulty of utilizing the lignin and decreasing the yield of sugars from polysaccharide degradation [45]. Numerous investigations have been conducted to elucidate the structure and nature of LCCs in plant cell walls, and until not so long ago the presence of these bonds have been debated due to the presence of only indirect observations [46, 47]. In recent years though, following improvements in LLC extraction methods and analysis through NMR have allowed for more concrete proof of their existence[21, 48-52]. Three types of LC bonds have been proposed: ester bonds to the glucuronic acid moieties of xylan, ether bonds to arabinose and xylose residues in xylan, and phenyl glycoside bonds to the reducing end of glucomannan [46, 53, 54]. In addition, lignin-ferulate cross-links between xylan and lignin have also been reported [55, 56]. Of the several types of LCCs described the ester bond between glucuronoxylan and lignin is the most relevant for this thesis, as separate works (**Paper I** and **II**) were conducted on glucuronoyl esterases, enzymes which have the proposed biological role of cleaving such linkages [57-59].

If the existence of the LCCs is not part of a debate anymore in the field, the nature of the of the ester bond between glucuronoxylan and lignin is still discussed. In particular, different species of ester bonds have been reported, that differ from one another for the position of the ester substituent relative to the ester bond, that would be α- or γ-linked (Figure 6) [48-52, 60, 61]. The γ-ester has been the predominantly reported species. It has been hypothesized that the observed occurrence of γ-ester LCCs instead of the initially reported benzyl esters, may be explained by a possible migration of the uronosyl group from the α to the γ position once formed during lignification (or even during sample preparation) [62]. Another possible explanation might come from an observation I could appreciate during the work described in **Paper II**, where the model substrate mimicking the γ-ester resulted much more stable than the substrate mimicking the α-ester, showing a much lower rate of autolysis when in the same conditions. It is difficult to infer the stability of these linkages in nature, because in reality they could be stabilized by a much complex

structure. But if the α-ester was less stable also in native LCCs than the γ-ester, that could partially explain why the latter has been the most observed species. Finally, it has been reported of possible direct linkages between lignin and pectin, suggested by the fact that depectination of wood prior LCC isolation did not result in a decrease in the amount of arabinose and galactose [49], notably found in pectin. It was then suggested that LC ester bonds could exist also between lignin and galacturonic acid [63]. A direct evidence of the existence of these linkages has not yet been reported, however there are several indications from the literature suggesting their existence. Furthermore, several CE15 enzymes showed capability of cleaving GalA-esters and some of these enzymes were even identified in pectin-targeting PULs. Altogether, I believe that these are strong indications that support the hypothesis of the existence of LCCs between lignin and pectin, but improvements in the techniques (described in 3.3.2) utilized for analyzing these complex structures are required for a chance to get a better understanding of them.

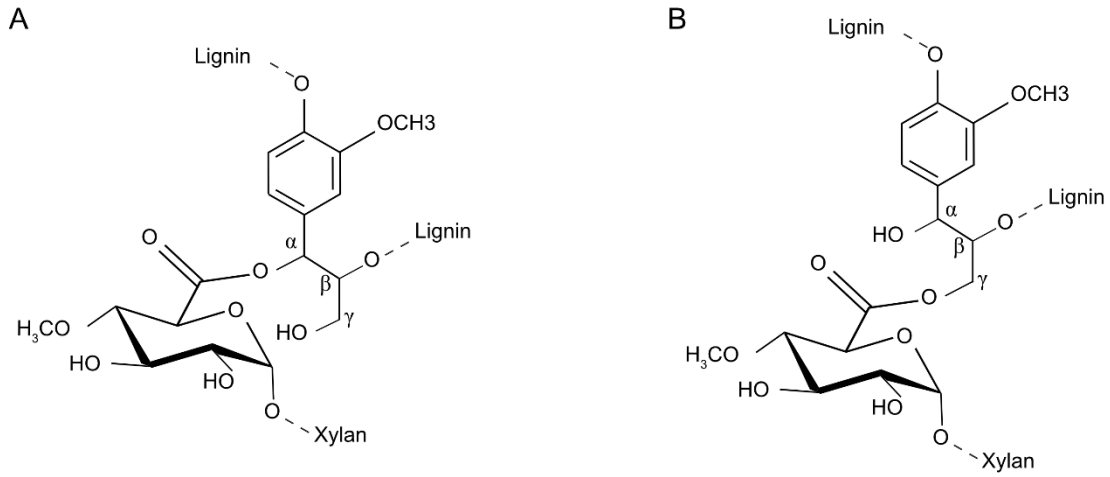

Figure 6. Structure of LC ester bonds between glucuronoxylan and lignin, either α-linked (A) or γ-linked (B) to the lignin benzyl group.

Chapter III: Enzyme discovery and characterization

Discovery and characterization of new enzymes is a long process that starts with defining a strategy to essentially pinpoint a target gene/protein of interest, followed by producing it as a soluble protein that once purified can be "characterized". The most common approach, which was applied in all the studied reported in this thesis, is to produce the proteins heterologously in a production host, but homologous protein production is also possible. Characterizing an enzyme can take several forms, depending on what information and properties one is looking for. However, in this text I will talk about what we call a structure-function characterization, which here refers to biochemical characterization by performing activity measurements and then structural studies with the aim to try to correlate specific structural features to the observed activity. The strategies and methods that can be applied to discover novel enzymes and perform such characterization are discussed in this chapter, but particular attention will be given to those related to the work presented in this thesis.

3.1 Enzyme discovery

Discovering new enzymes with properties of interest for industrial settings is a major driving force behind research into lignocellulose degrading organisms, but it is also a fundamental approach for expanding the existing scientific knowledge on microbiology, plant cell wall structure, enzyme families and activities, microbial evolution, and even carbon cycling in the microbiome. Nowadays the cost and time requirements for DNA sequencing have been greatly reduced, leading to an exponentially increasing availability of known or predicted protein sequences: the latest UniProtKB/TrEMBL protein database statistics [64] available, released earlier this year, counted almost 250 million sequence entries. With such amounts of data, it has become quite straightforward to discover new enzymes with predicted desired characteristics. Databases like this provide lists of computationally annotated predicted proteins, obtained by analyzing new open reading frames (ORFs), translating protein-encoding regions, and comparing the sequences to characterized proteins in order to predict functions. Not all predicted protein domains can be assigned to an already existing category however, and in fact according to the Pfam protein families database [65], over 20% of the domains are described as being of "unknown function", and although more difficult to work with, can represent a source of interesting new activities or functions. Indeed, even proteins with predicted functional domains may carry specificity, efficacy, or stability features that are not fully predictable. Biochemical characterization is therefore vital for full understanding. As such, looking for proteins carrying the function of interest requires researchers to have some screening methods that allow the detection of such function, in the case of an enzyme activity that would be an assay to measure it.

Consulting various databases, such as CAZy (introduced in chapter IV) is an established procedure adopted for enzyme discovery, as one can easily access putative protein-encoding genes, analyze their genomic context and find information on the source organism and on the environment from which it was isolated. Of course, the identified enzyme(s) needs to be produced and characterized experimentally in order to confirm whether it has the predicted function and properties, and this is not necessarily an easy procedure and can often be an unsuccessful endeavor. On top of all the challenges that one could encounter during experimental work, problems could also arise by the fact that putative genes can be misannotated. And this can represent a particular problem for example for genes of eukaryotic origins, where intron/exon prediction tools can be imprecise, and the resulting polypeptide chain may not be produced successfully by the production strain chosen. Such problems can be mitigated for example by cloning from cDNA where the splicing has been fixed by the native host [66].

An example of a useful resource for CAZymes discovery is the polysaccharide utilization locus database (PULDB; http://www.cazy.org/PULDB/ [67]), which lists predicted and literature-derived PULs. Consulting this database or other similar ones like dbCANPUL (https://bcb.unl.edu/dbCAN_PUL/ [68]) can very helpful for gaining further information on novel proteins. For example, many CAZy families can include several different enzyme activities, so-called polyspecific families, but if the gene encoding the predicted enzyme is located in a PUL targeting a specific polysaccharide, this gives an indication on its possible substrate/activity. Such an investigation of PULs was at the basis of our rationale behind the project described in **Paper I**. In fact, the only activity associated so far to enzymes belonging to CE15 is that of glucuronoyl esterases targeting ester bonds in glucuronoxylan. But the enzymes studied in this project were found in putative PULs predicted to target polysaccharides other than xylan, such as pectin.

Another, more structured but less straightforward approach for looking for novel enzyme functions consist of environmental sampling and screening for the function of interest within the microbial community growing on a material of interest. It is possible to enrich such a sample by adding a substrate of interest in abundance, thus giving an advantage to specific microorganisms carrying the function of interest (eg. adding xylan to a sample to increase the chances of identifying xylanases). With some level of luck, and following metagenomic-, transcriptomic- or proteomic analyses, or a combination thereof, can lead to identification and/or isolation of potential strains or enzymes of interest. If it is already known that an organism with a sequenced genome can perform the function of interest, environmental sampling can of course be bypassed, and instead the organism can be directly grown in the conditions that should induce the production of the protein(s) performing the desired function (see again the xylan example above). At this point -omics can be utilized to identify interesting genes, transcripts, proteins, possibly being involved in the function of interest. The large datasets generated through such efforts represent significant lists of potential candidate enzymes, and tracking down the 'correct' protein responsible for an observed activity is not always easy.

Regarding this last approach, I worked on such a project during my second year of the PhD, though it unfortunately had to be abandoned. This digression will be relevant not only to show were some of my efforts were directed during the years, but also to elucidate the potential of this enzyme discovery approach and to highlight the risks associated with it.

Briefly, the thermophilic fungus *Malbranchea cinnamomea* was found capable of growing on galactomannans (backbone of β-1,4 linked mannose with side chains of α-1,6-galactosyl units) as the only carbon source, despite the lack of enzymes from the glycoside hydrolase families (GH) 27 or 36, the main CAZyme families with reported α-galactosidase activity, in its genome. The secretome of the fungus was thus assayed for α-galactosidase activity using a colorimetric assay (described in section 3.3.1), showing positive results. Our hypothesis was that the detected activity could be associated to an enzyme belonging to a new family of GHs non yet reported. The aim of the project was then to isolate, identify and characterize the protein(s) responsible for the α-galactosidase activity. Different fractions of the secretome, collected from the medium with *M. cinnamomea* growing on galactomannan, were obtained using chromatography in an attempt to enrich the protein(s) of interest. The elution fractions were then prepared for proteomic analysis. However, after analyzing the proteomic data, no unique protein was found to be present in the active fractions. This would not represent a dead end in general, given that multiple proteins could be responsible for the observed activity, and four proteins were identified as possible targets, on the basis of the relevance of their predicted function, to be produced and characterized. Based on the time constraints and risks of producing several proteins of unknown function from a non-conventional fungus, the project was put on hold.

3.2 Enzyme purification strategies

In order to perform a thorough study of the protein of interest, it is often necessary to produce and purify it. This is needed for example when one wants to be sure that the detected function is attributable solely to the protein of interest, and no other proteins in the production host. Plus, it may be required for technical purposes when applying specific techniques, and this is the case for example for crystallization purposes. The choice of the purification strategy is determined by the biochemical nature of the protein, on its size, charge, shape, hydrophobicity, etc. For the enzymes produced during the studies described in this thesis, various purification strategies of Fast Protein Liquid Chromatography (FPLC) were utilized, such as Immobilized Metal Ion Affinity Chromatography (IMAC), ion exchange and size exclusion chromatography.

Immobilized Metal Ion Affinity Chromatography

IMAC is a type of affinity chromatography in which metal ions are immobilized in the stationary phase and the recombinant protein in the mobile phase is generally engineered to have a metal ion binding site. In this way, when the target protein is loaded onto the column, it forms high-affinity specific interaction with the matrix and gets separated from the rest of the molecules in the mobile phase. The elution can be nonspecific when a buffer with high salt concentration or with relatively extreme pH values is used, reducing specific interactions; or it can be specific if the buffer in use contains a high concentration of an analogous of the metal ion binding site, which will compete with the specific interactions. This purification technique is widely adopted because of the high selectivity that it provides. In this study the recombinant proteins were engineered to have a hexahistidine tag (His6-tag) and the column used has a resin made of highly cross-linked agarose beads

coupled to a chelating group pre-charged with nickel ions (Ni2+). At a basic pH the histidine residues of the tag are deprotonated and strongly bind to the Ni2+ exposed on the stationary phase, so the recombinant protein remains attached to the resin until imidazole is added to the mobile phase to compete with the histidines and the protein elutes.

Ion exchange chromatography

The ion exchange is another type of affinity chromatography that can be used to separate molecules on the basis of their surface charges. The stationary phase is charged itself, and that allows the formation of electrostatic interaction between the matrix and analytes in the mobile phase. The molecules that interact with the resin are thus adsorbed, while the others elute as the mobile phase runs through the column. There are two different ways in which this technique can be performed: in case of cation exchange, the resin in the column is negatively charged and can bind positively charged analytes; in case of anion exchange, the resin is positively charged and can bind negatively charged molecules. For proteins, the surface charge depends on its isoelectric point (pI) and the pH of the solution. So, once the pI (theoretical or empirical) of a target protein is known, the proper type of matrix can be chosen, as well as the ideal buffer to equilibrate the column with. Regardless of the type of resin used, the proteins can be eluted using a buffer with relatively high concentrations of salt to compete with the electrostatic interaction between the analytes and the matrix. A gradient with increasing concentrations of salt can be used for the elution process, so the weakest interactions will be disrupted first, and then progressively also the stronger ones are affected as the salt concentration increases. This technique is a valid first option for a purification method, for proteins for which IMAC is not available. An example is when trying to isolate a homologous protein with the desired characteristic from the proteome of the organism producing it. This is what I had to do when working on the project (later abandoned) mentioned in the previous section. In that case, ion exchange chromatography was performed on the secretome of M. cinnamomea in an attempt to enrich the sample with the protein of interest. To do so, I collected the eluates in different fractions correlated to their charge and assayed each fraction for α-galactosidase activity.

Size exclusion chromatography

Size exclusion chromatography, also called gel filtration, is a technique that does not rely on any kind of interaction between the matrix and the analytes; the column instead operates as a molecular sieve, separating the analytes in the mobile phase on the basis of their hydrodynamic radius, which can be defined as the radius of an ideal sphere and depends on a protein's molecular weight and shape. The stationary phase is a porous matrix constituted of small porous beads; smaller molecules enter the pores and are slowed down by the thread-like architecture of the beads, while bigger molecules are not able to penetrate the pores and thus migrate faster along the column. When this technique is applied in protein purification, it has to be considered that two molecules with the same molecular weight can migrate along the column with a different speed. A globular protein would have a smaller hydrodynamic radius and thus would have a longer elution time, while a fibrous protein

would have a larger hydrodynamic radius and thus a shorter elution time. Typically, however, gel filtration is used to separate globular proteins based on molecular weight.

Ammonium sulfate precipitation

Ammonium Ammonium sulfate is an inorganic salt with a high solubility that can be utilized as a protein precipitant. This protein purification method thus allows to separate proteins by altering their solubility in solution, which varies according to the ionic strength of the solution – in this case, the ammonium sulfate concentration. At higher concentrations of salt, the protein solubility is decreased, and when the ionic strength is high enough, the proteins will fall out of solution, an effect called "salting out". Proteins that fall out of solution this way are generally not denatured, thus can be recovered by collecting the precipitant and resolubilize it in a solution with lower ionic strength. Different proteins will fall out of solution at different levels of ionic strength, thus this method can be used to partially purify or concentrate proteins of interest. This methodology was utilized during the work described in **Paper I**, for the purification of PiCE15, a protein that proved very promising from an activity point of view, but also very problematic to be purified. In fact, it would apparently bind very tight to almost any resin I tried, from that of prepacked IMAC or ion exchange columns to alternative resins that were tried. The final strategy that I developed for this protein was to do a first round of purification of the E. coli crude extract through ammonium sulfate precipitation, resolubilize the precipitated protein and buffer exchange to remove the salt, and then perform size exclusion chromatography.

3.3 Enzyme assays

The activity of enzymes can be measured in different ways using various methodologies, depending on the type of reaction catalyzed, on the type of available substrates (e.g. simple model substrates or complex biomass) and the available analytical methods to measure directly or indirectly the product formation). In this section I describe the methodologies that were most relevant for the work presented in the thesis.

3.3.1 Monitoring enzyme activity on model substrates

Spectrophotometric assays are commonly used to study enzyme kinetics by measuring the amount of light (in the UV/visible region) absorbed or scattered during a reaction. These assays are suitable when studying small, simple substrates that do not generate high levels of background absorbance and can often be run continuously. The restriction to simpler (often synthetic) substrates is a disadvantage, especially when the biological substrate is supposed to be far more complex, as is the case with lignocellulose. One could argue that drawing too many conclusions from activities measured on model substrates (even if mimicking the naturally occurring structure) could be misleading because the structural differences between native and model substrate could significantly affect the activity of the enzyme assayed. More specifically for this thesis, this is especially valid for the work presented in **Paper I** and **II**, where CE15 enzymes were mostly characterized on model substrates. That said, the measurement of enzymatic parameters using model substrates allows for a more straightforward characterization of the enzymes of interest, providing a relatively rapid way of gaining an indication of what the activity on the native substrate

could be, and enabling the possibility of easily comparing the activities among different studied enzymes.

A continuous spectrophotometric assay now well established for measuring glucuronoyl esterase activity was utilized in **Paper I** and **II** to obtain kinetic parameters for the activities of the studied enzymes on model substrates. This is a coupled assay, meaning that the activity of the enzyme cannot be detected directly and thus the reaction is coupled with another enzyme that is able to generate a measurable and quantifiable product. The activities of these CE15 enzymes were assayed in a coupled fashion together with uronate dehydrogenase (UDH) on commercially available ester substrates of uronic acids (GlcA or GalA) (Figure 7). in the reaction, the GE cleaves the ester linkage generating uronic acid (and an alcohol), which is then oxidated to d-glucarate or d-galactarate in the presence of NAD^+, which is thus reduced to NADH. NADH absorbs UV light at 340 nm, and the increase in absorbance is proportional to the GE activity.

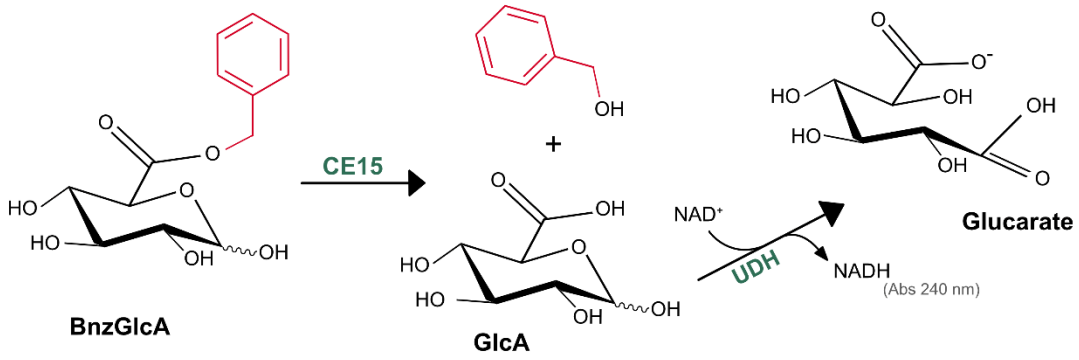

Figure 7. Schematic overview of the reaction in the GE coupled assay using BnzGlcA as substrate. The hydrolysis of the benzyl ester by the CE15 enzymes, allows UDH to form glucarate, with the simultaneous formation of NADH. The increase in absorbance upon NADH formation is detected at 340 nm

During the work described in **Paper I**, after assaying the activity of the CE15 enzymes, we hypothesized that the enzymes could possibly act as pectin methylesterases, and the assay set up to assess that was also a coupled enzyme assay. We utilized polymethylgalacturonan as a substrate, and the second enzyme was a pectate lyase from the polysaccharide lyase family 1 (PL1), which is able to act only on de-methylated polygalacturonan. Specifically, the PL1 enzyme catalyzes the eliminative cleavage of α-1,4-D-galacturonan to give oligosaccharides with 4-deoxy-α-D-galact-4-enuronosyl groups at their non-reducing ends. These unsaturated oligosaccharides that are released absorb strongly at 235 nm. The assay proved to be quite challenging, due to the high level of background noise at that wavelength given by plastic material (UV-transparent plates needed to be used) but also from the

peptide bonds and aromatic amino acids found in the enzymes themselves [65]. My initial tests performed using stopped assays made me think that there was indeed pectin methylesterase (PME) activity in my samples, because I could see an increase in absorbance when adding a CE15 enzyme to the reaction mixture, and additionally the more enzyme I would add, the higher the absorbance increase would be. It was only after trying to follow the activity continuously that I realized the hard truth, that no activity could be detected. The assay was also slightly modified for different substrates – extracted pectins that would represent a more naturalistic substrates, but still no activity was detected when adding any of the CE15 enzymes studied in **Paper I**. On the contrary, our positive control reaction, after trying different pectin methylesterases, was possible to be easily followed in a continuous assay.

Spectrophotometric assays can also be colorimetric, when it is possible to take advantage of substrates that upon modification by an enzyme, turn into colored compounds with peak absorbance at a specific wavelength. In some cases, synthetic substrates exist that allow for direct measurements of the activity of the enzyme of interest, but it is also possible to have coupled assays to indirectly measure that activity, following the colored compound generated by a second enzyme. The latter is the type of assay that was used in the study reported in **Paper III**, in order to characterize the activity of an alcohol oxidase. The assay performed was coupled with a horseradish peroxidase (HRP) and utilized the chromogenic substrate 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS). When the alcohol oxidase oxidizes its substrate it concomitantly produces hydrogen peroxide, which in turn is used by HRP to oxidize ABTS (Figure 8). Oxidized ABTS is green colored and can be measured at 420 nm.

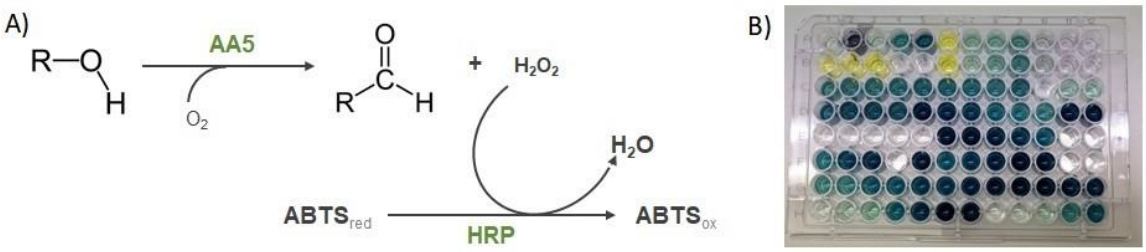

Figure 8. A) Schematic overview of the coupled colorimetric assay utilized for measuring alcohol oxidase activity. B) Photo of a plate at the end of a continuous assay to determine kinetic parameters. showing different intensities (and thus absorbance) of color, proportional to the enzyme activity.

Other very common assays for CAZymes that were not extensively used in the studied described in this thesis are the dinitrosalicylic acid (DNSA) assay and the para-nitrophenol (pNP) assay. The first is a widely adopted technique for the estimation of reducing sugars

and it is based on the capability of 3,5-dinitrosalicylic acid to react with the free carbonyl group in carbohydrates. In an alkaline solution, the DNSA is reduced to 3-amino,5-nitrosalicylic acid while the anomeric functional group is oxidized to a carboxyl group. The reduced product, 3-amino,5-nitrosalicylic acid, is typically characterized by an orange color but most importantly, it strongly absorbs light at a wavelength of 540 nm and that can be measured with a spectrophotometer. During the reaction, one molecule of DNSA reacts with one anomeric functional group, so the measured absorbance value is proportional to the amount of reducing sugar present in the solution. It can be used to study the activity of CAZymes because the hydrolytic cleavage of a polysaccharide will generate more reducing ends, which can be detected with this method. It is a convenient and relatively inexpensive method, but it has a relatively low specificity, and blanks must be run diligently for the colorimetric results to be interpreted correctly and accurately. The sensitivity is not very high as well, and the range in which the absorbance increases linearly and it's proportional to the amount of reducing sugar in solution is relatively limited.

The pNP assay on the other hand, requires the substrate (of interest for detecting the desired activity) to be linked to the chromogenic group 4-nitrophenol. Such model substrates are luckily commercially available for many CAZymes and are widely adopted. The enzyme of interest generally attacks the substrate and releases the 4-nitrophenol group, which from slightly alkaline pH values, is of yellow color and absorbs at 405-410 nm. Thus, the intensity can be detected and is proportional to the enzyme activity. This assay was used to screen for acetylxylan esterase activity in the study reported in **Paper III**, and to screen for α-galactosidase activity for the unfinished project that was mentioned in section 3.1.

3.3.2 Activity measurements on biomass

To assess the activity of an enzyme towards complex, natural occurring substrates can be much more laborious compared to the assays described in the previous section. In the study reported in **Paper I**, we sought to detect possible activity of CE15 enzymes on pectin rich biomass. Our hypotheses were that either the CE15 enzymes could act as pectin methylesterases, or that they could possibly target putative ester linkages between pectin and lignin. But directly measurement of such activities in a complex heterogeneous substrate is not at all straightforward and, in our case, would have required extensive external collaboration at a cost in time and effort not compatible within the available timeframe. We decided to try to assess these activities indirectly, by measuring the sugar release of a commercial enzyme cocktail (Ultraflo) supplemented by our enzymes. The rationale would be, for the putative pectin-LCC-cleaving activity, that adding a CE15 enzyme to the reaction mixture would result in a boost in released sugar by the commercial cocktail, as shown in previous studies [69]. A putative pectin methylesterase activity should not directly boost the cocktail (which do not include pectinase activity according to the manufacturer) so we added the previously mentioned PL1 to the reaction mixture. The PL1 activity would then be boosted by a possible CE15 enzyme-mediated demethylation of pectins, and the activity of the PL1 would boost the sugar release from the commercial cocktail by making the biomass more accessible to its enzymes and help degrade pectin polymers into mono-/oligosaccharides.

The monosaccharides in the aqueous fraction released from complex biomass by enzymes, such as an enzyme cocktail, can be quantified by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), also referred to as ion chromatography (IC). This is a highly sensitive method performed under strongly alkaline conditions, where carbohydrates are partially deprotonated and thus able to interact with the positively charged column. [70]. Known standards must be included to obtain qualitative and quantitative information about the products formed. The availability of standards is therefore often a limiting factor, although this was not the case for the evaluation we performed in **Paper I**. Another useful method for quantitative analysis is mass spectrometry (MS) that can also be coupled to IC in what is referred to as IC-MS, a setup that unfortunately was not available to me. In an MS, analytes are analyzed as ions, which are then injected into a chamber where they are subjected to an electric and/or magnetic field that accelerates and deflects them on the basis of their mass-to-charge ratio (m/z). The detection of the analytes, based on m/z, can then be visualized as a mass spectrum, with signal intensities proportional to the number of times a specific m/z has been detected. The advantages of using IC-MS are higher sensitivity and specificity compared to that of other chromatographic detectors [71], and it also enables to obtain information on enzymatic cleavage patterns in terms of fragment size, linkages and branching.

Another methodology that could be relevant here is nuclear magnetic resonance (NMR) spectroscopy, which allows to visualize chemical structures by observing local magnetic fields around atomic nuclei with an odd atomic number. The spins of these nuclei are normally randomly oriented, but in a magnetic field they align to the field. When subjected to a perturbance of the correct frequency, the nuclei can "spin-flip" aligning against the field, and they then produce an electromagnetic signal upon relaxation. The magnetic properties of a nucleus change with its chemical environment. Something that is defined as "chemical shift" (ratio of the detected frequency to the reference frequency for that nucleus) and can be utilized to reconstruct the chemical composition around the atom. Different NMR techniques, such as 1D ^{31}P NMR and 2D ^{1}H and ^{13}C NMR, have been utilized to detect LCCs and is what makes them relevant for this thesis. To study distribution of functional groups in lignin and/or LCCs samples, ^{31}P NMR requires the derivatization of hydroxyls groups with a phosphitylation reagent [72]. This way it is possible to determine hydroxyl groups attached to aliphatic, phenolic and carboxylic moieties, the latter being specifically important to detect the formation of new free GlcA from the hydrolysis of LC bonds by CE15 enzymes. With this methodology however, it is not possible to gain information on the type of ester bond (α- or γ-ester), or about the structure of the ester substituents.

2D NMR observes the chemical shifts of different types of nuclei directly bound together, which translate in better resolution and enables to gain structural information on ester LCCs that could not be obtained with ^{31}P NMR. It is worth of notice that with this technique, the detected γ-ester signal appears to be overlapped with that of other structures in the analyzed samples, something that might be correlated with the challenges in detecting signal of the α-ester, which are seldom reported [48-52, 61] During the work described in **Paper I**, we received samples of extracted lignin, which contained LCCs according to 2D NMR, from external collaborator. We sought to treat the samples with our CE15 enzymes and then

characterize the samples again through 2D NMR to assess changes in the structure imputable to the enzymatic treatment. However, several issues encountered, specifically with solubility, made not possible to use any of the samples that were prepared, and due to time constraints this collaboration did not pursue further.

3.4 Protein structure determination

Determining a protein structure can give access to very valuable information on an enzyme and how it functions, including configuration of the active site and the catalytic residues, or the possible ligand/substrate binding sites. The structure of an enzyme can also be obtained in complex with their substrate or another ligand, and in this case, it is possible to observe how the interaction between the two occurs, what residues are involved and possibly even identify the determinants of the substrate specificity of the enzyme. Obtaining the structure of the enzyme in complex can be challenging, and often one can find only the product of the reaction bound to the enzymes instead, that is the case for structures we solved and are presented in **Paper I**. There are several methodologies to obtain protein structures, but this section will focus only on those utilized in the work described in this thesis.

3.4.1 X-ray protein crystallography

X-ray protein crystallography is perhaps still the most established method utilized for obtaining tridimensional structures of protein. At the time of writing of this thesis, in the Protein Data Bank (PDB, http://www.wwpdb.org/ [73]) the number of structures obtained from X-ray crystallography experiments are almost 177,000. Curiously, after contributing to an initial exponential growth of the database, nowadays the number of new X-ray crystallography determined structures being published seem to have reached a plateau (PDB). As the name says, this method requires obtaining a crystal of the protein of interest. In order to crystallize a protein, a first strong requirement is for it to be as pure as possible, because to have a high resolution of the final structure, one would want to obtain crystals where only the protein of interested is aligned with the same orientation in the crystal lattice, and possibly to avoid contaminating the sample with the wrong protein. It is generally also good to use the pure protein at high concentrations, which can strongly benefit the crystallization process. The procedure requires mixing of the protein of interest with a series of precipitant solutions. Several different crystallization conditions are usually set up, with varied concentration of precipitant and protein; the aim is to slowly induce the protein to fall out of solution in an ordered crystal form instead of disordered precipitation. Obtaining a crystal is only the first stage of this method and it can require a significant amount of time, but it extremely variable from protein to protein, and can span from seconds to even years. As an example, for the two structures I solved during my PhD studies, *Pv*CE15 (described in **Paper I**) had already formed usable crystals after a few days, while for *Mc*AA9F (**Paper IV**) crystal formation took more than a year. By usable I mean that the crystals had a regular shape that would suggest the protein molecules are orderly disposed in the crystal lattice. There are several shapes of crystals that have been reported and that are associated with a certain quality, though these will not be discussed here. Crystals that look promising can be collected and brought to a synchrotron facility for

analysis, where they are subjected to x-ray beams. These then hopefully diffract, and the resulting diffraction patterns can be collected.

For transportation, the crystal itself must be frozen in liquid nitrogen, usually in some sort of cryoprotectant solution to prevent damage. Despite the "quality assurance" check the researcher can do, one never knows if a crystal would diffract and how good it would diffract until they reach the beamline and "shoot" the crystal with x-rays. Even if the crystal behaves excellently and provides a beautiful diffraction pattern, this alone unfortunately does not provide researchers with all the information needed to obtain a protein structure [74]. The physical reasons behind this will not be discussed here as we would go too far off the topic discussed in this thesis (an excellent source for detailed information is the textbook *Biomolecular Crystallography* by Bernhard Rupp), but I will just mention that the diffracted x-rays give information about their amplitudes when collected, but not their phases, and both are required to reconstruct a protein structure from the collected diffraction pattern. Generally, the most straightforward way to obtain the missing phases, is to use the known phase information from a previously discovered protein structure. Typically, one proceeds by trying to use the phase information of a group of candidates, which are protein with high sequence similarity with the protein of interest which also have a tridimensional structure available. If this method is not successful, the phase information must be obtained experimentally, generally by incorporating heavy atoms into the crystal, by soaking or by direct incorporation into the protein during microbial growth when it is produced. Another option, only recently made available by the progress in the field, consists of utilizing structure models of the protein of interest predicted by algorithms like AlphaFold2 (described in 3.4.2). Once all these information are collected, it's just a matter computational power, *luck* and effort to solve the protein structure from the available dataset(s).

3.4.2 Protein structure prediction

Recently, a number of different powerful protein prediction tools have become available, that allow the prediction of protein structures directly from sequence. Before diving into the most advanced platforms that we are used to normally utilize nowadays, it is also worth mentioning that other tools have been around for a few years although possibly were not as extensively and easily adopted by all researchers. SWISS-MODEL was the first fully automated server for prediction of tridimensional protein structure through homology modelling [75]. Upon uploading of the amino acid sequence of interest, the software would perform sequence alignments to proteins of known structure and attempt to generate a structure for the uploaded query, based on close homologs. The software would also optimize the structure to minimize energy levels and resolve any conflict, reporting statistics describing the estimated accuracy. Another tool based on homology modelling that has seen extensive use is Phyre2 [76]. Its functioning is similar to that just described above for SWISS-MODEL, where a user inputs an amino acid sequence, and the server searches for homologs and attempts to build a model. And additional feature PHyre2 has is to perform *ab initio* modelling for regions of the protein of interest for which a clear template is not available.

Transform-restrained Rosetta, is a tool that differs from the other two just described because it performs *de novo* protein structure prediction [77]. With the input of the amino acid sequence of the protein of interest, a deep neural network predicts the inter-residue

geometries, which are then used as restraints to guide the structure prediction proceeding by assembling regions of short, modelled peptides, undergoing cycles of refinements based on energy minimization. The software still takes advantage of homology modeling by using automatically homologous templates, when available, as additional inputs.

ESMFold 2 is a protein structure prediction tool recently made available by Meta. It is language model-based, trained on 15 billion parameters of protein structures, and requires only a single input sequence, without needing a multiple sequence alignment. The strengths of this tool, according to the authors is capability to predict accurate structures for proteins with limited sequence homologs, and being up to an order of magnitude faster than the competitors. In fact, it has proven to be a very computationally efficient tool, and has been used to predict the structure of all proteins from an extensive meta-genomics database utilizing a relatively limited amount of resources (in time and computation power) [78]. At CASP15, this method however performed significantly worse than AlphaFold2 in terms of accuracy.

Perhaps the most successful prediction algorithm is the aforementioned AlphaFold2, a neural network-based model from DeepMind (Alphabet) trained on all protein structures deposited in the PDB. It proceeds by building multiple sequence alignments doing intensive searches of protein databases using homology detection methods, a process that is very computation intensive. Since its launch, the database of protein structures predicted with this tool has seen exponential growth. In 2021 the achievement of 350,000 protein structure predictions was being celebrated [79] and at the time of writing this thesis, the number of predicted structures is over 200 million.

It has been impressive to directly observe the fast growth of these protein structure prediction methods, especially of AlphaFold that I ended up using the most. At the beginning of my PhD studies, I only used tools like Phyre2, which was useful to get a first prediction of a putative enzyme structure, that could give useful info on which constructs to design for cloning. However, for particular domains (e.g. of unknown function) with no homologous sequence available, it was little informative. I worked on a project that was then abandoned, which goal was to characterize a putative large protein identified in the Bacteroidetes *Cytophaga hutchinsonii*, predicted to have a CE15 domain, followed by a domain of unknown function. The unknown domain had a size comparable with that of an enzyme and thus we thought it could show catalytic activity. After several attempts I was able to clone the domain of unknown function, but was never able to produce it solubly. Later, AlphaFold predicted that the domain was in fact a stretch of immunoglobulin repeats and not a discreet enzyme domain. This is only an example of how I feel tools like AlphaFold have revolutionized protein (and enzyme) research. Since summer 2022 it is also possible to download directly from UniProt the AlphaFold2 predicted structure of most protein sequences, greatly facilitating the construct design and cloning process.

Of course, AlphaFold does not represent a complete solution to the protein folding problem, as it is demonstrated by the high amount of far-from-perfect predictions, but it does represent "a giant leap" in protein prediction, with major implications also for the field of experimental protein determination. Considering all the efforts required for X-ray protein crystallography experiments, and the fact that not all end successfully, it is certainly appealing to rely on predicted models instead of pursuing experimental protein structure determination. One example is given by the cover picture of this thesis, where the experimentally determined protein structure of *Pv*CE15 (**Paper I**) and the structure predicted by AlphaFold are compared. In 2019-2020 significant efforts were necessary to experimentally determine the protein structure of *Pv*CE15, and a couple of years later

AlphaFold predicted a model that looked almost exactly the same. The major differences can be appreciated in a loop region that goes from Pro156 to Ser172, however this is a likely a loop with high mobility, and the even in the determined structure, the quality of the map in that area is not great, and for one of the two molecules in the asymmetric unit, we were not able to model a portion of 10 residues of this loop. Seeing the accuracy of this model was impressive and to be frank it felt like it was diminishing the impact and value of our solved structure. However, our efforts were repaid when soaking of the *Pv*CE15 crystals with ligands such as GlcA and GalA was performed. So personally, I believe that the importance of X-ray crystallography experiments has been partially reduced, at least for the determination of apo-enzyme structures. I say partially because there are protein/sequences for which structure prediction remains problematic. As for now though, structure determination through X-ray protein crystallography remains the main way to obtain the structure of an enzyme in complex with a ligand. It is true that the utilization of tools to computationally simulate the binding of ligands to an enzyme is spreading and the results are progressively getting more compelling. It would be very interesting to follow the developments of this field and see how structure prediction and simulations would integrate experimental methodologies.

Chapter IV: Carbohydrate-active enzymes

The wide array of enzymes active on the polymers found in plant biomass are part of the carbohydrate-active enzymes (CAZymes). CAZymes are classified in the carbohydrate-active enzymes database (CAZy – www.cazy.org [80]) into five different main classes: glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyltransferases (GTs) and auxiliary activities (AAs). In each class, the enzymes are grouped into different families on the basis of the amino acid sequence similarity, which also govern their three-dimensional folds. Enzymes belonging to the same family can have different activities, which are identified by the corresponding enzyme commission (EC) number. The papers included in this thesis focus on CEs and AAs and the families of interest will be discussed in more detail in this chapter, while all the other classes of the CAZy database will not be discussed. A broad overview of the enzymes involved in the degradation of relevant lignocellulose components will also be given first. In CAZy, also carbohydrate-binding modules (CBMs) can be found, which are domains with a non-catalytic carbohydrate-binding activity, generally linked to CAZyme domains.

4.1 Lignocellulose degrading enzymes

4.1.1 Cellulose degrading enzymes

The current model for cellulose degradation is a process involving three different groups of glycoside hydrolases, being endoglucanases (mainly CAZy families GH5, GH9, GH12, and GH45) [81], cellobiohydrolases (GH6 and GH7) [82], and β-1,4-glucosidases (GH1 and GH3) [83], as well as AA enzymes – the lytic polysaccharide monooxygenases (LPMOs) (Figure 9). Briefly, endoglucanases (EG) cleave the internal β-1,4-linked glycosidic bonds in amorphous cellulose, while cellobiohydrolases (CBH) release cellobiose from individual chains from either the reducing or non-reducing ends, and finally the β-1,4-glucosidases (BG) hydrolyze soluble cello-oligosaccharides into glucose from the non-reducing end [84]. More recently, LPMOs have been defined as non-hydrolytic cellulose active proteins [85] which can boost cellulose degradation by oxidatively disrupting crystalline cellulose regions and forming more chain ends, which increases the accessibility and substrate availability for the other enzymes involved in the process [86-88]. LPMOs are object of the study described in **Paper IV**, and later on, a section of this chapter will be dedicated to these enzymes.

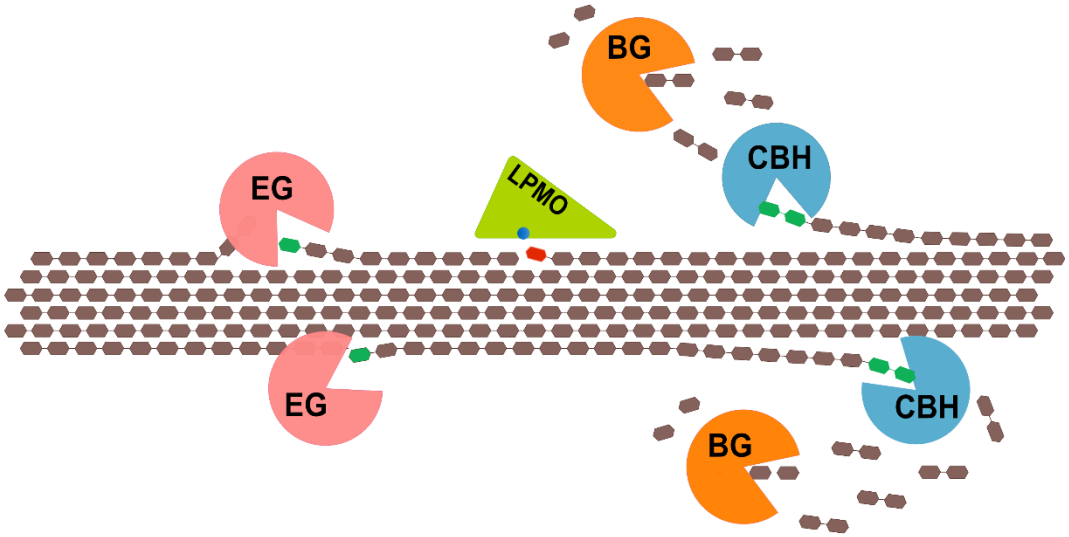

Figure 9. Schematic overview of the model for enzymatic degradation of cellulose. The mode-of-action of LPMOs, endoglucanase, cellobiohydrolase and β-1,4-glucosidase are described in the main text. Hydrolysis reactions are represented by coloring the involved glucose monomers/dimers in green. For the oxidative action of LPMOs, red color is used.

4.1.2 Hemicellulose (xylan) degrading enzymes

Hemicelluloses highly differ in their types of backbone and substitution, so their degradation requires various enzymes with different substrate specificities. In this section, only xylan-degrading enzymes are described, being more relevant for the contents of this thesis. The depolymerization of the xylan backbone is catalyzed by endo-β-1,4-xylanases (mainly found in GH10 and GH11, but also GH5, GH12, GH30, and GH43), and the released xylo-oligosaccharides are further hydrolyzed by β-xylosidases (e.g. GH3, GH43 or GH54) [89-91]. Depending on the type of substituents and decoration linked to the backbone, various additional enzymes can be required, such as α-arabinofuranosidases (e.g. GH43, GH51, GH54), α-glucuronidases (e.g. GH67 and GH115), and acetyl xylan esterases (CE1 and CE3) [22]. LCCs connecting hemicelluloses to the lignin polymers also deserve to be considered in association with hemicellulose degradation (or decoupling), where glucuronoyl esterases (CE15) have been proven to play a role [57, 69].

4.1.3 Pectin degrading enzymes

As described in 2.2.3, different types of pectin exist, which differ from one another mostly for the nature of the side chain substituted to the homogalacturonan backbone. This is true except for RGI which backbone can include rhamnose monomers. Simpler substitutions that can be found on homogalacturonan are methylations and acetylation that are removed by pectin methylesterases (CE8) and pectin acetylesterase (CE12 and CE13) [92]. The degradation of the backbone can happen through hydrolysis or by β-elimination. Endo-polygalacturonases (GH28) are responsible for the hydrolytic depolymerization of homogalacturonan, and the released oligos are further degraded by exo-polygalacturonases (GH28) [93]. Hydrolysis of RGI requires endo- and exo-acting rhamnogalacturonan hydrolases (GH28) , α-rhamnosidases (GH78), unsaturated glucuronyl hydrolases (GH88),

and unsaturated rhamnogalacturonan hydrolases (GH105) [93]. Pectin lyase (PL1) and pectate lyase (PL1, PL3, PL9, PL10) are responsible for the eliminative depolymerization of the methylated and unmethylated homogalacturonan, respectively [94]. Rhamnogalacturonan lyases (PL4 and PL11) instead, do so in RGI [95]. More complex pectins like RGI and II require also accessory enzymes to remove the side chains and provide access for the main-chain hydrolysing/pectinolytic enzymes. For example, α-arabinofuranosidases (GH51 and GH54), β-galactosidases (GH2 and GH35), and β-xylosidases (GH3 and GH43), while endoarabinanases (GH43), exoarabinanases (GH93), and β-endogalactanases (GH53) [92].

4.2 CAZymes studied in this thesis

This section will focus on three different families of CAZymes that were the object of the studies described in this thesis.

4.2.1 Carbohydrate esterase (family 15)

Characterized members of CE15 are glucuronoyl esterases which proposed biological role is to cleave ester bonds between glucuronoxylan and lignin in LCCs (Figure 10), and are therefore regarded as promising tools to aid in the extraction of single polymers of hemicellulose, and also in the saccharification of lignocellulose biomass [69, 96]. Several members have been characterized since the discovery of the first GE, but CE15 remains a relatively underexplored family and more knowledge is needed on how these enzymes interact and function with their substrate on a molecular level. The glucuronoyl esterase activity is the only main activity reported thus far in CE15. A feruloyl esterase trace activity for CuGE from *Cerrena unicolor* [97] and a handful acetyl esterase trace activities [98, 99] have been reported. Further exploration of the diversity within the family could potentially reveal new activities so far not reported.

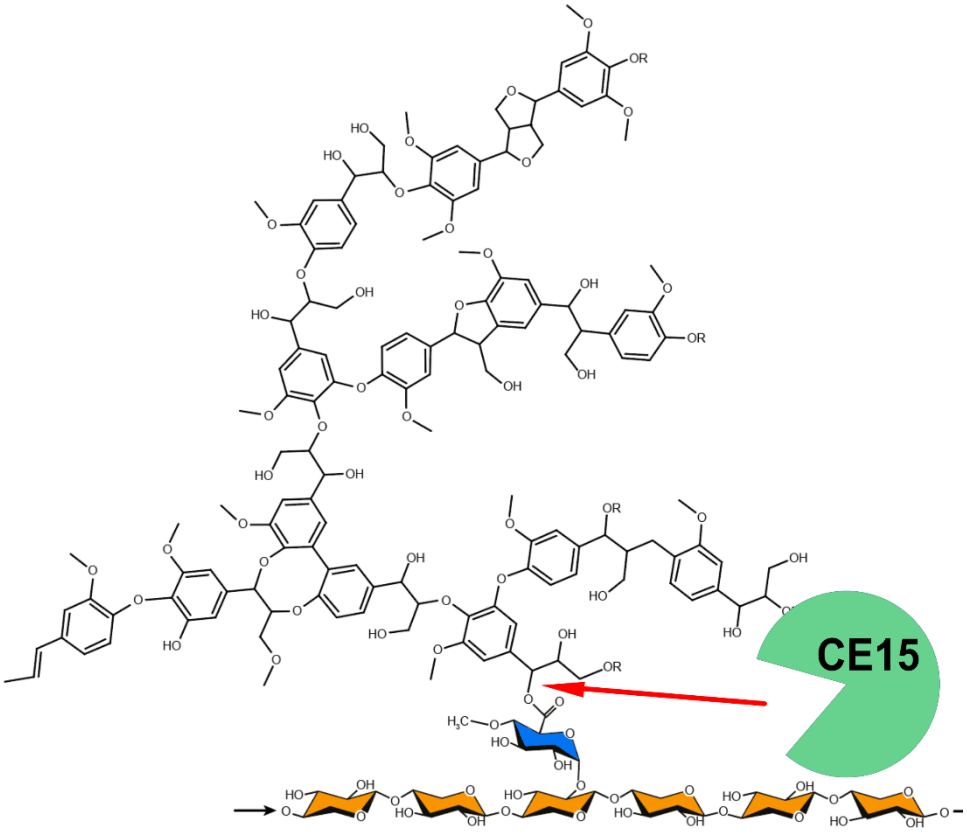

Figure 10. Schematic representation of the action of a CE15 enzyme on the LC bond between glucuronoxylan and lignin. Xylan is colored in orange, GlcA is colored in blue. Figure courtesy of Johan Larbsrink.

The first GE was discovered in 2006 from the fungus *Schizophyllum commune* (*Sc*GE) and characterized on a range of synthetic glucuronic acid esters [100]. Shortly after, another GE of fungal origin was characterized from *Trichoderma reesei* (*Tr*GE or Cip2) [101], and later on this enzyme was the first structurally solved from the family [102]. Homologous sequences have been progressively found in both fungal and bacterial genomes, leading to the formation of the CE15.

At the point this thesis is being written, 53 CE15 enzymes have been characterized (Table 1) [103], although only 22 are reported in CAZy, and a total of 8 three-dimensional structures have been experimentally determined. The count includes members studied in the work presented here (**Paper I** and **II**), which contributed with the biochemical characterization of 8 novel members and with the determination of one protein structure.

Table 1. All the known characterized CE15 enzymes listed in alphabetical order.

Protein name	Microrganism	Characterization	References
Bacteria (22)			
*As*CE15	*Alistipes shahii*	Model compounds	**Paper I**
*Bi*CE15	*Bacteroides intestinalis*	Model compounds	**Paper I**
*Be*CE15A	*Bacteroides eggerthii*	Model compounds	Kmezik, Krska, et al
*Ck*GE15A	*Caldicellulosiruptor kristjansonii*	Benzyl D-glucuronate	Krska, 2020
*Df*CE15A	Dyadobacter fermentans	Model compounds	**Paper II**
*Df*CE15B		Model compounds	
*Df*CE15C		Model compounds	
MZ0003	Uncultured marine bacterium	Model compounds	De Santi, et al., 2016
*Ot*CE15A	*Opitutus terrae* DSM 11246	Model compounds	Bååth, Mazurkewich, et al., 2018
*Ot*CE15B		Model compounds	
*Ot*CE15C		Model compounds	
*Ot*CE15D		Model compounds	
*Pd*CE15	*Phocaeicola dorei*	Model compounds	**Paper I**
*Pi*CE15A	*Parapedobacter indicus*	Model compounds	**Paper I**
*Pv*CE15	*Phocaeicola vulgatus*	Model compounds	**Paper I**
*Sl*CE15A	*Spirosoma linguale* DSM 74	Model compounds, activity on biomass	Bååth, Mazurkewich, et al., 2018
*Sl*CE15B		Model compounds	
*Sl*CE15C		Model compounds	
*Su*CE15A	*Solibacter usitatus* DSM 15142	Model compounds, activity on biomass	Bååth, Mazurkewich, et al., 2018
*Su*CE15B		Model compounds	
*Su*CE15C		Model compounds, activity on biomass	
TtCE15A	*Teredinibacter turnerae* DSM 11142	Model compounds	Bååth, Mazurkewich, et al., 2019

Eukarya (31)			
*Aa*GE1	*Acremonium alcalophilum*	Model compounds	Bååth et al., 2016 Hüttner et al., 2017
*Af*GE	*Aspergillus fumigatus*	Model compounds	Huynh et al., 2018
*Afu*GE	*Armillaria fuscipes*	Activity on biomass	Mosbech et al, 2019
*Ai*GE1	*Ascobolus immersus* RN 42	Model compounds	Dilokpimol et al., 2018

*Am*GE1	*Apiospora montagnei* NRRL 25634	Model compounds	Dilokpimol et al., 2018
*Bd*GE1	*Botryosphaeria dothidea*	Model compounds	Dilokpimol et al., 2018
*Cc*GE1	*Coprinopsis cinerea*	Model compounds	Dilokpimol et al., 2018
CesA	*Ruminococcus flavefaciens*	Polymeric substrate	Biely et al., 2015
*Cs*GE	*Ceriporiopsis subvermispora*	Model compounds	Lin et al., 2018
*Cu*GE	*Cerrena unicolor*	Activity on biomass	D'Errico et al., 2015 D'Errico et al., 2016 Mosbech et al., 2018
*Ds*GE1	*Dichomitus squalens*	Model compounds	Dilokpimol et al., 2018
*Ds*GE2		Model compounds	Dilokpimol et al., 2018
*Hs*GE1	*Hypholoma sublateritium*	Model compounds	Dilokpimol et al., 2018
*Lm*GE1	*Leptosphaeria maculans* JN3	Model compounds	Dilokpimol et al., 2018
*Nc*GE	*Neurospora crassa* FGSC987	Model compounds	Huynh & Arioka, 2016 Huynh et al., 2018
*Pa*GE1	*Podospora anserina* S mat+	Model compounds, activity on LCCs	Katsimpouras et al., 2014 Sunner et al., 2015
*Pa*GE3		Model compounds	
*Pc*GE	*Phanerochaete carnosa*	Model compounds	Tsai et al., 2012 Gandla et al., 2014
*Pe*GE	*Pleurotus eryngii*	Model compounds	Lin et al., 2018
*Pi*GE1	*Serendipita indica* DSM 11827	Model compounds	Dilokpimol et al., 2018
*Pir*GE1	*Piromyces* sp. E2	Model compounds	Dilokpimol et al., 2018
*Pr*GE1	*Penicillium rubens* Wisconsin 54-1255	Model compounds	Dilokpimol et al., 2018
*Ps*GE	*Punctularia strigosozonata*	Activity on biomass	Mosbech et al, 2019
*Sc*GE	*Schizophyllum commune* H4-8	Model compounds Polymeric substrate	Špániková et al, 2006 Špániková et al., 2007 Wong et al., 2012 Ďuranová et al., 2009 Biely et al., 2015 D'Errico et al., 2015 D'Errico et al., 2016
*Sh*GE1	*Stereum hirsutum* FP-91666	Model compounds	Dilokpimol et al., 2018
*Sn*GE1	*Stagonospora nodorum* SN15	Model compounds	Dilokpimol et al., 2018
*St*GE1	*Thermothelomyces thermophilus* ATCC 42464	Model compounds	Vafiadi et al., 2009
*St*GE2	*Thermothelomyces thermophilus* ATCC 42464	Model compounds	Topakas et al., 2010 Charavgi et al., 2013 Katsimpouras et al., 2014

			Sunner et al., 2015 Nylander et al., 2016
*Tr*GE / Cip2	*Trichoderma reesei* QM6A (*Hypocrea jecorina*)	Model compounds, polymeric substrate, activity on biomass	Li et al., 2007 Ďuranová et al., 2009 Pokkuluri et al., 2011 Biely et al., 2015 D'Errico et al., 2016
*Tt*GE	*Thielavia terrestris*	Activity on biomass	Tang et al, 2019 Mosbech et al, 2019
*Wc*GE1	*Wolfiporia cocos*	Model compounds,	Hüttner et al., 2017

4.2.1.1 CE15 phylogeny

All the CE15 sequences now collected in CAZy are from bacteria and fungi, with the exception of a handful identified in archaeal genomes. They are widespread in nature and phylogenetic studies from previous years have showed how fungal sequences would cluster in two major clades [104], while the bacterial targets that constitute the vast majority of the family, would be spread into a larger region, with more diverse clades with as low as 15% sequence identity [69]. At the time, the available sequences were only a fraction (239 to 265) of the one now collected in CAZy, however, the phylogenetic analysis that was done in **Paper II** by aligning 747 entries, is mostly consistent with the previous reports (Figure 11). One major observation is that if in the past it was thought that the diversity of the tree was broadly sampled, now we could see that the majority of the characterized enzymes are actually limited to two main regions, with a large section in the middle with no characterized members. Our work described in **Paper II** though, began to bridge this gap between these two regions. It is noteworthy that despite the fungal sequences representing a minority of the family and being closely clustered in the tree, they have been relatively more studied compared to the bacterial targets.

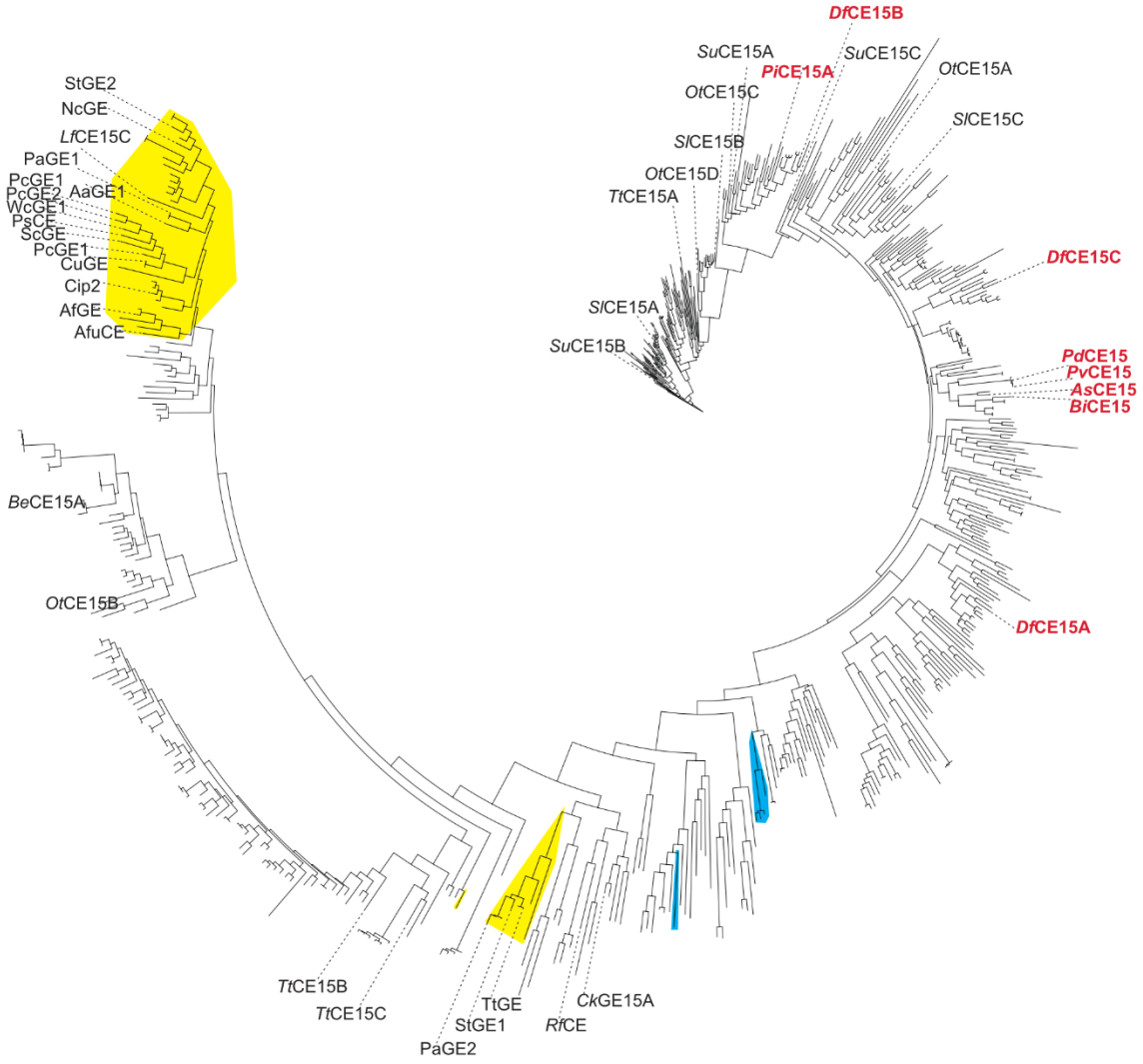

Figure 11. Phylogenetic tree of CE15 built by using all the sequences available on CAZy. Charachterized CE15 enzymes are labelled. Novel CE15 enzymes characterized for the first time in the work described in this thesis are labelled in bold red. The two fungal clades are highlighted in yellow, the two small archaeal clades are highlighted in blue.

It is also noteworthy that the diversity within the family applies non only to enzymes from different sources. Several microorganisms encode multiple putative CE15 enzymes (CAZy) that can be very phylogenetically distant from each other, which may suggest they serve potential different biological roles. These enzymes could be targeting LCCs from different sources with specific structures of lignocellulose. One possible explanation for this multiplicity, as we have shown in **Paper II**, is that different enzymes within the same organisms could have different specificity towards the type of ester linkage that connects hemicellulose and lignin, and possibly their production is regulated on the basis of the relative abundance of the esters in the substrate the microorganism is trying to degrade.

4.2.1.2 Substrate specificity of CE15 enzymes

One struggle of characterizing CE15 enzymes is the lack of accessible natural substrate as well as established analytical methods for assaying their proposed biological function. However chromatographic [104] and spectroscopic [69, 105] assays were developed to allow a more straightforward assessment of the ability of these enzymes to cleave smaller model compounds mimicking the ester LC bond (described in 3.3.1). The continuous coupled assay described in 3.3.1 has been widely used for the characterization of bacterial targets in commercially available esters of glucuronic acid and galacturonic acid. In particular benzyl-, allyl-, and methyl-substituted GlcA were utilized (BnzGlcA, AllylGlcA and MeGlcA; **Figure** 12A-C), as well as galacturonic acid methyl esters (MeGalA, Figure 12D). of these, BnzGlcA has been regarded as the most relevant because with its larger alcohol moiety represents also a better lignin portion of the ester LCCs. In addition to these "canonical" simple model substrates also galacturonic acid benzyl ester (BnzGalA, Figure 12E) and 3-phenylpropyl glucuronate (3ppGlcA, Figure 12F) were utilized in the studies described in this thesis. Specifically, BnzGalA allowed a compelling characterization of novel CE15 enzymes on a substrate mimicking a pectin-lignin ester bond (**Paper I**), whereas 3ppGlcA instead, mimics a γ-ester LC bond, and in combination with BnzGlcA (α-ester) it enabled comparison of substrate specificities of several CE15 enzymes towards the two different type of ester bonds (**Paper II**).

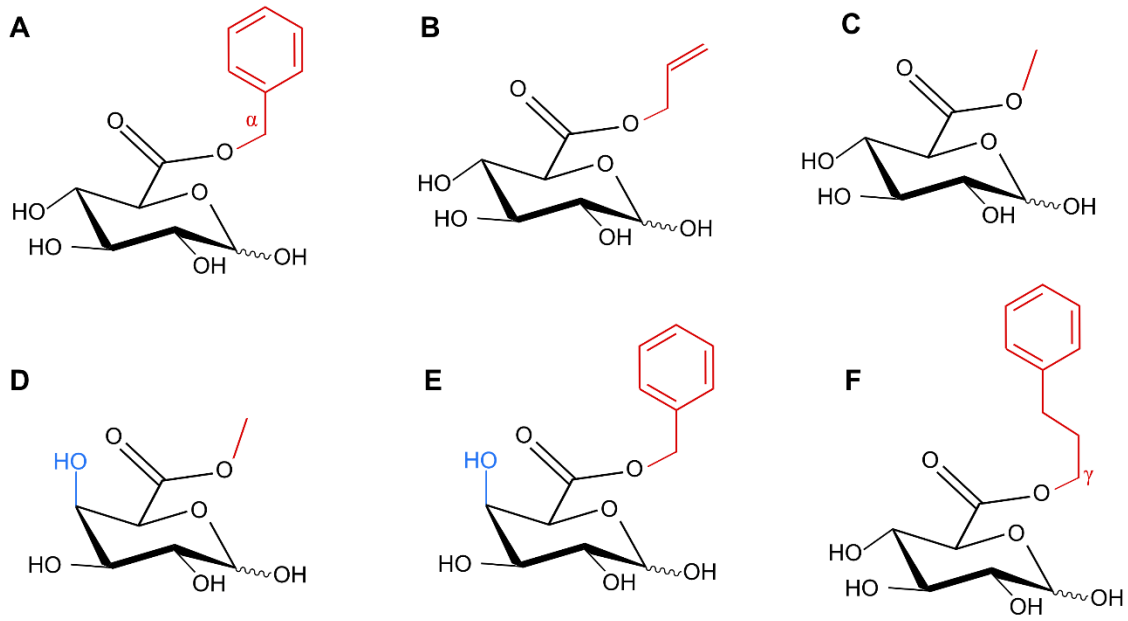

Figure 12. Commercially available model substrates used for assaying the activity of CE15 enzymes: BnzGlcA (A), AllylGlcA (B), MeGlcA (C), MeGalA (D), BnzGalA (E) and 3ppGlcA (F).

The variety of substrates discussed above allowed, throughout several studies, to assess the specificity of many CE15 enzymes also in regards of several stuctural factors. In the early years, assays performed using simple methyl esters of GlcA and its epimers on fungal enzymes [100, 101, 106], suggested that the recognition of the GlcA moiety was crucial for the enzymes given the lack of activity on substrates based on another epimer such as GalA [107, 108]. On the other hand, more recent studies on bacterial targets have shown how certain enzymes do not make any distinguishment for the nature of the sugar moiety [69]. The 4-O-methylation on GlcA has also been reported as an important feature for fungal enzymes, considered a requirement for the recognition of the compound [107, 109-111]. The complexity of the lignin-mimicking moiety was also shown to play an important role for several enzymes, which exhibit increased affinity towards bulkier substrates [69, 110, 112]. The xylan portion connected to the anomeric carbon of GlcA on the contrary, has not been found to be significantly affecting the activity of the CE15 enzymes tested [109, 113]. Lastly, before our work described in **Paper II**, only limited kinetic data was available for γ-linked substrates and it hadn't been possible to conclude weather the species of the ester bond could play a significant role in affecting GE activity [111, 114]. A direct comparison between one α- and one γ-linked substrate was performed only for CuGE, showing a slightly preference for the α-ester [110]. Such a variety of substrates, and the lack of a unified characterization pipeline, has the consequence of making it difficult to do a direct comparison between all the CE15 enzymes characterized thus far. Partially because full kinetics measurements are not available for most of GEs, and even when they are, they do not necessarily refer to the same substrates.

4.2.1.3 Original work – comparison of linkage specificity.

Our work presented in **Paper II** aimed to assess the substrate specificity towards the relative position of the ester substituent of the lignin moiety, in α- or γ-linked substrates. Since, as mentioned in chapter II, more studies have reported the occurrence of both species of ester in nature, we wanted to directly compare several CE15 enzymes the majority of which had been previously characterized, with the addition of few newly discovered targets, to investigate whether the species of ester bond can affect the activity. The kinetic parameters can be found in Table 1 of **Paper II** which also illustrates the preference exhibited by each of the enzymes studied.

To summarize the results, what stands out immediately is that all of the fungal enzymes tested exhibited a preference towards the α-ester, in accordance with what was reported for *Cu*GE [110] (which was also assayed in this work). It is worth to point out that the activities of the fungal enzymes were significantly lower than that of the majority of the bacterial enzymes, which is most likely due to the lack of 4-*O*-methylation on the substrates used. This may raise some doubts regarding the reliability of the assay performed, though, in this work we were interested to the ratio between the activities (expressed as k_{cat}/K_M) determined on the two substrates rather than the absolute values. Given that the substrates have the same sugar moiety, the only difference between the two is the positioning of the bound phenyl group relative to the ester linkage. Thus, I believe that it is safe to conclude that the observed differences in activity are attributable to this sole feature that differentiate the two substrates. That said the differences between the activity reported on the 4-*O*-methylated substrates and the non-methylated ones, are of course important and justify the

conclusions drawn in the previous works, but they are likely within one or two orders of magnitude, as observed for *Cu*GE. Regarding bacterial enzymes, we could see that there was a big variability in substrate preference among the enzymes: six enzymes (of which three were characterized for the first time) preferred the γ-ester, four did not show a clear preference, and five preferred the α-ester. Maybe not so surprising, the enzymes exhibiting similar substrate specificities were typically phylogenetically close (Figure 2 in **Paper II**). In particular, the enzymes with a preference for the for the γ-ester, and the enzymes that seem to not discriminate between the two ester species, tended to group in one region of the tree, while the α-ester-preferring bacterial enzymes were relatively close to the fungal clades. It would be interesting to characterize additional enzymes positioned in the middle, uncharacterized part of the tree, to evaluate their specificities. It is worth noticing also that a certain variability could be observed among enzymes originating from the same microorganism, often showing different specificities like for those from Opitutus terrae: out of four enzymes, two seem to be α-preferring, one γ-preferring, and the remaining one appears to not discriminate between the two substrates.

4.2.1.4 Original work – PUL-derived aldouronoyl esterase

In **Paper I**, we studied five putative CE15 enzymes found in different PULs that appeared to not target the expected xylan. The hypothesis was that such enzymes could host new substrate specificity, perhaps towards GalA-based ester compounds, since some of the targets were found in putative PULs predicted to target pectin (which is rich in GalA). The enzymes were mostly phylogenetically close, but we observed surprising differences in activity. The biochemical characterization, which was performed on the canonical commercial GE substrates plus the new benzyl galacturonate (BnzGalA), revealed that the three enzymes found in pectin targeting PULs showed kinetics on BnzGalA comparable with those on BnzGlcA. PdCE15 and PvCE15 exhibited only a 2-fold decrease in kcat/KM when acting on the GalA-based substrate, while for PiCE15 the value was equal, within error, for the two substrates. The other two CE15 enzymes of the study instead, located in PULs likely not targeting pectin, showed a decrease in activity on BnzGalA up to two orders of magnitude compared to their respective activity on BnzGlcA. These findings gave new flame to the hypothesis that these three CE15 enzyme would target ester bonds in pectin.

We then decided to investigate the ability of these CE15 enzymes to remove methyl ester substituents from homogalacturonan, thus acting as PMEs. The assay coupling PME and PL1 activity (described in section 3.3.1) was then employed on a variety of pectin substrates with different degree of methylation to assess the activity of two of the more efficient enzymes on GalA esters, *Pv*CE15A and *Pi*CE15A. however, no pme activity was observed for any of the two. We then sought to assess whether these enzymes could boost the hydrolysis, by a commercial cocktail, of pectin-rich biomass from different sources. The assay described in 3.3.1 was performed for *Pv*CE15, but no boosting in sugar release was observed, though our positive control with a bona fide PME did improve sugar release. These results indicate that our CE15 enzymes are not PMEs.

4.2.1.5 CE15 structural studies

10 CE15 enzymes have been experimentally solved by x-ray crystallography. Four are of fungal enzymes: *Cu*GE from *Cerrena unicolor* (PDB: 6RTV, &RU1, 6RU2, RV7, 6RV8 and 6RV9) [115], *Lf*CE15C from *Lentithecium fluviatile* (PDB: 8B48) [116], StGE2 from *Thermothelomyces thermophila* (PDB: 4g4g, 4g4i and 4g4j), and *Tr*GE (Cip2) from *T. reesei* (PDB: 3pic) [102]. The other six are of bacterial targets and are MZ0003 cloned from a marine metagenome (PDB: 6ehn) [117], *Ot*CE15A from *Opititus terrae* (PDB: 6grw, 6gs0, 6SYR, 6SYU, 6SYV, 6SZO, 6T0E, 6T0I, and 7B7H) [69, 118], *Su*CE15C from *Solibacter usitatus* (PDB: 6gry and 6gu8) [69], *Tt*CE15A from *Teredinibacter turnerae* (PDB: 6hsw) [119], *Ck*CE15A (PDB: 7NN3) [120]. In **Paper I,** I report the new structure of *Pv*CE15 from *Phocaeicola vulgatus*. All these structurally determined CE15 enzymes conform to the α/β-hydrolase superfamily fold, consisting of a three-layer αβα sandwich, that hosts the active site, comprising the catalytic triad serine-histidine-aspartate/glutamate, in a pocket located above the main parallel β-sheet. All the solved structures confirm that GEs are serine-type hydrolases, with the conserved nucleophilic serine in the consensus sequence: G-X-S-R-X-G-K. While the basic histidine is highly conserved, two different positions have been reported for the acidic residue. Some enzymes, like *Tt*CE15A have it at the canonical α/β-hydrolase position at the end of strand β7 [115, 119], others like *Cu*GE exhibit it at the end of strand β6 (Figure 13). It is also possible that an enzyme has an acidic residue in both positions, like *Ot*CE15A.

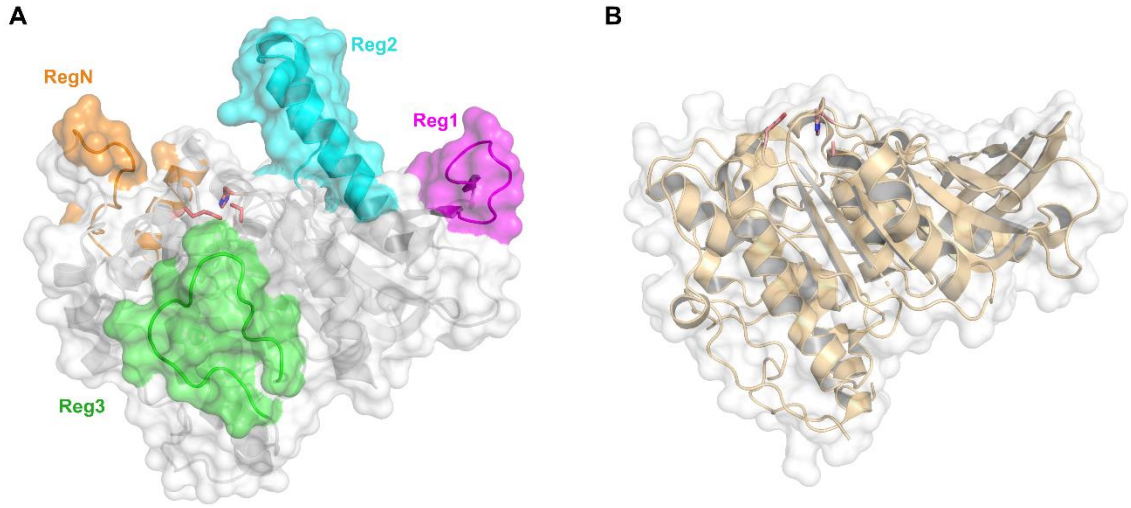

Figure 13. Comparison of the overall fold of bacterial and fungal CE15 enzymes. A) Structure of *Tt*CE15A (PDB ID: 6hsw), the extended regions and the ridges they create are colored in orange (RegN), magenta (Reg1), cyan (Reg2), and green (Reg3). B) Structure of *Cu*GE (PDB ID: 6rv9) showing the relatively alt surface of the fungal CE15 enzymes. In both panels, the side chains of the residues of the catalytic triad are colored in pink.

Main differences between the fungal structures and the bacterial ones can be appreciated starting from the overall fold. The fungal enzymes show an open and relatively flat substrate-binding surface, while solved bacterial structures have up to four inserted regions, namely Reg1, Reg2, Reg3 and RegN (Figure 13), where the latter two can form high ridges that create a deep pocket where the active site is located. This may suggest a difference in the type of natural substrate targeted, with fungal enzymes acting on larger and/or more intact LCC polymers thanks to their surface-exposed active sites, while bacterial enzymes could act on smaller/fragmented LC esters [69, 121]. Reg1 and Reg3 have analogous shapes and high sequence similarity in the bacterial CE15 enzymes structurally determined so far, but Reg2 can vary significantly like for example in *Tt*CE15A where it contains a much longer helix, forming a deeper active site pocket together with RegN, located in opposite position, that cannot be found in other determined structures. The side of Reg2 facing the inner side of the pocket can contain hydrophobic residues that have been proposed to be involved in the interaction with lignin [69]. On the opposite side of the pocket, RegN exhibit several hydrophilic residues that have been suggested to potentially play a role in the interaction with the carbohydrate moiety of the substrate [119].

While there are still no reported structures of Michaelis complexes or of the whole substrate binding the enzyme in the correct position for catalysis, some structures with a CE15 enzyme in complex with a product of their reaction have been obtained (for *St*GE2, *Ot*CE15A, *Cu*GE and now *Pv*CE15 reported in **Paper I**). Important conserved residues identified this way are an arginine next to the catalytic serine stabilizing the so-called "oxyanion hole", and other highly conserved residues supporting the glucuronate binding, such as tryptophan, glutamate, and lysine found interacting respectively with O2, O2 and O3, and O4 hydroxyls of the carbohydrate moiety (Figure 14). In addition, a hydrophobic patch nearby the afore-mentioned lysine has been suggested to be important for the correct positioning of the 4-O-methoxy moiety. A conserved phenylalanine was observed in bacterial structures and proposed to be involved in binding of the lignin moiety [69, 117, 119] and docking simulations with a benzyl ester of 4-Omethyl-glucuronoxylotriose confirmed the possibility of stacking interaction of the aromatic residue with the benzyl moiety of the substrate.

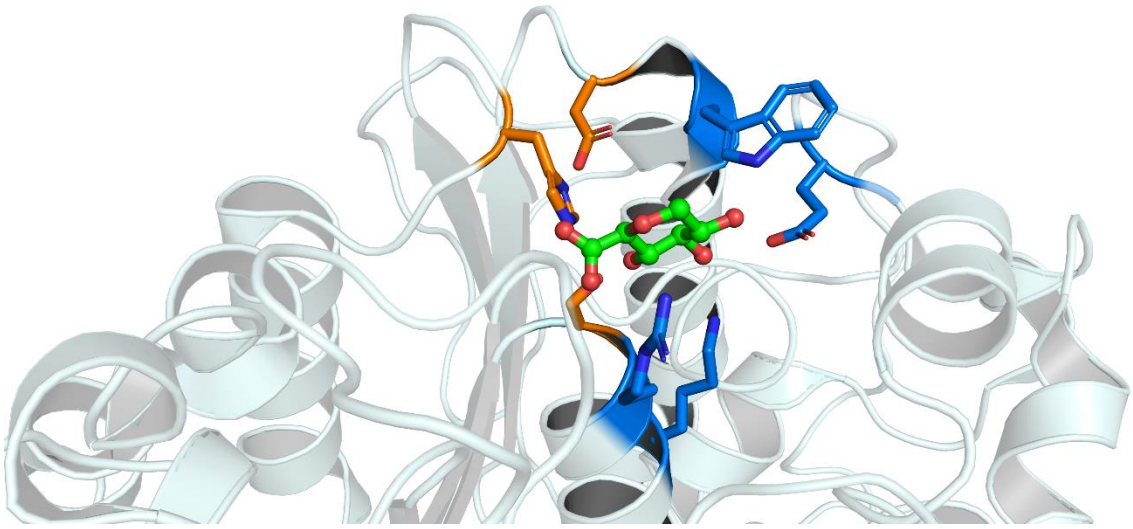

Figure 14. Details of the active site of OtCE15A (PDB ID: 6syr). The residues of the catalytic triad are colored in orange. Other conserved residues important for the binding of GlcA (green) are colored in blue. Trp is hydrogen-bonding to the hydroxyl group in C2. Glu is involved in two hydrogen bonds with the hydroxyls of C2 and C3. Two hydrogen bonds also for Lys, with hydroxyls in C3 and C4. Arg contributes in forming the oxyanion hole.

4.2.1.6 Original work – CE15 structural studies

In **Paper I** we solved three different structures of PvCE15, one of the apo-enzyme, one in complex with GlcA and one complex with GalA. The information we gained from these, confirmed several findings of previous studies relative to the binding of the carbohydrate moiety, but possible determinants of the specificity exhibited by PvCE15 were unfortunately limited. Looking at the GalA-ligand structure (Figure 15A), we could appreciate that GalA is sitting in the active site in a different binding pose than the ”usual” pose of GlcA (Figure 15B) with the pyranose ring flipped resulting in the anomeric hydroxyl pointing into the cleft and positioned where the GlcA O4 is found. This orientation for GalA had already been observed with OtCE15A, which had shown similar activities on MeGlcA and MeGalA, as well as BnzGlcA [69]. It is worth to point out that the OtCE15A structure was obtained using a mutated variant in which the catalytic serine was substituted with an alanine, but the PvCE15 structure with the WT protein gives additional support that there could be a conserved mechanism for these substrates that requires this specific binding pose.

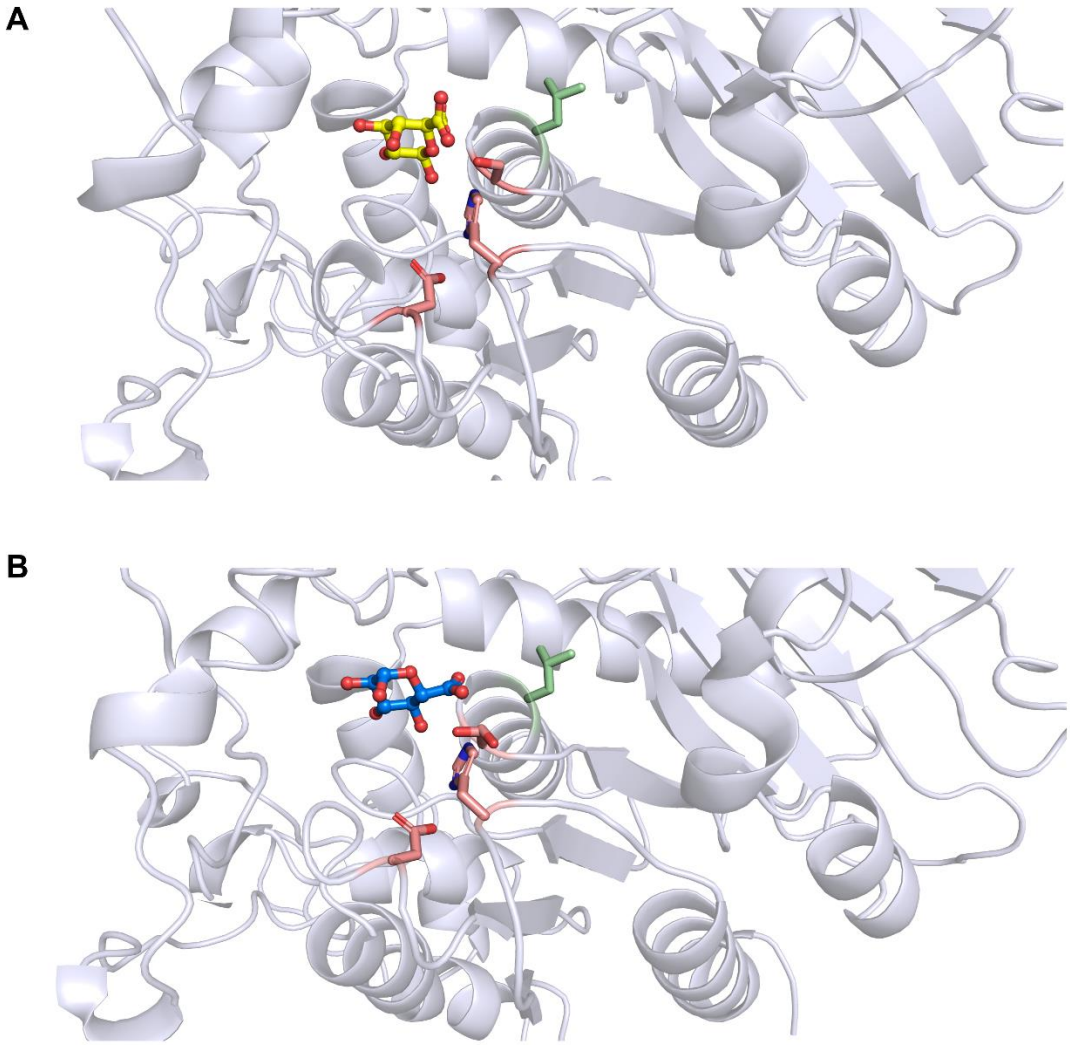

Figure 15. Comparison of the binding pose of GalA and GlcA. A) Structure of *Pv*CE15 in complex with GalA, colored in yellow. B) Structure of *Pv*CE15 in complex with GlcA, colored in blue. In both panels, the side chains of the residues of the catalytic triad are colored in pink, and the leucine conserved in enzymes efficient on GalA esters is colored in green,

The structures here obtained were aligned with those of other CE15 enzymes characterized on MeGalA and/or BnzGalA in this or in previous works. For enzymes for which an experimentally determined structure was not available, we utilized Alphafold2 predicted models in order to expand the array of enzyme to be used for a structural comparison. What emerged is that a leucine residue (Leu 266 in *Pv*CE15 – Figure 15) appears to be conserved in characterized CE15 members with a relatively good activity on BnzGalA and/or MeGalA, while other targets with lower efficiency on GalA-esters present bulkier residues such as histidine, tyrosine or tryptophan in the corresponding position. This leucine residue had been previously postulated to be possibly contributing to the MeGalA specificity after

a mutation study with *Ot*CE15A showed decreased catalytic efficiency towards MeGalA for a variant in which the leucine had been substituted with a tyrosine[69]. It is still unclear though how this leucine residue could be involved in determining the substrate specificity as it is located further away from the carbohydrate moiety especially in the peculiar "flipped" binding pose of GalA and appears to not interact with it. This suggested that although this leucine plays a role has shown by the mutation study, there could be other structural features contributing to the GalA specificity, as suggested by the fact that the afore-mentioned *Ot*CE15A variant had impaired activity on MeGalA but still higher compared to that of CE15 enzymes not carrying this leucine. In fact, there is another feature that appears to be correlated with a relatively high level of activity on MeGalA: which is the presence of a phenylalanine or tyrosine in the corresponding position of Phe 311 in *Pv*CE15 which is associated with a stricter preference towards GlcA-esters, though mutations of this residue were not pursued.

In the study described in **Paper II**, we also used several Alphafold2 predicted models as we sought to do an extensive structural comparison between all the targets we selected to be characterized, to possibly define structural features contributing to the specificity towards the α- or γ-ester. What we found first, was that three bacterial enzymes (*Tt*CE15B, *Ot*CE15B and *Be*CE15A) showed an open fungal-like substrate binding surface, and interestingly also had a similar preference for the α-ester, in contrast with what was observed for other bacterial CE15 enzymes with large inserts. Among the enzymes selected for our study, almost all the enzymes that possess Reg2 either do not seem to differentiate between the two substrates tested, or have a preference for 3ppGlcA (γ-ester), with the exception of three targets that showed a higher activity for the α-ester. While the Reg2 ridge seems to be important for the recognition of bulkier ester substituents, it is clearly not the only determinant of substrate specificity. Out of the CE15 enzymes in this study reported in **Paper II**, four bacterial enzymes and all the fungal enzymes curiously do not have the previously mentioned phenylalanine in Reg2 proposed to be involved in lignin binding, and showed a preference for BnzGlcA. These results suggest that this Phe is important for the specificity towards the γ-ester. We assayed the previously constructed F174A variant of *Tt*CE15A to probe this, and saw a higher activity for the mutant on BnzGlcA than that on 3ppGlcA. The mutation thus inverted the substrate preference of the enzyme, from γ- to α-preferring, suggesting that somehow that residue is playing a stronger role in the recognition of the γ-ester than it is for the α-ester.

Once again, having a structure of the enzyme in a Michaelis complex or at least in complex with the substrate could help in shedding some light on the matter. However, that is a luxury we do not have so to be able to visualize how the whole substrate would sit on the active site, I manually drawn the structure of the "missing" lignin moiety of 3ppGlcA and BnzGlcA substrates in PyMOL, superimposing it to the corresponding position of the glucuronate of the aldouronic acid XUXXR in the structure 6rv9 of *Cu*GE (Figure 4 panel D, E, F of **Paper II**). This drawn substrate XUXXR-3ppGlcA could then be superimposed to the structure of any of the enzymes of the study, it could be seen that the phenylalanine is placed in a parallel position in respect to the benzyl group of the γ-ester substrate, potentially making pi stacking interactions. Conversely, the superimposition of the modelled α-ester shows that the benzyl group in this case would be sitting further away from the phenylalanine, closer to the active site. Now given how these findings were

obtained, even though the drawn orientation of the lignin moiety is one of the possible with the lowest energy level, they should be taken with a grain of salt before drawing strong conclusion on why the phenylalanine is more important in the recognition of the γ-ester than it is for the α-ester. It is in fact important to consider that structure of the enzyme we obtain (both experimentally and computationally) is just a "photography" of a certain state of the enzyme, that in reality is a flexible molecule prone to adjustments especially when binding to a ligand. Thus it's not impossible to imagine that the active site cleft could slightly rearrange so that the phenylalanine could make stacking interaction also with the benzyl group of the α-ester (providing this is indeed its role in the binding of the natural substrate).

4.2.2 Auxiliary activity family 5

Enzymes from Auxiliary Activity Family 5 (AA5) are mononuclear copper-radical oxidases (CROs) that perform the oxidation of their substrates using oxygen as the final electron acceptor and producing hydrogen peroxide. This family of enzymes has gained increasing attention in the recent years because they do not require external co-factors such as flavin adenine nucleotide (FAD) that usually has a role in hydrogen transfer in other oxidative enzymes [122, 123]. This family is further classified into two subfamilies on the basis of amino acid sequence similarity. Subfamily AA5_1 comprises characterized (methyl)-glyoxal oxidases (GlOx, EC1.2.3.15) [124] , active on a variety of simple aldehydes, α-hydroxycarbonyl or dicarbonyl compounds generating the corresponding carboxylic acids [125-128], and galactose oxidases (GalOx, EC1.1.3.9) [129, 130]. Subfamily AA5_2 instead comprehends a larger number of activities: GalOx [131-133], raffinose oxidase (RafOx, EC1.1.3.-) [134], aryl alcohol oxidase (AAO, EC1.1.3.7) [135, 136] and general alcohol oxidase (AlcOx, EC1.1.3.13) [137-142]. The higher number of activities associated with AA5_2 could also be correlated with the fact that this subfamily has been relatively more studied, with 28 members characterized to date, against 12 only for subfamily 1, despite CAZy only reporting 17 members characterized (CAZy) [143]. Thus, it may be fair to suppose that more activities could be possibly comprised in the subfamily 1 but have not been discovered and/or reported yet. It is also worth to point out that several of the characterized AA5 members have been found to be very promiscuous, exhibiting a broad substrate specificity[144], but in several studies the activity of the enzymes was assayed only on a limited number of substrates and the main activity/function was assigned according to the highest efficiency observed within the substrates tested. Rationally, is likely that some activity-based naming assigned to AA5 enzymes may be imprecise.

The first CRO to be characterized was a GalOx from *Fusarium graminearum* [132], which represents also the most extensively studied one, object of many structural, biochemical and mechanistic works [125, 131, 144-148]. *Fgr*GalOx oxidize the primary hydroxyl group on C6 of galactose monosaccharides to the corresponding aldehyde [146, 149], and is also active on galactose-containing oligosaccharides and polysaccharides [145, 150, 151]. The first crystal structure of an AA5 enzyme to be experimentally determined was that of *Fgr*GalOx [152, 153], and it allowed to define the typical overall fold of a β-propeller constituted by seven β-strands (Figure 16). At the central area of the propeller sits the copper ion, coordinated by two tyrosines and two histidines. One of the copper-

coordinating tyrosines is covalently bound to the sulfur atom of a cysteine, forming a dimer through the typical thioether bond, that contributes to the rigidity of the active site. In the active site area, hydrophobic residues can be found that have been proposed to be involved in interacting with substrates [154].

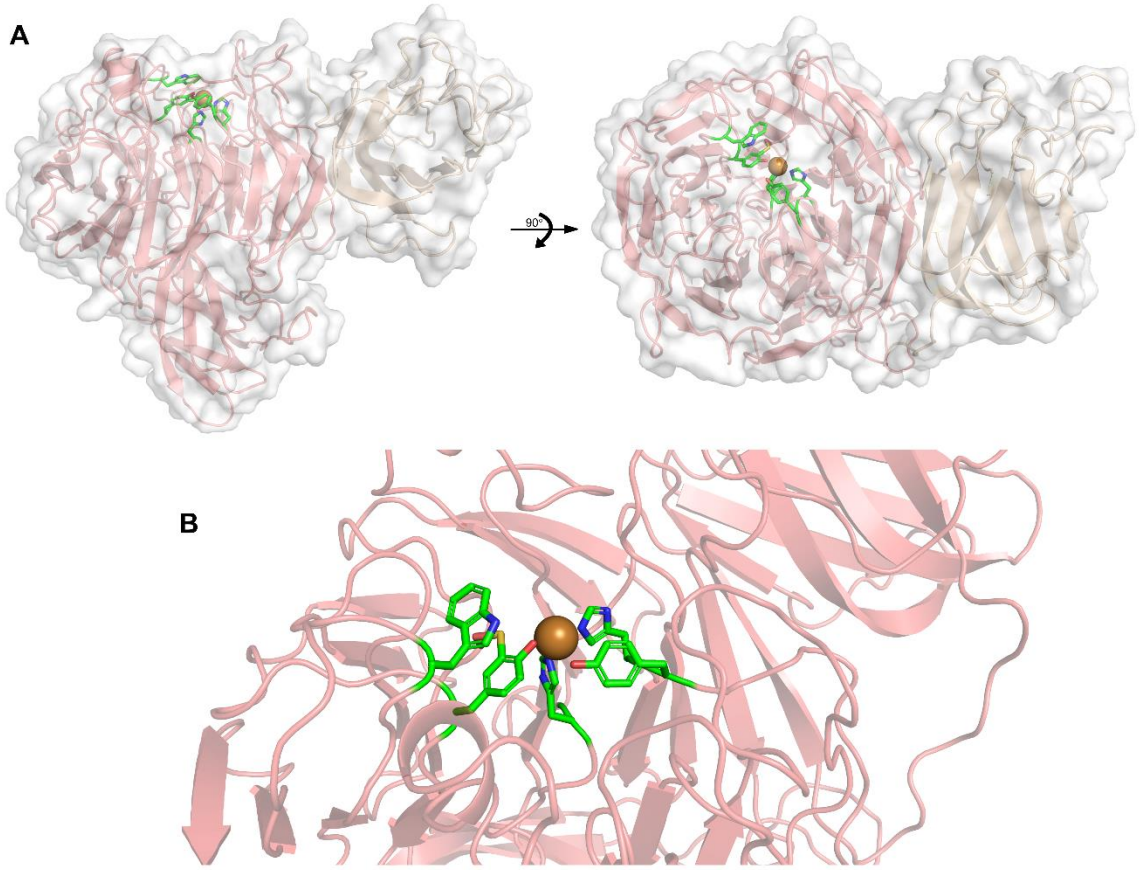

Figure 16. Typical overall fold of AA5 enzymes, here represented by *Fgr*GalOx (A). B) The typical copper ion coordination site. Important residues of the active site are colored in green.

4.2.2.1 Original work

While CAZy has identified putative AA5 members across all kingdoms of life, the vast majority of characterized AA5 members to date are from fungi. Despite a third of the AA5 sequences being from bacteria, only four bacterial enzymes have been studied, one of which is *Bp*AlcOx from *Burkholderia pseudomallei* that we reported in the study described in **Paper III**. One of the reasons fungal AA5_2 enzymes have been extensively studied was the initial interest towards the GalOx activity, regarded as a promising tool to be integrated in existing synthesis processes [145] and also for its potential utilization polysaccharide derivatization/functionalization [151]. At the beginning of this project, our efforts were also directed towards two putative fungal GalOx. We only managed, not without issues, to successfully clone one of the two targets in *Pichia pastoris*, which we were also able to

produce solubly. However, we decided to stop after a small-scale production, once we did not detect any activity on galactose and other simple alcohol. One could argue that we could have pursued this further by producing the enzyme again and in larger quantities, in order to test its activity on wider array of substrates, especially considering the broad substrate specificity within the family that I mentioned earlier in this section. That said, the amount of labor needed to produce the enzyme in higher amounts is not insignificant, and there was also a chance that the protein I was producing, although soluble, was inactive for a variety of reasons. A chance that could be higher than the chance we simply were not screening for activity with the correct substrate. To conclude, even if we had pursued and found activity on some alcohol substrate, chances were that the novelty we had have brought to the field would have been limited, considering the recent attention the fungal AA5 enzymes have received. Instead, we focused our efforts on a bacterial CRO from *B. pseudomallei*, a peculiar multidomain enzyme composed of a putative AA5 domain and a putative carbohydrate esterase family 3 (CE3) domain. Why a CRO, possibly a GalOx, would be fused together with an esterase domain was puzzling, as was the possible substrate targeted by such an enzyme, so we aimed to characterize both domains in order to gain some insights on their functions.

The phylogenetic analysis (Figure 1 in **Paper III**, page 4) showed that the *Bp*AlcOx domain does not fall within any of the two subfamilies, and is located into a bacterial clade, not close to any of the other characterized bacterial CROs which is not unexpected considering only three bacterial members had been characterized before. From the initial assays we could observe that the AA5 domain was not an obvious GalOx, as its activity on monosaccharides appeared very low compared to that on simple primary alcohols. We gradually expanded the array of substrate to be tested and measured kinetic parameters on several types of primary alcohols, from sugars, sugar alcohols, alkyl alcohols, to aryl alcohol, polyols, and furans, with the highest catalytic efficiency being displayed on 1,3-propanediol (a polyol), furans, and longer alkyl alcohols.

The results were quite puzzling. First of all, the activity in monosaccharide turned out to be so low because the enzyme was actually acting only on the open chain form of the sugars which represents only a minor fraction of the compounds in solution (relative abundances at equilibrium can be found at Table S4 in **Paper III**). While this preference for open chain forms, and the lack of activity on aldose and furanose form, seems consistent with the higher activity displayed on longer alkyl alcohols, it clashes with the fact that the enzymes exhibited high activity on furans like furfuryl alcohol and hydroxymethylfurfural. And again, the relatively high efficiency on 1,3-propanediol may be explained by the presence of two primary hydroxyl group in the molecule, but then the activity on longer polyols is three orders of magnitude lower, up to 300-fold lower. This is in contrast with the observed progressively higher activity on alkyl alcohol with longer chain. *Bp*AlcOx proved to have a very broad substrate specificity like other AA5 enzymes before [144] but it is hard to define a preferred class of molecules given the results were not fully consistent, and the "real" substrate for this enzyme remains enigmatic.

Our attempts to crystallize the *Bp*AlcOx, either the full length or the catalytic domain only, were not successful. Thus we decided to take advantage of the powerful protein structure prediction tools available today to visualize and analyze its structure. We utilized three different tools, being AlphaFold2, ESMFold 2, and transform-restrained Rosetta (all

described in 3.3.2), and all gave reasonably high confidence models for both the AA5 and CE3 domains (Figure S6, **Paper III**) and all look very similar to each other, and the overall fold of the AA5 domain is conform with that of observed in experimentally determined AA5 structures, described earlier in this section.

4.2.2.3 Original work – investigation of the CE3 domain

As mentioned above, *Bp*AlcOx domain is fused together to a putative CE3 domain in *B. pseudomallei*. Characterized CE3 enzymes are acetylxylan esterases that hydrolyze the ester bond between acetic acid and hemicellulose in plant biomass. The pNP assay described in 3.3.1 was used to assess the acetylxylan esterase activity of the predicted *Bp*CE3 domain. No activity was observed for the full-length enzyme and for the CE3 domain produced alone. Comparison of the predicted structure to the only two experimentally determined structures available of CE enzymes, Ces3 from *Acetivibrio thermocellus* [155] and of *Tc*AXE from *Talaromyces cellulolyticus* [156], showed that the expected Asp/Glu-His-Ser catalytic triad typical of CE3 enzymes was disrupted. In fact, two inactivating mutations were found: both in the catalytic histidine and serine, which are instead replaced by a tyrosine (Tyr741) and an alanine (Ala567), respectively (Figure 17). These drastic changes in crucial residues of the active site, could explain the lack of acetylxylan esterase activity observed. We then generate a double mutant gene by site-directed mutagenesis, with substitutions (A567S and Y741H) aimed in reverting the expected catalytic triad. However, the produced mutant still did not show any acetylxylan esterase activity. These results suggest that additional molecular determinants involved in the catalysis are absent in the putative *Bp*CE3 domain. From the predicted model structure, it is possible to see that in fact *Bp*CE3 has several tyrosine residues located around the catalytic triad (Figure 17), which are not found in the other two structures available for CE3 enzymes. These tyrosine residues could play a role in hindering the active site and impeding the binding of a substrate. This would be a possible explanation for the lack of any acetylxylan esterase activity in the mutant enzyme carrying the canonical active site. However, all these modifications that can be observed in this putative BpCE3 domain, could suggest that the enzyme may be evolved, losing the capability of catalysis, to play another biological function.

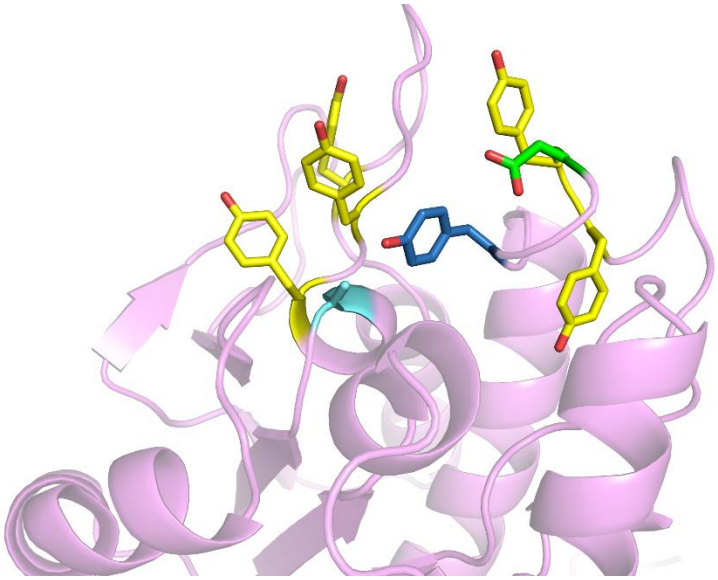

Figure 17. The "active site" area of the putative *Bp*CE3 domain, as predicted by AlphaFold2. The conserved aspartic acid of the active site is colored in green. The non-conserved alanine (instead of serine) and tyrosine (instead of histidine) are colored in cyan and blue respectively. The several histidine residues surrounding the active site area are colored in yellow.

4.2.3 Auxiliary activity family 9

Enzymes of the auxiliary Activity family 9 are LPMOs, metalloenzymes that utilize a copper ion to catalyze the oxidative cleavage of oligosaccharides and polysaccharides [157, 158]. LPMOs are actually classified in CAZy in eight different families: AA9, AA10, AA11, AA13, AA14, AA15, AA16 and AA17, but this section will focus on AA9, as it is the most relevant for this thesis. AA9 LPMOs are known to be able to target the most recalcitrant lignocellulosic polysaccharides, rendering the crystalline substrate more accessible to glycoside hydrolases [159, 160]. Their action has been reported to boost the performances of cellulase in lignocellulose hydrolysis, with increases in saccharification yields of 30-40%, and LPMOs from this family have been added to commercial cocktails for lignocellulose hydrolysis [161-163]. It is known that LPMOs cleave glycosidic bonds by incorporating an oxygen atom in the substrate[164], but the exact catalytic mechanism remains under debate, with two reaction models being proposed . The consensus is that these enzymes use activated oxygen species to perform a redox reaction with the transfer of two electrons via the Cu atom coordinated by the so-called histidine brace in the active site. According to one model, the enzymes use molecular oxygen for the catalysis, and electrons are delivered by an external donor [157, 165], while the other model proposes that hydrogen peroxide is utilized instead [164, 166]. Noteworthy, recent studies point to the latter model only, reporting peroxide to give orders of magnitude faster reaction kinetics [167-169].

LPMOs have also been classified in subgroups on the basis of their regioselectivity. There are C1-oxidizing enzymes that generate aldonic acids as products, and C4-oxidizing enzymes that generate ketoaldoses as products. The last subgroup comprises more promiscuous enzymes that can perform both C1-type and C4-type oxidations, leading to varying mixtures of the two product types (Figure 1 in **Paper IV**, page 2). Initially AA9 LPMOs were shown to only oxidize cellulose, but later reports of oxidative cleavage of soluble cello-oligosaccharides [170], xyloglucan and mixed β-1,3 and β-1,4-linked glucan [171], glucomannan, glucuronoxylan and arabinoxylan [172-178]. *Malbranchea cinnamomea* is a fungus from which several LPMOs have been characterized, including McAA9F which was the object of our structural study described in **Paper IV** [177]. The gene coding for this enzyme was one of nine that were found to be upregulated during growth on wheat bran, xylan and cellulose [179]. Four of the nine LPMOs from M. cinnamomea were successfully produced heterologously in P. pastoris and characterized, and exhibited different substrate specificities. McAA9F was reported to be predominantly a C4-oxidizing enzyme, but was also able to produce a mixture of C1- and C1/C4-oxidized cellooligosaccharides when active on phosphoric acid-swollen cellulose (PASC). Furthermore, it was reported to produce C4-oxidized oligosaccharides from tamarind xyloglucan (TXG) and was also found active on soluble cellooligosaccharides (cellohexaose), generating C4-oxidation products.

At the time of writing this thesis, the CAZy database reports 21 protein structures of AA9 LPMOs, including that of *Mc*AA9F that was described in **Paper IV**. All experimentally determined AA9 structures adopt an immunoglobulin-like β-sandwich fold with the individual β-strands connected by long loops that can have one to several helical insertions of variable length (Figure 18). Seven to ten β-strands constitute the two β-sheets of the core structure. The formation of disulfide bonds is regarded to be essential for AA9 enzymes to obtain their common fold [180, 181], so that all structurally determined AA9 members have a conserved disulfide bond to stabilize the β-sheet core. A second disulfide bond, important for stabilizing the C-terminal area [182], is less conserved and was found only in seven AA9 structures so far, including that of *Mc*AA9F described in **Paper IV**.

A variable number of loops can participate in forming the typical flat substrate-binding surface of LPMOs. The most relevant ones in defining the surface have been named as L2, L3, LS, L8 and LC, which have been associated with substrate recognition [183]. L2 is recognized to be the most diverse region, comprising different numbers and types of amino acid, and different secondary structures [184]. It generally contains multiple surface-located aromatic amino acids, like Tyr 24 in *Ta*AA9A, which are expected to take part in the binding of the polysaccharides[185]. Some AA9 LPMOs can include an L3 loop between β-strands 3 and 4 facing L2, as well as a loop C-terminal (LC) and a loop short (LS) regions, which are more conserved and can also contain solvent-exposed aromatic residues [185].

The substrate-binding surface hosts the so-called histidine brace, comprised by two histidine residues (one is always His1, the other for example is His76 in *Pc*AA9D/GH61D) that coordinates the copper atom, and together with a conserved tyrosine (Tyr160 in *Pc*AA9D) constitute the primary coordination sphere. Notably the N-terminal His1 involved in the Cu coordination, has been shown to be methylated when LPMOs are

obtained through homologous protein production in their fungal hosts [184, 186], which has been proposed to play a role in conveying protection against oxidative damage [187]. This peculiar post-translational modification is not observed when LPMOs are produced in bacteria or heterologous hosts such as *P. pastoris* [171, 176, 188]. Two more residues around the active site are also found in all AA9 structures available, a histidine and a glutamine (His149 and Gln158 *Pc*AA9D), which are part of the secondary copper coordination sphere that forms a conserved hydrogen-bonding network, and have been suggested to be involved in O2 activation [189, 190].

4.2.3.1 Original work

Our work reported in **Paper IV** described the structure of *Mc*AA9F from *M. cinnamomea*, and aimed to identify possible determinants of its peculiar substrate specificity. The overall structure of *Mc*AA9F conforms to the expected AA9 fold and has all the named loops L2, L3, LS, L8 and LC (Figure 18), a characteristic only found in xx other AA9 LPMOs. The histidine brace was conserved and composed by His1, His81 and Tyr169 coordinating the copper ion. A peculiar structural feature that is worth mentioning here is a succinimide positioned in a turn connecting the strands β1 and β2, in place of Asp10. This feature usually forms as a result of the main-chain N atom cyclizing onto the γ-carbon of an asparagine/aspartate side chain. It has been reported to affect protein stability and even functionality, depending on its location in the protein [191]. However, in our structure the succinimide is located in the N-terminal region, in a relatively "unimportant" area of the protein, and one possible effect it could have, is to confer more rigidity to the interaction between the two strands connected by the turn where it is placed. It is also important to point out that its formation is strongly dependent on a variety of factors, such as pH, temperature, buffer composition, and of course protein sequence (which in this case we can say was indeed favorable). So, it is relatively safe to conclude that this succinimide was occurring in the crystallization conditions used to obtain the structure, and might not have been present in the conditions used for assays in the previous study where it was characterized. And even if it was, no apparent implications for the enzyme activity were observed [177].

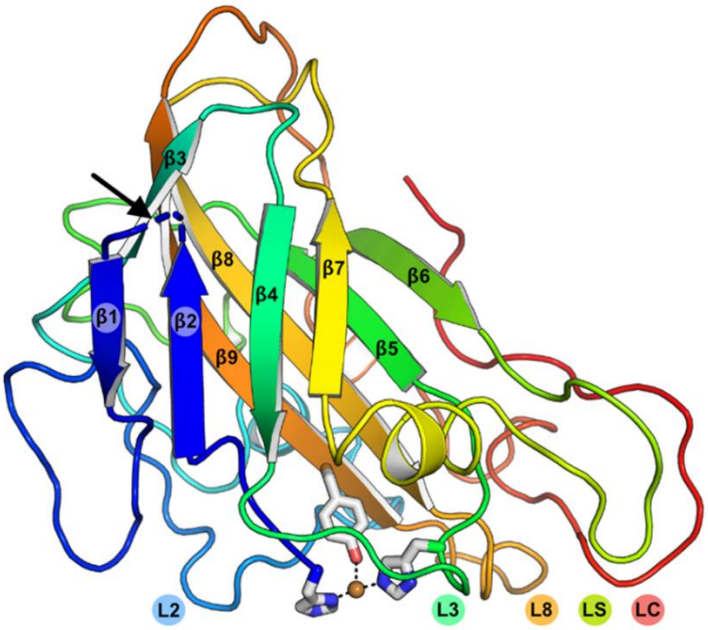

Figure 18. Overall structure of *Mc*AA9F. The colors change progressively from blue to red from the N-terminus to the C-terminus. The loops composing the substrate-binding face are labelled with their assigned name. The residues coordinating the copper ion (brown) are shown in white. Figure taken from **Paper IV** under the CC BY 4.0 license.

We then compared the structure of *Mc*AA9F to those available in the PDB for other AA9 LPMOs, and interesting characteristics could be observed in some of the loops constituting the substrate-binding surface. Other AA9 LPMOs that are C1/C4-oxidizing, exhibit an extended L2 that has been proposed to participate interactions with the substrate [184], and to be involved in the determination of regioselectivity by governing C4-oxidation in some AA9 enzymes [183]. In contrast, *Mc*AA9F, despite being a C1/C4-oxidizing enzyme, presents a shorter L2. Altogether, this suggests that the molecular determinants that define C4-oxidative regioselectivity in the AA9 may not depend upon factors in L2. In fact, it has been already suggested that this regioselectivity could be determined by the positioning of the enzyme on substrates, rather than by phylogenetic relationships [192], and it was proposed that AA9 LPMOs oxidize the linkage that is closer to the copper ion[193]. Personally, I believe it can be concluded that the regioselectivity is a result of the combination of multiple factors. That is because the way a substrate would bind to the enzymes, thus determining which atom would sit closer to the copper ion, will depend by the nature and structure of the substrate, as well as by the structure of the surface of the enzyme, the residues exposed towards the solvent and involved in the interaction with the substrate. The very fact that AA9 enzymes (like *Mc*AA9F) can exhibit different regioselectivity on different substrates, corroborates this hypothesis.

Another peculiarity of *Mc*AA9F, is its very short L3 loop compared with that of other AA9 enzymes active on soluble substrates, such as *Cv*AA9A from *Collariella virescens* (Figure 19). The presence of an extended L3 has in fact been associated with the activity on soluble substrates such as cellooligosaccharides and xyloglucan [194, 195]. Specifically, it has been shown that the extended L3 loop contributes to the formation of a ridge protruding from the binding surface, which could be involved in binding soluble substrates [192]. At the same time, such protrusion could impede the activity on crystalline substrates by affecting the typical flatness of LPMOs, which is proposed to be one of the key factors allowing the activity on flat, crystalline fibers[192]. This is in contrast with what was observed for *Mc*AA9F, active on both crystalline cellulose and on soluble substrates despite its short L3 loop aligned to the surface level. This provides and evidence that the lack of a longer, protruding L3 is not sufficient to justify the lack of activity on soluble substrates, suggesting that more features might be involved in defining the substrate specificity.

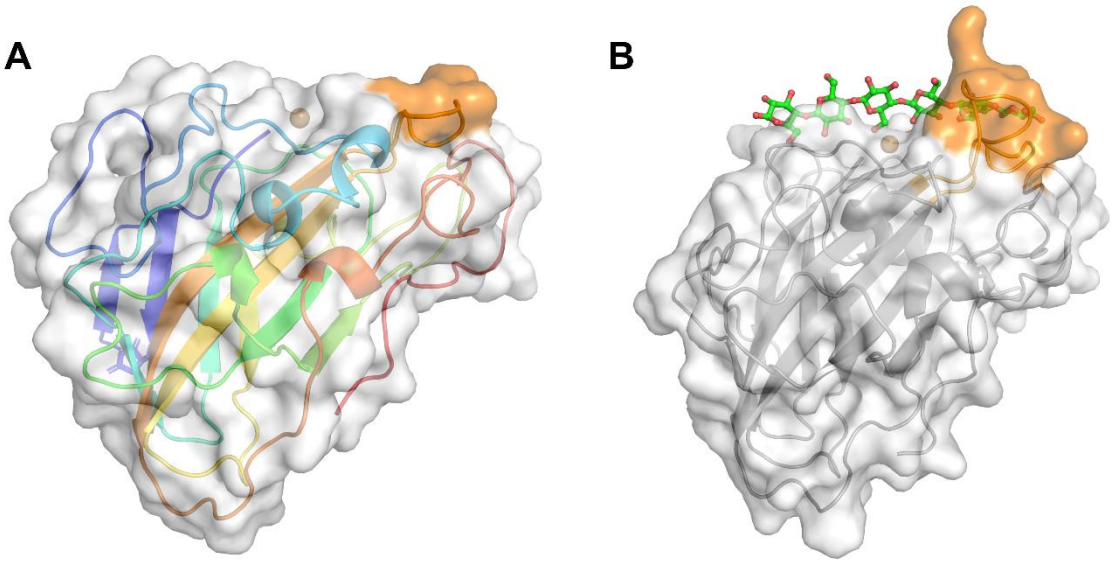

Figure 19. Comparison of the substrate-binding surfaces of *Mc*AA9F (A) and *Cv*AA9A (PDB ID: 6yde). The surface area generated by the L3 loop is colored in orange for both enzymes. The ridge on the surface of *Cv*AA9A is evident compared to the flat surface of *Mc*AA9F. This ridge has been proposed to be involved in interacting with oligosaccharides, such cellohexaose, here colored in green.

The superimposition of the structures of *Mc*AA9F to that of other enzymes in complex with the oligosaccharide cellohexaose, has allowed us to visualize possible points of interaction with such substrate (Figure 3 in **Paper IV**, page 5). The two enzymes' structures were that of *Ls*AA9A from *Lentinus similis* and *Cv*AA9A, which bound to cellohexaose differently, from subsites −4 to +2 the former, and from subsites −3 to +3 the latter, but in both the orientations and positions of the glucose monomers are the same. In *Mc*AA9F, most of the key residues interacting with the oligosaccharide from the +1 to −3 subsites are conserved

(or functionally similar). The main differences can be observed from the +2 subsite towards the oligo's reducing end. The area of the surface of *Mc*AA9F binding this portion of the substrate is defined by L2 and L3, which characteristics were described above. In particular, *Ls*AA9A and *Cv*AA9A in this area have amino acid residues protruding from the surface that are able to bind to the substrate, while in *Mc*AA9F L3 is significantly shorter and there are no residues within a distance suitable for non-covalent interactions. This would suggest that the conserved point of interaction from subsites +1 to −3 are sufficient for determining the specificity towards cellooligosaccharides. Once again in contrast with the previously proposed role of L3 [192].

It is important though to make a few considerations when drawing such conclusions. First of all, what we observed was a superimposition of a substate from other PDB files with the *Mc*AA9F structure we solved. Thus there is no certainty that, in reality, the mode of binding of cellohexaose with *Mc*AA9F would be the same as that observed here. In fact, *Ls*AA9A and *Cv*AA9A bind differently to this substrate, so it cannot be excluded that an additional, still unreported, mode of binding exists. That said, even assuming that the superimposition would be a correct representation of the real positioning of cellohexaose onto the surface of *Mc*AA9F, it is worth to point out that the displacements of flexible regions (especially the loops) and the orientations of the side chains we reported here, are just those that are visible in the "snapshot" obtained from the crystal. Thus it has to be taken into account that some parts of the molecule could rearrange to bind to the substrate. this is for example something that we could have had the chance to see in case we had obtained a structure in complex with cellohexaose.

Chapter V: Conclusions

The work presented in this thesis focused primarily on the discovery and characterization of CAZymes from underexplored families. This chapter evaluates whether the aims stated in chapter 1 section 3 have been reached.

The first aim was to investigate the role of putative CE15 enzymes that were found in PULs predicted to target other polysaccharides, such as pectin, rather than the expected glucuronoxylan, which is the proposed natural target of this family of enzymes. To address this aim, five CE15 enzymes encoded in PULs from different bacterial species were produced and biochemically characterized (**Paper I**). Three enzymes were found to be able to cleave pectin esters-mimicking model substrates (GalA esters) with comparable efficiency to that shown when acting on conventional model substrates mimicking the natural substrate (GlcA esters). We then assayed the activity of these enzymes on methylated pectin substrates to assess whether they could act as pectin methylesterases, but no activity was found. Indicating that this is not the function that the enzymes are performing when acting on pectin. Analogously, when we tried to assess the capacity of these enzymes to boost the hydrolysis, by a commercial cocktail, of pectin-rich biomass from different sources, no consistent boosting was observed. It is clear though that if organisms are expending resources to produce such enzymes when acting on these specific polysaccharides, like pectin, they must have some sort of benefits from it. Furthermore, the relative position of these PUL-derived CE15 enzymes, close to each other, would suggest the existence of a specific clade hindering a novel, unreported, "specialized" function. But what could that be targeting? It is true that, for example in pectin, there is a minimum amount of glucuronic acid that can be found on the side chains, and it is plausible that such residues could interact with lignin in forming LCCs similarly as observed in glucuronoxylans. If this is the function these enzymes are performing in pectin, that it no surprise that no activity was detected with the boosting assay we setup in the work described in **Paper I**. In fact, the substrates used are derived from sources with low content of lignin, and that together with the low occurrence of GlcA, make a putative LC ester between the two a rare bond, difficult to detect. Especially with the type of assay we used, because the benefit in removing such a rare bond, to the saccharification of the biomass would be difficult to quantify. Another possibility is that these enzymes might be targeting only terminal esterified uronic acid moieties, in the backbone or in the side chains of the pectin polymers. If this was the case, no new site would be liberated for a the PL1 used in our assay, resulting in no boosting and explaining our results.

The second aim of this thesis was to assess whether CE15 enzymes could exhibit a certain preference towards the two different species of ester bonds (α and γ) that have been reported LC bonds between glucuronoxylan and lignin. Several previously characterized members of this family, together with three newly discovered and produced, were characterized on two model substrates mimicking the two ester species (**Paper II**). Our results defined the substrate specificity profile, of the enzymes tested, towards the α- and the γ-ester. The phylogenetic analysis provided initial evidence that the observed substrate specificities could be phylogenetically related, with enzymes grouped in the same area of the three

showing similar specificity. Finally, a structural analysis using both experimentally determined structures and predicted models, allowed us to identify the possible determinants of the substrate specificity. The results from this study elucidate (at least for bacterial enzymes) one of the possible explanations for the multiplicity of CE15 enzymes observed in many microorganisms. If the α- and the γ-ester coexist in nature, it is likely that their relative abundance would change depending on the source and the area of the plant cell wall, and it is possible that an organism would produce only the CE15 enzymes specific for a certain ester encountered in the biomass it is growing on.

The third aim was to investigate the role of a putative multicatalytic enzyme predicted to be constituted by a AA5 domain fused together to a putative CE3 domain. The full-length constructs and the two separate domains were produced and studied (**Paper III**). The AA5 domain was characterized on a wide array of primary alcohols substrates, and it showed to have a broad specificity. It exhibited a higher efficiency only on a handful of substrates with different characteristics (Furans, di-alcohol, and alkyl alcohol with longer chain), a quite puzzling results that gives little to no indication on the possible biological role of this enzyme. Possibly, the natural substrate of the enzyme was not within our array of compounds, and in fact, the activity levels on most of the substrates were very low, with most of the reactions unsaturable even at very high concentrations up to 500 mM. Also, the activities on furans, despite relatively high within the substrates tested in this study, are still at least 10-fold lower than previously reported for other enzymes [137]. One could argue that we could have tested more substrates, but the reality is that finding new relevant substrates is not easy considering we still have no indication on the biological function of the enzyme. And this is valid for CROs in general. It has been proposed that these enzymes could be producing hydrogen peroxide as a co-substrate for other enzymes; something that would partially explain the broad substrate specificity of many AA5 enzymes. In fact, by being less selective, it would be easier for the enzyme to find a substrate to oxidize, thus producing hydrogen peroxide. Regarding *Bp*AlcOx specifically, perhaps the natural function has to be researched within *B. pseudomallei* pathogenicity. It could be that the enzyme we produce could be involved in procuring oxidative damage favoring the infection. The activity of the full-length enzyme was comparable to that of the single domain on the substrates that were tested, thus excluding a possible role of the additional domain in affecting the oxidative reaction performed by the AA5 enzyme. Assays to detect an acetylxylan esterase activity were performed both using the full-length enzyme and the CE3 domain, but no activity was observed. Structural analysis on predicted models of the enzymes allowed to see that the CE3 domain carries two inactivating mutations in the active site, as well as numerous other modifications in its surrounding, that can explain the absence of acetylxylan esterase activity.

The fourth aim of this thesis was to gain insights into the possible structural determinants of the substrate specificity observed for a AA9 LPMO from M. cinnamomea. The tridimensional structure of the enzyme was solved from the collected dataset, and it was compared to that of other AA9 LPMOs (**Paper IV**). One of the most intriguing observations is the lack of a protruding ridge from the substrate-binding surface, that has been reported to play a major role in interaction with soluble substrates, despite *Mc*AA9F being active on oligosaccharides. A superimposition with other structures in complex with cellohexaose, allowed to identify which are the putative points of contact of the enzyme with a soluble

substrate, suggesting that these could be sufficient for the binding. Speculatively speaking, this enzyme could have evolved to be able to be active on soluble substrates without the ridge, keeping a flat surface and thus remaining active also on crystalline substrates. A broader specificity could represent a sort of advantage for the host microorganism, being able to target multiple substrates with one enzyme. However, this is somehow in contrast with the fact that *M. cinnamomea* has 8 additional genes encoding AA9 enzymes. The multiplicity of AA9 enzymes has in fact often been "explained" has a way of having enzymes targeting different substrates. Another partial planation for this multiplicity could be the fact that is not unusual for different AA9 enzymes from the same fungus to show different regioselectivity [176, 196].

Chapter VI: Future perspectives

Despite the insights presented in this thesis, there is still significant work to be done in the investigation of CE15 enzymes. No new activity towards pectin was demonstrated and still no explanation for the presence of CE15 encoding genes in PULs targeting non-xylan polysaccharides was given. It is worth to point out that the lack of any boosting effect on the commercial cocktail acting on pectin, could also be due to the way the assay was set up. In fact, we performed a multi-point stopped assay, without optimizing it to follow the kinetics of the hydrolysis performed by the commercial cocktail. This way, it may be that at the time point we are measuring the sugar release, the hydrolysis of the available substrate has already gone to completion thus possibly hiding any boosting effect of the CE15 enzyme. Unfortunately, further optimizing the reaction would have required significant time and efforts that were not compatible with the available timeframe, but could definitely be an option for future works. However, methodologies that would allow the direct assessment of the activity enzymes would be preferable, and in this direction, the development of more straightforward methods of analyzing the LCC content in a substrate to be able to observe their cleavage by CE15 enzymes, could definitely be of help to gain further insight on other possible biological roles that this family of enzymes might be hindering.

Our work in **Paper II** have helped shedding light on the substrate specificity on a wide selection of bacterial and fungal CE15 enzymes towards two different types of ester linkages, α- and γ-ester, giving another possible explanation to the multiplicity of CE15 sequences observed in certain microorganisms. Showing how different CE15 enzyme from the same species can have a preference for a specific type of ester, though always being active on both. A more straightforward methodology for analyzing the LCC content in biomass would be of great use also here, to quantify the abundance of the two species of esters and investigate any difference in the capability of hydrolyzing them by α-preferring and γ-preferring enzymes. In fact, the model substrates extensively used in **Paper I** and **II** represent a great tool to obtain kinetic parameters of multiple targets and compare them to identify potential promising and interesting candidates. But to confirm native activity on biomass it is necessary to use develop and combine multiple, complementary analytical strategies. Methods that directly detect changes in the lignocellulosic structure upon uronoyl esterase treatment (like NMR and imaging techniques) could be combined with techniques that analyze enzymatically released products (such as HPAEC-PAD and MS). However, several challenges need to be addressed in the development of new methodologies. The low abundance of the bonds of interest makes it already difficult to detect and characterize them extensively, but the problem is amplified by the current methods of extraction and sample preparation which can break or alter these bonds, and by the current technique of analysis that often suffer from high background noise attributable to signals from other bonds.

Our efforts in the work described in **Paper III** lead to a biochemical characterization of one of the first bacterial AA5 enzyme, contributing to investigate the diversity of a family that so far has seen little exploration in the bacteria domain. However, the characterization of one single new member is not enough as still a lot remains unknown. Many of the AA5

enzymes characterized to date have shown, like *Bp*AlcOx, a broad substrate specificity, given very little information on their possible natural substrate and biological role. Further studies and possibly the adoption of high-throughput methods to screening activity on a wide range of substrates could help in gaining more insight in regards of the natural function of these enzymes. A deeper investigation of the genomic surroundings of AA5 sequences and on the microorganisms could help in selecting more relevant substrates to focus the screening on. Similarly, we could not draw any conclusion on the *Bp*AlcOx biological role, but possibly it has to be researched within *B. pseudomallei* and its mode of action in nature. Being a human pathogen (BSL 3), possibly *Bp*AlcOx could be involved, for example, in the infection process by causing oxidative damage. What still remains puzzling is also the role of the predicted CE3 domain attached to *Bp*AlcOx in the protein produce natively by *B. pseudomallei*. Our results showed in **Paper III** suggest that this domain might have evolved to fulfill another function rather than that of an acetylxylan esterase. Perhaps the observed modifications in the surroundings of the active site (eg. multiple tyrosine residues) are constituting a metal (possibly copper, as a cofactor of AA5 enzymes) binding domain with a specific role in the *Bp*AlcOx natural function. Multiple studies could be done in order to investigate this hypothesis. The easiest option would probably be to determine the melting temperature of the *Bp*CE3 domain before and after the binding of copper, with the idea that this ion could stabilize the protein. However, not all proteins are stabilized upon binding to a metal ion, and in this case no effect on thermal stability can be appreciated. Another, more complicated option, could be to perform isothermal titration colorimetry to study directly if there is an interaction between the CE3 domain and a metal ion.

Acknowledgments

I don't think i can (and have to) explain how completing a PhD can be challenging in so many ways. I cannot imagine myself going through all these years without the help, support, presence of all the people that, in one way or another, have been around me. I would like to take the chance to thank (in no particular order) some of the people that have supported me through this journey of my life.

Of course, I would not be here trying to wrap up more than 4 years' worth of studies, if Johan in the first place had not accepted me (again) in this group. Thank you for the opportunity you gave me. My gratitude is also for all your efforts as a supervisor. You have always been there and available when I needed help. From allowing me some flexibility in organizing my work to face personal problems, to smaller things like signing last minute applications afterhours. But also, I realize I must not have been the easiest of the candidate to supervise, with my somewhat scarce communication, and most of all my reluctancy in reaching out for help. Thank you for the patience and the "dexterity" you had to use to supervise me and our projects.

I would like to thank also Lauren and Pernilla, who acted as my co-supervisor and examiner, respectively. For every feedback and every piece of advice that contributed to improving my work and my thesis. Despite me not being particularly good in reaching out for help and asking for feedbacks, you have always had a response super quickly and provided great inputs. Thank you for contributing in improving my work and anything that I produced from it.

A thank you is also due to the members of my defence committee, Bastien, Jane, Mats, Mikael, and my opponent Maija, for dedicating the time to read and (by the time anyone reads) approve this thesis, and for all the discussion and questions that will come during the defence. A special mention to Yvonne as well, for accepting to be the chairperson and "hosting" this event.

A big appreciation to Anna, Anne-Lise, Gunilla, Erica, Barbro, Jenny for their work on the administration side of the department. And to Julia, Thenia, Vijay, and Pun for keeping the lab together and ensure that everything would run as smooth as possible, for thinking about everyone's safety and comfort and always being prepositive in trying to improve the working environment. Everyone might appear always busy prioritizing their research, but the amount of work you all do behind the scenes is of undeniable importance for everyone to succeed. Thank you!

The work described in this thesis is the result of the efforts of several people, and not only mine. I would like to thank all of my co-authors Leila, Sanchari, Gisela, Tom, Miriam, Scott, Kris, and Johan. I appreciate all the contributions you made for all that is included in this thesis. Miriam, you arrived here visiting only for a few months, thank you for adapting so quickly to all the changes in the procedures and habits and yet manage to deliver so much. Tom, thank you for making it super easy to share the load of work and responsibilities on our project, for all the discussions, ideas and thoughts that we shared in and out of the the lab. I really enjoyed this collaboration and had a lot of fun, I appreciate it. Scott, after a quick overview on things at the beginning of this journey, our research paths divided for you to explore other interests, only for rejoin for the second part of this PhD and resulted in being co-authors in three of the papers presented here in this thesis. I enjoyed working side by side with you, sharing tasks and even plate-reader time. I think I should also take the chance to thank you for everything I learned from you during my MSc thesis here, without that I do not know how I would have endured through all these years.

I would also like to thank the Maple group and all IndBio, including present and former members that have been around throughout the years (far too numerous to call out individually). People have

come and gone, but everyone has always been helpful and given their contribution to the nice work environment we have. I appreciate this.

A special mention goes to *hey Google*, then *Extended* to *Alphabet* group. The first part of PhD would have not been the same without all the fun, experiences, and support that has come from you during a variety of activities in and outside campus. Thank you for being there Sir Dan, Daniel, Elena, Gerard, Jae, Javi, Jeroen, Otavio, Rike, Thenia, and Vero. Analogously I would like to thank also the later M*editerraneans* group, for all the shared emotions during "late" lunches at work and activities all together. Thank you, Cecilia, Elena, Ibai, Luca, Mauri, Vero and later on Fabio and Kamesh.

A special thought for Alessio and Saul, for all that we did together, for all the laughs, for making the best out of all the situations we found ourselves into… for making any WWSC event a great experience to remember. Que guapi!

I also would like to say grazie to all my friends back home. Despite the distance, which has forced me to miss many events and celebrations, we have always managed to find the time to group together and have fun, whether with simple grigliate at somebody's place, dinners and aperitivi, trips to the lake, to the sea, or the mountains and even abroad. I'm grateful for all the time and all the things we experienced together and knowing that could count on you guys has always been of strong help and support. Thank you, Alessandro, Alessia, Edoardo, Efrem, Ekin, Francesca, Riccardo, Stefan8 e Stefania.

I would have never made it till the end if it wasn't for Giada. All the years that we have spent apart from each other have not been easy, for none of us, but never have I doubt your unconditional support along the road. Thank you for every call you answered, for every time you took the chance to visit me here in Sweden, for all the memories we were able to share despite living most of our time separated. Thank you for having the courage of deciding to leave everything you had back home and move abroad to stay with me, to support me in what probably is supposed the be most stressful time of the PhD, but that thanks to you, it wasn't. Thank you for all the last days of writing spent in isolation with me. I'm extremely grateful to have you by my side.

Ora uno speciale ringranziamento a tutta la mia famiglia. In particolar modo ai miei genitori, Enrica e Robeto. Senza di voi non avrei potuto mai neanche immaginare di arrivare a fare un dottorato, ed é merito vostro se adesso sono qui che lo sto completando, preparandomi per la difesa della tesi. Grazie per avermi concesso l'educazione necessaria per andare cosí lontano, anche se questo ha volute dire non essere lí con voi per molti compleanni e gite di famiglia. Grazie per avermi dato sempre il vostro supporto e per avere sempre creduto in me durante questo capitolo della mia vita. Grazie anche a Marzia e Mattia, sempre pronti per organizzare qualcosa da fare, per tutti i bei momenti passata insieme in questi anni, che hanno aiutato a togliermi un sacco di stress. Grazie a tutti per la famiglia che abbiamo, che mi ha sempre fatto sentire supportato, e amare il tempo passato in visita a casa.

References

1. Mbow, H.-O.P., Reisinger, A., et al. , *Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.* (SR2). Ginevra, IPCC, 2017.
2. Hainsch, K., et al., *Energy transition scenarios: What policies, societal attitudes, and technology developments will realize the EU Green Deal?* Energy, 2022. **239**: p. 122067.
3. De Bhowmick, G., A.K. Sarmah, and R. Sen, *Lignocellulosic biorefinery as a model for sustainable development of biofuels and value added products.* Bioresource technology, 2018. **247**: p. 1144-1154.
4. Amoah, J., et al., *Bioenergy and biorefinery: feedstock, biotechnological conversion, and products.* Biotechnology journal, 2019. **14**(6): p. 1800494.
5. Hassan, S.S., G.A. Williams, and A.K. Jaiswal, *Lignocellulosic biorefineries in Europe: current state and prospects.* Trends in biotechnology, 2019. **37**(3): p. 231-234.
6. Sousa, A.F., et al., *Biobased polyesters and other polymers from 2, 5-furandicarboxylic acid: a tribute to furan excellency.* Polymer chemistry, 2015. **6**(33): p. 5961-5983.
7. Abdelraheem, E.M., et al., *Biocatalysis explained: from pharmaceutical to bulk chemical production.* Reaction Chemistry & Engineering, 2019. **4**(11): p. 1878-1894.
8. Bornscheuer, U.T. and R.J. Kazlauskas, *Hydrolases in organic synthesis: regio-and stereoselective biotransformations*. 2006: John Wiley & Sons.
9. Hüttel, W. and M. Müller, *Regio-and stereoselective intermolecular phenol coupling enzymes in secondary metabolite biosynthesis.* Natural Product Reports, 2021. **38**(5): p. 1011-1043.
10. Meyer, H.-P., et al., *The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC).* Catalysis Science & Technology, 2013. **3**(1): p. 29-40.
11. Sheldon, R.A., D. Brady, and M.L. Bode, *The Hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis.* Chemical science, 2020. **11**(10): p. 2587-2605.
12. Anastas, P. and N. Eghbali, *Green chemistry: principles and practice.* Chemical Society Reviews, 2010. **39**(1): p. 301-312.
13. SJÖSTRÖM, E., *Wood polysaccharides.* Wood chemistry, 1993: p. 51-70.
14. Pauly, M. and K. Keegstra, *Cell-wall carbohydrates and their modification as a resource for biofuels.* The Plant Journal, 2008. **54**(4): p. 559-568.
15. Alberts, B., *Molecular biology of the cell 4th edition.* (No Title), 2002.
16. Shewry, P.R., *Biochemistry & Molecular Biology of Plants. B.B. Buchanan, W. Gruissem and R.L. Jones (eds), 2000.* Plant Growth Regulation, 2001. **35**(1): p. 105-106.
17. Houston, K., et al., *The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions.* Frontiers in plant science, 2016. **7**: p. 984.
18. Falcioni, R., et al., *Cell wall structure and composition is affected by light quality in tomato seedlings.* Journal of Photochemistry and Photobiology B: Biology, 2020. **203**: p. 111745.
19. Loix, C., et al., *Reciprocal interactions between cadmium-induced cell wall responses and oxidative stress in plants.* Frontiers in plant science, 2017. **8**: p. 1867.
20. Carpita, N.C. and M.C. McCann, *Redesigning plant cell walls for the biomass-based bioeconomy.* Journal of Biological Chemistry, 2020. **295**(44): p. 15144-15157.
21. Nishimura, H., et al., *Direct evidence for α ether linkage between lignin and carbohydrates in wood cell walls.* Scientific reports, 2018. **8**(1): p. 6538.
22. Van Dyk, J. and B. Pletschke, *A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—factors affecting enzymes, conversion and synergy.* Biotechnology advances, 2012. **30**(6): p. 1458-1480.
23. Nelson, D., *Lehninger principles of biochemistry: Macmillan.* Lehninger principles of biochemistry: Macmillan, 2008.

24. George, J. and S. Sabapathi, *Cellulose nanocrystals: synthesis, functional properties, and applications.* Nanotechnology, science and applications, 2015: p. 45-54.
25. Henriksson, G. and H. Lennholm, *4. Cellulose and Carbohydrate Chemistry*, in *Volume 1 Wood Chemistry and Wood Biotechnology*, E. Monica, G. Göran, and H. Gunnar, Editors. 2009, De Gruyter: Berlin, New York. p. 71-100.
26. Scheller, H.V. and P. Ulvskov, *Hemicelluloses.* Annual review of plant biology, 2010. **61**: p. 263-289.
27. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: a review.* Bioresource technology, 2010. **101**(13): p. 4775-4800.
28. Terrett, O.M. and P. Dupree, *Covalent interactions between lignin and hemicelluloses in plant secondary cell walls.* Current Opinion in Biotechnology, 2019. **56**: p. 97-104.
29. Shakeel, U., et al., *Fundamentals of Lignin-Carbohydrate Complexes and Its Effect on Biomass Utilization.* Emerging Technologies for Biorefineries, Biofuels, and Value-Added Commodities, 2021: p. 133-155.
30. Fu, G.-Q., et al., *Isolation, purification, and potential applications of xylan.* Production of materials from sustainable biomass resources, 2019: p. 3-35.
31. Pereira, C.S., et al., *Effects of xylan side-chain substitutions on xylan–cellulose interactions and implications for thermal pretreatment of cellulosic biomass.* Biomacromolecules, 2017. **18**(4): p. 1311-1321.
32. Voragen, A.G., et al., *Pectin, a versatile polysaccharide present in plant cell walls.* Structural Chemistry, 2009. **20**: p. 263-275.
33. Mohnen, D., *Pectin structure and biosynthesis.* Current opinion in plant biology, 2008. **11**(3): p. 266-277.
34. Caffall, K.H. and D. Mohnen, *The structure, function, and biosynthesis of plant cell wall pectic polysaccharides.* Carbohydrate research, 2009. **344**(14): p. 1879-1900.
35. Lawoko, M., *Unveiling the structure and ultrastructure of lignin carbohydrate complexes in softwoods.* International Journal of Biological Macromolecules, 2013. **62**: p. 705-713.
36. Hartmann, M. and D. Kaplan, *Biopolymers from renewable resources.* Kaplan, DL, Ed, 1998. **367**: p. 629-638.
37. Shrotri, A., H. Kobayashi, and A. Fukuoka, *Catalytic conversion of structural carbohydrates and lignin to chemicals*, in *Advances in catalysis*. 2017, Elsevier. p. 59-123.
38. Ragauskas, A.J., et al., *Lignin valorization: improving lignin processing in the biorefinery.* science, 2014. **344**(6185): p. 1246843.
39. Upton, B.M. and A.M. Kasko, *Strategies for the conversion of lignin to high-value polymeric materials: review and perspective.* Chemical reviews, 2016. **116**(4): p. 2275-2306.
40. Weng, C., X. Peng, and Y. Han, *Depolymerization and conversion of lignin to value-added bioproducts by microbial and enzymatic catalysis.* Biotechnology for Biofuels, 2021. **14**(1): p. 1-22.
41. Menezes, F.F., et al., *Depolymerization of enzymatic hydrolysis lignin: Review of technologies and opportunities for research.* Fuel, 2023. **342**: p. 127796.
42. Tramontina, R., et al., *Consolidated production of coniferol and other high-value aromatic alcohols directly from lignocellulosic biomass.* Green chemistry, 2020. **22**(1): p. 144-152.
43. Li, M., Y. Pu, and A.J. Ragauskas, *Current understanding of the correlation of lignin structure with biomass recalcitrance.* Frontiers in chemistry, 2016. **4**: p. 45.
44. Oliveira, D.M., et al., *Lignin plays a key role in determining biomass recalcitrance in forage grasses.* Renewable Energy, 2020. **147**: p. 2206-2217.
45. Henriksson, G., *What are the biological functions of lignin and its complexation with carbohydrates?* Nordic Pulp & Paper Research Journal, 2017. **32**(4): p. 527-541.
46. Koshijima, T. and T. Watanabe, *Association between lignin and carbohydrates in wood and other plant tissues*. 2013: Springer Science & Business Media.
47. Brunow, G. and K. Lundquist, *Functional groups and bonding patterns in lignin (including the lignin-carbohydrate complexes).* Lignin and lignans: advances in chemistry, 2010. **1**.
48. Balakshin, M., et al., *Quantification of lignin–carbohydrate linkages with high-resolution NMR spectroscopy.* Planta, 2011. **233**: p. 1097-1110.

49. Balakshin, M.Y., E.A. Capanema, and H.-m. Chang, *MWL fraction with a high concentration of lignin-carbohydrate linkages: Isolation and 2D NMR spectroscopic analysis.* 2007.
50. Giummarella, N., et al., *Structural features of mildly fractionated lignin carbohydrate complexes (LCC) from spruce.* Rsc Advances, 2016. **6**(48): p. 42120-42131.
51. Yuan, T.-Q., et al., *Characterization of lignin structures and lignin–carbohydrate complex (LCC) linkages by quantitative 13C and 2D HSQC NMR spectroscopy.* Journal of agricultural and food chemistry, 2011. **59**(19): p. 10604-10614.
52. Sapouna, I. and M. Lawoko, *Deciphering lignin heterogeneity in ball milled softwood: unravelling the synergy between the supramolecular cell wall structure and molecular events.* Green Chemistry, 2021. **23**(9): p. 3348-3364.
53. Tarasov, D., M. Leitch, and P. Fatehi, *Lignin–carbohydrate complexes: properties, applications, analyses, and methods of extraction: a review.* Biotechnology for biofuels, 2018. **11**(1): p. 1-28.
54. Fengel, D., G. Wegener, and A. Greune, *Studies on the delignification of spruce wood by organosolv pulping using SEM-EDXA and TEM.* Wood Science and Technology, 1989. **23**: p. 123-130.
55. Grabber, J.H., J. Ralph, and R.D. Hatfield, *Cross-linking of maize walls by ferulate dimerization and incorporation into lignin.* Journal of agricultural and food chemistry, 2000. **48**(12): p. 6106-6113.
56. Iiyama, K., T.B.-T. Lam, and B.A. Stone, *Covalent cross-links in the cell wall.* Plant physiology, 1994. **104**(2): p. 315.
57. Mosbech, C., et al., *The natural catalytic function of Cu GE glucuronoyl esterase in hydrolysis of genuine lignin–carbohydrate complexes from birch.* Biotechnology for biofuels, 2018. **11**: p. 1-9.
58. d'Errico, C., et al., *Improved biomass degradation using fungal glucuronoyl-esterases-hydrolysis of natural corn fiber substrate.* J Biotechnol, 2016. **219**: p. 117-23.
59. Arnling Bååth, J., et al., *A glucuronoyl esterase from Acremonium alcalophilum cleaves native lignin-carbohydrate ester bonds.* FEBS Lett, 2016. **590**(16): p. 2611-8.
60. Imamura, T., et al., *Ester linkages between lignin and glucuronic acid in lignin-carbohydrate complexes from Fagus crenata.* Phytochemistry, 1994. **37**(4): p. 1165-1173.
61. Nishimura, H., et al., *Direct evidence of α ester linkages between lignin and glucuronoxylan that reveal the robust heteropolymeric complex in plant cell walls.* 2022.
62. Li, K. and R.F. Helm, *Synthesis and rearrangement reactions of ester-linked lignin-carbohydrate model compounds.* Journal of Agricultural and Food Chemistry, 1995. **43**(8): p. 2098-2103.
63. Wang, P., J. Fan, and Y. Xie, *Synthesis and characterization of pectindehydrogenation polymer complex by isotopic labeling method.* Cellul. Chem. Technol, 2013. **47**: p. 401-408.
64. Consortium, T.U., *UniProt: the Universal Protein Knowledgebase in 2023.* Nucleic Acids Research, 2022. **51**(D1): p. D523-D531.
65. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future.* Nucleic acids research, 2016. **44**(D1): p. D279-D285.
66. Ahmed, S., S. Riaz, and A. Jamil, *Molecular cloning of fungal xylanases: an overview.* Applied microbiology and biotechnology, 2009. **84**: p. 19-35.
67. Terrapon, N., et al., *PULDB: the expanded database of Polysaccharide Utilization Loci.* Nucleic Acids Res, 2018. **46**(D1): p. D677-d683.
68. Ausland, C., et al., *dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates.* Nucleic Acids Research, 2020. **49**(D1): p. D523-D528.
69. Arnling Bååth, J., et al., *Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion.* Biotechnology for biofuels, 2018. **11**(1): p. 1-14.
70. Cataldi, T.R., C. Campa, and G.E. De Benedetto, *Carbohydrate analysis by high-performance anion-exchange chromatography with pulsed amperometric detection: the potential is still growing.* Fresenius' journal of analytical chemistry, 2000. **368**: p. 739-758.

71. Pitt, J.J., *Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry.* The Clinical Biochemist Reviews, 2009. **30**(1): p. 19.
72. Argyropoulos, D.S., *Heteronuclear NMR spectroscopy of lignins*. 2010: Taylor and Francis Group, LLC: Boca Raton, FL.
73. Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide Protein Data Bank.* Nature Structural & Molecular Biology, 2003. **10**(12): p. 980-980.
74. Rupp, B., *Biomolecular crystallography: principles, practice, and application to structural biology*. 2009: Garland Science.
75. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic acids research, 2018. **46**(W1): p. W296-W303.
76. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature protocols, 2015. **10**(6): p. 845-858.
77. Du, Z., et al., *The trRosetta server for fast and accurate protein structure prediction.* Nat Protoc, 2021. **16**(12): p. 5634-5651.
78. Mitchell, A.L., et al., *MGnify: the microbiome analysis resource in 2020.* Nucleic acids research, 2020. **48**(D1): p. D570-D578.
79. Hatch, V., *DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins.* https://www.embl.org/news/science/alphafolddatabase-launch/, 2021.
80. Drula, E., et al., *The carbohydrate-active enzyme database: functions and literature.* Nucleic Acids Res, 2022. **50**(D1): p. D571-d577.
81. Payne, C.M., et al., *Fungal cellulases.* Chemical reviews, 2015. **115**(3): p. 1308-1448.
82. Reese, E.T., *A microbiological process report; enzymatic hydrolysis of cellulose.* Appl Microbiol, 1956. **4**(1): p. 39-45.
83. Bischof, R.H., J. Ramoni, and B. Seiboth, *Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei.* Microbial cell factories, 2016. **15**(1): p. 1-13.
84. Lynd, L.R., et al., *Microbial cellulose utilization: fundamentals and biotechnology.* Microbiology and molecular biology reviews, 2002. **66**(3): p. 506-577.
85. Ekwe, E., et al., *Non-hydrolytic cellulose active proteins: research progress and potential application in biorefineries.* Industrial Biotechnology, 2013. **9**(3): p. 123-131.
86. Lenfant, N., et al., *A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9.* Carbohydrate research, 2017. **448**: p. 166-174.
87. Horn, S.J., et al., *Novel enzymes for the degradation of cellulose.* Biotechnology for biofuels, 2012. **5**(1): p. 1-13.
88. Hemsworth, G.R., et al., *Lytic polysaccharide monooxygenases in biomass conversion.* Trends in biotechnology, 2015. **33**(12): p. 747-761.
89. Kormelink, F., et al., *Purification and characterization of three endo-(1, 4)-β-xylanases and one β-xylosidase from Aspergillus awamori.* Journal of Biotechnology, 1993. **27**(3): p. 249-265.
90. Kormelink, F.J., et al., *Mode of action of the xylan-degrading enzymes from Aspergillus awamori on alkali-extractable cereal arabinoxylans.* Carbohydrate Research, 1993. **249**(2): p. 355-367.
91. Saha, B.C. and R.J. Bothast, *Enzymology of xylan degradation*. 1999, ACS Publications.
92. Martens-Uzunova, E.S. and P.J. Schaap, *Assessment of the pectin degrading enzyme network of Aspergillus niger by functional genomics.* Fungal Genetics and Biology, 2009. **46**(1): p. S170-S179.
93. van den Brink, J. and R.P. de Vries, *Fungal enzyme sets for plant polysaccharide degradation.* Applied microbiology and biotechnology, 2011. **91**: p. 1477-1492.
94. Sutherland, I.W., *Polysaccharide lyases.* FEMS microbiology reviews, 1995. **16**(4): p. 323-347.
95. Azadi, P., et al., *The backbone of the pectic polysaccharide rhamnogalacturonan I is cleaved by an endo hydrolase and an endo lyase.* Glycobiology, 1995. **5**(8): p. 783-789.
96. Gandla, M.L., et al., *Expression of a fungal glucuronoyl esterase in Populus: effects on wood properties and saccharification efficiency.* Phytochemistry, 2015. **112**: p. 210-220.

97. Mosbech, C., et al., *The natural catalytic function of CuGE glucuronoyl esterase in hydrolysis of genuine lignin–carbohydrate complexes from birch.* Biotechnology for Biofuels, 2018. **11**(1): p. 71.
98. De Santi, C., N.P. Willassen, and A. Williamson, *Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine arctic metagenome.* PLoS One, 2016. **11**(7): p. e0159345.
99. Aurilia, V., et al., *Three multidomain esterases from the cellulolytic rumen anaerobe Ruminococcus flavefaciens 17 that carry divergent dockerin sequences.* Microbiology (Reading), 2000. **146 (Pt 6)**: p. 1391-1397.
100. Špániková, S. and P. Biely, *Glucuronoyl esterase–novel carbohydrate esterase produced by Schizophyllum commune.* FEBS letters, 2006. **580**(19): p. 4597-4601.
101. Li, X.-L., et al., *Identification of genes encoding microbial glucuronoyl esterases.* FEBS letters, 2007. **581**(21): p. 4029-4035.
102. Pokkuluri, P.R., et al., *Structure of the catalytic domain of glucuronoyl esterase Cip2 from Hypocrea jecorina.* Proteins, 2011. **79**(8): p. 2588-2592.
103. Agger, J.W., et al., *New insights to diversity and enzyme–substrate interactions of fungal glucuronoyl esterases.* Applied Microbiology and Biotechnology, 2023: p. 1-11.
104. Monrad, R.N., et al., *Glucuronoyl esterases: diversity, properties and biotechnological potential. A review.* Critical Reviews in Biotechnology, 2018. **38**(7): p. 1121-1136.
105. Sunner, H., et al., *Glucuronoyl Esterase Screening and Characterization Assays Utilizing Commercially Available Benzyl Glucuronic Acid Ester.* Molecules, 2015. **20**(10): p. 17807-17.
106. Topakas, E., et al., *Functional expression of a thermophilic glucuronoyl esterase from Sporotrichum thermophile: identification of the nucleophilic serine.* Applied microbiology and biotechnology, 2010. **87**: p. 1765-1772.
107. ĎURANOVÁ, M., et al., *Fungal glucuronoyl esterases and substrate uronic acid recognition.* Bioscience, biotechnology, and biochemistry, 2009. **73**(11): p. 2483-2487.
108. Wong, D.W., et al., *Functional cloning and expression of the Schizophyllum commune glucuronoyl esterase gene and characterization of the recombinant enzyme.* Biotechnology research international, 2012. **2012**.
109. Špániková, S., et al., *Synthetic esters recognized by glucuronoyl esterase from Schizophyllum commune.* Archives of Microbiology, 2007. **188**(2): p. 185-189.
110. d'Errico, C., et al., *Enzymatic degradation of lignin-carbohydrate complexes (LCCs): Model studies using a fungal glucuronoyl esterase from Cerrena unicolor.* Biotechnology and bioengineering, 2015. **112**(5): p. 914-922.
111. Huynh, H.H., et al., *A novel glucuronoyl esterase from Aspergillus fumigatus—the role of conserved Lys residue in the preference for 4-O-methyl glucuronoyl esters.* Applied microbiology and biotechnology, 2018. **102**: p. 2191-2201.
112. Hüttner, S., et al., *Characterisation of three fungal glucuronoyl esterases on glucuronic acid ester model compounds.* Applied Microbiology and Biotechnology, 2017. **101**(13): p. 5301-5311.
113. Biely, P., et al., *Glucuronoyl esterases are active on the polymeric substrate methyl esterified glucuronoxylan.* FEBS letters, 2015. **589**(18): p. 2334-2339.
114. Huynh, H.H. and M. Arioka, *Functional expression and characterization of a glucuronoyl esterase from the fungus Neurospora crassa: identification of novel consensus sequences containing the catalytic triad.* The Journal of general and applied microbiology, 2016. **62**(5): p. 217-224.
115. Ernst, H.A., et al., *The structural basis of fungal glucuronoyl esterase activity on natural substrates.* Nature Communications, 2020. **11**(1): p. 1026.
116. Mazurkewich, S., et al., *Structural and functional investigation of a fungal member of carbohydrate esterase family 15 with potential specificity for rare xylans.* Acta Crystallographica Section D: Structural Biology, 2023. **79**(6).
117. De Santi, C., et al., *Structural insight into a CE15 esterase from the marine bacterial metagenome.* Scientific Reports, 2017. **7**(1): p. 17278.

118. Mazurkewich, S., et al., *Structural and biochemical studies of the glucuronoyl esterase OtCE15A illuminate its interaction with lignocellulosic components.* Journal of Biological Chemistry, 2019. **294**(52): p. 19978-19987.
119. Bååth, J.A., et al., *Structure–function analyses reveal that a glucuronoyl esterase from Teredinibacter turnerae interacts with carbohydrates and aromatic compounds.* Journal of Biological Chemistry, 2019. **294**(16): p. 6635-6644.
120. Krska, D., et al., *Structural and Functional Analysis of a Multimodular Hyperthermostable Xylanase-Glucuronoyl Esterase from Caldicellulosiruptor kristjansonii.* Biochemistry, 2021. **60**(27): p. 2206-2220.
121. Raji, O., et al., *The coordinated action of glucuronoyl esterase and α-glucuronidase promotes the disassembly of lignin-carbohydrate complexes.* FEBS Lett, 2021. **595**(3): p. 351-359.
122. Serrano, A., J. Carro, and A.T. Martínez, *Chapter Seven - Reaction mechanisms and applications of aryl-alcohol oxidase*, in *The Enzymes*, P. Chaiyen and F. Tamanoi, Editors. 2020, Academic Press. p. 167-192.
123. Urlacher, V.B. and K. Koschorreck, *Pecularities and applications of aryl-alcohol oxidases from fungi.* Applied Microbiology and Biotechnology, 2021. **105**: p. 4111-4126.
124. Daou, M. and C.B. Faulds, *Glyoxal oxidases: their nature and properties.* World J Microbiol Biotechnol, 2017. **33**(5): p. 87.
125. Kersten, P. and D. Cullen, *Copper radical oxidases and related extracellular oxidoreductases of wood-decay Agaricomycetes.* Fungal Genet Biol, 2014. **72**: p. 124-130.
126. Kersten, P.J. and T.K. Kirk, *Involvement of a new enzyme, glyoxal oxidase, in extracellular H2O2 production by Phanerochaete chrysosporium.* J Bacteriol, 1987. **169**(5): p. 2195-201.
127. Daou, M., et al., *Heterologous production and characterization of two glyoxal oxidases from Pycnoporus cinnabarinus.* Appl Environ Microbiol, 2016. **82**(16): p. 4867-75.
128. Whittaker, M.M., et al., *Glyoxal oxidase from Phanerochaete chrysosporium is a new radical-copper oxidase.* J Biol Chem, 1996. **271**(2): p. 681-7.
129. Šola, K., et al., *RUBY, a putative galactose oxidase, influences pectin properties and promotes cell-to-cell adhesion in the seed coat epidermis of Arabidopsis.* The Plant Cell, 2019. **31**(4): p. 809-831.
130. Šola, K., et al., *Expression patterns and functional characterization of Arabidopsis galactose oxidase-like genes suggest specialized roles for galactose oxidases in plants.* Plant and Cell Physiology, 2021. **62**(12): p. 1927-1943.
131. Cooper, J.A., et al., *Galactose oxidase from Polyporus circinatus, Fr.* J Biol Chem, 1959. **234**(3): p. 445-8.
132. Avigad, G., et al., *The D-galactose oxidase of Polyporus circinatus.* J Biol Chem, 1962. **237**: p. 2736-43.
133. Whittaker, M.M. and J.W. Whittaker, *Catalytic reaction profile for alcohol oxidation by galactose oxidase.* Biochemistry, 2001. **40**(24): p. 7140-8.
134. Andberg, M., et al., *A novel Colletotrichum graminicola raffinose oxidase in the AA5 Family.* Appl Environ Microbiol, 2017. **83**(20).
135. Mathieu, Y., et al., *Discovery of a fungal copper radical oxidase with high catalytic efficiency toward 5-hydroxymethylfurfural and benzyl alcohols for bioprocessing.* ACS Catalysis, 2020. **10**(5): p. 3042-3058.
136. Cleveland, M., et al., *Two Fusarium copper radical oxidases with high activity on aryl alcohols.* Biotechnol Biofuels, 2021. **14**(1): p. 138.
137. Yin, D.T., et al., *Structure-function characterization reveals new catalytic diversity in the galactose oxidase and glyoxal oxidase family.* Nat Commun, 2015. **6**: p. 10197.
138. Oide, S., et al., *Carbohydrate-binding property of a cell wall integrity and stress response component (WSC) domain of an alcohol oxidase from the rice blast pathogen Pyricularia oryzae.* Enzyme and Microbial Technology, 2019. **125**: p. 13-20.
139. Ribeaucourt, D., et al., *Identification of copper-containing oxidoreductases in the secretomes of three Colletotrichum species with a focus on copper radical oxidases for the*

biocatalytic production of fatty aldehydes. Appl Environ Microbiol, 2021. **87**(24): p. e0152621.
140. Cleveland, M.E., et al., *A survey of substrate specificity among Auxiliary Activity Family 5 copper radical oxidases.* Cell Mol Life Sci, 2021. **78**(24): p. 8187-8208.
141. Ribeaucourt, D., et al., *Comprehensive insights into the production of long chain aliphatic aldehydes using a copper-radical alcohol oxidase as biocatalyst.* ACS Sustainable Chemistry & Engineering, 2021. **9**(12): p. 4411-4421.
142. Bissaro, B., et al., *Tandem metalloenzymes gate plant cell entry by pathogenic fungi.* Sci Adv, 2022. **8**(51): p. eade9982.
143. Fong, Jessica K. and H. Brumer, *Copper radical oxidases: galactose oxidase, glyoxal oxidase, and beyond!* Essays in Biochemistry, 2022.
144. Koschorreck, K., S. Alpdagtas, and V.B. Urlacher, *Copper-radical oxidases: A diverse group of biocatalysts with distinct properties and a broad range of biotechnological applications.* Engineering Microbiology, 2022. **2**(3): p. 100037.
145. Parikka, K., E. Master, and M. Tenkanen, *Oxidation with galactose oxidase: multifunctional enzymatic catalysis.* Journal of Molecular Catalysis B: Enzymatic, 2015. **120**: p. 47-59.
146. Whittaker, J.W., *Free radical catalysis by galactose oxidase.* Chemical Reviews, 2003. **103**(6): p. 2347-2364.
147. Whittaker, J.W., *Galactose oxidase.* Advances in Protein Chemistry, 2002. **60**: p. 1-49.
148. Humphreys, K.J., et al., *Galactose oxidase as a model for reactivity at a copper superoxide center.* Journal of the American Chemical Society, 2009. **131**(13): p. 4657-4663.
149. Whittaker, J.W., *The radical chemistry of galactose oxidase.* Archives of Biochemistry and Biophysics, 2005. **433**(1): p. 227-239.
150. Siebum, A., et al., *Galactose oxidase and alcohol oxidase: scope and limitations for the enzymatic synthesis of aldehydes.* Journal of Molecular Catalysis B: Enzymatic, 2006. **41**(3-4): p. 141-145.
151. Parikka, K., et al., *Oxidation of polysaccharides by galactose oxidase.* Journal of Agricultural and Food Chemistry, 2010. **58**(1): p. 262-271.
152. Ito, N., et al., *Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase.* Nature, 1991. **350**(6313): p. 87-90.
153. Ito, N., et al., *Three-dimensional structure of galactose oxidase: an enzyme with a built-in secondary cofactor.* Faraday Discussions, 1992. **93**(0): p. 75-84.
154. Ito, N., et al., *Crystal Structure of a Free Radical Enzyme, Galactose Oxidase.* Journal of Molecular Biology, 1994. **238**(5): p. 794-814.
155. Correia, M.A., et al., *Crystal structure of a cellulosomal family 3 carbohydrate esterase from Clostridium thermocellum provides insights into the mechanism of substrate recognition.* J Mol Biol, 2008. **379**(1): p. 64-72.
156. Watanabe, M., et al., *Crystal structure of an acetylesterase from Talaromyces cellulolyticus and the importance of a disulfide bond near the active site.* FEBS Lett, 2015. **589**(11): p. 1200-6.
157. Vaaje-Kolstad, G., et al., *An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides.* Science, 2010. **330**(6001): p. 219-222.
158. Morgenstern, I., J. Powlowski, and A. Tsang, *Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family.* Briefings in functional genomics, 2014. **13**(6): p. 471-481.
159. Kracher, D., et al., *Extracellular electron transfer systems fuel cellulose oxidative degradation.* Science, 2016. **352**(6289): p. 1098-1101.
160. Wang, B., et al., *Activation of O2 and H2O2 by lytic polysaccharide monooxygenases.* ACS Catalysis, 2020. **10**(21): p. 12760-12769.
161. Martínez, A.T., *How to break down crystalline cellulose.* Science, 2016. **352**(6289): p. 1050-1051.
162. Costa, T.H., et al., *Demonstration-scale enzymatic saccharification of sulfite-pulped spruce with addition of hydrogen peroxide for LPMO activation.* Biofuels, Bioproducts and Biorefining, 2020. **14**(4): p. 734-745.

163. Müller, G., et al., *The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail.* Biotechnology for biofuels, 2018. **11**(1): p. 1-17.
164. Bissaro, B., et al., *Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2.* Nature chemical biology, 2017. **13**(10): p. 1123-1128.
165. Beeson, W.T., et al., *Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases.* Journal of the American Chemical Society, 2012. **134**(2): p. 890-892.
166. Hangasky, J.A., A.T. Iavarone, and M.A. Marletta, *Reactivity of O2 versus H2O2 with polysaccharide monooxygenases.* Proceedings of the National Academy of Sciences, 2018. **115**(19): p. 4915-4920.
167. Rieder, L., et al., *Fast and specific peroxygenase reactions catalyzed by fungal mono-copper enzymes.* Biochemistry, 2021. **60**(47): p. 3633-3643.
168. Kuusk, S. and P. Väljamäe, *Kinetics of H2O2-driven catalysis by a lytic polysaccharide monooxygenase from the fungus Trichoderma reesei.* Journal of Biological Chemistry, 2021. **297**(5).
169. Stepnov, A.A., V.G. Eijsink, and Z. Forsberg, *Enhanced in situ H2O2 production explains synergy between an LPMO with a cellulose-binding domain and a single-domain LPMO.* Scientific Reports, 2022. **12**(1): p. 6129.
170. Isaksen, T., et al., *A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides.* Journal of biological chemistry, 2014. **289**(5): p. 2632-2642.
171. Borisova, A.S., et al., *Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity.* Journal of Biological Chemistry, 2015. **290**(38): p. 22955-22969.
172. Agger, J.W., et al., *Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation.* Proceedings of the National Academy of Sciences, 2014. **111**(17): p. 6287-6292.
173. Hegnar, O.A., et al., *Quantifying oxidation of cellulose-associated glucuronoxylan by two lytic polysaccharide monooxygenases from Neurospora crassa.* Applied and Environmental Microbiology, 2021. **87**(24): p. e01652-21.
174. Frommhagen, M., et al., *Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase.* Biotechnology for Biofuels, 2015. **8**(1): p. 1-12.
175. Sun, P., et al., *Extending the diversity of Myceliophthora thermophila LPMOs: Two different xyloglucan cleavage profiles.* Carbohydrate Polymers, 2022. **288**: p. 119373.
176. Bennati-Granier, C., et al., *Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina.* Biotechnology for biofuels, 2015. **8**: p. 1-14.
177. Hüttner, S., et al., *Specific xylan activity revealed for AA9 lytic polysaccharide monooxygenases of the thermophilic fungus Malbranchea cinnamomea by functional characterization.* Applied and Environmental Microbiology, 2019. **85**(23): p. e01408-19.
178. Tõlgo, M., et al., *Comparison of six lytic polysaccharide monooxygenases from Thermothielavioides terrestris shows that functional variation underlies the multiplicity of LPMO genes in filamentous fungi.* Applied and Environmental Microbiology, 2022. **88**(6): p. e00096-22.
179. Hüttner, S., et al., *Combined genome and transcriptome sequencing to investigate the plant cell wall degrading enzyme system in the thermophilic fungus Malbranchea cinnamomea.* Biotechnology for biofuels, 2017. **10**(1): p. 1-16.
180. Tan, T.-C., et al., *Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation.* Nature communications, 2015. **6**(1): p. 7542.
181. Calderaro, F., et al., *Characterization of an AA9 LPMO from Thielavia australiensis, TausLPMO9B, under industrially relevant lignocellulose saccharification conditions.* Biotechnology for biofuels, 2020. **13**(1): p. 1-17.

182. Gaber, Y., et al., *Heterologous expression of lytic polysaccharide monooxygenases (LPMOs).* Biotechnology advances, 2020. **43**: p. 107583.
183. Vu, V.V., et al., *Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases.* Journal of the American Chemical Society, 2014. **136**(2): p. 562-565.
184. Li, X., et al., *Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases.* Structure, 2012. **20**(6): p. 1051-1061.
185. Vaaje-Kolstad, G., et al., *Structural diversity of lytic polysaccharide monooxygenases.* Current opinion in structural biology, 2017. **44**: p. 67-76.
186. Quinlan, R.J., et al., *Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components.* Proceedings of the National Academy of Sciences, 2011. **108**(37): p. 15079-15084.
187. Petrović, D.M., et al., *Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation.* Protein Science, 2018. **27**(9): p. 1636-1650.
188. Hemsworth, G.R., et al., *The copper active site of CBM33 polysaccharide oxygenases.* Journal of the American Chemical Society, 2013. **135**(16): p. 6069-6077.
189. Span, E., et al., *The role of the secondary coordination sphere in a fungal polysaccharide monooxygenase. ACS Chem Biol 12: 1095–1103*. 2017.
190. O'Dell, W.B., P.K. Agarwal, and F. Meilleur, *Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase.* Angewandte Chemie International Edition, 2017. **56**(3): p. 767-770.
191. Liu, T., et al., *Identification, Efficacy, and Stability Evaluation of Succinimide Modification With a High Abundance in the Framework Region of Golimumab.* Frontiers in Chemistry, 2022. **10**: p. 826923.
192. Frandsen, K.E., et al., *The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases.* Nature chemical biology, 2016. **12**(4): p. 298-303.
193. Zhou, H., et al., *Distinct interaction of lytic polysaccharide monooxygenase with cellulose revealed by computational and biochemical studies.* The Journal of Physical Chemistry Letters, 2020. **11**(10): p. 3987-3992.
194. Simmons, T.J., et al., *Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates.* Nature communications, 2017. **8**(1): p. 1064.
195. Frandsen, K.E., et al., *Identification of the molecular determinants driving the substrate specificity of fungal lytic polysaccharide monooxygenases (LPMOs).* Journal of Biological Chemistry, 2021. **296**.
196. Sun, P., et al., *Configuration of active site segments in lytic polysaccharide monooxygenases steers oxidative xyloglucan degradation.* Biotechnology for Biofuels, 2020. **13**: p. 1-19.