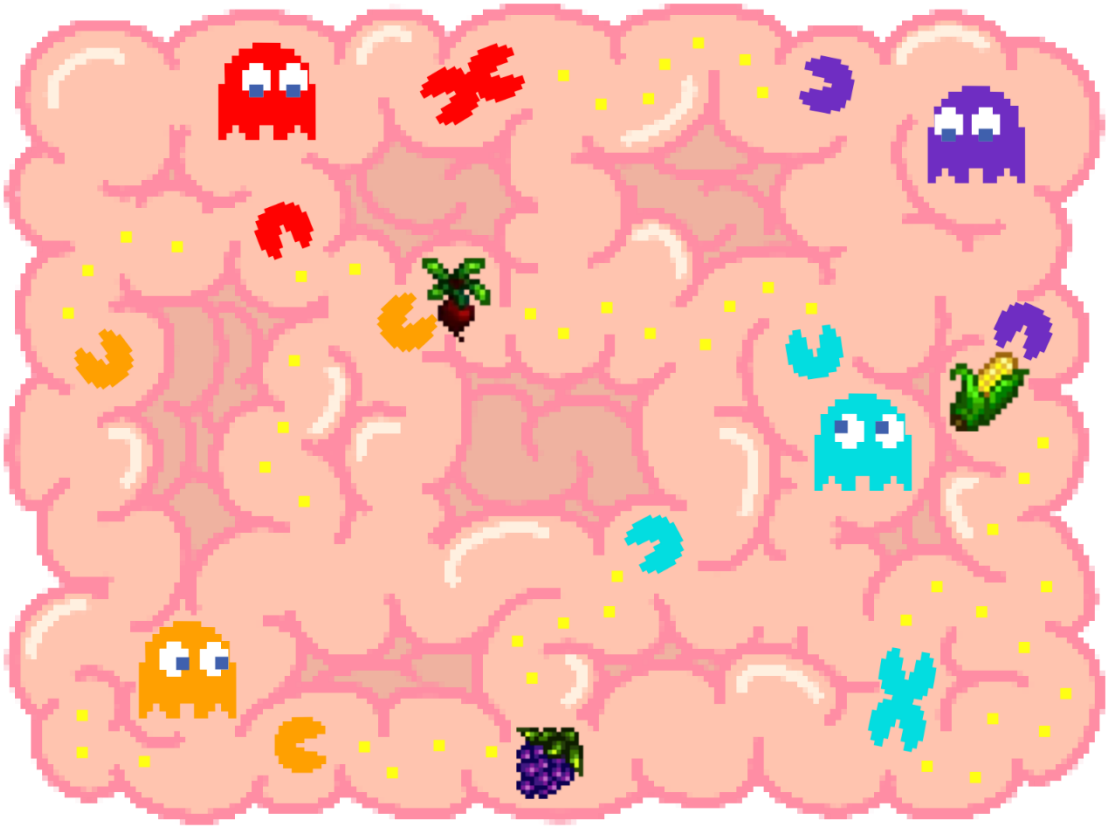
THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Exploring and exploiting plant biomass degradation by Bacteroidetes

Cathleen Kmezik



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
Chalmers University of Technology
Gothenburg, Sweden 2021

Exploring and exploiting plant biomass degradation by Bacteroidetes
Cathleen Kmezik

ISBN 978-91-7905-532-5

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4999
ISSN 0346-718X

Division of Industrial Biotechnology
Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: + 46 (0) 31 772 10 00

Cover: Bacteroidetes (ghosts) residing in the gastrointestinal tract produce enzymes (pac-mans) to hunt for simple (yellow circles) and complex (beet, corn, grape) plant biomass derived carbohydrates.
The artwork was inspired by the classic arcade game “Pac-Man” (Bandai Namco Entertainment). The beet, corn, and grape pixel art was taken from the game “Stardew Valley” (Eric Barone).

Printed by Chalmers Reproservice,
Gothenburg, Sweden 2021

The struggle is real, but the hustle is deep.
— Alyssa Edwards

Für Opa.

Preface

This dissertation serves as partial fulfillment of the requirements for obtaining the degree of Doctor of Philosophy at the Department of Biology and Biological Engineering at Chalmers University of Technology. The work was funded by the Swedish Research council (Dnr 2016-03931) and the Novo Nordisk Foundation (Grant number NNF17OC0027648). The PhD studies were carried out between February 2017 and August 2021 under the supervision of Assoc. Prof. Johan Larsbrink and co-supervision of Prof. Lisbeth Olsson. The thesis was examined by Prof. Carl Johan Franzén.

The majority of the work in this thesis was carried out at the Division of Industrial Biotechnology (Indbio) at Chalmers University of Technology. Nuclear magnetic resonance experiments were conducted at the Department of Chemistry and Chemical Engineering by Dr. Alexander Idström and Prof. Lars Evenäs. Mass spectrometry experiments were executed at Chalmers Mass Spectrometry Infrastructure (CMSI) by Marina Armeni and Dr. Otto Savolainen. Crystallization experiments were performed at Gothenburg University and X-ray diffraction data were collected at the MAX IV Laboratory by Dr. Scott Mazurkewich. Thin layer chromatography experiments were carried out by Dr. Lauren McKee at KTH Royal Institute of Technology.

Cathleen Kmezik
August, 2021

Exploring and exploiting plant biomass degradation by Bacteroidetes

Cathleen Kmezik

Division of Industrial Biotechnology
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

Bacteroidetes bacteria have evolved to become excellent biomass degraders. They achieved this by applying carbohydrate-active enzymes (CAZymes) and organizing genes connected to the degradation of specific polysaccharides into discrete gene cassettes, so-called polysaccharide utilization loci (PULs). Consequently, CAZymes and PULs may hold the potential to improve biomass valorization processes in biorefineries and to advance our understanding of human and livestock gut health.

CAZymes are extremely diverse in activity and structure, and for some enzyme families only little is known to date. For example, certain carbohydrate esterases (CEs) combine multiple catalytic domains within one protein, resulting in multicatalytic enzyme architectures, and the properties of these have been little explored. In this thesis, I present biochemical data showcasing the existence of intramolecular synergy between the active domains of multicatalytic CEs (*Bo*CE6-CE1). The observed intramolecular synergy facilitated more efficient degradation of xylan-rich biomass compared to non-multicatalytic CEs, giving a possible explanation as to why multicatalytic CEs exist in the genomes of Bacteroidetes species. Well-defined activity profiles of several here characterized CEs support the hypothesis that each catalytic domain fulfills an individual role during concerted plant biomass degradation, explaining why some PULs encode multiple CEs from the same enzyme family. Further, the investigated CEs cleaved xylan decorations and increased the activity of xylanase-mediated biomass degradation up to 20-fold (*Fj*CE6-CE1). During the investigation of the CAZyme repertoire of different species I also identified a remarkably active and promiscuously acting acetyl xylan esterase (*Dm*CE6A), as well as a rare enzyme architecture that may offer new insights into the multitude of interacting enzyme activities necessary to degrade plant biomass (*Be*CE15A-Rex8A).

PULs encode a plethora of CAZymes and have been shown to be vital for the glycan degradation abilities of Bacteroidetes species. However, the investigation of PULs is aggravated by their usually large size, which often limits the scope of genetic studies. In this thesis, I present a new method for the transfer of PULs between Bacteroidetes species, thus expanding the tools available for the identification and characterization of PULs and their components. The PUL transfer was demonstrated for a previously characterized mixed-linkage β-glucan utilization locus and conferred the ability to metabolize mixed-linkage β-glucan to the receptor strain.

Keywords: Bacteroidetes, polysaccharide utilization locus, PUL transfer, carbohydrate-active enzymes, multidomain enzymes, carbohydrate esterases, plant biomass degradation, xylan

List of publications

This thesis is based on the following papers:

I **Kmezik C**, Bonzom C, Olsson L, Mazurkewich S and Larsbrink J (2020). Multimodular fused acetyl-feruloyl esterases from soil and gut Bacteroidetes improve xylanase depolymerization of recalcitrant biomass. *Biotechnology for Biofuels*, 13, 1-14.
DOI: 10.1186/s13068-020-01698-9

II **Kmezik C**, Mazurkewich S, Meents T, McKee LS, Idström A, Armeni M, Savolainen O, Brändén G and Larsbrink J (2021). A polysaccharide utilization locus from the gut bacterium *Dysgonomonas mossii* encodes functionally distinct carbohydrate esterases. *Journal of Biological Chemistry*, 296, p.100500.
DOI: 10.1016/j.jbc.2021.100500

III **Kmezik C***, Krska D*, Mazurkewich S and Larsbrink J (2021). Characterization of a novel multidomain CE15-GH8 enzyme encoded by the human gut bacterium *Bacteroides eggerthii*. *Manuscript*

IV **Kmezik C***, Porter NT*, Pope PB, Koropatkin N, Martens E and Larsbrink J (2021). Enabling metabolism of mixed-linkage *β*-glucan in *Bacteroides thetaiotaomicron* by transfer of a polysaccharide utilization locus using a new versatile method. *Manuscript*

*These authors contributed equally.

Reprints were made with permission from the respective publisher.

Contribution summary

I First author. I planned and performed most of the experiments, analyzed the data, interpreted the results and wrote the manuscript.

II First author. I contributed in conceiving and designing the study. I performed the experimental work together with my co-authors. I analyzed the data, interpreted the results and wrote the manuscript.

III First author (shared). I performed the experimental work, analyzed the data and interpreted the results together with my co-authors. I wrote the manuscript.

IV First author (shared). I contributed in conceiving and designing the study. I performed the experimental work, analyzed the data and interpreted the results together with my co-authors. I wrote the manuscript.

Table of contents

Abbreviations

AX	arabinoxylan
AXE	acetyl xylan esterase
*Bo*XylS	small xylan utilization locus of *B. ovatus*
*Bo*XylL	large xylan utilization locus of *B. ovatus*
CAZy	carbohydrate-active enzymes database
CAZyme	carbohydrate-active enzyme
CBM	carbohydrate-binding module
CE	carbohydrate esterase
CE1	carbohydrate esterase family 1
CE6	carbohydrate esterase family 6
CFU	colony forming units
ChiUL	chitin utilization locus
DNSA	3,5-dinitrosalicylic acid
FAE	feruloyl esterase (ferulic acid esterase)
GAX	glucuronoarabinoxylan
GE	glucuronoyl esterase
GH	glycoside hydrolase
GH11	glycoside hydrolase family 11
gpPUL	Gram-positive polysaccharide utilization locus
GX	glucuronoxylan
HGM	human gut microbiota
HPAEC-PAD	high-performance anion-exchange chromatography with pulsed amperometric detection
LCC	lignin-carbohydrate complex
LPMO	lytic polysaccharide monooxygenase
MaltO	maltooligosaccharide
*β*ManUL	*β*-mannan utilization locus
MLG	mixed-linkage *β*-glucan
MLGUL	mixed-linkage *β*-glucan utilization locus
PorUL	porphyran utilization locus
PUL	polysaccharide utilization locus
PULDB	polysaccharide utilization locus database
Rex	reducing-end-xylose releasing *exo*-oligoxylanase
RS	reducing sugar
SCFA	short-chain fatty acid
SGBP	surface glycan binding protein
Sus	starch utilization system
TR	target region
XO	xylooligosaccharide
XyGUL	xyloglucan utilization locus

1. Introduction

Microorganisms, such as bacteria from the phylum Bacteroidetes, employ enzymes to degrade biomass in their natural habitat. Researching these enzymes and how they are utilized by Bacteroidetes advances our understanding of natural biomass degradation, which in turn may find application in biomass valorization processes and gut health. The overall aim of my PhD thesis was to explore the potential of the bacterial phylum Bacteroidetes for the degradation of plant biomass.

1.1. Structure of this thesis

This thesis consists of two parts. In the first part, I summarize my research results and put them in a larger context. First, I introduce the reader to microorganisms and their ability to enzymatically degrade biomass, as well as biorefineries in which plant biomass can be used as a feedstock (chapter 1). Then, I explain the structure of lignocellulose in general and xylan in detail and establish an understanding for the challenging recalcitrance of this substrate (chapter 2). I examine various enzyme activities that are important in degrading lignocellulose (chapter 3) and subsequently broaden the focus to discuss how microbes metabolize lignocellulose on the example of polysaccharide utilization loci (PULs) in Bacteroidetes (chapter 4). I also consider current limitations and give an outlook on future research (chapter 5). Lastly, I evaluate whether the aims of my thesis stated in chapter 1 section 7 have been addressed (chapter 6).

Chapters 1 to 4 each contain a summary at the end. Abbreviations are introduced only once in text, but a compilation of all abbreviations used can be found at the beginning of the thesis. If used in figure captions, abbreviations are explained.

The second part of this thesis consists of my scientific papers. Paper I discusses the biochemical properties of two multicatalytic carbohydrate esterases encoded by *Bacteroides ovatus* and *Flavobacterium johnsoniae*. The biochemical characterization of all carbohydrate esterases present in a specific PUL encoded by *Dysgonomonas mossii* is the topic of paper II. In paper III, the biochemical properties of a carbohydrate-active enzyme with an unusual multicatalytic enzyme architecture encoded by *Bacteroides eggerthii* are presented. *B. ovatus*, *F. johnsoniae*, and *D. mossii* are all members of the bacterial phylum of Bacteroidetes. In paper IV a new method for the transfer of PULs between Bacteroidetes species is described.

1.2. Microorganisms evolved carbohydrate-active enzymes to degrade organic matter

Life on Earth in the form of single-cell microorganisms has been shown to date back 3.5 billion years (1). During their long existence on Earth, microorganisms adapted to different habitats and in the process diversified into archaea, bacteria, and eukaryotes. Microorganisms influence the habitats they live in and are often vital for the balance of their respective ecosystems. For example, in soils microorganisms such as bacteria and fungi decompose organic matter making nutrients like nitrogen, phosphorus, and sulfur available for plants, thus closing the respective nutrient cycles (2,3). Enzymes are essential for the degradation abilities of microorganisms as they catalyze the various chemical reactions taking place during the decomposition of organic matter. With the advent of larger multicellular organisms, microorganisms started to also colonize habitats provided by these hosts, for example the gastrointestinal tract. Microorganisms inhabiting the gastrointestinal tract started to form their own ecosystems in which diversification led to the occurrence of thousands of species specialized on different nutrient sources and ecological niches. A prevalent bacterial phylum present in the gastrointestinal tract of mammals is Bacteroidetes (4,5), and many of its members have specialized on the degradation of diet or host-derived glycans by developing a variety of carbohydrate-active enzymes (CAZymes) (6). Gut bacteria like Bacteroidetes also excrete short-chain fatty acids which contribute to the hosts caloric intake and have health-promoting effects (7-11). As such, CAZymes are naturally relevant for gut health. CAZymes have also attracted interest for various industrial applications, for example for efficient degradation of organic matter in biorefineries.

1.3. Biorefineries are envisioned to facilitate a circular and biobased economy

To date, fossilized carbon resources such as petroleum oil, natural gas, and coal supply us with energy in the form of fuel, but also provide important precursors to the chemical industry for the production of many products such as plastics, fertilizers, and waxes (12). Unfortunately, fossilized carbon resources are limited and their consumption releases greenhouse gases which facilitate climate change. The limitations of fossilized carbon resources and the problems they create require a shift towards a circular economy that is based on sustainable and renewable resources.

To effectively replace fossilized carbon resources, both the energy providing fuels, and the materials/chemicals derived from fossilized carbon have to be substituted (13). Biorefineries are envisioned to fulfill this role by converting renewable and ideally sustainably grown biomass into energy carriers, materials, and chemicals (Fig. 1). While the overall principle is the same for all biorefineries, the individual

process steps depend on the composition of the biomass feedstock and the product portfolio.

Generally, plant biomass based biorefinery processes start with the pretreatment of the feed material, followed by enzymatic hydrolysis during which sugars are released. These sugars can be metabolized by microorganisms in a fermentation step to produce for example biofuel. The fermentation products usually undergo several downstream processing steps to reach desired concentration, purity or other process relevant properties.

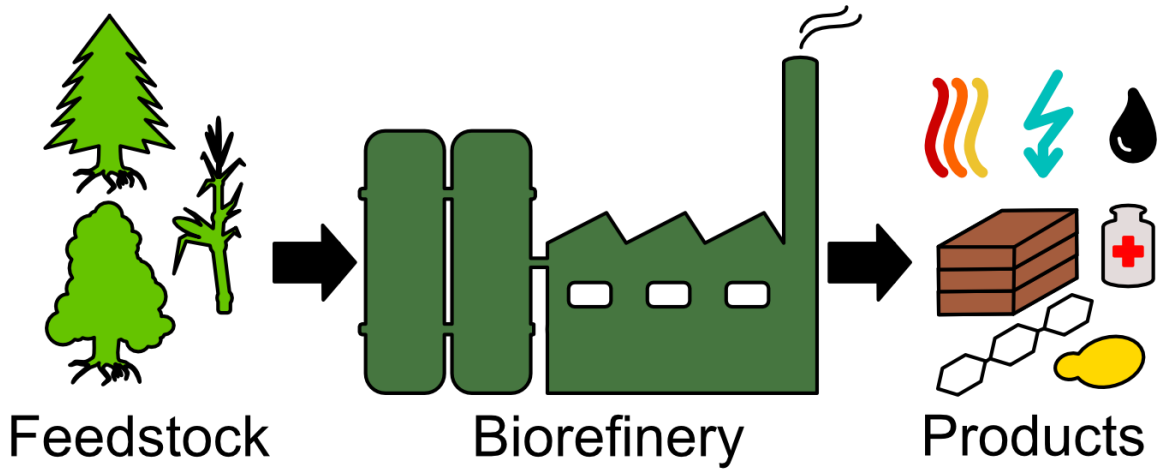


Figure 1: The biorefinery concept. Renewable and sustainably grown biomass serves as feedstock and is converted in the biorefinery into a multitude of products including energy carriers (heat, electricity, and biofuels) and chemicals/materials (platform chemicals, pharmaceuticals, polymers, and microbial cell mass).

Plant biomass based biorefineries produce a wide range of products in order to make efficient use of all components of the feedstock. Sugars are generally regarded as low-value products, but can be used for the microbial production of platform chemicals, pharmaceuticals, microbial cell mass (for example yeast for baking, protein for food and feed industry), and biofuels. Platform chemicals are a group of small molecules such as ethanol, glycerol, succinic acid, and lactic acid (14). They are the building blocks of many chemical synthesis processes and therefore of great importance for countless products of the chemical industry.

Out of the different energy carriers that can be produced in biorefineries biofuels are the closest to large scale implementation. Biofuels have been continuously developed and improved over the last few decades and are generally divided into generations, currently ranging from first to fourth generation biofuels. In 2019, 5% of worldwide energy consumption (29 exajoules) was covered through biofuels (15). Most of the biofuels produced to date are first generation biofuels. First generation biofuels are based on crops rich in sugars and starch, and grown on arable land (16,17). Second

generation biofuels are produced from forestry or agricultural biomass grown on marginal land or as byproduct of crop production on arable land (16,17). While far less second generation biofules are produced compared to first generation biofuels, intense research efforts are undertaken to make second generation biofuels more competitive. In 2015, the worldwide capacity of second generation biofuels reached 1,400 million liters compared to 85 billion liters of first generation biofuels (18). By 2018, global biofuel production reached 154 billion liters (19), demonstrating a rapid increase in production.

Selective processing in biorefineries can allow for the preservation of the polymer character of the biomass substrate, enabling the production of a variety of more complex products such as composite materials, carbon fibers, dispersants, and activated carbon (20-24). There are three major types of polymers present in plant biomass: cellulose, hemicellulose, and lignin. In reference to these three polymers, plant biomass is also called lignocellulose (25). The composition of lignocellulose is discussed in more detail in chapter 2.

1.4. Plant biomass is a renewable feedstock for biorefineries

Biomass, or in other words organic matter sourced from plants or animals, is the largest source of renewable carbon available to us (26). Our ability to make efficient use of this resource will be instrumental in the establishment of a circular and biobased economy. To date, the most commonly used feedstocks for biorefineries include corn, sugarcane, and sugar beet (27). These crops are high in simple sugars and more complex carbohydrates, a beneficial trait for biorefinery feedstocks, but they are also a food source and grown on arable land. Plants that are not used as food sources such as poplar, switchgrass, and elephant grass are attractive alternative candidates (28,29). In general, plants or crops used for biorefinery feedstocks should be fast growing, have short rotation times, and little need for fertilizers and artificial irrigation (26). Growing crops and plants specifically for biorefineries is highly efficient and delivers a relatively uniform feedstock. Still, arable land is used for their cultivation. To circumvent high usage of arable land many biorefinery concepts are designed to instead utilize agricultural and forestry waste as feedstock. Such waste streams are abundant, for example estimated 1.5×10^9 tons of crop waste are produced each year worldwide (30). However, the composition of these waste streams can be very diverse and include different plant species and plant tissues, which makes efficient processing a challenge. As a result, there is an ongoing need for developing new or modified biorefinery concepts and improving process steps. This is especially true regarding the breakdown of the biomass feedstock into building blocks (generally sugar monomers) that can be further converted into the desired products.

The breakdown of biomass can be achieved by physio-chemical means, through enzymatic hydrolysis, or a combination thereof. Generally, enzymatic hydrolysis allows for milder process conditions, produces less side products and toxic waste streams, and is more specific. However, the utilization of enzymes also poses challenges. Enzymes are pH, temperature, pressure, and solvent sensitive. Further, the production of enzymes can be costly and they are often identified as one of the key bottlenecks in lignocellulosic biorefineries (31). Ideally, enzymes applied in lignocellulosic biorefineries should be produced in high yields, be storable, be highly active under the chosen process conditions, retain activity for the duration of the whole hydrolysis process, and be reusable (for example by separation from product stream and cycling back into the hydrolysis step). Finding enzymes fulfilling all these requirements is challenging, but thankfully nature offers a vast repertoire of enzymes that can be mined to find suitable candidates and advance our general understanding of enzymatic biomass degradation processes.

In this thesis, I investigated and biochemically characterized several CAZymes exhibiting activity on lignocellulosic material with interesting but understudied properties. The CAZymes belonged to the enzyme class of carbohydrate esterases (CEs) and were sourced from PULs encoded in the genomes of Bacteroidetes species. Due to their natural habitat these bacteria have evolved highly efficient biomass degrading enzymes. While the initial focus of my research was to contribute to the understanding and application of such enzymes for example in biorefineries, the work with gut bacteria also entailed the potential of gaining insights that could benefit mammalian gut health.

1.5. Aims of the thesis

The overall aim of my thesis was to explore the potential of the bacterial phylum Bacteroidetes for the degradation of plant biomass. As Bacteroidetes is a large bacterial phylum the study was limited to a handful of species. Further, I define the biomass degradation potential of Bacteroidetes here as their CAZyme and PUL repertoire. While various enzymes with hydrolytic activity on plant biomass, such as corn cob and beechwood, were investigated in this thesis, I focused on enzymes belonging to carbohydrate esterase families 1 (CE1) and carbohydrate esterase family 6 (CE6). My overall aim can be divided into more specific aims which are discussed below.

PULs are gene clusters encoding the proteins needed to deconstruct specific polysaccharides. The xylan targeting PULs of Bacteroidetes encode many CAZymes targeting the xylan backbone, but only few CEs which often belong to CE1 and CE6. CE1 and CE6 members are little explored with regards to their role within the PUL. The first aim was to investigate the role of PUL encoded CEs.

Generally, CEs are less abundant than for example xylanases in PULs. Some PULs encode only a single CE or none while others encode multiple CE enzymes. The second aim was to investigate the occurrence of multiple CEs within the same PUL.

An interesting feature of some PUL encoded CEs is that they consist of multiple catalytic domains. Multicatalytic enzymes are in general little studied and multicatalytic CEs had not been studied at all prior to publication of **paper I**. The third aim was to investigate PUL encoded multicatalytic CEs and find a possible explanation for this enzyme architecture.

Until this point the specific aims focused on the action of CAZymes, which play a central role in microbial biomass degradation. In Bacteroidetes, the numerous CAZymes necessary to degrade a polysaccharide are often found organized into PULs. PUL research is frequently aggravated by the large size of PULs. The fourth aim of this thesis was to develop a method to transfer PULs between different species to expand the available genetic toolbox. In connection with this the fifth aim was to prove that the sole transfer of a PUL also transfers the degradation ability conferred by the respective PUL.

1.6. Main points of chapter 1

- Microorganisms such as Bacteroidetes evolved carbohydrate-active enzymes that can degrade lignocellulose.
- Biorefineries enable production of energy carriers, chemicals, and materials from renewable feedstocks such as lignocellulosic material.
- The enzymatic hydrolysis of lignocellulose is a bottleneck in biorefineries.
- The main aim of this thesis was to explore the potential of the bacterial phylum Bacteroidetes regarding their encoded carbohydrate-active enzymes and polysaccharide utilization loci for the degradation of plant biomass.

2. Plant biomass is composed of a network of heterogenous polymers

Plant biomass is a renewable and abundant resource that is composed of lignocellulose. Lignocellulose is available as agricultural waste (for example hulls and straw from rice, corn, wheat, sugar cane, switchgrass, sorghum) and in the form of wood chips and other forestry residues (for example from birch, poplar, spruce, pine), making it a promising feedstock for biorefineries and second generation biofuel production (32). As mentioned earlier, lignocellulose or lignocellulosic material are umbrella terms for several plant biomass derived polymers. These carbon-based biopolymers vary substantially in their composition, properties, and abundance and this multi-level heterogeneity contributes to lignocellulose being a challenging substrate for biorefineries (25,32).

In this chapter, I introduce terminology frequently used when discussing lignocellulose. Further, the structure of lignocellulose is discussed on different levels including plant physiology (structure of plant cell walls) and chemical composition in different species (hardwoods, softwoods, and grasses). Lastly, the main polymers present in lignocellulose (cellulose, hemicellulose, and lignin) are described with a focus on the hemicellulose xylan.

2.1. Lignocellulose is highly variable and recalcitrant towards degradation

During their evolution plants have adapted to inhabit vastly different environments and are as a result incredibly diverse in their physiology. Classification systems for plants mirror this complexity of physiological traits and group species based on for example whether a species is vascular or non-vascular, seed- or spore producing, and flowering or non-flowering. Notably, all to date identified plants with potential for biofuel production are vascular except for algae. Describing a comprehensive classification of plants is beyond the scope of this thesis, but an overview of commonly used terms, when discussing lignocellulose as a resource, is given in Table 1. Many of the described terms are used throughout this thesis.

The most frequently used classification when focusing on the enzymatic degradation of plant biomass is the division into hardwoods, softwoods, and grasses. Although this classification does not originate from chemical composition, members belonging to the same group often share general trends, for example hardwoods contain relatively high amounts of xylan but little mannan, while softwoods generally contain more mannan than xylan (33). Still, the chemical composition of lignocellulosic material also depends on other factors such as species and in some cases season.

Moreover, when describing soft- and hardwoods, mechanical pressure can also influence physiology leading to the formation of compression and tension wood (34). Compression wood is usually high in lignin and low in cellulose and galactoglucomannan content. Tension wood generally contains higher amounts of cellulose than "normal" wood (35).

Table 1: Frequently used terminology describing lignocellulosic material in the context of biorefineries.

Terminology	Description
Woody vs herbaceous plants	Woody plants have hard stems that survive above ground in winter. Herbaceous plants have soft stems that die back, typically during winter. Herbaceous plants can be annual, biennial, and perennial.
Annual, biennial, and perennial plants	Plants with life cycles spanning one year, two years and more than two years, respectively.
Angiosperms vs gymnosperms	While both are seed-bearing plants, angiosperms (also called flowering plants) form their seeds in the ovaries of their flowers. Gymnosperms (also called non-flowering plants) develop their seed in unisexual cones.
Coniferous, deciduous, and evergreen plants	Coniferous plants are a subset of gymnosperms and have needle- or scale-like leaves. Deciduous plants shed their leaves/needles, opposite to evergreens that keep their leaves/needles all year around.
Monocot vs dicots	Monocots and dicots are seed-bearing plants, but monocots have one cotyledon (embryonic leaf) and dicots have two cotyledons.
Hardwood	Hardwood is wood from dicot trees. Examples are walnut, poplar, birch, and beech.
Softwood	Softwood is wood from coniferous trees. Examples are pine, cedar, and spruce.
Grasses	Grasses are monocot flowering plants. Examples are corn, rice, wheat, and switchgrass.

Plant cell walls make up the majority of biomass present in plants (36) and are composed of primary and secondary cell wall layers. The first layer, the primary cell wall (P1), determines the shape and size of the cell and borders the middle lamella which connects neighboring cells (Fig. 2). The secondary cell wall is typically composed of three secondary cell wall layers (S1-S3) and provides strength and rigidity. The S3 layer is connected to the plasma membrane. The cell wall layers vary in their chemical composition and their thickness. The S2 layer accounts on average for 80% of the cell wall biomass and is therefore the layer with the highest industrial relevance (37,38).

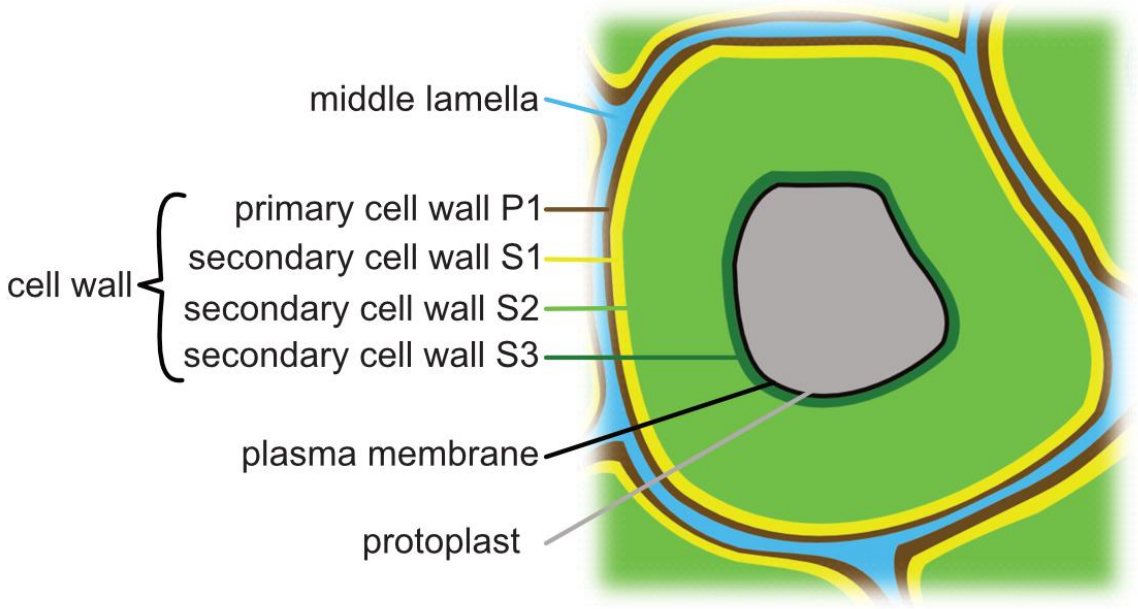


Figure 2: Simplified representation of the plant cell wall anatomy. The middle lamella connects neighboring cells. The primary cell wall (P1) develops first and determines the size and shape of the growing plant cell. Secondary cell wall layers (S1-S3) are developed later in woody plant tissue and supply rigidity, strength, and protection (39).

2.2. The main building blocks of lignocellulose are cellulose, hemicellulose, and lignin

The main constituents of secondary plant cell walls are cellulose, hemicellulose, and lignin. Cellulose is often organized into microfibrils that are embedded into a matrix of hemicellulose and lignin (40). The composition of lignocellulose from beech, spruce, and corn cob is shown as examples for hardwood, softwood, and grasses, respectively, in Fig. 3 (41).

The composition of lignocellulose varies with species but follows a few general trends. Firstly, cellulose is usually the most abundant polymer in lignocellulose. Secondly, the lignin content is generally highest in softwoods, followed by hardwoods and grasses. Thirdly, lignin content varies widely with species, especially in grasses. For instance, corn cob and tobacco leaf have relatively low lignin contents (15% each), while wheat straw and switchgrass have lignin contents similar to that of the softwood species spruce (21%) (41,42).

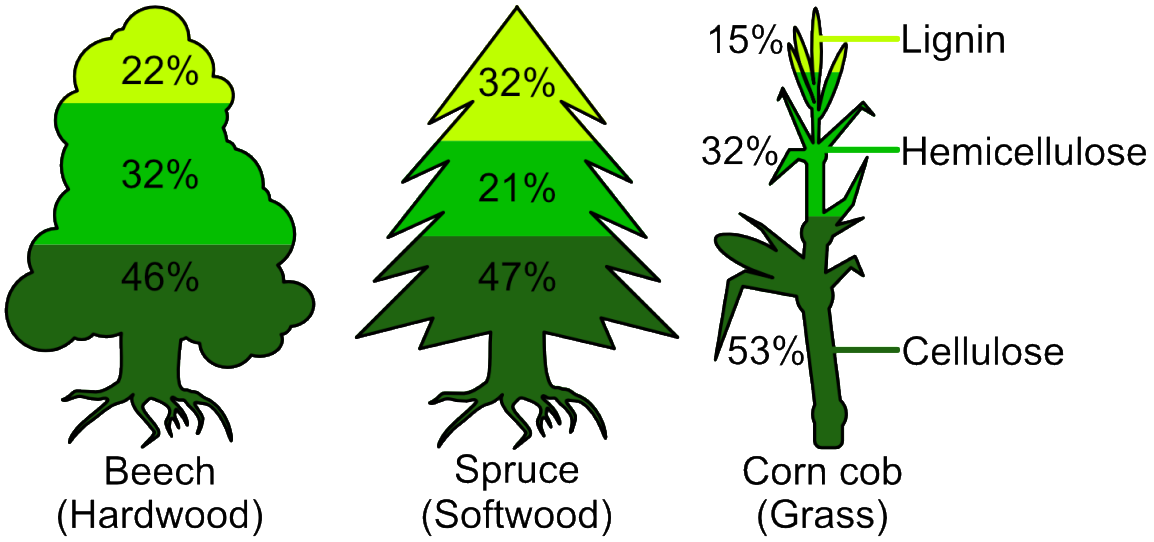


Figure 3: Cellulose, hemicellulose and lignin composition of beech, spruce, and corn cob. Presented values are based on Demirbaş *et al.* (41). While lignocellulose composition varies with species, the lignin content of softwoods is usually higher than in hardwoods and grasses. Cellulose is generally the most abundant polymer in lignocellulose.

Cellulose

Cellulose is a linear polysaccharide consisting of glucose units that are linked via *β*-1,4-glycosidic linkages. Glycosidic bonds are covalent linkages that join the anomeric carbon in carbohydrates to other compounds, such as hydroxyl groups of other carbohydrates or alcohols. The smallest repeating unit of cellulose is cellobiose, which is composed of two glucose units with the second glucose moiety axially rotated by 180°. The degree of polymerization depends on the species and the treatment method used for extraction (43). Plant biomass is often pretreated to facilitate subsequent enzymatic hydrolysis. For example, the degree of polymerization of cellulose in corn stover can range from 7,000 repeating units for ammonia fiber expansion pretreatment to 2,700 repeating units for pretreatment with dilute acid (44,45).

The glucose units of neighboring cellulose polymers can form hydrogen bonds between their hydroxyl groups, resulting in the formation of microfibrils, which provide strength to the plant cell wall. Regions in microfibrils can either have a high or low degree of order, resulting in crystalline or amorphous regions, respectively (33,46). While highly crystalline regions are especially challenging for enzymatic breakdown, enzymes called lytic polysaccharide monooxygenases (LPMOs) can target these ordered regions and break the cellulose chains (47).

Hemicellulose

The term hemicellulose describes a range of different heteropolymers, which can vary substantially in chemical composition and type of glycosidic bonds. The hemicellulose backbone can consist of different monosaccharides, such as xylose, glucose, mannose, arabinose, rhamnose, glucuronic acid, fucose, and galactose (48). Hemicelluloses are usually named after the most abundant monosaccharide building blocks present. For instance, xyloglucan consists mainly of xylose and glucose. While highly diverse, hemicelluloses can be divided into four groups (24):

- **Xylans:** Xylans are the most common hemicelluloses (49) and the third most abundant polymer on earth after cellulose and chitin (50,51). In this thesis, xylan rich biomass in the form of corn cob and Japanese beechwood was used in enzymatic assays to determine the activity of different CEs (**papers I - III**), and a more detailed description of xylan can be found in chapter 2 section 3.
- **Mannans:** The backbone of mannans consists either of β-1,4-linked mannose (galactomannans, linear mannan) or both mannose and glucose moieties (glucomannans). While linear mannan (from ivory nut) has no decorations, galactose units are linked to the mannose moieties of the galactomannan and glucomannan backbone via α-1,6 glycosidic bonds. The amount of galactose present governs solubility and its distribution pattern influences properties like viscosity (24,52). Galactoglucomannans combine structural motifs of galactomannans and glucomannans. They are the main hemicellulose present in the secondary cell wall of softwoods and can be found in lower abundance in the primary cell wall of hardwoods and grasses (49). Galactoglucomannan can be acetylated (53).
- **Xyloglucans:** The backbone of xyloglucan consists, similar to cellulose, of β-1,4-linked glucose units. Contrary to cellulose, the glucan backbone in xyloglucan has repeating patterns of α-1,6 linked xylose moieties attached to it, typically resulting in 3 out of 4 glucose moieties of the backbone being linked to xylose. Additional sugars can be attached to these xylose units, and their identity depends on the xyloglucan source and influences the physiochemical properties of the polymer. For instance, in tamarind seeds galactose is linked via β-1,2 glycosidic bonds to xylose moieties giving tamarind xyloglucan its gelling properties (54,55). Xyloglucan is found mainly in the primary cell wall and in some seeds (56,57).
- **Mixed-linkage β-glucans (MLGs):** Similar to cellulose the backbone of MLG consists of β-1,4 linked glucose units, but unlike cellulose also contains β-1,3 glycosidic bonds resulting in a non-linear structure (24). The ratio of β-1,3 and β-1,4 linkages varies. For barley very different β-1,3: β-1,4 ratios have been reported (58,59). MLG is most abundant in grasses such as cereals (60) and does not form branches (24)

As mentioned the abundance of these different kinds of hemicellulose depends on species and tissue. Within the lignocellulose matrix, hemicelluloses form hydrogen

bonds with cellulose (61), surround cellulose microfibrils, and can be covalently linked to lignin (48), thus strengthening the plant cell wall.

Lignin

Lignin is an aromatic polymer consisting of relatively randomly distributed units of monolignols, covalently linked through radical coupling mechanisms. The three most abundant monolignols present in lignin are *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol. Coniferyl aldehyde is predominant in softwood (95%), while hardwood usually contains a mixture of coniferyl aldehyde and sinapyl aldehyde (26). In grasses sinapyl, coniferyl, and *p*-coumaryl aldehyde are most abundant (62). A multitude of linkages are present in lignin, but in softwoods and hardwoods, the *β-O*-4 ether bond is predominant (26). Overall, lignin forms a three-dimensional network around hemicelluloses and cellulose that contributes to cell wall rigidity and provides water-resistance as well as protection against microbial degradation to the cell wall (63).

Lignin can be covalently linked to hemicelluloses and cellulose forming so-called lignin-carbohydrate complexes (LCC). So far, eight different types of lignin-carbohydrate bonds have been identified (64). In herbaceous plants ferulate and diferulate esters forming linkages to lignin are most abundant (65). To date, not much is known about the occurrence of ferulate linkages in woody biomass, but their presence has been reported in the bark of softwoods (64,66,67). The hemicellulose glucuronoarabinoxylan is known to form LCCs via ferulic acid and glucuronic acid groups (68).

Other components of lignocellulose

Besides cellulose, hemicellulose, and lignin, other components are also present in lignocellulose, but in much lower abundance. Such components are pectins, lipids, proteins, extractives, and ash (26,32). Pectins are typically charged polysaccharides enriched in α-1,4-linked galacturonic acid units, but also contain other linkages and monosaccharide types. Pectins function as hydrating agents and “glue” in the lignocellulose network (69,70). Extractives are compounds that can be extracted from lignocellulose by various polar or non-polar solvents and include terpenes, resins, and tannins (26). Extractives have been shown to possess antimicrobial activity and likely negatively impact the microbial degradation of lignocellulose (71).

2.3. The three main classes of xylan are glucuronoxylan, arabinoxylan, and glucuronoarabinoxylan

Xylan is extremely diverse in sugar composition and glycosidic linkage types. Complete enzymatic hydrolysis of xylan consequently requires a host of different enzyme activities. The woody tissue of dicots and the lignified tissue of monocots may contain 25-35% xylan, with seeds and cereal grains reaching up to 50% (72). The backbone of xylan consists of *β*-1,4-linked xylose units which are decorated with different carbohydrate and non-carbohydrate groups. These decorations determine the three major classes of xylan: glucuronoxylan (GX), arabinoxylan (AX), and glucuronoarabinoxylan (GAX; Fig. 4).

In addition to the xylan backbone, GX contains 4-*O*-methyl D-glucuronic acid units that are found *α*-1,2-linked to the xylose units in the main chain (73,74). AX contains L-arabinofuranose moieties that are attached via *α*-1,2- and *α*-1,3-linkages to the xylan backbone (75,76). GAX is the most complex class of xylan and includes several decorations on the backbone such as *α*-1,2-linked 4-*O*-methyl D-glucuronic acid, *α*-1,3-linked L-arabinose, and *α*-1,2-, *α*-1,3- and *β*-1,3-linked xylose. Arabinose moieties in GAX can be further linked to D-xylose and L-galactose or via ferulic acid to other GAX chains or lignin (68,77,78). LCCs can also be formed between glucuronic acid and lignin (79). The numerous different linkages present in GAX make it an especially challenging substrate for degradation processes involving enzymatic hydrolysis.

Most polymers in the plant cell wall stick together non-covalently and covalent bonds between for example hemicellulose and lignin are comparatively rare. Still, the previously described cross-linking in LCCs, for example via feruloyl groups and glucuronic acid, make such covalent bonds important targets for enzymatic degradation of lignocellulose for disentangling of the complex lignocellulose matrix. Enzymes with the potential of cleaving ferulate esters and a putative glucuronoyl esterase possibly targeting glucuronic acid-lignin linkages were investiged in **papers I – III** and chapter 3.

Another type of non-carbohydrate xylan decorations is acetylation. Acetyl groups can impede the action of xylan degrading enzymes and are common in GX and GAX at the C2 and/or C3 position of xylose. In GX 70-80% of xylose units can be *O*-acetylated (80). GAX from bagasse and wheat have been shown to contain acetyl groups on 30 % and 10 % of their xylose units, respectively. In these two lignocellulose sources the C3 position was more often acetylated than the C2 position, while simultaneous acetylation on both positions (C2 and C3) was quite rare (81). Enzymes acting on acetylated xylan are described in more detail in chapter 3 and were also investigated in **papers I & II**.

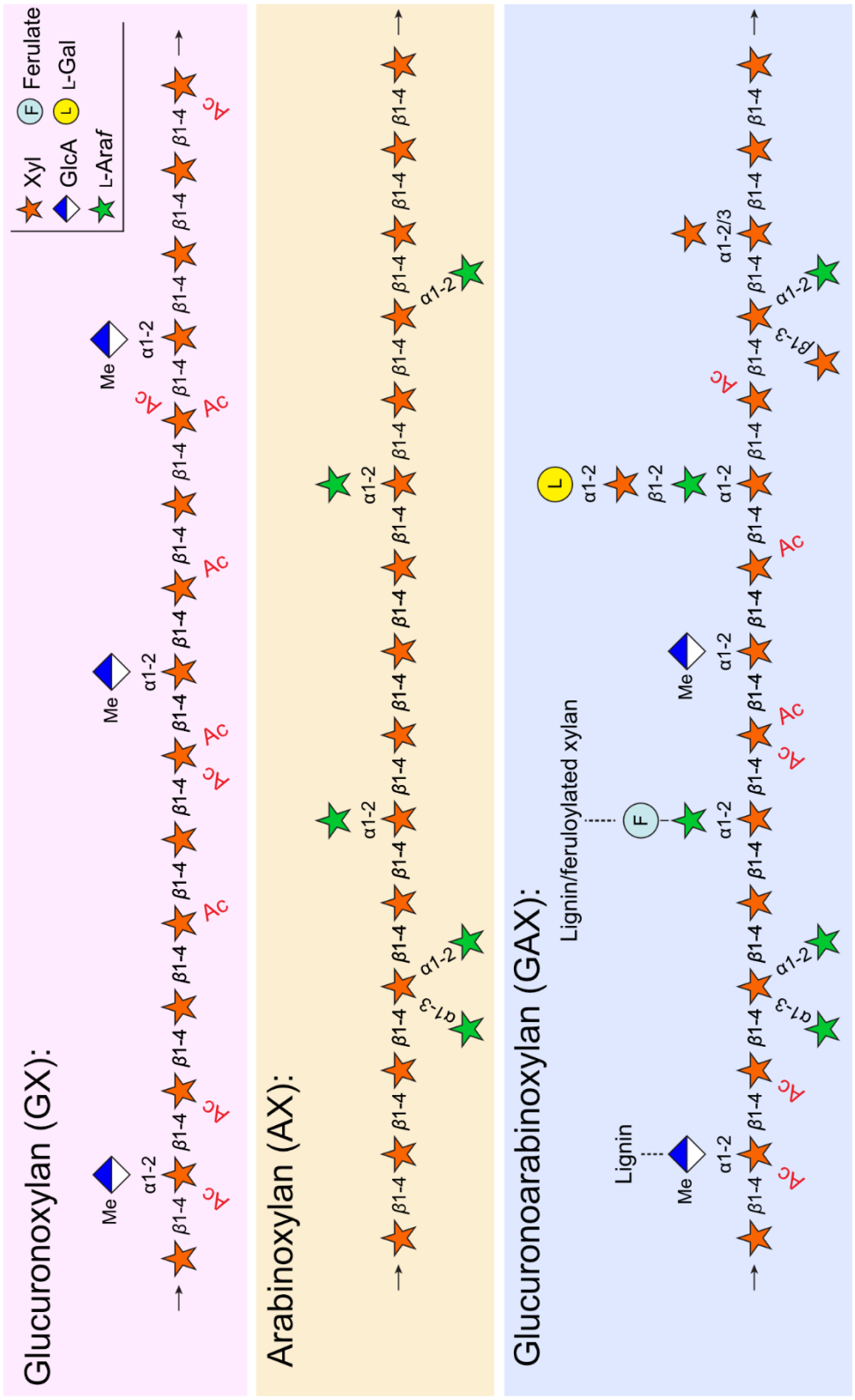


Figure 4: The three main classes of xylan. The figure was adapted from Rogowski *et al.* (104). Glucuronoxylan (GX), arabinoxylan (AX), and glucuronoarabinoxylan (GAX) have the *β*-1,4-linked xylose backbone in common, but carbohydrate and non-carbohydrate decorations differ. Acetyl groups are common on the C2 and C3 positions of xylose units in GX and GAX and symbolized by "Ac" (red). AX is not known to be acetylated in softwood, but might be acetylated in other species.

Besides GX, AX, and GAX, two additional terms for xylan are sometimes used: homoxylans and heteroxylans. Homoxylans are non-branched polysaccharides with a backbone consisting of only *β*-1,3 or *β*-1,4 linked xylose units or both, and are found in red and green algae in which they substitute cellulose in the cell wall (24). Heteroxylans contain the for xylan typical *β*-1,4 linked xylose backbone and are heavily substituted with oligosaccharide side-chains making them highly viscous. Heteroxylans have been isolated from cereal bran, seeds, and mucilages (49).

2.4. Xylan is an abundant but challenging substrate

Xylan is an attractive substrate for saccharification processes and polymer production due to its abundance. Xylan sources are available in the form of forestry products (hardwood, softwood), forestry waste (shrubs, branch tips, foliage), and crop side products or waste streams. Examples for crop side products rich in xylan are wheat straw (82), sunflower hulls (83), corn cob (84), olive pulp (85), and corn hulls/bran (86). The abundance of the different classes of xylan varies in these sources, but in general GAX is the most abundant xylan class in grasses, while GX is the most abundant class in hardwood. In softwoods xylan only constitutes 5-10% of dry biomass with the most abundant xylan being arabino-4-*O*-methyl glucuronoxylan (similar to GAX but with less arabinose). The main hemicellulose in softwood is acetylated galactoglucomannan with amounts ranging between 20 and 25 % (33,87,88).

Still, for the efficient utilization of xylan major obstacles have to be overcome as xylan and in a broader sense lignocellulose are very difficult substrates to degrade. This recalcitrance towards hydrolysis is rooted in the layered structure of the plant cell wall and the complex polymer network forming these layers. The entanglement of cellulose, hemicellulose, and lignin and the diversity within these polymers in composition, linkages, and decorations demand the action of many enzymatic activities. Enzymes active on xylan are discussed in the next chapter.

2.5. Main points of chapter 2

- Lignocellulose is composed of various heterogenous polymers forming a complex matrix.
- The heterogeneity of its components and their physical arrangement contribute to the recalcitrance of lignocellulose towards degradation.
- The main building blocks of lignocellulose are cellulose, hemicellulose, and lignin and their abundances vary among species.
- Hemicellulose is an umbrella term for several heteropolymers of which xylan is the most abundant in many industrially relevant species.
- Based on the presence or absence of certain carbohydrate and non-carbohydrate decorations on the xylan backbone, three classes of xylan are distinguished: glucuronoxylan, arabinoxylan, and glucuronoarabinoxylan.
- Due to its various decorations, glucuronoarabinoxylan is the xylan class that is the most challenging to enzymatically degrade.

3. The enzymatic degradation of xylan requires a multitude of specialized carbohydrate-active enzymes

As established in the previous chapter, xylan has an highly complex structure and consists of a plethora of carbohydrate and non-carbohydrate moieties that are connected via diverse linkages. This structural complexity of xylan aggravates its enzymatic degradation in several ways. For one, motifs or linkages might be physically obstructed by other structural motifs, thus limiting enzyme access. Secondly, the variety of linkages and chemical groups within the xylan structure requires a multitude of enzymes with specialized activities, which further have to act in concert to achieve full degradation.

Enzymes acting on xylan, and more widely on carbohydrates in general, are called CAZymes and are discussed in this chapter. The structural diversity of carbohydrates is mirrored in the diversity of CAZymes. After giving a brief overview over CAZymes, I discuss the action of selected GHs as these enzymes work in conjunction with CEs, which are the focal point of this chapter. CEs are themselves very diverse and I focus on CEs that were characterized as part of this thesis (belonging to CE1, CE6, and CE15). Many of the here discussed CEs are comprised of multiple catalytic domains, an enzymatic architecture referred to as “multicatalytic”, which has previously been little described in literature (Fig. 5). The investigated CEs were also encoded in PULs, a factor which is discussed in chapter 4.

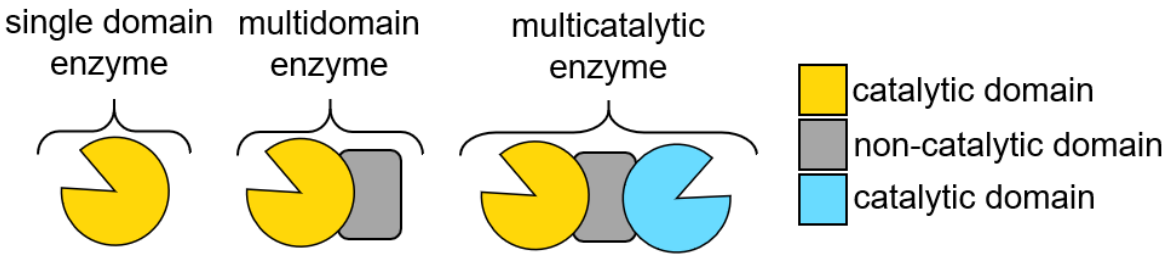


Figure 5: Enzyme architecture of multicatalytic enzymes. Multicatalytic enzymes are composed of multiple domains with distinct catalytic functions, whereas multidomain enzymes only harbor one catalytic domain but have additional non-catalytic domains. The different domains in multicatalytic and multidomain enzymes can be connected by linker regions (not shown here). Multicatalytic CEs are discussed in detail in chapter 3 section 7.

3.1. Carbohydrate-active enzymes are numerous and highly diverse

As their name implies, CAZymes act on carbohydrates and are as such crucial in both the enzymatic build-up and degradation of polysaccharides, which are for example found in lignocellulose. Accordingly, the investigation of CAZymes has implications for processes in which lignocellulose is enzymatically degraded, including industrial biomass valorisation processes, gut health, and carbon cycling in ecosystems.

Due to their great importance thousands of CAZymes have been researched to date and it has been shown that these enzymes are highly specific and highly diverse. CAZyme data is collected in various online databases. For my thesis I often utilized the Carbohydrate-Active enZYmes database (CAZy; www.cazy.org; (89)). CAZy collects genomic and biochemical data of CAZymes and is linked to other databases such as the Protein Data Bank (PDB; https://www.rcsb.org; (90)), in which structural data are collected. Contrary to the enzyme commission number system, where classification is based on substrate specificity, enzymes listed on CAZy are classified based on amino acid sequence similarity. Generally, high sequence similarity implies similar protein folds and similar enzyme activity (89). In other words, enzymes with high sequence similarity often target the same or similar bonds. This allows for the activity prediction of unknown enzymes based on sequence similarity to previously biochemically characterized enzymes. Due to recent advances in genome sequencing, sequencing data is much more readily available than biochemical characterization data of enzymes. As a result, CAZy offers the classification of thousands of CAZymes into classes, families, and subfamilies without requiring in-depth biochemical characterization of each of them (Table 2). Still, biochemical characterization of enzymes remains crucial, especially for enzyme families that are not well studied, such as the CE families investigated in this thesis.

Table 2: The five classes of CAZymes in CAZy.

CAZyme class	Activity	Enzyme families within the CAZyme class
Glycoside hydrolases*	Hydrolyze glycosidic bonds.	171
Glycosyltransferases	Form glycosidic bonds.	114
Polysaccharide lyases*	Cleave glycosidic bonds non-hydrolytically through β-elimination.	41
Carbohydrate esterases	Hydrolyze carbohydrate ester linkages.	19
Auxiliary activities	Catalyze redox reactions on carbohydrates.	17

The presented information was obtained from CAZy (accessed July 2021; (89)). All CAZyme classes are ordered into families based on sequence similarity. CAZyme classes marked with an asterisk additionally contain families divided into subfamilies in CAZy.

3.2. Glycoside hydrolases target glycosidic bonds in xylan

Glycoside hydrolases (GHs) cleave glycosidic linkages and act on xylan on both the sugar backbone and sugar side chains. Depending on their activity, GHs cleave either between carbohydrate moieties or between carbohydrate moieties and non-carbohydrate decorations (91). The diversity of these linkages is mirrored in the diversity of GHs with 171 GH families listed in CAZy to date, many of which are polyspecific ((89); accessed July 2021).

An example of a polyspecific GH family is family 11 (GH11). Members of GH11 have been shown to possess *endo-β*-1,4-xylanase or *exo-β*-1,4-xylosidase activity (89). In my thesis work, I used Xyn11A, a commercially available GH11 *endo-β*-1,4-xylanase from the anaerobic fungus *Neocallimastix patriciarum* in xylanase boosting experiments (**papers I - III**). The boosting experiments served as a measurement for CE activity and are discussed in more detail in chapter 3 section 5. As typical for *endo*-acting enzymes, Xyn11A cleaves linkages "within" the polysaccharide backbone and produces oligosaccharides of varying lengths as a result. In contrast, *exo*-acting enzymes such as *exo-β*-1,4-xylosidases attack linkages at the non-reducing end of the polysaccharide, releasing monosaccharides in a successive manner (92,93).

Reducing-end-xylose releasing *exo*-oligoxylanase (Rex) enzymes are another example for *exo*-acting GHs. Rex enzymes target linkages at the reducing end of xylooligosaccharides (XOs) while showing no or minimal activity on xylan (94). Only a handful of enzymes with Rex activity have been discovered and described to date and all of them have been categorized into GH8 ((94-96); **paper III**). In this thesis, I investigated a multicatalytic Rex enzyme encoded by *B. eggerthii* (*Be*Rex8A; **paper III**). The supplementation of Xyn11A with *Be*Rex8A led to increased xylan hydrolysis of GAX rich corn cob biomass, likely due to the complementary action of *endo*- (Xyn11A) and *exo*-acting (*Be*Rex8A) enzymes (Fig. 6).

Interestingly, xylanases and Rex enzymes are very similar in amino acid sequence to the extent that Rex form a branch within the xylanase clade of characterized GH8 members. Both *endo*-xylanases and Rex enzymes possess the same glutamate-aspartate-aspartate catalytic triad ((96); **paper III**). It has been hypothesized that the differing substrate specificities of xylanases and Rex are the result of obstruction of the active site cleft in Rex ((97); **paper III**). In *Pb*Rex8A, a Rex encoded by *Paenibacillus barcinonensis* BP-23, a Leu320-His321-Pro322 loop close to the active site was suggested to be responsible for the enzyme's preference for XOs (98). Similarly, in *Be*Rex8A an arginine residue was hypothesized to obstruct part of the active domain, possibly restricting access to smaller XOs with no decorations or side groups (**paper III**). However, when creating enzyme variants in which the loop size of *Pb*Rex8A was reduced and the arginine residue in *Be*Rex8A was replaced by a smaller alanine residue, no activity was observed on xylan akin to the wild type versions ((98); **paper III**). It remains unclear if all Rex share a specific structural motif that is responsible for activity on XOs and inactivity towards xylan. The

biochemical characterization, and especially structural elucidation, of more Rex enzymes could help solve this puzzle.

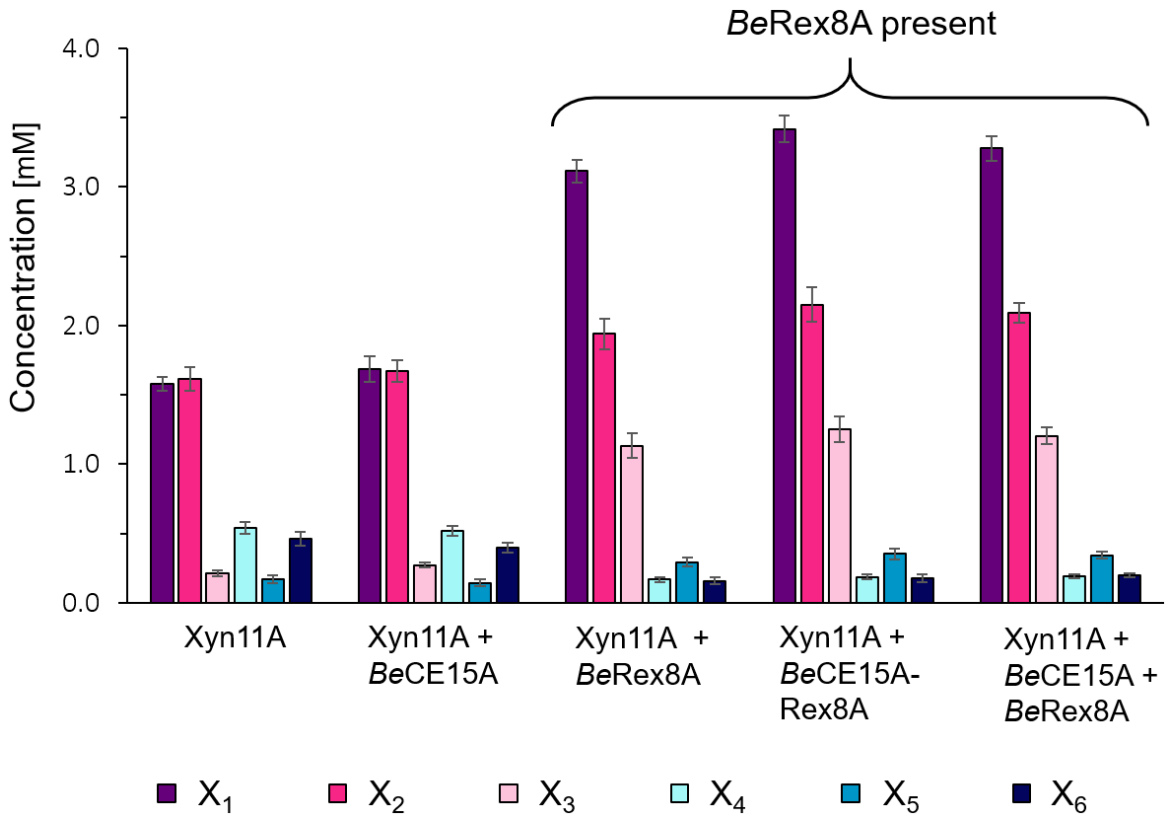


Figure 6: Xylooligosaccharide production profiles from corn cob biomass hydrolysis by Xyn11A supplemented with different enzyme constructs from *B. eggerthii*. The figure was adapted from paper III. When Xyn11A (*endo*-*β*-1,4-xylanase from *N. patriciarum*) was supplemented with a Rex in the form of *Be*Rex8A, *Be*CE15A-Rex8A or an equimolar mix of *Be*CE15A and *Be*Rex8A xylan hydrolysis was increased. Concentrations of X_1 (xylose), X_2 (xylobiose), X_3 (xylotriose), X_4 (xylotetraose), X_5 (xylopentaose), and X_6 (xylohexaose) were measured after 30 h of incubation using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Data are presented as average of three replicates with standard errors of the mean.

3.3. Carbohydrate esterases target ester-linked xylan decorations

As GHs cleave the carbohydrate backbone and side chains of polysaccharides, members of this enzyme class are of central importance for the degradation of xylans. Generally, many different CAZymes work in concert to reach complete degradation of xylan, especially in the case of the complex GAX. Here, CEs facilitate GH activity by cleaving non-carbohydrate decorations from the xylan backbone, thus increasing accessibility for GHs (Fig. 7). Examples for xylan decorations targeted by CEs are ester-linked acetic acid (acetylation), ferulic acid (feruloylation), and glucuronic acid (99). In the next sections of this chapter, I discuss examples of CEs targeting each of these decorations.

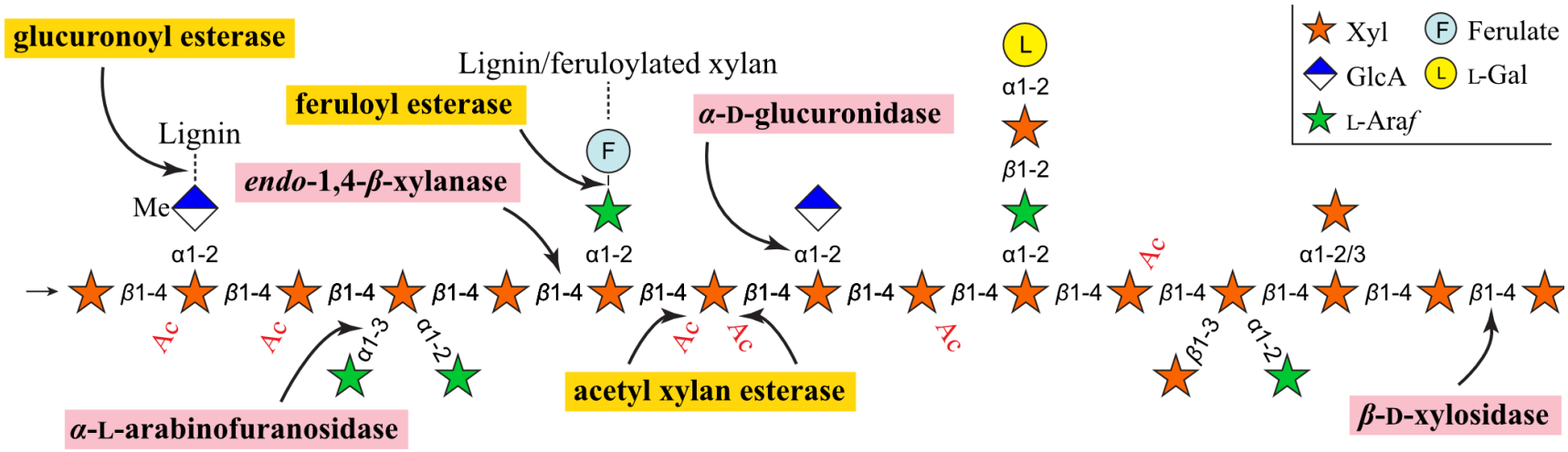


Figure 7: Selection of important enzymatic activities involved in the hydrolysis of glucuronoarabinoxylan. The figure was adapted from paper I. Glucuronoarabinoxylan is a highly decorated form of xylan and for its complete enzymatic hydrolysis many different enzyme activities are required to act in concert. Glycoside hydrolases are highlighted in pink and carbohydrate esterases are highlighted in yellow. Acetyl groups are marked with “Ac” in red.

3.4. Enzymes belonging to carbohydrate esterase family 15 show varying glucuronoyl esterases activities

Glucuronoyl esterases (GEs) are CEs that target the ester linkage between (xylan-bound) glucuronic acid and lignin (Fig. 7). In chapter 3 section 2, I described the Rex enzyme *Be*Rex8A of *B. eggerthii*. *Be*Rex8A is the C-terminal domain of the multicatalytic *Be*CE15A-Rex8A enzyme (Fig. 8). The N-terminal domain of *Be*CE15A-Rex8A was predicted to be a CE15 domain (*Be*CE15A). All to date characterized CE15 enzymes have been categorized as GEs (89).

Figure 8: Enzyme architecture of the multicatalytic *Be*CE15A-Rex8A. The figure was adapted from paper III. *Be*CE15A-Rex8A (locus tag HMPREF1016_02164) is encoded by *B. eggerthii* and its domains are drawn to scale: signal peptide (black), CE15 (teal), and GH8 (light pink).

*Be*CE15A showed only minor activity on GE model substrates compared to previously studied GEs (**paper III**). In general, model substrates are useful for establishing enzyme activities for uncharacterized enzymes, as well as for comparing activities to identify promising candidates for further investigation/application. However, model substrates often only model part of the native substrate of the enzyme. The low activity of *Be*CE15A on the here tested GE model substrates indicates that the enzyme either targets a linkage different to many other characterized GEs or that the tested model substrates did not mimic the linkage targeted by the enzyme sufficiently.

Interestingly, while the observed activity of *Be*CE15A on GE model substrates was much lower than for most other characterized GEs, it was similar to that of the previously characterized *Ot*CE15B from the soil bacterium *Opitutus terrae* (100). *Be*CE15A and *Ot*CE15B share an amino acid substitution of the arginine residue close to the catalytic serine that is found conserved in other GE enzymes (Fig. 9). This arginine residue is hypothesized to contribute in forming the oxyanion hole in the active site and stabilizing the transition state during catalysis (100-102). In *Be*CE15A the residue is instead a non-polar phenylalanine residue and in *Ot*CE15B a tyrosine residue is found in the equivalent position (100). Creation of enzyme variants of *Be*CE15A in which the phenylalanine was replaced with arginine (F231R) and tyrosine (F231Y) did not improve activity on GE model substrates but instead led to a complete loss of activity (**paper III**). With our current understanding of GEs,

the activity and sequence discrepancies of *Be*CE15A and *Ot*CE15B cannot be fully explained, but more research for instance in the form of structural studies of many more GE members might elucidate whether these enzymes have a different role in biomass turnover to classical GEs and if activities present in CE15 members are more diverse than currently known.

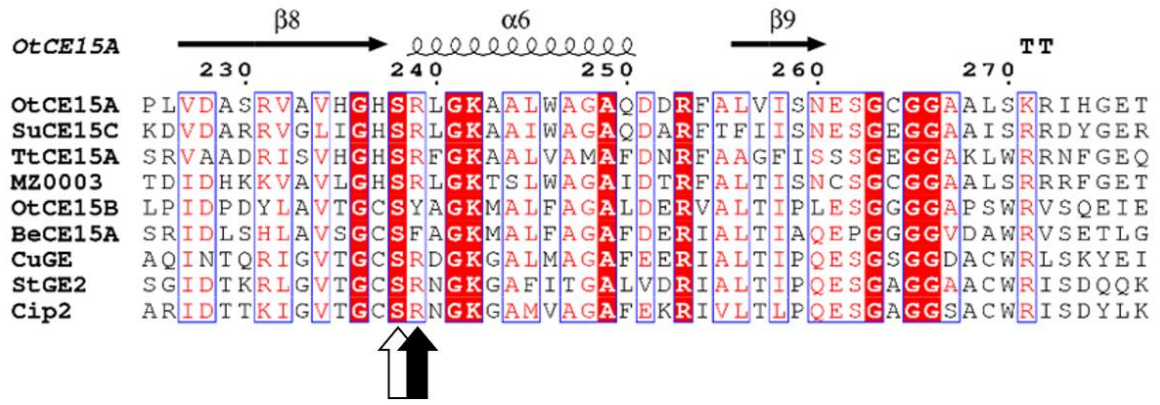


Figure 9: Sequence-based alignment of *Be*CE15A showing the region of the catalytic serine and the neighboring phenylalanine residue. The figure was adapted from paper III. The secondary structural elements of *Ot*CE15A (101), a glucuronoyl esterase from *Opitutus terrae* with published structure (PDB accession 6SYU) served as template. The catalytic serine is marked with a white arrow, the usually conserved arginine residue that is substituted in *Be*CE15A with phenylalanine and in *Ot*CE15B with tyrosine is marked by a black arrow.

3.5. Enzymes belonging to carbohydrate esterase family 6 target acetyl groups

To date, all CE6 members listed on CAZy have been characterized as acetyl xylan esterases (AXEs), which are enzymes cleaving acetate groups from the xylan backbone (89). Acetylation is a common decoration in xylan and as such it is no surprise that CE6 enzymes are encoded in the genomes of various bacteria and eukaryotes, including the fungus *N. patriciarum* (103). Still, not many CE6 enzymes have been biochemically characterized and only one structure has been published to date (104). The published structure originates from At4g34215 which is encoded by the eudicot plant *Arabidopsis thaliana* and has a low sequence similarity to the CE6 enzymes investigated here. In this thesis, I contribute to our understanding of CE6 enzymes with the biochemical characterization of three CE6 domains originating from different Bacteroidetes species: *Bo*CE6, *Fj*CE6, and *Dm*CE6A. *Bo*CE6 and *Fj*CE6 are part of multicatalytic enzymes encoded by *B. ovatus* and *F. johnsoniae*, respectively. *Dm*CE6A is a single-catalytic domain enzyme encoded by *D. mossii* (Fig. 10).

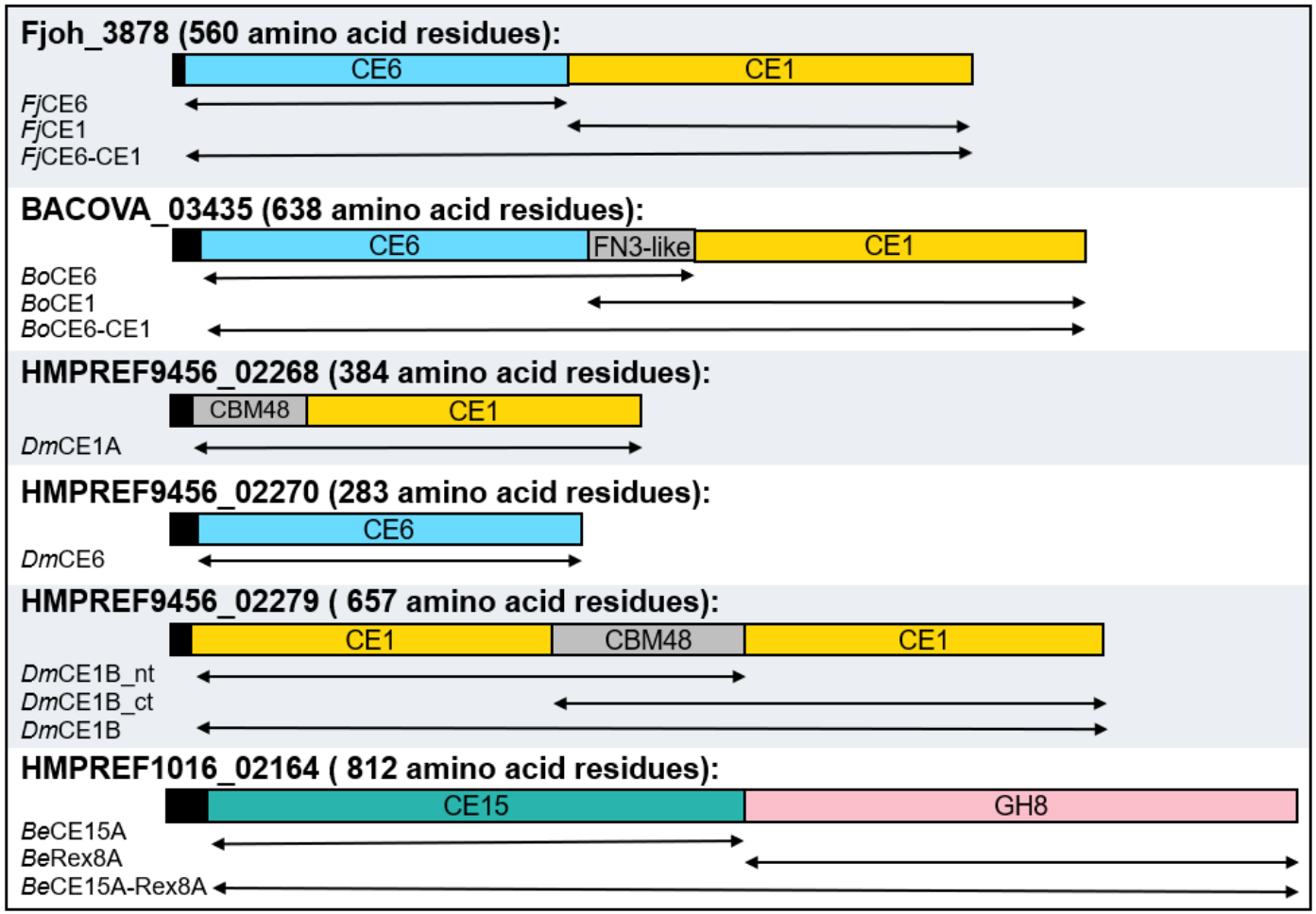


Figure 10: Enzyme architecture of enzymes characterized in this thesis. The figure is compiled from data presented in papers I - III. For each gene, enzyme encoding domains are drawn to scale and locus tag as well as number of amino acid residues are given. Signal peptides are shown in black and were excluded during enzyme production. Catalytic active domains are shown in blue (CE6), yellow (CE1), teal (CE15) or light pink (GH8) and non-catalytic domains are shown in gray. Arrows indicate which domains of the full-length enzyme are present in each enzyme construct.

The enzymes *Bo*CE6, *Fj*CE6, and *Dm*CE6A all harbored the typical serine-histidine-aspartate catalytic triad and were active on model substrates routinely used to test for acetyl esterase activity (*p*-nitrophenyl acetate, *p*NP-Ac; 4-methylumbelliferyl acetate, 4-MU-Ac). Comparison of *Dm*CE6A to *Bo*CE6 and *Fj*CE6 revealed a superior activity on 4-MU-Ac with a 30,000-fold higher catalytic efficiency (k_{cat}/K_M) (Table 3, **papers I & II**). This superior catalytic efficiency of *Dm*CE6A over *Bo*CE6 and *Fj*CE6 can be largely attributed to a four orders of magnitude higher turnover number (k_{cat}). The K_M of *Dm*CE6A on 4-MU-Ac was 15- and 19-fold lower than those previously reported for acetyl xylan esterase I and II from *Thermoanaerobacterium* sp. strain JW/SL-YS485, respectively (105). The catalytic efficiency of *Tm*AxeA, an AXE encoded by the hyperthermophilic bacterium *Thermotoga maritima*, on *p*NP-Ac was with 580 $s^{-1}mM^{-1}$ comparable to that of *Dm*CE6A (106). The K_M and k_{cat} of *Tm*AxeA were reported with 0.12 mM and 70 s^{-1}, respectively. Both *Tm*AxeA and *Dm*CE6A seem to be exceptionally active on *p*NP-Ac compared to other characterized AXE enzymes. For example, they both have much lower K_M values on *p*NP-Ac than *O*AxeA from *Orpinomyces* sp. strain PC-2 (K_M of 0.9 mM; (107)) and higher turnover numbers than reported for BD-FAE, a multicatalytic CE isolated from a metagenome sample (k_{cat} value of 0.89 s^{-1}; (108)). As mentioned earlier, model substrates are useful for initial activity screens, but have a limited ability to simulate native substrates. For instance, *p*NP-Ac and 4-MU-Ac only model the ester-linked acetate group but not an attached glycan moiety. Structural motifs surrounding the enzymatically targeted bond can play an important role in target recognition but can also hamper access of the enzyme. As a result, the observed high activity of *Dm*CE6A on model substrates might not translate to high activity on more natural substrates.

For this reason, *Bo*CE6, *Fj*CE6, and *Dm*CE6A were also tested on ball-milled corn cob biomass. Corn cob biomass is rich in GAX (109) and the xylose moieties in GAX can by acetylated at the C2 and C3 positions (110,111). AXE activity can be measured using different methods. In this thesis, I mainly chose to monitor the activity of *Bo*CE6, *Fj*CE6, and *Dm*CE6A by comparing their abilities to boost the xylanase activity of Xyn11A (**papers I & II**). While AXE activity is measured indirectly using this approach, it allows for a setup that is more similar to the native environment of these enzymes, in which they act in concert with other CAZymes. *Bo*CE6, *Fj*CE6, and *Dm*CE6A were able to boost the xylanase activity of Xyn11A on corn cob biomass. *Bo*CE6 boosted xylanase activity 1.4-fold, *Fj*CE6 1.4-fold, and *Dm*CE6A 2.8-fold, respectively (Fig. 11 A-C). *Dm*CE6A yielded the highest boost in xylanase activity, which is in agreement with the high activity that was observed for *Dm*CE6A on AXE model substrates.

Acetylation is also common in GX, which can be found in abundance in hardwoods such as beechwood (112). Japanese beechwood biomass was used as native substrate in additional Xyn11A boosting studies with *Fj*CE6 (**paper I**). In these experiments, the supplementation of reactions with *Fj*CE6 increased the xylanase activity of

Table 3: Kinetic parameters of investigated CE1 and CE6 enzymes. The enzymes are encoded by *F. johnsoniae*, *B. ovatus*, and *D. mossii*. The presented data was taken from papers I & II.

Enzyme	Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
*Fj*CE6	*p*NP-Ac	6.9 ± 2.5	57 ± 12	8.3 ± 3.5
	4-MU-Ac	0.15 ± 0.02	6.7 ± 0.2	43.4 ± 4.4
*Fj*CE1	*p*NP-Ac	0.95 ± 0.14	5.6 ± 0.3	5.9 ± 0.9
	4-MU-Ac	0.07 ± 0.01	0.61 ± 0.03	8.7 ± 1.7
	MFA	0.11 ± 0.02	2.5 ± 0.2	23.1 ± 4.5
	MSA	0.14 ± 0.02	1.8 ± 0.1	13.0 ± 2.2
	MCA	0.34 ± 0.10	0.83 ± 0.14	2.7 ± 1.0
	M*p*CA	0.39 ± 0.12	1.5 ± 0.3	3.9 ± 1.4
*Bo*CE6	*p*NP-Ac	1.8 ± 0.2	36.0 ± 2.2	20.1 ± 2.7
	4-MU-Ac	0.027 ± 0.006	1.2 ± 0.1	42.8 ± 10.0
*Bo*CE1	*p*NP-Ac	1.4 ± 0.1	0.70 ± 0.02	0.50 ± 0.04
	4-MU-Ac	0.04 ± 0.01	0.28 ± 0.01	7.1 ± 1.6
	MFA	not saturable up to 0.3 mM		0.011 ± 3×10-4
	MSA	not saturable up to 0.3 mM		0.043 ± 0.017
*Dm*CE1A	*p*NP-Ac	4.4 ± 0.67	1.03 ± 0.07	0.23 ± 0.04
	4-MU-Ac	0.55 ± 0.06	38.6 ± 1.26	70 ± 8
	MFA	not saturable up to 0.4 mM		0.12 ± 0.007
	MSA	0.18 ± 0.04	0.13 ± 0.01	0.70 ± 0.2
*Dm*CE1B	*p*NP-Ac	3.5 ± 0.70	12.92 ± 1.1	3.7 ± 0.81
	4-MU-Ac	0.37 ± 0.06	870 ± 3	2300 ± 400
	MFA	not saturable up to 0.4 mM		5.0 ± 0.35
	MSA	0.36 ± 0.10	5.0 ± 0.83	14 ± 4.5
	M*p*CA	not saturable up to 0.4 mM		0.59 ± 0.02
*Dm*CE1B_nt	*p*NP-Ac	1.2 ± 0.24	6.0 ± 0.40	5.1 ± 1.1
	4-MU-Ac	0.30 ± 0.04	630 ± 22	2100 ± 300
	MFA	0.16 ± 0.03	1.8 ± 0.13	11 ± 2.0
	MSA	0.29 ± 0.07	2.7 ± 0.34	9.2 ± 2.4
	M*p*CA	0.31 ± 0.09	0.61 ± 0.10	2.0 ± 0.7
*Dm*CE1B_ct	*p*NP-Ac	3.8 ± 0.77	0.15 ± 0.0015	0.041 ± 0.009
	4-MU-Ac	0.42 ± 0.10	11 ± 0.71	26 ± 6.6
*Dm*CE6A	*p*NP-Ac	0.18 ± 0.01	59 ± 0.87	320 ± 20
	4-MU-Ac	0.029 ± 0.003	38 ×103 ± 1200	$1.3\times10^6 \pm 2\times10^5$

All enzyme constructs were assayed on the full range of acetyl- and feruloyl esterase substrates (*p*NP-Ac, 4-MU-Ac, MFA, MSA, MCA, and M*p*CA), but kinetic parameters are only given where activity could be detected and was possible to measure. The data represent the average of three replicates with standard errors of the mean. Data were fitted to the Michaelis-Menten equation using OriginPro software. The catalytic efficiency (k_{cat}/K_M) was determined using linear regression for reactions that could not be saturated without exceeding the detection limit.

Xyn11A up to 20-fold (Fig 11 D). The observed minimal activity of Xyn11A when *Fj*CE6 was not present indicates that the xylan in Japanese beechwood is heavily decorated with acetyl groups, thus prohibiting substrate access for the xylanase.

The activity profiles of the here investigated CE6 enzymes on model and native substrates illustrate for one the diversity in activity within this enzyme family in AXE activity. Secondly, the presented data also shows the potential for enzyme discovery in environments where biomass is actively decomposed. Enzymes like *Bo*CE6, *Dm*CE6A (both encoded by gut Bacteroidetes species), and *Fj*CE6 (encoded by a soil Bacteroidetes species) can advance our understanding on how biomass is enzymatically degraded on a basic research level. These enzymes could further find application in industrial enzyme cocktails tailored for the degradation of biomass rich in complex xylan.

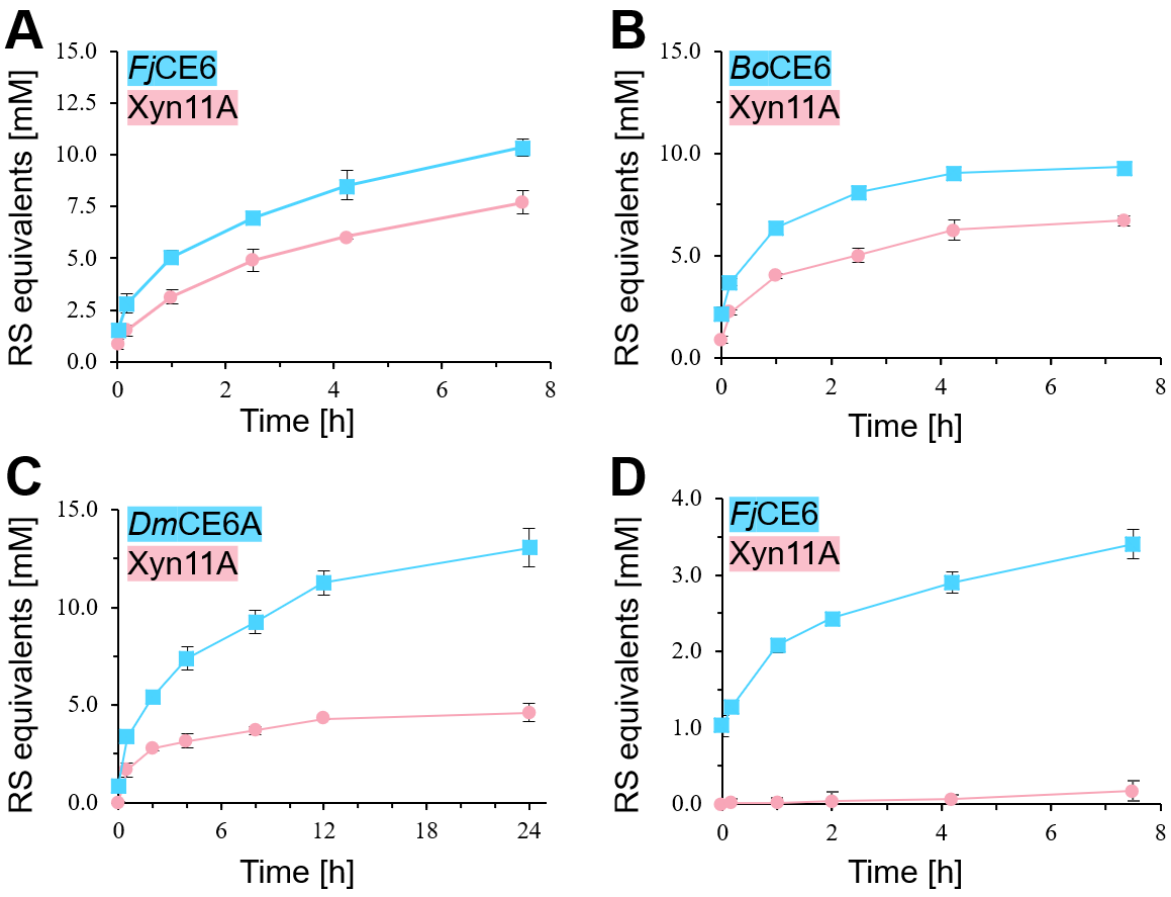


Figure 11: Effect of CE6 supplementation on xylanase hydrolysis activity. Data was taken from papers I &II. The *endo*-*β*-1,4-xylanase Xyn11A (light pink) was supplemented with different carbohydrate esterases (blue) during incubation on corn cob biomass: **(A)** *Fj*CE6, **(B)** *Bo*CE6, and **(C)** *Dm*CE6A. **(D)** Xyn11A was supplemented with *Fj*CE6 on Japanese beechwood biomass. The amount of reducing sugar (RS) equivalents released over time was used to quantify xylan hydrolysis. Data are presented as average of three replicates with standard errors of the mean.

3.6. Enzymes belonging to carbohydrate esterase family 1 can target both acetyl and feruloyl groups

Among CE families, CE1 harbors the most diverse activity portfolio with in total seven distinct enzyme activities. This diversity suggests that enzymes of CE1 play various roles during biomass degradation. The two most relevant activities in CE1 for xylan degradation are AXE and feruloyl esterase (FAE) activity (89) and the enzymes investigated in this thesis were screened for both of these activities.

FAEs are known to target feruloyl esters in GAX (Figure 7), which are found in commelinid monocots (68,77). The amount of ferulic acid present in GAX is moderate with for example up to 3% of the dry weight in corn bran (113). However, as mentioned before ferulic acid moieties can form covalent linkages to lignin, which are important targets for loosening up the complex lignocellulose matrix. As lignin is highly heterogenous the enzymatically recognized motifs, including the ferulic acid ester linkages to lignin, are likely also very heterogenous and thus many specialized FAEs may be required to degrade the totality of these linkages. This hypothesis is a possible explanation for the diversity in activity present among FAEs, which was also observed for CE1 enzymes characterized in this thesis.

Overall, five distinct CE1 domains were characterized, four of which were part of three multicatalytic enzymes (*Fj*CE6-CE1, *Bo*CE6-CE1, and *Dm*CE1B; Fig. 10). *Dm*CE1B contains an N- and a C-terminal CE1 domain, which are designated as *Dm*CE1B_nt and *Dm*CE1B_ct, respectively. *Dm*CE1A was the only CE1 enzyme that was not multicatalytic, but contained besides the CE1 domain also a possible carbohydrate-binding module (CBM) domain.

*Fj*CE1, *Bo*CE1, *Dm*CE1A, *Dm*CE1B_nt, and *Dm*CE1B_ct showed moderate to minor activity on *p*NP-Ac and 4-MU-Ac, with activities on 4-MU-Ac being generally more pronounced (Table 3, **papers I & II**). The activity profiles on the FAE model substrates (methyl ferulate, MFA; methyl sinapate, MSA; methyl caffeate, MCA; methyl *p*-coumarate, M*p*CA) were distinct and considerable compared to AXE activities. Interestingly, *Bo*CE1 and *Dm*CE1A had similar activity profiles, being only active on MFA and MSA with a slight preference for MSA. This observation coincides with both enzymes harboring a serine-histidine-glutamate catalytic triad. Contrary to this, *Fj*CE1 and *Dm*CE1B_nt were active on MFA, MSA, M*p*CA and in the case of *Fj*CE1 also on MCA (FAE model substrates are mentioned is descending order of activity), while harboring a serine-histidine-aspartate catalytic triad, which is also common in serine proteases and lipases (114). Catalytic dyads (serine-histidine) have also been reported for CEs but are limited to CE2 (103,115,116). The correlation of catalytic triad residues and preference for FAE model substrates may be a coincidence and in-depth structural investigations would be necessary to evaluate this observation.

*Dm*CE1B_ct was an outlier among the tested CE1 enzymes as it was not active on any FAE model substrates (Table 3, **paper II)**. Moreover, *Dm*CE1B_ct showed only minor activity on the AXE model substrates. **Paper II** however presents the successfully solved crystal structure of *Dm*CE1B_ct, which consists of one CE1 domain appended to one CBM domain (Fig. 12 A).

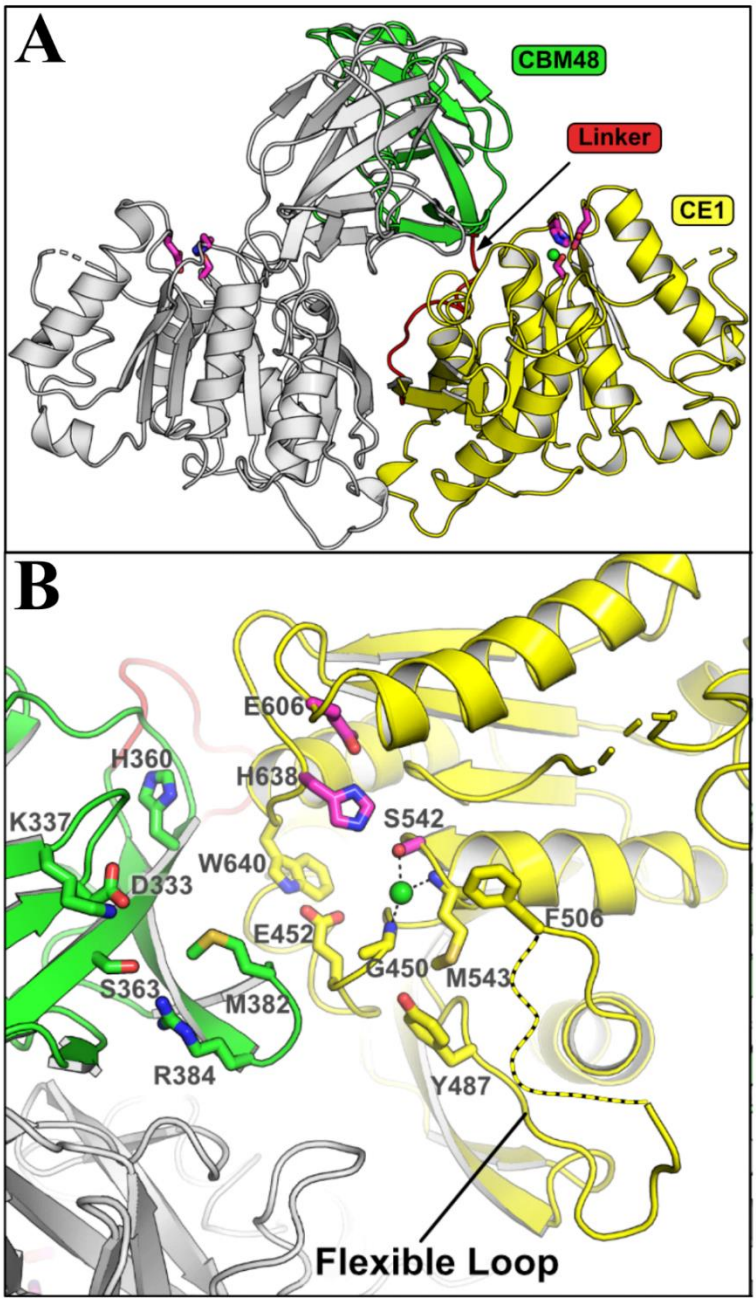


Figure 12: Structure of *Dm*CE1B_ct. The figure was adapted from paper II. **(A)** The overall structure of the proposed dimeric biological unit of *Dm*CE1B_ct with one monomer shown in gray, the other monomer colored in green (CBM domain) and yellow (CE1 domain), and the linker region connecting the CBM and CE1 domains shown in red. The glutamic acid, histidine, and serine residues of the catalytic triad are colored in magenta and the chloride ion found in the active site of one of the protomers is shown as a green sphere. **(B)** Close-up view of the active site of *Dm*CE1B_ct with marked flexible loop.

While the crystal structure suggests that *Dm*CE1B_ct forms dimers, it remains unclear if this is also the case under physiological conditions and influences xylan degradation. The structure of *Dm*CE1B_ct revealed the presence of a flexible loop close to the active site similar to the previously described wtsFae1A and wtsFae1B, which might promote substrate binding (117). However, not all of the residues of the flexible loop could be resolved in *Dm*CE1B_ct (Fig. 12 B). In *Am*Fae1A and *Bi*Fae1A a *β*-clamp has been found capping the active site and has been associated with a preference for single feruloyl moieties over diferulate esters (118,119). Diferulate crosslinks can join neighboring feruloylated GAX chains and thus create recalcitrant matrices (78). A similar *β*-clamp was not present in *Dm*CE1B_ct. While the native target of *Dm*CE1B_ct remains elusive the presented structure might help future studies to elucidate its true substrate.

3.7. Multicatalytic enzymes can benefit from intramolecular synergy

Due to the complexity of xylan and especially GAX, lignocellulolytic microorganisms apply a plethora of enzymes to achieve complete or at least partial hydrolysis. Most of these CAZymes harbor one active site, but some enzymes contain multiple catalytic domains, like the here investigated *Bo*CE6-CE1, *Fj*CE6-CE1, *Dm*CE1B, and *Be*CE15A-Rex8A. The benefit of producing such a multicatalytic enzyme over several single-catalytic enzymes is still not well understood.

Multicatalytic enzyme architectures exist outside of CAZymes. An example for this is the human fatty acid synthase, which is a homodimer consisting of many active domains catalyzing the synthesis of the saturated fatty acid palmitate from acetyl-CoA and malonyl-CoA (120,121). Cellulosomes, which are multi-protein complexes targeting lignocellulose, can be equally complex though they are not covalently connected. Instead, cellulosomes comprise a scaffolding protein in which different enzymatic subunits dock into to form a large multi-enzyme complex. Cellulosome-producing bacteria have been shown to be able to degrade highly recalcitrant polysaccharides, during which they release soluble saccharides which can support the growth of saccharolytic microbes (122,123). Examples of multicatalytic GHs include the cellulolytic CelA from *Caldicellulosiruptor bescii* and the chitinolytic ChiA from *F. johnsoniae*. Both of these enzymes have been shown to be highly efficient thanks to harboring complementary *endo*- and *exo*-acting catalytic domains (124,125). However, multicatalytic CEs have not been previously described.

To assess whether the multicatalytic enzyme architecture of the CEs *Bo*CE6-CE1, *Fj*CE6-CE1, *Dm*CE1B, and *Be*CE15A-Rex8A benefits xylan hydrolysis, I compared the ability of each active domain with their full-length enzyme versions (Fig. 10; **papers I - III**).

When Xyn11A was supplemented with *Bo*CE6 or *Bo*CE1, an equimolar mix of *Bo*CE6 and *Bo*CE1 or the full-length enzyme (*Bo*CE6-CE1), xylanase activity was boosted on corn cob biomass (Fig. 13). Interestingly, Xyn11A activity was facilitated to a higher degree when supplemented with *Bo*CE6-CE1 than with an equimolar mix of *Bo*CE6 and *Bo*CE1, which is illustrated by the highest release of reducing sugar equivalents observed. I hypothesize that the reason behind the synergistic effect is the ability of the full-length *Bo*CE6-CE1 to simultaneously cleave acetyl groups in close proximity to feruloyl groups (**paper I**). Such structural motifs have been shown to be present in corn fiber (126), in which closely positioned feruloyl groups inhibited the AXE action of FjoAcXE, an enzyme encoded by *F. johnsoniae* (127).

The observed intramolecular synergy of *Bo*CE6-CE1 was the first instance of direct proof of a beneficial effect of the enzyme architecture of multicatalytic CE enzymes. The other three here investigated multicatalytic CEs (*Fj*CE6-CE1, *Dm*CE1B, and *Be*CE15A-Rex8A) did not exhibit a similar intramolecular synergy on corn cob biomass when supplemented to Xyn11A (**papers I - III**). The reason for the lack of observed synergy is not clear and more research is necessary to illuminate the differences between these enzymes. One possibility is that the tested corn cob biomass did not contain the structural motif targeted by both active domains of these enzymes. To observe synergy between the active domains of multicatalytic enzymes all the domains involved must be able to act on the substrate at hand to a sufficient degree, meaning that orientation and proximity of the active domains must match the targeted linkages and vice versa to complement each other. The orientation of domains in multicatalytic and multidomain enzymes can be influenced by linker regions. Catalytic and non-catalytic domains (for example CBMs) of multicatalytic enzymes are often connected by linker regions. These regions vary in length for example ranging between 30 to 180 residues (128). Linkers may give multicatalytic enzymes great flexibility, but could also form unpredicted secondary structure elements or may be glycosylated resulting in a more rigid protein structure. Another possibility why synergy was not observed could be that the targeted motif was not present in sufficient abundance to quantify the enzymes' action with the detection methods applied. Lastly, a lack of observed synergy could also be rooted in simply no intramolecular synergy existing between the active domains of *Fj*CE6-CE1, *Dm*CE1B, and *Be*CE15A-Rex8A. This would demand a new explanation as to why these enzyme activities are combined into one enzyme unit. The fact that multicatalytic enzymes, for example with CE6-CE1 architectures, are not isolated cases but frequently encoded in the genomes of Bacteroidetes (89) implies a biological reason and warrants further investigation. Contrary to the CE6-CE1 enzyme architecture, the combination of Rex and CE15 domains into one multicatalytic enzyme is rare and confined to very few members of the Bacteroidetes phylum including *Bacteroides* sp. NSJ-48, *B. stercoris*, *B. gallinarum*, *Prevotella* sp. BP1-148, and *Prevotella* sp. BP1-145 (89).

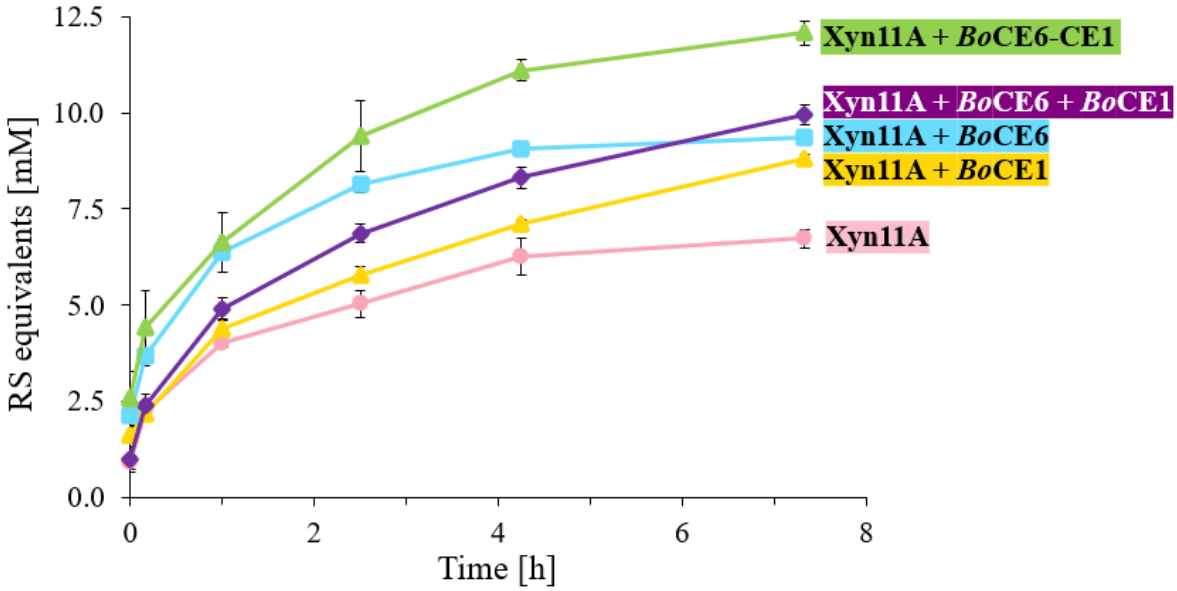


Figure 13: Hydrolysis of corn cob biomass by Xyn11A supplemented with different carbohydrate esterase constructs encoded by *B. ovatus*. The figure was adapted from paper I. Reducing sugar (RS) equivalents were measured by 3,5-dinitrosalicylic acid (DNSA) assays and data are presented as average of three replicates with standard errors of the mean. Xylan hydrolysis was higher when Xyn11A was supplemented with the full-length *Bo*CE6-CE1 enzyme than when supplemented with an equimolar mix of *Bo*CE6 and *Bo*CE1 demonstrating an intramolecular synergistic effect between the catalytic domains when fused into one multicatalytic enzyme.

3.8. Bacteroidetes encode loci of carbohydrate-active enzymes relevant for enzyme discovery and application

Bacteroidetes are well versed glycan degraders and investigation of their degradation abilities and strategies holds the potential of advancing our understanding of lignocellulose hydrolysis, a process with a multitude of applications. For this reason, I mined the genomes of different Bacteroidetes species for little studied and interesting enzyme architectures and activities. Investigation and biochemical characterization of the CEs described above revealed that multicatalytic enzymes are a strategy employed by the Bacteroidetes phylum to degrade biomass. Research on multicatalytic CEs is still in its infancy and their importance is still unclear. In this thesis, the mining of multicatalytic enzymes led to the discovery of highly active domains from both "standard" and multicatalytic enzymes. *Dm*CE6A was not only a highly active AXE but also showed promiscuous behavior towards acetylation position and acetylated monosaccharide in subsequent thin layer chromatography and nuclear magnetic resonance studies (**paper II**). Another example is the earlier described *Fj*CE6, which boosted xylanase activity up to 20-fold on Japanese

beechwood (**paper I**). AXEs have been shown to be promising candidates for improving the saccharification of lignocellulosic material for biofuel production (129), as have FAEs (130-132). FAEs have also been applied in the pulp and paper industry to increase delignification and bleaching (133,134), and in feed production to increase digestibility and caloric value (135,136). Besides their relevance for "cleaning up" ester-linked compounds, FAEs also receive attention for the release of ferulic acid. Ferulic acid itself is a valuable product that has been shown to have antioxidant properties (137,138). Further, ferulic acid can serve as precursor in the production of styrenes, polymers, catechol (for example used in photography), and vanillin, which is an important aromatic compound used in food, pharmaceutical, and cosmetic industries (139-141).

Besides practical application the investigation of CEs encoded by Bacteroidetes species also holds the potential to discover yet unknown enzyme activities or at least pointing out remaining knowledge gaps as exemplified by the *Be*CE15A enzyme. The continued investigation of such enzymes may increase our knowledge of enzyme architectures and structural motifs present in plant biomass.

In summary, CAZymes encoded by Bacteroidetes and as described in this thesis, are a valuable resource that can help us expand our understanding of enzymatic plant biomass degradation. Enzyme discovery is of course possible in all kinds of organisms, but Bacteroidetes have a special feature that allows prediction of enzyme activities by taking the genomic neighborhood of the encoding gene into account. This special feature of Bacteroidetes of co-locating genes encoding CAZymes and related functions into PULs is the focus of the next chapter.

3.9. Main points of chapter 3

- Many different enzyme activities have to act in concert to hydrolyze xylan.
- Glycoside hydrolases cleave glycosidic bonds between carbohydrates and non-carbohydrates in the xylan backbone and xylan side-chains.
- The action of glycoside hydrolases can be hampered by xylan decorations, for example ester-linked acetic acid and ferulic acid.
- A multitude of specialized carbohydrate esterase cleave these decorations and as a result facilitate xylanase mediated hydrolysis of xylan.
- Multicatalytic enzymes harbor several distinct active domains and can profit from intramolecular synergy, as shown for the multicatalytic carbohydrate esterase *Bo*CE6-CE1.
- Carbohydrate-active enzymes in general, and PUL and Bacteroidetes encoded carbohydrate-active enzymes in particular, are a valuable resource for advancing the basic understanding of enzymatic biomass degradation and improving existing applications.

4. Plant biomass degradation by Bacteroidetes species is facilitated by polysaccharide utilization loci

In the previous chapter I outlined how various enzyme activities target different motifs found in xylan and demonstrated the benefit of employing complementary enzyme activities together. Analogous to industrial enzyme cocktails, lignocellulolytic microbes produce an assortment of CAZymes to metabolize plant biomass. As enzyme production is costly for the microbial cell, microbes employ different strategies to make efficient use of their CAZymes. In this thesis, I focus on the Bacteroidetes and their strategy for degrading polysaccharides.

This chapter introduces the role of the Bacteroidetes for mammalian gut health. The structure of PULs is explained on the basis of the archetypical starch utilization system. After discussing recent advances in PUL-related research, the pICKUP method is introduced, a new tool for PUL transfers that was developed as part of this thesis. Lastly, the validity of the pICKUP method is demonstrated by transferring the mixed-linkage *β*-glucan utilization locus between two Bacteroidetes species.

4.1. Bacteroidetes metabolize dietary fiber in the human gut

Bacteria of the Bacteroidetes phylum are Gram-negative, non-spore-forming, and rod-shaped. Its members can be divided into six classes: Bacteroidia, Chitinophagia, Cytophagia, Flavobacteriia, Saprospiria, and Sphingobacteria (142). Within Bacteroidetes, members of the Bacteroidia class, especially belonging to the *Bacteroides* genus, have received much attention in research.

Bacteroidetes excel in metabolizing biopolymers and have established themselves in numerous environments such as sediments, soils, sea water, fresh water, and the gastrointestinal tract and skin of animals (143). In the mammalian gut Bacteroidetes metabolize dietary glycans (starches, fructans, pectins, and hemicelluloses) and host-derived glycans (*O*- and *N*-linked glycans and glycosaminoglycans) (144), and are one of two dominant phyla present, with the other being the mostly Gram-positive Firmicutes (4,5). This thesis focuses on the degradation of hemicelluloses by Bacteroidetes, but an excellent overview of the variety and breadth of glycans metabolized by this phylum is given in Porter *et al.* (6).

The abundance of Bacteroidetes in the gut depends on the diet of the host. It has been shown that Bacteroidetes are more abundant in omnivores and herbivores than in carnivores (4). The number of Bacteroidetes and bacteria in general is also influenced by the location of residence within the gastrointestinal tract. Whereas low concentrations of bacteria reside in the stomach and small intestine in humans, colonic microbial cell

densities reach as high as 10^{12} colony forming units (CFU)/mL (145), making the mammalian colon one of the most densely populated environments on Earth (146,147).

Some HGM members are thought to influence and often benefit many of their hosts' vital functions such as development of the intestinal epithelium and lymphoid structures, fecundity, metabolism, immunity, and behavior (148-153). Bacteroidetes are generally associated with a healthy adult human gut microbiota (HGM) composition and different members of this phylum have been shown to play important roles in human gut health and disease (154). Because of their beneficial role it has been proposed that Bacteroidetes form a relationship of mutualism rather than commensalism with their human host (155), though it should be noted that certain Bacteroidetes species have also been shown to be opportunistic pathogens (154).

In the colon, Bacteroidetes metabolize dietary fiber (indigestible by the host) and in return excrete short-chain fatty acids (SCFAs) in their salt forms, such as butyrate (butyric acid) and acetate (acetic acid). SCFAs can be absorbed and metabolized by the host and serve as a nutrient source (7,8) (Fig. 14).

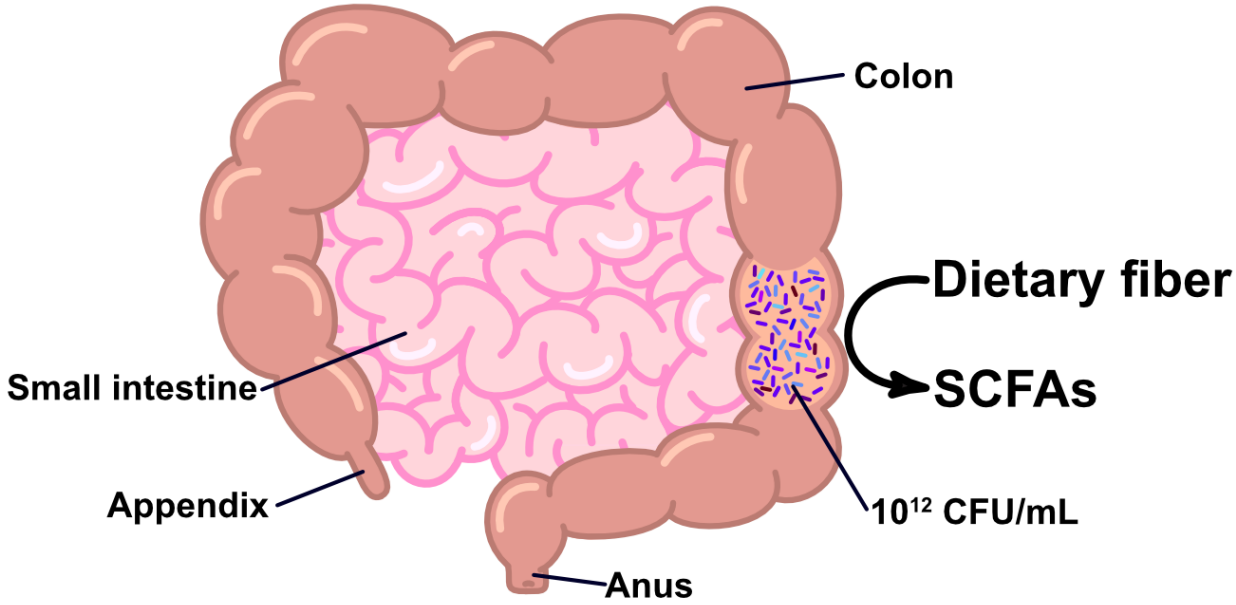


Figure 14: In the human gut resident microbes convert dietary fiber into short-chain fatty acids (SCFAs). The human gut is one of the most densely populated environments on Earth reaching up to 10^{12} colony forming units (CFU) per mL in the colon (145-147). The two dominant bacterial phyla of the human gut microbiome are Bacteroidetes and Firmicutes (4,5).

SCFAs have been correlated with a number of health benefits. For example the SCFAs butyrate, propionate, and acetate have been suggested to promote gut integrity (9-11). In the healthy human gut a layer of mucus separates the epithelium and luminal environment, thus creating a physical barrier between the host and microbiota (11). Disruption of gut integrity has been associated with diseases such as inflammatory bowel

disease, irritable bowel syndrome, and colorectal cancer (156,157). While acetate itself has been linked to various health benefits (158) it has also been hypothesized that Firmicutes scavenge acetate produced by Bacteroidetes to produce and excrete butyrate and propionate (159).

4.2. Bacteroidetes tackle polysaccharide diversity with strategic gene organisation and regulation: the polysaccharide utilization locus

Bacteroidetes employ a special strategy to degrade the extraordinarily diverse and complex structure of dietary fiber. To fully appreciate this strategy it is important to understand the challenges of their ecological environment:

- The human gut is densely populated resulting in fierce competition among HGM members.
- Substrate availability is dictated by the diet and dietary schedule of the host.
- Readily digestible diet components are absorbed by the host before reaching the colon.
- The human diet can be highly variable, which can make specialization on a certain substrate for example in the form of a specific polysaccharide a challenge for long-term colonization of this environment.
- Dietary fiber structures are diverse and require multiple CAZymes acting together to be degraded completely.
- Due to their size, polysaccharides cannot be taken up by the microbial cell without at least some preliminary extracellular enzymatic degradation. Secreted enzymes and partly digested polysaccharides might be used or taken up by competitors.

Despite these challenges Bacteroidetes thrive in the human gut by employing a strategy allowing for efficient protein use and flexible substrate response in the form of PULs. A PUL has been defined as a gene cluster encoding the proteins necessary to sense, bind, and degrade a specific polysaccharide (160). PUL systems have also been identified in species outside of the Bacteroidetes phylum. For instance, Firmicutes possess so-called Gram-positive PULs (gpPULs) with the major differences that gpPULs do not contain SusC/D-like gene pairs and generally harbor fewer CAZymes than Bacteroidetes-type PULs (161-165). As this thesis focuses on Bacteroidetes, I will limit the discussion to classical Bacteroidetes-type PULs.

4.3. The starch utilization system serves as archetype for polysaccharide utilization systems

The first PUL described was the starch utilization system (Sus) of the gut bacterium *Bacteroides thetaiotaomicron* (166,167). Since its discovery, the Sus has been researched in great detail and nowadays serves as a blueprint for similar gene systems within the Bacteroidetes phylum. The multitude of scientific works contributing to the elucidation of the Sus of *B. thetaiotaomicron* has been summarized in several excellent reviews, including publications by Martens *et al.* (168) and Foley *et al.* (169). Here, I will briefly describe the Sus as an example to illustrate the setup and working principle of PULs in general (Fig. 15).

The Sus is composed of the eight genes *SusRABCDEFG* which are hypothesized to be expressed at all times at a low level and upregulated upon sensing specific motifs of the target polysaccharide (170,171). The target polysaccharide of the Sus is starch, a polysaccharide composed of α-1,4- and α-1,6-linked glucose moieties (172). The disaccharide unit of starch is maltose. Imported maltose and maltooligosaccharides (MaltOs) bind to SusR, a sensory protein spanning the inner cell membrane, which then triggers upregulation of the rest of the Sus (166,173). SusE and SusF aid in the formation of the Sus complex (174) and together with the CBM58 domain of SusG bind extracellular starch to the cell surface (175). The GH13 domain (amylase) of SusG cleaves bound starch into MaltOs (175), which are then transported across the outer cell membrane into the periplasm by the SusC/D complex. The SusC/D complex acts as a "pedal-bin" transporter including a pore in which SusC is the TonB-dependent transporter and SusD assists binding and forms the lid on the pore (176-178). In the periplasm, MaltOs are further processed by the CAZymes SusA (neopullulanase) and SusB (α-glucosidase) to glucose, which is transported into the cytoplasm by an undefined glucose transporter (172,179-181). The Sus does not encode a glucose transporter, but genes encoding such transporters can be found in the genome of *B. thetaiotaomicron*. Possibly due to glucose being a pervasive nutrient source for *B. thetaiotaomicron* glucose transporter genes are not included in the Sus. Other PULs may encode specific monosaccharide transporters (124,182).

Due to differences in the structure of polysaccharides targeted by PULs, the repertoires of encoded CAZymes and carbohydrate binding proteins differ, but the functional setup with initial but limited hydrolysis on the cell surface followed by full degradation within the periplasms is ubiquitous among PULs in the HGM studied so far (183). Further, the presence of homologs to the SusC/D pair is conserved in PULs. These tandem SusC/D-like proteins serve as signature motifs for the automatic identification of PULs in the genome sequences of Bacteroidetes (184). While PULs appear to be highly efficient systems for the degradation of polysaccharides, not all species of the Bacteroidetes phylum encode PULs. Future research might reveal alternative polysaccharide degradation systems applied by Bacteroidetes.

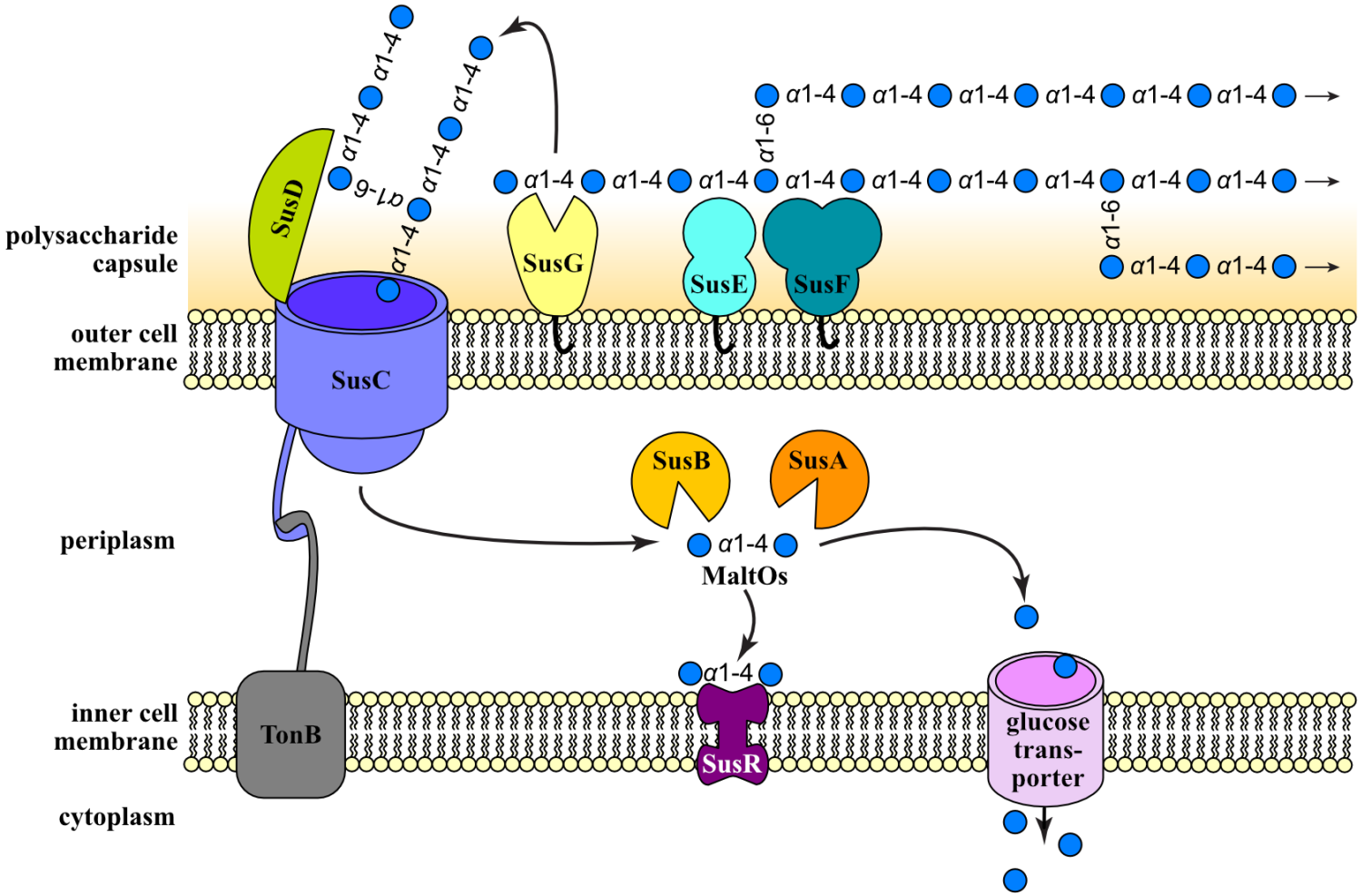


Figure 15: Components of the starch utilization system (Sus). The figure was adapted from Foley *et al.* (167). Starch is bound to the cell surface by SusE, SusF, and SusG. SusG also cleaves bound starch into maltooligosaccharides (MaltOs), which are transported through the SusC pore with assistance from SusD (binds MaltOs) and TonB (pulls away the pore plug). MaltOs (here symbolized by maltose) are further metabolized by SusA and SusB into glucose, which is transported into the cytoplasm by sugar transporters that are not encoded in the Sus gene cluster. Maltose also activates the regulatory protein SusR and initiates a positive feedback loop.

4.4. Multicatalytic enzymes may play important roles during polysaccharide utilization locus mediated glycan degradation

Over the past decade a plethora of PULs have been identified, mainly based on sequence similarity to the *B. thetaiotaomicron* Sus (182,184), but also through upregulation of gene clusters during growth on different glycans (185). Predicted and literature-derived PULs are collected in the polysaccharide utilization locus database (PULDB; http://www.cazy.org/PULDB/; (182)). To date, many proteins and especially CAZymes encoded in PULs have been studied. However, multicatalytic enzymes encoded in PULs received only little attention and this thesis aims to help close this gap in knowledge. As mentioned before in chapter 3, the enzymes studied in this thesis originated from PULs from different species. An overview of the genes present in these PULs is given in Fig. 16. Here, I describe these enzymes in a larger context that is not focused on their individual enzyme activities. In short, the enzymes *Bo*CE6-CE1 and *Fj*CE6-CE1 were sourced from the *Bo*XylL of *B. ovatus* and the *Fj*XylUL of *F. johnsoniae*, respectively (**paper I)**; *Dm*CE1A, *Dm*CE1B, and *Dm*CE6A were encoded in the predicted PUL 17 of *D. mossii* (**paper II**) and *Be*CE15A-Rex8A was part of predicted PUL 27 of *B. eggerthii* (**paper III**).

Previously the *Bo*XylL of *B. ovatus* has been characterized with a focus on GH enzymes, surface glycan binding proteins (SGBPs), and SusD-like and SusC-like proteins (112), but the only CE present in the PUL was overlooked. This CE was investigated in **paper I** and designated as *Bo*CE6-CE1. It has been shown that the presence of *Bo*XylL is crucial for the ability of *B. ovatus* to grow on GAX originating from corn, sorghum, and rice (112). A second PUL encoded by *B. ovatus* (*Bo*XylS) that was characterized in the same study was shown to be essential for the degradation of the less complex GX (birchwood) and AX (wheat) and did not encode any CEs. The here presented characterization of *Bo*CE6-CE1 as a multicatalytic AXE/FAE and its activity on GAX-rich corn cob biomass complements the findings by Rogowski *et al.* (112) and highlights the importance of decoration-cleaving enzyme activities for the efficient degradation of complex xylan through PULs.

While there is no synteny between *Bo*XylL and *Fj*XylUL, both PULs share a number of enzyme activities associated with the degradation of GAX, such as putative *β*-xylosidases (GH3, GH43), *endo*-xylanases (GH10), *α*-L-arabinofuranosidases (GH43) and *α*-glucuronidases (GH115), as well as the multicatalytic CE6-CE1 proteins studied here (Fig. 7). The presence of GAX-targeting CAZymes together with the previously reported ability of *F. johnsoniae* to grow on wheat arabinoxylan and methyl-glucuronoxylan (186) suggests that this Bacteroidete may metabolize GAX. The ability of *Fj*CE6-CE1 to boost xylanase activity on GAX-rich corn cob biomass supplies further evidence to support this hypothesis (**paper I**).

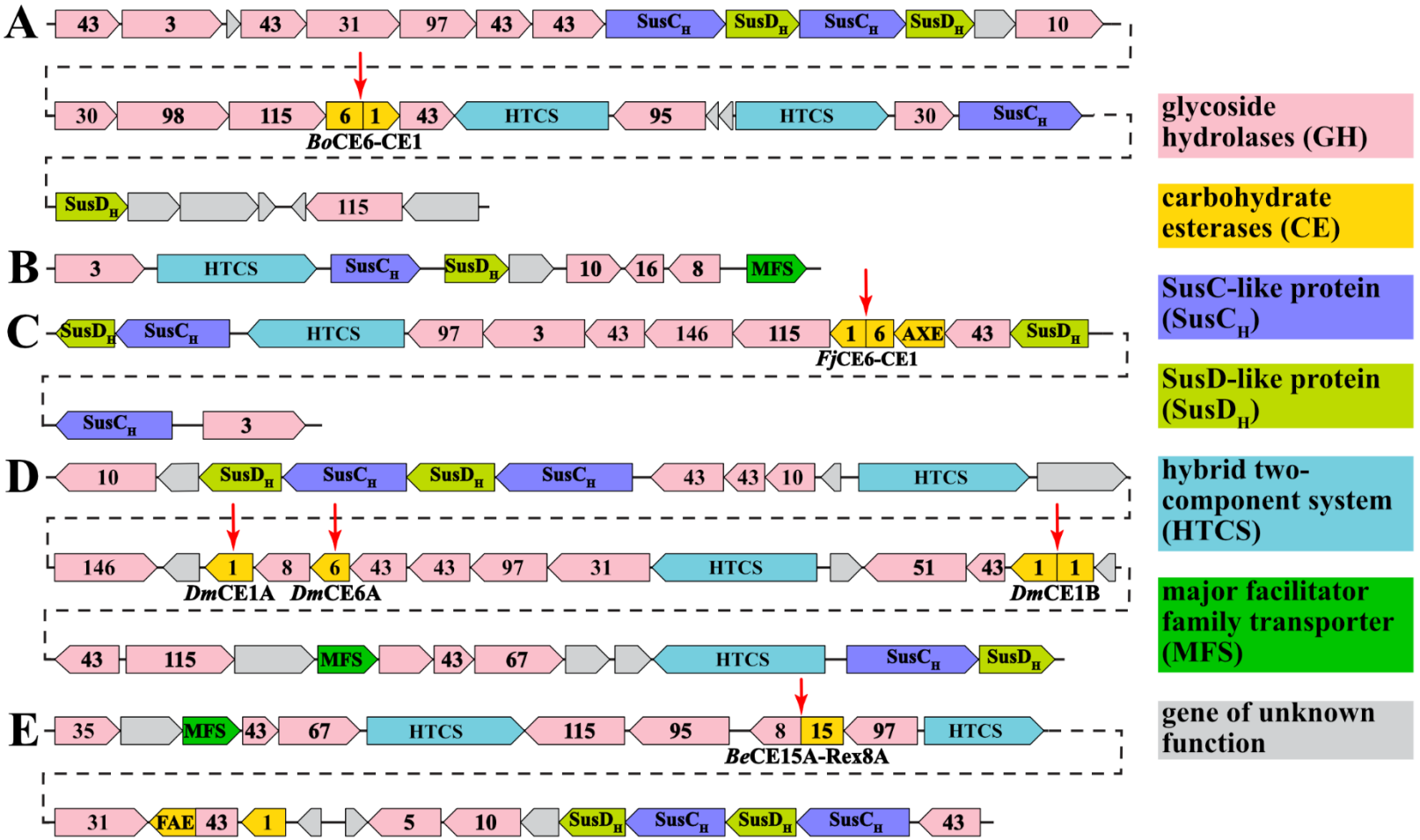


Figure 16: Gene composition of **polysaccharide utilization loci (PULs) encoding enzymes that were characterized in this thesis**. The figure is adapted from papers I - III. PUL architectures of **(A)** *Bo*XylL of *B. ovatus*, **(B)** *Fj*XylUL-I and **(C)** *Fj*XylUL-II of *F. johnsoniae*, **(D)** predicted PUL 17 of *D. mossii*, and **(E)** predicted PUL 27 of *B. eggerthii* were extracted from the PULDB (180). Genes and intergenic sequences are drawn to scale except for areas marked by dashed lines. Genes of biochemically characterized proteins are marked by red arrows.

The predicted PUL 17 of *D. mossii* encodes various GH families associated with the enzymatic degradation of complex xylan. Additionally, this PUL encodes three distinct CE enzymes, one of which is multicatalytic. Investigation of these CEs and their individual domains revealed that each CE domain had a distinct activity profile (**paper II**). This suggests that each CE domain fulfills a unique role during the degradation of the polysaccharide targeted by predicted PUL 17. Moreover, the fact that all three CEs are part of the same PUL indicates that these enzymes are produced at the same time. The abundance of CEs (and the resulting ability to metabolize highly complex xylan) could give *D. mossii* a competitive advantage over other HGM members, which might not be as well equipped to metabolize highly decorated GAX. However, it is not clear if *D. mossii* acts in a "selfish" manner (importing decorated GAX oligosaccharides) or whether it secretes *Dm*CE1A, *Dm*CE1B, and *Dm*CE6A. Secretion of these enzymes could enable cross-feeding of other HGM members, which might support HGM composition stability. More research for example in the form of proteomics and co-culture studies is necessary to establish the interaction of *D. mossii* with the rest of the HGM community.

Multicatalytic enzymes are not limited to PULs. Vice versa not every PUL encodes multicatalytic enzymes. The two most common multicatalytic enzyme architectures encoded in PULs listed in the PULDB consists of multiple CE domains or multiple GH domains. Several examples of fused GH and polysaccharide lyase domains also exist in the PULDB (182). The general prevalence of multicatalytic enzymes throughout different PULs and different Bacteroidetes species implies that multicatalytic enzymes may play important roles during PUL concerted glycan degradation.

4.5. The holistic investigation of polysaccharide utilization loci is challenging

Since its beginnings marked by the discovery of the Sus in the early 90's by the Salyers lab, PUL research has come a long way (166,167). It became apparent that PULs are widespread among Bacteroidetes and that other phyla can harbor systems similar to Bacteroidetes-style PULs.

Today, several online resources exist that facilitate PUL research. One of them is dbCANPUL (http://bcb.unl.edu/dbCAN_PUL/), a database collecting information on experimentally validated CAZyme containing gene clusters (187). Another useful resource is the previously mentioned PULDB, which predicts PULs from published genome sequences based on a number of requirements, one of which is the presence of SusC/D homologues (182). The PULDB allows for a quantitative assessment of the PUL repertoire of Bacteroidetes species. Generally, Bacteroidetes species encode varying numbers of PULs. The model organisms *B. ovatus* and *B. thetaiotaomicron*

encode around 100 PULs each (185). Bacteroidetes may preferentially target some polysaccharides over others following a PUL priority order when multiple polysaccharides are encountered in their environment. While there is a strong correlation between number of PULs encoded in a species' genome and number of polysaccharides metabolized (185), the prediction of PULs is still limited by the extent of our current understanding of PULs. For instance, the PULDB lists many PULs containing multiple SusC/D-homologues and the role of this seeming redundancy remains unclear. Further, the PULDB also lists PULs only consisting of a single SusC/D homologue. These "lonely" SusC/D homologues could be involved in the degradation of non-carbohydrates or might be co-regulated with distantly placed CAZymes through yet unknown transcriptional regulation (188). Complicating things further are indications that PULs might not act entirely independently from their genetic context. For instance, it has been shown that PUL activation can lead to upregulation of genes involved in the formation of capsular polysaccharides (189). Further, it has been demonstrated that during PUL mediated degradation of porphyran (190,191) and carrageenan (192) not all activated genes are present in one locus.

The holistic characterization of PULs can contribute to fill these knowledge gaps and improve future predictions. However, thorough investigations of whole PULs are still rare (Table 4), likely due to the challenging blend of disciplines including genetics, enzymology, microbiology, transcriptomics, proteomics, and crystallography that have to be successfully applied together in order to build a comprehensive picture of the investigated PUL.

Table 4: Examples of comprehensive PUL characterization studies to date.

PUL	Species	Target	Characterization
XyGUL	*B. ovatus*	xyloglucan	Genetic and biochemical characterization of eight GHs and two SGBPs, including the structure of *Bo*GH5A (193).
*Bo*XylL and *Bo*XylS	*B. ovatus*	xylan	Genetic, biochemical, and structural characterization of various GHs and SGPB (112). Biochemical characterization of *Bo*CE6-CE1 (**paper I**).
*β*ManUL	*B. ovatus*	*β*-galactomannan	Genetic, biochemical, and structural characterization of *Bo*Gal36A (194), *Bo*Man26A, and *Bo*Man26B (195).
ChiUL	*F. johnsoniae*	chitin	PUL gene deletions, structures, and binding profiles of SusD-like proteins and biochemical characterization of chitinases ChiA and ChiB (124). Additional characterization of ChiA (196). Structural investigation of ChiA, ChiB, and GH20 (197).
MLGUL	*B. ovatus*	mixed-linkage *β*-glucans	Biochemical characterization of *Bo*GH16$_{MLG}$, *Bo*GH3$_{MLG}$, and BACOVA_02738 (GH3), including structure of *Bo*GH16$_{MLG}$ (198). Biophysical, structural, and *in vivo* characterization of *Bo*SGBP$_{MLG}$-A and *Bo*SGBP$_{MLG}$-B (199). Proof that MLGUL transfer also transfers MLG degradation ability (**paper IV**).
*α*ManUL1, *α*ManUL2, *α*ManUL3	*B. thetaiotaomicron*	*α*-mannan	Genetic and biochemical characterization of multiple PUL-encoded proteins supplemented with localization studies (200).
RG-II-UL	*B. thetaiotaomicron*	rhamnogalacturonan-II	Genetic, biochemical, and structural characterization of multiple PUL encoded proteins (201).

Naming of the PULs follows the suggested nomenclature in McKee *et al.* (188).

4.6. Transferring polysaccharide utilization loci is a promising but challenging method to answer new research questions

When establishing the specificity of a PUL three approaches are mainly used, often in combination. One is defining the PUL boundaries through identification of co-regulated genes by comparison of transcription levels when a species is grown on different glycan substrates (185). The second approach includes biochemical, and if possible structural, characterization of PUL encoded proteins, especially CAZymes,

as done in this thesis. The third approach focuses on the deletion of the investigated PUL or its parts in connection with growth studies. For example it was shown that the deletion of the PUL genes BACOVA_02738–02745 of *B. ovatus* led to an abolishment of growth on MLG (198).

Due to the numerous proteins they encode, PULs are typically very large (often 15 - 100 kb), which impedes their investigation and can limit genetic studies to deletions. While PUL or gene deletions are undoubtedly useful, they cannot address certain research questions. The introduction of PULs into Bacteroidetes species, for example by transferring PULs between strains, could however provide answers for the following questions:

- Does the transfer of a PUL likewise transfer a degradation ability?
- Do genetic context and evolutionary distance influence whether a PUL accepting species can utilize the donated PUL?
- Is the priority order of PULs conserved among different species?

The transfer of PULs between Bacteroidetes species is believed to occur naturally in the gastrointestinal tract of mammals as well as in other habitats. This hypothesis is based on the observed synteny, meaning highly similar gene organization, of PULs between closely and distantly related HGM species (112,124,185,193,198). Despite this, there are currently only two examples in the literature of successful transfers of PULs or larger PUL segments in a controlled environment (excluding **paper IV**). Speaking from personal experience, this may partly be because transferring PULs between different Bacteroidetes species is extremely difficult.

The first instance was only recently published (in 2018) and was part of a mouse study focusing on the potential of PULs for the engraftment of beneficiary strains into the gut microbiota (202). A rare porphyran utilization locus (PorUL) encoded by a *B. ovatus* strain, and likely received through lateral gene transfer from a marine organism, was transferred into *B. stercoris*. The same study also attempted to transfer the PorUL into *B. thetaiotaomicron*, but achieved only complete genomic integration for a shortened version of the PorUL (40 vs 60 kb), once more exemplifying the difficulty of controlled PUL transfers (202). The abundance of the resulting mutant strain in the mouse gut was shown to directly correlate with the amount of porphyran supplemented in the diet and by supplying its own metabolic niche in the form of porphyran the exogenous strain was able to assert itself (202). In the second instance an antibiotic-independent vector system that is based on inulin utilization was developed (203). Here, only a part of the full PUL was transferred which was also comparatively small to most full-length PULs. In both instances NBU2 integrase-expressing plasmids were applied. NBU2 is a mobilizable genetic element native to members of the *Bacteroides* genus. It allows for the integration at two specific *att* sites in the genome of *B. thetaiotaomicron*, which are depleted during integration (204,205).

4.7. The pICKUP method expands the toolbox for investigating polysaccharide utilization loci

The difficulties in recreating PUL transfers in a controlled environment make it obvious that the details of this process are still poorly understood. Therefore, it is all the more important to expand basic knowledge, introduce new tools, and generate more data related to PUL transfers. **Paper IV** of this thesis presents a new tool for the transfer of PULs between Bacteroidetes species. The pICKUP method relies on the equally named vector that was created within the studies leading to **paper IV**.

The "empty" pICKUP vector harbors various genetic elements allowing for propagation in three different hosts (*Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacteroides* species). In short, the yeast *S. cerevisiae* is used to assemble the pICKUP vector. *E. coli* is used to propagate the vector and conjugate it into the receiving strain (*Bacteroides*). A detailed overview of all genetic elements present in pICKUP as well as their respective functions can be found in **paper IV**. In addition to the vector backbone, the "loaded" pICKUP vector also contains the PUL from the donor strain and two target regions (TRs). TRs are PCR amplified regions homologous to the site of integration in the genome of the receiving strain.

In a first step, the restriction digested pICKUP backbone, the PCR amplified PUL (can be split into several parts if necessary due to large size), and the TRs are assembled in yeast, taking advantage of its remarkable ability to fuse DNA fragments via homologous recombination (Fig. 17). The assembled pICKUP-TR-PUL is then extracted from yeast and transformed into *E. coli*. Using *E. coli* as carrier of the pICKUP-TR-PUL entails a multitude of advantages:

- High plasmid concentrations of pICKUP-TR-PUL can be achieved.
- Straight-forward plasmid extraction resulting in high purity plasmid that is suitable for sequencing, and
- *E. coli* can conjugate with *Bacteroides*.

In a last step, *E. coli* harboring pICKUP-TR-PUL is co-cultured with the receiving *Bacteroides* strain, in which the plasmid is taken up and integrated into the genome via bacterial conjugation.

The pICKUP method has several advantages. Firstly, the site of integration into the genome of the PUL receiving strain can be chosen by selecting appropriate TRs. Selectable integration allows for more control over the PUL transfer process, limiting unfavorable gene disruption and enabling the transfer of multiple PULs, as there is no limit to integration sites as it is the case in systems that rely on *att* sites. Secondly, marker genes of the pICKUP plasmid are not present in the final mutant strain. The pICKUP vector includes a range of marker genes, some of which increase the fitness

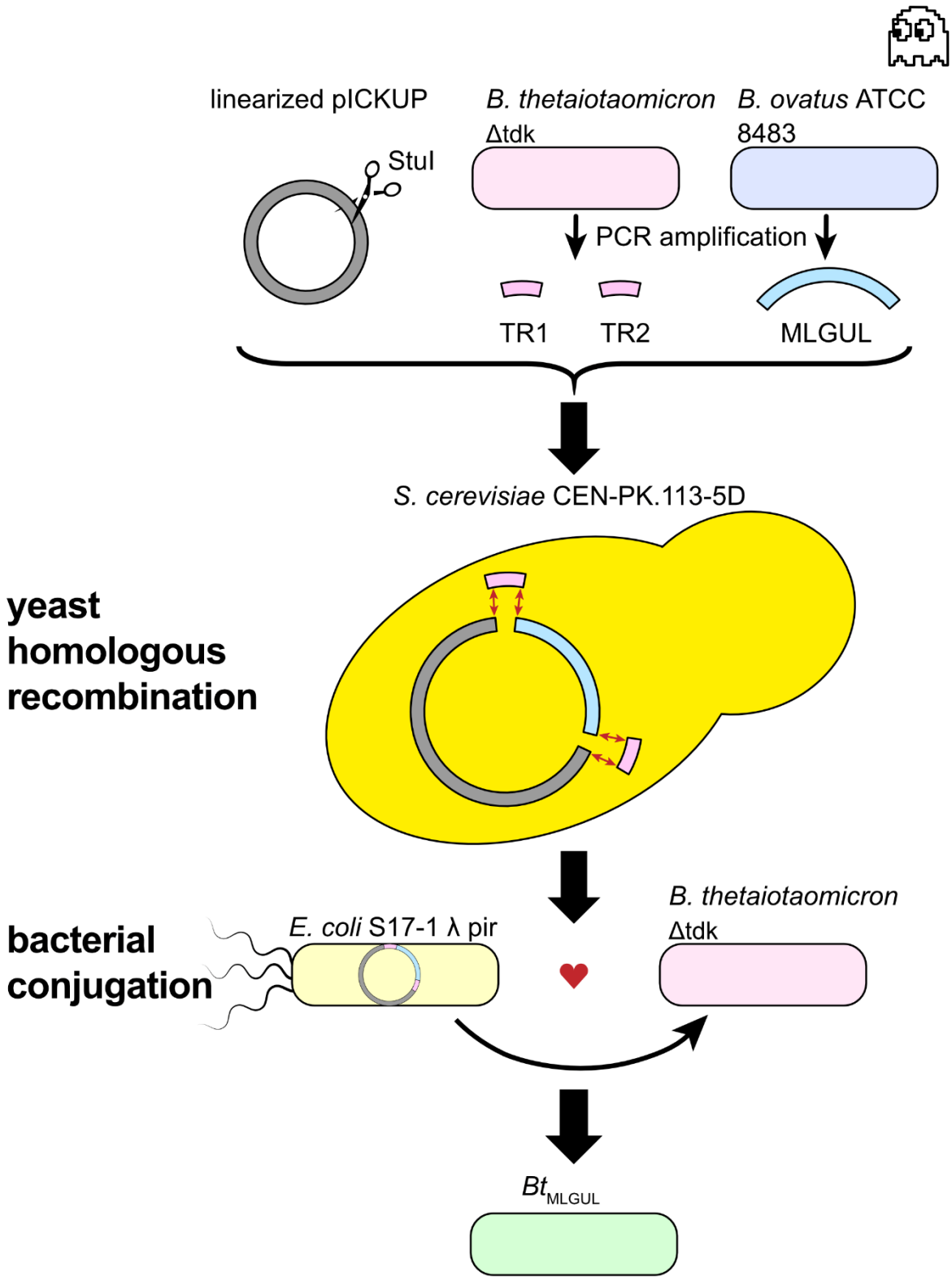


Figure 17: Schematic representation of PUL transfer steps. Enzymatically digested (StuI) pICKUP, PCR amplified target regions (TR1 and TR2) which are homologous to the site of integration of the acceptor strain, and the PCR amplified PUL (here mixed-linkage *β*-glucan utilization locus; MLGUL) from the donor strain are assembled in yeast by homologous recombination (marked by red arrows). The resulting pICKUP-TR-PUL vector is extracted from the yeast cells and transformed into *E. coli*, from which it is transferred into the genome of the acceptor strain by bacterial conjugation.

of the receiving strain (antibiotic resistance genes) and others crippling it (counterselection genes). The marker genes are crucial for reducing background and enabling the selection of:

- yeast assembled pICKUP plasmids that contain the PUL and TRs over empty pICKUP vectors.
- *E. coli* that harbors pICKUP-TR-PUL over *E. coli* without plasmid.
- *Bacteroides* that integrated pICKUP-TR-PUL into their genome over wild-type *Bacteroides*.
- *Bacteroides* that retained the PUL and expelled the pICKUP backbone.

Not permanently integrating marker genes into the genome is a big advantage, as the same markers can be used over and over to introduce PULs into the same strain, and do not promote the spread of antibiotic resistance.

An important parameter in PUL transfer research is transfer efficiency. It is not known how frequent PUL transfers are under physiological conditions. The fact that microbial cell densities are extremely high in the gut and that Bacteroidetes strains possess a characteristic PUL repertoire (while also sharing many PULs) might indicate that PUL transfer events in nature are relatively rare. It is also possible that the PUL repertoire of strains has a degree of plasticity, resulting in the periodical removal of PULs that are for example not frequently used or of low priority to the encoding strain. To date there is no research available supporting either of these hypotheses. Future studies, evaluating the impact of PUL size, PUL contents, evolutionary distance between donor and receiving strain, and conjugation conditions may help understand PUL transfer efficiency under physiological conditions and improve PUL transfer efficiency under controlled conditions.

4.8. Transfer of the mixed-linkage *β*-glucan utilization locus also transfers the associated phenotype

For the demonstration of the PUL transfer (**paper IV**), two of the most researched Bacteroidetes species belonging to the HGM, *B. thetaiotaomicron* and *B. ovatus*, were used (183,185). In earlier studies, the investigation of the PUL inventory of these species revealed that both encoded many highly similar PULs, and as a result both species were able to metabolize many of the same glycans (185). However, these two species still populate different ecological niches in the HGM, with *B. thetaiotaomicron* metabolizing certain pectins and host-derived *O*-glycans that are part of the host's mucus layer (170), while being unable to metabolize hemicellulose found in plant biomass. In contrast, *B. ovatus* favors hemicelluloses as substrates (185).

One kind of hemicellulose metabolized by *B. ovatus* is MLG. MLG is a relatively simple polysaccharide composed of *β*-1,3 and *β*-1,4 linked glucose units and is typically found in cereals (206). The MLGUL of *B. ovatus* was associated through RNA sequencing and deletion studies with the bacterium's ability to degrade MLG (Table 4). When the MLGUL was deleted from its genome, *B. ovatus* lost the ability to grow on MLG, indicating that this PUL is central for the ability of *B. ovatus* to metabolize this hemicellulose (198). *B. thetaiotaomicron* is unable to grow on MLG.

In **paper IV** the MLGUL of *B. ovatus* ATCC 8483 was transferred into the genome of *B. thetaiotaomicron* Δtdk using the previously described pICKUP method. The resulting strain Bt_{MLGUL} gained the ability to grow on MLG. Bt_{MLGUL} grew to comparable optical densities as *B. ovatus* on MLG, albeit displaying a prolonged lag-phase (Fig. 18). The MLGUL transfer showcases that the developed pICKUP method can be applied to successfully transfer full PULs. Further, it proves that the sole transfer of the MLGUL conferred the MLG metabolizing phenotype. The observed lag-phase may result from processes influencing the ability of Bt_{MLGUL} to utilize the MLGUL, such as PUL priority and/or regulation. These yet unidentified processes and interactions might be encoded in other parts of the Bacteroidetes genome and thus may differ in Bt_{MLGUL} from *B. ovatus*. More research is necessary to evaluate the mechanisms causing the observed lag-phase. Likewise, it remains unclear if all PULs transfer complete degradation abilities, as was observed in the case for the MLGUL transfer and MLG degradation.

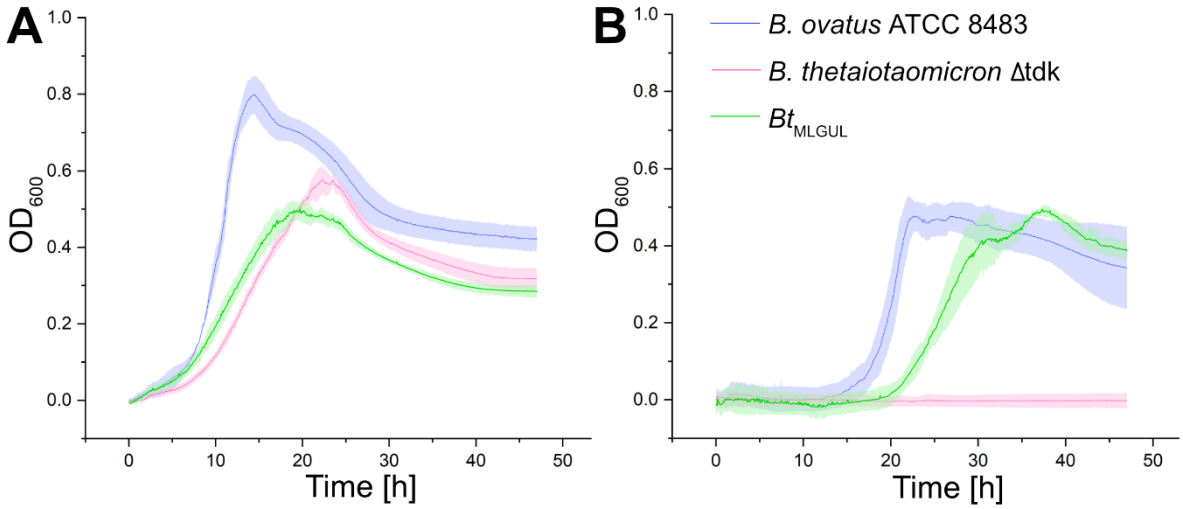


Fig. 18: Comparative growth on minimal medium supplemented with different carbon sources. The figure was taken from paper IV. **(A)** *B. ovatus* ATCC 8483 (blue), *B. thetaiotaomicron* Δtdk (pink), and Bt_{MLGUL} (green) were grown on glucose. **(B)** The wildtype *B. ovatus*, but not *B. thetaiotaomicron*, exhibited growth on mixed-linkage *β*-glucan (MLG). By receiving the mixed-linkage *β*-glucan utilization locus (MLGUL) from *B. ovatus*, Bt_{MLGUL} gained the ability to grow on MLG. Data are presented as the average of three replicates with standard errors of the mean. The optical density (OD) was measured at 600 nm.

4.8. Main points of chapter 4

- Bacteroidetes metabolize complex polysaccharides found in dietary fiber with the help of polysaccharide utilization loci (PULs).
- PULs encode proteins necessary to sense, bind, and degrade a specific polysaccharide or polysaccharide motif.
- Transferring PULs is a crucial step to address many PUL-related research questions, but is very challenging.
- The pICKUP method is a genetic tool developed for the site-directed transfer of PULs between species, using a combination of yeast homologous recombination and bacterial conjugation. Resulting mutants only retain PUL parts but no selection markers.
- The pICKUP method enabled the transfer of the mixed-linkage *β*-glucan utilization locus from *B. ovatus* into *B. thetaiotaomicron* Δtdk, which also transferred the corresponding mixed-linkage *β*-glucan metabolizing phenotype.

5. Outlook

This chapter first discusses current prospects and limitations of CAZymes, with a focus on CEs, and gives an outlook on future research. Then, the limitations and advantages of the pICKUP method, which was developed as part of this thesis, are considered. Lastly, the relevance and potential of PUL research for the advancement of basic knowledge and future applications is discussed.

To date, thousands of CAZymes have been studied resulting in the publication of vast amounts of biochemical and structural data. Despite that, our understanding of CAZymes is far from complete. Enzyme families such as CE1 demonstrate that CAZymes match the complexity and heterogeneity of their lignocellulose substrates with an equal high variety in activities and structures. The exploration and characterization of CAZymes is often aggravated by the availability of model substrates and well-defined natural substrates. The work underlying this thesis also suffered to a certain degree from such limitations, as for example the corn cob biomass that was utilized in xylanase assays is known to be rich in GAX (109), but its exact chemical structure is not known and might additionally vary from batch to batch. Another challenge in enzyme characterization is the acquisition of good structural data as many proteins simply cannot be crystallized (with reasonable efforts).

Nonetheless, further exploration of the depth of enzyme activities found among CAZymes will eventually enable us to reliably predict exact enzyme activities from protein sequences and structures. Such knowledge will also be useful for enzyme modelling and the creation of artificial enzymes.

In the past CEs have not received as much attention as enzymes targeting glycosidic linkages in xylan. However, the ability of CEs to target xylan decorations and facilitate the action of GH enzymes such as xylanases as demonstrated in this thesis, makes members of this enzyme class interesting targets for further studies focusing on the elucidation of enzymatic biomass degradation.

Under physiological conditions enzymatic biomass degradation is not performed by a single enzyme, but by a mix of enzymes that influence each other. As such, studies investigating the interaction of multiple CAZymes working together are becoming increasingly important. With this in mind, multicatalytic enzymes are especially interesting as their enzyme architectures may give pointers as to what enzyme activities work well together and should be combined or even fused. Similarly, the CAZymes encoded in PULs exemplify enzyme activities needed to degrade a specific polysaccharide motif and may inspire future enzyme cocktail compositions. The pICKUP method may facilitate such research.

In this thesis the pICKUP method was developed to expand the tools available for the investigation of PULs by facilitating the transfer of whole PULs, PUL parts or large

DNA fragments in general. The functionality of the pICKUP method was demonstrated by transferring the MLGUL of *B. ovatus* into *B. thetaiotaomicron* Δtdk. The MLGUL is a relatively small PUL "only" spanning 8 genes (locus tags BACOVA_02739 to BACOVA_02747) corresponding to a length of 16 kb. While the MLGUL transfer is sufficient for a proof of concept study, more research is needed to establish whether the pICKUP method also facilitates the transfer of larger PULs.

A current limitation of the pICKUP method is that counterselection against reverted wildtype mutants relies on the absence of thymidine kinases in the genome of the receiving Bacteroidetes strain. This is also the reason why *B. thetaiotaomicron* Δtdk was utilized in this study. Future substitution of this selection process may increase flexibility in picking the PUL receiving Bacteroidetes strain. Further, the cloning of the PUL in chunks suitable for PCR amplification can be quite labor intensive. An overhauled pICKUP method may replace PCR amplified PUL parts with target regions homologous to the ends of the PUL. In theory, yeast should then be able pick up the PUL from sheared genomic DNA during vector assembly.

The pICKUP method has two big advantages. Firstly, the integration site in the genome of the PUL receiving strain is selectable. This level of control ensures that PUL integration does not interrupt genes if not desired. Further, multiple PULs may be introduced and the PULs can be placed in sequence or scattered throughout the genome of the receiving strain. Secondly, final mutants only carry the PUL, but no parts of the pICKUP backbone, thus no marker genes such as antibiotic resistances are accumulated. These characteristics are particularly useful when building up a desired PUL repertoire in a receiving strain.

Full or partial PUL transfers can advance our understanding of PUL mediated glycan degradation. For example, the transfer of multiple PULs from Bacteroidetes strain A to Bacteroidetes strain B coupled with growth studies on mixed glycans may give insights into whether PUL priority is conserved among Bacteroidetes species. Another example are studies involving the transfer of the same PUL into increasingly distantly related Bacteroidetes species. Such studies could help establish if evolutionary distance and genetic context (interaction of the integrated PUL with the receptor genome) are important factors influencing the ability of a Bacteroidetes strain to accept and/or utilize a foreign PUL. Lastly, the transfer of PULs from little-studied species into established model organisms such as *B. ovatus* or *B. thetaiotaomicron* may simplify and accelerate the genetic and biochemical characterization of PULs or PUL parts including encoded CAZymes.

A stepwise tailoring of PUL repertoires enables the creation of "designer" strains, which could be engineered to degrade a specific biomass feed composition. Such strains could find application in wastewater treatment facilities and biorefineries, where they may improve biomass hydrolysis. Bacteroidetes are known to secrete SCFAs like succinic acid in the human gut (7). A combination of Bacteroidetes

mediated biomass degradation with the microbial production of platform chemicals (for example succinic acid) could improve the economic feasibility of biorefineries, which are central for the implementation of a bioeconomy. Lastly, designer strains may also be tailored to support the gut health of humans and livestock in the form of probiotic strains. The potential of rare PULs for long-term engraftment of Bacteroidetes species into the HGM has already been demonstrated (202).

6. Conclusions

This chapter evaluates whether the aims stated in chapter 1 section 7 have been reached. The first aim was to investigate the role of PUL encoded CEs. To address this aim, various CEs encoded in different PULs from the Bacteroidetes species *B. ovatus*, *F. johnsoniae*, *D. mossii*, and *B. eggerthii* were biochemically characterized. Investigated CE1 and CE6 enzymes improved xylanase mediated xylan-rich biomass hydrolysis indicating that these enzymes support the activity of PUL encoded xylanases (**papers I & II**). Additionally, several of the investigated CEs demonstrated characteristics that highlight the potential of PUL encoded CEs for enzyme discovery. For example, the enzyme *Dm*CE6A was shown to be highly active and showcases that PUL encoded CEs may be a valuable resource for the exploration of industrially relevant enzymes (**paper II**). The multicatalytic enzyme *Be*CE15A-Rex8A harbors a rare and novel enzyme architecture. While the exact target of this enzyme could not be determined, the existence of *Be*CE15A-Rex8A is an indication for the existence of yet unknown structural motifs in plant biomass. Thus, PUL encoded CEs may also advance our basic understanding of enzymatic plant biomass degradation and plant biomass structures.

The second aim was to investigate the occurrence of multiple CEs in the same PUL. Predicted PUL 17 of *D. mossii* encodes three CE enzymes (including one multicatalytic CE), which were all biochemically characterized revealing that each CE domain had a distinct activity profile on model and native substrates. This suggests that each CE fulfills an individual role during the degradation of the xylan motif that is targeted by the encoding PUL (**paper II**).

The third aim was to investigate PUL encoded multicatalytic CEs and find a possible explanation for this enzyme architecture. Studies on corn cob biomass revealed that supplementing the multicatalytic enzyme *Bo*CE6-CE1 in its full-length form yielded a higher boost to xylanase activity than when supplemented in the form of an equimolar mix of the N- and C-terminal domains (**paper I**). Possibly, the combination of several CE activities into one multicatalytic enzyme allows for the degradation of complex motifs within xylan decorations that cannot be efficiently degraded by single enzyme activities. The reason behind this may be close proximity of the catalytic domains in multicatalytic CEs, which is otherwise not guaranteed under physiological conditions. To the best of my knowledge, *Bo*CE6-CE1 is the first reported case of intramolecular synergy for multicatalytic CEs.

The fourth aim of this thesis was to develop a method to transfer PULs between different species to expand the available genetic toolbox. A combination of yeast homologous recombination and bacterial conjugation was applied to create the pICKUP method, which allows for site-directed integration of gene cassettes, for example PULs, into the genome of the receiving Bacteroidetes strain (**paper IV**). The pICKUP method was successfully applied by transferring the MLGUL from

B. ovatus ATCC 8483 into *B. thetaiotaomicron* Δtdk, resulting in the creation of the strain Bt_{MLGUL}.

Comparative growth studies on mixed-linkage *β*-glucan using *B. ovatus* ATCC 8483, *B. thetaiotaomicron* Δtdk, and Bt_{MLGUL} showed that the transfer of the MLGUL initiated an MLG metabolizing phenotype (**paper IV**). This experiment addressed the fifth aim, which was to prove that the sole transfer of a PUL also transfers the degradation ability conferred by the respective PUL.

The overall aim of this thesis was to explore the potential of the bacterial phylum Bacteroidetes for the degradation of plant biomass. I showed that Bacteroidetes and PUL encoded CEs (some of which are multicatalytic) are active on lignocellulosic material. Some of the investigated CEs exhibited favorable activity profiles with potential for industrial application (for example *Dm*CE6A), and harbored little studied enzyme architectures advancing our basic understanding of cooperative enzymatic hydrolysis (*Be*CE15A-Rex8A). Besides their CAZymes, Bacteroidetes are known to gain many of their biomass degradation capabilities from the PULs they encode. While the characterization of PULs themselves beyond encoded CEs was not part of this thesis, the pICKUP method was developed to facilitate future PUL research. Overall, this thesis further supports the Bacteroidetes phylum being a rich and important resource for research and applications related to microbial biomass degradation.

Acknowledgments

During my PhD I had the pleasure to visit the University of Michigan. Thank you to Matt, Nicole, and Eric who were so welcoming and supportive, and made my research stay not only productive but also a really fun experience I remember fondly.

Thank you to Jenny, Julia L, Julia W, Thenia, Helén, Vijay, Pun, Gunilla, Anne-Lise, Erica, and Barbro who silently worked in the background, so I didn't have to deal with many lab organization, administration, and accounting related tasks.

I am a firm believer in sharing knowledge. Thank you to everyone that goes through the extra (often financial) trouble of publishing open access. On a completely unrelated note I would also like to thank Alexandra Elbakyan.

Research is very much a collaborative effort and during my PhD I had the pleasure of working together with many excellent scientists. My deepest gratitude goes to my collaborators: Nathan, Dan, Scott, Tomke, Alex, Cyrielle, Lisbeth, Johan, Lauren, Marina, Otto, Gisela, Phil, Eric, and Nicole. Thank you for contributing to this thesis!

Special thanks go to Alex, who worked so tirelessly on squeezing out usable data from the NMR experiments. Working together with people from different fields has its pitfalls, but you made it so easy. Thank you!

Tomke, you have been an amazing master student. Thank you for making teaching you and working with you a throughout pleasant and rewarding experience. I know you are going to do great in your PhD, too.

Nathan, you and me are probably the only two people who truly understand just how challenging the PUL transfer project really is. Thank you for being such an exceptionally uncomplicated collaborator on such a difficult project. Your patience, kindness, and sharp mind were and always will be appreciated.

Dan, as you can see I didn't go with your suggestion of just writing "Thank you Dan" and nothing else for my acknowledgments… I have to admit that some of your suggestions can be pretty amazing though. Thank you for being a great colleague, collaborator, and most of all a really awesome friend.

Scott, we are going way back. I remember you showing me around the lab in my first week as a PhD student. 4+ years later we still work together and now we have our names on three shared publications. Thank you for being this amazing blend of capable scientist, patient teacher, and cheerful friend.

Writing a PhD thesis is a big task. Thank you Vijay, Amanda, Scott, Nathan, Tom, Calle, Lisbeth, and Johan for helping making this thesis better by giving constructive feedback.

I would also like to express my gratitude towards my PhD thesis committee members Henrik Stålbrand, Jean-Guy Berrin, Jens Eklöf, Cecilia Geijer, and my opponent Birte Svensson, who agreed to partake in my PhD defense.

Special thanks go also to Calle and Lisbeth for being my examiner and co-supervisor, respectively. Thank you for supporting me and giving me feedback throughout this journey. I feel that I have grown a lot during my PhD years and you have definitely fostered this process.

Johan, thank you for being an amazing supervisor. I am grateful that you were consistently open, trusting, and understanding towards me. You always made time for me and your feedback was always top-notch. You pushed me to do better and also never failed to appreciate a good joke. I am glad you are my “Doktorvater”.

People come and go as we move locations to grow and follow our ambitions, but the memories stay. Thank you to everyone in Indbio for contributing to a great working environment and creating good memories. Special thanks for being part of my PhD journey go to Ben, Avlant, Josh, Helén, Paulo, Marlous, Eric MB, Pia, Erik, Wiebke, Bene, Veronica G, Max, Mauri, Monika, Dan, David, Scott, Nathan, Amanda, Jonas, Rike, Silfa, Duncan, Pawel, Tom, Kamesh, Pun, Veronica B, and Fei.

Ich möchte mich auch bei meiner ganzen Familie bedanken. Einen Doktor zu machen dauert lange und in den Jahren die ich im Ausland gelebt habe, habe ich so manches Familientreffen verpasst und war physisch abwesend bei vielen Geburtstagsparties. Trotzdem habt ihr nie das Interesse an mir verloren und unsere Verbindung ist immer noch stark. Oma und Opa, danke dass ihr mich immer angefeuert habt, auch wenn ich vielleicht ein bisschen mehr in der Weltgeschichte rumtingel als euch lieb ist. Mama, Papa und Jani, ihr habt vielleicht nicht den hardcore naturwissenschaftlichen Hintergrund um die Details meiner Forschung zu verstehen, aber ihr habt immer mich verstanden. Danke, dass ihr immer für mich da seid.

References

1. Schopf, J. W., Kitajima, K., Spicuzza, M. J., Kudryavtsev, A. B., and Valley, J. W. (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proceedings of the National Academy of Sciences* **115**, 53-58
2. Saccá, M. L., Caracciolo, A. B., Di Lenola, M., and Grenni, P. (2017) Ecosystem services provided by soil microorganisms. in *Soil Biological Communities and Ecosystem Resilience* (Lukac, M., Grenni, P., and Gamboni, M. eds.), Springer. pp 9-24
3. Jacoby, R., Peukert, M., Succurro, A., Koprivova, A., and Kopriva, S. (2017) The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Frontiers in plant science* **8**, 1617
4. Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., and Gordon, J. I. (2008) Evolution of mammals and their gut microbes. *Science* **320**, 1647-1651
5. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005) Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638
6. Porter, N. T., and Martens, E. C. (2017) The critical roles of polysaccharides in gut microbial ecology and physiology. *Annual Review of Microbiology* **71**, 349-369
7. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R., and White, B. A. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbioly* **6**, 121-131
8. Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host & Microbe* **4**, 447-457
9. Ferreira, T. M., Leonel, A. J., Melo, M. A., Santos, R. R., Cara, D. C., Cardoso, V. N., Correia, M. I., and Alvarez-Leite, J. I. (2012) Oral supplementation of butyrate reduces mucositis and intestinal permeability associated with 5-fluorouracil administration. *Lipids* **47**, 669-678
10. Burger-van Paassen, N., Vincent, A., Puiman, P. J., van der Sluis, M., Bouma, J., Boehm, G., Van Goudoever, J. B., Van Seuningen, I., and Renes, I. B. (2009) The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: implications for epithelial protection. *Biochemical Journal* **420**, 211-219
11. Tan, J., McKenzie, C., Potamitis, M., Thorburn, A. N., Mackay, C. R., and Macia, L. (2014) The role of short-chain fatty acids in health and disease. *Advances in Immunology* **121**, 91-119
12. Tripathy, D. B., Yadav, A., and Mishra, A. (2017) Applications of petrochemicals: a mini review. *Recent Advances in Petrochemical Science* **2**, 67-69
13. Nikolau, B. J., Perera, M. A. D., Brachova, L., and Shanks, B. (2008) Platform biochemicals for a biorenewable chemical industry. *The Plant Journal* **54**, 536-545
14. Bozell, J. J., and Petersen, G. R. (2010) Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chemistry* **12**, 539-554
15. BP. (2020) BP statistical review of energy 2020. Accessed on 24.07.2021. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/xlsx/energy-economics/statistical-review/bp-stats-review-2020-all-data.xlsx
16. Lee, R. A., and Lavoie, J.-M. (2013) From first-to third-generation biofuels: challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers* **3**, 6-11
17. Cheng, J. J., and Timilsina, G. R. (2011) Status and barriers of advanced biofuel technologies: a review. *Renewable Energy* **36**, 3541-3549

18. United Nations. (2016) Second generation biofuel markets: state of play, trade and developing country perspectives. Accessed on 24.07.2021. https://unctad.org/system/files/official-document/ditcted2015d8_en.pdf
19. International Energy Agency. (2018) Fuels & Technologies. Accessed on 24.07.2021. https://www.iea.org/fuels-and-technologies/bioenergy
20. Grishkewich, N., Mohammed, N., Tang, J., and Tam, K. C. (2017) Recent advances in the application of cellulose nanocrystals. *Current Opinion in Colloid & Interface Science* **29**, 32-45
21. Norgren, M., and Edlund, H. (2014) Lignin: recent advances and emerging applications. *Current Opinion in Colloid & Interface Science* **19**, 409-416
22. Vishtal, A. G., and Kraslawski, A. (2011) Challenges in industrial applications of technical lignins. *BioResources* **6**, 3547-3568
23. Lora, J. H., and Glasser, W. G. (2002) Recent industrial applications of lignin: a sustainable alternative to nonrenewable materials. *Journal of Polymers and the Environment* **10**, 39-48
24. Ebringerová, A. (2005) Structural diversity and application potential of hemicelluloses. in *Macromolecular Symposia*, Wiley Online Library
25. Sanderson, K. (2011) Lignocellulose: a chewy problem. *Nature* **474**, S12-S14
26. Shrotri, A., Kobayashi, H., and Fukuoka, A. (2017) Catalytic conversion of structural carbohydrates and lignin to chemicals. *Advances in Catalysis* **60**, 59-123
27. Balan, V. (2014) Current challenges in commercially producing biofuels from lignocellulosic biomass. *International Scholarly Research Notices Biotechnology* **2014**, 1-31
28. Hudiburg, T. W., Wang, W., Khanna, M., Long, S. P., Dwivedi, P., Parton, W. J., Hartman, M., and Delucia, E. H. (2016) Impacts of a 32-billion-gallon bioenergy landscape on land and fossil fuel use in the US. *Nature Energy* **1**, 1-7
29. Sannigrahi, P., Ragauskas, A. J., and Tuskan, G. A. (2010) Poplar as a feedstock for biofuels: a review of compositional characteristics. *Biofuels, Bioproducts and Biorefining* **4**, 209-226
30. Kim, S., and Dale, B. E. (2004) Global potential bioethanol production from wasted crops and crop residues. *Biomass and Bioenergy* **26**, 361-375
31. Patel, A., and Shah, A. (2021) Integrated lignocellulosic biorefinery: gateway for production of second generation ethanol and value added products. *Journal of Bioresources and Bioproducts* **6**, 108-128
32. Pauly, M., and Keegstra, K. (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal* **54**, 559-568
33. Sjostrom, E. (1993) *Wood chemistry: fundamentals and applications*, Gulf professional publishing
34. Blanchette, R. A., Obst, J. R., and Timell, T. E. (1994) Biodegradation of compression wood and tension wood by white and brown rot fungi. *Holzforschung* **48**, 34-42
35. McDonald, A., and Donaldson, L. (2001) Constituents in wood. *Encyclopedia of Materials: Science and Technology. Elsevier Science*, 9612-9616
36. Duchesne, L. C., and Larson, D. (1989) Cellulose and the evolution of plant life. *Bioscience* **39**, 238-241
37. Daniel, G. (2009) Wood and fibre morphology. in *Wood chemistry and wood biotechnology* (Ek, M., Gellerstedt, G., and Henriksson, G. eds.), de Gruyter. pp 45-70
38. Gibson, L. J. (2012) The hierarchical structure and mechanics of plant materials. *Journal of the Royal Society Interface* **9**, 2749-2766
39. Zhong, R., and Ye, Z.-H. (2015) Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant and Cell Physiology* **56**, 195-214

40. Nishimura, H., Kamiya, A., Nagata, T., Katahira, M., and Watanabe, T. (2018) Direct evidence for α ether linkage between lignin and carbohydrates in wood cell walls. *Scientific Reports* **8**, 1-11
41. Demirbaş, A. (2001) Relationships between lignin contents and heating values of biomass. *Energy Conversion and Management* **42**, 183-188
42. Wang, H., Male, J., and Wang, Y. (2013) Recent advances in hydrotreating of pyrolysis bio-oil and its oxygen-containing model compounds. *ACS Catalysis* **3**, 1047-1070
43. Klemm, D., Heublein, B., Fink, H. P., and Bohn, A. (2005) Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition* **44**, 3358-3393
44. Kumar, R., Mago, G., Balan, V., and Wyman, C. E. (2009) Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. *Bioresource Technology* **100**, 3948-3962
45. Hallac, B. B., and Ragauskas, A. J. (2011) Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol. *Biofuels, Bioproducts and Biorefining* **5**, 215-225
46. Henriksson, G., and Lennholm, H. (2009) Cellulose and carbohydrate chemistry. in *Wood Chemistry and Wood Biotechnology* (Ek, M., Gellerstedt, G., and Henriksson, G. eds.), De Gruyter. pp 71-100
47. Horn, S. J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. (2012) Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels* **5**, 1-13
48. Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. *Annual Review of Plant Biology* **61**, 263-289
49. Ebringerová, A., and Heinze, T. (2000) Xylan and xylan derivatives–biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. *Macromolecular rapid communications* **21**, 542-556
50. Salmon, S., and Hudson, S. M. (1997) Crystal morphology, biosynthesis, and physical assembly of cellulose, chitin, and chitosan. *Journal of Macromolecular Science, Part C: Polymer Reviews* **37**, 199-276
51. Jeuniaux, C., Voss-Foucart, M.-F., Poulicek, M., and Bussers, J.-C. (1988) Sources of chitin, estimated from new data on chitin biomass and production. *Chitin and Chitosan*, 3-11
52. Moreira, L. (2008) An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology* **79**, 165-178
53. Capek, P., Alföldi, J., and Lišková, D. (2002) An acetylated galactoglucomannan from Picea abies L. Karst. *Carbohydrate Research* **337**, 1033-1037
54. Kozioł, A., Cybulska, J., Pieczywek, P. M., and Zdunek, A. (2015) Evaluation of structure and assembly of xyloglucan from tamarind seed (*Tamarindus indica* L.) with atomic force microscopy. *Food Biophysics* **10**, 396-402
55. Nishinari, K., Takemasa, M., Zhang, H., and Takahashi, R. (2007) Storage plant polysaccharides: xyloglucans, galactomannans, glucomannans. in *Comprehensive glycoscience* (Kamerling, H. ed.), Elsevier. pp 613-652
56. Hayashi, T. (1989) Xyloglucans in the primary cell wall. *Annual review of plant biology* **40**, 139-168
57. Kooiman, P. (1957) Amyloids of plant seeds. *Nature* **179**, 107-109
58. Lazaridou, A., Biliaderis, C., Micha-Screttas, M., and Steele, B. R. (2004) A comparative study on structure–function relations of mixed-linkage (1→ 3),(1→ 4) linear β-d-glucans. *Food Hydrocolloids* **18**, 837-855
59. Bengtsson, S., Åman, P., Graham, H., Newman, C. W., and Newman, R. K. (1990) Chemical studies on mixed-linked β-glucans in hull-less barley cultivars giving different hypocholesterolaemic responses in chickens. *Journal of the Science of Food and Agriculture* **52**, 435-445

60. Ojima, T. (2013) Polysaccharide-degrading enzymes from herbivorous marine invertebrates. in *Marine Enzymes for Biocatalysis* (Trincone, A. ed.), Elsevier. pp 333-371
61. Silveira, R. L., Stoyanov, S. R., Gusarov, S., Skaf, M. S., and Kovalenko, A. (2013) Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale forces that control cell wall strength. *Journal of the American Chemical Society* **135**, 19048-19051
62. Li, C., Zhao, X., Wang, A., Huber, G. W., and Zhang, T. (2015) Catalytic transformation of lignin for the production of chemicals and fuels. *Chemical Reviews* **115**, 11559-11624
63. Argyropoulos, D. S., and Menachem, S. B. (1998) Lignin. in *Biopolymers from Renewable Resources* (Kaplan, D. ed.), Springer. pp 292-322
64. Tarasov, D., Leitch, M., and Fatehi, P. (2018) Lignin-carbohydrate complexes: properties, applications, analyses, and methods of extraction: a review. *Biotechnology for Biofuels* **11**, 1-28
65. Zhao, X., Qi, F., and Liu, D. (2017) Hierarchy nano-and ultrastructure of lignocellulose and its impact on the bioconversion of cellulose. in *Nanotechnology for Bioenergy and Biofuel Production* (Rai, M., and da Silva, S. S. eds.), Springer. pp 117-151
66. Virgili, F., Pagana, G., Bourne, L., Rimbach, G., Natella, F., Rice-Evans, C., and Packer, L. (2000) Ferulic acid excretion as a marker of consumption of a French maritime pine (*Pinus maritima*) bark extract. *Free Radical Biology and Medicine* **28**, 1249-1256
67. Reiter, J., Strittmatter, H., Wiemann, L. O., Schieder, D., and Sieber, V. (2013) Enzymatic cleavage of lignin *β-O*-4 aryl ether bonds via net internal hydrogen transfer. *Green Chemistry* **15**, 1373-1381
68. de O. Buanafina, M. M. (2009) Feruloylation in grasses: current and future perspectives. *Molecular Plant* **2**, 861-872
69. Mualikrishna, G., and Tharanathan, R. N. (1994) Characterization of pectic polysaccharides from pulse husks. *Food Chemistry* **50**, 87-89
70. Flutto, L. (2003) Pectin: roperties and determination. in *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (Caballero, B. ed.), Academic Press, Oxford. pp 4440-4449
71. Khan, M., and Omoloso, A. (2003) Antimicrobial activity of extractives of *Sarcocephalus coadunatus*. *Fitoterapia* **74**, 695-698
72. Wilkie, K. C. (1979) The hemicelluloses of grasses and cereals. in *Advances in Carbohydrate Chemistry and Biochemistry* (Tipson, R. S., and Horton, D. eds.), Elsevier. pp 215-264
73. Teleman, A., Larsson, P. T., and Iversen, T. (2001) On the accessibility and structure of xylan in birch kraft pulp. *Cellulose* **8**, 209-215
74. Matsuo, N., Yoshida, S., Kusakabe, I., and Murakami, K. (1991) Chemical structure of xylan in cotton-seed cake. *Agricultural and Biological Chemistry* **55**, 2905-2907
75. Yoshida, S., Kuno, A., Saito, N., Aoyama, M., and Kusakabe, I. (1998) Structure of xylan from culms of bamboo grass (*Sasa senanensis* Rehd.). *Journal of Wood Science* **44**, 457-462
76. Shibuya, N., Misaki, A., and Iwasaki, T. (1983) The structure of arabinoxylan and arabinoglucuronoxylan isolated from rice endosperm cell wall. *Agricultural and Biological Chemistry* **47**, 2223-2230
77. Kato, Y., and Nevins, D. J. (1985) Isolation and identification of *O*-(5-*O*-feruloyl-*α*-L-arabinofuranosyl)-1(→3)-*O-β*-D-xylopyranosyl-(1→4)-d-xylopyranose as a component of *Zea shoot* cell-walls. *Carbohydrate Research* **137**, 139-150
78. Markwalder, H. U., and Neukom, H. (1976) Diferulic acid as a possible crosslink in hemicelluloses from wheat germ. *Phytochemistry* **15**, 836-837
79. Balakshin, M., Capanema, E., and Berlin, A. (2014) Isolation and analysis of lignin–carbohydrate complexes preparations with traditional and advanced methods: a review. *Studies in Natural Products Chemistry* **42**, 83-115

80. Bajpai, P. (2018) *Wood and fiber fundamentals*, Elsevier
81. de Carvalho, D. M., Martínez-Abad, A., Evtuguin, D. V., Colodette, J. L., Lindström, M. E., Vilaplana, F., and Sevastyanova, O. (2017) Isolation and characterization of acetylated glucuronoarabinoxylan from sugarcane bagasse and straw. *Carbohydrate Polymers* **156**, 223-234
82. Sun, R., Lawther, J. M., and Banks, W. B. (1998) Isolation and characterization of hemicellulose B and cellulose from pressure refined wheat straw. *Industrial Crops and Products* **7**, 121-128
83. Bazus, A., Rigal, L., Gaset, A., Fontaine, T., Wieruszeski, J.-M., and Fournet, B. (1993) Isolation and characterisation of hemicelluloses from sunflower hulls. *Carbohydrate Research* **243**, 323-332
84. Ebringerová, A., Hromadkova, Z., Alföldi, J., and Berth, G. (1992) Structural and solution properties of corn cob heteroxylans. *Carbohydrate Polymers* **19**, 99-105
85. Coimbra, M. A., Waldron, K. W., and Selvendran, R. R. (1994) Isolation and characterisation of cell wall polymers from olive pulp (*Olea europaea* L.). *Carbohydrate Research* **252**, 245-262
86. Doner, L. W., and Hicks, K. B. (1997) Isolation of hemicellulose from corn fiber by alkaline hydrogen peroxide extraction. *Cereal Chemistry* **74**, 176-181
87. Capek, P., Šutovská, M., Barboríková, J., Kazimierová, I., Fraňová, S., and Kopačová, M. (2020) Structural characterization and anti-asthmatic effect of *α*-l-arabino (4-*O*-methyl-*α*-d-glucurono)-*β*-d-xylan from the roots of *Rudbeckia fulgida*. *International Journal of Biological Macromolecules* **165**, 842-848
88. Timell, T. (1967) Recent progress in the chemistry of wood hemicelluloses. *Wood Science and Technology* **1**, 45-70
89. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**, D490-D495
90. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242
91. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853-859
92. Lee, C. C., Kibblewhite-Accinelli, R. E., Wagschal, K., Robertson, G. H., and Wong, D. W. (2006) Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles* **10**, 295-300
93. Naumoff, D. (2011) Hierarchical classification of glycoside hydrolases. *Biochemistry (Moscow)* **76**, 622-635
94. Honda, Y., and Kitaoka, M. (2004) A family 8 glycoside hydrolase from *Bacillus halodurans* C-125 (BH2105) is a reducing end xylose-releasing exo-oligoxylanase. *Journal of Biological Chemistry* **279**, 55097-55103
95. Lagaert, S., Van Campenhout, S., Pollet, A., Bourgois, T. M., Delcour, J. A., Courtin, C. M., and Volckaert, G. (2007) Recombinant expression and characterization of a reducing-end xylose-releasing exo-oligoxylanase from *Bifidobacterium adolescentis*. *Applied and Environmental Microbiology* **73**, 5374-5377
96. Hong, P. Y., Iakiviak, M., Dodd, D., Zhang, M., Mackie, R. I., and Cann, I. (2014) Two new xylanases with different substrate specificities from the human gut bacterium *Bacteroides intestinalis* DSM 17393. *Applied and Environmental Microbiology* **80**, 2084-2093
97. Jimenez-Ortega, E., Valenzuela, S., Ramirez-Escudero, M., Pastor, F. J., and Sanz-Aparicio, J. (2020) Structural analysis of the reducing-end xylose-releasing exo-oligoxylanase Rex8A from *Paenibacillus barcinonensis* BP-23 deciphers its molecular specificity. *The FEBS Journal* **287**, 5362-5374

98. Valenzuela, S. V., Lopez, S., Biely, P., Sanz-Aparicio, J., and Pastor, F. J. (2016) The glycoside hydrolase family 8 reducing-end xylose-releasing exo-oligoxylanase Rex8A from *Paenibacillus barcinonensis* BP-23 is active on branched xylooligosaccharides. *Applied and Environmental Microbiology* **82**, 5116-5124
99. Terrett, O. M., and Dupree, P. (2019) Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Current Opinion in Biotechnology* **56**, 97-104
100. Arnling Bååth, J., Mazurkewich, S., Knudsen, R. M., Poulsen, J.-C. N., Olsson, L., Lo Leggio, L., and Larsbrink, J. (2018) Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion. *Biotechnology for Biofuels* **11**, 1-14
101. Mazurkewich, S., Poulsen, J. N., Lo Leggio, L., and Larsbrink, J. (2019) Structural and biochemical studies of the glucuronoyl esterase *Ot*CE15A illuminate its interaction with lignocellulosic components. *Journal of Biological Chemistry* **294**, 19978-19987
102. Topakas, E., Moukouli, M., Dimarogona, M., Vafiadi, C., and Christakopoulos, P. (2010) Functional expression of a thermophilic glucuronyl esterase from *Sporotrichum thermophile*: identification of the nucleophilic serine. *Applied Microbiology and Biotechnology* **87**, 1765-1772
103. Dalrymple, B. P., Cybinski, D. H., Layton, I., McSweeney, C. S., Xue, G.-P., Swadling, Y. J., and Lowry, J. B. (1997) Three *Neocallimastix patriciarum* esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. *Microbiology* **143**, 2605-2614
104. Bitto, E., Bingman, C. A., McCoy, J. G., Allard, S. T., Wesenberg, G. E., and Phillips, G. N., Jr. (2005) The structure at 1.6 Å resolution of the protein product of the At4g34215 gene from *Arabidopsis thaliana*. *Acta Crystallographica Section D: Biological Crystallography* **61**, 630-635
105. Shao, W., and Wiegel, J. (1995) Purification and characterization of two thermostable acetyl xylan esterases from *Thermoanaerobacterium* sp. strain JW/SL-YS485. *Applied and Environmental Microbiology* **61**, 729-733
106. Drzewiecki, K., Angelov, A., Ballschmiter, M., Tiefenbach, K. J., Sterner, R., and Liebl, W. (2010) Hyperthermostable acetyl xylan esterase. *Microbial Biotechnology* **3**, 84-92
107. Blum, D. L., Li, X.-L., Chen, H., and Ljungdahl, L. G. (1999) Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. strain PC-2. *Applied and environmental Microbiology* **65**, 3990-3995
108. Hameleers, L., Penttinen, L., Ikonen, M., Jaillot, L., Fauré, R., Terrapon, N., Deuss, P. J., Hakulinen, N., Master, E. R., and Jurak, E. (2021) Polysaccharide utilization loci-driven enzyme discovery reveals BD-FAE: a bifunctional feruloyl and acetyl xylan esterase active on complex natural xylans. *Biotechnology for Biofuels* **14**, 1-16
109. Kabel, M., Carvalheiro, F., Garrote, G., Avgerinos, E., Koukios, E., Parajó, J., Gırio, F., Schols, H., and Voragen, A. (2002) Hydrothermally treated xylan rich by-products yield different classes of xylo-oligosaccharides. *Carbohydrate Polymers* **50**, 47-56
110. Kabel, M. A., de Waard, P., Schols, H. A., and Voragen, A. G. (2003) Location of *O*-acetyl substituents in xylo-oligosaccharides obtained from hydrothermally treated *Eucalyptus* wood. *Carbohydrate Research* **338**, 69-77
111. Gille, S., and Pauly, M. (2012) *O*-acetylation of plant cell wall polysaccharides. *Frontiers in Plant Science* **3**, 1-7
112. Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Basle, A., Morland, C., Day, A. M., Zheng, H., Rogers, T. E., Thompson, P., Hawkins, A. R., Yadav, M. P., Henrissat, B., Martens, E. C., Dupree, P., Gilbert, H. J., and Bolam, D. N. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications* **6**, 1-16

113. Saulnier, L., Crépeau, M.-J., Lahaye, M., Thibault, J.-F., Garcia-Conesa, M. T., Kroon, P. A., and Williamson, G. (1999) Isolation and structural determination of two 5, 5′-diferuloyl oligosaccharides indicate that maize heteroxylans are covalently cross-linked by oxidatively coupled ferulates. *Carbohydrate Research* **320**, 82-92
114. Biely, P. (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* **30**, 1575-1588
115. Montanier, C., Money, V. A., Pires, V. M., Flint, J. E., Pinheiro, B. A., Goyal, A., Prates, J. A., Izumi, A., Stålbrand, H., and Morland, C. (2009) The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol* **7**, e1000071
116. Topakas, E., Kyriakopoulos, S., Biely, P., Hirsch, J., Vafiadi, C., and Christakopoulos, P. (2010) Carbohydrate esterases of family 2 are 6-*O*-deacetylases. *FEBS letters* **584**, 543-548
117. Holck, J., Fredslund, F., Møller, M. S., Brask, J., Krogh, K. B. R. M., Lange, L., Welner, D. H., Svensson, B., Meyer, A. S., and Wilkens, C. (2019) A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan. *Journal of Biological Chemistry* **294**, 17339-17353
118. Gruninger, R. J., Cote, C., McAllister, T. A., and Abbott, D. W. (2016) Contributions of a unique β-clamp to substrate recognition illuminates the molecular basis of exolysis in ferulic acid esterases. *Biochemical Journal* **473**, 839-849
119. Wefers, D., Cavalcante, J. J. V., Schendel, R. R., Deveryshetty, J., Wang, K., Wawrzak, Z., Mackie, R. I., Koropatkin, N. M., and Cann, I. (2017) Biochemical and structural analyses of two cryptic esterases in *Bacteroides intestinalis* and their synergistic activities with cognate xylanases. *Journal of Molecular Biology* **429**, 2509-2527
120. Smith, S., Witkowski, A., and Joshi, A. K. (2003) Structural and functional organization of the animal fatty acid synthase. *Progress in Lipid Research* **42**, 289-317
121. Chakravarty, B., Gu, Z., Chirala, S. S., Wakil, S. J., and Quiocho, F. A. (2004) Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. *Proceedings of the National Academy of Sciences* **101**, 15567-15572
122. Bayer, E. A., Morag, E., and Lamed, R. (1994) The cellulosome—a treasure-trove for biotechnology. *Trends in Biotechnology* **12**, 379-386
123. Artzi, L., Bayer, E. A., and Moraïs, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* **15**, 83-95
124. Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G., Koropatkin, N. M., McBride, M. J., and Pope, P. B. (2016) A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnology for Biofuels* **9**, 1-16
125. Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., Yang, S. J., Resch, M. G., Adams, M. W., Lunin, V. V., Himmel, M. E., and Bomble, Y. J. (2013) Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* **342**, 1513-1516
126. Appeldoorn, M. M., de Waard, P., Kabel, M. A., Gruppen, H., and Schols, H. A. (2013) Enzyme resistant feruloylated xylooligomer analogues from thermochemically treated corn fiber contain large side chains, ethyl glycosides and novel sites of acetylation. *Carbohydrate Research* **381**, 33-42
127. Razeq, F. M., Jurak, E., Stogios, P. J., Yan, R., Tenkanen, M., Kabel, M. A., Wang, W., and Master, E. R. (2018) A novel acetyl xylan esterase enabling complete deacetylation of substituted xylans. *Biotechnology for Biofuels* **11**, 1-12
128. Krska, D., Mazurkewich, S., Brown, H. A., Theibich, Y., Poulsen, J.-C. N., Morris, A. L., Koropatkin, N. M., Lo Leggio, L., and Larsbrink, J. (2021) Structural and functional analysis of a multimodular hyperthermostable xylanase-glucuronoyl esterase from *Caldicellulosiruptor kristjansonii*. *Biochemistry* **60**, 2206-2220

129. Li, X.-L., Skory, C. D., Cotta, M. A., Puchart, V., and Biely, P. (2008) Novel family of carbohydrate esterases, based on identification of the *Hypocrea jecorina* acetyl esterase gene. *Applied and Environmental Microbiology* **74**, 7482-7489
130. Yu, P., McKinnon, J. J., Maenz, D. D., Olkowski, A. A., Racz, V. J., and Christensen, D. A. (2003) Enzymic release of reducing sugars from oat hulls by cellulase, as influenced by *Aspergillus* ferulic acid esterase and *Trichoderma* xylanase. *Journal of Agricultural and Food Chemistry* **51**, 218-223
131. Tabka, M. G., Herpoël-Gimbert, I., Monod, F., Asther, M., and Sigoillot, J.-C. (2006) Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme and Microbial Technology* **39**, 897-902
132. Faulds, C., Mandalari, G., Curto, R. L., Bisignano, G., Christakopoulos, P., and Waldron, K. (2006) Synergy between xylanases from glycoside hydrolase family 10 and family 11 and a feruloyl esterase in the release of phenolic acids from cereal arabinoxylan. *Applied Microbiology and Biotechnology* **71**, 622-629
133. Tapin, S., Sigoillot, J.-C., Asther, M., and Petit-Conil, M. (2006) Feruloyl esterase utilization for simultaneous processing of nonwood plants into phenolic compounds and pulp fibers. *Journal of Agricultural and Food Chemistry* **54**, 3697-3703
134. Record, E., Asther, M., Sigoillot, C., Pagès, S., Punt, P., Delattre, M., Haon, M., Van den Hondel, C., Sigoillot, J.-C., and Lesage-Meessen, L. (2003) Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Applied Microbiology and Biotechnology* **62**, 349-355
135. Chen, J., Fales, S. L., Varga, G. A., and Royse, D. J. (1995) Biodegradation of cell wall components of maize stover colonized by white-rot fungi and resulting impact on in-vitro digestibility. *Journal of the Science of Food and Agriculture* **68**, 91-98
136. Yu, P., McKinnon, J., and Christensen, D. (2005) Improving the nutritional value of oat hulls for ruminant animals with pretreatment of a multienzyme cocktail: in vitro studies. *Journal of Animal Science* **83**, 1133-1141
137. Kanski, J., Aksenova, M., Stoyanova, A., and Butterfield, D. A. (2002) Ferulic acid antioxidant protection against hydroxyl and peroxyl radical oxidation in synaptosomal and neuronal cell culture systems in vitro: structure-activity studies. *The Journal of Nutritional Biochemistry* **13**, 273-281
138. Tan, Z., and Shahidi, F. (2013) Phytosteryl sinapates and vanillates: chemoenzymatic synthesis and antioxidant capacity assessment. *Food Chemistry* **138**, 1438-1447
139. Topakas, E., Vafiadi, C., and Christakopoulos, P. (2007) Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry* **42**, 497-509
140. Rosazza, J., Huang, Z., Dostal, L., Volm, T., and Rousseau, B. (1995) Biocatalytic transformations of ferulic acid: an abundant aromatic natural product. *Journal of Industrial Microbiology and Biotechnology* **15**, 457-471
141. Priefert, H., Rabenhorst, J., and Steinbüchel, A. (2001) Biotechnological production of vanillin. *Applied Microbiology and Biotechnology* **56**, 296-314
142. Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020) List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology* **70**, 5607
143. Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M., and Michel, G. (2011) Environmental and gut Bacteroidetes: the food connection. *Frontiers in Microbiology* **2**, 1-16
144. Salyers, A. A., Vercellotti, J. R., West, S. E., and Wilkins, T. D. (1977) Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Applied and Environmental Microbiology* **33**, 319-322
145. Evaldson, G., Heimdahl, A., Kager, L., and Nord, C. (1982) The normal human anaerobic microflora. *The Scandinavian Journal of Infectious Diseases* **35**, 9-15

146. Simon, G. L., and Gorbach, S. L. (1984) Intestinal flora in health and disease. *Gastroenterology* **86**, 174-193
147. Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* **95**, 6578-6583
148. Shapira, M. (2016) Gut microbiotas and host evolution: scaling up symbiosis. *Trends in Ecology & Evolution* **31**, 539-549
149. Sommer, F., and Bäckhed, F. (2013) The gut microbiota—masters of host development and physiology. *Nature Reviews Microbiology* **11**, 227-238
150. Sison-Mangus, M. P., Mushegian, A. A., and Ebert, D. (2015) Water fleas require microbiota for survival, growth and reproduction. *The ISME journal* **9**, 59-67
151. Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012) Host-gut microbiota metabolic interactions. *Science* **336**, 1262-1267
152. McKenney, P. T., and Pamer, E. G. (2015) From hype to hope: the gut microbiota in enteric infectious disease. *Cell* **163**, 1326-1332
153. Sampson, T. R., and Mazmanian, S. K. (2015) Control of brain development, function, and behavior by the microbiome. *Cell Host & Microbe* **17**, 565-576
154. Wexler, H. M. (2007) Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews* **20**, 593-621
155. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920
156. Voltolini, C., Battersby, S., Etherington, S. L., Petraglia, F., Norman, J. E., and Jabbour, H. N. (2012) A novel antiinflammatory role for the short-chain fatty acids in human labor. *Endocrinology* **153**, 395-403
157. Tolhurst, G., Heffron, H., Lam, Y. S., Parker, H. E., Habib, A. M., Diakogiannaki, E., Cameron, J., Grosse, J., Reimann, F., and Gribble, F. M. (2012) Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein–coupled receptor FFAR2. *Diabetes* **61**, 364-371
158. González Hernández, M. A., Canfora, E. E., Jocken, J. W., and Blaak, E. E. (2019) The short-chain fatty acid acetate in body weight control and insulin sensitivity. *Nutrients* **11**, 1-32
159. Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., Verberkmoes, N. C., Hettich, R. L., and Gordon, J. I. (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences* **106**, 5859-5864
160. Bjursell, M. K., Martens, E. C., and Gordon, J. I. (2006) Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *Journal of Biological Chemistry* **281**, 36269-36279
161. Crost, E. H., Tailford, L. E., Le Gall, G., Fons, M., Henrissat, B., and Juge, N. (2013) Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *PloS one* **8**, e76341
162. Cockburn, D. W., Orlovsky, N. I., Foley, M. H., Kwiatkowski, K. J., Bahr, C. M., Maynard, M., Demeler, B., and Koropatkin, N. M. (2015) Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Molecular Microbiology* **95**, 209-230
163. Ze, X., David, Y. B., Laverde-Gomez, J. A., Dassa, B., Sheridan, P. O., Duncan, S. H., Louis, P., Henrissat, B., Juge, N., and Koropatkin, N. M. (2015) Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic *Firmicutes* bacterium *Ruminococcus bromii*. *MBio* **6**, e01058-01015

164. Sheridan, P. O., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., Duncan, S. H., O'Toole, P. W., K, P. S., and H, J. F. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microbial Genomics* **2**, e000043
165. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* **11**, 497-504
166. D'Elia, J. N., and Salyers, A. A. (1996) Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. *Journal of Bacteriology* **178**, 7180-7186
167. Tancula, E., Feldhaus, M. J., Bedzyk, L. A., and Salyers, A. A. (1992) Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *Journal of Bacteriology* **174**, 5609-5616
168. Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *Journal of Biological Chemistry* **284**, 24673-24677
169. Foley, M. H., Cockburn, D. W., and Koropatkin, N. M. (2016) The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cellular and Molecular Life Sciences* **73**, 2603-2617
170. Rogers, T. E., Pudlo, N. A., Koropatkin, N. M., Bell, J. S., Moya Balasch, M., Jasker, K., and Martens, E. C. (2013) Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures. *Molecular Microbiology* **88**, 876-890
171. Pereira, G. V., Abdel-Hamid, A. M., Dutta, S., D'Alessandro-Gabazza, C. N., Wefers, D., Farris, J. A., Bajaj, S., Wawrzak, Z., Atomi, H., and Mackie, R. I. (2021) Degradation of complex arabinoxylans by human colonic Bacteroidetes. *Nature Communications* **12**, 1-21
172. D'Elia, J. N., and Salyers, A. A. (1996) Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *Journal of Bacteriology* **178**, 7173-7179
173. Cho, K. H., Cho, D., Wang, G.-R., and Salyers, A. A. (2001) New regulatory gene that contributes to control of *Bacteroides thetaiotaomicron* starch utilization genes. *Journal of Bacteriology* **183**, 7198-7205
174. Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M., and Martens, E. C. (2012) Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism. *Journal of Biological Chemistry* **287**, 34614-34625
175. Koropatkin, N. M., and Smith, T. J. (2010) SusG: a unique cell-membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* **18**, 200-215
176. Bolam, D. N., and van den Berg, B. (2018) TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Current Opinion in Structural Biology* **51**, 35-43
177. Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Baslé, A., Firbank, S. J., Zheng, H., Robinson, C. V., Winterhalter, M., and Kleinekathöfer, U. (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411
178. Celia, H., Botos, I., Ni, X., Fox, T., De Val, N., Lloubes, R., Jiang, J., and Buchanan, S. K. (2019) Cryo-EM structure of the bacterial Ton motor subcomplex ExbB–ExbD provides information on structure and stoichiometry. *Communications Biology* **2**, 1-6
179. Anderson, K. L., and Salyers, A. A. (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *Journal of Bacteriology* **171**, 3192-3198

180. Smith, K. A., and Salyers, A. A. (1991) Characterization of a neopullulanase and an alpha-glucosidase from *Bacteroides thetaiotaomicron* 95-1. *Journal of Bacteriology* **173**, 2962-2968
181. Smith, K. A., and Salyers, A. A. (1989) Cell-associated pullulanase from *Bacteroides thetaiotaomicron*: cloning, characterization, and insertional mutagenesis to determine role in pullulan utilization. *Journal of Bacteriology* **171**, 2116-2123
182. Terrapon, N., Lombard, V., Drula, E., Lapebie, P., Al-Masaudi, S., Gilbert, H. J., and Henrissat, B. (2018) PULDB: the expanded database of polysaccharide utilization loci. *Nucleic Acids Research* **46**, D677-D683
183. Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide utilization loci: fueling microbial communities. *Journal of Bacteriology* **199**, e00860-00816
184. Terrapon, N., Lombard, V., Gilbert, H. J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647-655
185. Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., and Bolam, D. N. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biology* **9**, e1001221
186. McBride, M. J., Xie, G., Martens, E. C., Lapidus, A., Henrissat, B., Rhodes, R. G., Goltsman, E., Wang, W., Xu, J., Hunnicutt, D. W., Staroscik, A. M., Hoover, T. R., Cheng, Y. Q., and Stein, J. L. (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Applied and Environmental Microbiology* **75**, 6864-6875
187. Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., Zheng, B., and Yin, Y. (2021) dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Research* **49**, D523-D528
188. McKee, L. S., La Rosa, S. L., Westereng, B., Eijsink, V. G., Pope, P. B., and Larsbrink, J. (2021) Polysaccharide degradation by the Bacteroidetes—mechanisms and nomenclature. *Environmental Microbiology Reports*, 1-23
189. Martens, E. C., Roth, R., Heuser, J. E., and Gordon, J. I. (2009) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *Journal of Biological Chemistry* **284**, 18445-18457
190. Hehemann, J.-H., Kelly, A. G., Pudlo, N. A., Martens, E. C., and Boraston, A. B. (2012) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences* **109**, 19786-19791
191. Hehemann, J.-H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert, W., Michel, G., and Czjzek, M. (2012) Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *Journal of Biological Chemistry* **287**, 30571-30584
192. Ficko-Blean, E., Préchoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y., Stam, M., Génicot, S., Jam, M., and Calteau, A. (2017) Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nature Communications* **8**, 1-17
193. Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., Klinter, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., Martens, E. C., and Brumer, H. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502
194. Reddy, S. K., Bågenholm, V., Pudlo, N. A., Bouraoui, H., Koropatkin, N. M., Martens, E. C., and Stålbrand, H. (2016) A β-mannan utilization locus in *Bacteroides ovatus* involves a GH36 α-galactosidase active on galactomannans. *FEBS letters* **590**, 2106-2118
195. Bågenholm, V., Reddy, S. K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C. M., Aurelius, O., Rogers, T., Xiao, Y., and Logan, D. T. (2017) Galactomannan catabolism

conferred by a polysaccharide utilization locus of *Bacteroides ovatus*: enzyme synergy and crystal structure of a *β*-mannanase. *Journal of Biological Chemistry* **292**, 229-243
196. Kharade, S. S., and McBride, M. J. (2014) *Flavobacterium johnsoniae* chitinase ChiA is required for chitin utilization and is secreted by the type IX secretion system. *Journal of Bacteriology* **196**, 961-970
197. Mazurkewich, S., Helland, R., Mackenzie, A., Eijsink, V. G. H., Pope, P. B., Branden, G., and Larsbrink, J. (2020) Structural insights of the enzymes from the chitin utilization locus of *Flavobacterium johnsoniae*. *Scientific Reports* **10**, 1-11
198. Tamura, K., Hemsworth, G. R., Dejean, G., Rogers, T. E., Pudlo, N. A., Urs, K., Jain, N., Davies, G. J., Martens, E. C., and Brumer, H. (2017) Molecular mechanism by which prominent human gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides. *Cell Reports* **21**, 417-430
199. Tamura, K., Foley, M. H., Gardill, B. R., Dejean, G., Schnizlein, M., Bahr, C. M., Creagh, A. L., van Petegem, F., Koropatkin, N. M., and Brumer, H. (2019) Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cellular and Molecular Life Sciences* **76**, 4319-4340
200. Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J. L., Day, A., Pena, M. J., McLean, R., Suits, M. D., Boraston, A. B., Atherly, T., Ziemer, C. J., Williams, S. J., Davies, G. J., Abbott, D. W., Martens, E. C., and Gilbert, H. J. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169
201. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Basle, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y., O'Neil, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M. C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65-70
202. Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R., and Sonnenburg, J. L. (2018) An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**, 434-438
203. García-Bayona, L., and Comstock, L. E. (2019) Streamlined genetic manipulation of diverse *Bacteroides* and *Parabacteroides* isolates from the human gut microbiota. *MBio* **10**, e01762-01719
204. Wang, J., Shoemaker, N. B., Wang, G.-R., and Salyers, A. A. (2000) Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. *Journal of Bacteriology* **182**, 3559-3571
205. Bedzyk, L. A., Shoemaker, N. B., Young, K. E., and Salyers, A. A. (1992) Insertion and excision of Bacteroides conjugative chromosomal elements. *Journal of Bacteriology* **174**, 166-172
206. Ahmad, A., Anjum, F. M., Zahoor, T., Nawaz, H., and Dilshad, S. M. R. (2012) Beta glucan: a valuable functional ingredient in foods. *Critical Reviews in Food Science and Nutrition* **52**, 201-212