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

Philosophiae Doctor (PhD)
Thesis 2018:97

Exploring acetylated galactoglucomannan as a source of prebiotics

Uttesting av acetylert galaktoglukomannan som en kilde til prebiotika

Leszek Michalak

Exploring acetylated galactoglucomannan as a source of prebiotics

Uttesting av acetylert galaktoglukomannan som en kilde til prebiotika

Philosophiae Doctor (PhD) Thesis

Leszek Michalak

Faculty of Chemistry, Biotechnology and Food Science

Norwegian University of Life Sciences

Ås, 2018

ISSN: 1894-6402

ISBN: 978-82-575-1565-2

Thesis: 2018:97

Table of Contents

Table of contents.

Acknowledgments:

This thesis is the result of a little over three years of work, which began in May 2015 at the Protein Engineering and Proteomics group, and ended in October 2018 at the Bioprocess Technology and Biorefining group at NMBU, spanning research groups, time and (office) space. The project was funded by the Research Council of Norway grant no. 244259.

I want to first and foremost thank my supervisor, Bjørge Westereng for giving me the opportunity, for the enormous amount of teaching you contributed during my time at NMBU, and especially for always being available for an hour or two of 'just before I go' brainstorming. Secondly, I want to thank Vincent Eijsink for assembling and leading a fantastic group of people to work with and learn from. I would like to thank my co-supervisors: Svein Jarle Horn, for your contributions with expertise on biorefining, Margareth Øverland, for making the feeding trial a reality, Svein Halvor Knutsen, for teaching me a great deal about carbohydrate analytics, and Phillip Pope, for being patient with me while I took baby steps in the 'omics universe. I also owe a debt of gratitude to my honorary co-supervisor, Sabina Leanti La Rosa, for spending so much of your time teaching me about microbiology, protein expression, and tortellini folding. It's on paper - I owe a Sicilian a favour. Special thanks to Finn Lillelund Aachmann for your NMR wizardry and for taking things personally, and Magnus Øverlie Arntzen for your superhuman skills in analytics, especially when things go wrong.

Thanks to Leidy, Liv Torunn, Adrijana, Jon and everyone at Ås Gård who helped to assemble the feeding trial and disassemble the piglets. Anne Cath and Lasse, you have been the emergency response team every time I broke something in the lab, and for this, a big thank you! Thank you Jenny, Sophie, Zarah, Anikó, and Bastien, for answering a billion 'How do I…'s over the course of the last three years. Thanks to Chris for all your work and patience, Shaun for your help with the analytics. Thank you Line, John and Ricardo for helping with the biorefinery operations and thank you, Richard Dean Anderson for inspiring my bioprocessing solutions. Thanks to Berit and Kristine at BioChos for the Christmas parties! Thanks to Ben and Dejan for the moral support and extensive experimental work in carbohydrate processing we carried out in my kitchen!

We have a tough job people. The days are long, failures too frequent and too often undeserved. It can get very frustrating, but you made three and a half tough years seem like fun. For that, dear IKBM people, thank you!

Leszek

Acknowledgments.

“Laughter and tears are both responses to frustration and exhaustion. I myself prefer to laugh, since there is less cleaning up to do afterward.”

Kurt Vonnegut

Summary

The wood and paper industry in Norway and worldwide suffers from low demand and low prices. Developing consolidated biorefinery solutions to generate higher value products from woody biomass could be a viable alternative to pulp and paper industries. Paper is a low cost-high volume commodity, and modern biotechnology can produce high value compounds from the same feedstock. Lignocellulose based compounds can become food thickeners and stabilizers, specialty and platform chemicals or nutraceuticals/pharmaceutical ingredients. In the case of our project, extracted hemicellulose was tested as a potential prebiotic – a feed additive intended to improve the gastrointestinal health of weaned piglets by selectively enriching a population of gut commensal bacteria beneficial to the host. Since preventive antibiotics were banned in farming, considerable efforts have been made to prevent illness and lower the mortality rates in piglets. Intervention with prebiotics is an appealing means of counteracting this issue. Mannan based prebiotics were tested as a means to reduce the loss of efficiency and improve animal welfare.

In the course of the process development, a number of research areas were investigated beginning with the effect of steam explosion conditions on the features of the produced acetylated galactoglucomannan (GGM). Determining the impact of pretreatment severity and 'in stream' processing conditions on complexity of the produced hemicellulose were important steps for developing new processes aimed at using complex oligosaccharides for high value applications (Paper I). A large part of the research efforts focused on characterizing microbial enzymes that break down the GGM *in vivo*, with a special focus on the acetyl esterases. This allowed us to develop an understanding of the metabolism of prebiotics in the gut microbiome and new possibilities for enzymatic hydrolysis of mannans to tailored oligosaccharides. While studying enzymatic deacetylation of these hemicelluloses, we also documented in fine detail the phenomenon of acetyl migration in oligosaccharides (Paper II). This is an observation of high importance for large scale biorefining of mannans, as it affects the microbial utilization and enzymatic recalcitrance of the biomass. Finally, by testing the GGM preparation in a large scale feeding trial and the following multi 'omics' analysis, we gained an in depth understanding of the

effect of mannans on the gut microbiome of developing piglets. Through 16S rRNA amplicon analysis we monitored the development of the gut microbiome over time, and using metagenomics and metaproteomics we were able to identify the microbes degrading the prebiotics GGM preparation and the enzymatic processes they use (Paper III). These results are highly important for understanding the diet-microbiome interactions, and the microbial ecology of the mammalian gut. Identification of *Faecalibacterium* and *Roseburia* strains among the mannan degraders is of special importance, since these genera are considered beneficial for human gut health.

In its wide interdisciplinary scope, the project addressed a number of issues of industrial, economical and societal importance; the aim was to use a local feedstock, available in abundance in the Norwegian forests and as a waste product from lumbering to create a novel process that could be adapted by the pulp and paper industry. The project aimed at creating a product that would improve farming efficiency, animal welfare and possibly make a global improvement in food availability. At the same time, a functioning prebiotic could reduce the amount of antibiotics in feed in countries where it is still in practice. Production of spruce prebiotics was the largest non-food experiment conducted at NMBU at the time, and the first of its kind conducted in the newly built pilot scale biorefinery. Through extensive experimentation, we have developed a better understanding of the pilot plants capabilities, identified opportunities for expansion, and established methods now routinely used by other researchers and industrial partners at NMBU.

Sammendrag

Tre og papirindustrien i Norge og verden forøvrig lider av lav etterspørsel og lave priser. Utvikling av sammensatte bioraffineriløsninger for å generere høyverdiprodukter fra trebiomasse kan være et bærekraftig alternativ for industrien. Papir er et lavkost høy volum produkt laget fra trebiomasse. Moderne bioteknologi kan isteden lage høyverdiprodukter fra det samme råstoffet. Lignocellulose baserte forbindelser kan bli fortykningsmidler og stabilisatorer i mat, spesialkjemikalier, plattformkjemikalier eller ingredienser til farmasøytisk industri og kostilskuddsbransjen. I dette prosjektet har vi testet mulighetene for å benytte hemicellulose ekstrahert fra gran som prebiotika – en fôrtilsetning som skal forbedre tarmhelsen hos (avvendte) griser ved å selektivt stimulere gode bakterier. Siden bruk av profylaktisk antibiotika ble forbudt, har det blitt lagt ned enorme ressurser for å forhindre sykdom og redusere dødelighet hos husdyr. Et mulig alternativ for å redusere eller å eliminere disse problemene helt eller delvis er å bruke prebiotika som en fôrtilsetning. I dette prosjektet har vi testet mannan fra gran som et mulig alternativ for å forbedre dyrehelse og om mulig forhindre produksjonstap.

I dette svært tverrfaglige prosjektet har vi sett på flere forskningsområder. Vi ønsket å utvikle en prosess for å produsere hemicelluloser med høy strukturkompleksitet med høyest mulig acetyleringsgrad og galaktosesidekjeder. Det er kjent at intensitetsgraden av dampeksplosjon kan bidra til delvis ødeleggelse av karbohydrater, samtidig er det kjent at acetyleringer lett påvirkes av selv milde alkaliske betingelser. Vi har sett på hvordan struktursammensetningen av det acetylerte galaktoglukomannanet (GGM) påvirkes av pH i dampeksplosjon. Videre har vi også studert hva slags effekt intensitetsgraden i ulike prosesstrinn har på strukturkompleksiteten av hemicellulosen (artikkel 1 i denne avhandlingen).

Et annet forskningsområde vi har fokusert på er karakterisering av bakterielle enzymer som bryter ned acetylert galaktoglukomannan (GGM) *in vivo* i gris, der vi har hatt et særlig fokus på acetylesteraser. Dette bidro til at vi har utviklet en grunnleggende forståelse av hvordan metabolisme av prebiotika i tarmfloraen foregår og samtidig kunne utvikle enzymatisk hydrolyse for å lage skreddersydde oligosakkarider. Studiene av acetylesteraser har gitt oss grunnleggende forståelse av hvordan to ulike esteraser virker

sammen om å fjerne alle acetyleringer på komplekse mannaner. I tillegg har vi dokumentert i detalj hvordan acetyleringsmønsteret på oligosakkarider endres ved migrering indusert av ulik pH. Dette er en viktig observasjon som det må tas hensyn til i bioraffinering av mannan fordi det a) vil påvirke de teknologiske egenskapene ved mannan og b) påvirke bakteriell nedbryting dersom man tenker å bruke mannan som prebiotika (artikkel 2 i denne avhandlingen).

I siste del av prosjektet har vi produsert 50 kilo GGM og gjennomført et foringsforsøk på gris der vi har gjennomført en «multi-omics» tilnærming og fått en dyp forståelse av hvordan mannan påvirker tarmfloraen i smågris over tid inkludert en effekt av dosering. Vi brukte 16S rRNA amplicon analyse for å se på endring i tarmfloraen over tid, og ved å bruke metagenomikk og metaproteomikk kunne vi identifisere de bakteriene som brøt ned mannan og hvilke enzymatiske prosesser de bruker. Disse resultatene er svært viktige for å forstå interaksjoner og sammmenhenger mellom diett og tarmflora og den mikrobielle økologien i tarmen hos gris generelt (artikkel 3 i denne avhandlingen). Blant de mannannedbrytende bakteriene identifiserte vi to bakterierslekter av særlig interesse, nemlig *Faecalibacterium* and *Roseburia.* Det er økende enighet om at disse er gunstige for tarmhelsen hos mennesker.

Som en effekt av prosjektets brede tverrfaglige natur har vi adressert aspekter som er viktige industrielt, økonomisk og samfunnsmessig. Vi har brukt lokalt råstoff som er tilgjengelig i enorme mengder, og utviklet nye prosesser som kan bli innført i treforedlingsindustrien. Prosjektet har tatt sikte på å skape et produkt som kan forbedre effektiviten i husdyrproduksjon, dyrevelferd og muligens bidra til global økning i mattilgjengelighet ved å tilvirke mat av komponenter som normalt ikke inngår i mat og næringsmidler. Samtidig vil en fungerende prebiotika kunne redusere bruken av antibiotika i fôr i land der profylaktisk antibiotikabruk fortsatt praktiseres. Produksjon av 50 kilo mannan fra gran er det største eksperimentet utover konvensjonell matproduksjon som er gjort ved NMBU, og et pionerforsøk gjennomført ved det nye pilot bioraffineriet. Gjennom prosjektperioden har vi utviklet en bred erfaring om prosessering i det ny-etablerte bioraffineriet, og vært en driver for å identifisere flaskehalser og muligheter for nye anvendelser. Det er nå etablert metoder som brukes rutinemessig på NMBU, på ÅS Campus av andre forskere og prosjekter samt industripartnere.

Abbreviation

AA – Auxiliary activity.

Ac – Acetyl.

AGP – Antibiotic growth promoters.

AVM – *Aloe vera* mannan.

CAD – Charged aerosol detection

CAZy – Carbohydrate Active enZYmes database

CBM – Carbohydrate binding module

CD – Crohn's disease

CE – Carbohydrate esterase

DA – Degree of acetylation

DMSO - Dimethylsuphoxide

DP – Degree of polymerization

GAX – Glucuronoarabinoxylan

GC-FID – Gas chromatography – flame ionization detection

GGM – Galactoglucomannan

GH – Glycosyl hydrolase

Glc*p* – Glucopyranose

GM – Glucomannan

GT – Glycosyl transferase

HILIC – Hydrophilic interaction chromatography

HPLC – High performance liquid chromatography

HSQC – Heteronuclear Single Quantum Coherence

IBS – Irritable bowel syndrome

MALDI-ToF – Matrix assisted laser desorption ionization-time of flight

MAG – Metagenomics assembled genomes

MeGlcA – Methylglucuronic acid

MS – Mass spectrometry

NMR – Nuclear magnetic resonance

ORF – Open reading frame

OTU – Operational taxonomic unit

PA – Polysaccharide lyase

PUL – Polysaccharide utilization locus

SCFA – Short chain fatty acid

SE – Steam explosion

SEC – Size exclusion chromatography

Sus – Starch utilization system

UC – Ulcerative colitis

Xyl*p* – Xylopyranose

Abbreviations.

List of papers:

Paper I:

Effects of pH on steam explosion extraction of acetylated galactoglucomannan from Norway spruce. Leszek Michalak, Svein Halvor Knutsen, Ida Aarum, Bjørge Westereng

Manuscript submitted for publication.

Paper II:

A pair of esterases from a commensal gut bacterium completely deacetylate highly complex mannans. Leszek Michalak, Sabina Leanti La Rosa, Shaun Leivers, Åsmund Kjendseth Røhr, Finn Lillelund Aachmann, Bjørge Westereng

Manuscript in preparation.

Paper III:

Wood-derived galactoglucomannan promotes butyrate-producing microbes in the swine gut microbiome. Leszek Michalak, John Christian Gaby, Sabina Leanti La Rosa, Leidy Lagos, Johannes Dröge, Margareth Øverland, Phillip B. Pope, Bjørge Westereng

Manuscript in preparation.

Other publications by the author:

The human gut Firmicute *Roseburia intestinalis* is a primary degrader of dietary β-mannans. Sabina Leanti La Rosa, Leszek Michalak, Maria Louise Leth, Morten Ejby Hansen, Nicholas A. Pudlo, Robert Glowacki, Christopher Workman, Magnus Ø. Arntzen, Phillip B. Pope, Eric C. Martens, Maher Abou Hachem, Bjørge Westereng

Manuscript submitted for publication.

1 Introduction

1.1 Plant biomass

Terrestrial plant biomass is a ubiquitous, renewable feedstock for contemporary biorefining. Plant biomass can be divided into hardwood, softwood and monocot derived biomass, although all three share the main biochemical constituents: cellulose, hemicellulose and lignin. All three constitute the cell walls of plants, present in varying ratios and in various, complex networks depending on the plant and tissue of origin.

Cellulose is the main component of all three; it is the most abundant polymer in Nature and one of the most versatile resources used in many industries. Cellulose consists of long chains of β-1,4 linked glucopyranose (Glc*p*) units arranged into microfibrils. An exact structure of the microfibrils is not known; it is hypothesized that 18-24 of the glucose chains are arranged together to form a structure with a hexagonal cross-section. Cellulose, hemicellulose and lignin are together arranged in a complex three dimensional network forming the plant cell walls (Fig. 1). The relative abundance of constituents and their arrangement is what gives the plants biomass its functional properties – flexibility, rigidity, tensile strength and chemical resistance (Cosgrove, 2005). In most plant tissues, cellulose constitutes 35-50% of the dry weight (Cosgrove, 2005). Biorefining of cellulose was for a long time focused on extracting the biomass through pretreatments, breaking the polymer down to constituent glucose, followed by fermentation performed separately or simultaneously (Olsson et al., 2004). In saccharification, the crystalline cellulose fibers are broken down into fermentable constituents, while fermentation uses microbes to process the released carbohydrates to platform chemicals. Cellulose in the form of woody biomass and agricultural by-products (corn stover, sugarcane bagasse, wheat straw) are the canonical biorefining feedstocks, commercially used for biofuel and ethanol production.

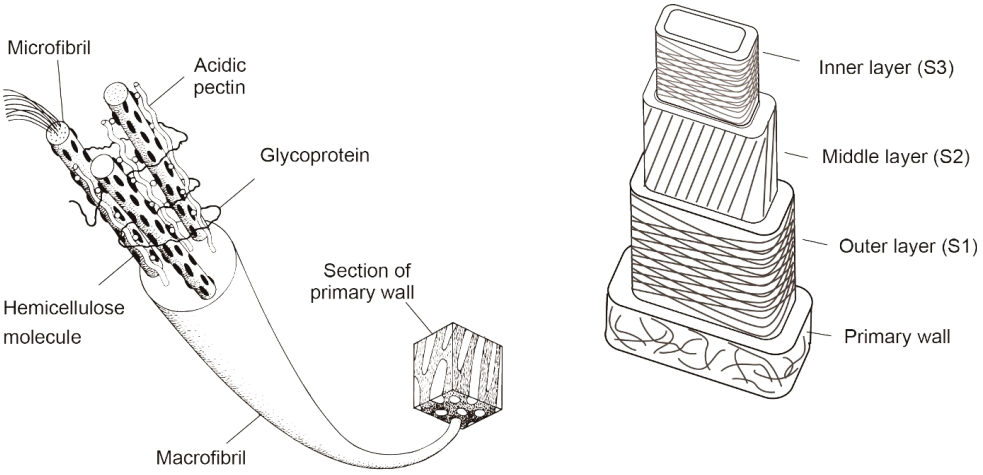


Figure 1. Arrangement of cellulose, hemicellulose and lignin in plant cell walls of lignocellulosic biomass. The primary cell wall consists of cellulose microfibrils woven together with hemicellulose, pectin and glycoproteins into macrofibrils. Secondary cell wall lines the primary wall from the cytosol side, and consists of an outer (S1), middle (S2) and inner layer (S3). Figure adapted from (Yang et al., 2013).

Hemicellulose is a broad term used to describe the carbohydrate fraction of plant biomass, which does not fall into the categories of cellulose, lignin and pectin. This group includes mannans, xylans, β-glucans and xyloglucans. The structural characteristic of hemicelluloses is the presence of β-1,4 linked monosaccharides in the backbone of the chains. Mixed linkage glucans are an exception to this rule, consisting of both β-1,3 and β-1,4 linked Glc*p* (Scheller and Ulvskov, 2010).

Xyloglucan is a polymer of β-1,4 linked glucose with α-1,6 linked xylose units (Xyl*p*), with fucose and galactose branching from the xylose (Hayashi, 1989) (Fig. 2). Arabinosylated variety of xyloglucans have been found in potatoes (Fry et al., 1993). The xyloglucan backbone is very rigid and is considered to grant plant tissues the tensile strength (Hayashi and Kaida, 2011). Xyloglucans are an example of hemicellulose with multiple high value applications, the adhesive properties of xyloglucans in solution are used in drug delivery (Kulkarni et al., 2017), and in wound healing (Ajovalasit et al., 2018).

Xylans are polymers of β-1,4 linked xylose units and is the prevalent hemicellulose in grasses and hardwoods. The backbone of xylan is highly decorated, commonly with α-1,2

linked glucuronosyl, 4-*O*-methyl glucuronosyl residues (Scheller and Ulvskov, 2010) (Fig. 2 and 3). Besides glucuronic acids, xylans are decorated with α-arabinose/acetic acid esters bound as 2-*O* and 3-*O*-acetylations (Fig. 3). Xylan is a highly relevant biorefinery feedstock for biochemicals and platform chemicals manufacture (Olsson et al., 2004). It is one of the common constituents of plant-based foods and animal feeds, although it is non-digestible for mammals. Humans and animals rely on their gastrointestinal tract microbes to break down the xylan and ferment the released sugars into short chain fatty acids (SCFAs).

Xylans are highly abundant in industrially relevant feedstocks for biorefining including grasses, corn stover and sugarcane bagasse. Since all of these feedstocks have been used in large scale bioethanol plants, the pretreatments and enzymatic processing steps of their utilization have been optimized and a considerable body of knowledge on enzymatic processes has been developed around these processes (Biely et al., 2016). Xylans resemble mannans in some respects: they are substituted by acetylations and branchings that require multiple hydrolytic enzymes to break down (Fig. 2 and 3). The body of knowledge developed around xylan breakdown has aided the research into mannan breakdown considerably, especially concerning acetylation and esterases removing the 2-*O*- and 3-*O*-monoacetylated Xyl*p* residues, some of which are also mannan active (Topakas et al., 2010a, Montanier et al., 2009). Xylans have also been used as prebiotics for poultry and monogastric animals (Singh et al., 2015).

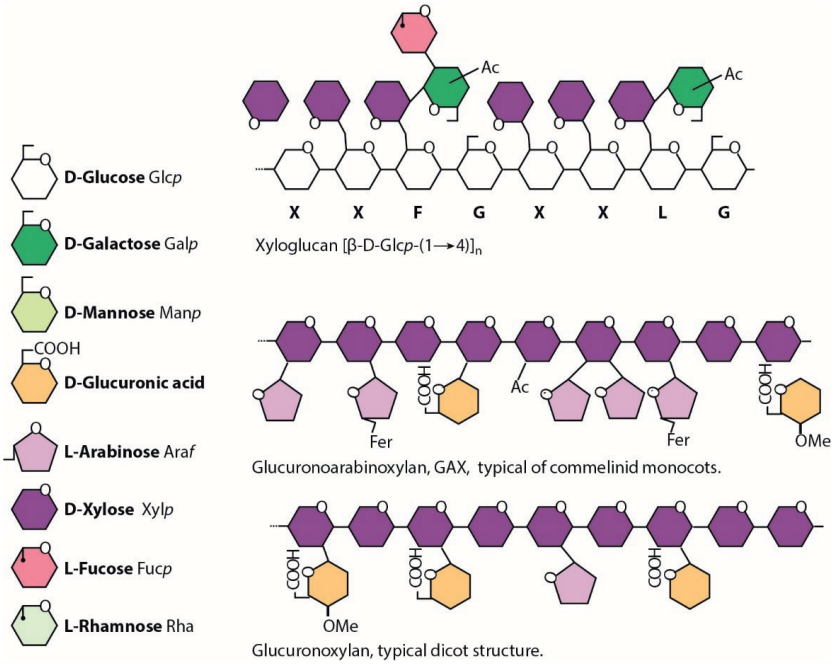


Figure 2. Structures of the most common hemicelluloses found in plant biomass: xyloglucan consists of a β-1,4 linked glucose units with α-1,6 linked xylose, with acetylated galactose and fucose present as branchings of xylose. Glucuronoarabinoxylan (GAX) and glucuronoxylan are variations of decorated xylans differing in the types of substitutions present. Figure adapted from (Scheller and Ulvskov, 2010).

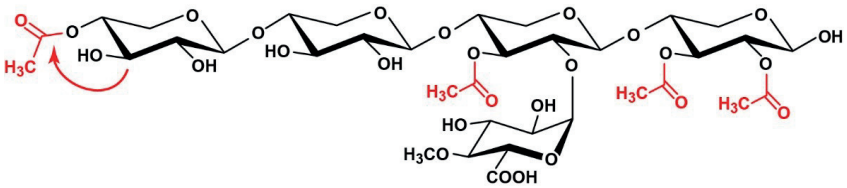


Figure 3. Detailed structure of the backbone of acetylated glucuronoxylan with 2-O and 3-O acetylations present in the plane of their respective xylose moiety. Acetylations can migrate 'around' the xylose molecule in a 2-O- -> 3-O and 3-O- -> 4-O- direction in the non-reducing end.

1.1.1 Mannans

Mannans are a hemicellulose based on a backbone of β-1,4 linked mannose (Man*p*) (Fig. 4). They are a prevalent form of hemicellulose in softwoods, legume seeds and coffee beans. In softwoods, about 20% of dry wood mass (Timell, 1967, Lundqvist et al., 2002) is made up of mannans. Glucomannans, such as konjac (*Amorphophallus konjac*), mannose residues are interspersed with β-1,4 linked D-Glc*p* in varying ratios. In galactomannans, such as carob (*Ceratonia siliqua*), the backbone is decorated with α-1,6-D-galactose (Gal*p*) substitutions (Tester and Al-Ghazzewi, 2013). Mannans can be 2-*O*, 3-*O* and 6-*O*-acetylated (Fig. 4). In *Aloe vera* mannan, a small degree of arabinose substitutions was observed (Simões et al., 2012).

Konjac, *Aloe vera*, ivory nut guar gum and carob mannans are all of industrial relevance (Singh et al., 2018). Guar gum, which is a powdered form of mannan from the guar bean pods, and ivory nut mannan are used as thickeners in the food industry (Du et al., 2012, Mudgil et al., 2014). Konjac has recently become available on the health food market as a nutraceutical and low-calorie alternative to starchy foods such as pasta and rice. *Aloe vera* mannan (AVM) has immunostimulatory properties and is a common ingredient in dietary supplements, topical medicine and skin care products (Simões et al., 2012).

1.1.2 Galactoglucomannan from Norway spruce (*Picea abies*)

Galactoglucomannan (GGM) is the main hemicellulose in Norway spruce (*Picea abies*). The backbone of this polysaccharide consists of β-1,4- D-Man*p* and β-1,4- D-Glc*p* residues present at varying ratios (Fig. 4). The mannan backbone is decorated with α-1,6-D-Gal*p* substitutions, prevalently attached to the Man*p*, and to a lesser extent on Glc*p* (Willfor et al., 2003, Lundqvist et al., 2002). GGM is commonly described using its monosaccharide ratio: Gal:Glc:Man. Two varieties of GGM are reportedly recovered from Norway spruce: a high galactose type with a monosaccharide ratio of 1:1:3, and a low galactose type with a ratio of 0.1:1:3. The former being the less prevalent (5-8% dry wood weight) than the galactose poor variety (10-15% dry wood weight) (Timell, 1967). The composition of extracted GGM varies depending on the severity of extraction and purification methods. For example, changes in the pH caused by NaOH impregnation in extraction with heat

fractionation have recovered GGM with Gal:Glc:Man ratio of 0.3:1:3.3 (pH 3.6 after fractionation) to 0.6:1:1.6 (pH 12.3 after fractionation). According to existing reports on spruce GGM, about 30% of the D-Man*p* residues are 2-*O*-, 3-*O*- and 6-*O*- esterified by acetylations (Lundqvist et al., 2002).

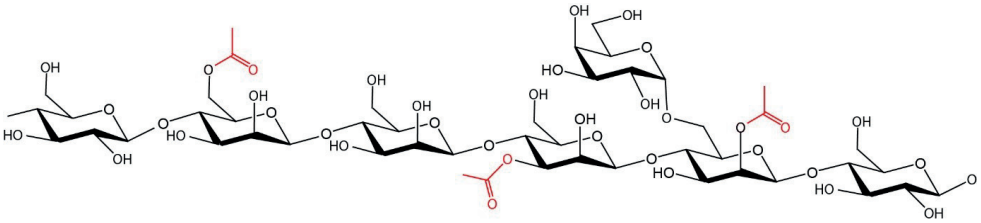


Figure 4. Structure of acetylated GGM: A β-1,4 lined Glcp and Manp backbone is decorated with α-1,6 galactosylations, and 2-O-, 3-O- and 6-O- acetylations. Figure adapted from (Arnling Bååth et al., 2018).

1.1.3 Acetylation of mannans

Acetylations of mannan are considered to be a structural feature enabling the fiber to interact with lignin (Cosgrove, 2005), and a form of defense of the secondary cell wall against hydrolytic breakdown by pathogens (Juturu and Wu, 2012). Acetylations inhibit hydrolytic enzymes, such as those secreted by plant pathogens, by obstructing the glycosidic bonds. In xylans, which also carry acetylations on their primary constituent, both the 2-*O*- and 3-*O*-acetylations are in the equatorial plane of the xylose molecules. Equatorial acetylations are structurally similar (Fig. 3) and can be accommodated in different orientations by a similar active site structure. This is why many acetyl esterases are active on both 2-*O*- and 3-*O*-acetylations in xylan (Nakamura et al., 2017). A unique characteristic of the *O*-acetyl groups in Man*p* units of the mannose chains is their relative orientation: the 3-*O*-acetylations are in the equatorial plane of the molecule, while 2-*O*-acetylations are axial (Fig. 4), and this orientation makes mannan acetylations inaccessible to many acetylxylan esterases. 2-*O*-acetylations are the more prevalent form of acetylation, the 2-*O*-:3-*O*- acetylation ratio in Norway spruce GGM is approximately 2.0:1.0 (Lundqvist et al., 2003). This feature of acetylated mannans makes them unlike any other hemicellulose component, a uniquely

challenging substrate for enzymatic breakdown. There are a number of studies describing occurrence of 6-*O*-acetylations, which are considered to be a result of migration caused by extraction methods involving high temperature (Xu et al., 2010). It is claimed that acetylations migrate 'clockwise' of the sugar ring (Roslund et al., 2008) and hence, the migration does not explain the presence of 6-*O*-acetylations on the Man*p* residues other than the non-reducing end. 6-*O*-acetylations as well as 4-*O*-acetylations present on the non-reducing end Man*p* were found in the GGM produced in this project. Presence of the 6-*O*-acetylations warrants further research into acetyl migration and hemicellulose synthesis in spruce.

Acetylation of oligosaccharides affects the solubility and viscosity of mannans in solution by restricting the formation of hydrogen bonds between oligosaccharides (Williams et al., 2000). This versatile behavior makes mannans an attractive ingredient for hydrocolloids (Willfor et al., 2008), thickeners or stabilizers (Mikkonen et al., 2009). Because of the effects on viscosity guar gum and konjac mannan are commonly used in the food and feed industry (Du et al., 2012, Mudgil et al., 2014).

From the perspective of this study, acetylation of the product was of paramount importance. A high degree of acetylation was hypothesized to nourish a subpopulation of the gut microbiome which had an efficient deacetylation apparatus, and hence provide a mechanism for selectivity (Bi et al., 2016). Extraction of mannan from the wood by steam explosion is facilitated by the release of acetate. Released acetate contributes to the autohydrolysis of lignocellulose during pretreatment. Therefore it was important to choose conditions that limit deacetylation, but at the same time result in a sufficiently effective extraction. Acetylation and its impact on extraction, processing and microbial utilization of GGM is the common theme of this work.

1.2 Biorefining of lignocellulosic biomass

Biorefining of lignocellulose is a broad term for all large-scale biotechnology endeavors aimed at tapping into the source of complex, renewable chemicals present in plant biomass. Applications and potential products of biorefining are endless, and the best-established ones are ethanol and biodiesel for transport fuel, feed and food ingredients, specialty cellulose,

lignin related products. The common pipeline for most biorefining approaches consists of various combinations of a pretreatment step, filtration, fractionation, a microbial fermentation or enzymatic hydrolysis, alternatively chemical functionalization, followed by product recovery or purification.

The purpose of pretreatments is to increase extractability of oligosaccharides and render the biomass more accessible to the enzymes or microbes in the following steps. In order to be a commercially viable option, it must be cheap, effective, and produce as little carbohydrate and lignin degradation products as possible. Chemical pretreatments use low or high pH in order to break down the biomass into its constituent sugars (Pedersen and Meyer, 2010). Sulfates, inorganic acids and hydroxides are often used in chemical pretreatments, resulting in high process severity.

1.2.1 Hydrothermal pretreatments and steam explosion

Hydrothermal pretreatment is an efficient technique which uses hot water or steam to induce an autohydrolysis of the cell wall, disrupt the cell wall material and ease the release of carbohydrates. Hydrothermal pretreatment yields can be improved by following the pretreatment with defibration to improve the surface to volume ratio and the yields from water extraction. This is usually achieved by an explosion at the end of the treatment, but can also be carried out with an equally effective subsequent refiner (mechanical disruption) step (Schütt et al., 2012). In a steam explosion (SE) reactor, the biomass is exposed to steam at high temperature and pressure. At these conditions, the acetic acid bound to hemicellulose fibers is to some degree released and hydrolyses glycosidic bonds generating shorter oligosaccharides from the polysaccharides (Rissanen et al., 2015). Shorter, partly deacetylated oligosaccharides are water soluble and can be water extracted. Steam explosion pretreatment may be conducted at different conditions, some more severe than others. The intensity of these conditions are characterized by a severity factor $R_0 = e^{(T_{exp} - 100)/14.75}$ (R. P. Overend, 1987), which is a value used to compare the results of treatments at different conditions, calculated from the reaction temperature, pressure, and the biomass residence time. To account for additives altering the pH of hydrothermal pretreatment, a more comprehensive combined severity factor $R'_0 = (10^{-pH}) * (t *$

$e^{(T_{exp}-100)/14.75}$) (Kabel et al., 2007, Chum et al., 1990) has been developed. The latter includes the contribution of reaction pH to the autohydrolysis in the course of the reaction. A higher degree of biomass breakdown is the result of higher severity pretreatment and usually translates to a higher degree of fermentability and improved yields. At the same time, the higher severity leads to lignin breakdown and generation of compounds inhibitory to enzymatic processing and microbial fermentation (Jönsson and Martín, 2016).

1.2.2 Production of prebiotic GGM from Norway spruce

In a SE based production pipeline, the biomass is first milled to increase its surface to volume ratio. Then, the biomass is steam exploded, resulting in a slurry containing water soluble oligosaccharides and residual solids (Fig. 6). Soluble oligosaccharides from the steam explosion slurry can be extracted by rinsing the biomass with water, and by pressing the liquid fraction out of the soaked biomass. High water volume to biomass ratio (approx. 10 L/kg) improves yields, while producing a high volume of dilute product. Ultrafiltration – filtration through membranes with small pores and 1-100 kDa molecular weight cutoff is an efficient way of fractionating the oligosaccharides and filtering out the potentially harmful carbohydrate and lignin breakdown products (Jönsson and Martín, 2016). Fractions of liquid slurry containing the oligosaccharides with desired characteristics can then be concentrated by nanofiltration with membranes of <1 kDa cutoff, which allow salts, monosaccharides and water to permeate. GGM in solution can be freeze dried or spray dried, however the latter method exposes the oligosaccharides to high (>200° C) temperatures, risking acetate migration.

A microbe's capability to degrade complex carbohydrates relies on how versatile its enzymatic machinery is. Since the conception of this project, it was hypothesized that by selecting production conditions resulting in complex GGM, the number of bacterial taxa capable of consuming the GGM will be reduced. Several commensal bacteria and some potentially health beneficial bacteria are very efficient polysaccharide degraders. A substrate of high complexity could therefore promote a healthy microbiota. Experimentation with steam explosion and ultrafiltration for processing GGM determined the conditions resulting in high number of galactose substitutions and high degree of

acetylation in the product. In the GGM production pipeline, steam exploded slurry was soaked in water at 70° C, which was then pressed out and collected for filtration. Solids were removed by filtering the liquid fraction through a 50 µm membrane, followed by fractionation on a 5 kDa membrane. Oligosaccharides retained by the 5 kDa membrane were concentrated and dried to become the GGM used in the feeding trial. Samples of products from each of these stages are presented in Fig. 5.

Figure 5. Samples taken at each step in the Norway spruce GGM production process. Left to right: dry wood chips, chips milled to <2mm size, steam explosion slurry, liquid fraction of the slurry recovered with the cider press, dried retentate from the 5 kDa cutoff, dried permeate of the 5 kDa filtration.

1.3 Enzymatic hydrolysis of complex mannan *in vivo* and *in vitro*

Enzymatic breakdown of polysaccharides into their constituent sugars is central in both biorefining and decomposition of plant tissue in Nature. In the biorefinery, the process is aimed at producing a highly specified type of end product and requires a specific enzyme cocktail. In polysaccharide degrading microbial ecosystems, such as the gut microbiome of animals, breakdown of complex polysaccharide such as GGM, requires a range of enzyme activities to be broken down into its monosaccharide constituents. A combination of endo β-mannanases (EC 3.2.1.78) break down the polymer (McDonald et al., 2009). In order to enable the activity of mannanases and mannosidases, the α-1,6 bound galactose must be removed by an α-galactosidase (3.2.1.22), and the 2-*O*-, 3-*O*- and 6-*O*-acetylations must be removed by acetyl esterases (3.1.1.72). Oligosaccharides are eventually trimmed down by subsequent removal of mannose and glucose residues by exo-acting β-mannosidases (EC 3.2.1.25) and β-glucosidases (EC 3.2.1.21) into mannobiose and mannosyl glucose,

which are broken down by mannosylglucose phosphorylases (EC 2.4.1.281), and mannobiose phosphorylases (EC 2.4.1.319). Structural representation of a mannooligosaccharide and the enzymes required to process it are depicted in Fig. 6. Constituents of GGM enter glycolysis as mannose, glucose, mannose-1-phosphate and galactose, while the acetate enters the metabolism as Acetyl-CoA.

Plant polysaccharide degraders have developed an arsenal of enzymes to harvest energy from plant biomass. These enzymes are highly valued tools for developing efficient biorefineries and novel, high value products from biomass. As such, they are the focus point of a great deal of research. The characterized enzymes and their structure-function relationships are collected in databases such as CAZy (Carbohydrate Active enZYmes) (Lombard et al., 2014b) and Candb (Yin et al., 2012).

1.3.1 CAZy database.

The CAZy database (Lombard et al., 2014a) classifies carbohydrate active enzymes based on their structural similarity. Enzymes in CAZy are assigned classes based on their general type of activity: glycosyl hydrolases (GH), carbohydrate esterases (CE), glycosyl transferases (GT), polysaccharide lyases (PL) and auxiliary activities (AA). Furthermore, there is a broad group of modular structures without enzymatic activity, namely carbohydrate binding modules (CBMs). Within each class, enzymes are divided into families based on structure similarity, or predicted structure inferred from sequence where structural data is missing. Large families containing enzymes with a wide range of activities are further divided into subfamilies, such as the 56 subfamilies of GH5 hydrolases (Aspeborg et al., 2012). CAZy classification is a useful tool in enzyme discovery and genome annotation. Based on sequence similarity to known and previously classified enzymes, the activity of a protein can be predicted from just the open reading frame (ORF) sequence. Gene clusters for degradation of complex polysaccharides like polysaccharide utilization loci (PULs) can be detected and their function predicted based on the similarity to known polysaccharide degrading enzymes and the proximity to known non-hydrolase genes (susC and susD) in the genome. Besides enzyme sequences, the CAZy database contains a PULDB section (Terrapon et al., 2018) containing annotated PULs from sequenced genomes. The PUL systems studied to date are primarily from the Bacteroides phylum, however, complete clusters from other phyla are beginning to emerge, such as xylan and mannan degrading clusters from the Firmicutes *Roseburia intestinalis* (La Rosa, 2018, Leth et al., 2018a)

Glycosyl hydrolases are enzymes breaking down the glycosidic bonds between two carbohydrate monomers. Hydrolysis occurs either with retention of the anomeric configuration in the newly

formed reducing end, or by its inversion. Retaining hydrolases cleave the glycosidic bond by double-displacement mechanism with a covalent intermediate between the glycosyl and the enzyme (Henrissat and Davies, 1997). Inverting hydrolases cleave the bond by single-displacement mechanism that undergoes an oxocarbenium ion transition state (Mccarter and Withers, 1994). There are 153 GH families classified currently, and the families involve a wide range of specialized enzymes from the genomes of a wide range of organisms. The high complexity of different carbohydrates (including a wide variety of decorations and branchings) requires a wide variety of enzymes to hydrolyze them.

1.3.2 Glycosyl hydrolases breaking down mannans.

For mannan hydrolysis, the relevant GH families are GH1, GH2, GH5, GH26, GH113 and GH134 β-mannanases and β-glucosidases/mannosidases, GH27 and GH36 α-galactosidases, GH130 β-mannan phosphorylases, and CE2 acetyl esterases (Moreira and Filho, 2008, Malgas et al., 2015). A summary of enzymatic activities required for breaking down GGM is presented in Fig. 6.

Both GH1 and GH2 have an $(\beta/\alpha)_8$ fold, and together contain about 36000 proteins. Many of these enzymes display β-mannosidase activities, which remove single Man*p* residues from the non-reducing end of oligosaccharides and are necessary for the final steps of polysaccharide breakdown, just before the monosaccharides enter their appropriate metabolic pathways (Fig. 10) (Chauhan and Gupta, 2017). Families GH5 and GH26 contain retaining β-endomannanases with a $(\beta/\alpha)_8$ barrel fold (Srivastava and Kapoor, 2017). GH5 is one of the largest families of glycoside hydrolases with 12457 members, primarily of bacterial origin (10014 sequences) (Aspeborg et al., 2012). GH5s described so far are active on xylans, xyloglucans, mannans, β-glucans and chitin. The versatility of enzymes clustered in this family indicates the various activities share a general fold that can be adapted to suit new roles. Because of the wide range of activities attributable to the same general fold, the family has been further divided into 56 families. Family GH26 contains mostly (1609/1680) bacterial enzymes, with activities on xylan, lichenan and mannan (Araki et al., 2000, Taylor et al., 2005). Family 134 contains 139 β-mannanases, mostly from eukaryotes. The first GH134 was characterized in *A. nidulans*, and is an inverting mannanase active on glucomannan, with low activity on substituted mannans (Shimizu et al., 2015). Family GH113 contains mannanases, mostly of bacterial origin, which also share a $(\beta/\alpha)_8$ fold and a retaining mechanism (Zhang et al., 2008).

The endo-acting β-1,4 mannanases from GH5 and GH26 family are crucial enzymes for breaking down mannans (Srivastava and Kapoor, 2017). These enzymes break down long mannan chains into short oligosaccharides of appropriate length for intracellular transport, generate more ends for the

exo-acting mannosidases and other enzymes that continue breaking down the carbohydrate. The two mannanases used in the course of this study were an *Aspergillus nidulans* GH5 (Dilokpimol et al., 2011) and a GH26 mannanase from *Roseburia intestinalis* (La Rosa, 2018).

Galactose substitutions of GGM are removed by α-1,6 galactosidases from families GH27 and GH36 remove the substitutions from galacto- and galactoglucomannans, making the β-1,4 linkages in the backbone more accessible to other enzymes. Presence of mannanase and mannosidase activities in many GH families reflects the adaptation of polysaccharide degraders to the complexity of mannan based polysaccharides. Enzymes breaking down the polysaccharide have limited specificities and often cannot access glycosidic bonds adjacent to decorations. Examples of this include mannanase activity on mannan decorated with α-1,6 galactosylations or acetylations (Arnling Bååth et al., 2018) and xylanase reactivity on xylans decorated with glucuronosylation, acetylations and arabinosylations (Leth et al., 2018a).

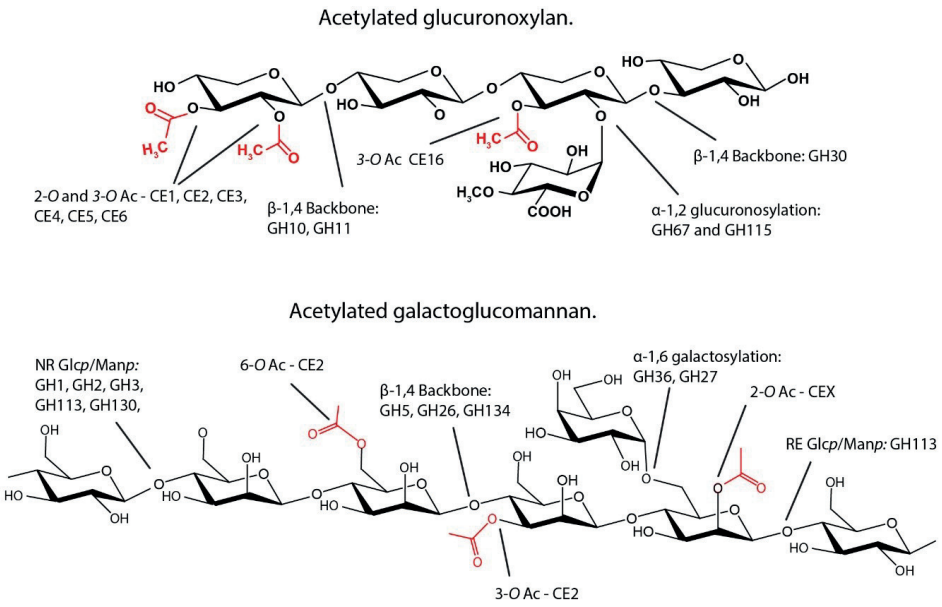


Figure 6. Schematic view of glucuronoxylan and GGM, with CAZy families containing enzymes necessary for the hydrolysis of the particular types of bonds. The backbone of glucoronoxylan is broken down by GH10 and GH11 β-xylanases with various abilities to accept backbone decorations. Glucuronic acids are removed by GH67 and GH115 glucuronosidases. GH30 family xylanases break down xylan specifically in the vicinity of glucuronosylations. Xylan acetylations are removed by esterases from families CE1, CE2, CE3, CE4, CE5, CE6, and when they are adjacent to a glucuronosylation, CE16. β-Mannanases from families GH5, GH26 and GH134 breakdown the mannan backbone. Terminal Manp residues on the reducing end are removed by GH113 β-mannosidases, while on the non-reducing end, the terminal residues are removed by GH1, GH2, GH5, GH113 and GH130 β-mannosidases (if the terminal residue is a Manp), or by GH1, GH3 and GH5 β-glucosidase (if the terminal residue is a Glcp). 2-O- acetylations are removed by CEX homologues, while the 3-O-, 6-O-, and 4-O- (in the reducing end) acetylations are removed by CE2 family esterases NR – non-reducing end, RE – reducing end.

1.3.3 Carbohydrate esterases active on mannans.

Carbohydrate esterases (CEs) cleave the ester bonds between carbohydrates and acyl groups such as the acetylation on mannans. All CEs that are active on mannan are so far grouped together with xylan esterases under the (EC 3.1.1.72) classification. CAZy currently contains 16 carbohydrate esterase families, of which the CE2 family currently is the most relevant for mannan deacetylation. Family CE2 esterases are all classified as acetyl xylan esterases (AcXEs) (EC 3.1.1.72). CE2 esterases share a two-domain structure, consisting of a GDSL2 hydrolytic domain, and an accessory jellyroll domain (Nakamura et al., 2017). Some CE2s have been suggested to be specific towards 6-*O*-acetylation (Topakas et al., 2010b). Two *Cellvibrio japonicus* esterases, *Cj*CE2B and *Cj*CE2C from the CE2 family have shown higher k_{cat} on glucomannan than xylan (Montanier et al., 2009).

Mannan esterase activity was described already in 1992 when six known polysaccharide degraders were surveyed: *T. reesei, A. awamori, A. oryzae, S. commune, Aureobasidium pullulans* and *Streptomyces olivochromogenes.* Culture filtrates of all of them contained acetyl glucomannan esterase activity, with *T. reesei* and *A. pullulans* having considerable deacetylation capabilities. However, no sequence data nor structure of these enzymes are available (Tenkanen et al., 1993).

1.3.4 Enzymatic breakdown of xylan.

Enzymatic hydrolysis of xylan has been researched in the context of using xylan rich biomass as a feedstock for bioethanol production (Dodd and Cann, 2009). Much like mannan, the basic structure of xylan in woody biomass consists of a β-1,4 linked backbone, with substitutions like glucuronosylations, arabinosylations and acetylations (Fig 6). Methyl glucuronic acids are a common substituent on birch and spruce xylan, approximately one in ten xylose residues is decorated with an α-1,2 linked glucuronic acid, which may or may not carry an *O*-methylation on carbon C4 (Biely et al., 2016). α-1,2 and α-1,3 linked arabinose is a common substituent of xylans in monocots, such as the industrially important switchgrass, corn stover and sugarcane bagasse (Scheller and Ulvskov, 2010). Glucuronic acids are removed from the backbone by α-glucuronidases from CAZy families GH67 and GH115. The backbone itself is hydrolyzed by xylanases from CAZy families GH10, GH11, and GH30. GH30 xylanases are dependent on the presence of glucuronic acids on the xylose residue penultimate to the reducing end and have very little activity on unsubstituted xylan (Šuchová et al., 2018). Acetylations on the xylan backbone are removed by acetylxylan esterases from families CE1, CE2, CE3, CE4, CE5, CE6 and CE16, which have their particular specificities towards 2-*O*- and 3-*O*-acetylations (Nakamura et al., 2017). β-xylosidases from families GH3, GH39, GH43 and GH52

remove single xylose moieties from the non-reducing ends of xylo-oligosaccharides. Together these enzymes are involved in the complete degradation of xylan, which may be utilized in various ways in a biorefinery approach, or which are utilized in microbes to generate entities that then enter the cellular metabolism.

1.3.5 *In vivo* breakdown of xylan and mannan in the gut.

Polysaccharide-degrading microbes in the gastrointestinal tract (GIT) have evolved sophisticated enzymatic toolboxes to harvest energy from complex dietary fibers (Flint et al., 2012). Strategies vary depending on the bacterial strain, type of glycan and the host. Two key aspects of polysaccharide degradation are enzymatic breakdown and transport of oligosaccharides into the cell. Enzymes breaking down polysaccharides can be secreted into the cell surroundings (Gilbert et al., 2008), present in the cytosol, or assembled into cellulosomes (Flint et al., 2008). Cellulosomes are protein aggregates consisting of a scaffoldin protein anchored to the extracellular matrix, to which multiple CAZymes are attached by cohesion-dockerin interactions. Cellulosomes are multifunctional tools for binding and hydrolysis of polysaccharides in the immediate vicinity of the bacterium (Fig. 7).

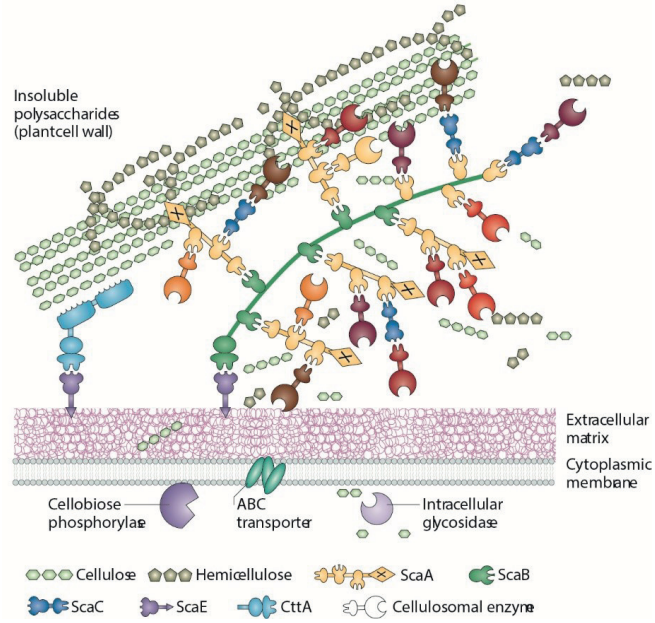


Figure 7. Schematic presentation of a cellulosome from Ruminococcus flavefaciens. ScaA, ScaB, ScaC and ScaE proteins assemble into a scaffold anchored in the cytoplasmic membrane by ScaA. Hydrolytic enzymes attach via dockerin – adhesion interactions forming a multifunctional aggregate with a wide range of specificities releasing processed oligosaccharides in the vicinity of ABC transporters, which facilitates internalization to the cytosol. Figure adapted from (Flint et al., 2008)

The two most dominant phyla in human and animal gut microbiomes, Firmicutes and Bacteroidetes, have developed two distinct strategies for ingesting carbohydrates. For Bacteroidetes, carbohydrate uptake requires a complex transport system spanning the outer membrane, periplasmic space and inner cell membrane. The starch utilization system (Sus) of *Bacteroides thetaiotaomicron* is a paradigm for carbohydrate uptake and breakdown by these gram negative bacteria (Shipman et al., 2000). Sus systems share a number of highly conserved genes responsible for glycan binding, breakdown and transport through the outer membrane and periplasmic space. Conservation of the crucial SusD (glycan binding and channeling towards the other surface enzymes), SusC (periplasmic space transporter), and SusR (transcription activator) proteins is used in metagenomics prospecting and genome annotation (Fig. 8). Expression of genes in a particular PUL is induced by the SusR – like protein in a positive feedback response to the target polysaccharide. PULs with Sus-like proteins transporting and breaking down mannans and xylans have been found in Bacteroidetes (Rogowski et al., 2015, Cuskin et al., 2015). These genes are located in PULs adjacent to polysaccharide degrading enzymes. Searching newly sequenced genomes and proteomes for Sus-like proteins is a common way of identifying polysaccharide degrading enzymes and determining the metabolic capacity of microbes. It is also a way of prospecting genomes and metagenomes for novel CAZymes for industrial applications.

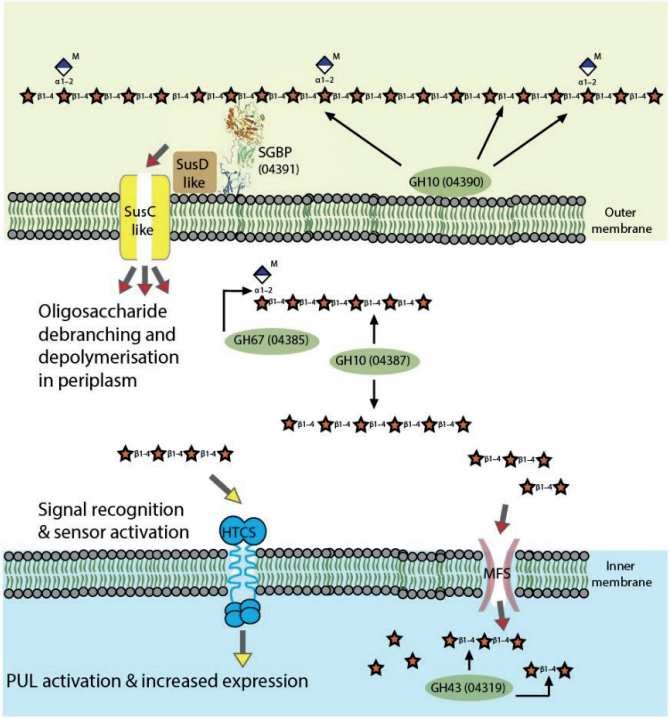


Figure 8. Bacteroidetes ovatus xylan PUL operates in a similar manner to the canonical starch utilization system of Bacteroides thetaiotaomicron. Glucuronoxylan is degraded by an outer membrane attached GH10 family endoxylanases, producing xylo-oligosaccharides which can then be transported by the SusC homologue into the periplasmic space. In the periplasmic space xylo-oligosaccharides are debranched and trimmed further into xylobiose, which is transported into the cytosol and hydrolyzed to xylose by a GH43 family xylosidase. Xylotetrose in the periplasmic space is detected by a hybrid two component system – HTCS, which induces transcription of the enzymes encoded in the PUL. Figure adapted from (Rogowski et al., 2015).

Roseburia intestinalis is a gram positive commensal bacterium from the Firmicutes phylum, present in the gut microbiomes of pigs and humans. *Roseburia* spp. are dietary fiber degraders that produce butyrate and have significant impact on the health of their hosts (Tamanai-Shacoori et al., 2017). Unlike *Bacteroidetes, Roseburia* have a different PULs system, and do not have a Sus-like transport systems, but rely on extracellular enzymes for breakdown of polysaccharides in the environment and ATP-binding cassette (ABC) transporters to internalize the oligosaccharides (Leth et al., 2018b). ABC transporters are a part of the polysaccharide utilization apparatus and are expressed in response to polysaccharides in the environment (Scott et al., 2011). In the course of this project we examined the mannan degradation capability of *R. intestinalis (La Rosa, 2018)*, manuscript under revision), while a group of our collaborators at DTU in Denmark examined the same strains ability to degrade complex xylan (Leth et al., 2018b). Breakdown of glucuronoxylans and GGM by *R. intestinalis* is shown in Fig. 9 and 10.

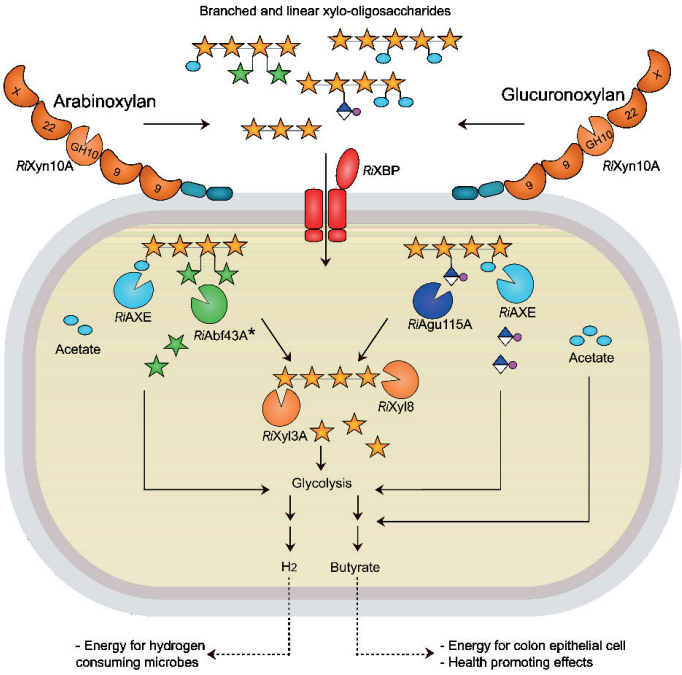


Figure 9. The xylan degradation apparatus of R. intestinalis: Xylans in the cells surroundings are broken down by a GH10 xylanase, RiXyn10A, on the cell surface. Xylo-oligosaccharides are transported by the RiXBP ABC transporter into the cytoplasm. In the cytosol, xylo-oligos are deacetylated by the RiAXE esterase, α-1,2 α-1,3 arabinosylations are removed by the α-L-arabinofuranosidase (RiAbf43A) and glucuronic acids are removed by a glucuronidase (GH115 RiAgu115A). Once the xylo-oligosaccharides are clear of decorations, they are hydrolyzed by GH3 and GH8 xylosidases into xylose. Xylose and arabinose are converted to xylulose 5-phosphate and enter the cellular metabolism via the pentose phosphate pathway, whereas methyl-glucuronic acid enters glycolysis as 2-oxo-3-deoxygalactonate 6-phosphate. Figure adapted from (Leth et al., 2018b).

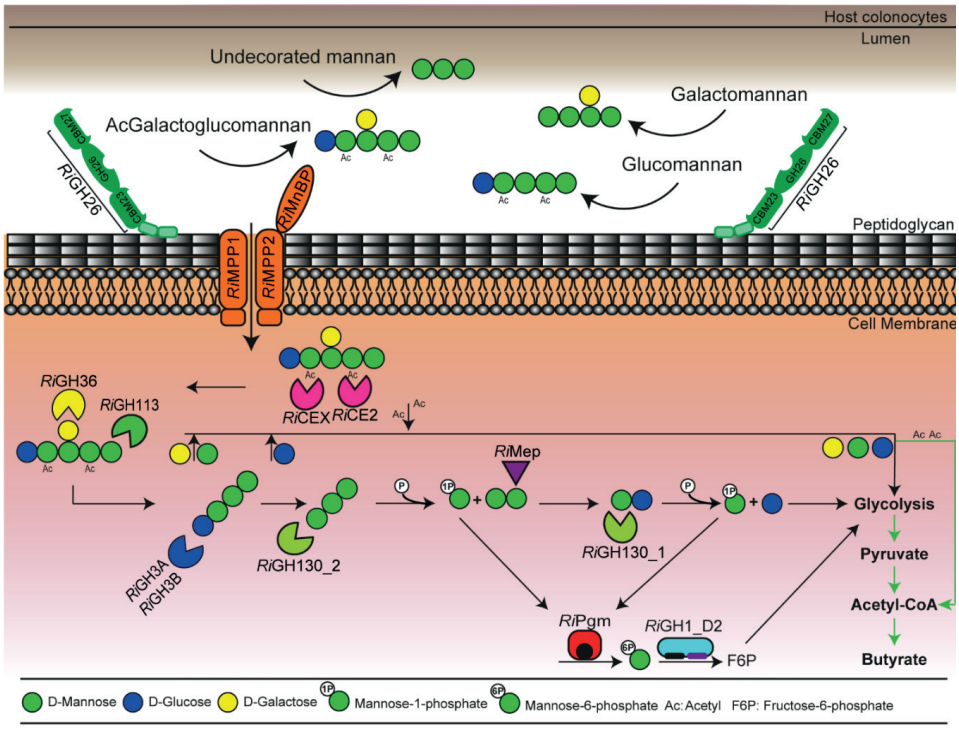


Figure 10. The mannan degradation apparatus of Roseburia intestinalis: mannans in the environment are hydrolyzed by a cell wall anchored extracellular GH26 β-mannanase, which breaks the polysaccharide into oligosaccharides suitable for membrane transport. Once the oligosaccharide is transported into the cytosol, acetylations are removed by the CEX and CE2 esterases, and galactose substitutions are removed by a GH36 α-galactosidase. The backbone of oligosaccharides is broken down by two GH3 β-glucosidases removing glucose from the non-reducing end, or a GH130 manno-oligosaccharide phosphorylase removing mannose from the non-reducing end in the form of mannose-1-phosphate. A GH113 mannosidase removes mannose from the reducing end. Mannose-1-phosphate released by GH130 manno-oligosaccharide phosphorylase (RiGH130_2) is converted into mannose-6-phosphate by a phosphoglucomutase, and then converted into fructose-6-phosphate by a GH1 isomerase. The mannan backbone is eventually broken down into mannobiose, which is converted to mannosylglucose and broken down by GH130 manno-oligosaccharide phosphorylase (RiGH130_1). The end products of mannan degradation are fructose-6-phosphate, galactose, glucose and acetate. The carbohydrates enter cellular metabolism via glycolysis, while acetate enters the metabolism as Acetyl-CoA, all ultimately becoming butyrate which is released from the bacterial cell and taken up by the gut epithelium.

1.4 The gastrointestinal tract microbiome and its interactions with the host

The lower gastrointestinal tract (GIT) of all animals is home to one of the most complex microbial ecosystem known in Nature. The primary role of the GIT microbiome is in digestion, fermenting complex polysaccharides commonly termed ‘dietary fibers’ into short chain fatty acids (SCFAs) and other compounds that the host GIT epithelium can absorb. GIT microbiomes of humans are dominated by *Bacteroides* and *Firmicutes*, as are those of pigs. While the human genome encodes 17 glycoside hydrolase (GH) enzymes involved in

carbohydrate nutrient metabolism (Cantarel et al., 2012), the genome of a common polysaccharide degrader in the GIT, *Bacteroides thetaiotaomicron* contains genes for 236 GH enzymes and 15 polysaccharide lyases (Flint et al., 2008). *B. thetaiotaomicron* is one example of a microbe with an extremely large GH toolbox, other members of *Bacteroidetes* average 136 GH genes per genome, while the more specialized *Firmicutes* average 39.6, just over twice that of their human hosts (El Kaoutari et al., 2013). Glycan metabolism in GIT microbiomes is vastly superior to that of mammalian hosts, and essential to the utilization of complex polysaccharides.

The GIT microbiome produces an array of compounds necessary for the host health. SCFAs are the primary microbial metabolite utilized by the host tissues. Butyric acid is especially important for gut health, as it is the preferential energy source for colonocytes (Bedford and Gong, 2018). Propionic acid has also been identified as beneficial to gut health by nourishing the epithelium. SCFAs in diet affect the regulatory T cells of the immune system, reduce the secretion of interleukin 18, increase mucus secretion, increase the proliferation and shedding of the gut epithelium, and reduce inflammation in the GIT (Rooks and Garrett, 2016). Furthermore, SCFAs as well as tryptophan are also neuroactive metabolites, which have an effect on the behavior and mental state of the host. Tryptophan is taken up from the lower GIT (Cryan and Dinan, 2012). Both SCFAs and tryptophan have been linked to significant changes in the behavior and gene expression in the brains of germ free mice (Diaz Heijtz et al., 2011). Besides SCFAs, the human gut microbiome also have the metabolic capacity to produce a number of vitamins (Magnúsdóttir et al., 2015).

Gut microbiota impacts a range of aspects of the host metabolism, such as fat storage (Backhed et al., 2004), hunger and feeding behaviors, and energy expenditure (Li et al., 2008). Reduced microbiome diversity and stability are considered detrimental to health (Cammarota et al., 2015). Distinct microbiome profiles have been linked to metabolic diseases, type 2 diabetes and obesity (Li et al., 2008). Besides the digestive tract, communication between the microbiome and the central nervous system has been characterized and implicated in the *in utero* development of the human nervous system (Sharon et al., 2016). Neurological diseases such as schizophrenia, depression and autism have been linked to specific microbiome shifts. GIT diseases such as Crohn's disease (CD), irritable bowel syndrome (IBS) and ulcerative colitis (UC) have also been linked to

microbiome dysbiosis. Multiple studies on the involvement of the GIT microbiome in these conditions have linked the diseases to a dysbiosis in a range of genera within the Proteobacteria and Bacteroidetes phyla. Some of these genera have been shown to produce pro-inflammatory metabolites such as hydrogen sulfide (Cammarota et al., 2015). Two of the bacterial genera studied in the course of this project: *Roseburia* and *Faecalibacterium* have also been linked to GIT diseases. Decreases in the abundance of *Roseburia* and *Faecalibacterium* that been linked to UC, CD (Machiels et al., 2014), and IBS (Willing et al., 2010).

1.4.1 The gut microbiome of pigs

The human microbiome has been intensively studied due to its implications in health and diseases. The pig gut microbiome has been studied as a potential entry point for interventions aimed at increasing the health and efficiency of livestock. GIT microbiome of pigs differs depending on their race, gender, age, and most importantly the composition of their feed (Frese et al., 2015). A meta-study looking at porcine microbiota has identified a number of bacteria taxa shared between >90% of samples from of the same GIT section (duodenum, jejunum, ileum, cecum, colon and feces), which can be considered a standard pig gut microbiome (Holman et al., 2017). Firmicutes and Bacteroidetes phyla accounted for nearly 85% of the total 16S rRNA gene sequences from the 939 samples used in the study. *Prevotella, Clostridium, Alloprevotella*, and *Ruminococcus* were present in >99% of faeces samples used in the meta-study. At the genus level, *Clostridium*, *Blautia*, *Lactobacillus*, *Prevotella*, *Ruminococcus*, *Roseburia*, the RC9 gut group, and *Subdoligranulum* were found in more than 90% of samples. Composition of this core microbiome samples is summarized in Fig. 11.

1.4.2 Antibiotics in pig feed

Since 1940s, animal protein factors (APF) such as *Streptomyces aureofaciens* fermentation byproducts were used to improve the yields in chicken farming (Stokstad et al., 1949). Success with poultry production lead to extending the use of animal protein factor (APF) and antibiotic growth promoters (AGPs) to pork farming (Gaskins et al., 2002). The specific functionality of

APF was eventually determined to be caused by the antibiotics that controlled the microbiome of the animals and exerted systemic effects (Gonzalez Ronquillo and Angeles Hernandez, 2017). Antibiotics such as penicillin, ionophores, tertracyclins and streptogramins (Brown et al., 2017) have been added to pig and poultry feed to improve efficiency and reduce the risks of infections (Looft et al., 2012a). The European Commission has issued a ban on the addition of antibiotics and antibiotic growth promoters in animal feed (Regulation (EC) No 1831/2003) which became effective in 2006. Even before that, some countries banned AGPs due to concerns about human health and increasing antibiotic resistance (Sweden in 1986, Denmark in 1999) (Holt et al., 2011). Appearance of antibiotic resistance as a result of AGP in farming is a serious issue, since bacteria are capable of sharing genes for antibiotic resistance by means of horizontal transfer. Strains of vancomycin resistant *Enterococcus faecium* have been detected and the rise in resistance linked to the use of avoparcin in farming (Bager et al., 1997).

Since the ban on AGPs, farming industry has faced significant challenges in maintaining animal health and efficiency. Antibiotics contained in feed improved the efficiency by controlling microbial ecosystem in the gastrointestinal tract. The lack of antibiotic control on the gut microbiome is especially damaging during weaning, the process of switching the animals from the mother's milk to plant derived feed. Losses in farming are not just monetary, removing antibiotics from feed has resulted in a decline in animal health and an increase in piglet mortality resulting from post-weaning anorexia and diarrhea. Piglet mortality in antibiotic-free farming can be as high as 17% (Lallès et al., 2007). The exact mechanism of growth promotion by antibiotics in feed is still not understood, although a number of studies have shown that antibiotics have, not surprisingly, shifted the composition of the GIT microbiome (Holman and Chénier, 2014, Looft et al., 2012b).

1.4.3 The pig gut microbiome and the microbial population shift during weaning

The porcine gut microbiome, like that of any mammal, begins its development as a milk-oligosaccharide degrading community, and develops into the mature, polysaccharide degrading microbiome as the adult food is gradually introduced (Frese et al., 2015). Weaning in Nature should occur at about four months of the piglets' life, once the immune system of

the piglet is more developed, and happens by gradual introduction of new types of foods. In intensive farming, piglets are weaned at about 3-4 weeks of life, and even as early as at 2 weeks of age in some cases (Gresse et al., 2017). This dietary shift is abrupt and coincides with separation from the sow and introduction to pens with other animals from outside the litter. All this happens in a period when the immune system of the piglet is still developing and is reliant on immune factors present in milk (Lalles et al., 2007). The stress often results in low to no food and water intake in the first 24-48 hours after the piglet is removed from the sow, which causes an inflammation to the gastrointestinal tract epithelium. Weaning results in transient changes in gut permeability to toxins and hormones, reduction in villi height and in nutrient absorption. All the above mentioned factors contribute to a lower feed conversion in the following days, and in overall lower efficiency of meat production (Gresse et al., 2017).

During weaning, the gut microbiome undergoes an abrupt shift from the milk degrading Bacteroides, Bifidobacteria and Lactobacilli to complex polysaccharide degrading Firmicutes and Bacteroidetes (Gresse et al., 2017). While the mature, established microbiome is a symbiotic ecosystem with enormous positive impact on the health of the host, the transition period is associated with high risk of infections from pathogens. Animals with a fully matured and nourished gut microbiome are less susceptible to infections as the incoming pathogens are outcompeted by the commensal bacteria.

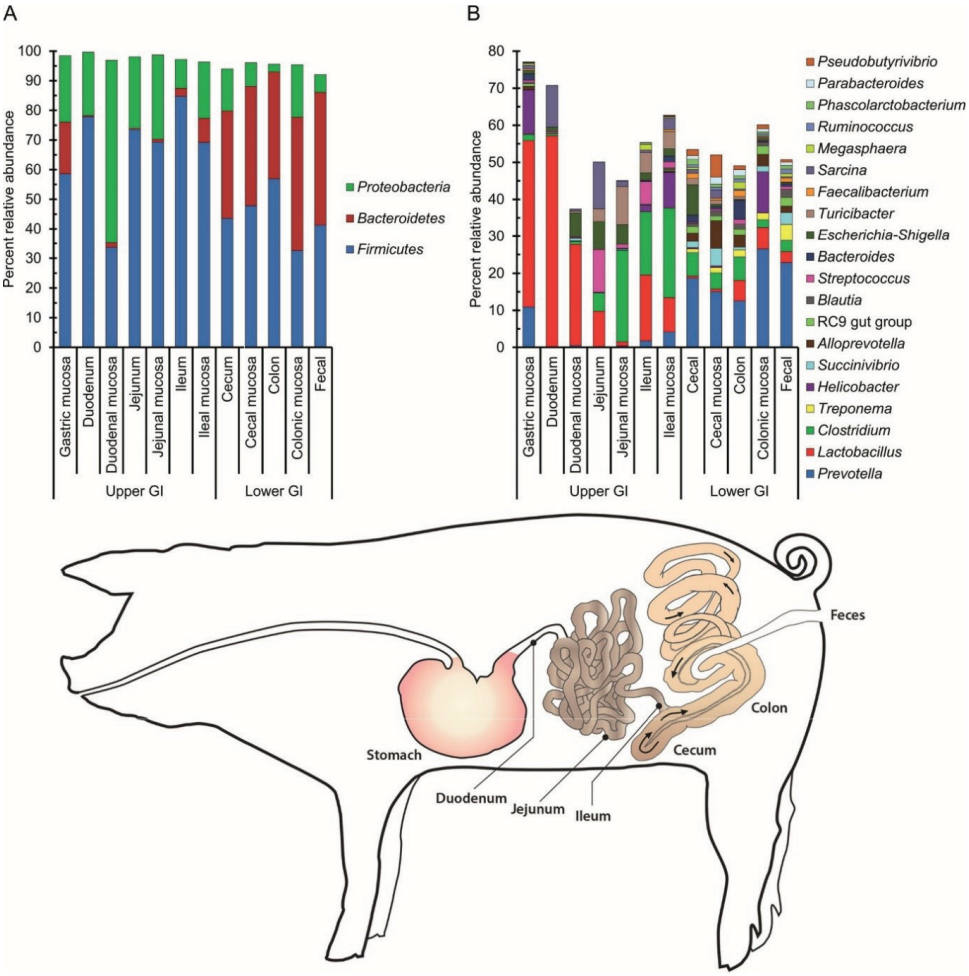


Figure 11. The pig gastrointestinal tract consists of the stomach, duodenum and jejunum in the upper GIT, cecum and colon in the lower GIT as sketched in the lower part of the figure. A: Bar plots representing the relative abundances of the three dominant phyla: Proteobacteria, Bacteroides and Firmicutes in each GIT compartment. B: The same distribution represented at the genus level shows much more divergence between sites. Lactobacilli are more prevalent in the upper GI, while Prevotella dominate the lower GI and fecal samples. Lower GI is also more diverse than the upper GI section. The microbiome composition differs between the GIT compartment lumen and the mucosal layer, showing that the microbial diversity extends along and across the GIT. The figure is adapted from (Holman et al., 2017)

1.5 Multi-omics analysis of microbial communities

In depth analysis of complex microbial ecosystems such as the gut microbiome are based on nucleic acids and protein sequencing assisted by bioinformatics (Kuczynski et al., 2011). The vast majority of microbes inhabiting the gut cannot be cultured *in vitro* or analyzed with conventional microbiological methods. High throughput sequencing of genetic material and mass spectrometry based sequencing of proteins extracted directly from the host gut are culture-independent approaches that allow the analysis of the entire microbial community and its metabolism.

Genes encoding the 16S subunit of ribosomes are ubiquitous in Nature, and contain variable regions which are mutation hotspots. Based on the sequence of the variable regions, taxonomical assignment of the bacterium can be inferred with relatively high certainty. Amplified DNA is sequenced using the Illumina MiSeq platform, and analyzed with bioinformatics pipelines such as Qiime (Caporaso et al., 2010), which can remove errors, chimeric sequences arising in PCR and allow for statistical interpretation of the sequence data. 16S rRNA gene analysis provides an information regarding the relative abundance of each operational taxonomic unit (OTU). Taxonomy is assigned using the closest taxonomic unit and the species level taxonomy is retrieved when possible. However, because of the resolution of the 16S analysis, taxonomical assignment at the species level is not always feasible (e.g. for new or unknown microbes) and the classification might occur at a higher taxonomic level (genus, family). 16S analysis produces an overview of the microbial community, allowing for determination of dominant taxa and identification of pathogens.

Whole metagenome shotgun sequencing gives a more in-depth overview of the bacterial community, and has the potential to sequence and assemble genomes from new, previously unseen and uncultivated microbes. The total nucleic acids content of a sample is extracted and sequenced, producing a complex mixture of short reads which are then assembled into contigs. Based on taxa-specific sequence features such as *k*-mer and GC content, contigs are binned together into metagenomics assembled genomes (MAGs). Once assembled, MAGs are annotated with taxonomy by sequence comparison with sequences in reference databases, such as MiGA (Rodriguez et al., 2018). Protein sequences are annotated with predicted function based on homology to sequences of characterized enzymes. At this stage, identification of known carbohydrate active enzyme (CAZymes), starch utilization system

(Sus) -like proteins and clusters of polysaccharide degrading enzymes sheds a light on the bacteria with polysaccharide degradation potential. Hemicellulose degradation machinery, such as those described in Fig. 8, 9 and 10, can be located within the genomes. For in depth studies, these degradation clusters may be recombinantly expressed, in parts or all proteins, for *in vitro* characterization. Collection of MAGs from a sample serves as a blueprint for the metabolic capacity of the microbial ecosystem.

Metaproteomics analysis detects the levels of protein expression, which can then be mapped to enzyme coding sequences present in particular MAGs. Metaproteomic analysis begins with an extraction and purification of the total protein content of the sample. Proteins are fractionated by gel electrophoresis and then digested into peptides with a proteolytic enzyme with a known cleavage site (i.e. trypsin). HPLC-MS/MS (Karpievitch et al., 2010) is used to determine the mass of the peptides in the hydrolysate and quantify each of the peptides. *In silico* analyses are then used to identify the sequences of the peptides and the proteins from the MS/MS spectra, either via *de novo* approach or by matching against a database. The former determine the peptide sequence based on the mass of the peptide obtained in MS1 and the fragmentation spectra obtained in MS2 without prior knowledge of the sequence. The latter generates a theoretical list of peptides from a known protein database and try to match the experimental spectra against the theoretical list. Expression levels are inferred from peptide abundance based on the intensity of the MS spectra. Protein sequences can be related to the sequences of genes in MAGs. Linking protein expression data with the genomes adds a dynamic and functional aspect to the genetic potential shown by metagenomics. There is considerable redundancy and ambiguity among enzyme sequences in the CAZy database, and not all microbes with a predicted mannanase in the genome will utilize the mannan. Proteomic-based annotation of MAGs can be used to find out which sequences are expressed in response to polysaccharide, and reconstruct the metabolic pathways affected by GGM (De Filippo et al., 2012). Production of SCFAs and other metabolites can be reconstructed in the same manner. Integrated multi-omics are the leading approach for functional characterization of human and animal microbiomes (Zoetendal et al., 2008).

In the feeding trial described in paper III, 16s rRNA analyses were used to investigate the shifts in relative abundance of particular taxa in response to GGM. Total nucleic acid content of digesta was extracted and amplified using PCR primers for the V4 variable region, which

is a common target for amplification (Yang et al., 2016, Frese et al., 2015), and was previously successfully used to characterize porcine GIT microbiome. 16s rRNA analysis of fecal samples collected weekly throughout the feeding trial allowed us to investigate the post-weaning development of the microbiome under the influence of GGM and without it. 16S rRNA analysis of digesta from compartments along the GIT identified the spatial distribution of the microbial community. Once 16S rRNA analysis established that colon was the GIT compartment where GGM effects were occurring, samples of colon were picked for shotgun metagenomics. Sequences obtained were assembled with metaSPAdes (Nurk et al., 2017) to produce metagenome assembled genomes (MAGs). MAG content was assigned taxonomy and searched for the presence of mannan degradation genes in the genomes of colon commensal bacteria. A subset of colon samples was used for metaproteomics analysis. Detected proteins were annotated to their predicted MAGs of origin, creating an integrated omics picture of a microbial community. A summary of the omics pipeline used in paper III is presented in Fig.12.

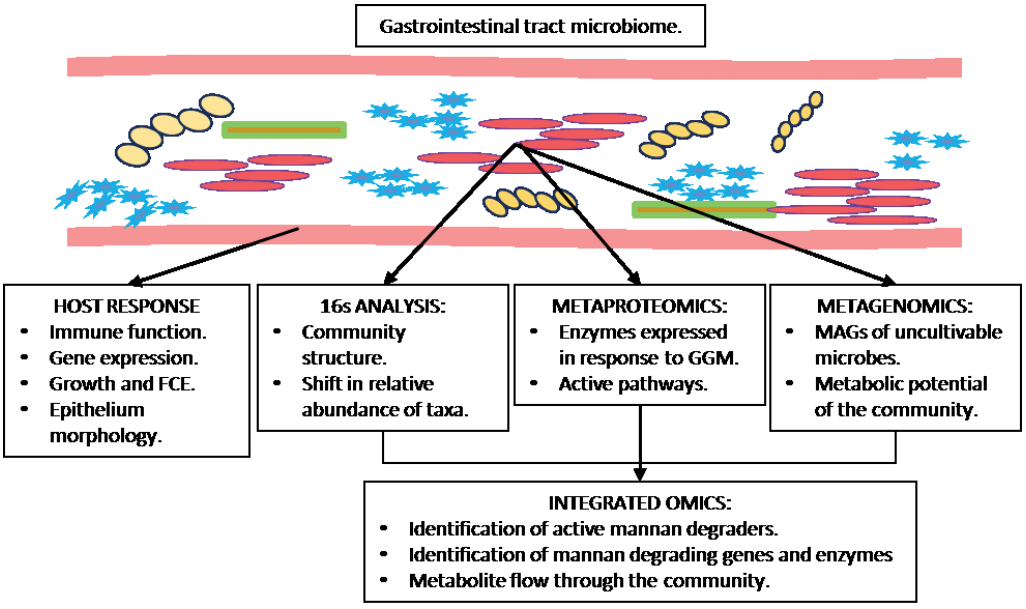


Figure 12. The pipeline of the integrated omics analysis used in the characterization of the GIT microbiome and the effects of mannan inclusion in diet as used in paper III.

1.6 Analysis involved in enzyme and carbohydrate characterization

Hemicellulose consists of many highly similar building blocks, and requires high resolution analytics to characterize. Analysis of mannans is complicated by the high similarity of its constituents: all hexose monosaccharides have identical mass and differ only in the orientation of single hydroxyl groups. In the course of this project, a number of analytical methods was used to characterize the mannans as a bulk compound (the hydrothermal extraction product) and fine structures in oligosaccharides to determine enzyme preferences.

Monosaccharide composition of GGM and other hemicellulose components released in steam explosion was determined by gas chromatography – flame ionization detection (GC-FID). GC can separate monosaccharides, and provides highly quantitative data. Carbohydrates in the sample were hydrolyzed with concentrated H_2SO_4, derivatized to alditol acetates prior to analysis according to (Englyst et al., 1994). Analytical and preparative scale size exclusion chromatography (SEC) coupled with refractive index detection was used to determine the oligosaccharide size range and concentrations in a quantitative manner. SEC allows for an easy identification of the DP range in a sample based on comparisons between analytes and standards of known DP. The method is simple and requires minimal sample preparation, although it proved to be very sensitive to the presence of salt and small molecule contaminants released in steam explosion. In production of tailored oligosaccharides, preparative scale SEC was used as a first step. Fractions with a narrow size range were collected, and subsequently separated by structural features on the preparative HILIC and fractionated to become substrates for enzyme assays. Analysis of acetate content of oligosaccharides, dry biomass, as well as SCFA content of digesta samples were all conducted using HPLC-UV with a REZEX ROA-Organic Acid ion exclusion column. This method was cheap, robust, reproducible, and sensitive enough to be used for measuring acetate release in time course reactions with esterases.

To evaluate the products of steam explosion trials, a number of colorimetric carbohydrate analytics methods were used. These included colorimetric determination of total carbohydrate concentrations according to (Dubois et al., 1956) and reducing sugars according to (Miller, 1959). The Dubois method is based on measuring the color intensity (λ=490 nm) of carbohydrate samples treated with sulfuric acid and phenol. The Miller

method estimates the molarity of reducing ends of oligosaccharides in a solution based on their reaction with 3,5-dinitrosalicylic acid, and measuring the absorbance of products at λ=575 nm.

Matrix assisted laser desorption-ionization time of flight (MALDI-ToF) (Karas et al., 1985) analysis has been the most widely used analytical method in the course of this project. In MALDI-ToF, the mass to charge ration (*m/z*) of the compounds in the analyte and their relative amounts are devised from the way the analytes travel in an electrical field. Analytes are ionized with a laser beam and are moving towards the detector through an electric field in vacuum. Their speed is proportional to the magnetic attraction of the charge on the analyte, and inversely proportional to its mass, which results in larger compounds traveling slowly. Amounts of the analyte reaching the detector are recorded as peak intensities. MALDI-ToF is a useful tool to determine the length range of oligosaccharides in a sample, degree of acetylation, and the presence of other substituents such as feruloylations or glucuronosylations. MALDI-ToF can discriminate between hexose and pentose based oligosaccharides however oligosaccharides containing different numbers of the same monosaccharides cannot be discriminated since hexoses are isobaric (exactly the same mass). The disadvantage of this method is that the peak intensities are relative and poorly correlate to factual abundance in the sample (Grant et al., 2003). Since the masses and *m/z* of all hexose are identical, MALDI-ToF does not provide any information on the composition or structure of the oligosaccharides.

Structural features of oligosaccharides, such as the orientations of hydroxyl groups, acetylation sites and the presence of branchings affect retention of oligosaccharides in hydrophilic interaction chromatography (HILIC) (Alpert, 1990). In HILIC, the hydrophilic analytes are applied to a column with a polar stationary phase as a solution containing a high amount (75%) organic solvent, usually acetonitrile. Under these conditions, analytes adhere to the stationary phase, and a water enriched layer of solvent forms on the stationary phase. A gradient of increasingly polar solvent is passed through the column, and the carbohydrates gradually detach from the stationary phase into the water-enriched layer and into the liquid phase and elute based on their size and structural features. HILIC elution segregated the manno-oligosaccharides based on the presence and location of branchings and acetylations. In our analytical pipeline, HILIC was coupled with an ion trap mass spectrometer and a

charged aerosol detector (CAD). Splitting the flow of analyte into a CAD detector provided a good quantitative view of the amounts of analytes, while MS identified the exact masses of the analytes. The HILIC –MS/CAD required more sample preparation than MALDI-ToF, but provided more quantitative and structural information about the analyzed oligosaccharides. Another advantage of the HILIC-CAD analysis is that it was easily scaled to a preparative column. At the preparative scale, the amount of sample applied could be increased to tens of milligrams, meaning that multiple milligrams of oligosaccharides with highly defined structures could be recovered as fractions in a single run. Collected fractions could then be analyzed further by NMR, to determine their exact structure. Pairing of HILIC purification and NMR characterization of products allowed for some 'rules of thumb' to be developed about the impact of oligosaccharide fine structure (location of acetylations, monosaccharide composition) and HILIC retention behavior.

Characterization of lignin in steam explosion samples was performed at NMBU using a Bruker Ascend 400 spectrometer (400 MHz). Lignin was extracted from the dried, extracted solids using DMSO and analyzed by Heteronuclear Single Quantum Coherence (HSQC) spectroscopy recorded with a spectral width of 0 – 12 ppm and 0 – 250 ppm in ^{1}H and ^{13}C, respectively. For the characterization of acetylation of GGM and esterase activity, NMR experiments were analyzed on a BRUKER AVIIIHD 800 MHz by our collaborators at the Department of Biotechnology and Food Science at NTNU. Reactions were also followed with time resolved by recording 1D proton NMR spectra at 5 minute intervals with a total of 200 time points.

In the process part of the project, multiple optimization runs of steam explosion extraction and purification of GGM were conducted and evaluated for the desired characteristics. DP, DA, oligosaccharide size distribution and total yields were the criteria. Those were evaluated by analytical SEC, MALDI-ToF, ion exclusion HPLC, and colorimetric methods. Colorimetric methods have proven to be superior to HPLC when analyzing steam explosion products with high salt content. When SEC and MS failed, DP of samples was estimated from the proportion of total carbohydrate content and content of reducing sugars in a sample.

During esterase activity characterization, presence of acetylated oligosaccharides, acetate content of samples, effects of oligosaccharide structure on enzyme activity, and the

preference for particular acetylations were assessed by NMR, MALDI-ToF and HILIC-MS. Preparative HILIC-CAD/MS and NMR had great resolution and proved indispensable in assessing acetylation patterns. To determine the presence of acetylations in substrates and their removal by esterases, HSQC NMR spectra were taken before addition of enzyme and at the reaction endpoint. In time resolved experiments, NMR was able to measure the reaction rates for deacetylation of each particular acetylation. Preparative scale HILIC was used to produce some of the substrates used for NMR analysis, producing very homogenous and tailored substrates, such as the 2-*O*- acetylated mannotriose used for characterizing the acetyl migration in response to temperature and pH.

2. Project aims

The main aims of the project were to develop a steam explosion based process pipeline for production of highly decorated GGM from Norway spruce. In the process, develop the NMBU pilot biorefinery, and research the steam explosion process with a focus complex oligosaccharides. Once a process was develop, the goal was to scale it up produce sufficient amounts of GGM for a large scale feeding trial. In parallel to the process research, we aimed at gaining in-depth understanding about the microbial utilization of GGM by investigating the mannan degradation apparatus in gut commensal bacterium, *Roseburia intestinalis.* Finally, we aimed to evaluate the effects of GGM in feed on the health and microbiome composition of weaned piglets, search for novel enzymes, pathways and microbes implicated in the utilization of mannans.

3. Main results and discussion

Manuscripts contained in this thesis describe the consecutive stages of our interdisciplinary effort into evaluating Norway spruce GGM as a potential prebiotic. Early experiments with steam explosion conditions led us to realize that the complexity of produced oligosaccharides was the result of lower severity pretreatment. Based on the understanding of the role of dietary fibers in stimulating the GIT microbiome, increased complexity seemed to be a highly desired characteristic. A substrate containing increased amounts of acetylation, galactosylations and DP would selectively promote growth of gut microbes with a more adapted mannan degrading apparatus. This hypothesis lead us to exploring the possibility of mitigating pretreatment severity by adjusting the reaction pH.

Results of this experiment were described in paper I. This paper highlighted the role of hemicellulose acetylation in hydrothermal pretreatment and extraction. Autohydrolysis of hemicellulose by acetate released from mannan and xylan has proven to have a high impact on the structural features of the extracts. Adjustment of reaction pH resulted in vastly different products without changing the residence time, pressure or temperature. Each one of the six buffered conditions used in the experiment released a product with a different monosaccharide composition, and each one was different from the unbuffered control (paper I, Fig. 3). Acetate, however, was not preserved on the extracted oligosaccharides, nor remained on the solid biomass. Acetylations are labile to higher pH, when increasing reaction pH, more acetylations were removed from the hemicellulose, while at the same time the increased pH inhibited polysaccharide autohydrolysis by the released acetate and decreased the yields (paper I Fig. 2). Results from this paper pose a strong argument for using the combined severity factor R'_0, rather than R_0, which does not account for the pH and its inhibition of autohydrolysis.

In parallel to the GGM production experiments, the project explored enzymatic breakdown of mannan by one of the key gut microbes, *Roseburia intestinalis.* An in-depth analysis of its mannan utilization operon (Fig. 10), described in (La Rosa, 2018) revealed two novel acetylesterases. Both esterases have shown activity on mannans, and no activity on xylan. *Ri*CE2 was not active on xylan despite belonging to an acetylxylan esterase CE2 family. *Ri*CEX is the first characterized esterase of its kind, with a unique, exclusive activity on the

axially oriented 2-*O*-acetylation present in mannans. *Ri*CE2 is able to remove acetylations from carbons 3-*O*-, 4-*O*- and 6-*O*-, however it requires *Ri*CEX to remove the 2-*O*-acetylations from double 2-*O*-, 3-*O*- and 2-*O*-, 6-*O*- acetylated mannose units. Structural characterization of *Ri*CEX revealed that the unique activity is the result of a novel two domain structure (paper II, Fig. 1) consisting of an SGNH superfamily hydrolase domain, and a CBM35 family domain, which together form a vise-like grip on manno-oligosaccharides. A tryptophan residue (Trp326) of the CBM35 domain is located above the Ser41 –His193 –Asp190 active site, and orients the substrate towards the active site by aromatic stacking. A sequence homology search identified 357 esterases with the same domain architecture as that of *Ri*CEX, and with catalytic residues and the Trp326 present in a 100% consensus alignment sequence. These are potentially members of a new carbohydrate esterase CAZy family.

The mannan deacetylation apparatus of *Roseburia intestinalis* is the first one of its kind described in this much detail and may serve as a 'template' for degradation of complex carbohydrates by Firmicutes. Similar pairs of esterases consisting of a *Ri*CEX homologue and a CE2 family esterase were found in the metaproteome of colon microbiota samples collected from piglets fed a diet with 4% GGM (paper III, supplementary table 1 &2). Finding homologues of the two esterases expressed in a microbiome sample indicates that the broad specificity CE2 and 2-*O*-acetylation – exclusive CEX homologue pair might provide a paradigm for deacetylation of mannans by bacteria.

In the experiments involving transacetylation of oligosaccharides, we observed a high impact of reaction pH on the yields of acetylated oligosaccharides. We attributed this effect to acetyl migration. With *Ri*CEX being a 2-*O*-acetyl specific esterase, we produced transacetylated oligosaccharides, purified them using preparative scale HILIC, and used NMR to examine the effects of pH and temperature on acetyl migration. Our results show that acetyl migration occurs in response to exposure to pH above neutral (pH 7.4 in the experiment), and by exposure to temperatures as low as 60° C for one hour. These results provide important knowledge for selecting good processing conditions when producing mannans for applications requiring acetylation (thickeners, stabilizers, and prebiotics). Steam explosion and ultrafiltration both expose mannans to temperatures above 60°C, and affect the distribution of acetylations.

A feeding trial showed that Norway spruce GGM has a positive effect on the gut microbiome of weaning pigs (Paper III). A thorough analysis of immune response biomarkers, blood cells composition, gene expression and intestinal epithelium morphology showed no adverse effects on the piglets. The prebiotic effect was located in the lower GIT – cecum and colon. The GGM in pig feed selectively increased the populations of bacterial taxa considered to be health beneficial, such as *Faecalibacterium prausnitzii* and *Prevotella* (paper III, Fig. 4 & 5). The genus *Prevotella* responded with the highest increase in relative abundance, as compared with corresponding samples from control diet piglets. The relative abundance of many operational taxonomic units (OTUs) changed between the mannan inclusion levels in a dose-dependent manner, showing that the presence of GGM was the major driver in the GIT microbiome. Shotgun metagenomics analysis of samples from colons of control and 4% GGM fed piglets has reconstructed 355 MAGs, while metaproteomics of a subset of colon samples identified 8515 proteins. Mapping the protein expression to the MAGs identified multiple MAGs with differential protein expression in response to GGM in feed. MAGs representing *Roseburia intestinalis, Faecalibacterium prausnitzii* and *Prevotella* were among the ones with differential protein expression, with *F. prausnitzii* and *Roseburia* MAGs showing expression of enzymes from their respective mannan degradation operons. A more in-depth analysis is still in progress at the time of writing this thesis, and the dataset generated in the feeding trial holds promise to identify more mannan degraders and possibly, novel mannan degradation systems.

4. Future prospects

Besides the manuscripts contained in this thesis, the work involved in this project has produced a wealth of knowledge and experience for everyone involved, and opened new avenues for future research on biorefining, mannan active enzymes and prebiotics. Production of GGM at the NMBU biorefinery has vastly improved our knowledge of the capacity and limitations of this facility. Mitigation of steam explosion severity in the narrow range between the control samples (pH 3.7) and pH 4.5 generated mannans with a higher DP and DA. This range should be further explored as a means of producing highly complex oligosaccharides.

The success of the feeding trial warrants further research into the impact of GGM on the GIT microbiome. The two mannan degraders, *Roseburia intestinalis, Faecalibacterium prausnitzii,* identified in paper III are implicated in a wide range of GIT diseases in humans. At the functional level, porcine microbiomes highly resemble human microbiomes, with 96% of KEGG-orthology functionalities of the human microbiome present in the pig microbiome (Xiao et al., 2016). Furthermore, mannan degradation is one of the conserved pathways in the human microbiome (Lloyd-Price et al., 2017). This similarity could mean that a prebiotic effect on the same taxa of microbes might be elicited in the human GIT, and serve to improve gut health of humans.

5. References

AJOVALASIT, A., SABATINO, M. A., TODARO, S., ALESSI, S., GIACOMAZZA, D., PICONE, P., DI CARLO, M. & DISPENZA, C. 2018. Xyloglucan-based hydrogel films for wound dressing: Structure-property relationships. *Carbohydrate Polymers,* 179**,** 262-272.

ALPERT, A. J. 1990. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A,* 499**,** 177-196.

ARAKI, T., HASHIKAWA, S. & MORISHITA, T. 2000. Cloning, sequencing, and expression in Escherichia coli of the new gene encoding beta-1,3-xylanase from a marine bacterium, Vibrio sp. strain XY-214. *Appl Environ Microbiol,* 66**,** 1741-3.

ARNLING BÅÅTH, J., MARTÍNEZ-ABAD, A., BERGLUND, J., LARSBRINK, J., VILAPLANA, F. & OLSSON, L. 2018. Mannanase hydrolysis of spruce galactoglucomannan focusing on the influence of acetylation on enzymatic mannan degradation. *Biotechnology for Biofuels,* 11**,** 114.

ASPEBORG, H., COUTINHO, P. M., WANG, Y., BRUMER, H., 3RD & HENRISSAT, B. 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol,* 12**,** 186.

BACKHED, F., DING, H., WANG, T., HOOPER, L. V., KOH, G. Y., NAGY, A., SEMENKOVICH, C. F. & GORDON, J. I. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A,* 101**,** 15718-23.

BAGER, F., MADSEN, M., CHRISTENSEN, J. & AARESTRUP, F. M. 1997. Avoparcin used as a growth promoter is associated with the occurrence of vancomycin-resistant Enterococcus faecium on Danish poultry and pig farms. *Preventive Veterinary Medicine,* 31**,** 95-112.

BEDFORD, A. & GONG, J. 2018. Implications of butyrate and its derivatives for gut health and animal production. *Animal Nutrition,* 4**,** 151-159.

BI, R., BERGLUND, J., VILAPLANA, F., MCKEE, L. S. & HENRIKSSON, G. 2016. The degree of acetylation affects the microbial degradability of mannans. *Polymer Degradation and Stability,* 133**,** 36-46.

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

BROWN, K., UWIERA, R. R. E., KALMOKOFF, M. L., BROOKS, S. P. J. & INGLIS, G. D. 2017. Antimicrobial growth promoter use in livestock: a requirement to understand their modes of action to develop effective alternatives. *International Journal of Antimicrobial Agents,* 49**,** 12-24.

CAMMAROTA, G., IANIRO, G., CIANCI, R., BIBBO, S., GASBARRINI, A. & CURRO, D. 2015. The involvement of gut microbiota in inflammatory bowel disease pathogenesis: potential for therapy. *Pharmacol Ther,* 149**,** 191-212.

CANTAREL, B. L., LOMBARD, V. & HENRISSAT, B. 2012. Complex carbohydrate utilization by the healthy human microbiome. *PLoS One,* 7**,** e28742.

CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PEÑA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TURNBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature methods,* 7**,** 335-336.

CHAUHAN, P. S. & GUPTA, N. 2017. Insight into microbial mannosidases: a review. *Crit Rev Biotechnol,* 37**,** 190-201.

CHUM, H. L., JOHNSON, D. K., BLACK, S. K. & OVEREND, R. P. 1990. Pretreatment-Catalyst effects and the combined severity parameter. *Applied Biochemistry and Biotechnology,* 24**,** 1.

COSGROVE, D. J. 2005. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology,* 6**,** 850-861.

CRYAN, J. F. & DINAN, T. G. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci,* 13**,** 701-12.

CUSKIN, F., LOWE, E. C., TEMPLE, M. J., ZHU, Y., CAMERON, E. A., 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., MUNŌZ-MUNŌZ, J. L., DAY, A., PEÑA, M. J., MCLEAN, R., SUITS, M. D., BORASTON, A. B., ATHERLY, T., ZIEMER, C. J., WILLIAMS, S. J., DAVIES, G. J., ABBOTT, D. W., MARTENS, E. C. & GILBERT, H. J. 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature,* 517**,** 165.

DE FILIPPO, C., RAMAZZOTTI, M., FONTANA, P. & CAVALIERI, D. 2012. Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Briefings in Bioinformatics,* 13**,** 696-710.

DIAZ HEIJTZ, R., WANG, S., ANUAR, F., QIAN, Y., BJORKHOLM, B., SAMUELSSON, A., HIBBERD, M. L., FORSSBERG, H. & PETTERSSON, S. 2011. Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A,* 108**,** 3047-52.

DILOKPIMOL, A., NAKAI, H., GOTFREDSEN, C. H., BAUMANN, M. J., NAKAI, N., ABOU HACHEM, M. & SVENSSON, B. 2011. Recombinant production and characterisation of two related GH5 endo-beta-1,4-mannanases from Aspergillus nidulans FGSC A4 showing distinctly different transglycosylation capacity. *Biochim Biophys Acta,* 1814**,** 1720-9.

DODD, D. & CANN, I. K. O. 2009. Enzymatic deconstruction of xylan for biofuel production. *Global change biology. Bioenergy,* 1**,** 2-17.

DU, X. Z., LI, J., CHEN, J. & LI, B. 2012. Effect of degree of deacetylation on physicochemical and gelation properties of konjac glucomannan. *Food Research International,* 46**,** 270-278.

DUBOIS, M., GILLES, K. A., HAMILTON, J. K., REBERS, P. A. & SMITH, F. 1956. Colorimetric method for determination of sugars and related substances. *Analytical Chemistry,* 28**,** 350-356.

EL KAOUTARI, A., ARMOUGOM, F., GORDON, J. I., RAOULT, D. & HENRISSAT, B. 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol,* 11**,** 497-504.

ENGLYST, H. N., QUIGLEY, M. E. & HUDSON, G. J. 1994. Determination of dietary fibre as non-starch polysaccharides with gas–liquid chromatographic, high-performance liquid chromatographic or spectrophotometric measurement of constituent sugars. *Analyst,* 119**,** 1497-1509.

FLINT, H. J., BAYER, E. A., RINCON, M. T., LAMED, R. & WHITE, B. A. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol,* 6**,** 121-31.

FLINT, H. J., SCOTT, K. P., DUNCAN, S. H., LOUIS, P. & FORANO, E. 2012. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes,* 3**,** 289-306.

FRESE, S. A., PARKER, K., CALVERT, C. C. & MILLS, D. A. 2015. Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome,* 3**,** 28.

FRY, S. C., YORK, W. S., ALBERSHEIM, P., DARVILL, A., HAYASHI, T., JOSELEAU, J.-P., KATO, Y., LORENCES, E. P., MACLACHLAN, G. A., MCNEIL, M., MORT, A. J., GRANT REID, J. S., SEITZ, H. U., SELVENDRAN, R. R., VORAGEN, A. G. J. & WHITE, A. R. 1993. An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiologia Plantarum,* 89**,** 1-3.

GASKINS, H. R., COLLIER, C. T. & ANDERSON, D. B. 2002. Antibiotics as growth promotants: Mode of action. *Animal Biotechnology,* 13**,** 29-42.

GILBERT, H. J., STALBRAND, H. & BRUMER, H. 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol,* 11**,** 338-48.

GONZALEZ RONQUILLO, M. & ANGELES HERNANDEZ, J. C. 2017. Antibiotic and synthetic growth promoters in animal diets: Review of impact and analytical methods. *Food Control,* 72**,** 255-267.

GRANT, G. A., FRISON, S. L., YEUNG, J., VASANTHAN, T. & SPORNS, P. 2003. Comparison of MALDI-TOF Mass Spectrometric to Enzyme Colorimetric Quantification of Glucose from Enzyme-Hydrolyzed Starch. *Journal of Agricultural and Food Chemistry,* 51**,** 6137-6144.

GRESSE, R., CHAUCHEYRAS-DURAND, F., FLEURY, M. A., VAN DE WIELE, T., FORANO, E. & BLANQUET-DIOT, S. 2017. Gut Microbiota Dysbiosis in Postweaning Piglets: Understanding the Keys to Health. *Trends in Microbiology,* 25**,** 851-873.

HAYASHI, T. 1989. Xyloglucans in the Primary Cell Wall. *Annual Review of Plant Physiology and Plant Molecular Biology,* 40**,** 139-168.

HAYASHI, T. & KAIDA, R. 2011. Functions of Xyloglucan in Plant Cells. *Molecular Plant,* 4**,** 17-24.

HENRISSAT, B. & DAVIES, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol,* 7**,** 637-44.

HOLMAN, D. B., BRUNELLE, B. W., TRACHSEL, J. & ALLEN, H. K. 2017. Meta-analysis To Define a Core Microbiota in the Swine Gut. *mSystems,* 2.

HOLMAN, D. B. & CHÉNIER, M. R. 2014. Temporal changes and the effect of subtherapeutic concentrations of antibiotics in the gut microbiota of swine. *FEMS Microbiology Ecology,* 90**,** 599-608.

HOLT, J. P., VAN HEUGTEN, E., GRAVES, A. K., SEE, M. T. & MORROW, W. E. M. 2011. Growth performance and antibiotic tolerance patterns of nursery and finishing pigs fed growth-promoting levels of antibiotics. *Livestock Science,* 136**,** 184-191.

JÖNSSON, L. J. & MARTÍN, C. 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technology,* 199**,** 103-112.

JUTURU, V. & WU, J. C. 2012. Microbial xylanases: Engineering, production and industrial applications. *Biotechnology Advances,* 30**,** 1219-1227.

KABEL, M. A., BOS, G., ZEEVALKING, J., VORAGEN, A. G. J. & SCHOLS, H. A. 2007. Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw. *Bioresource Technology,* 98**,** 2034-2042.

KARAS, M., BACHMANN, D. & HILLENKAMP, F. 1985. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry,* 57**,** 2935-2939.

KARPIEVITCH, Y. V., POLPITIYA, A. D., ANDERSON, G. A., SMITH, R. D. & DABNEY, A. R. 2010. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *The annals of applied statistics,* 4**,** 1797-1823.

KUCZYNSKI, J., LAUBER, C. L., WALTERS, W. A., PARFREY, L. W., CLEMENTE, J. C., GEVERS, D. & KNIGHT, R. 2011. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet,* 13**,** 47-58.

KULKARNI, A. D., JOSHI, A. A., PATIL, C. L., AMALE, P. D., PATEL, H. M., SURANA, S. J., BELGAMWAR, V. S., CHAUDHARI, K. S. & PARDESHI, C. V. 2017. Xyloglucan: A functional biomacromolecule for drug delivery applications. *International Journal of Biological Macromolecules,* 104**,** 799-812.

LA ROSA, S. L., MICHALAK, L., LETH, ML., EJBY MH., PUDLO NA., GLOWACKI R., WORKMAN C., POPE PB., ARNTZEN, MØ., MARTENS, E., HACHEM, MA., WESTERENG, B. 2018. The Human Gut Firmicute Roseburia intestinalis is a Primary Degrader of Dietary β-Mannans. *Manuscript*.

LALLES, J. P., BOSI, P., SMIDT, H. & STOKES, C. R. 2007. Nutritional management of gut health in pigs around weaning. *Proc Nutr Soc,* 66**,** 260-8.

LETH, M. L., EJBY, M., WORKMAN, C., EWALD, D. A., PEDERSEN, S. S., STERNBERG, C., BAHL, M. I., LICHT, T. R., AACHMANN, F. L., WESTERENG, B. & ABOU HACHEM, M. 2018a. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nature Microbiodilogy,* 3**,** 570-580.

LI, M., WANG, B., ZHANG, M., RANTALAINEN, M., WANG, S., ZHOU, H., ZHANG, Y., SHEN, J., PANG, X., ZHANG, M., WEI, H., CHEN, Y., LU, H., ZUO, J., SU, M., QIU, Y., JIA, W., XIAO, C., SMITH, L. M., YANG, S., HOLMES, E., TANG, H., ZHAO, G., NICHOLSON, J. K., LI, L. & ZHAO, L. 2008. Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 2117-2122.

LLOYD-PRICE, J., MAHURKAR, A., RAHNAVARD, G., CRABTREE, J., ORVIS, J., HALL, A. B., BRADY, A., CREASY, H. H., MCCRACKEN, C., GIGLIO, M. G., MCDONALD, D., FRANZOSA, E. A., KNIGHT, R., WHITE, O. & HUTTENHOWER, C. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature,* 550**,** 61-66.

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

LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2014b. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res,* 42**,** D490-5.

LOOFT, T., JOHNSON, T. A., ALLEN, H. K., BAYLES, D. O., ALT, D. P., STEDTFELD, R. D., SUL, W. J., STEDTFELD, T. M., CHAI, B., COLE, J. R., HASHSHAM, S. A., TIEDJE, J. M. & STANTON, T. B. 2012a. In-feed antibiotic effects on the swine intestinal microbiome. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 1691-1696.

LOOFT, T., JOHNSON, T. A., ALLEN, H. K., BAYLES, D. O., ALT, D. P., STEDTFELD, R. D., SUL, W. J., STEDTFELD, T. M., CHAI, B. L., COLE, J. R., HASHSHAM, S. A., TIEDJE, J. M. & STANTON, T. B. 2012b. In-feed antibiotic effects on the swine intestinal microbiome. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 1691-1696.

LUNDQVIST, J., JACOBS, A., PALM, M., ZACCHI, G., DAHLMAN, O. & STALBRAND, H. 2003. Characterization of galactoglucomannan extracted from spruce (Picea abies) by heat-fractionation at different conditions. *Carbohydrate Polymers,* 51**,** 203-211.

LUNDQVIST, J., TELEMAN, A., JUNEL, L., ZACCHI, G., DAHLMAN, O., TJERNELD, F. & STÅLBRAND, H. 2002. Isolation and characterization of galactoglucomannan from spruce (Picea abies). *Carbohydrate Polymers,* 48**,** 29-39.

MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., EECKHAUT, V., BALLET, V., CLAES, K., VAN IMMERSEEL, F., VERBEKE, K., FERRANTE, M., VERHAEGEN, J., RUTGEERTS, P. & VERMEIRE, S. 2014. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63**,** 1275-83.

MAGNÚSDÓTTIR, S., RAVCHEEV, D., DE CRÉCY-LAGARD, V. & THIELE, I. 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in Genetics,* 6.

MALGAS, S., VAN DYK, J. S. & PLETSCHKE, B. I. 2015. A review of the enzymatic hydrolysis of mannans and synergistic interactions between beta-mannanase, beta-mannosidase and alpha-galactosidase. *World Journal of Microbiology & Biotechnology,* 31**,** 1167-1175.

MCCARTER, J. D. & WITHERS, S. G. 1994. Mechanisms of Enzymatic Glycoside Hydrolysis. *Current Opinion in Structural Biology,* 4**,** 885-892.

MCDONALD, A. G., BOYCE, S. & TIPTON, K. F. 2009. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Research,* 37**,** D593-D597.

MIKKONEN, K. S., TENKANEN, M., COOKE, P., XU, C., RITA, H., WILLFÖR, S., HOLMBOM, B., HICKS, K. B. & YADAV, M. P. 2009. Mannans as stabilizers of oil-in-water beverage emulsions. *LWT - Food Science and Technology,* 42**,** 849-855.

MILLER, G. L. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry,* 31**,** 426-428.

MONTANIER, C., MONEY, V. A., PIRES, V. M., FLINT, J. E., PINHEIRO, B. A., GOYAL, A., PRATES, J. A., IZUMI, A., STALBRAND, H., MORLAND, C., CARTMELL, A., KOLENOVA, K., TOPAKAS, E., DODSON, E. J., BOLAM, D. N., DAVIES, G. J., FONTES, C. M. & GILBERT, H. J. 2009. The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol,* 7**,** e71.

MOREIRA, L. R. S. & FILHO, E. X. F. 2008. An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology,* 79**,** 165-178.

MUDGIL, D., BARAK, S. & KHATKAR, B. S. 2014. Guar gum: processing, properties and food applications-A Review. *Journal of Food Science and Technology-Mysore,* 51**,** 409-418.

NAKAMURA, A. M., NASCIMENTO, A. S. & POLIKARPOV, I. 2017. Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation,* 1**,** 35-51.

NURK, S., MELESHKO, D., KOROBEYNIKOV, A. & PEVZNER, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res,* 27**,** 824-834.

OLSSON, L., JØRGENSEN, H., KROGH, K. & ROCA, C. 2004. *Bioethanol Production from Lignocellulosic Material*.

PEDERSEN, M. & MEYER, A. S. 2010. Lignocellulose pretreatment severity – relating pH to biomatrix opening. *New Biotechnology,* 27**,** 739-750.

R. P. OVEREND, E. C. 1987. Fractionation of lignocellulosics by steam-aqueous pretreatments. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences,* 321**,** 523-536.

RISSANEN, J. V., GRENMAN, H., XU, C. L., KROGELL, J., WILLFOR, S., MURZIN, D. Y. & SALMI, T. 2015. CHALLENGES IN UNDERSTANDING THE SIMULTANEOUS AQUEOUS EXTRACTION AND HYDROLYSIS OF SPRUCE HEMICELLULOSES. *Cellulose Chemistry and Technology,* 49**,** 449-453.

RODRIGUEZ, R. L., GUNTURU, S., HARVEY, W. T., ROSSELLO-MORA, R., TIEDJE, J. M., COLE, J. R. & KONSTANTINIDIS, K. T. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res,* 46**,** W282-w288.

ROGOWSKI, A., BRIGGS, J. A., MORTIMER, J. C., TRYFONA, T., TERRAPON, N., LOWE, E. C., BASLÉ, 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. & BOLAM, D. N. 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications,* 6**,** 7481.

ROOKS, M. G. & GARRETT, W. S. 2016. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol,* 16**,** 341-52.

ROSLUND, M. U., AITIO, O., WARNA, J., MAAHEIMO, H., MURZIN, D. Y. & LEINO, R. 2008. Acyl group migration and cleavage in selectively protected beta-d-galactopyranosides as studied by NMR spectroscopy and kinetic calculations. *J Am Chem Soc,* 130**,** 8769-72.

SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. *In:* MERCHANT, S., BRIGGS, W. R. & ORT, D. (eds.) *Annual Review of Plant Biology, Vol 61.*

SCHÜTT, F., WESTERENG, B., HORN, S. J., PULS, J. & SAAKE, B. 2012. Steam refining as an alternative to steam explosion. *Bioresource Technology,* 111**,** 476-481.

SCOTT, K. P., MARTIN, J. C., CHASSARD, C., CLERGET, M., POTRYKUS, J., CAMPBELL, G., MAYER, C.-D., YOUNG, P., RUCKLIDGE, G., RAMSAY, A. G. & FLINT, H. J. 2011. Substrate-driven gene expression in Roseburia inulinivorans: Importance of inducible enzymes in the utilization of inulin and starch. *Proceedings of the National Academy of Sciences,* 108**,** 4672-4679.

SHARON, G., SAMPSON, T. R., GESCHWIND, D. H. & MAZMANIAN, S. K. 2016. The Central Nervous System and the Gut Microbiome. *Cell,* 167**,** 915-932.

SHIMIZU, M., KANEKO, Y., ISHIHARA, S., MOCHIZUKI, M., SAKAI, K., YAMADA, M., MURATA, S., ITOH, E., YAMAMOTO, T., SUGIMURA, Y., HIRANO, T., TAKAYA, N., KOBAYASHI, T. & KATO, M. 2015. Novel beta-1,4-Mannanase Belonging to a New Glycoside Hydrolase Family in Aspergillus nidulans. *J Biol Chem,* 290**,** 27914-27.

SHIPMAN, J. A., BERLEMAN, J. E. & SALYERS, A. A. 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol,* 182**,** 5365-72.

SIMÕES, J., NUNES, F. M., DOMINGUES, P., COIMBRA, M. A. & DOMINGUES, M. R. 2012. Mass spectrometry characterization of an Aloe vera mannan presenting immunostimulatory activity. *Carbohydrate Polymers,* 90**,** 229-236.

SINGH, R. D., BANERJEE, J. & ARORA, A. 2015. Prebiotic potential of oligosaccharides: A focus on xylan derived oligosaccharides. *Bioactive Carbohydrates and Dietary Fibre,* 5**,** 19-30.

SINGH, S., SINGH, G. & ARYA, S. K. 2018. Mannans: An overview of properties and application in food products. *International Journal of Biological Macromolecules,* 119**,** 79-95.

SRIVASTAVA, P. K. & KAPOOR, M. 2017. Production, properties, and applications of endo-β-mannanases. *Biotechnology Advances,* 35**,** 1-19.

STOKSTAD, E. L., JUKES, T. H. & ET AL. 1949. The multiple nature of the animal protein factor. *J Biol Chem,* 180**,** 647-54.

ŠUCHOVÁ, K., KOZMON, S., PUCHART, V., MALOVÍKOVÁ, A., HOFF, T., MØRKEBERG KROGH, K. B. R. & BIELY, P. 2018. Glucuronoxylan recognition by GH 30 xylanases: A study with enzyme and substrate variants. *Archives of Biochemistry and Biophysics,* 643**,** 42-49.

TAMANAI-SHACOORI, Z., SMIDA, I., BOUSARGHIN, L., LOREAL, O., MEURIC, V., FONG, S. B., BONNAURE-MALLET, M. & JOLIVET-GOUGEON, A. 2017. Roseburia spp.: a marker of health? *Future Microbiology,* 12**,** 157-170.

TAYLOR, E. J., GOYAL, A., GUERREIRO, C. I., PRATES, J. A., MONEY, V. A., FERRY, N., MORLAND, C., PLANAS, A., MACDONALD, J. A., STICK, R. V., GILBERT, H. J., FONTES, C. M. & DAVIES, G. J. 2005. How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides: structure and activity of a Clostridium thermocellum lichenase, CtLic26A. *J Biol Chem,* 280**,** 32761-7.

TENKANEN, M., PULS, J., RATTO, M. & VIIKARI, L. 1993. ENZYMATIC DEACETYLATION OF GALACTOGLUCOMANNANS. *Applied Microbiology and Biotechnology,* 39**,** 159-165.

TERRAPON, N., LOMBARD, V., DRULA, E., LAPEBIE, P., AL-MASAUDI, S., GILBERT, H. J. & HENRISSAT, B. 2018. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res,* 46**,** D677-d683.

TESTER, R. F. & AL-GHAZZEWI, F. H. 2013. Mannans and health, with a special focus on glucomannans. *Food Research International,* 50**,** 384-391.

TIMELL, T. E. 1967. Recent progress in the chemistry of wood hemicelluloses. *Wood Science and Technology,* 1**,** 45-70.

TOPAKAS, E., KYRIAKOPOULOS, S., BIELY, P., HIRSCH, J., VAFIADI, C. & CHRISTAKOPOULOS, P. 2010b. Carbohydrate esterases of family 2 are 6-O-deacetylases. *FEBS Letters,* 584**,** 543-548.

WILLFOR, S., SJOHOLM, R., LAINE, C., ROSLUND, M., HEMMING, J. & HOLMBOM, B. 2003. Characterisation of water-soluble galactoglucomannans from Norway spruce wood and thermomechanical pulp. *Carbohydrate Polymers,* 52**,** 175-187.

WILLFOR, S., SUNDBERG, K., TENKANEN, M. & HOLMBOM, B. 2008. Spruce-derived mannans - A potential raw material for hydrocolloids and novel advanced natural materials. *Carbohydrate Polymers,* 72**,** 197-210.

WILLIAMS, M. A. K., FOSTER, T. J., MARTIN, D. R., NORTON, I. T., YOSHIMURA, M. & NISHINARI, K. 2000. A Molecular Description of the Gelation Mechanism of Konjac Mannan. *Biomacromolecules,* 1**,** 440-450.

WILLING, B. P., DICKSVED, J., HALFVARSON, J., ANDERSSON, A. F., LUCIO, M., ZHENG, Z., JÄRNEROT, G., TYSK, C., JANSSON, J. K. & ENGSTRAND, L. 2010. A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology,* 139**,** 1844-1854.e1.

XIAO, L., ESTELLE, J., KIILERICH, P., RAMAYO-CALDAS, Y., XIA, Z., FENG, Q., LIANG, S., PEDERSEN, A. O., KJELDSEN, N. J., LIU, C., MAGUIN, E., DORE, J., PONS, N., LE CHATELIER, E., PRIFTI, E., LI, J., JIA, H., LIU, X., XU, X., EHRLICH, S. D., MADSEN, L., KRISTIANSEN, K., ROGEL-GAILLARD, C. & WANG, J. 2016. A reference gene catalogue of the pig gut microbiome. *Nat Microbiol***,** 16161.

XU, C., LEPPÄNEN, A.-S., EKLUND, P., HOLMLUND, P., SJÖHOLM, R., SUNDBERG, K. & WILLFÖR, S. 2010. Acetylation and characterization of spruce (Picea abies) galactoglucomannans. *Carbohydrate Research,* 345**,** 810-816.

YANG, F., MITRA, P., ZHANG, L., PRAK, L., VERHERTBRUGGEN, Y., KIM, J. S., SUN, L., ZHENG, K. J., TANG, K. X., AUER, M., SCHELLER, H. V. & LOQUE, D. 2013. Engineering secondary cell wall deposition in plants. *Plant Biotechnology Journal,* 11**,** 325-335.

YANG, H., HUANG, X., FANG, S., XIN, W., HUANG, L. & CHEN, C. 2016. Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Scientific Reports,* 6**,** 27427.

YIN, Y., MAO, X., YANG, J., CHEN, X., MAO, F. & XU, Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res,* 40**,** W445-51.

ZHANG, Y., JU, J., PENG, H., GAO, F., ZHOU, C., ZENG, Y., XUE, Y., LI, Y., HENRISSAT, B., GAO, G. F. & MA, Y. 2008. Biochemical and structural characterization of the intracellular mannanase AaManA of Alicyclobacillus acidocaldarius reveals a novel glycoside hydrolase family belonging to clan GH-A. *J Biol Chem,* 283**,** 31551-8.

ZOETENDAL, E. G., RAJILIĆ-STOJANOVIĆ, M. & DE VOS, W. M. 2008. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut,* 57**,** 1605-1615.

Paper I

Effects of pH on steam explosion extraction of acetylated galactoglucomannan from Norway spruce.

Leszek Michalak[1], Svein Halvor Knutsen[2], Ida Aarum[1], Bjørge Westereng[1*]

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

[2] Nofima, Norwegian Institute of Food, Fishery and Aquaculture Research, PB 210, N-1431 Ås, Norway

* Corresponding author, bjorwe@nmbu.no

Keywords:

Steam explosion, pH control, Norway spruce, mannan, galactoglucomannan, acetylation, hemicellulose, hydrothermal extraction.

Abstract:

Background: Acetylated galactoglucomannan (GGM) is a complex hemicellulose found in softwoods such as Norway spruce (*Picea abies).* GGM has a large potential as a biorefinery feedstock and source of oligosaccharides for high value industrial applications. Steam explosion is an effective method for extraction of carbohydrates from plant biomass. Increasing the reaction pH reduces the combined severity (R'_0) of treatment, affecting yields and properties of extracted oligosaccharides. In this study, steam explosion was used to extract oligosaccharides from Norway spruce wood chips soaked with sodium citrate and potassium phosphate buffers with pH of 4.0-7.0. Yields, monosaccharide composition of released oligosaccharides and biomass residue, their acetate content and composition of their lignin fraction were examined to determine the impact of steam explosion buffering on the extraction of softwood hemicellulose.

Results: Reducing the severity of steam explosion resulted in lower yields, although the extracted oligosaccharides had a higher degree of polymerization. Higher buffering pH also resulted in a higher fraction of xylan in the extracted oligos. Oligosaccharides extracted in buffers of pH >5.0 were deacetylated. Buffering lead to a removal of acetylations from both the extracted oligosaccharides and the hemicellulose in the residual biomass. Treatment of the residual biomass with a GH5 family mannanase from *Aspergillus nidulans* was not able to improve the GGM yields. No hydroxymethylfurfural formation, a decomposition product from hexoses, was observed in samples soaked with buffers at pH higher than 4.0.

Conclusions: Buffering the steam explosion reactions proved to be an effective way to reduce the combined severity (R'_0) and produce a wide range of products from the same feedstock at the same physical conditions. The results highlight the impact of chemical autohydrolysis of hemicellulose by acetic acid released from the biomass in hydrothermal pretreatments. Lower combined severity results in products with a lower degree of acetylation of both the extracted oligosaccharides and residual biomass. Decrease in severity appears not to be the result of reduced acetate release, but rather a result of inhibited autohydrolysis by the released acetate. Based on the results presented, the optimal soaking pH for fine tuning properties of extracted GGM is below 5.0.

Background:

Steam explosion (SE) is an effective and scalable method for solubilizing hemicellulose from plant biomass, applicable to a wide range of biorefinery feedstocks. SE extraction was successfully used as pretreatment for production of biogas from hay (1), sugarcane bagasse (2) and corn stover (3), birchwood (4) as well as the production of ethanol from spruce bark (5) and many other platform chemicals from a wide range of lignocellulose feedstocks (6).

Steam explosion combines hydrothermal treatment of biomass with defibration by a rapid release of pressure at the end of the process. These two processes are independent of each other, and results comparable with SE have been obtained by hydrothermal treatment with a mechanical refining step, as long as the treatment severity was the same (7). In the course of the hydrothermal pretreatment, a major part of the hemicellulose and lignin present in the secondary cell wall lamellae is separated from the adjacent cellulose microfibrils, and becomes water soluble (8). At the same time, some of the acetate naturally linked to the xylan and mannan in the lignocellulose is released and contributes to the autohydrolyis of biomass. Release of acetic acid is the reason for the low pH usually seen in the SE product slurry. Properties of SE treated material depend on a range of factors, the most important being the residence time and temperature in the vessel. Impact of temperature on the material is described by the severity factor $R_0 = e^{(T_{exp} - 100)/14.75}$ (9). A combined severity factor $R'_0 = (10^{-pH}) * (t * e^{(T_{exp} - 100)/14.75})$ (10) was developed to include the contribution of H^+ to the hydrolysis process. This combined severity factor was previously used to predict and compare the severities of treatments where pH, rather than temperature or residence time, was the variable (11, 12). Mitigating the severity of pretreatment by controlling pH is a potential means of fine-tuning the products.

A number of factors besides temperature and residence time also play a role, such as the biomass particle size and the rate of steam and liquid diffusion through the particle, the ratio of solids to liquid loaded into the SE vessel and the chemicals brought in from upstream processing stages. During SE treatment, acetylated hemicellulose releases acetic acid, which decreases the pH and facilitates chemical hydrolysis of polysaccharides. Acetate mediated autohydrolysis depends on the diffusion of liquid through the biomass particles (13). Diffusion rate depends on the particle size and the surface to volume ratio. The final pH of the product slurry after hydrothermal pretreatment depends on the composition of the liquid fraction, its amount and buffering capacity. The intricacies of hemicellulose breakdown in hydrothermal pretreatment, and difficulties in the analysis of the process are brilliantly explained by Rissanen et al. (13).

For inclusion in microbial fermentation, conditions are usually selected with the aim of highest possible breakdown of biomass, while keeping the formation of chemicals inhibitory to enzymatic hydrolysis or fermentation to a minimum (14, 15). In literature pertaining SE and pre-treatments fermentability and end-product yields are often selected as the main evaluation criteria, favoring high severity conditions often using acids or sulphates as additives (6, 16). These high severity conditions yield oligosaccharides with low degree of polymerization (DP), low degree of acetylation (DA) and fewer branchings, which require fewer enzymes for hydrolysis to monosaccharides. For GGM, this means a partial or complete deacetylation and removal of galactose sidechains. In contemporary biorefining focused on production of higher value chemicals such as food and feed ingredients, nutraceuticals (17) or hydrocolloids (18), controlled extraction conditions yielding high molecular mass and high complexity can be a more attractive pretreatment option. With the right enzyme toolbox, further tailoring and breakdown into constituent monomers is easy to achieve, while synthesis of highly branched and decorated polysaccharides in large amounts is almost impossible. Obtaining more complex hemicelluloses is of interest for several reasons: more complex products with novel physicochemical properties open doors to new applications, higher complexity may improve selectivity in microbial degradation (19), and increase the biodiversity of gut microbiomes when used as prebiotics. More complex oligosaccharides that more closely resemble *in vivo* hemicellulose would also make attractive substrates for studying activity of carbohydrate active enzymes.

GGM is the main hemicellulose in Norway spruce (*Picea abies*). It is a complex hemicellulose consisting of a backbone of β-(1→4)-D-Man*p* and β-(1→4)-D-Glc*p* residues with α-(1→6)-D-Gal*p* branches, prevalently attached to the Man*p*, and to a lesser extent on Glc*p* (19). An estimated 30% of the D-Man*p* residues of spruce GGM are *O*-acetylated at carbons 2, 3 and 6, as well as 4 in the non-reducing ends of oligosaccharides (19). Acetylation of spruce mannan is a particularly important feature, since it affects the accessibility of mannans to microbes, and the physicochemical properties of mannans in solution. At the same time, release of acetylations from hemicellulose and hydrolysis of polysaccharides by the released acetate is a crucial process for the solubilization of hemicellulose (8).

In this study, SE extraction was carried out with pH control resulting in a mitigation of treatment severity. Six experimental conditions at five pH levels as well as a control sample using water only were used for SE to yield significantly different oligosaccharides in the extract. The relationship between the combined severity factor and the product composition was evaluated by assessing the yields, apparent DP, oligosaccharide acetylation, monosaccharide composition of products and biomass residue, MALDI-ToF MS analysis of extracted oligosaccharides, NMR analysis

of lignin released, and analysis of susceptibility of biomass residue to treatment with a GH5 mannanase.

Results and discussion:

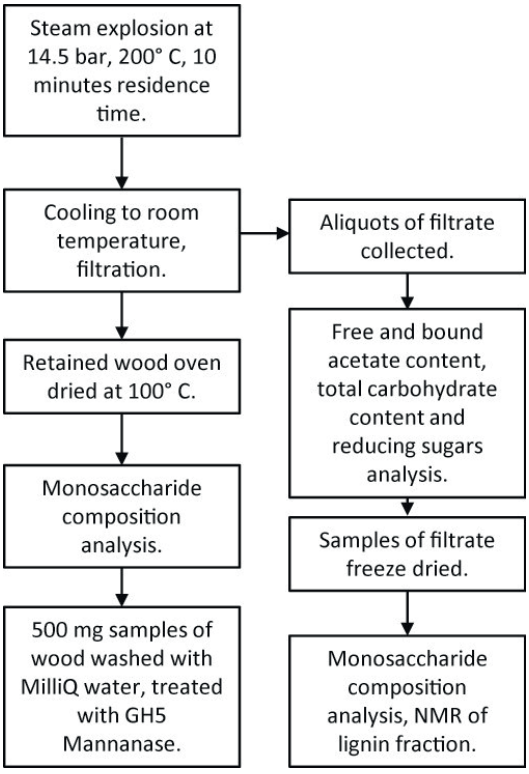


Figure 1. Flowchart of sample treatment and analyses carried out. The steam exploded wood chips were transferred from the collection vessel to plastic buckets and allowed to cool, pH measurements were taken once the slurry reached room temperature. Water was then added to aid extraction. Samples were mixed and transferred to funnels laid with Whatman B1 filters. Aliquots of this filtrate were used for quantification of acetate content, total carbohydrate content and reducing sugars. Samples of the filtrate were freeze dried and used for monosaccharide composition analysis, and analysis of lignin by NMR. SE wood retained by the filteres was dried at 100°C for 36-48 hours, until steady weight was reached. Samples of dried extracted wood were used for monosaccharide composition and enzymatic hydrolysis.

A detailed description of sample handling and analysis pipeline is illustrated in the flowchart above (Fig. 1). Citrate and phosphate based buffers were selected due to their respective buffer ranges and temperature stability. In all samples except the citrate pH 4.0, the pH has dropped after SE due to release of acetate from the wood (table 1). Higher buffer concentrations would be necessary to keep the post-SE pH exactly as the soaking buffers, however this would cause more interference with downstream analysis. The range of buffers resulted in combined

severities ranging from 0.004 to 0.519 in the buffer controlled samples. Non-buffered controls had the highest R'_0 at 1.68-1.75. The wide range of calculated R'_0 is entirely attributable to the buffered conditions, since other conditions in the reaction were the same. The large difference in R'_0 between the buffered samples illustrates the room for adjustment and possibility for fine-tuning granted by pH controlled extractions.

Table 1. Sample treatments, slurry pH after steam explosion and the combined severity factors calculated as in (11) which determine severities based on the pH after the treatment. In all samples buffering the SE reaction has resulted in final pH higher (pH 4.22- 6.32) than that of the control samples (average pH 3.70). 0.5M citrate and 1M phosphate at pH 6.0 resulted in different final pH, highlighting the difference in the buffering capacity between citrate and phosphate.

Buffer:	*Average pH:*	*St. dev.*	*Average Combined Severity R'_0:*	*St. dev.*	*Man:Glc:Gal ratio:*	*Bound acetate µmole/mg carbohydrate.*
MilliQ H_2O Control	3.703	0.009	1.707	0.037	1.88: 1: 0.28	0.303
0.5M Citrate pH 4.0	4.227	0.009	0.511	0.011	2.71: 1: 0.39	0.306
0.5M Citrate pH 5.0	4.973	0.041	0.092	0.009	1.64: 1: 0.51	0.112
0.5M Citrate pH 6.0	5.553	0.012	0.024	0.001	0.67: 1: 0.53	n.d.
1M Phosphate pH 6.0	5.317	0.045	0.042	0.004	1.39: 1: 0.23	n.d.
1M Phosphate pH 6.5	5.917	0.017	0.010	0.000	0.19: 1: 0.11	n.d.
1M Phosphate pH 7.0	6.303	0.031	0.004	0.000	0.23: 1: 0.16	n.d.
						-

Yields and composition of extracted hemicellulose:

Higher severity treatment yielded higher amounts of solubilized carbohydrates, with the highest yield of 17.4 % average based on dry wood weight for the non-buffered samples (Fig. 2A). Yields dropped for the buffered samples, with only the citrate pH 4.0 among the buffered samples (average R'_0 = 0.511) being close to the non-buffered sample (13.1 % average yields). The total yield of soluble carbohydrates dropped rapidly with the decreasing R'_0 although the yields remained over 4 % (4.4 % for the potassium phosphate pH 7.0 buffered samples, R'_0=0.0045). Yields from the three least severe treatments (sodium citrate pH 6.0, average R'_0 = 0.0241; and potassium phosphate pH 6.5 and 7.0, R'_0 = 0.0104 and R'_0 = 0.0045, respectively) shift very slightly (6.1 % for citrate pH 6.0, 5.2 % and 4.4 % for phosphate pH 6.5 and 7.0, respectively) despite a considerable

drop in the R'_0. This decrease in efficiency with increasing pH was attributed to reaction pH being higher than the pK_a of acetic acid (4.76). Under these conditions, the reactivity of acetic acid and its contribution to autohydrolysis of hemicellulose are markedly decreased. Characteristics of products from these low severity treatments illustrate a baseline for extraction in a SE reaction with a minor contribution of autohydrolysis. The extracts approximate the products of an extraction with steam and temperature only.

In order to assess yields as well as the degree of hemicellulose breakdown occurring during extraction, total carbohydrate content of each sample was determined using the phenol-sulphuric acid method of Dubois (20). Concentrations of reducing sugars were estimated by Miller's dinitrosalycilic acid assay (21). For comparison of severity effects on the estimated length of oligosaccharides in the soluble fraction, the ratio of total carbohydrates to reducing sugars was used as an approximation for the DP of the solubilized oligosaccharides (Fig. 2A). Comparison of yields and DP of extracted oligosaccharides shows the increase of average DP (from 2.52 at R'_0 = 1.707 to 5.22 at R'_0 = 0.004), accompanied by a reduction in yields (decrease from 17.4 % of dry wood weight at R'_0 = 1.707 to 4.4 % at R'_0 = 0.004). An overview of oligosaccharide length and sample composition is presented in MALDI-ToF MS spectra (Fig. S2). Multiple oligosaccharides with *m/z* over 1000 are present in all samples, despite the comparison of total to reducing sugars indicating the average DP range to be between 2.52 (control) to 6.99 (potassium phosphate pH 6.5). This apparent discrepancy is due to the fact that MALDI-ToF was not able to detect monosaccharides and cleary visualize the oligosaccharides <750 *m/z* due to high background from the salts and other contaminants in the samples.

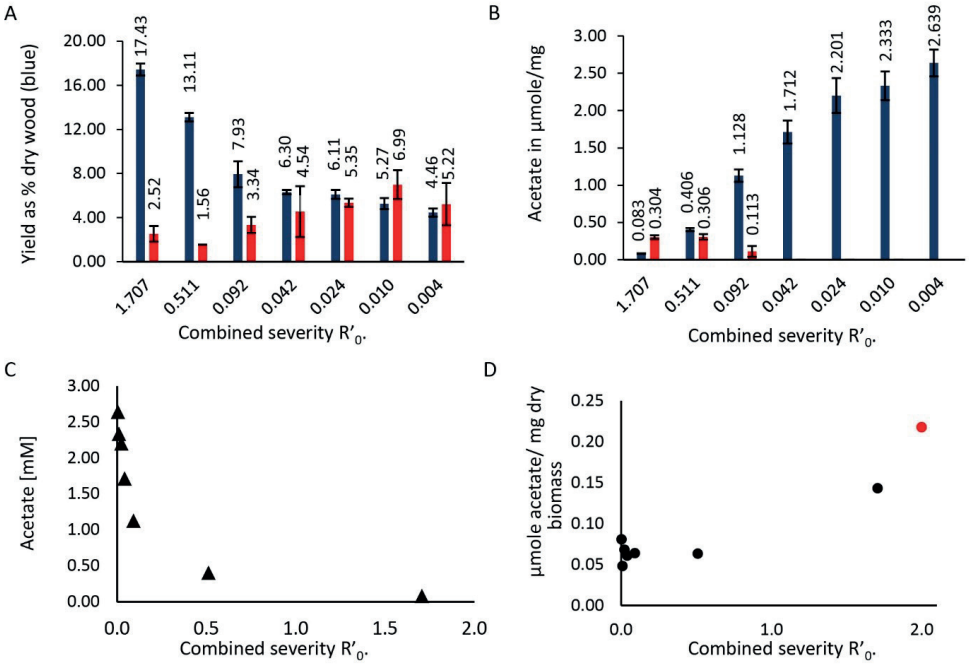


Figure 2. A: Percentage yields of total carbohydrates (blue bars) from dry wood mass, and the average DP of extracted oligosaccharides (red bars), error bars show standard deviation between technical replicates. B: Bar chart of acetate in filtrate (blue bars), and acetate released from the oligosaccharides in solution (red bars) after KOH treatment. Error bars indicate the standard deviation between technical replicates. C: Scatterplot of acetate content of filtered samples at the various severities. D: Scatterplot of acetate content of dried biomass residue, dry wood raw material in red.

Composition of extracted hemicellulose:

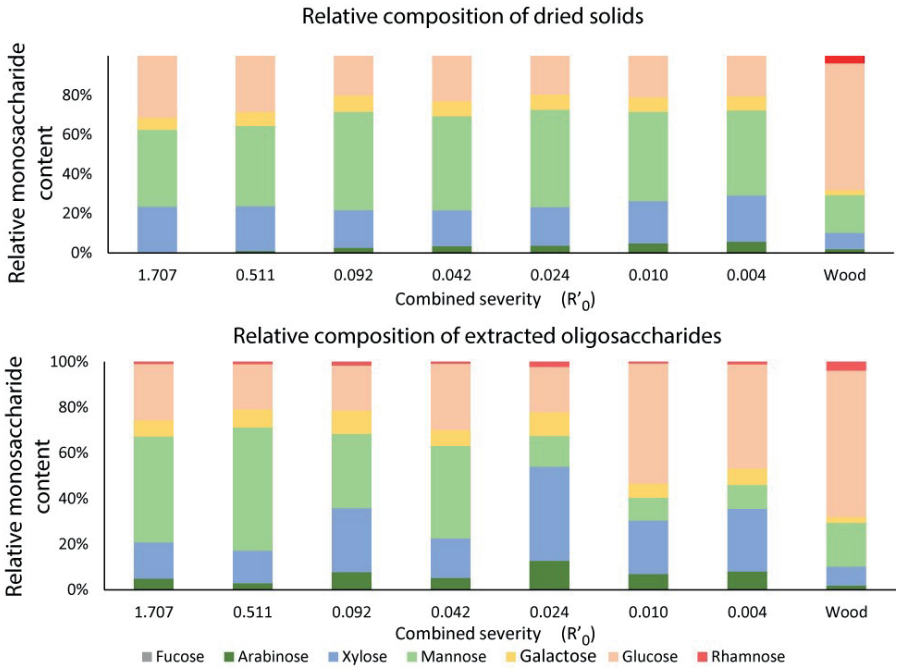
Beside the yields and apparent DP, buffering the SE reaction had an impact on the composition of extracted oligosaccharides. A comparison of the monosaccharide composition of all samples and the extracted wood is summarized in Fig. 3, and supplementary tables S2 and S3. At lower severity, more xylo-oligosaccharides were released, with only the citrate buffered and control samples yielding GGM as the predominant hemicellulose. No rhamnose was detected in the dried solids biomass residue after any treatment. Loss of arabinose in the high severity samples can be attributed to hydrolysis observed previously in low pH extractions (22). In samples buffered with pH 6.0, 6.5 and 7.0 phosphate, the relative content of mannose in the solubilized carbohydrate fraction was several times lower than that of xylose (Fig. 3, supplementary table s2). Galactose content and the apparent Gal:Man ratio have increased with decreasing severity, although we were unable to ascertain if the galactose was bound to GGM oligosaccharides, or was present as

monosaccharides resulting from debranching of GGM in the cell wall. The decrease in efficiency of extraction over the wide range of severities is apparent in Fig. 3, the content of hemicellulose left in dried biomass residue increases with the decrease in combined severity.

The gradual shift from extraction of GGM towards xylan and glucuronoxylan is illustrated in Fig. 4. In order to clear the MALDI-ToF spectrum and avoid ambiguity of *m/z* assignment (such as in the case of peak 1097 m/z, which appears in the hexose and pentose series), aliquots of the extracts were deacetylated by adding 100mM NaOH. The control sample and sodium citrate pH 4.0 samples spectra contain predominantly hexose peaks, while xylooligosaccharide peaks are dominant in the spectra of citrate pH 5.0 and 6.0 samples.

Extraction buffered with sodium citrate at pH 4.0 produced the highest relative content of GGM in the filtrate (Fig. 3) and with the highest degree of acetylation of extracts (Fig 2B and 5). The apparent increase in relative mannan content in soluble fraction of citrate pH 4.0 buffered samples ($R'_0 = 0.511$) comes at a reduction of yield from 17.3% to 13.1% compared to the control sample (Fig. 2A). The corresponding dried biomass residue samples have a very similar monosaccharide distribution: 37.56% mannose and 6.01 % galactose for the control sample residue, 39.51% mannose and 6.92 % galactose for citrate pH 4.0 (table s2). SE with citrate pH 4.0 buffering appears more selective towards mannan, while the unbuffered control had a higher overall efficiency.

The Man:Glc:Gal ratio (table 1) is an indication of complexity of yielded mannooligosaccharides. In high severity hydrothermal extraction, the α-(1→6)-D-Gal*p* branchings of GGM are cleaved off (23). For the Norwegian Spruce (*Picea abies*) the Man:Glc:Gal ratios reported in literature is 4:1:0.1 3.8:1:0.4 (19, 24). The ratio varies based on the wood and extraction methods. When the GGM constituent ratios are considered, buffering with citrate at pH 5.0 has yielded the best results, nearly doubling the galactose content of the extracted oligos from control samples (table 1). The ratios were 1.88: 1: 0.28 Man:Glc:Gal in the control samples and 1.64:1:0.51 in the citrate pH 5.0. At the same time, citrate at pH 5.0 increased the apparent DP of the oligosaccharides from 2.52 to 3.34 (Fig. 2A). The improvement in Man:Glc:Gal ratio was accompanied with a pronounced decrease in yield (7.93% for citrate pH 5.0), and the mannose content of the extract (32.55% for citrate pH 5.0 vs 46.41% for control). Citrate pH 5.0 extracts contained 28.09% xylose, nearly twice as much as the control (15.94% for xylose) (table s2, Fig. 3), and had nearly three times lower degree of acetylation (Fig. 2B).



233

234 *Figure 3. Top: Relative monosaccharide composition of carbohydrates in the dried, washed solids. Bottom: Relative*
235 *monosaccharide composition of carbohydrates in aqueous extracts of steam exploded wood. The composition of*
236 *untreated spruce chips raw material (wood) is provided for comparison.*

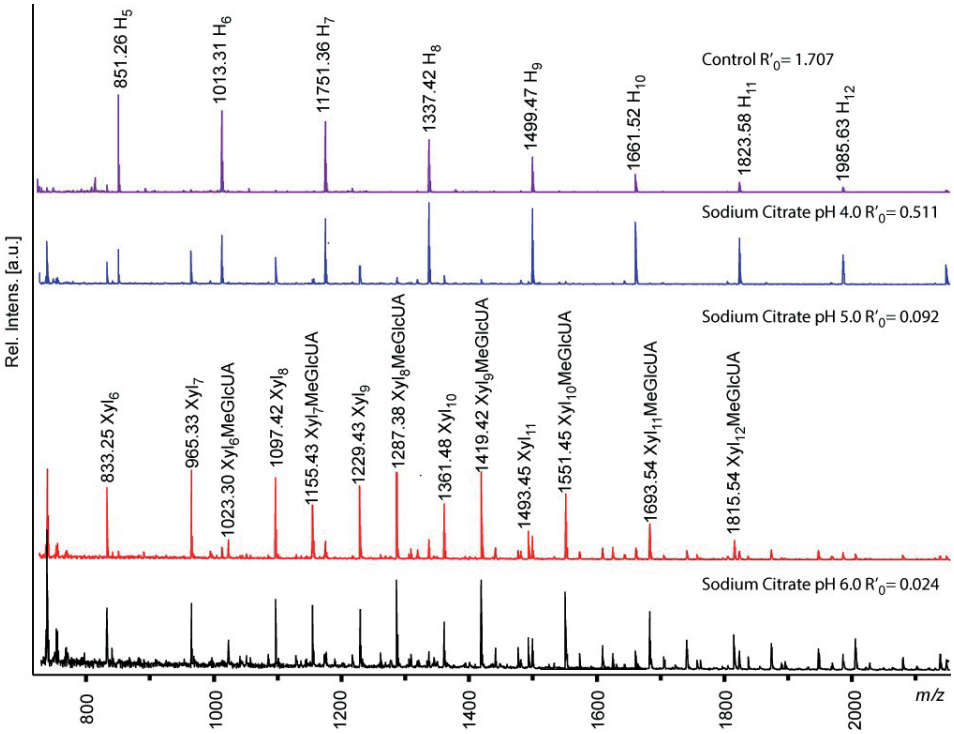


Figure 4. MALDI-ToF MS Spectra of extract samples deacetylated with NaOH. In the control and sodium citrate pH 4.0 samples, GGM peaks are the main components, with small xylooligosaccharide peaks alongside GGM in the sodium citrate pH 4.0. In sodium citrate pH 5.0 and 6.0 the dominant peaks are the xylooligosaccharides and methylglucuronic acids.Xyl – xylose, H-hexose, MeGlcUA – methylglucuronic acid, Ac-acetylation.

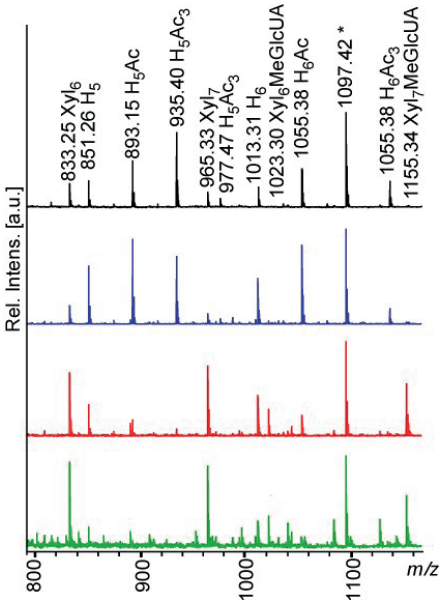


Figure 5. MALDI-TOF MS spectra of extracted oligosaccharides from the buffer control (black), sodium citrate pH 4.0 (blue), sodium citrate pH 5.0 (red), sodium citrate pH 6.0 (green). Peak labelled 1097.42 is either a double acetylated mannohexose or non-acetylated octapentose. Xyl – xylose, H-hexose, MeGlcUA – methylglucuronic acid, Ac-acetylation.*

Acetate content of soluble fractions:

Acetate content in the filtrate decreased quickly with increasing severity (Fig. 2C). The same trend was apparent in analysis of acetate content in biomass residue. Biomass from buffered samples contained between 0.064-0.040 µmole of acetate per mg of biomass (Fig. 2D), while the control samples contained 0.142 µmole of acetate per mg. Dried biomass from control samples retained 65.8% of the acetate measured in wood raw material (0.142 µmole vs 0.218 µmole of acetate per mg). Since it is difficult to estimate the factual DP of oligosaccharide products, acetylation values were calculated as µmole of acetate per mg of solubilized carbohydrates.

In the severity range between the control samples and the samples buffered with sodium citrate pH 6.0, hemicellulose peaks seen in MALDI-ToF MS gradually became deacetylated (Fig. 5). The relative intensities of peaks corresponding to acetylated mannooligosaccharides indicate that the highest content of acetylated mannooligos was extracted in the control sample. Acetylated mannooligos are the majority of peaks in the control and sodium citrate pH 4.0 samples, and disappear in sodium citrate pH 6.0 samples.

Aliquots of the aqueous extracts were treated with KOH to deacetylate the oligosaccharides in solution. KOH treatment removed the acetylations on oligos in solution, and allowed for comparison between the free acetate and bound acetate. Only the Citrate pH 4.0, 5.0, and the control samples contained appreciable amounts of acetate bound to carbohydrates (Fig. 2B, table 1). Despite the fact that high pH and low severity conditions yield more acetate per mg of released hemicellulose, the acetate was present free in solution. Whether this occurred as a result of pH in the SE vessel or occurred during storage of the sample (since buffer solution is still present) is unclear. From previous, unpublished experimental results at the same scale, as well as pilot scale where over 700 kg of spruce was processed, we know that storage at the control sample pH (3.6 - 4.0) did not cause a deacetylation even at ambient temperatures, for two to four weeks.

Acetylation of extracted oligosaccharides is a characteristic crucial for their physicochemical properties. The DA affects water solubility, susceptibility to enzymatic hydrolysis and availability as a carbon source for microbes. Release of acetate during hydrothermal pretreatment is one of the mechanisms of cell wall breakdown, and a decrease in severity would be expected to correlate with a decrease in the acetate released and in the amounts of acetate bound to oligosaccharides. This was however, not the case as more acetate was released with higher buffer pH. This may be due to de-esterification which is accelerated at higher pHs (25).

Enzymatic treatment of solid residue.

Enzymatic hydrolysis was tested as a means to assist the release of hemicellulose from wood treated with SE in conditions of inhibited autohydrolysis. Samples of dried residual biomass were treated with a GH5 family endomannanase from *Aspergillus nidulans* (26) to find out if severity of SE had an effect on the availability of hemicellulose in the steam exploded wood to hydrolytic enzymes. Even at low combined severity, the hemicellulose matrix is exposed to extreme conditions and undergoes defibration in the pressure release. These conditions were hypothesized to open the secondary cell wall matrix and render the hemicellulose more accessible to mannanases. GH5 family mannanases have been shown to be more efficient on less acetylated substrates (27), and since a large part of the acetate was removed in the steam explosion, it was hypothesized that a hydrolytic enzyme could to a larger extent access the residual mannan and thus improve the yields of manno-oligosaccharides. However, mannanase treatment of dried residual biomass from SE did not release appreciable amounts of manno-oligosaccharides, indicating that mannan in the biomass residue remains largely inaccessible to hydrolytic enzymes, regardless of the material being acetylated (high severity) or non-acetylated (low severity). While

there was an apparent effect of the mannanase when the relative content of carbohydrates in enzyme treatment solution was analyzed (Fig. 5), the only observable effect was a slight increase in the combined galactose and mannose fraction of the released oligosaccharides as compared to the control sample incubated at the same conditions in buffer without the enzyme. Enzymatic treatment with this enzyme was not a viable means of improving the yields of low severity SE.

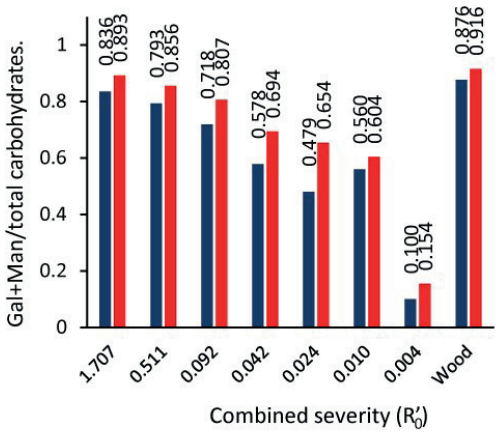


Figure 6. Content of galactose and mannose as a fraction of total carbohydrates extracted with GH5 mannanase treatment. Red bars represent the Gal+Man fraction in mannanase treated samples; blue bars represent control samples with no enzyme.

NMR analysis of lignin content in the solubilized fraction:

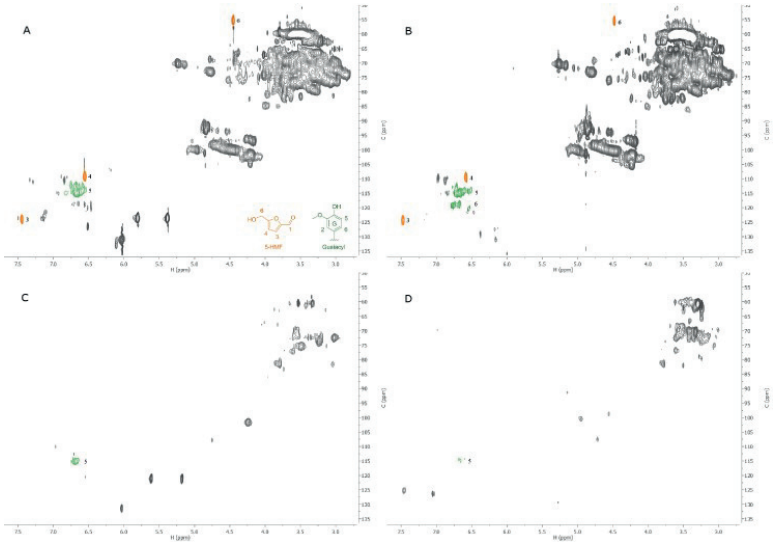


Figure 7. HSQC 2D NMR Spectra of lignin content in biomass residues: (A) sodium citrate pH 4.0 buffered sample, (B) no buffer control, (C) sodium citrate pH 6.0 and (D) potassium phosphate pH 7.0. 5-hydroxymethylfurfural (5-HMF) and Guaicyl are depicted in the lower right of panel A, signals are colored and numbered according to the structures they relate to.

The HSQC 2D-NMR experiments taken of the solvable fraction of the samples shows the proton-carbon corresponding peaks. The spectra mainly contain carbohydrate signals, however there are detectable amounts of aromatic signals in all the samples (Fig. 6). The C_5/H_5-signal for guaiacyl unit (G5) at 114.9/6.7 ppm has the highest intensity in the sample citrate buffer pH 4 (Fig. 6A) and in the control (Fig. 6B), and only the control sample shows the C_6/H_6-signal for guaiacyl unit (G6) at 118.6/6.7. Control sample (Fig. 6A) had a pH of 3.7 after steam explosion. Both the control (Fig. 8A) and citrate pH 4.0 (Fig. 6B) samples were steam exploded at a lower pH (3-4) and the degradation of lignocellulose is more intense for both in comparison to the higher pH steam exploded samples, citrate pH 6 (Fig. 6C) and phosphate pH 7 (Fig. 6D). During SE the lignin undergoes hydrolysis and degrades into smaller units of lignin (28). These units should be detectable in the solvable fraction if they are small enough. With a lower pH, as in sample A and B the hydrolysis is more extensive and lignin was detected in the solvable fraction (Fig. 6). In addition to the signals from degraded lignin, there were some signals from dehydrated carbohydrates in the form of 5-hydroxymethylfurfural (5-HMF, (29)), these were again only visible in citrate buffered sample pH 4 (A) and in the control (B) (table 2). The pH is therefore important for control of both lignin and carbohydrate degradation.

Table 2. Determination of the $^{13}C/^{1}H$ correlation signals acquired in 2D-NMR HSQC of the samples and semi quantitative analysis of lignin. Based on the summarized integrated areas of 5-HMF and guaiacyl relative to co-extracted mannose, signals is calculated per 100 mannose C1/H1 signal (%).

Label	δ_C/δ_H (ppm)	Assignment	A (Citrate pH 4)	B (control)
G5	114.9/6.7	C_5/H_5 in a Guaiacyl unit	66 %	22 %
G6	118.6/6.7	C_6/H_6 in a Guaiacyl unit	—	12 %
F3	124.1/7.5	C_3/H_3 in a 5-HMF unit	28 %	8 %
F4	109.4/6.6	C_4/H_4 in a 5-HMF unit	41 %	9 %
F6	55.4/4.5	C_6/H_6 in a 5-HMF unit	-42 %	-9 %

The signals for 5-HMF (F) and guaiacyl (G) unit were integrated in NMR with the C_1/H_1 signal of mannose as an internal reference signal (30). In the citrate buffer pH 4 sample the G5-signals was 66% (calculated per 100 mannose C_1/H_1, table 2) in comparison to control which had only 22%. This means that the relative amount of lignin is higher in the citrate buffer than in the control, even though the final pH in control sample was lower, as is expected based on existing research (31). As the initial pH in the control was not 3.7 before SE, the degree of hydrolysis seems to be more severe with continuously low pH. The same effect of more severe degradation is also detected with the carbohydrate fraction, as there is more 5-HMF, a common decomposition product of hexoses (32), in citric buffer (A) than control (B) (table 2). Besides the effect on properties of extracted oligosaccharides, inhibition of polysaccharide autohydrolysis in samples soaked with buffers >5.0 prevented the formation of HMF.

Optimal pH range for the production of acetylated galactoglucomannan.

From the wide range of combined severities tested in this study, between R'_0=1.707 and R'_0=0.092 (controls, citrate pH 4.0 and 5.0 buffered samples) appears to be the best range for production of acetylated GGM. Extracts within this range contained acetylated oligosaccharides with varying DP, DA and Man:Glc:Gal ratios. At the same time, only the control and citrate pH 4.0 samples contained detectable levels of HMF. In the range between unbuffered and pH 5.0, buffering can mitigate the deacetylation, autohydrolysis and formation of HMF, at the cost of yield. As seen in the comparison between the control sample and citrate pH 4.0 the apparent loss in yield is partly due to increased specificity towards mannan extraction. Some general trends are apparent in the data presented here: increased combined severity results in higher yields and higher degree of acetylation of extracted oligosaccharides, while at the same time reducing the degree of

polymerization. Further experiments into steam explosion production of tailored oligosaccharides from Norway spruce should be focused on this severity range.

Conclusions:

Introducing buffers to a steam explosion reaction has shown to be an efficient approach for mitigating the severity of the treatment, and production of a wide range of oligosaccharides from the same feedstock at the same temperature and pressure. Vast differences in monosaccharide composition, oligosaccharide size and degree of acetylation of the solubilized carbohydrate fraction were caused by the difference in pH. Notably, higher pH resulted in more pronounced deacetylation of residual biomass and extracted oligosaccharides.

Altering the pH did not reduce the severity by preventing the acetate release from the biomass, but by limiting acid hydrolysis of hemicellulose. Buffering mitigates the reactivity of acetate once it is released. The results show that the role of temperature and pressure is mainly to create conditions where autohydrolysis can occur. When the autohydrolysis was inhibited by buffering, the yields dropped and the breakdown of oligosaccharides was reduced. This study clearly shows that pH largely affects product composition and yields. It has been argued that pH has more impact on SE (12) reactions than temperature or pressure, and the results presented here support this claim.

Materials and Methods:

Buffers:

1M sodium citrate and 2M potassium phosphate buffers were prepared by mixing 1M solutions of sodium citrate (Sigma-Aldrich, Germany) and citric acid (Sigma-Aldrich, Germany), and 2M solutions of di- and mono-basic potassium phosphate (Sigma-Aldrich, Germany) were mixed to reach the desired pH. Citrate buffers produced were pH 4.0, 5.0, and 6.0, phosphate pH was 6.0, 6.5 and 7.0. The higher concentration of phosphate buffers was used to counteract the poor pH retention after SE in the phosphate buffered samples observed in initial trial experiments (unpublished).

Wood:

Dry Norway spruce (*Picea abies*) wood was milled using a hammer mill with a 2 mm sieve. 500 gram samples of spruce chips were soaked with buffers and MilliQ water in a 1: 1: 1 (g: mL: mL) ratio prior to SE. Water was added to ensure the buffers were thoroughly mixed into the wood, resulting in final buffer concentrations of 0.5 M for sodium citrate and 1 M for potassium phosphate. The wood chips were stirred until the sample was thoroughly soaked and transferred into the SE reactor.

Steam Explosion and extraction of water-soluble material:

Soaked spruce chips were hydrothermally treated in a steam explosion unit (Cambi, Asker, Norway) consisting of a 20 L pressure vessel and a flash tank with collection bucket. Steam was generated in a 25 kW electric boiler (Parat, Flekkefjord, Norway). The steam explosion unit is described in detail in (33). Treatment conditions were 200° C, 14.5 bar, biomass residence time was 10 minutes.

Handling of extracts and residuals:

After SE, water was added, the slurry was stirred for extraction and filtered through a whatman B1 filter paper (Sigma Aldritch, Norway). The residual water-insoluble material was

squeezed to release the remaining soluble oligosaccharides, which were combined with the extract. Aliquots were frozen to determine extract yield and to supply samples for carbohydrate, lignin and acetyl analysis. 200 mL of each sample was freeze dried for the analysis of constituent neutral monosaccharides (GC) and uronic acid (colorimetry) of the released oligosaccharides. Insoluble materials were dried in an oven at 100°C for 36-48 hours, to constant weight, then milled on a cutter mill (Retsch, Haan, Germany) with a 0.5 mm sieve.

Poly- and oligosaccharide constituent sugars, carbohydrate content and reducing sugar in extract and non-soluble residuals:

Concentration of carbohydrates in solution were quantified according to the Dubois method (20), and reducing sugars content according to the Miller method (21). Calibration curves for both colorimetric methods were based on glucose. Constituent monosaccharide of residuals and extracts were quantified by GC via alditol acetates after acid hydrolysis (34) and uronic acids in the hydrolysates were determined by a colorimetric assay (35).

MALDI-ToF analysis:

MALDI-ToF analysis of hydrolysis product was conducted on an UltraFlextreme MALDI-ToF instrument (Bruker Daltonics GmbH, Germany) equipped with a nitrogen 337 nm laser beam. Samples were prepared by applying 2 µL of a 9mg/mL solution of 2,5-dihydroxybenzoic acid (Sigma-Aldrich, Germany) in 30% acetonitrile (VWR) to an MTP 384 ground steel target plate (Bruker Daltonics GmbH, Germany), adding 1 µL of sample (0.1-1 mg/mL) and mixing the drop with the pipette. Sample drops were then dried under a stream of warm air.

Acetate content analysis:

For the analysis of free acetate content in solution, the filtered liquid fraction washed from the biomass was diluted 1:2 with MilliQ water (to measure acetate in solution) or 100 mM KOH (to release the acetate bound to the oligosaccharides). 50 µL samples of the liquid phase were collected and analyzed by HPLC. All values were corrected for the concentration of oligosaccharides in solution and exact weight of biomass in the sample.

For the analysis of acetate content in the biomass residue, 100 mg ±10 % samples of the dried, milled residue were soaked overnight with 500 µL of 0.1 M KOH, left in a thermomixer (Eppendorf, Oslo, Norway) overnight at 1000 rpm, 40 °C. After 18 hours, 500 µL of MilliQ water was added to the samples, which were then mixed by vortexing and spun down at x10000 g, for 5 minutes, and analyzed by HPLC. All values were corrected for the concentration of oligosaccharides in solution and exact weight of biomass in the sample.

HPLC:

Acetate content was analysed by HPLC using a REZEX ROA-Organic Acid H+ (Phenomenex, Torrance, California, USA) 300x7.8mm ion exclusion column, isocratic elution with 0.6 mL/min 4mM H_2SO_4 at 65 °C and UV detection at 210 nm.

Enzymatic treatment:

Milled, dry samples were washed with water to remove remaining soluble carbohydrates and buffer salts from the SE slurry, dried and resuspended in 25 mL of 50 mM sodium acetate buffer at pH 5.5. A GH5 family mannanase from *Aspergillus nidulans* (26) was applied to the sample with loadings of 0.01 mg : 1mg (1%), 0.1 mg : 1mg (10%) and 0.3mg : 1mg (30%) of enzyme : mannan in samples, based on an estimate of 20 % of the substrate being mannan. Samples were left in a shaking incubator overnight at 50° C, which is the optimum temperature for enzymatic activity.

NMR of lignin fraction:

The NMR spectra were recorded on a Bruker Ascend 400 spectrometer (400 MHz) at 320 K using a 5 mm PABBO probe. The samples (45 mg) were dissolved in DMSO-d_6 (1 mL), sonicated for 30 min and filtered through glass wool to remove any undissolved particles directly into the NMR-tube. Two of the samples did not fully dissolve. The Heteronuclear Single Quantum Coherence (HSQC) spectroscopy recorded with a spectral width of 0 – 12 ppm and 0 – 250 ppm in ^{1}H and ^{13}C, respectively. The number of scans for both were 512 at 27°C. For the ^{1}H-^{13}C parameters the relaxation time was 1.5 s and the free induction decay dimensions was 2048 and 256, while the number of scans were 120 at 27°C. Integrations were done with MestReNova (version 9.1.0), where the C1 of mannose were used as an internal reference.

Acknowledgements

Norwegian Research council grant no. 244259, 208674/F50, 270038 and 226247 supported this work.

We thank Hanne Zobel for her assistance in carbohydrate analysis and CAMBI for the support in design of steam explosion units.

Supplementary

A sample of Norwegian spruce extracted by milled wood lignin (36) (MWL) was run as a reference standard for HSQC NMR.

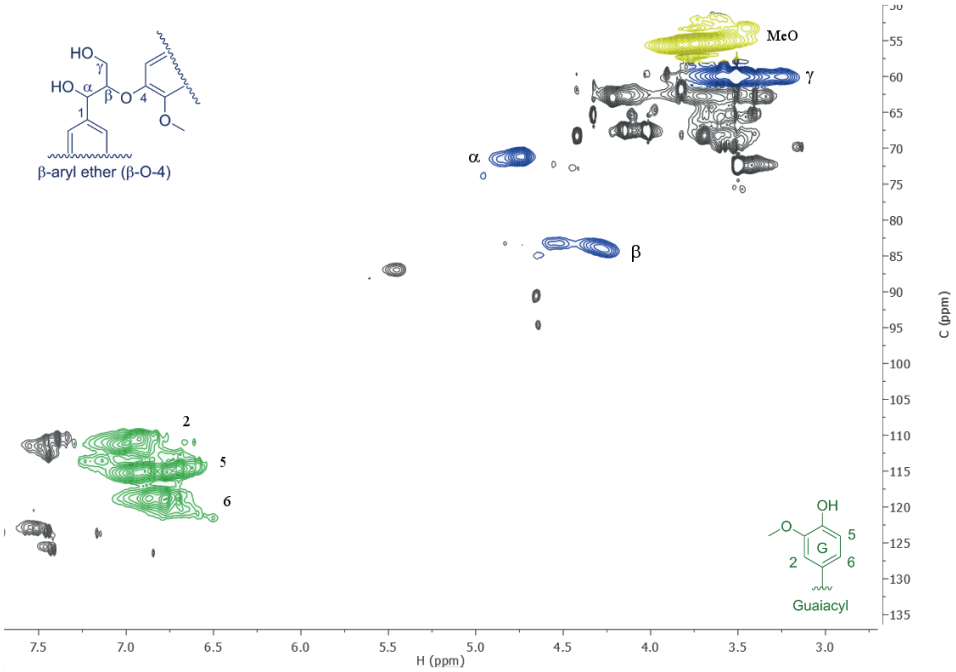


Figure S1. 2D-NMR HSQC of Norwegian spruce lignin (Milled Wood Lignin extracted (MWL)), focused on ^{1}H: 2.7 – 7.7 and ^{13}C: 50.0 – 135.4 ppm.

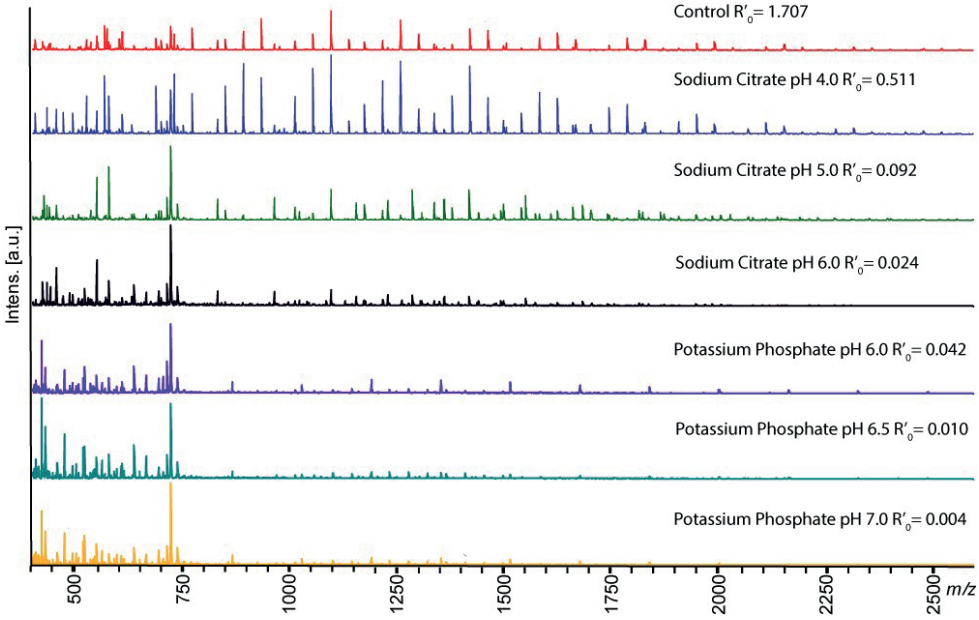


Figure S2. MALDI-ToF spectra of extracted oligosaccharides samples from all treatment. Relative intensities show the most prevalent oligosaccharide sizes to be in the 1000-1500m/z range (DP6-DP9 for hexoses) and highly acetylated in the control and citrate pH 4.0 samples. In further treatments the hexose peaks are gradually replaced with xylooligosaccharide peaks at much higher intensities and with no acetylations. The peak at 723 m/z is a persistent contamination.

Table S1. Determination of the $^{13}C/^{1}H$ correlation signals acquired in the 2D-NMR HSQC spectrum of MWL spruce.ß-O-4 reflect the ß-aryl ether as sketched in Fig. S1, G2, -5 and -6-reflects the aromatic signals and MeO reflects the methoxyl-group in the guaiacyl units sketched in Fig. S1.

Label	δ_C/δ_H (ppm)	Assignment
MeO	55.7/3.7	C/H is methoxyl- in C_3 position on a guaiacyl unit
ß–O–4α	71.9/4.8	C_α/H_α from β–O–4 lignin bonding pattern
ß–O–4ß	86.1/4.1	C_β/H_β from β–O–4 lignin bonding pattern
ß–O–4γ	59.7/3.5	C_γ/H_γ from β–O–4 lignin bonding pattern
G2	111.0/7.0	C_2/H_2 in a Guaiacyl unit
G5	114.9/6.8	C_5/H_5 in a Guaiacyl unit
G6	118.9/6.8	C_6/H_6 in a Guaiacyl unit

527 *Table S2. Monosaccharide composition of carbohydrates in freeze dried aliquots of filtered, water soluble fractions.*

Sample:	*Combined severity $R'_{0:}$*	*Rha*	*Fuc*	*Ara*	*Xyl*	*Man*	*Gal*	*Glc*
Control	1.707	0.90	n.d.	4.95	15.94	46.41	7.10	24.68
Citrate pH 4.0	0.511	0.95	n.d.	2.97	14.30	54.02	7.87	19.89
Citrate pH 5.0	0.092	1.60	n.d.	7.82	28.09	32.55	10.17	19.78
Phosphate 6.0	0.042	0.75	n.d.	5.29	17.27	40.59	6.95	29.14
Citrate pH 6.0	0.024	2.22	n.d.	12.72	41.35	13.40	10.52	19.79
Phosphate pH 6.5	0.010	0.63	n.d.	6.99	23.38	10.10	6.11	52.79
Phosphate 7.0	0.004	1.08	n.d.	8.06	27.49	10.47	7.16	45.75

528

529

530 *Table S3. Monosaccharide composition of dried residual biomass.*

Sample:	*Combined severity $R'_{0:}$*	*Rha*	*Fuc*	*Ara*	*Xyl*	*Man*	*Gal*	*Glc*
Control	1.707	n.d.	n.d.	n.d.	22.62	37.56	6.01	30.20
Citrate pH 4.0	0.511	n.d.	n.d.	0.94	22.09	39.51	6.92	27.60
Citrate pH 5.0	0.092	n.d.	n.d.	2.49	18.51	48.33	8.14	19.37
Phosphate pH 6.0	0.042	n.d.	n.d.	3.28	17.63	46.15	7.42	22.16
Citrate pH 6.0	0.024	n.d.	n.d.	3.58	18.91	47.62	7.57	18.84
Phosphate pH 6.5	0.010	n.d.	n.d.	4.70	20.56	43.42	7.07	20.11
Phosphate pH 7.0	0.004	n.d.	n.d.	5.39	22.36	41.22	6.90	19.48
Wood	N/A	3.90	0.00	1.95	8.72	19.89	2.68	66.77

531

532

533

534

535

536

Table S4 Acetate present in filtered slurry and released from solubilized carbohydrates in KOH treatment. Data from Figure 2B,C and D.

Combined severity R'_0:	*Acetate in solution in µmole/mg carbohydrate.*	*Alkali released acetate in µmole/mg carbohydrate.*	*Acetate in biomass residue in µmole/mg biomass.*
1.707	0.083	0.303	0.143
0.511	0.406	0.306	0.063
0.092	1.128	0.112	0.064
0.042	1.712	n.d.	0.061
0.024	2.201	n.d.	0.068
0.010	2.333	n.d.	0.048
0.004	2.639	n.d.	0.081
Wood	-	-	0.218

Declarations:

Ethics approval and consent to participate: Not applicable

Competing interests: None of the authors have any competing interests in the content of this manuscript.

Consent for publication: All authors consent to publication.

Availability of data and material: All raw data and material samples available upon reasonable request.

Authors' contributions: Experimental design (L.M., S.H.K. and B.W.) and conducting the experiments (L.M. S.H.K. and I.A.). All authors contributed in writing the manuscript.

References:

1. Bauer A, Lizasoain J, Theuretzbacher F, Agger JW, Rincon M, Menardo S, et al. Steam explosion pretreatment for enhancing biogas production of late harvested hay. Bioresource Technology. 2014;166:403-10.
2. Vivekanand V, Olsen EF, Eijsink VGH, Horn SJ. Methane Potential and Enzymatic Saccharification of Steam-exploded Bagasse. Bioresources. 2014;9(1):1311-24.
3. Lizasoain J, Trulea A, Gittinger J, Kral I, Piringer G, Schedl A, et al. Corn stover for biogas production: Effect of steam explosion pretreatment on the gas yields and on the biodegradation kinetics of the primary structural compounds. Bioresource Technology. 2017;244:949-56.
4. Biely P, Cziszarova M, Uhliarikova I, Agger JW, Li XL, Eijsink VGH, et al. Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. Biochimica et Biophysica Acta General Subjects. 2013;1830(11):5075-86.
5. Kemppainen K, Inkinen J, Uusitalo J, Nakari-Setala T, Siika-aho M. Hot water extraction and steam explosion as pretreatments for ethanol production from spruce bark. Bioresource Technology. 2012;117:131-9.
6. Sun Y, Cheng JY. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresource Technology. 2002;83(1):1-11.
7. Schutt F, Westereng B, Horn SJ, Puls J, Saake B. Steam refining as an alternative to steam explosion. Bioresource Technology. 2012;111:476-81.
8. Ramos LP. The chemistry involved in the steam treatment of lignocellulosic materials. Quimica Nova. 2003;26(6):863-71.
9. R. P. Overend EC. Fractionation of lignocellulosics by steam-aqueous pretreatments. Philosophical Transactions of the Royal Society of London Series A, Mathematical and Physical Sciences. 1987;321(1561):523-36.
10. Chum HL, Johnson DK, Black SK, Overend RP. Pretreatment-Catalyst effects and the combined severity parameter. Applied Biochemistry and Biotechnology. 1990;24(1):1.
11. Kabel MA, Bos G, Zeevalking J, Voragen AGJ, Schols HA. Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw. Bioresource Technology. 2007;98(10):2034-42.
12. Pedersen M, Meyer AS. Lignocellulose pretreatment severity - relating pH to biomatrix opening. New Biotechnology. 2010;27(6):739-50.
13. Rissanen JV, Grenman H, Xu CL, Krogell J, Willfor S, Murzin DY, et al. Challenges in understanding the simultaneous aqueous extraction and hydrolysis of spruce hemicelluloses. Cellulose Chemistry and Technology. 2015;49(5-6):449-53.

14. Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? In: Olsson L, editor. Biofuels. Advances in Biochemical Engineering-Biotechnology. 1082007. p. 67-93.

15. Jönsson LJ, Martín C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresource Technology. 2016;199:103-12.

16. Singh J, Suhag M, Dhaka A. Augmented digestion of lignocellulose by steam explosion, acid and alkaline pretreatment methods: A review. Carbohydrate Polymers. 2015;117:624-31.

17. Tester RF, Al-Ghazzewi FH. Mannans and health, with a special focus on glucomannans. Food Research International. 2013;50(1):384-91.

18. Willför S, Sundberg K, Tenkanen M, Holmbom B. Spruce-derived mannans – A potential raw material for hydrocolloids and novel advanced natural materials. Carbohydrate Polymers. 2008;72(2):197-210.

19. Lundqvist J, Teleman A, Junel L, Zacchi G, Dahlman O, Tjerneld F, et al. Isolation and characterization of galactoglucomannan from spruce (Picea abies). Carbohydrate Polymers. 2002;48(1):29-39.

20. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for determination of sugars and related substances. Analytical Chemistry. 1956;28:350-6.

21. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Analytical Chemistry. 1959;31(3):426-8.

22. Westereng B, Michaelsen TE, Samuelsen AB, Knutsen SH. Effects of extraction conditions on the chemical structure and biological activity of white cabbage pectin. Carbohydrate Polymers. 2008;72(1):32-42.

23. Lundqvist J, Jacobs A, Palm M, Zacchi G, Dahlman O, Stålbrand H. Characterization of galactoglucomannan extracted from spruce (Picea abies) by heat-fractionation at different conditions. Carbohydrate Polymers. 2003;51(2):203-11.

24. Willför S, Sjöholm R, Laine C, Roslund M, Hemming J, Holmbom B. Characterisation of water-soluble galactoglucomannans from Norway spruce wood and thermomechanical pulp. Carbohydrate Polymers. 2003;52(2):175-87.

25. Kravtchenko TP, Arnould I, Voragen AGJ, Pilnik W. Improvement of the Selective Depolymerization of Pectic Substances by Chemical Beta-Elimination in Aqueous-Solution. Carbohydrate Polymers. 1992;19(4):237-42.

26. Dilokpimol A, Nakai H, Gotfredsen CH, Baumann MJ, Nakai N, Abou Hachem M, et al. Recombinant production and characterisation of two related GH5 endo-beta-1,4-mannanases from Aspergillus nidulans FGSC A4 showing distinctly different transglycosylation capacity. Biochimica Et Biophysica Acta-Proteins and Proteomics. 2011;1814(12):1720-9.

27. Arnling Bååth J, Martínez-Abad A, Berglund J, Larsbrink J, Vilaplana F, Olsson L. Mannanase hydrolysis of spruce galactoglucomannan focusing on the influence of acetylation on enzymatic mannan degradation. Biotechnology for Biofuels. 2018;11(1):114.

28. Heikkinen H, Elder T, Maaheimo H, Rovio S, Rahikainen J, Kruus K, et al. Impact of Steam Explosion on the Wheat Straw Lignin Structure Studied by Solution-State Nuclear Magnetic Resonance and Density Functional Methods. J Agric Food Chem. 2014;62(43):10437-44.

29. Li J, Henriksson G, Gellerstedt G. Carbohydrate reactions during high-temperature steam treatment of aspen wood. Appl Biochem Biotechnol. 2005;125(3):175.

30. Sette M, Wechselberger R, Crestini C. Elucidation of lignin structure by quantitative 2D NMR. Chemistry (Weinheim an der Bergstrasse, Germany). 2011;17(34):9529-35.

31. Li Y, Liu W, Hou Q, Han S, Wang Y, Zhou D. Release of Acetic Acid and Its Effect on the Dissolution of Carbohydrates in the Autohydrolysis Pretreatment of Poplar Prior to Chemi-Thermomechanical Pulping. Industrial & Engineering Chemistry Research. 2014;53(20):8366-71.

32. García-Aparicio MP, Ballesteros I, González A, Oliva JM, Ballesteros M, Negro MJ. Effect of inhibitors released during steam-explosion pretreatment of barley straw on enzymatic hydrolysis. Applied Biochemistry and Biotechnology. 2006;129(1):278-88.

33. Horn SJ, Nguyen QD, Westereng B, Nilsen PJ, Eijsink VGH. Screening of steam explosion conditions for glucose production from non-impregnated wheat straw. Biomass Bioenerg. 2011;35(12):4879-86.

34. Vestby LK, Møretrø T, Ballance S, Langsrud S, Nesse LL. Survival potential of wild type cellulose deficient Salmonella from the feed industry. BMC Veterinary Research. 2009;5(1):43.

35. Englyst HN, Quigley ME, Hudson GJ. Determination of dietary fibre as non-starch polysaccharides with gas-liquid chromatographic, high-performance liquid chromatographic or spectrophotometric measurement of constituent sugars. The Analyst. 1994;119(7):1497-509.

36. Crestini C, Melone F, Sette M, Saladino R. Milled Wood Lignin: A Linear Oligomer. Biomacromolecules. 2011;12(11):3928-35.

Paper II

A Pair of esterases from a commensal gut bacterium completely deacetylate highly complex mannans.

Authors

Leszek Michalak[1], Sabina Leanti La Rosa[1], Shaun Leivers[1], Åsmund Kjendseth Røhr[1], Finn Lillelund Aachmann[2], Bjørge Westereng[1*]

[1] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås

[2] NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Sem Sælands vei 6-8, N-7491 Trondheim, Norway

* Corresponding author, Bjørge Westereng, bjorwe@nmbu.no

Abstract

Mannans and xylans are decorated with acetylations that protect them from degradation by glycoside hydrolysases. Mannans are widely present in human and animal diets, as fiber from leguminous plants and as thickeners and stabilizers in processed foods. Furthermore, mannans are highly abundant in a range of biomasses and are attractive compounds for biorefining. Deacetylation of mannans is central in the breakdown of dietary fibers by commensal gut bacteria and in biorefining. Mannan degrading pathways include deacetylation and are among the conserved functions of the human microbiome. There are many fully characterized acetylxylan esterases (AcXEs), however, the enzymes deacetylating mannans are less understood. Acetylations on xylan are present in the equatorial plane of the sugar ring, whereas in most mannans, the prevalent acetylation is on the axially oriented 2C-hydroxyl. This arrangement makes the acetylation inaccessible for most carbohydrate esterases. Here we present two carbohydrate esterases, *Ri*CE2 and *Ri*CEX, from the Firmicutes *Roseburia intestinalis,* which together deacetylate complex galactoglucomannan (GGM). The 3D-structure of *Ri*CEX with a mannopentaose in the active site

shows that the accessory domain of *Ri*CEX forms a confined complex in where the axial C2-OH group on mannose is directed towards the Ser41 in the catalytic triad. Cavities on the *Ri*CEX surface may accept galactosylations at the C6 positions of both residues adjacent to the mannose residue being deacetylated. In depth characterization of the two enzymes using time resolved NMR, and LC-MS shows that they work in a complementary manner, with the *Ri*CEX exclusively removing 2-*O*-acetylations on any sugar residue in an oligosaccharide, including double acetylated mannoses, while the broad specific *Ri*CE2 is active on 3-*O*-, 4-*O*- and 6-*O*- acetylation. Activity of *Ri*CE2 is dependent on *Ri*CEX removing 2-*O*- acetylations from double acetylated mannose.

Significance statement

Acetylations protect hemicellulose from enzymatic degradation, and thus constitute a protective adaptation for plants as well as plant biomass recalcitrance in biorefineries. Removal of acetylations is thus a key step towards utilization of as a carbon source for human and animal gut microbiomes and in the biorefinery. We present two highly substrate specific acetyl (galacto)glucomannan esterases (AGMEs) from a gut commensal Firmicutes which cooperatively deacetylate complex galactoglucomannan. We present detailed insight into mannan deacetylation based on the crystal structure of an AGME with mannopentaose bound in the active site. *Ri*CEX has a unique two-domain architecture including a CBM35 binding domain. This discovery increases our understanding of gut microbes utilization and biorefining of complex carbohydrates.

Introduction

Roseburia intestinalis

Gut microbes are highly adapted to fermenting complex carbohydrates. Along with Bacteroidetes, Firmicutes are the main polysaccharide degrading commensal bacteria in the human gut (1, 2). Some clades of Firmicutes, such as Clostridiales XIV are dietary fiber degraders producing short chain fatty acids (SCFAs) in the lower gastrointestinal tract of their hosts and are considered beneficial to the gut health (3-5). Bacteroidetes are generalists with a large number of glycoside hydrolases (6), they are able to adapt to a wide range of polysaccharides and break them down using periplasmic enzymes. On the other hand, Firmicutes are considered specialists, highly adapted towards a more limited range of carbohydrates, and break down glycans in the cytoplasm (1, 7). β-mannans constitute a small but significant part of human diets and mannan degradation is identified as one of the core pathways in the human gut microbiome (8). The gram-positive Firmicutes *Roseburia intestinalis*, commonly found in human and animal gastrointestinal tracts, has a sophisticated degradation apparatus for complete β-mannan degradation (9).

Acetylated mannans

β-Mannans are among the most important hydrocolloids used in food processing. Carob, guar and konjac mannans are commonly used as food stabilizers that are fermented by gut commensal bacteria in the lower gastrointestinal tract. Humans are also exposed to mannans in leguminous plants, *Aloe vera* in cosmetics, nutraceuticals and coffee (10). Mannans are polysaccharides present in the secondary cell walls of plants. The basic structure of their chains are β-1,4 linked D-Man*p* units, which are the only constituent of mannan. In glucomannans, such as konjac (*Amorphophallus konjac*) glucomannan, mannose residues are interspersed with β-1,4 linked D-Glc*p* in varying ratios. Galactomannans, found in the carob nut (*Ceratonia siliqua*), consist of a β-1,4-D-Man*p* backbone decorated with α-1,6-D-Gal*p* branchings (11). Mannans may be *O*-acetylated at the hydroxyl groups attached to the carbons 2-*O*-, 3-*O*-, and 6-*O*- acetylated on any sugar residue in the chain, as well as 4-*O*- acetylated in the non-reducing end of the chain.

In softwoods, such as the Norway spruce (*Picea abies*), galactoglucomannan (GGM) is the main hemicellulose in the secondary cell wall, accounting for ~20% of the dry wood mass (12, 13). GGM is a polysaccharide consisting of a backbone of β-1,4- D-Man*p* and β-1,4- D-Glc*p* residues, in varying ratios depending on the source (13, 14). This glucomannan backbone is decorated with α-1,6- D-Gal*p*

residues, prevalently attached to the Man*p*, and at a low level on Glc*p* (15). Previous reports on the composition of mannan in spruce describe two types of mannan with different monosaccharide compositions. One with a high galactose content (Gal:Glc:Man ratio of 1:1:3) which constitutes about 5-8% of dry wood weight, and a more prevalent, low galactose type (Gal:Glc:Man ratio of 0.1:1:3), which constitutes 10-15% dry weight (12). The composition of GGM in biomass extracts varies depending on the severity of extraction and purification methods. In spruce GGM, about 30% of the D-Man*p* residues are 2-*O*, 3-*O* and 6-*O* esterified by acetylations (13).

Aloe vera mannan (AVM) is another type of GGM commonly found in cosmetics and nutraceuticals. AVM contains fewer galactosylations albeit, a much higher degree of acetylation than spruce, and minor amounts of arabinose substitutions (16). The AVM degree of acetylation is higher than in spruce containing many double acetylated mannose units (16).

The role of acetylations of the mannan backbone in plants is thought to be a defense mechanism developed as a part of an evolutionary arms race against pathogens, and to act as structural support by enabling hemicellulose interactions with lignin (17, 18). A unique characteristic of the *O*-acetyl groups in Man*p* units of the GGM chain is their relative orientation: 3-*O*-acetylations are in the equatorial plane of the sugar ring while 2-*O*- acetylations are axial. 2-*O*- and 3-*O*- acetylations are also present on xylan – the most common hemicellulose in dietary plants, although they are both present in the equatorial plane of the Xyl*p*. There are several reports of 6-*O* acetylations of mannan. The 6-*O* acetylations are considered to be a result of extraction methods involving high temperature or pH, which cause migration of acetyl groups (19).

Acetylations have significant impact on the processing potential of mannans. Acetylations improve the solubility of the oligosaccharides by restricting the formation of hydrogen bonds between oligosaccharides in solution (20). At the same time, the limited availability to microbes make acetylated mannans selective prebiotics (11). Effects of mannan acetylations on the viscosity of solutions make it an attractive constituent in hydrocolloids (21), thickeners or stabilizers (10, 22) in the food and feed industry where mannans such as guar gum and Konjac are commonly used (23, 24). The immunostimulatory properties of the common nutraceutical and cosmetics ingredient *Aloe vera* has been linked to its high degree of acetylation (16). In the context of biorefining, acetylations limit fermentability (11) and degradability and thus pose a limitation on the utilization of mannans as a feedstock for fermentation. Deacetylation of mannan by chemical pretreatment or enzymatic treatment is known to improve fermentation yields (25). Further research on the effect of hemicellulose acetylation is highly relevant in order to understand degradation of important dietary

compounds, and the development of new processes such as producing environmentally friendly hemicellulose based plastics (26) and including new feedstock into biorefineries.

Carbohydrate esterases

The esterases deacetylating xylan and mannan characterized so far have been classified together as acetylxylan esterases (EC 3.1.1.72). Carbohydrate esterases are common in bacterial and fungal genomes and often among genes encoding enzymes required for hemicellulose degradation. Their presence in the genetic clusters encoding polysaccharide active enzymes highlights their biological role as accessory enzymes enabling the bacteria to process the carbohydrate backbone. CEs are classified into 16 families according to the CAZy database (27), in which members of families 1-7 and 16 have been reported to be active on carbohydrates present in plant biomass. In the CAZy database, the enzymes are catalogued based on structure similarity and are, with the exception of family CE4, all serine-histidine hydrolases (28). The CE2 family share a two-domain architecture consisting of a catalytic SGNH hydrolase superfamily domain with a conserved catalytic dyad, and an accessory jelly roll domain (29). CE2 family esterases characterized so far have shown a wide range of deacetylation activities. Some of the CE2 family esterases are specific to 6-*O*- acetylations (30) when transacetylating mannose with vinyl acetate as the acetate donor. Others have activity on 2-*O*-, 3-*O*- and 4-*O*- acetylated xylan and 3-*O*-acetylations on xylose that also are substituted with methylated glucuronic acids on 2-*O*- (31), and 6-*O*-acetylated mannan (28). Deacetylation of konjac glucomannan by CE2 family esterase from *Cellvibrio japonicus* has been reported (32).

Enzymatic deacetylation of mannans has previously been described in literature (33, 34), but no esterase exclusively active on mannan has been reported. The structural features necessary for 2-*O*-acetylation selectivity have not been described either. Currently there are no entries for AGMEs in the ENZYME database (35) nor in CAZy (36). Acetyl esterases active on mannans reported in literature so far are classified as acetyl xylan esterases (32). Here, we report the detailed functional and structural characterization of *Ri*CE2 and *Ri*CEX - two mannan acetyl esterases, isolated from *Roseburia intestinalis*. The esterases have complementary substrate specificities and act in tandem to deacetylate glucomannan and galactoglucomannan. Furthermore, we report the first structure of an AGME with mannopentaose co-crystallized in the active site, revealing how the two-domain architecture facilitates specificity towards the axially oriented 2-*O*-acetylation. To the best of our knowledge, this structure is also the first carbohydrate esterase of any kind crystallized with a substrate.

Results and discussion

*Ri*CE2 is a CE2 family carbohydrate esterase with a galactose binding superfamily accessory domain

A gene cluster in *Roseburia intestinalis* encoding mannan degrading enzymes contains two acetyl esterases: *Ri*CEX and *Ri*CE2. *Ri*CE2 is a 349 aa CAZy family 2 esterase (41% sequence similarity to the xylan esterase Axe2C of *Cellvibrio japonicus* (32)). Analysis of the *Ri*CE2 sequence with InterProScan (37) identified residues 163-375 in the C terminal of *Ri*CE2 as an SGNH hydrolase endoglucanase-E-like fold, while residues 1-159 constitute a galactose binding superfamily domain. This two domain architecture is characteristic of CE2 family esterases (29).

To identify homologous domains, Hidden Markov Models (HMMs) of the accessory domains *Ri*CE2 was used for searching protein sequence databases (detailed description in the supplementary material). Searching the UniProtKB database (38) database with an HMM of the *Ri*CE2 N terminal domain identified 115 proteins that had a high similarity to the query HMM (>78.8%). Hits came mainly from two phyla: 87 in Firmicutes and 22 in Actinobacteria – another phylum of gram-positive polysaccharide degraders (39). 61 of these domains were in the same arrangement as in *Ri*CE2: with a C terminal Lipase_GDSL2 domain, suggesting that the accessory domain may play an important role in hydrolysis or substrate recognition and this organization seem to be particular feature for Firmicutes.

*Ri*CEX is a GDSL hydrolase/esterase with a novel accessory domain

*Ri*CEX is a 372 aa enzyme with very low sequence homology to any characterized esterases (20% sequence similarity with acetyl xylan esterase Axe2 of *Geobacillus stearothermophilus* (40)). Residues 35-209 in the N terminal of the *Ri*CEX sequence was identified by BlastP searches (41) as an SGNH hydrolase superfamily fold. InterProScan (37) was not able to predict a function for the 163 aa C-terminal domain of *Ri*CEX.

Using an HMM of the accessory domain of *Ri*CEX returned 79 significant (>45.2% identity) hits in the UniProtKB database (38). All 79 hits were from the Firmicutes phylum, 76 of which appeared in the same protein architecture as in the *Ri*CEX - as C terminal domains of Lipase_GDSL2 hydrolases.

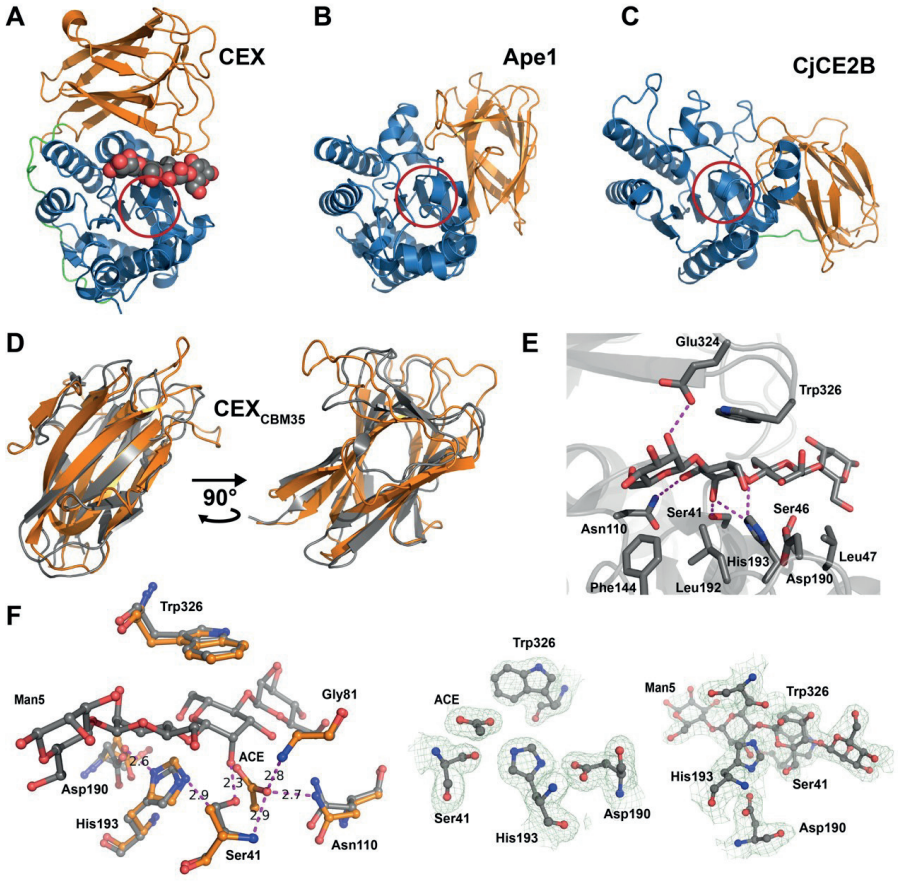


Figure 1. A) The structure of RiCEX, with four of the mannopentaose residues (spheres) visible in the electron density map. The catalytic domain (blue) is connected to the CBM35 domain (orange) by linker region (green). Panels B) and C) display the structural homologs Ape1 and CjCE2B that are closest to RiCEX in the PDB database, respectively. The catalytic domains (blue) in B) and C) are oriented to superimpose with the RiCEX catalytic domain in A). The red rings in the panels A-C indicate where the residues of the catalytic triad are located. In panel D) the RiCEX-CBM35 domain (orange) is superimposed on the CtCBM35 domain (gray) and shown at two rotations demonstrating the highly similar folds. E) Several charged or polar residues form hydrogen bonds to the mannopentaose ligand (Ser41, Asn110, His193 and Glu324). Ser46 may interact with the ligand through a water bridge. The hydrophobic residues Leu47, Phe144, Leu192 and Trp326 are part of the ligand-protein contact surface. Trp326 plays an important role, stacking on top of the mannose residue in the active site. Panel F) displays a superposition of the RiCEX with acetate (orange carbons) or mannopentaose (gray carbons) in the active sites. The corresponding electron density maps are shown to the right in panel F). In both structures Asp190 form a hydrogen bond with His193. In the acetate bound structure the His193-ε2N is hydrogen bonded with the Ser41-OH O-atom (2.9 Å). The O-atom of the acetate molecule pointing towards the oxyanion hole is hydrogen bonded by Ser41-N (2.9 Å), Asn110-Nδ2 (2.7 Å) and Gly81-N (2.8 Å). In the mannopentaose bound structure, the hydrogen bond between His193- ε2N and the Ser41-OH O-atom is replaced with a Ser41-OH O-atom hydrogen bond to the C2-OH group of the mannose residue in the esterase catalytic site (2.3 Å), implying that the Ser41 side chain change conformation during catalysis.

Structure of *Ri*CEX

We were able to obtain crystal structures for the *Ri*CEX in absence and presence of mannopentaose at 1.75 (PDBid 6hfz) and 2.4 Å (PDBid 6hh9) resolution, respectively. The enzyme (Fig 1A) consists of two domains that are connected by a linker (green) of about 20 amino acids. The N-terminal domain (ca. 210 amino acids, blue) displays an α/β-hydrolase (SGNH-hydrolase) fold while the C-terminal domain (ca. 140 amino acids, orange) comprises a jelly-roll beta-sandwich fold. Structural alignment against the Protein Data Bank using PDBe fold (42) using the complete structure returned hits with low scores, indicating significant structural differences. The two structures with the highest scores were Ape1, a peptidoglycan *O*-acetylesterase from *Neisseria meningitidis* (Fig 1B) (PDBid 4k40, Q-score 0.17) (43), and the CE2B carbohydrate esterase from *Cellvibrio japonicus* (Fig 1C) (PDBid 2w9x, Q-score 0.12) (32). In panels A, B and C the catalytic SGNH domains are aligned, and it is apparent that the second domain adopts different positions in space. These differences are not due to domain flexibility. For the Ape1 enzyme, the second domain is inserted in the middle of the SGNH domain sequence and is kept in place by two short linkers consisting of 3-4 amino acids. The second domain in the *Cj*CE2B enzyme, located at the N-terminal end of the SGNH domain, also adopt a very different position than found for the C-terminal domain in *Ri*CEX, connected to the SGNH domain via a linker (eight residues).

When performing structural alignments with the isolated, C-terminal and non-catalytic domain of *Ri*CEX (orange), the highest scoring protein is a carbohydrate binding module, CBM35 (gray), from *Clostridium thermocellum* (Fig 1D) (PDBid 2w1w, Q-score 0.39)(44). CBM35 modules have previously been demonstrated to bind decorated mannans (source CAZy, CBM35). The $CBM35_{RiCEX}$ domain is apparently involved in substrate recognition and binding, forming a lid over the SGNH-domain active site. Several hydrophobic amino acids, Ile79, Phe86, Ala89, Ile263, Leu330, Leu332 and Ile367, contribute to stabilize the inter-domain interaction together with the three pairs of charged or polar amino acid residues Glu119-His338, His116-Glu337 and Gln85-Thr334.

In (Fig. 1E) we show interactions between *Ri*CEX and the mannopentaose ligand. Two amino acids from the $CBM35_{RiCEX}$ domain contribute to ligand binding, namely the Glu324 and Trp326 residues. The Glu324 forms a hydrogen bond to the *O*2 hydroxyl group at the mannose at the non-reducing end next to the active site. Aromatic amino acids are often involved in carbohydrate interactions, and in *Ri*CEX the Trp326 side chain stack on top of the mannose unit in the active site. Interestingly, the two-domain arrangement result in cavities or clefts on each side of the active site (Fig. 2A), indicated by the two green and magenta patches. Both these cavities are arranged so that galactose residues linked to the mannose backbone at the *O*6 position would point towards these

spaces. This feature may both affect the affinity for galactose decorated mannans and influence the specificity of the enzyme. In the active site the three residues Ser41, His193 and Asp190 form the hydrogen-bonded catalytic triad and the amide nitrogen atoms of Ser41 and Gly81 and the Nδ2 of Asn110 line the oxyanion hole (indicated by dotted magenta lines to the acetate molecule) (Fig 1F). The high 2-*O*- acetylation specificity of *Ri*CEX can be explained from the ligand bound structure. Amino acids from both of the SGNH and CBM35 domains form specific interactions with the mannopentaose and aligns it such that the C2-OH group is only 2.3 Å from the Ser41-OH group (O-O distance), close to the oxyanion hole. In the apo- *Ri*CEX structure we can clearly see an acetate molecule in the oxyanion hole (Fig 1F). Such binding of acetate in absence of carbohydrate substrate has also been observed for Ape1 (43). The binding of acetate indicates how intermediates in the reaction may be positioned and provides structural insight useful for further mechanistic studies.

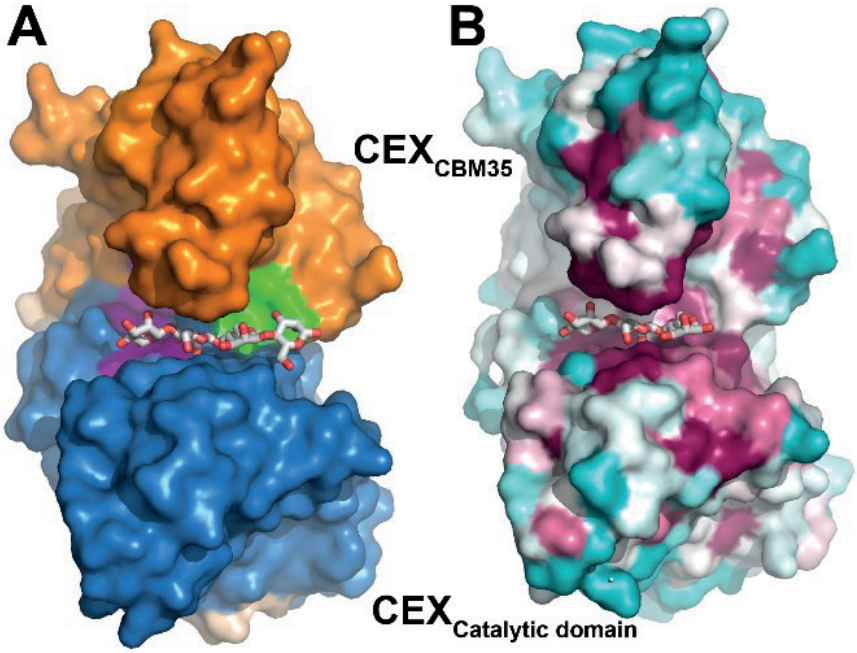


Figure 2. A) the magenta and green patches in the interface between the RiCEX catalytic domain and the CBM35 domain indicate cavities on the enzyme surface that may bind galactose residues decorating GGM at the C6 position. In panel B) the conservation score derived by ConSurf is projected on the RiCEX surface. The magenta patches indicate highly conserved residues and they are concentrated around the substrate binding site.

*Ri*CEX and *Ri*CE2 deacetylate complex GGM in a cooperative manner

The two esterases were tested on four types of mannan, birch glucuronoxylan, chitpentaose, and a mixture of cellulose monoacetate oligosaccharides. By selecting a wide range of mannose-based substrates was used to test the impact of galactose (Norway spruce mannan), multiple acetylations on single Man*p* (*Aloe vera* mannan) and different acetylation distributions on enzyme activity (konjac mannan, and its chemically acetylated version). Hydrolysates of GGM were generated by treatment with a GH26 family mannanase from the same genetic locus as the esterases (9). Neither *Ri*CE2 nor *Ri*CEX were active on birch xylan, cellulose monoacetylatete or chitopentaose (Fig. S1). The esterases showed the same pattern of deacetylation on all mannans; there was a partial deacetylation seen whenever one of the esterases was used on its own (red and blue traces Fig. 3 and S1). When the two esterases were used together, a near complete deacetylation was observed (purple traces Fig. 3 and S1).

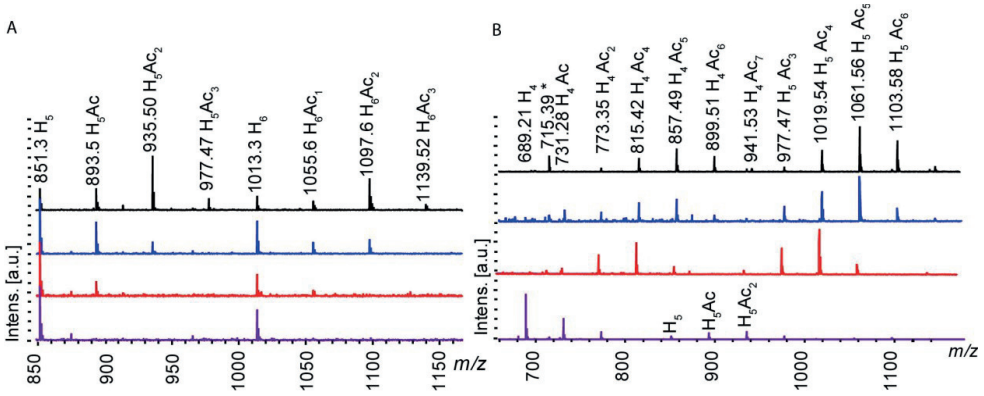


Figure 3. MALDI-ToF spectra of endpoint deacetylation reactions A: 10 mg/mL Spruce GGM (untreated in black) was treated with RiCE2 (in blue), RiCEX (in red) and the two enzymes in combination (purple). B: Aloe vera mannan treated with RiCE2 (in blue), RiCEX (in red) and the two enzymes in combination (purple). On both substrates when either of the esterases was used, a decrease in the intensitites of peaks signifying acetylated oligosaccharides was observed. Using the two enzymes in combination lead to a complete deacetylation of spruce mannan, and a near complete deacetylation of aloe vera mannan. H= hexose, Ac = acetylation

*Ri*CEX and *Ri*CE2 deacetylate complex mannans in a complementary manner

*Ri*CEX exclusively removed all 2-*O*-acetylations on AV and GGM (Fig 4B and 4F, respectively), whereas *Ri*CE2 removed the 3-*O*- and 4-*O*- acetylations as well as some of the 6-*O*- acetylations, leaving behind 2-*O*-acetylations and a minor part of the 6-*O*-acetylations (Fig. 4C and 4G). Adding *Ri*CEX at the end point of the *Ri*CE2 reaction lead to a complete deacetylation of both substrates (Fig.

4D and 4H). When the order of enzyme additions was changed, the end result was the same. The two esterases have interdependent specificities, and are both required for complete deacetylation of mannans.

*Ri*CEX is highly specific towards the axially oriented 2-*O*-acetylation on both single and double acetylated mannose residues. *Ri*CE2 is a broad specific esterase capable of removing 3-*O*- and 4-*O*-acetylations from single and multiply acetylated mannose residues, as well as 6-*O*-acetylations from single acetylated mannose, after the *Ri*CEX removes the 2-*O*-acetylation. *Ri*CE2 was able to remove the 6-*O*-acetylations, but only when the 2-*O*-acetylations were removed. Notably, CE2 is much slower when 2-*O*-acetylations are present, whereas CEX apparent activity is not affected if acetylations at other positions are present (Table 1).

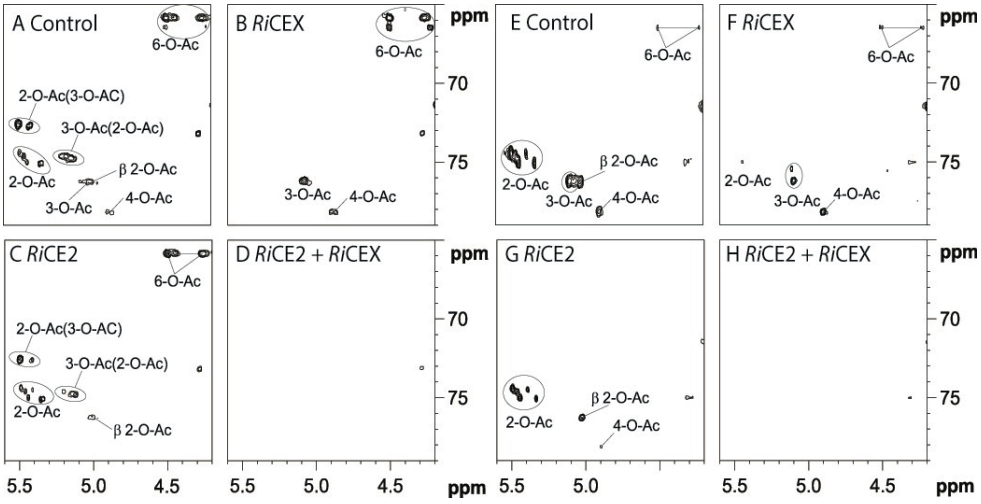


*Figure 4. HSQC NMR Spectra of Aloe vera mannan GH26 hydrolysate (panels A-D) and Norway spruce mannan GH26 hydrolysate (panels E-H) treated with the esterases. A: Aloe vera hydrolysate without enzyme addition, contained a high degree of acetylation on all possible positions, and a high degree of double acetylation. B: When treated with Ri*CEX, *the peaks corresponding to 2-O- acetylations dissappear, and the shift values for peaks corresponding to 3-O- and 6-O-acetylations change due to the removal of C2 acetylations from double acetylated mannoses. C: Treatment with Ri*CE2 *removed all the non-reducing end 4-O- acetylations, the majority of 3-O- acetylations, and some of the 6-O-. D: Treatment with both esterases at the same time removed nearly all acetylations, with minor peaks for 6-O- remaining. E: Spruce mannan hydrolysate without enzyme addition contained prevalently 2-O-, some 3-O- acetylations, and a lower degree of acetylation on carbon 4-O- and 6-O-. F, G: Both enzymes exhibited similar activity on spruce mannan as on AV. H: Notably, at the end of reaction with both enzymes, there were more acetylations remaining on spruce than on aloe vera. The higher amount of 2-O- and 3-O- acetylations on spruce, a substrate that has a higher galactose content implies that the presence of galactosylations may partly restrict the oligosaccharides entry into the active sites.*

*Ri*CEX reactivity was not affected by the presence of 3-*O*-, 4-*O*- and 6-*O*-acetylations, with near identical turnover rate when applied directly to AV solution as when applied to a CE2 treated AV (Table 1). On the other hand, CE2 applied directly to the AV solution has shown a marked decrease in turnover compared to the reaction on Aloe vera after *Ri*CEX treatment. These results demonstrate that the two enzymes have a different spatial dependency on acetyls in non-catalytic positions. The first sample allowed for the determination of *Ri*CE2 activity towards 3-*O*-acetylation and double 2-*O*-, 3-*O*- acetylation, as well as the activity of *Ri*CEX towards 2-*O*-acetylation, with no 3-*O*-acetylations interfering. Sample 2 allowed for the determination of activity of *Ri*CEX toward 2-*O*- and double 2-*O*-, 3-*O*- acetylations, as well as activity of *Ri*CE2 towards 3-*O*-, 6-*O*-, and 3-*O*-, 6-*O*- double acetylations. The presence of 3-*O*- and 6-*O*-acetylations on the same mannose residues containing 2-*O*-acetylation did not affect the k_{cat} of *Ri*CEX. However, the catalytic activity of *Ri*CE2 has more than doubled when 2-O- acetylations were removed (Table 1). The data demonstrate that *Ri*CE2 is a generalist with an ability to remove 3-*O*-, 4-*O*- and 6-*O*-acetylations, whereas *Ri*CEX is the specialist esterase, removing the 2-*O*-acetylations. *In vivo*, the enzyme pair would act in complement, with the specialist, *Ri*CEX, removing the 2-*O*-acetylations which limit the activity of the generalist *Ri*CE2.

Table 1. Turnover rate in time resolved NMR analysis of deacetylation determined as initial rates measured over 15 minutes. Deacetylation of GH26 hydrolysate of Aloe vera *was followed by time resolved NMR in order to determine the preferences of the esterases towards a particular acetylation. Two identical samples of hydrolysate were prepared and treated with 1) 62.5 nM loading if* Ri*CE2 for 16 hours, then 10 nM loading of* Ri*CEX 2) 10 nM loading of* Ri*CEX for 16 hours, then 62.5 nM loading of* Ri*CE2.*

Sample 1:	**CEX**	**CE2 after CEX**
k_{cat} [s^{-1}]	4.78	4.67
Sample 2:	**CE2**	**CEX after CE2**
k_{cat}[s^{-1}]	1.77	4.48

Transacetylation experiments with vinyl acetate

Transacetylation of oligosaccharides with vinyl acetate as donor has been used to determine the of carbohydrate esterases positional preference of acetylation (46). *Ri*CE2 and *Ri*CEX were able to transacetylate manno-oligosaccharides, and galactomanno-oligosaccharides, but had no activity on xylo-oligosaccharides Elevated pH and temperature (>90°C) can induce acetyl migration on mannose and xylose (47-49), thus we put a lot of attention to minimize the migration in reaction conditions used for the characterization of our esterases. We utilized the strong preference for 2-O acetylatylations and transacetylation behavior to generate oligosaccharides that we strictly 2-O-acetylated. pH 5.9 allowed for the enzymatic reaction to occur but at the same time prevented

migration (Fig. S3). A detailed description of acetyl migration in response to temperature and pH is provided in the supplementary information (Fig. S3).

Catalytic activities of *Ri*CEX and *Ri*CE2

The catalytic parameters of the esterases vary slightly (Fig. 6). *Ri*CEX has a pH optimum of 7.0, while *Ri*CE2 has a pH optimum of pH 7.5. *Ri*CEX was active and stable in the entire pH range from 5.0 to 8.5, with a linear release of acetate over the reaction course of 15 minutes. *Ri*CE2 on the other hand, was very sensitive to pH, and became inactive within 10 minutes at pH 8.0 and within 4 minutes at pH 8.5. For *Ri*CEX the temperature optimum is 35 °C, and the enzyme was stable up to 45 °C. For *Ri*CE2, the temperature optimum is 25 °C, and the enzyme was sensitive to temperature, deactivating quickly at 55 °C.

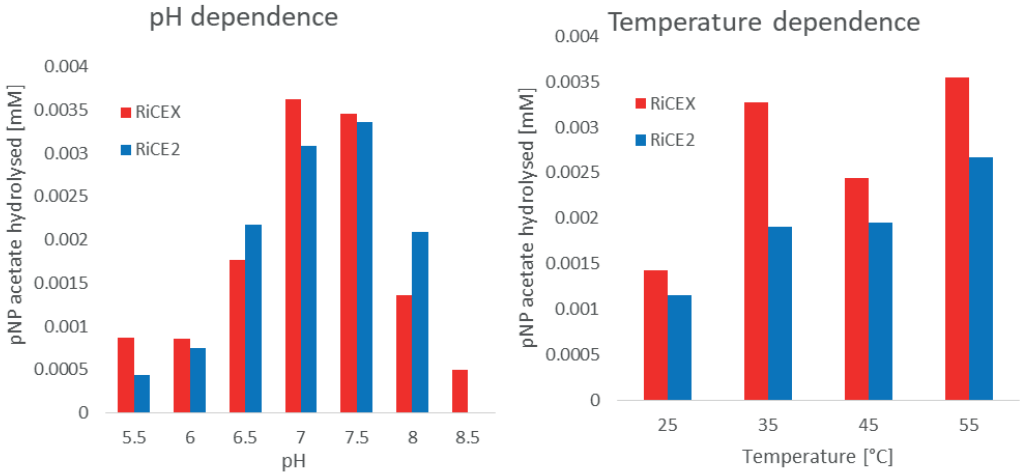


Figure 6. Bar charts of pNP acetate hydrolysis dependence on reaction pH and temperature. pNP release measured after 270 seconds at pH range 5.0-8.5 shows the pH optima of 7.0 for RiCEX and 7.5 for RiCE2. Temperature dependence was measured by measuring the pNP release after 210 seconds at pH 7.5 and temperature range 25-55° C. Both enzymes showed the highest deacetylation rate at 55° C, however neither of them was stable for more than 5 minutes at 45 and 55° C.

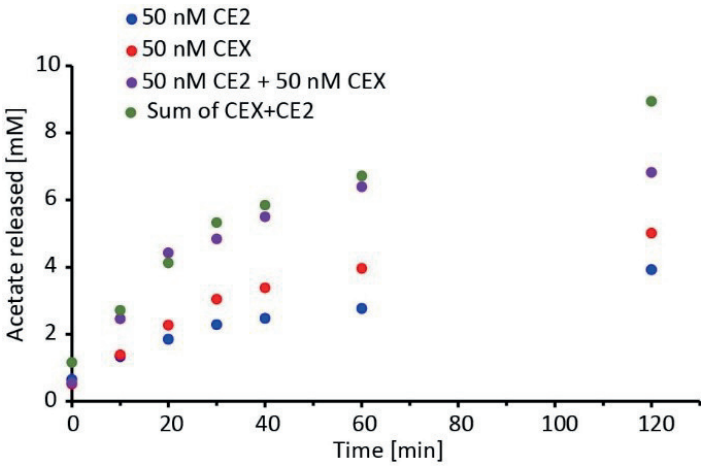


Figure 7. Plots of acetate release in a timecourse reaction with 50nM loading of RiCEX (red), RiCE2 (blue), both enzymes together (purple), and a sum of acetate released in the RiCE2 and RiCEX treatments for each timepoint (green) for comparison. RiCEX has liberated more acetate from the spruce mannan and at a faster rate in the initial 40 minutes of the reaction. There was no appreciable increase in acetate release rate between the two enzyme treatment compared to the sum of two single enzyme treatments, indicating that while the enzymes do have complementary activities, they do not exhibit a synergistic effect.

Table 2. Specific activity of the two esterases towards spruce mannan and pNP acetate, and the respecitve turnover rates of the enzymes in the same reactions. Activities and rates were calculated from the initial timepoints (0-10 minutes) of reactions with 10 mg/ml spruce mannan and 0.5 mM pNP acetate, both in sodium phosphate buffer at pH 7.0 and 30°C.

	Esterase activity in nmole min-1 µg-1				**Turnover rate k_{cat} in s^{-1}**			
	Spruce GGM		pNP Acetate		Spruce GGM		pNP Acetate	
***Ri*CE2**	62.74	SD 2.12	155.94	SD 3.65	46.01	SD 1.56	114.35	SD 2.67
***Ri*CEX**	51.40	SD 4.25	0.19	SD 0.012	36.32	SD 0.30	0.14	SD 0.01

Concluding remarks

This study provides detailed insight into enzymatic deacetylation of mannans by an important human gut commensal bacterium. Deacetylation is a crucial step in utilization of mannans by the gut microbiota, and this pair of *Roseburia* esterases provide a paradigm for this metabolic function. The deacetylation apparatus of *Roseburia* consists of highly specialized 2-*O*-acetylations specific *Ri*CEX, and *Ri*CE2 which can remove all types of GGM acetylations, once the 2-*O*-acetylations have been removed. Both enzymes are necessary for complete deacetylation, and their activities are complementary. Both esterases have a two-domain structure with an SGNH superfamily hydrolytic domain and an accessory domain – a galactose binding superfamily domain in *Ri*CE2 and a CBM35 in *Ri*CEX. The accessory domains of these enzymes are present in multiple gram-positive polysaccharide degraders and could be a specific adaptation of common hydrolytic enzyme domains to complex glycans.

The potential for industrial application of this enzyme pair is quite apparent. With the activity on gluco- and galactoglucomannans, the pair could be applied to various mannan-containing feedstocks to reduce the recalcitrance of mannans or to rationally design oligosaccharides with distinct acetylation pattern via specific deacetylation or transacetylation.

Acknowledgments:

Cellulose monoacetate with a degree of acetylation of 0.6 was a kind gift from Qi Zhou. A chemically acetylated konjac glucomannan prepared according to Bååth et al. (50) was kindly provided by Francisco Villaplana.

Materials and Methods

1. Enzymes.

*Ri*CEX (SGNH/GDSL hydrolase family protein [*Roseburia intestinalis*] NCBI Reference Sequence: WP_006855599.1) is a 372 amino acids, 42.5 kDa serine-histidine-asparagine hydrolase (active residues: Ser-41; Asp-190; His-193) with a predicted pI of 5.28. It shares a very low, 20% similarity with an octameric acetyl xylan esterase Axe2 of *Geobacillus stearothermophilus* (40). The amino acid sequence of *Ri*CEX is:

MEYQIKYENGIANRGCLYRLKKVMDRAKAGEALNIAFLGGSITQGSLSSKPELCYAYHVYEWWKKTFPQADFTYI
NAGIGGTTSQFGVARAEADLLSKEPDFVIIEFSVNDDSTEHFMETYEGLVRKVYTSKTKPAVLLVHNVFYNNGANA
QLMHGRIARHYNLPAVSMQSTIYPEVVAGRIENREITPDDLHPNDAGHALVASVITYFLDKVKTEDATEQSEPDY
PAPLTKNTYEKSIRHQNSDENVVCHGFVADTSAQRDITDCFKHGWTASKKGDSITLDVEGCNISVQYRKSVKLPA
PVAEIIVDGDAEHAVRLDANFDETWGDKLELDTILEHGENKVHKVEVRLTETHENDAVPFYLVSVIGSSEK

*Ri*CE2 (hypothetical protein ROSINTL182_05483 [*Roseburia intestinalis* L1-82]), Gen Bank accession no. EEV02560.1 is a 40.5-kDa serine-histidine-asparagine hydrolase (active residues: Ser-134, Asp-319, His-322) with a predicted pI of 5.27. It shares 41 % similarity with the xylan esterase Axe2C of *Cellvibrio japonicus* (32). It consists of 349 amino acids:

MIIKPDNEMLSYSGRIDFDDRLAPVLVYACSSIGMKFAGTSLKAVIANHRSCWTNELGYFIDGEQKRFTLSSDEEKK
TYTLAEGLSEGTHELLLFKRMDSCHTFTFYGFEIDDGAEVLPLPEKPKRKMEFFGDSVSCGEVSEAVAYVGKPDPE
HDGEYSNSWYSYAWMTARKLNAQIHDTSQGGISLLDDTGWFAAPHYKGVESCYDKIEYHPDLGPTKQWDFSKY
VPHVVVVAIGQNDNHPVDYMAEDYDSEKSKNWRKHYQAFIEKLMQLYPKAQIILATTILCHDKSWDRSIDEVCT
RIGSERVHHFLYTKNGSGTPGHIRIPEAEQMSDELAAYINSLGDAVWES

2. Sequence analysis.

Aminoacid sequences of the non-hydrolase parts of *Ri*CEX and *Ri*CE2 sequences (highlighted yellow above) were used for a protein vs protein database BlastP search. The HMM was in turn used with a HMMER tool to search the UniProtKB (https://www.uniprot.org/help/uniprotkb) database for characterized related protein sequences and their taxonomy. A detailed description of the HMM building process is in the supplementary.

3. Substrates.

GGM from Norway spruce (*Picea abies*) and glucoronoxylan from Birch (*Betula pubescens*) were produced in house from dried wood chips. The wood was milled into <2mm particles, and steam exploded at 200°C (10 minutes reactor residence time) in 5-6 kg batches. The liquid soluble fraction containing the hemicelluloses was extracted by washing the steam-exploded material with MilliQ water in a 50µm pore WE50P2VWR bag filter (Allied filter systems, England). The liquid fraction of hemicellulose was then filtered through a 5 kDa spiral wound Polysulphone/polyethersulphone polyester ultrafiltration membrane (GR99PE, Alfa Laval, Denmark) using a GEA pilot scale filtration system Model L (GEA, Denmark). The fraction retained by the membrane was concentrated, collected and freeze-dried to become the R5K sample. Birch xylan was produced according to the protocol described in Biely *et al*. (51). 6^1-α-D-Galactosyl-mannotriose; 6^3, 6^4-α-D-Galactosyl-mannopentaose, penta-N-acetylchitopentaose, mannotriose, mannotetrose, mannopentaose, konjac and carob mannans (Megazyme, Ireland). *Aloe vera* mannan (Acemannan; Elicityl, France).

4. Activity analysis.

Activities of the enzymes were checked by adding 1 µL of enzyme stock solution (~1µM) to 10 mg/mL solutions of carbohydrate substrates listed above. All substrate solutions were prepared with 20mM sodium phosphate pH 5.9 to prevent acetyl migration. Reactions were ran overnight in 20mM sodium phosphate pH 5.9 at 30°C with 700 rpm shaking. In some analyses, MilliQ water was used instead of buffer to reduce background signals in MALDI-ToF MS as both enzymes were found to be active without buffers. Sequential enzymatic treatment was conducted by filtering hydrolysis products through a pre-washed 1 mL Amicon Ultracel 3kDa ultrafiltration device (Merck KGaA, Germany) in a benchtop Thermo Pico21 centrifuge (Thermo Scientific, USA), at 12.000 g and collecting the permeate. Permeate and retentate were analyzed on electrophoresis gels (Mini-protean TGX stain-free gels, Biorad, Norway) to ensure successful enzyme removal.

5. Gene cloning and expression.

The genes encoding the *Ri*CE2 and *Ri*CEX enzymes were amplified from *R. intestinalis* genomic DNA using the primer pairs CE2_up/CE2_down and CEX_up/CEX_down, respectively (Table S1). Fragments were cloned into the pNIC-CH expression vector with a C-terminal hexa-histidine tag by ligation-independent cloning (LIC) (52), giving the constructs pNIC-*Ri*CE2 and pNIC-*Ri*CEX. Transformants were verified by sequencing. The two proteins were expressed in *Escherichia coli* BL21(DE3) cells harboring the appropriate recombinant plasmids. The recombinant *E. coli* strains

were pre-cultured overnight in Luria Bertani (LB) broth supplemented with 50 µg/mL kanamycin (Sigma-Aldrich, Germany) and then used to inoculate 500 mL of medium consisting of 450 mL LB, 50 µg/mL kanamycin and 50 mL of potassium phosphate buffer (0.17 M KH_2PO_4, 0.72 M K_2HPO_4). Protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM for 16 h at 23°C. Cells were harvested by centrifugation (6000 g for 10 minutes) and resuspended in 30 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10 mM imidazole). A cell-free extract was prepared by pulsed sonication and centrifugation at 15000 g for 15 minutes. The supernatant containing the soluble proteins was collected and filtered with 0.22-µm syringe filters. Recombinant *Ri*CE2 and *Ri*CEX were purified by immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). Protein purity was verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein concentration was determined using the Bradford assay (Bio-Rad) with bovine serum albumin as a standard.

For crystallographic studies, seleno-L-methionine substituted *Ri*CEX was produced by first transforming pNIC-*Ri*CEX into *E. coli* B834(DE3) cells via the heat-shock method and plating the bacteria onto LB agar supplemented with 50 µg/mL kanamycin. Recombinant cells were grown in Medium A according to the EMBL protocol (EMBL protocol), supplemented with 50 µg/mL methionine (Sigma-Aldrich, Germany) and 50 µg/mL kanamycin at 37°C until an OD_{600} of 0.8 was reached. At this point, cells were harvested by centrifugation at 6000 g for 15 minutes and resuspended in an equal amount of kanamycin-supplemented Medium A. Temperature was adjusted at 23 °C and, following a 2 h starvation period, the flask was supplemented with 50 µg/mL seleno-L-methionine (Sigma-Aldrich, Germany). After 30 min of further incubation, seleno-L-methionine substituted *Ri*CEX expression was induced with 0.5 mM IPTG and cultures were allowed to grow for an additional 48 h before being collected. Cells were harvested by centrifugation at 5000 g for 10 minutes, cell pellets were resuspended in 30 mL of 50 mM Tris buffer containing 10 mM Imidazole and 500 mM NaCl and lysed by sonication. The cell lysate was centrifuged at 12000 g for 15 minutes and the supernatant containing the proteins was collected, sterile filtered and purified as described below.

Table S1: Primers used in this study. 5' extension sequences used for molecular cloning are underlined.

Gene	Primer (5' -3')
CEX_up	**F:**TTAAGAAGGAGATATACTATGGAATATCAAATTAAATACGAAAACGGC
CEX_down	**R:**AATGGTGGTGATGATGGTGCGCTTTTTCAGAGGAACCAATGACAGAC
CE2_up	**F:**TTAAGAAGGAGATATACTATGAAACGTGTGATGGAGTGTCCG
CE2_down	**R:**AATGGTGGTGATGATGGTGCGCAGATTCCCAGACTGCATCCC

6. Protein purification

The filtered protein solutions were applied to a 5 mL Histrap column (GE Healthcare, Norway), washed with the binding buffer (50 mM Tris-HCl pH 8.0 with 200 mM NaCl and 5 mM Imidazole) until a steady UV reading was reached. The bound protein was then eluted with a linear gradient of 0-75% elution buffer (50 mM Tris-HCl pH 8.0 with 200 mM NaCl and 500 mM Imidazole) over five column volumes (CVs). The elution gradient was followed by an immediate increase to 100 % B that was kept for 5 CV. Fractions corresponding to the peaks were collected using automated fractionation and the content of each fraction was analyzed using Mini-Protean TGX Stain-Free Gels (Biorad, Norway). Fractions containing the protein of correct size were pooled and concentrated using Amicon Ultra Centrifugal device with a 3000 NMWL membrane (Merck KGaA, Germany). The buffer was exchanged to 20 mM Tris-HCl pH 8.0 with 200 mM NaCl (for *Ri*CEX) or 20 mM sodium phosphate pH 5.9 with 200 mM NaCl (for *Ri*CE2) and the protein solution was further purified with a HiLoad 16/600 Superdex 75 pg size exclusion chromatography column. The protein was eluted with the same buffers, isocratically at 0.8 mL/min. Fractions (1.0 mL) were collected and purity checked on Mini-Protean TGX Stain-Free Gels (Biorad, Norway).

7. Temperature and pH optima using 4-nitrophenyl acetate.

In order to determine the pH and temperature optima for the enzymes, reactions with 0.5 mM pNP acetate were prepared using 50 mM sodium phosphate buffer (Sigma-Aldrich, Germany) in the pH range 5.5-7.5, and 50 mM Tris-HCl (Sigma-Aldrich, Germany) in the pH range 8.0-9.0. pH optimum reactions were ran at 30 °C. Temperature optima were determined by adding the enzymes to a sample mix with 0.5 mM pNP acetate in 50mM sodium phosphate pH 7.5 at 25, 35, 45 and 55 °C. Samples were analyzed in a 96 well plate reader at 60 s intervals, with intermittent shaking. Due to the difference in deacetylation rate of pNP acetate by the two enzymes, 0.1 nM final concentration of *Ri*CE2 and 0.1 µM of *Ri*CEX were used in the pNP acetate experiments. Standard plots were prepared at each pH. To determine the optimum pH, 99 µL of sample mixture containing 0.5 mM of 4-nitrophenyl acetate in each of the buffer were added to the wells of a 96 well plate. 1 µL of enzyme solution was then added to the sample mixture and, the reaction was followed by measuring the absorbance at 405 nm at one minute intervals for 15 minutes. Triplicates of enzyme reactions with two blanks were prepared for each condition.

For the optimum temperature, samples were prepared and measured in the same manner, with 50 mM sodium phosphate buffer pH 7.5 used as buffer and temperatures of 25°C, 35°C, 45°C and 55°C.

536

537 8. HPLC Measurement of acetate release

538 Acetate content was analyzed on an RSLC Ultimate 3000 (Dionex, Norway) HPLC using a REZEX
539 ROA-Organic Acid H+ 300x7.8mm ion exchange column at 65°C, 5 µL injection volume, with isocratic
540 elution using 0.6 mL/min of 4mM H_2SO_4 as mobile phase and a UV detector set to 210 nm.

541

542 9. Crystallography.

543 *Ri*CEX and *Ri*CE2 crystallization conditions were screened using several commercial high
544 throughput 96 condition sitting drop screens, using 20 mg/mL, 10 mg/mL and 5 mg/mL solutions of
545 protein in a 1:1 ratio with ready-mixed mother liquors (20 nL of each). Screening plates were set up
546 using a mosquito HTS liquid handling robot (ttp labtech, UK). Crystals were observed in spot G8 of
547 the INDEX HT screen (Hampton Research, USA), containing 0.2M ammonium acetate, HEPES pH 7.5
548 and 25% w/v PEG 3350. Hanging drop optimization grids were manually set up with 0.2 M ammonium
549 acetate, HEPES pH 7.0, 7.5 and 8.0, 20%, 25% and 30% w/v PEG 3350. Crystallization liquor was
550 mixed 2 µL:2 µL and 1 µL:1 µL in the hanging drops, with additional 2 µL or 1 µL of 5 mg/mL solution
551 of mannotriose, mannotetrose and mannopentose (Megazyme, Ireland) for co-crystallization.

552 Crystals were transferred to a cryo-solution containing 35 % glucose in mother-liquor before
553 flash freezing in liquid nitrogen. Diffraction data were collected at beamlines ID23-1 and ID-29 at the
554 European synchrotron Research Facility in Grenoble, France. The initial structure was solved by
555 single-wavelength anomalous diffraction (SAD) using selenomethionine to obtain an anomalous
556 signal. Data was processed by XDS (53) and scaled by AIMLESS (54). The Phenix software package
557 (55) was used to phase (AUTOSOL) (56) and build (AUTOBUILD) (57) the first structure. Subsequent
558 structures were solved by molecular replacement (Phaser) (58) and refined using REFMAC (59).
559 Model manipulations were carried out using Coot (60) and molecular graphics were generated using
560 Pymol2 (Schrödinger).

561

562

563

564

565

Table S2 Crystal data, data collection, and refinement statistics

[a] *Values for outer shell in parenthesis,*

[b] $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$.

[c] $R_{cryst} = \Sigma(||F_{obs}| - |F_{calc}||)/\Sigma|F_{obs}|$

[d] R_{free} *is the* R_{cryst} *value calculated on the 5 % reflections excluded for refinement.*

	Crystal data	
	*Ri*CEX	*Ri*CEX-mannopentaose
Space group	P 1 2_1 1	
Crystal parameters	a = 75.12, b = 135.52, c = 85.13	a = 75.48, b = 136.69, c = 85.41
	α = 90, β = 115.09, γ = 90	α = 90, β = 113.86, γ = 90
	Data collection	
X-ray source	ESRF, ID23-1	ESRF, ID23-1
Resolution (Å)[a]	48.2-1.75 (1.78-1.75)	48.6-2.40 (2.53-2.40)
Wavelength (Å)	0.97531	0.97625
Temperature (K)	100	100
Number of unique reflections	155949 (7703)	61043 (8942)
Completeness[a]	100 (99.9)	98.7 (99.2)
Redundancy[a]	6.7 (6.4)	3.5 (3.5)
CC half[a]	0.998 (0.815)	0.995 (0.713)
I/ σ(I)[a]	11.5 (2.2)	8.2 (2.0)
R_{sym}[b]	0.079 (0.631)	0.097 (0.593)
	Refinement statistics	
R_{cryst}[c]	0.190	0.181
R_{free}[d]	0.232	0.253
Wilson B-factor (Å^2)	25.4	42.1
Ramachandran plot,in most favored/other allowed regions (%)	96/4	95/5
Standard Uncertainty (Maximum Likelyhood):	0.093	0.224
Added waters	808	487
PDB code	6HFZ	6HH9

572 10. MALDI-ToF analysis:

573 MALDI-ToF analysis of hydrolysis product was conducted on an UltraFlextreme MALDI-ToF
574 instrument (Bruker Daltonics GmbH, Germany) equipped with a nitrogen 337 nm laser beam.
575 Samples were prepared by mixing 2 µL of a 9 mg/mL solution of 2,5-dihydroxybenzoic acid (Sigma-
576 Aldrich, Germany) in 30% acetonitrile (VWR) to an MTP 384 ground steel target plate (Bruker
577 Daltonics GmbH, Germany) and 1 µL of hydrolysate (0.1-1 mg/mL). Sample drops were then dried
578 under a stream of warm air.

579

580 11. Hydrophilic interaction liquid chromatography mass spectrometry (HILIC-LC-MS)

581 HILIC-LC-MS was conducted on a Dionex UltiMate 3000 RSLC nano system (Thermo Scientific,
582 Waltham, USA), using an Xbridge Amide 3.5 µm particle size 4.6x250mm column (Waters, Norway).
583 Five µL injections of analytes in 75 % acetonitrile (VWR) were eluted with 75 % acetonitrile in MilliQ
584 water from 0 - 5 minutes, followed by a gradient of 75 – 50 % acetonitrile between 5-30 minutes, 10
585 minutes of 50 % acetonitrile, and 5 minutes of 75 % acetonitrile, all at a flowrate of 0.6 mL/min. A
586 50:50 T-split after the column divided the eluted analyte between a Corona Ultra CAD detector
587 (Thermo Scientific, USA) and a Velos Pro ion trap mass spectrometer (Thermo Scientific, USA).

588

589 12. Preparative HPLC.

590 Preparative chromatography was conducted using an Agilent 1260 Infinity preparative
591 chromatography system with an XBridge BEH prep OBD 5 µm particle size 30x250mm column.
592 Analytical HILIC method was scaled up to 17.5 mL/min flow with 3.85 mL injections of 1-5 mg/mL
593 carbohydrate concentrations. Elution started with a step of 0 – 3.57 min 75 % acetonitrile, followed
594 by a linear gradient of 3.58 – 14.28 min 75 % - 50 % acetonitrile, a step of 14.29 - 21.42 min 50 %
595 acetonitrile, and a final step of 21.43 – 33 minutes at 75 % acetonitrile. Fractions were collected as
596 one minute time slices, acetonitrile evaporated in a fume hood and liquid fractions freeze-dried.

597

598 13. Transacetylation.

599 Transacetylation of oligosaccharides was conducted using vinyl acetate (Thermo scientific, USA)
600 as acetate donor. Enzymes were added to oligosaccharide solutions with concentrations from 1 – 10
601 mg/mL, and a volume of vinyl acetate equal to 20 – 50 % sample volume was added. The samples
602 were left in a thermomixer (Eppendorf, Norway) shaking at 600 rpm overnight, then moved to a
603 freezer at -20 °C. The vinyl acetate, which remained liquid on top of the frozen aqueous phase was

removed from the samples, and 96% ethanol was added until the final concentration exceeded 80% in order to deactivate the enzymes. Samples were thawed by vortexing and filtered through a pre-washed 1mL Amicon Ultracel 3kDa ultrafiltration device (Merck KGaA, Germany) to remove the enzymes completely and minimize the risk of deacetylation. Oligos were then dried in an Eppendorf Concentrator plus (Eppendorf, Norway) at room temperature.

For the purpose of preparative chromatography, samples were frozen, liquid layer of vinyl acetate was removed and samples were then diluted with acetonitrile to a 75 % concentration, foregoing the filtration step.

14. NMR

To reduce the interference of the water signal the substrate (GH 26 treated spruce galactoglucomannan R5k) was dissolved in 99.9% D_2O (Sigma-Aldrich, Germany) and lyophilized. Similarly, 10 mL 40 mM phosphate buffer pH 5.9 and 250 mM phosphate buffer pH 8.0 were lyophilized and the powder were dissolved in 10 mL 99.9% D_2O.

For the time-resolved NMR recordings: 4-5 mg of R5K spruce hydrolyzed with the *Ri*GH26 mannanase or *Ri*CEX transacetylated mannotriose were dissolved in 500 µL 40 mM phosphate buffer pD 5.9 (99.9% D_2O) and transferred to an 5 mm NMR tube. The sample was preheated in the NMR spectrometer for ~10 min. Hereafter all recording parameters were set prior to the time-resolved NMR experiment. 2 or 5 µL of enzyme solution (to a final concentration of 1 µM *Ri*CEX or 10µM *Ri*CE2) was added to the preheated substrate and mixed by inverting the sample three times. The sample was then immediately inserted into the preheated NMR spectrometer and the experiment was started (time from adding the enzyme to the first spectra has been recorded was between 3-4 min. totally). The recorded spectrum is a pseudo-2D type experiment recording a 1D proton NMR spectrum with weak water suppression (Based on Bruker 1D proton setup for metabolomics noesygppr1d) every 5 min with in total 200 time points. The recorded 1D proton spectrum contains 32K data points and has a spectral width of 10 ppm, 24 scans, and pre-saturation during relaxation delay and 10 ms mixing time with spoil gradient and relaxation delay of 1 s (total recording time of 89s).

To monitor the effect of temperature and pH on acetyl migration 2 mg transacetylated mannotriose was dissolved in 500 µL 40 mM phosphate buffer pD 5.9 (99.9% D2O) and it served as an control sample where 1D proton and 2D ^{13}C heteronuclear single quantum coherence (HSQC) with multiplicity editing spectra were recorded. The sample was split into 3 samples of 160 µL each and transferred into 3mm NMR tubes. Hereafter, the first sample was heated to 60 °C for 60 min. In the

second sample pD was adjusted pD 7.4 by adding 20 µL of 250 mM phosphate buffer pD 8.0 and in the third sample the pD was adjusted pD 7.4 by adding 20 µL of 250 mM phosphate buffer pD 8.0 and heated to 60 °C for 60 min. A 1D proton and 2D ^{13}C HSQC spectra were recorded at 25 °C for each of the samples.

All homo and heteronuclear NMR experiments were recorded on a BRUKER AVIIIHD 800 MHz (Bruker BioSpin AG, Switzerland) equipped with a 5 mm cryogenic CP-TCI. All NMR recordings were performed at 37°C. For chemical shift assignment of *Ri*CEX transacetylated mannotriose, the following spectra were recorded: 1D proton, 1D proton with presaturation during relaxation delay and 10 ms mixing time with spoil gradient, 2D double quantum filtered correlation spectroscopy (DQF-COSY), 2D total correlation spectroscopy (TOCSY) with 70 ms mixing time, 2D ^{13}C HSQC, 2D ^{13}C Heteronuclear 2 Bond Correlation (H2BC), 2D ^{13}C HSQC-[^{1}H,^{1}H]TOCSY with 70 ms mixing time on protons and 2D heteronuclear multiple bond correlation (HMBC) with BIRD filter to suppress first order correlations. The water signal to 4.75 ppm (at 25 °C, pH 5.5 (61)) was used as chemical shift reference for protons, while ^{13}C chemical shifts were referenced indirectly, based on the absolute frequency ratios (62). The spectra were recorded, processed and analyzed using TopSpin 3.5 software (Bruker BioSpin AG, Switzerland).

Supplementary

To investigate the substrate specificity of the esterases, their activity was tested on a wide range of relevant substrates. Activity was only observed on mannose based oligosaccharides, and the patterns of activity were always similar: a partial deacetylation when either of esterases was, and a near complete deacetylation when both enzymes were used. MALDI-ToF spectra illustrating the results of these experiments are presented in Fig. 3 and Fig. S1.

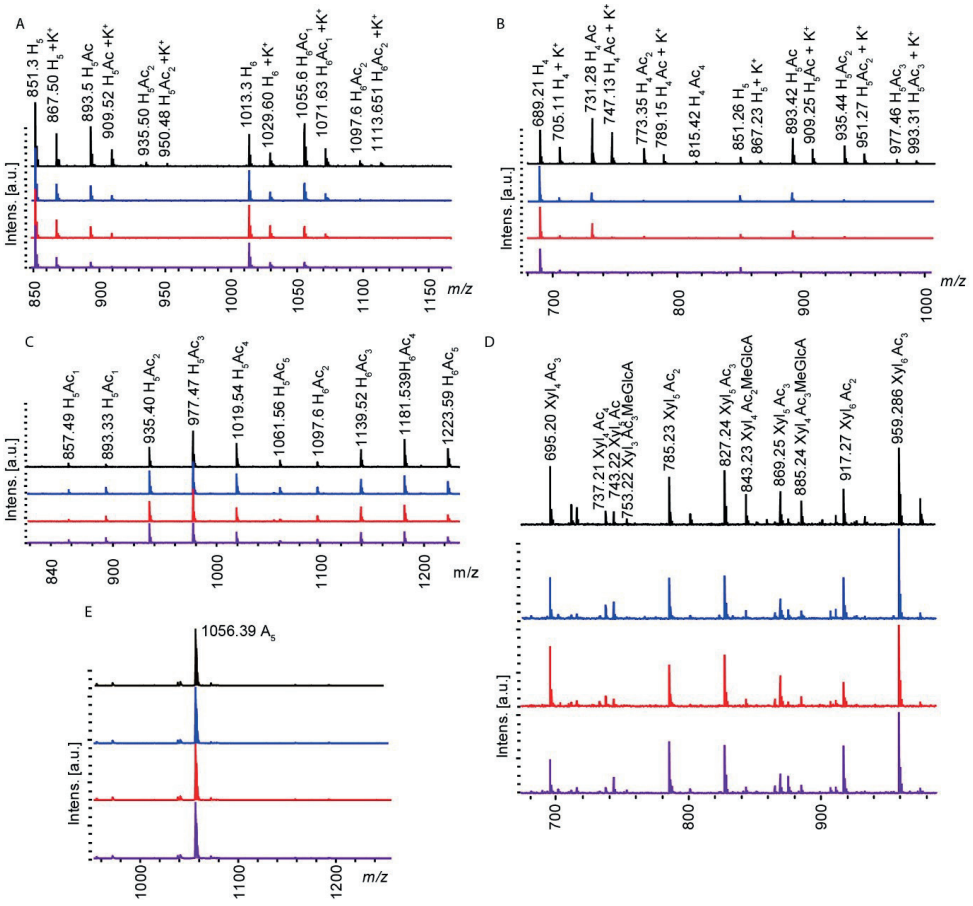


Figure S1. A: Konjac glucomannan deacetylated with RiCE2 (in blue), RiCEX (in red), and both enzymes combined (in purple). B: Chemically acetylated Konjac glucomannan deacetylated with RiCE2 (in blue), RiCEX (in red) and both enzymes combined (in purple). C: Acetylated cellulose oligosaccharides treated with RiCE2 (in blue), RiCEX (in red) and both enzymes combined (in purple), the spectra show no signs of enzymatic activity. D: Birch xylan oligosaccharides treated with RiCE2 (in blue), RiCEX (in red) and both enzymes combined (in purple) showing no apparent activity on the xylo-oligosaccharides. E: Chitopentaose (penta-N-acetylchitopentaose) (Megazyme, Ireland) treated with RiCE2 (in blue), RiCEX (in red) and both enzymes combined (in purple), showing no signs of activity. H-hexose, X-xylose, Ac-acetylations, Me-methylations, GlcA-glucuronic acid. All masses represent sodium adducts unless marked with K+, unlabeled peaks represent background signals from the sample matrix.

HMM building

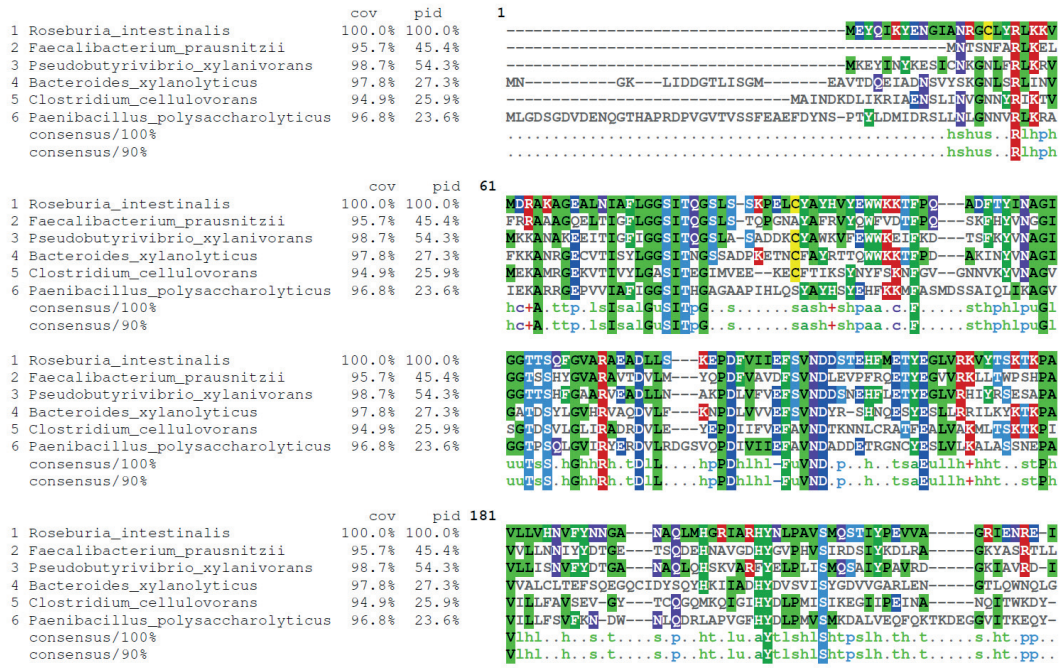
In order to build HMMs, the phmmer tool (63) on the European Bioinformatics Insititute website (www.ebi.ac.uk) was used to search for homologous sequences in the reference proteomes database. Homologous sequences were used to generate an HMM (64) which was then used to search the Uniprotkb (38) database for homologous sequences. Results of the HMM search in the Uniprotkb are described in the main text.

The phmmer search using the 126 aa in the N terminal domain of *Ri*CE2 yielded 187 hits using default settings. 142 of these contained the homologous sequence upstream of a C terminal Lipase GDSL2 hydrolase domain (including the *Ri*CE2 N terminal domain), and five upstream of a C terminal Lipase GDSL domain. 29 hits came from proteins with unknown domain architecture, and the remaining 11 from various multidomain proteins. All of the hits came from bacteria, with the vast majority originating from Bacteroidetes (70 hits), Firmicutes (54 hits), and Proteobacteria (30 hits). When using a search with higher stringency in phmmer (significance e value for sequence and hit 10^{-18}) yielded 30 sequences (28 Firmicutes, 2 Actinobacteria), 14 of which appeared upstream of Lipase_GDSL2 domains, 3 upstream of Lipase_GDSL domains. The remaining 14 sequences were part of proteins of unknown architecture. These 30 sequences were used to produce a HMM, which was used to search the Uniprotkb databse. Results of this search are described in the main text.

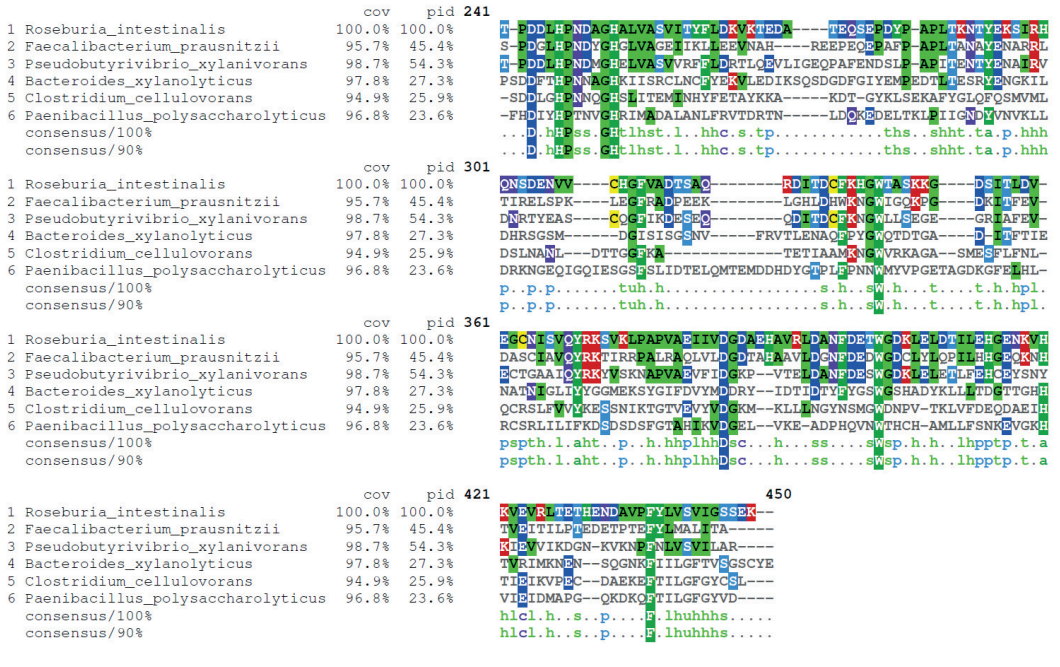
For *Ri*CEX, the search for sequences similar to the CBM35 domain using a protein vs protein database tool phmmer (default settings of 0.01 significance value for sequence, 0.03 for hit, search against reference proteomes database) resulted in 42 sequences with significant similarity of which 40 came from Firmicutes. 39 of those came from proteins that consisted of the C terminal domain followed by a Lipase_GDSL_2 domain, the remaining two were from proteins with unknown domain architecture. *S*ignificance e values for hit and sequence were adjusted to 1^{-19} to eliminate the lowest similarity hits (similarity <61%), leaving 34 hits used to build a HMM, which was then used to search the Uniprotkb (38). Results of this search are described in the main text.

A phmmer search in the reference proteomes using the whole sequence of *Ri*CEX with an e value for significant hit at 1^{-20} resulted in 46 sequences with matching domain architecture. These were used for HMM building. A search in the Uniprotkb database using this HMM yielded 381 hits with an e value $<1^{-50}$, 357 of these had a two domain GDSL2_Lipase + CBM 35 architecture. Most of the hits came from bacteria (355), of which the majority from Firmicutes (327). 8 results came from Bacteroidetes. The sequences of those 357 sequences were aligned to identify conserved aminoacids. All three catalytic residues (serine 41, aspartic acid 190, histidine 193) and the tryptophan 326 of the clamping domain that orients the substrate in the active site by aromatic

715 stacking were present in the 100% consensus sequence. Aligned sequences of five esterases
716 homologous to *Ri*CEX from known polysaccharide degraders as well as 100% and 90% consensus
717 sequences from the alignment of the 357 sequences are presented in Figure S2.



718



719

720 *Figure S2. Sequence alignment of RiCEX and five selected homologous proteins from representatives of various bacterial*
721 *taxa. From the 357 sequences found in UniprotKb database using an HMM of the RiCEX, sequences with 100% and 90%*
722 *consensus are displayed in the lower part of the alignment. All three catalytic residues (serine 41, aspartic acid 190,*
723 *histidine 193) as well as the tryptophan 326 of the CBM35 domain are present in the 100% consensus sequences.*

Acetyl migration

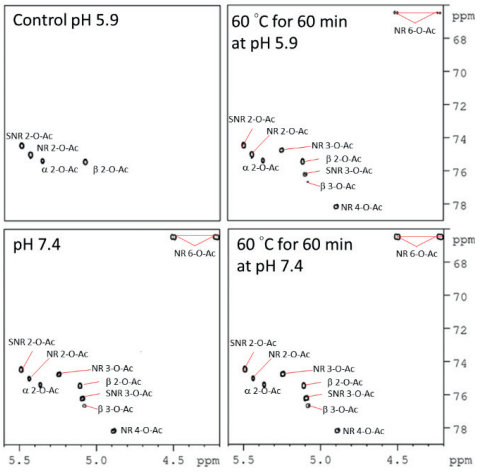


Figure S3. Acetyl migration on RiCEX transacetylated oligosaccharides. RiCEX acetylated mannotriose transacetylated in 20 mM sodium phosphate at pH 5.9 (top left panel), at pH 5.9 after exposure to 60°C for one hour (top right), after pH was increased to 7.4 by adding sodium phosphate (bottom left), and after exposure to both high temperature and pH 7.4. NR - nonreducing end mannose, SNR – Mannose prior to nonreducing end, α and β – anomeric configuration of the reducing end mannose.

For the purpose of acetyl esterase characterization, it was crucial to determine whether the results observed in deacetylation and transacetylation reactions were the products of enzymatic activity or acetyl migration. Elevated pH and temperature can induce acetyl migration on mannose and xylose (47-49). From experiments conducted on monosaccharides, it is known that the acetyl groups migrate at high pH. In both D-glucose and D-galactose acetylations appear to move in a 'clockwise' (2-*O*->3-*O*->4-*O*…) direction at pD >7 (65, 66), while at pD <3.0 D-galactose was shown to deacetylate without the acetyl group migrating (66). Presence of 6-*O*-acetylations on spruce GGM was previously described as a product of high temperature during processing in the extraction (19). Results presented here indicate that migration could only account for the 6-*O*-acetylations present on the non-reducing end of mannans, and that 6-*O*-acetylation in Norway spruce GGM is a natural occurrence.

To find out the extent of acetyl migrations on oligosaccharides, we exposed a solution of mannotriose transacetylated by *Ri*CEX to high temperature and pH. At pH 5.9 and 20°C *Ri*CEX exclusively transacetylates the 2-*O*-acetylations on the reducing end, non-reducing end and the intrachain Man*p* of mannotriose (Fig. S4 A). Exposure to pD 7.4 (Fig S4 C), 60°C for one hour pD 5.9 (Fig S4 B), or pD 7.4 at 60°C (Fig S4 D), caused a decrease in the signals for C2 bound acetylations an appearance of signals for 3-*O*-, 4-*O*- and 6-*O*-acetylations, the latter two only in the non-reducing end mannose. In 2-*O*-acetylated mannotriose, migration was limited to individual Man*p* residues ie.

749 not crossing glycosidic linkages, which is in agreement with previous findings about acetyl migration
750 in xylan (48). In an oligosaccharide, the acetylations on the non-reducing end migrates in the same
751 'clockwise' 2-*O*-→6-*O*- direction as described before on Gal*p* monosaccharides (47). The glycosidic
752 bond is preventing migration from 3-*O*-→ 6-*O*- and thus limit the migration from 2-*O*-→3-*O*- for the
753 reducing end and the intra-chain Man*p*.

754 This finding is especially important for hemicellulose biorefining with enzymatic
755 deacetylation steps – since conditions throughout the process can quickly change the acetyl
756 distribution. Migration induced by a short exposure to just 60 °C, at pH 5.9 as well as previously
757 published data on migration caused by heating (19) imply that the distribution of acetylations
758 present in steam exploded hemicellulose do not represent the distribution of acetylations present in
759 the hemicellulose *in vivo.*

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776 **References**

777 1. Cockburn DW & Koropatkin NM (2016) Polysaccharide Degradation by the
778 Intestinal Microbiota and Its Influence on Human Health and Disease. *Journal of*
779 *molecular biology* 428(16):3230-3252.

780 2. Flint HJ, Scott KP, Duncan SH, Louis P, & Forano E (2012) Microbial degradation of
781 complex carbohydrates in the gut. *Gut microbes* 3(4):289-306.

782 3. Tamanai-Shacoori Z*, et al.* (2017) Roseburia spp.: a marker of health? *Future*
783 *Microbiol.* 12(2):157-170.

784 4. Flint HJ, Duncan SH, & Louis P (2017) The impact of nutrition on intestinal bacterial
785 communities. *Current Opinion in Microbiology* 38:59-65.

786 5. Duncan SH & Flint HJ (2013) Probiotics and prebiotics and health in ageing
787 populations. *Maturitas* 75(1):44-50.

788 6. El Kaoutari A, Armougom F, Gordon JI, Raoult D, & Henrissat B (2013) The
789 abundance and variety of carbohydrate-active enzymes in the human gut
790 microbiota. *Nature reviews. Microbiology* 11(7):497-504.

791 7. Leth ML*, et al.* (2018) Differential bacterial capture and transport preferences
792 facilitate co-growth on dietary xylan in the human gut. *Nature Microbiology*
793 3(5):570-580.

794 8. Lloyd-Price J*, et al.* (2017) Strains, functions and dynamics in the expanded Human
795 Microbiome Project. *Nature* 550(7674):61-66.

796 9. La Rosa SL, Michalak, L., Leth, ML., Ejby MH., Pudlo NA., Glowacki R., Workman C.,
797 Pope PB., Arntzen, MØ., Martens, E., Hachem, MA., Westereng, B. (2018) The
798 Human Gut Firmicute Roseburia intestinalis is a Primary Degrader of Dietary β-
799 Mannans. *Manuscript*.

800 10. Singh S, Singh G, & Arya SK (2018) Mannans: An overview of properties and
801 application in food products. *International Journal of Biological Macromolecules*
802 119:79-95.

803 11. Tester RF & Al-Ghazzewi FH (2013) Mannans and health, with a special focus on
804 glucomannans. *Food Research International* 50(1):384-391.

805 12. Timell TE (1967) Recent progress in the chemistry of wood hemicelluloses. *Wood*
806 *Science and Technology* 1(1):45-70.

807 13. Lundqvist J*, et al.* (2002) Isolation and characterization of galactoglucomannan from
808 spruce (Picea abies). *Carbohydrate Polymers* 48(1):29-39.

14. Willför S*, et al.* (2003) Characterisation of water-soluble galactoglucomannans from Norway spruce wood and thermomechanical pulp. *Carbohydrate Polymers* 52(2):175-187.

15. Lundqvist J*, et al.* (2003) Characterization of galactoglucomannan extracted from spruce (Picea abies) by heat-fractionation at different conditions. *Carbohydrate Polymers* 51(2):203-211.

16. Simões J, Nunes FM, Domingues P, Coimbra MA, & Domingues MR (2012) Mass spectrometry characterization of an Aloe vera mannan presenting immunostimulatory activity. *Carbohydrate Polymers* 90(1):229-236.

17. Biely P (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* 30(6):1575-1588.

18. Cosgrove DJ (2014) Re-constructing our models of cellulose and primary cell wall assembly. *Current Opinion in Plant Biology* 22:122-131.

19. Xu C*, et al.* (2010) Acetylation and characterization of spruce (Picea abies) galactoglucomannans. *Carbohydrate Research* 345(6):810-816.

20. Williams MAK*, et al.* (2000) A Molecular Description of the Gelation Mechanism of Konjac Mannan. *Biomacromolecules* 1(3):440-450.

21. Willfor S, Sundberg K, Tenkanen M, & Holmbom B (2008) Spruce-derived mannans - A potential raw material for hydrocolloids and novel advanced natural materials. *Carbohydrate Polymers* 72(2):197-210.

22. Mikkonen KS*, et al.* (2009) Mannans as stabilizers of oil-in-water beverage emulsions. *LWT - Food Science and Technology* 42(4):849-855.

23. Du XZ, Li J, Chen J, & Li B (2012) Effect of degree of deacetylation on physicochemical and gelation properties of konjac glucomannan. *Food Research International* 46(1):270-278.

24. Mudgil D, Barak S, & Khatkar BS (2014) Guar gum: processing, properties and food applications-A Review. *Journal of Food Science and Technology-Mysore* 51(3):409-418.

25. Selig MJ, Adney WS, Himmel ME, & Decker SR (2009) The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes. *Cellulose* 16(4):711-722.

26. Mikkonen KS & Tenkanen M (2012) Sustainable food-packaging materials based on future biorefinery products: Xylans and mannans. *Trends in Food Science & Technology* 28(2):90-102.

27. Lombard V*, et al.* (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *The Biochemical journal* 432(3):437-444.

28. Biely P (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnol Adv.* 30.

29. Nakamura AM, Nascimento AS, & Polikarpov I (2017) Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation* 1(1):35-51.

30. Topakas E*, et al.* (2010) Carbohydrate esterases of family 2 are 6-O-deacetylases. *FEBS letters* 584(3):543-548.

31. Biely P*, et al.* (2014) Trichoderma reesei CE16 acetyl esterase and its role in enzymatic degradation of acetylated hemicellulose. *Biochim Biophys Acta.* 1840.

32. Montanier C*, et al.* (2009) The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol* 7(3):e71.

33. Tenkanen M, Puls J, Ratto M, & Viikari L (1993) Enzymatic deacetylation of galactoglucomannans. *Applied Microbiology and Biotechnology* 39(2):159-165.

34. Tenkanen M, Thornton J, & Viikari L (1995) An acetylglucomannan esterase of Aspergillu oryzae – purification, characterization and role in hydrolysis of O-acetyl galactoglucomannan. *Journal of Biotechnology* 42(3):197-206.

35. Bairoch A (2000) The ENZYME database in 2000. *Nucleic acids research* 28(1):304-305.

36. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42.

37. Finn RD*, et al.* (2017) InterPro in 2017—beyond protein family and domain annotations. *Nucleic acids research* 45(Database issue):D190-D199.

38. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic acids research* 45(D1):D158-D169.

39. Binda C*, et al.* (2018) Actinobacteria: A relevant minority for the maintenance of gut homeostasis. *Digestive and Liver Disease* 50(5):421-428.

40. Lansky S*, et al.* (2014) A unique octameric structure of Axe2, an intracellular acetyl-xylooligosaccharide esterase from Geobacillus stearothermophilus. *Acta Crystallographica Section D* 70(2):261-278.

41. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25(17):3389-3402.

42. Krissinel E & Henrick K (2005) Multiple alignment of protein structures in three dimensions. *Lect Notes Comput Sc* 3695:67-78.

43. Williams AH*, et al.* (2014) Visualization of a substrate-induced productive conformation of the catalytic triad of the Neisseria meningitidis peptidoglycan O-acetylesterase reveals mechanistic conservation in SGNH esterase family members. *Acta Crystallogr D* 70:2631-2639.

44. Montanier C*, et al.* (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences* 106(9):3065-3070.

45. Ashkenazy H*, et al.* (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* 44(W1):W344-W350.

46. Kremnicky L, Mastihuba V, & Cote GL (2004) Trichoderma reesei acetyl esterase catalyzes transesterification in water. *Journal of Molecular Catalysis B-Enzymatic* 30(5-6):229-239.

47. Roslund MU*, et al.* (2008) Acyl Group Migration and Cleavage in Selectively Protected β-d-Galactopyranosides as Studied by NMR Spectroscopy and Kinetic Calculations. *Journal of the American Chemical Society* 130(27):8769-8772.

48. Kabel MA, de Waard P, Schols HA, & Voragen AG (2003) Location of O-acetyl substituents in xylo-oligosaccharides obtained from hydrothermally treated Eucalyptus wood. *Carbohydr Res* 338(1):69-77.

49. Mastihubová M & Biely P (2004) Lipase-catalysed preparation of acetates of 4-nitrophenyl β-d-xylopyranoside and their use in kinetic studies of acetyl migration. *Carbohydrate Research* 339(7):1353-1360.

50. Arnling Bååth J*, et al.* (2018) Mannanase hydrolysis of spruce galactoglucomannan focusing on the influence of acetylation on enzymatic mannan degradation. *Biotechnology for Biofuels* 11(1):114.

51. Biely P*, et al.* (2013) Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1830(11):5075-5086.

52. Aslanidis C & de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research* 18(20):6069-6074.

53. Kabsch W (2010) XDS. *Acta Crystallographica Section D* 66(2):125-132.

54. Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D* 69(7):1204-1214.

55. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* 66(2):213-221.

56. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D* 65(6):582-601.

57. Terwilliger TC*, et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D* 64(1):61-69.

58. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40(4):658-674.

59. Vagin AA*, et al.* (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D* 60(12 Part 1):2184-2195.

60. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D* 66(4):486-501.

61. Cavanagh J, Fairbrother WJ, Palmer AG, & Skelton NJ (2007) *Protein NMR Spectroscopy (Second Edition)* (Academic Press, Burlington) pp vii-x.

62. Zhang H, Neal S, & Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. *Journal of biomolecular NMR* 25(3):173-195.

63. Potter SC*, et al.* (2018) HMMER web server: 2018 update. *Nucleic acids research* 46(W1):W200-W204.

64. Finn RD*, et al.* (2015) HMMER web server: 2015 update. *Nucleic acids research* 43(W1):W30-W38.

65. Brecker L, Mahut M, Schwarz A, & Nidetzky B (2009) In situ proton NMR study of acetyl and formyl group migration in mono-O-acyl D-glucose. *Magnetic resonance in chemistry : MRC* 47(4):328-332.

66. Roslund MU*, et al.* (2008) Acyl group migration and cleavage in selectively protected beta-d-galactopyranosides as studied by NMR spectroscopy and kinetic calculations. *J Am Chem Soc* 130(27):8769-8772.

Paper III

Wood-derived galactoglucomannan promotes butyrate-producing microbes in the swine gut microbiome

Leszek Michalak[1], John Christian Gaby[1], Leidy Lagos[2], Sabina Leanti La Rosa[1], Johannes Dröge[3] Margareth Øverland[2], Phillip B. Pope[1,2], Bjørge Westereng[1]

[1] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] Faculty of Biosciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[3] Department for Mathematical Sciences, Chalmers University of Technology, Göteborg

Introduction

The microbiome in the gastrointestinal tract (GIT) has a profound impact on host physiology. Bacteria present in the GIT affect the immune and endocrine systems, digestion, feed efficiency, and nourish the gut epithelium with the fermentation products. Bacteria produce key nutrients including vitamins (1), short chain fatty acids (SCFAs) such as acetate, propionate and butyrate (2, 3) as well as signaling molecules that affect host tissues and gut homeostasis (4). Changes in microbiota have been associated with many diseases such as type 2 diabetes (5), inflammatory bowel disease and colorectal cancer (6). Research show that the gut microbiome is also implicated in obesity (7) neurological (8), psychological and immune function (9). The microbiome composition varies depending on the GIT site, health of the host and diet content and it shifts rapidly in response to incoming nutrients (10). Some of the commensal bacteria are capable of degrading mucin and causing inflammation in the absence of alternative carbon sources (11).

Microbiome function is also a critical factor that influences the efficiency of farming and animal health in the meat industry. In pigs, the process of switching from sow's milk to plant based feed (known as weaning), elicits a rapid, diet driven shift in the GIT microbiome (12). This puts the animals at high risk of swine dysentery (13) or infection by intestinal pathogens such as enterotoxigenic *E.coli* (ETEC) and *Salmonella enterica* (14). Weaning in pigs is also associated with a reduced feed intake, which in combination with the high mortality results in poor performance and large losses to farmers. Risks of infection during weaning and growth have been kept low by the inclusion of antibiotics, antibiotic growth promoters (AGPs) and 'animal protein factors' (APFs), which has been routine practice since the 1940s (15). Since the ban on AGP in farm animal feed was imposed by the European Commission (EC) in 2003 (Regulation No. 1831/2003), animal welfare has

declined while piglet mortality resulting from post-weaning anorexia and diarrhea has increased and piglet mortality in antibiotic-free farming can be as high as 17% (16). Besides a systemic effect, AGPs had pronounced effects on shaping the gut microbiome post weaning. Currently, functional feed ingredients are being pursued to counteract the loss of efficiency and improve the health of the animals (17) without contributing to the global increase in antibiotic resistant bacteria.

Prebiotics are feed ingredients that are resistant to the hosts digestive apparatus and selectively enrich groups of bacteria in the GIT that confer health benefits to the host (18). Prebiotics as well as more classical, non-antibiotic small molecule interventions aimed at improving the gut microbiome are being pursued as therapeutic targets (19). Galactoglucomannan (GGM) is the main hemicellulose in the secondary cell wall of Norway spruce (*Picea abies*), and is a complex polysaccharide consisting of a backbone of β-1,4-D-Man*p* and β-1,4-D-Glc*p* residues. This backbone is decorated with α-1,6-D-Gal*p* residues and acetylations. Approximately 35% of the D-Man*p* residues in GGM are esterified by 2-*O*- and 3-*O*-acetylations (20). The presence of acetylations is a defense mechanism against plant pathogens, and it renders the GGM difficult for microbes to utilize as a carbon source (21). Mannan degradation requires a large set of carbohydrate active enzymes (CAZymes), including endo-active mannanases (GH5, GH26, GH113), exo-active mannosidases (GH1, GH2,GH5), α-1,6 galactosidases (GH27) and acetyl esterases capable of deacetylating the axially oriented 2-*O*-acetylations (www.cazy.org) (22). In the context of a prebiotic intervention, the complexity of GGM is potentially a significant advantage. Since the oligosaccharides can only be utilized by specific taxa within Firmicutes (23) and *Bacteroidetes* (24, 25), manno-oligosaccharides are a finely targeted prebiotic. Bacteroidetes along with Firmicutes and Proteobacteria (26), are the most prevalent bacterial phyla in the matured porcine gut with others present in varying abundance (27). Amongst the phyla present are multiple known and well-studied plant polysaccharide degraders (12, 28). Bacteria capable of degrading GGM will have an advantage over non-mannan degraders, leading to faster growth and GIT colonization, outcompeting pathogens and increased levels of production of the host-beneficial microbial products. In the human microbiomes, mannan degradation has been shown to be a conserved function, despite the generally low levels of mannans present in most diets (29).

This study has evaluated the potential prebiotic effects of GGM oligosaccharides on gut microbiome maturation in weaned piglets. Four separate cohorts of twelve post-weaning piglets were fed a pelleted feed that contained varying inclusion levels of Norway spruce GGM. A no-GGM control diet and three diets containing 1%, 2% and 4% GGM were fed semi-*ad libitum* to determine the GGM level necessary to elicit an effect, and to evaluate dose response. Fecal samples were collected before prebiotic administration (when piglets were assigned to pens), and subsequently at

days 7, 14, 21 and 27 during the feeding trial. The piglets were sacrificed and GIT content samples taken after 28 days. We conducted an integrated, multi-omics analysis in order to identify the effects of GGM supplementation on the gut microbiome structure, assess the GGM degradative capacity that is encoded within the inherent metagenomics assembled genomes (MAGs) and detect the active mannan-degrading populations. Finally, we looked into the effect of GGM on the immune system, with special focus on T lymphocytes, which are responsible for linking innate and adaptive immune responses against pathogens. The role of T cells was investigated along with their involvement in cytokine and chemokine networks.

Microbiome population structure of the piglet GIT microbiome varies over space and time:

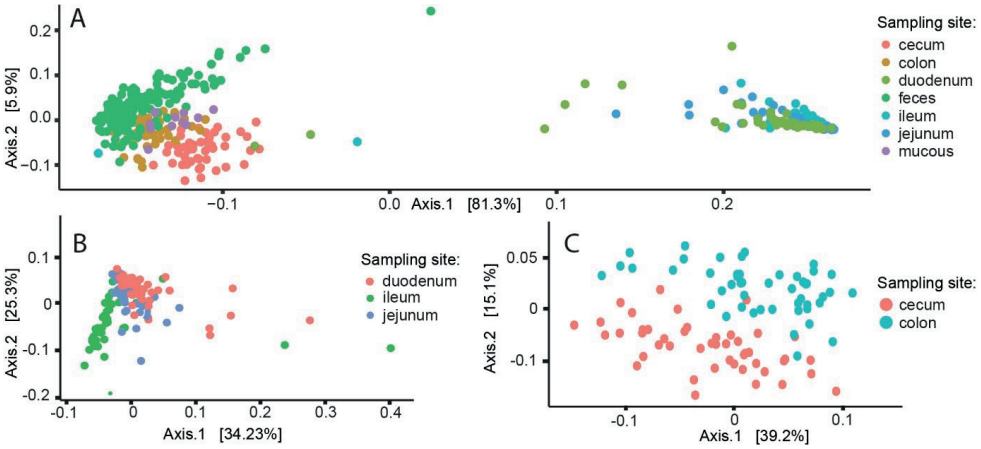


Figure 1. Microbiota composition of samples from different sites in the GIT, presented as a multidimensional scaling (MDS) ordination of weighted UNIFRAC distances. A: Samples from all GIT compartments and feces, labelled according to GIT compartment of origin. The two clusters separate samples from upper GIT (duodenum, jejunum, ileum) in the right side of the plot, from fecal samples and lower GIT and (colon, cecum, colon mucus). B: MDS ordination of weighted UNIFRAC distances of a subset of upper GIT samples. C: MDS ordination of weighted UNIFRAC distances of a subset of lower GIT samples.

We used 16S rRNA gene amplicon analysis to investigate the spatial differences that influence the population structure of piglet GIT microbiome. The various gastrointestinal tract sites that were sampled at day 28 had distinguishable microbial communities (Fig. 1), which were grouped based on microbiome location and without apparent bias by pen, piglet gender, or litter (Fig. S1). In particular, microbiomes inherent to the cecum and colon had clearly distinct structure.

To investigate temporal changes, fecal samples were collected at one-week intervals to follow the development of GIT microbiome as it matured under the influence of the GGM-

supplemented diet. Previous work on the GIT microbiome development shows that the microbiome composition post weaning is strongly dependent on the presence of dietary fibers (30). We hypothesized that inclusion of GGM in the diet from the first meal post weaning would shift the final composition of the microbiome and aid colonization of the GIT by a polysaccharide degrading community. Inclusion of GGM in the feed changed the microbiome composition from the 14 day onwards (Fig. 2). Fecal microbiomes from the piglets fed the control diet were distinct from those recovered from the 4% inclusion level, whereas those recovered from the 1% and 2% GGM diets were observed to cluster inbetween (Fig. 2).

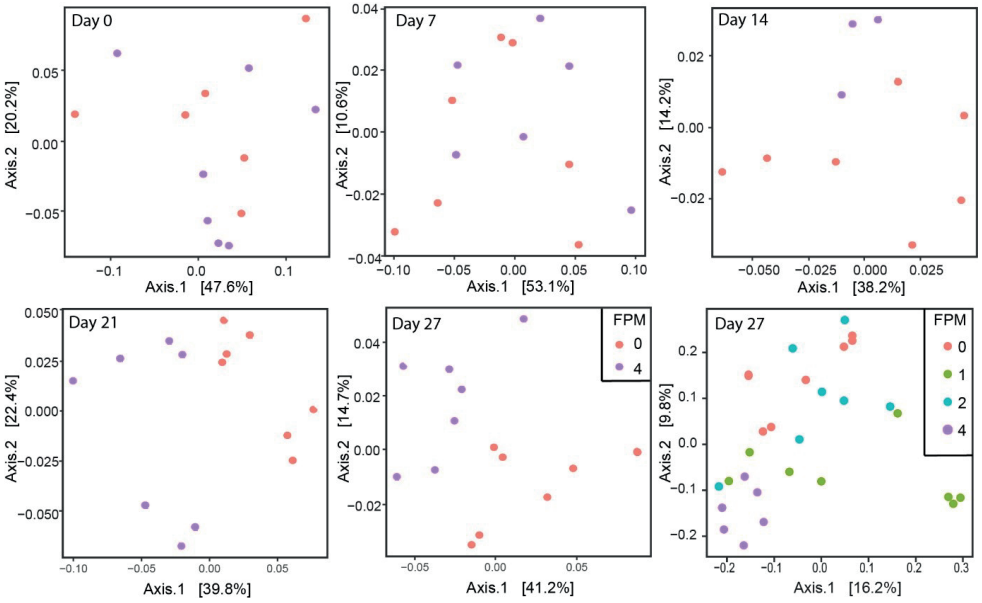
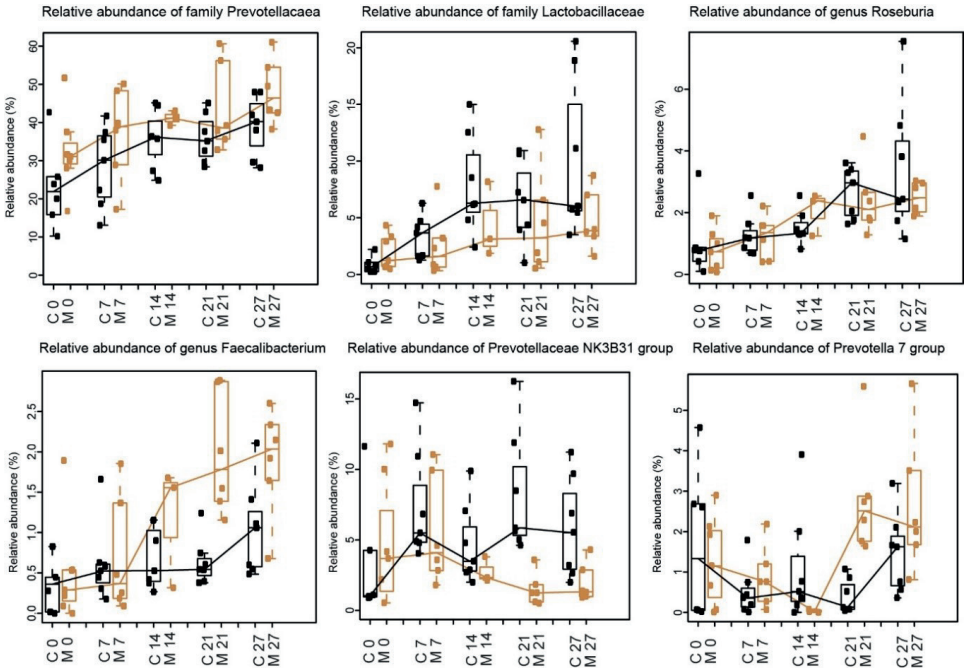


Figure 2. Ordination plots of Bray-Curtis distances between microbiome communities from fecal samples collected during the trial. From day 14 onwards, the control samples (magenta dots) separated from the 4% GGM samples (purple dots). The panel in the lower right corner shows the ordination of feces samples from all GGM inclusion levels. FPM – Feed percent GGM.



105

106 *Figure 3. Relative abundance of selected bacterial families and genera in the fecal samples. Families Prevotellaceae and*
107 *Lactobacillaceae became the most abundant in the mature microbiome. The increase was accelerated by inclusion of GGM*
108 *in diets. Genus Roseburia, a known butyrate producing polysaccharide degrader, has increased over time in both control*
109 *and GGM samples. Increase in relative abundance of genus Faecalibacterium was accelerated with inclusion of GGM. While*
110 *the family Prevotellaceae became the most dominant and was enhanced by GGM inclusion, the effect was elicited on*
111 *subpopulations of the family, with some Prevotella OTUs decreasing in abundance (such as the NK3B31 group) or being*
112 *slowed down in onset (the group Prevotella 7) in response to GGM. C – control group M – group fed 4% GGM in diet,*
113 *numbers in X-axis labels indicate sampling day.*

114 Looking specifically at microbiome taxa present in temporally collected fecal samples,
115 switching from milk to plant based diets resulted in a higher abundance of Prevotellaceae,
116 Lactobacillaceae, Ruminococaceae (genus *Roseburia)* and Lachnospiroceae (genus *Faecalibacterium*)
117 (Fig. 3). Prevotellaceae became the most dominant family in the fecal samples during the four weeks
118 feeding trial. Inclusion of GGM accelerated the increase in relative abundance of *Faecalibacterium,*
119 while at the same time, the onset of *Lactobacillus* OTUs was inhibited by inclusion of GGM.
120 Examination of the relative abundance of dominant families and genera in spatially distinct samples
121 from sites along the GIT length revealed that the shaping of the microbiome is localized to the lower
122 GIT (cecum and colon) (Fig. 4). Family Lactobacillaceae was the most dominant in the upper GIT
123 (jejunum, duodenum, ileum). In cecum and colon, like in the fecal samples at the final day,
124 Prevotellaceae was the dominant family. *Faecalibacterium* and *Roseburia* were also found in the
125 colon and cecum, with *Faecalibacterium* increasing in abundance with GGM inclusion, and *Roseburia*
126 decreasing in abundance in 4% GGM samples. Looking specifically at the microbiome differences
127 resulting from varying GGM inclusion levels, a total of 273 OTUs were listed as differentially

abundant between the control, and the GGM inclusion samples. In particular, relative abundance of many OTUs revealed a dose response effect in the colon and cecum, with abundance increasing or decreasing with varying GGM inclusion levels (Fig. 4). The most pronounced response was observed in the *Prevotella* genus, with one of the OTUs increasing from 4% to 12% between the control and 4% GGM inclusion in both colon and cecum.

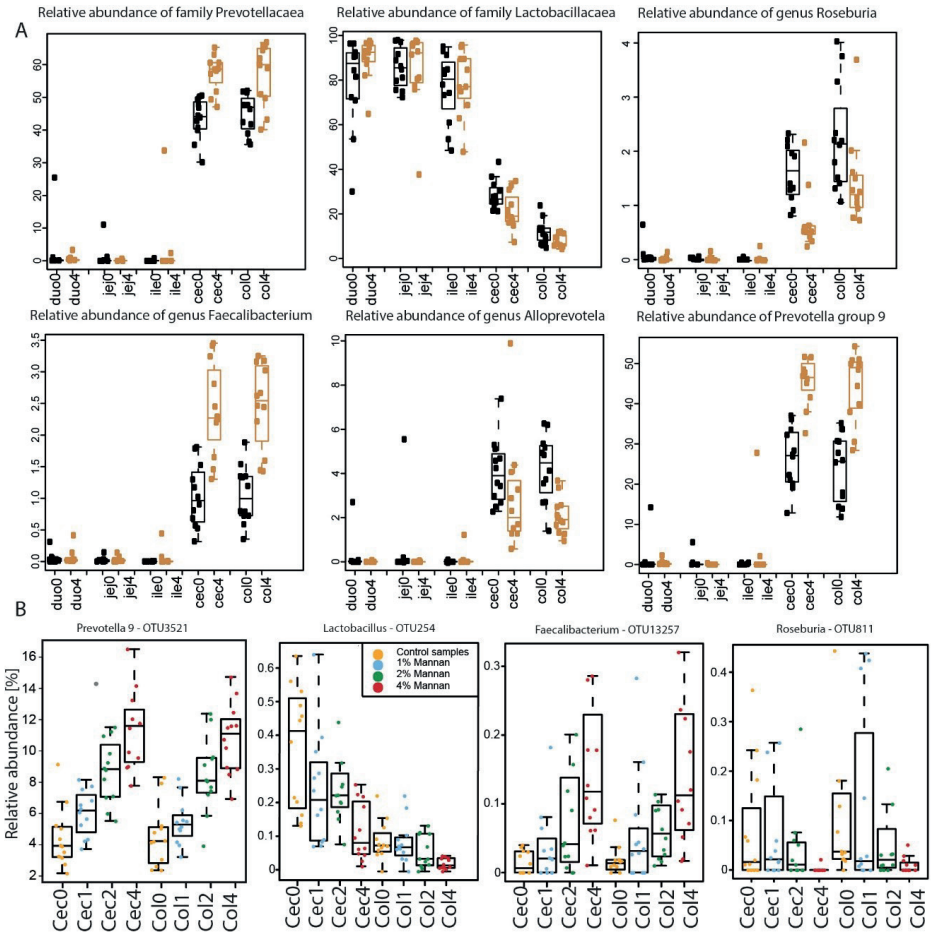
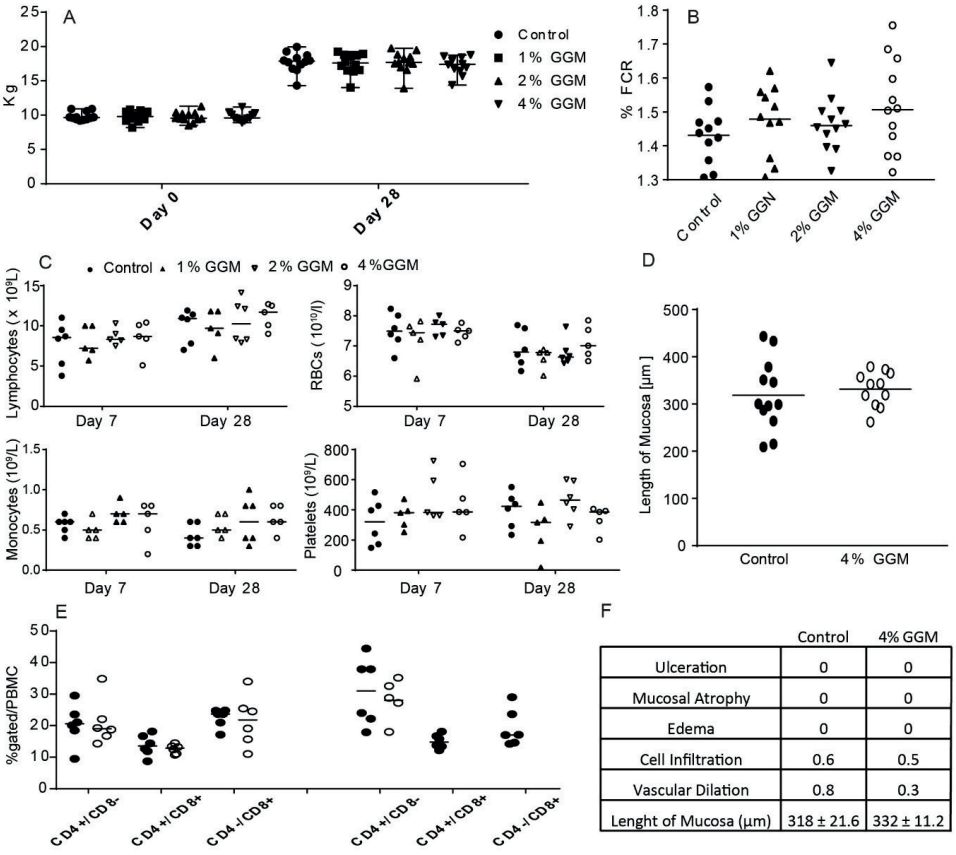


Figure 4. Changes in relative abundance of selected taxonomic groups in response to GGM inclusion. A: Relative abundance of dominant families and genera across the GIT compartments compared between the control group (in black) and 4% GGM inclusion samples (orange). As in the fecal samples, Prevotellaceae was found to be the most dominant family in the lower GIT. Lactobacillaceae dominates the duodenum, jejunum and ileum. In the lower GIT, inclusion of GGM decreased the relative abundance of Roseburia and Alloprevotela. The genus Faecalibacterium was enhanced by the inclusion of GGM. As in the fecal samples, the dominance of Prevotellaceae is the result of shifts of relative abundance within the family, with Prevotella group 9 being the most abundant and most enhanced by GGM inclusion. Abbreviations: duo – duodenum, jej – jejunum, ile – ileum, cec – cecum, col – colon. Numbers 0 next to compartment is the control, 4 is the 4% GGM inclusion. All x axes are %.B:Relative abundance of selected OTUs in the colon and cecum in response to various levels of GGM inclusion. Prevotella 9 OTU3521 has shown the highest relative abundance and dose of all OTUs. Lactobacillus abundance decreased with GGM inclusion. OTU13257 belonging to the genus Faecalibacterium was one of many Feacalibacterium OTUs increasing with the GGM inclusion. Despite being an active mannan degrader, Roseburia OTUs have decreased in abundance with increasing %GGM.

Effects of GGM on host physiology



	Control	4% GGM
Ulceration	0	0
Mucosal Atrophy	0	0
Edema	0	0
Cell Infiltration	0.6	0.5
Vascular Dilation	0.8	0.3
Lenght of Mucosa (µm)	318 ± 21.6	332 ± 11.2

Figure 5. A: Average weight of piglets at the start of the feeding trial and at sampling day shows no statistically significant difference in Dunnett's multiple comparisons test (p=0.05). B: Feed conversion rate for piglets in the four % GGM inclusion levels shows no statistically significant difference Dunnett's multiple comparisons test (p=0.05). C: Neither the numbers of lymphocytes, RBCs, monocytes nor platelets were significantly affected by the inclusion of GGM in diet (Tukey's multiple comparisons, p=0.05). D: Colon epithelium morphology: length of mucosal layer shows no significant difference between the controls and 4% GGM inclusion. E: Flow cytometry analysis of T cell populations in piglets from the control vs 4% GGM inclusion group shows no statistically significant shift in any of the subpopulations (two-way ANOVA, p=0.05). Y-axis - abundance of cells in percent of peripheral blood mononuclear cells F: Evaluation of colon morphology according to the scoring system described by Fabian et al. (30) indicates no significant difference between the control and 4% GGM inclusion level.

Regardless of inclusion level, GGM-diets were not observed to change the average weight of piglets at the conclusion of the experiment, or change the feed conversion rate in a statistically significant manner (Fig. 5, table S1). Hematology analysis was performed at day 7 and 28, and although the number of monocytes was slightly higher in the groups fed with 2% and 4% GGM, the difference was not statistically different (Fig. 5). The number of the other cellular components of blood and hemoglobin content was not significantly different between the groups at 7 (table S2) or 28 days (table S3). The effect of GGM on the subpopulations of immune cells and leukocytes isolated

from blood at days 7 and 28 post feeding was also analyzed by flow cytometry (Fig. 5). No difference in the number of conventional T cell (CD45+/CD3+/CD4+/CD8+), regulatory T cells or Natural Killers cells (data not shown) was noted. Serum immunity (IFNγ, GMCSF, IL-1A, IL1B, IL-1RA, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12, IL-18 and TNFα) of piglets fed the control diet and the 4% GGM diet was not significantly affected (table S4), there were no statistical difference in serum levels among the treatments at day 28 in this experiment. Regarding intestinal morphology, colon evaluation was performed using the scoring of Fabian *et al* (31), evaluating five criteria: ulceration, mucosal atrophy, edema, inflammatory cell infiltration and vascular dilation. The Fabian's score average for the analyzed samples were 0.3 for the control diet and 0.2 for the diet supplemented with 4% GGM, both groups are within normal ranges (Fig. 6). Piglets fed with control diet showed the lowest average intestinal mucosa height with 318 µm and 332 µm, respectively. Overall, there were no significant difference in the intestinal morphology due to supplementation of GGM on the diet. Finally, the presence of GGM had no impact on the total SCFA content of the digesta (Fig. 6) or its pH. Total levels of SCFAs have increased over time in the fecal samples. No effects on the relative content of butyrate (a *Faecalibacterium* metabolite) or propionate (a *Prevotella* metabolite*)* in the samples were observed.

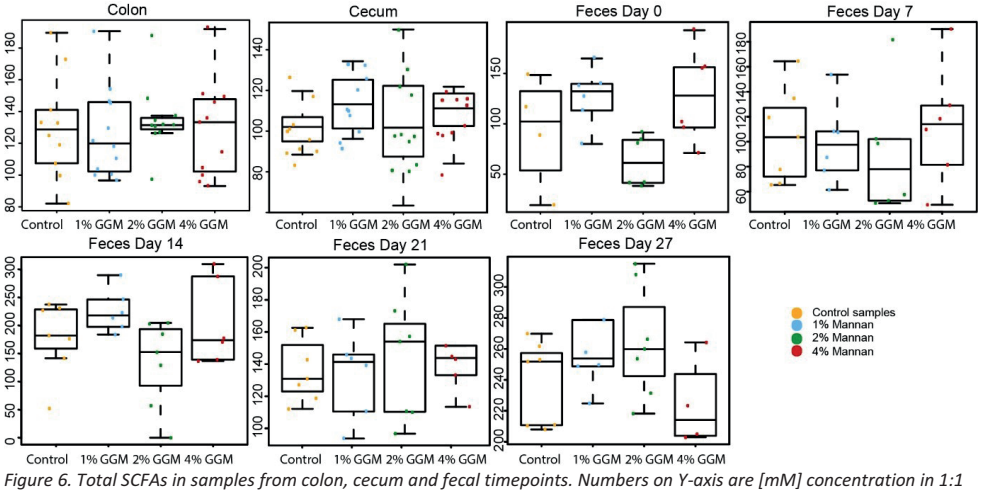


Figure 6. Total SCFAs in samples from colon, cecum and fecal timepoints. Numbers on Y-axis are [mM] concentration in 1:1 digesta: eluent extracts. Inclusion of GGM has not affected the SCFA content of digesta.

Meta-omic analysis of piglet GIT microbiomes fed 4% GGM and control diets

To expand our understanding of the key microbiota that were actively metabolizing the GGM diet, we first performed shotgun metagenomics sequencing on all 24 colon digesta samples (12 x control, 12 x 4 % GGM). Subsets of the control (n=4) and 4% GGM diet groups (n=4) were then

selected at random, and subjected to metaproteomics analysis. Metagenomic sequencing yielded between 5.1-26.4 gigabase pairs per sample, with a mean of 8.8 gigabase pairs (table S5). Assembly and genome binning resulted in 355 MAGs that had completeness higher than 73%, and a subset of 145 had >90% completeness (Fig. 7B). Analysis of the MAGs present in each sample further supported our 16S rRNA gene analysis, reiterating that the microbiomes from piglets fed the control and 4% GGM diets were distinct (Fig. 7C). Metaproteomic analysis was conducted on random colon digesta samples from which we identified a total of 8515 proteins with varying detection levels determined via LFQ scores (table S6). As expected, hierarchical clustering of all identified proteins from all samples showed that the differences in detection patterns are the result of GGM inclusion (Fig. 7). Metagenomic and metaproteomic analysis indicate that the presence of GGM alters the metabolic potential of the microbiome.

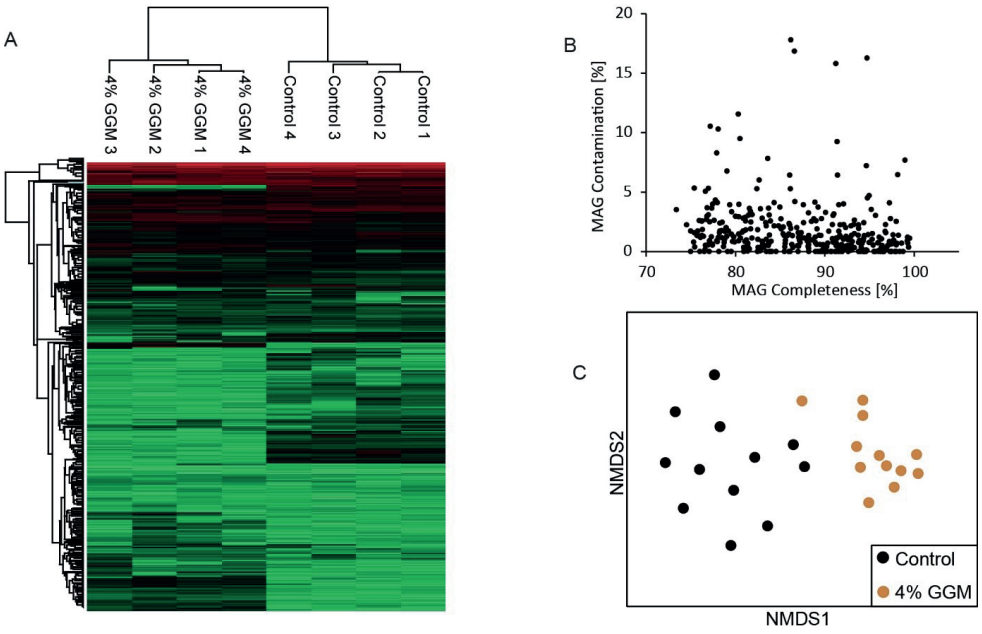


Figure 7. Metagenomic and metaproteomic analysis of the colon microbiome population of control and 4% GGM fed piglets A: Heatmap showing log transformed intensity of the 8515 proteins identified in the metaproteomic analysis. Based on patters of Label-free quantification (LFQ) detection patterns, the samples from control and 4% GGM fed piglets cluster together. B: Completeness % of assembled MAGs vs their contamination. C: Non-metric multidimensional scaling (NMDS) ordination plot of sample distances calculated with MASH based on the MAGs present in each sample.

The 35 MAGs with the highest mean of detected proteins in 4% GGM samples are listed in table S5. MAGs that showed differential protein detection levels between 4% GGM and control samples

were suspected to represent specific populations of mannan-degraders. In particular, MAGs 191, 196, 41 and 243 had an increase in the number of detected proteins in response to the 4 % GGM diet (table S6). Moreover, MAGs 191 and 196 were predicted to represent the genus *Prevotella*, whereas MAG 41 and MAG 243 were affiliated to the *Roseburia* (table S7) and *Faecalibacterium prausnitzii* (table s8), all four being known polysaccharide degraders. *Prevotella* MAGs attracted our attention since they represented the most abundant and the fastest responding genera in the 16S rRNA gene amplicon analysis, while *Roseburia* and *Faecalibacterium* have been long recognized as beneficial bacteria with a high impact on health (32, 33). Besides mannan degradation, the *F. prausnitzii* enzymes involved in butyrate production were also detected at a higher intensity in the 4% GGM inclusion (Table S9).

Mannan is utilized by populations affiliated to commensal bacteria that are associated with gut health

Annotation of MAG 41 revealed a mannan utilization cluster containing a similar set of mannan degrading enzymes to a previously reported locus from *Roseburia intestinalis* L1-82 (23). Specifically, the MAG 41 mannan operon encoded a phosphoglucomutase, two GH130 family enzymes, a GH26 with two carbohydrate binding modules and two carbohydrate esterases, which were all detected at higher levels in the 4% GGM samples compared to controls (Fig. 8, table S7). These proteomic detection patterns infer that the *Roseburia*-affiliated MAG 41 is actively degrading the prebiotic in the colon. The CBM27-GH26-CBM23 mannanase in the *Roseburia*-affiliated MAG 41 is the only extracellular protein in the locus, and is suspected to play a crucial role to enable this population to utilize GGM. The GH26 breaks down mannans in the cells immediate environment into oligosaccharides that are suitable for transport into the cell. The CBM27-GH26-CBM23 of MAG 41 is homologous to the GH26 in *R. intestinalis* L1-82 (48% identity over 87% of the sequence), and can be presumed to fulfill the same function in MAG 41.

A crucial step in the utilization of mannans as energy source is the deacetylation of 2-*O*-, 3-*O*- and 6-*O*- of Man*p*. The cluster in *R. intestinalis* L1-82, as well as the one in MAG 41 encode two carbohydrate esterases: a CE2 family esterase and a CEX – a mannan esterase which has not been assigned a CAZy family as yet. Family CE2 esterases characterized thus far have shown activity on mannan, and have shown to be active on the 6-*O*-acetylation. The *Ri*CE2 of *R. intestinalis* L1-82 has shown activity on 3-*O*-, 4-*O*- and 6-*O*-acetylations, and is mannan specific (ongoing work in our laboratory). It shares 63% identity (over 99% of sequence) with the CE2 found in MAG 41. CEX of MAG 41 shares 65% identity with the *Ri*CEX of *R. intestinalis* L1-82. The active site residues and the

Trp326 tryptophan, which orients the substrate manno-oligosaccharide in the active site are conserved, indicating that the MAG 41 CEX could also be a 2-*O*-acetyl specific esterase.

Besides the GH26 and esterases, a phosphoglucomutase, a multiple-sugar binding protein, two GH130 family enzymes: 4-*O*-β-D-mannosyl-D-glucose phosphorylase and a β-1,4-manno-oligosaccharide phosphorylase were also differentially detected. Surprisingly, the GH113, GH36 (β-mannosidase, α-galactosidase in *R. intestinalis* L1-82, respectively) were not detected in any of the samples. Taken together, these results confer that the *R. intestinalis* mannan utilization locus is a paradigm for mannan degradation by gut commensals found in mammalian digestive systems.

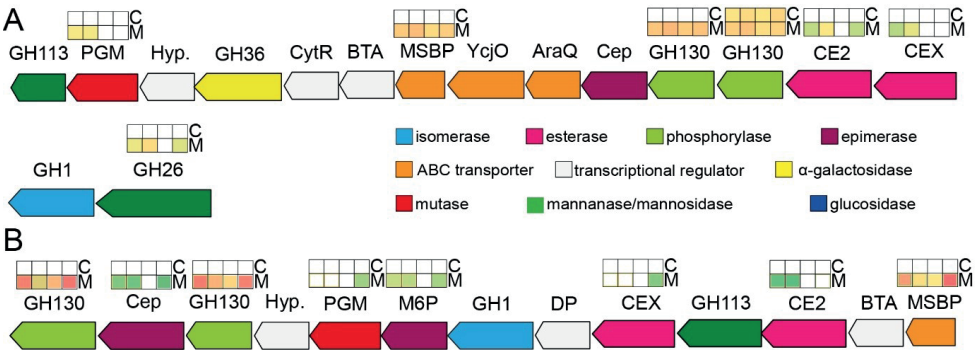


Figure 8. Mannan degradation apparatus of the R. intestinalis and F. prausnitzii representatives A: Enzymes in the mannan degradation cluster from the Roseburia intestinalis (MAG41). The heatmaps above detected enzymes show the LFQ detection levels for the four control (C) pigs and 4% GGM (M) pigs. B: Sequence of the mannan degradation cluster and relative detection heatmaps for the Faecalibacterium prausnitzii (MAG 243). Coloring in heatmaps is based on the LFQ values from tables S7 and S8, white – no detection, green to red coloring signifies low to high LFQ detection.

The *Faecalibacterium*–affiliated MAG 243 also encodes a mannan utilization locus, which showed no detectable expression in the control samples, while being highly detectable at 4% GGM inclusion (Fig. 8, table S8). Two GH130 manno-oligophosphorylases, a mannose 6-phosphate isomerase, phosphoglucomutase and two carbohydrate esterases represent highly detectable hits in the proteome corresponding to MAG 243. The presence of a CE2 esterase accompanied by an unclassified family esterase highly resembles the deacetylation machinery of *Roseburia*. The CEX of *F. prausnitzii* is 46% similar to *Ri*CEX of *Roseburia intestinalis* (over 95% of sequence), and contains the conserved active site residues and an aromatic stacking tryptophan, which are associated with 2-*O*- acetylation specificity. Proteins from the butyrate production pathway from MAG 243 were detected in both the controls and the 4% GGM, with higher levels in the latter (Table S9). This

suggests that MAG 243 is alike *F. prausnitzii* and appears to be a butyrate producer capable of utilizing various carbon sources.

MAGs 191 and 196 represent the genus *Prevotella*, which has been the most differentially abundant in the 16S rRNA gene analysis of the bacterial community. MAG 191 had the highest number of proteins detected in the 4% GGM samples, and a 1.84 -fold increase in mean detected proteins in 4% GGM samples over the controls (table S5). The fold increase in detection of proteins from MAG 196 was 1.41. MAG 191 expressed a Mannose-6-phosphate isomerase in response to the mannan in diet, while MAG 196 expressed a GH130 hydrolase and a SusC homologue. GH130 hydrolases are implicated in mannan degradation, although they do not appear to be capable of extracellular mannan breakdown of GGM for membrane transport.

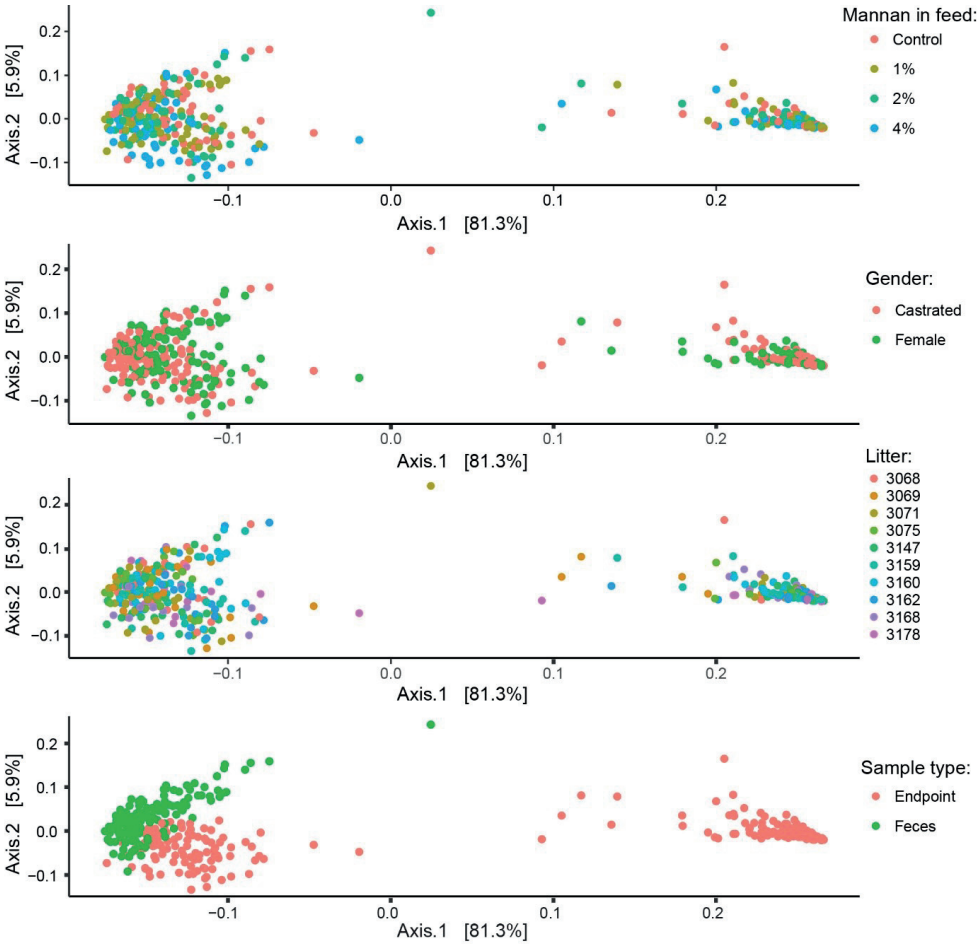
Conclusions

Observed effects of the GGM intervention were primarily on the composition of GIT microbiome in the lower digestive tract. Inclusion of GGM in the pig diet indeed increased the abundance of polysaccharide degrading OTUs, with the family *Prevotellaceae* responding the quickest and becoming the most dominant in the fecal samples from day 7 in 4% GGM inclusion fecal samples. In contrast, a coinciding decrease in the abundance of *Lactobacilli* was observed. *Prevotellaceae* abundance increased with GGM inclusion levels in a dose-dependent response. *Prevotellaceae*-affiliated *MAGs* demonstrated differential protein detection patterns, suggesting that their metabolism is effected by the addition of GGM. Despite the clear response of *Prevotellaceae*-affiliated populations, no clear mannan-degrading apparatus has thus far been detected in the corresponding MAGs, suggesting that these populations are using alternative, uncharacterized mechanisms to degrade GGM or that they are taking advantage of the metabolism from other microbial populations that are actively degrading the fiber.

Despite apparent low abundance in the 16s amplicon analysis, our functional meta-omic analyses have shown that *Roseburia* are primary mannan degraders in the colon. The increased relative abundance of *F. prausnitzii*-affiliated populations in the 16s amplicon analysis was further supported by the metaproteomics data that showed *F. prausnitzii* as an active mannan degrader, which further iterates the special importance of these populations. *F. prausnitzii* depletion is implicated in colorectal cancer, Crohn's disease, inflammatory bowel syndrome, ulcerative colitis and a number of other diseases (34, 35). A study on humans has named a reduction in *F. prausnitzii* and *R. hominis* a contributing factor to ulcerative colitis and Crohn's disease (36). Applications of *F. prausnitzii* as a therapeutic probiotic are currently pursued. The selective, dose dependent

303 enrichment of known beneficial butyrate-producing populations affiliated to *Roseburia* spp. and *F.*
304 *prausnitzii* in the lower GIT warrants further research on possible applications of GGM prebiotics for
305 humans.

306 **Supplementary**



307

308 *Figure S1. Weighted UNIFRAC distances MDS plot colored by A: mannan content in feed B: piglet gender C: litter D:Sample*
309 *type. The samples cluster together by sampling site (in Figure 1, main text), with no bias from any of the variables listed*
310 *here.*

311

312

313

314

315

316

Supplementary table 1. Growth performance of weaned piglets fed diets supplemented with GMOS showed no significant change between GGM inclusion groups in a Dunnett's multiple comparisons test (p value in the table).

	Control	1 % GGM	2% GGM	4% GGM	SEM	*P*
Initial BW (kg)	9.833	9.783	9.75	9.8	0.26	0.999
Final BW (kg)	17.74	17.51	17.73	17.26	0.26	0.95
Week 1:						
ADG (g)	138.1	147.6	145.2	125	44	0.99
ADFI (g)	145.5	154.9	163.8	150	14.8	0.98
F :G	1.606	1.474	2.067	1.436	0.64	0.99
Week 2:						
ADG (g)	198.8	188.1	198.8	191.7	43.51	0.99
ADFI (g)	336	326	339.2	331.4	21.83	0.99
F :G	2.24	1.15	2.51	2.28	0.95	0.53
Week 3:						
ADG (g)	369.6	401.1	429.2	388.5	30.7	0.87
ADFI (g)	527.1	541.5	574.8	535.7	24.51	0.97
F :G	1.44	1.38	1.34	1.46	0.09	0.99
Week 4:						
ADG (g)	422.3	367.4	366.9	359.9	26.9	0.06
ADFI (g)	581	571.9	585	577.5	23.7	0.99
F :G	1.41	1.62	1.61	1.62	0.11	0.16

BW, body weight ; ADG, average daily gain ; ADF1, average daily feed intake ; F :G, feed :gain ratio.

328 *Supplementary table 2. Hematology results at day 7. P-values are shown in parenthesis.*

	Control Mean± SD	1% GGM Mean± SD	2% GGM Mean± SD	4% GGM Mean± SD
White blood cells (x 10^9 /l)	15.1 ± 2.3	14.2 ± 2.3 (0.91)	16.4 ± 2 (0.76)	16.6 ± 3.3 (0.69)
Red blood cells (x 10^1 /l)	7.5 ± 0.5	7.2 ± 0.7 (0.66)	7.6 ± 0.3 (0.95)	7.4 ± 0.2 (0.99)
Hemoglobin (g/l)	119.2 ± 15.9	121.2 ± 6.9 (0.99)	122.4 ± 11.7 (0.98)	115 ± 24.9 (0.96)
Mean corpuscular volume (fl)	48.3 ± 4.4	50.8 ± 7.4 (0.86)	48.8 ± 4.7 (0.99)	46.4 ±7.9 (0.94)
Mean corpuscular hemoglobin content (g/l)	329 ± 8.4	334.8 ± 4.5 (0.64)	328.6 ± 7.2 (0.99)	330 ± 12.6 (0.99)
RDW (%)	23.9 ± 3.6	22.8 ± 3.3 (0.96)	24.2 ± 3.7 (0.99)	26.6 ± 6.4 (0.67)
Neutrophils (x 10^9 /l)	6.3 ± 1.3	5.4 ± 1.4 (0.74)	6.7 ± 2.1 (0.96)	7.1 ± 1.5 (0.8)
Lymphocytes (x 10^9 /l)	7.8 ± 2.5	8 ± 1.7 (0.99)	8.7 ± 0.9 (0.82)	8.5 ± 1.8 (0.87)
Monocytes (x 10^9 /l)	0.6 ± 0.1	0.5 ± 0.1 (0.83)	0.7 ± 0.1 (0.39)	0.6 ± 0.2 (0.97)

329

330 *Supplementary table 3: Hematology results at day 27. P-values are shown in parenthesis.*

	Control Mean± SD	1% GGM Mean± SD	2% GGM Mean± SD	4% GGM Mean± SD
White blood cells (x 10^9 /l)	17.4 ± 2.5	18.2 ± 5.1 (0.98)	19.2 ± 3.7 (0.82)	20.1 ± 4.8 (0.63)
Red blood cells (x 10^1 /l)	6.9 ± 0.6	6.6 ± 0.3 (0.61)	6.8 ± 0.4 (0.93)	7.1 ± 0.4 (0.83)
Hemoglobin (g/l)	110 ± 5.9	109.4 ± 6.5 (0.99)	107.8 ± 2.1 (0.89)	110.6 ± 8.4 (0.99)
Mean corpuscular volume (fl)	50.1 ± 1.6	53.3 ± 4.3 (0.44)	51 ± 2.1 (0.96)	48.7 ±5 .2 (0.87)
Mean corpuscular hemoglobin content (g/l)	317.3 ± 6.3	312.2 ± 4.8 (0.44)	312.7 ± 4.7 (0.48)	319.8 ± 4.7 (0.86)
RDW (%)	21.8 ± 2.3	20.5 ± 1.7 (0.84)	21.2 ± 2.1 (0.97)	23.9 ± 5 (0.62)
Neutrophils (x 10^9 /l)	6.5 ± 1.1	7.5 ± 3.7 (0.89)	7.6 ± 1.2 (0.86)	7.7 ± 3.9 (0.83)
Lymphocytes (x 10^9 /l)	10 ± 1.9	9.7 ± 2.1 (0.99)	10.5 ± 2.4 (0.94)	11.2 ± 1.4 (0.69)
Monocytes (x 10^9 /l)	0.4 ± 0.1	0.6 ± 0.1 (0.58)	0.6 ± 0.2 (0.27)	0.6 ± 0.1 (0.22)

331

Supplementary table 4. Serum immunity responses of weaned piglets fed diets control compared to supplemented with 4% GMOS showed no significant changes in a two way ANOVA.

	Control [ng/ml]	4% GGM [ng/ml]	SEM	*P*
IFNγ	3.374	3.824	0.2179	0.3664
IL1a	0.02333	0.021	0.1756	>0.9999
IL1b	0.195	0.2113	0.2341	>0.9999
IL1ra	0.1857	0.157	0.1884	>0.9999
IL4	0.2533	0.36	0.2015	>0.9999
IL6	0.07429	0.061	0.1884	>0.9999
IL8	0.1356	0.105	0.1756	>0.9999
IL10	0.1344	0.161	0.1756	>0.9999
IL12	0.8589	0.695	0.1756	0.9916
IL18	0.4678	0.494	0.1756	>0.9999
TNFα	0.09286	0.1411	0.1927	>0.9999

Supplementary table 5. Total reads and sequence length obtained in the whole metagenome sequencing of each sample. 0M and 4M in sample id designate 0% and 4% GGM content in diet.

sample	total reads per sample	total bases per sample	total Gbases per sample
01_0M	149875718	22481357700	22.48
02_0M	39328376	5899256400	5.89
03_0M	41490606	6223590900	6.22
04_0M	50328468	7549270200	7.54
05_0M	51520620	7728093000	7.72
06_0M	44035782	6605367300	6.60
07_0M	71111130	10666669500	10.66
08_0M	38528542	5779281300	5.78
09_0M	48341876	7251281400	7.25
10_0M	43270632	6490594800	6.50
11_0M	86191960	12928794000	12.93
12_0M	60690558	9103583700	9.10
		0M total:	108.71
13_4M	176308520	26446278000	26.45
14_4M	46804332	7020649800	7.02
15_4M	53833158	8074973700	8.07
16_4M	49041518	7356227700	7.35
17_4M	61654754	9248213100	9.24
18_4M	42908204	6436230600	6.43
19_4M	34987350	5248102500	5.25
20_4M	44732402	6709860300	6.71
21_4M	33761876	5064281400	5.06
22_4M	55012540	8251881000	8.25
23_4M	45067490	6760123500	6.76
24_4M	40208238	6031235700	6.03
		4M total:	102.65
		mean:	112.98
		median:	110.05
		max:	125.13
		min:	5.06

Supplementary table 6. Means and standard deviation of numbers of detected proteins mapped to MAGs. The table presents the 35 MAGs with the largest difference in proteins expressed in 4% GGM vs control.

MAG no.	Mean detected proteins in control samples:	St. Dev.	Mean detected proteins in 4% mannan:	St. Dev.	-fold increase in 4% vs control.
191	126.75	6.95	232.75	10.63	1.84
53	110.75	23.14	195.50	12.23	1.77
285	68.75	4.35	176.50	23.01	2.57
113	81.75	5.85	139.25	19.62	1.70
225	159.00	12.83	132.00	18.46	0.83
170	75.75	4.92	131.50	22.96	1.74
272	95.00	9.93	122.75	17.88	1.29
138	67.25	6.24	111.25	9.22	1.65
34	125.50	8.43	108.75	13.70	0.87
13	163.00	17.57	107.50	8.39	0.66
294	47.75	3.77	106.00	20.41	2.22
98	39.50	5.97	105.75	31.92	2.68
112	67.75	2.50	104.50	9.11	1.54
183	60.25	2.50	103.25	18.01	1.71
45	61.25	3.50	102.75	13.87	1.68
2	110.25	13.72	99.75	6.95	0.90
307	38.75	5.68	99.00	24.95	2.55
36	90.25	5.32	97.00	8.72	1.07
196	68.25	2.22	96.25	8.06	1.41
63	100.50	4.93	95.25	8.26	0.95
289	68.25	3.86	94.50	12.40	1.38
329	93.00	10.98	94.25	6.40	1.01
224	56.25	2.36	93.75	5.74	1.67
41	52.50	5.07	93.50	30.99	1.78
137	76.25	8.42	91.00	10.23	1.19
48	21.75	0.96	89.50	16.50	4.11
84	90.25	34.32	87.75	11.93	0.97
198	77.25	9.32	87.50	11.12	1.13
292	97.25	22.97	86.50	10.75	0.89
298	62.00	3.65	84.25	9.18	1.36
64	61.25	2.36	83.25	8.22	1.36
150	12.50	1.29	82.75	9.22	6.62
271	65.75	5.12	79.50	7.05	1.21
243	31.50	2.08	76.50	12.79	2.43

Supplementary Table 7. LFQ intensities of proteins from the Roseburia intestinalis MAG 041 mannan degradation locus. Coloring in the table is the base for heatmaps in Fig. 8 white – no detection, green to red coloring signifies low to high LFQ detection.

MAG 041 (*Roseburia intestinalis*)								
LFQ intensities								
Ctrl Pig 1	Ctrl Pig 2	Ctrl Pig 3	Ctrl Pig 4	4% Pig 1	4% Pig 2	4% Pig 3	4% Pig 4	Protein
								GH113
				26.46	26.31			Phosphoglucomutase
								hypothetical protein
								GH36
								HTH-type transcriptional repressor CytR
								Bifunctional transcriptional activator
				28.85	29.68	27.69	28.59	Multiple sugar-binding protein
								Inner membrane ABC transporter permease protein YcjO
								L-arabinose transport system permease protein AraQ
								Cellobiose 2-epimerase
27.03	27.71	27.06	28.10	28.66	29.36	27.47	28.34	GH130
				29.13	29.98	28.79	29.33	GH130
				24.33	26.89		24.17	CE2
				23.89	25.74			CEX
								GH1
				26.04	27.64		25.44	CBM27-GH26-CBM23

Supplementary Table 8. LFQ intensities of proteins from the Faecalibacterium prausnitzii MAG 243 mannan degradation locus. Coloring in the table is the base for heatmaps in Fig. 8 white – no detection, green to red coloring signifies low to high LFQ detection.

MAG 243 (*Faecalibacterium praustnitzii*) mannan degradation locus.								
LFQ intensities								
Ctrl Pig 1	Ctrl Pig 2	Ctrl Pig 3	Ctrl Pig 4	4% Pig 1	4% Pig 2	4% Pig 3	4% Pig 4	Protein
				29.55	27.73	28.58	29.76	GH130
				23.68	22.97		22.91	Cellobiose 2-epimerase
				29.62	28.65	28.09	29.52	GH130
								Hypothetical protein
							24.69	Phosphoglucomutase
				26.07	25.57		24.97	Putative mannose-6-phosphate isomerase YvyI
								GH1
								Dipeptidase
							23.80	CEX
								GH113
				22.35	22.09			CE2
								Catabolite control protein A
				29.02	27.97	27.84	29.93	Multiple sugar-binding protein

Supplementary Table 9. LFQ intensities of proteins from the Faecalibacterium prausnitzii MAG 243 butyrate production locus.

MAG 243 (*Faecalibacterium praustnitzii*) butyryl CoA:Acetate pathway.								
LFQ intensities								
Ctrl Pig 1	Ctrl Pig 2	Ctrl Pig 3	Ctrl Pig 4	4% Pig 1	4% Pig 2	4% Pig 3	4% Pig 4	Protein
29.39	29.68	30.06	30.18	31.20	30.80	30.49	31.27	Acryloyl-CoA reductase electron transfer subunit beta
29.55	29.29	29.94	29.98	30.92	30.25	29.47	30.61	Acryloyl-CoA reductase electron transfer subunit gamma
28.70	29.46	29.92	29.71	30.99	30.73	29.73	30.50	Acyl-CoA dehydrogenase, short-chain specific
29.19	30.21	30.11	30.00	31.66	30.58	30.79	31.55	3-hydroxybutyryl-CoA dehydrogenase
25.15	24.95	27.21	26.65	28.45	29.10	27.01	28.03	Short-chain-enoyl-CoA hydratase
29.80	30.06	30.69	30.43	31.94	31.66	30.71	31.66	Acetyl-CoA acetyltransferase

Materials and Methods:

Animals, diets and experimental design

A total of forty eight cross bred piglets (Landrace x Yorkshire), 24 male and 24 female, with an average initial body weight (BW) of 9,8 ± 0,5 kg, weaned at 28 days of age were sorted by litter, sex and weight and randomly divided into 12 groups of four animals each. However, pigs were housed individually during meal time. The animals were housed in an environmentally controlled facility with plastic flooring and a mechanical ventilation system. The temperature of the room was maintained at 22º C. Animal care protocols and experimental procedures were approved by the Norwegian Animal Research Authority, approval no. 17/9496, FOTS ID 11314 and treated according to institutional guidelines.

Piglets were fed cereal-based diets containing increasing levels of GGM in the diets (1, 2 and 4%). Diets were formulated to meet or exceed the requirements for indispensable amino acids and all other nutrients (NRC, 2012). The composition of diets is listed in Table 1.

Pigs were fed semi-ad libitum twice a day at a feeding level equal to about 5% of body weight. To evaluate growth performance, the BW of each pig was recorded at the beginning and once a week and feed consumption was recorded on an individual pig basis during the experiment to calculate weight gain and feed intake. After each meal, feed leftovers were registered and dried, and subtracted from the total feed intake.

Production of GGM

GGM oligosaccharides for the feeding trial were produced from Norway spruce chips milled with a hammer mill to <2 mm size. Wood chips were then steam exploded at NMBUs Volebekk pilot scale plant. Steam explosion was conducted in batches of approximately 6kg, at 200° C, 14.5 bar pressure, with a residence time of 10 minutes. The pH in the collected biomass slurry after steam explosion was ~3.7, which corresponds to a combined severity factor R'_0=1.706 for the process. Steam exploded wood was collected in 50 L plastic buckets that were topped up with hot (~70° C) water. The slurry was transferred to a 60L Speidel cider press and the liquid fraction was pressed out. Milled wood was collected, soaked in hot water again, and pressed for the second time. The liquid fraction was collected and recirculated through a bag filter 50µm pore WE50P2VWR (Allied filter systems, England) partly filled with the wood particles as a filter aid. Once free of floating wood particles, the liquid fraction of hemicellulose was filtered through a 5-kDa spiral wound Polysulphone/polyethersulphone ultrafiltration membrane, GR99PE polyester (Alfa Laval, Denmark) using a GEA pilot scale filtration system Model L (GEA, Denmark). The fraction retained by the

membrane was concentrated by nanofiltration using a TriSep XN 45, which had a higher efficiency for permeating water. Concentration was followed by vacuum evaporation. Concentrated samples were freeze-dried and homogenized with a grain mill.

Fecal scoring

During the experiment, fecal consistency was assessed using a scoring system developed by Pedersen and Toft (37) to improve and help standardize current protocols for clinical characterization of fecal consistency. The scoring was based on the following 4 consistency categories: score 1 = firm and shaped, score 2 = soft and shaped, score 3 = loose and score 4 = watery. Samples with score 3 or 4 are considered diarrheic. Daily fecal scores for each pen were recorded throughout the trial.

pH measurements

The pH of digesta samples from duodenum, jejunum, ileum, cecum and colon were measured immediately after slaughter. Samples were placed in universal containers and pH measurements made using a pH meter.

Blood sampling and flow cytometry

Blood samples were collected from the same pigs, three piglets per diet at 0, 7 and 27 feeding days. The blood samples were taken 1-2 hours post-prandial by venapuncture in the jugular vein while pigs are kept on their backs. Non-heparinised and K_3EDTA vacuum tubes (Beckman Dickson Vacutainer System) were used to recollect serum and whole blood. Serum was isolated immediately by centrifugation at 1,500 x *g* at 4°C for 15min. Serum samples were split in PCR-tubes (200 µl) and stored at -80°C until analysis. For hematological and clinical chemistry analyses, 6 piglets per diet were included. Hematological analyses were performed with an Advia® 2120 Hematology System using Advia 2120 MultiSpecies System Software and clinical chemistry analyses were performed with Advia 1800 Chemistry System (both from Siemens AG Healthcare Sector).

For flow cytometry analysis, whole blood were diluted 1:1 in RPMI 1640 and kept on ice until single cells isolation. For the isolation of peripheral blood mononuclear cells (PBMCs) blood were purified by centrifugation in a ficoll gradient (Kreuzer et al. 2012). Briefly, isolated PBMCs were incubated with Fixable Yellow Dead Cell Stain Kit (Life Technologies, Thermo Fisher Scientific Inc.) followed by primary monoclonal antibodies (mAbs), brief incubation with 30% normal pig serum to block Fc-receptors, and finally fluorescence-labeled goat-anti-mouse secondary antibodies. To detect the intracellular CD3 epitope, surface-labeled cells were permeabilized with Intracellular Fixation and Permeabilization Buffer Set (eBioscience, Affymetrix Inc.) according to the manufacturer's

441 instructions. Labeled cells were analyzed in a Gallios flow cytometer and data were processed using
442 Kaluza 1.5 software (both Beckman Coulter, Inc.). Cell gates were designed to select for single and
443 viable mononuclear cells. Defined markers were used to identify the different immune
444 subpopulations. For monocytes, antibodies against CD45, CD3, CD14, CD163 and MHCII were used.
445 To determinate regulatory T cells the following antibodies were used: CD45, CD3, TCR γ/δ, CD4, CD8,
446 FOXp3 and CD25. To identified T and NK cells we used CD45, CD8, NKp46, CD4, CD8, Ki67 and CD27.

447

448 **Analysis of Serum Cytokines: MULTIPLEX**

449 Expression of GMCSF, IFNG, IL-1A, IL1B, IL-1RA, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12, IL-18 and TNFα were
450 measured in serum samples using MILLIPLEX MAP Porcine Cytokine and Chemokine Magnetic Bead
451 Panel - Immunology Multiplex Assay (Merck Millipore) following the manufacturer instructions.

452

453 **Small Intestine Morphology**

454 The samples of small intestine were collected on day 0 and 28 for determination of intestinal
455 morphology and integrity. Intestinal morphological measurements included the following indices:
456 villus hight (VH), crypt depth (CD) and VH:CD. Mean values of VH, CD and their ratio were calculated.
457 Histology evaluation was performed by the Veterinary Histophalogy Center, VeHiCe, Chile.

458

459 **SCFA analysis**

460 Samples of digesta from duodenum, jejunum, ileum, cecum and colon of individual pigs were
461 collected for SCFA analysis. 250 mg or 250µL of sample, depending on the source site were mixed
462 1:1 with 4mM H_2SO_4, homogenized by shaking at room temperature for one hour, and centrifuged
463 at 12000xg for 10 minutes. The supernatant was collected with a syringe and filtered through a 0.22
464 µm pore syringe filter. Samples were stored at -20 °C and centrifuged at 12000xg before transferring
465 aliquots into HPLC vials for analysis. SCFA content was analyzed by HPLC using a REZEX ROA-Organic
466 Acid H+ (Phenomenex, Torrance, California, USA) 300x7.8mm ion exclusion column, isocratic elution
467 with 0.6 mL/min 4mM H_2SO_4 at 65 °C and UV detection at 210 nm.

468

469

470

Composition and chemical content of basal diets

Ingredients (%)		**Calculated content (g)**	
Wheat	51	Calculated contents Metabolizable energy, MJ/kg	
Barley	20	DM	884
Soybean meal	8	Crude protein	176,91
Oats	6	Crude fiber	34
Soy oil	4	Digestable crude protein	156.11
Fish meal	2	Starch	431.28
Potato protein	2	Crude fat	55
Corn gluten	1.78	Calcium	9.13
Calcium phosphate	1.13	Phosphorous	5.82
Limestone	1	Digestable phosphorous	3.73
*Vilomix	1	Sodium	2.51
Lysine	0.69	Chloride	4.59
Salt	0.39	Lysine	13.04
Sodium	0.36	Methionine + Cysteine	7.27
Threonine	0.25	Digestable Lys	12.06
Methionine	0.15	Digestable Met + Cys	6.57
Valine	0.14	Methionine	4.21
Tryptophan	0.075	Threonine	8.11
*Vilomix	0.14	Tryptophan	2.70
Salt	3.9	Valine	9.03

GGM added to the base diet consisted predominantly of mannan and a small part of xylan. Monosaccharide composition analysis showed: 0.9% rhamnose, 2.7% arabinose, 13.7% xylose, 58.9% mannose, 14.9% glucose and 9.4% galactose. The Man:Glc:Gal ratio in the GGM was 4:1:0.6, and the DA=0.36. Estimated DP range of oligosaccharides in this mixture was between 2- 10. GGM was assigned net energy value zero.

*VilomixMineral premix and vitamin mineral premix provided the following per kilogram of diet: vitamin A, 12 000 IU; vitamin D_3, 3200 IU; vitamin E, 80 IU; vitamin K_3, 25 mg; vitamin B_1, 25 mg; vitamin B_2, 65 mg; vitamin B_6, 5 mg; vitamin B_{12}, 0.5 mg; niacin, 45 mg; pantothenic acid, 20 mg; folic acid 15 mg; biotin 0.15 mg. §Fe,150 mg; Cu, 125 mg; Zn, 150 mg; Mn, 30 mg; I, 0.3 mg; Se, 0.3 mg.

Microbial Sampling

Fecal samples were collected from 6 piglets per experimental group (n=24) at days 0, 7, 14, 21, and 27 post-weaning. At the end of the trial, all piglets (n=48) were sacrificed, and samples were collected from the lumen of the duodenum, jejunum, ileum, cecum, and colon. Samples were obtained within 15 minutes of sacrifice and were flash frozen in liquid nitrogen and stored at -80 ºC until DNA extraction.

DNA Extraction

DNA was extracted with a MagAttract PowerMicrobiome DNA/RNA Kit (MO BIO Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer instructions, except for the bead beating step where we used a FastPrep-96 Homogenizer (MP Biomedicals LLC., Santa Ana, CA, USA) at maximum intensity for a total of 2 minutes in 4 pulses of 30s with a 5 minute cooling period between each pulse. A KingFisher Flex DNA extraction robot was used for the automated steps of the protocol. The extracted nucleic acids were quantified with a Qubit Fluorimeter and the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -80 ºC.

16S Amplicon Sequencing and Analysis

16S amplicon sequence data was obtained for all fecal and intestinal samples. The V3-V4 region of the 16S rRNA gene was PCR amplified using the primers Pro341F (5'-CCT ACG GGN BGC ASC AG-3') and Pro805R (5'-GAC TAC NVG GGT ATC TAA TCC-3'), to which the MiSeq adaptors were additionally incorporated on the 5' ends (38). The 25 µL PCR reactions consisted of 1X iProof High-Fidelity Master Mix (Biorad, Hercules, CA, USA), 0.25 µM primers, and 5 ng template DNA. PCR thermal cycling began with a hot start step at 98 °C for 180 s and was followed by 25 cycles of 98 °C denaturation for 30 s, 55 °C annealing for 30 s, and 72 °C extension for 30 s, followed by a final, 300 s extension step at 72 °C. Amplicons were individually purified with AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA) and indexed with the Nextera XT Index Kit v2 (Illumina, San Diego, CA, USA) according to the Illumina protocol for 16S metagenomic sequencing library preparation. Next, equal volumes from each indexing reaction were pooled together, and the pool was purified with AMPure XP beads. The purified amplicon pool was then quantified with a Qubit Fluorimeter, diluted, mixed with 15% PhiX Control v3 (Illumina), and denatured according to the aforementioned Illumina protocol. The denatured library was sequenced on the Illumina MiSeq platform using the MiSeq

Reagent Kit v3 (600 cycle). Data were output from the sequencer as demultiplexed FASTQ format files.

Processing of the data was done with a combination of standalone programs, QIIME (39) MOTHUR (40) and the R package Phyloseq (41) on an ASUS laptop with an Intel Core i7-6700HQ, 2.60GHz quad core processor and 16 GB RAM running Biolinux 8 (42) with the Ubuntu 14.04 LTS operating system. To process the data, the paired end reads for each sample were merged with PEAR (43), specifying a minimum assembly length 400, maximum assembly length 575, minimum overlap 50, and no statistical test. Then, PRINSEQ (44) version 0.20.4 was used to filter low quality reads by requiring a minimum quality score of 10 for all bases and a minimum mean quality of 30. Primer sequences were trimmed in MOTHUR version 1.36.1, and chimeric sequences were identified and filtered out using QIIME version 1.9.1. Next, open reference $OTU_{0.97}$ clustering (45) was performed with VSEARCH (46) version 2.3.2 and the Silva database (47) release 128 as the taxonomy reference. Then, the QIIME core diversity analyses script was run. Differentially abundant taxa were identified in both cecum and colon for the control vs. 4% GGM samples using both the MetagenomeSeq fitZIG and DESeq2 negative binomial algorithms via the QIIME wrapper. The OTU table, phylogenetic tree, representative sequences, and taxonomy from QIIME were incorporated along with the sample metadata into a Phyloseq version 1.22.3 object in R for data exploration and visualization.

Whole Metagenome Sequencing and Analysis

Whole metagenome sequencing was performed at the Norwegian Sequencing Centre on 2 lanes of the Illumina HiSeq 4000 to generate 2 X 150 paired-end reads. TruSeq PCR-free libraries were prepared for 12 control and 12 GGM (4%) samples from the colon. All 24 samples were run in both lanes to eliminate the potential for lane-specific sequencing bias. FASTQ format files were received from the sequencing center, and prior to assembly, these were quality filtered with Trimmomatic (48) version 0.36 whereby TruSeq adaptor sequences were eliminated, sequences were required to have an average quality score above 20, leading and trailing bases with quality below 20 were removed, sequences with average quality score below 15 in a 4-base sliding window were trimmed, and the minimum read length was required to be 36 bases. Individual sample assembly was accomplished with metaSPAdes (49) version 3.11.1. MegaHIT (50) version 1.1.3 was used for co-assembly of all 24 samples together as well as co-assembly of the 12 control samples together and the 12 4% GGM samples together. MetaBAT (51)version 0.26.3 was used to bin the assemblies, and dRep (52) was used to dereplicate the multiple assembly and binning combinations to produce an optimal set of MAGs. MASH (53) version 2.0 used to compare the similarity of the 24

metagenomes by calculating pairwise Jaccard distances which were imported into R for NMDS ordination and visualization. Completeness and contamination was determined for each MAG using CheckM (54) version 1.0.7. Feature and functional annotation were completed with the Prokka pipeline (55) version 1.12, and the predicted protein sequences from all 355 MAGs were concatenated to create the metaproteomics reference database. The GH5 and GH26 family mannanases were identified among the MAGs using the HMM models from the dbCAN database (56) version 6 in conjunction with the search command in the HMMer package (www.hmmer.org) version 3.1b1.

Metaproteomics

Proteins were extracted from each sample in quadruplicate by the following method. An aliquot (1 g) of colon digesta from pigs fed either a control diet or a diet supplemented with 4% β-mannan was dissolved 1:1 (w/v) in 50 mM TrisHCl, pH 8.4.

Lysis was performed using a bead beating approach whereby glass beads (size ≤ 106 µm) were added to the colon digesta slurry and cells were disrupted in 3 x 60 second cycles using a FastPrep24 (MP Biomedicals, Santa Ana, CA, USA). Debris were removed by centrifugation at 16.600 x *g* for 20 minutes and proteins were precipitated overnight in 16% ice cold TCA. The next day, proteins were dissolved in 100 µL 50 mM TrisHCl, pH 8.4 and concentration was determined using the Bradford protein assay (Bradford Laboratories, USA) using bovine serum albumin as a standard. Fifty milligrams of protein was prepared in SDS sample buffer, separated by SDS-PAGE using an Any-kD Mini-PROTEAN gel (Bio-Rad Labaoratories, Hercules, CA, USA) and stained using Coomassie Brilliant Blue R250. The gel was cut into 6 slices and reduced, alkylated and digested as described previously (57). Prior to mass spectrometry, peptides were desalted using C_{18} ZipTips (Merck Millipore, Darmstadt, Germany) according to the manufacturer's instructions.

To analyze the proteins in the culture fluid, we used the previously described FASP procedure (58). In brief, denaturing, alkylation and digestion were accomplished by binding the sample to a filter and subsequently passing through 8M urea, 50 mM iodoacetamide and 2 µg trypsin in Tris-HCl, pH 7.8. Trypsination was performed overnight on the filter, and peptides were collected the next day by centrifugation as these would now pass through the filter. Peptides were desalted using C_{18} ZipTips as described above.

The peptides were analysed by nanoLC-MS/MS as described previously, using a Q-Exactive hybrid quadupole orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany) (57), and the acquired raw data was analysed using MaxQuant (59) version 1.4.1.2. Proteins were quantified using the MaxLFQ algorithm (60). The data was searched against a sample-specific database (602.947 protein

sequences), generated from the 355 metagenome assembled genomes (MAGs), and against the genome of *Sus scofa* (40.708 sequences). In addition, common contaminants such as human keratins, trypsin and bovine serum albumin were concatenated to the database as well as reversed sequences of all protein entries for estimation of false discovery rates. Protein N-terminal acetylation, oxidation of methionine, conversion of glutamine to pyro glutamic acid, and deamination of asparagine and glutamine were used as variable modifications, while carbamidomethylation of cysteine residues was used as a fixed modification. Trypsin was used as digestion enzyme and two missed cleavages were allowed. All identifications were filtered in order to achieve a protein false discovery rate (FDR) of 1% using the target-decoy strategy. For a protein to be considered valid, we required the protein to be both identified and quantified in both replicates, and in addition, we required at least one unique peptide per protein and at least two peptides in total for every protein. The output from MaxQuant was further explored in Persues version 1.6.0.7 where filtering, data transformation, and imputation were performed, and visualizations including heatmaps, hierarchical clustering, and volcano plots (for identification of differentially abundant proteins between the mannan and control groups) were made.

References:

1. Magnusdottir S, Ravcheev D, de Crecy-Lagard V, & Thiele I (2015) Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in Genetics* 6.
2. Haenen D*, et al.* (2013) A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine. *Journal of Nutrition* 143(3):274-283.
3. Sung J*, et al.* (2017) Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature communications* 8:15393.
4. Rooks MG & Garrett WS (2016) Gut microbiota, metabolites and host immunity. *Nature reviews. Immunology* 16(6):341-352.
5. Musso G, Gambino R, & Cassader M (2011) Interactions Between Gut Microbiota and Host Metabolism Predisposing to Obesity and Diabetes. *Annual Review of Medicine, Vol 62, 2011,* Annual Review of Medicine, ed Caskey CT), Vol 62, pp 361-380.
6. Sears CL & Garrett WS (2014) Microbes, Microbiota, and Colon Cancer. *Cell Host & Microbe* 15(3):317-328.
7. Turnbaugh PJ*, et al.* (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444(7122):1027-1031.
8. Sampson Timothy R & Mazmanian Sarkis K (2015) Control of Brain Development, Function, and Behavior by the Microbiome. *Cell Host & Microbe* 17(5):565-576.
9. Thaiss CA, Zmora N, Levy M, & Elinav E (2016) The microbiome and innate immunity. *Nature* 535(7610):65-74.
10. David LA*, et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505(7484):559-+.
11. Desai MS*, et al.* (A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* 167(5):1339-1353.e1321.
12. Martens EC*, et al.* (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS biology* 9(12):e1001221.
13. Burrough ER (2017) Swine Dysentery:Etiopathogenesis and Diagnosis of a Reemerging Disease. *Veterinary Pathology* 54(1):22-31.
14. Fairbrother JM, Nadeau E, & Gyles CL (2005) Escherichia coli in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Animal health research reviews / Conference of Research Workers in Animal Diseases* 6(1):17-39.
15. Stokstad EL, Jukes TH, & et al. (1949) The multiple nature of the animal protein factor. *The Journal of biological chemistry* 180(2):647-654.
16. Lallès J-P, Bosi P, Smidt H, & Stokes CR (2007) Nutritional management of gut health in pigs around weaning. *Proceedings of the Nutrition Society* 66(2):260-268.
17. de Lange CFM, Pluske J, Gong J, & Nyachoti CM (2010) Strategic use of feed ingredients and feed additives to stimulate gut health and development in young pigs. *Livestock Science* 134(1-3):124-134.
18. Gibson GR & Roberfroid MB (1995) DIETARY MODULATION OF THE HUMAN COLONIC MICROBIOTA - INTRODUCING THE CONCEPT OF PREBIOTICS. *Journal of Nutrition* 125(6):1401-1412.
19. Garber K (2015) Drugging the gut microbiome. *Nature biotechnology* 33(3):228-231.

20. Xu C*, et al.* (2010) Acetylation and characterization of spruce (Picea abies) galactoglucomannans. *Carbohydrate research* 345(6):810-816.
21. Bi R, Berglund J, Vilaplana F, McKee LS, & Henriksson G (2016) The degree of acetylation affects the microbial degradability of mannans. *Polymer Degradation and Stability* 133:36-46.
22. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* 42(Database issue):D490-495.
23. La Rosa SL, Michalak, L., Leth, ML., Ejby MH., Pudlo NA., Glowacki R., Workman C., Pope PB., Arntzen, MØ., Martens, E., Hachem, MA., Westereng, B. (2018) The Human Gut Firmicute Roseburia intestinalis is a Primary Degrader of Dietary β-Mannans. *Manuscript*.
24. Kawaguchi K*, et al.* (2014) The mannobiose-forming exo-mannanase involved in a new mannan catabolic pathway in Bacteroides fragilis. *Archives of microbiology* 196(1):17-23.
25. Cuskin F*, et al.* (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517(7533):165-169.
26. Holman DB, Brunelle BW, Trachsel J, & Allen HK (2017) Meta-analysis To Define a Core Microbiota in the Swine Gut. *mSystems* 2(3).
27. Xiao L*, et al.* (2016) A reference gene catalogue of the pig gut microbiome. *Nature Microbiology* 1:16161.
28. Tamura K*, et al.* (2017) Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides. *Cell reports* 21(2):417-430.
29. Lloyd-Price J*, et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550(7674):61-66.
30. Frese SA, Parker K, Calvert CC, & Mills DA (2015) Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome* 3(1):28.
31. Fabia R*, et al.* (1993) The effect of exogenous administration of Lactobacillus reuteri R2LC and oat fiber on acetic acid-induced colitis in the rat. *Scandinavian journal of gastroenterology* 28(2):155-162.
32. Miquel S*, et al.* (2013) Faecalibacterium prausnitzii and human intestinal health. *Current Opinion in Microbiology* 16(3):255-261.
33. Tamanai-Shacoori Z*, et al.* (2017) Roseburia spp.: a marker of health? *Future Microbiol.* 12(2):157-170.
34. Ferreira-Halder CV, Faria AVdS, & Andrade SS (2017) Action and function of Faecalibacterium prausnitzii in health and disease. *Best Practice & Research Clinical Gastroenterology* 31(6):643-648.
35. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, & Martinez-Medina M (2017) Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. *The Isme Journal* 11:841.
36. Machiels K*, et al.* (2014) A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 63(8):1275-1283.
37. Pedersen KS & Toft N (2011) Intra- and inter-observer agreement when using a descriptive classification scale for clinical assessment of faecal consistency in growing pigs. *Prev Vet Med* 98(4):288-291.

38. Takahashi S, Tomita J, Nishioka K, Hisada T, & Nishijima M (2014) Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PloS one* 9(8):e105592.
39. Caporaso JG*, et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature methods* 7(5):335-336.
40. Schloss PD*, et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537-7541.
41. McMurdie PJ & Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* 8(4):e61217.
42. Field D*, et al.* (2006) Open software for biologists: from famine to feast. *Nature biotechnology* 24(7):801-803.
43. Zhang J, Kobert K, Flouri T, & Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30(5):614-620.
44. Schmieder R & Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863-864.
45. Rideout JR*, et al.* (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2:e545.
46. Rognes T, Flouri T, Nichols B, Quince C, & Mahe F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
47. Yilmaz P*, et al.* (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic acids research* 42(Database issue):D643-648.
48. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
49. Nurk S, Meleshko D, Korobeynikov A, & Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome research* 27(5):824-834.
50. Li D, Liu CM, Luo R, Sadakane K, & Lam TW (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10):1674-1676.
51. Kang DD, Froula J, Egan R, & Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165.
52. Olm MR, Brown CT, Brooks B, & Banfield JF (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *Isme j* 11(12):2864-2868.
53. Ondov BD*, et al.* (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* 17(1):132.
54. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, & Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 25(7):1043-1055.
55. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069.
56. Huang L*, et al.* (2018) dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic acids research* 46(D1):D516-d521.

57. Arntzen MO, Karlskas IL, Skaugen M, Eijsink VG, & Mathiesen G (2015) Proteomic Investigation of the Response of Enterococcus faecalis V583 when Cultivated in Urine. *PloS one* 10(4):e0126694.
58. Wisniewski JR, Zougman A, Nagaraj N, & Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6(5):359-362.
59. Cox J & Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26(12):1367-1372.
60. Cox J*, et al.* (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* 13(9):2513-2526.

ISBN: 978-82-575-1565-2
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

101005 / ANDVORDGRAFISK.NO