A mechanistic overview of ruminal fibre digestion.

Adrian E. Naas and Phillip B. Pope

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

Contact email: phil.pope@nmbu.no

Abstract

Ruminants have co-evolved with symbiotic rumen microbiota, which readily convert ingested plant fibres into the nutrients they need to sustain their growth and maintenance. Fibre degradation within the rumen microbiome has been attributed to a limited number of cultivable representatives, which has restricted our ability to understand the different enzymatic machineries that exist. However, via a combination of culturing, meta-omics, bioinformatics, biochemistry and enzymology, we are beginning to expand our insight into the different fibre-digesting strategies that rumen microbiota employ. We discuss findings from studies on well-known Ruminococcus, Fibrobacter and Prevotella isolates, as well as those from poorly understood and as-yet uncultured Bacteroidetes lineages. Collectively, these approaches have revealed new mechanistic information related to the hydrolytic capacity of cellulosomes, free enzymes, outer membrane vesicles, polysaccharide utilization loci and large multi-modular enzymes, which are generating deeper insights into the intricate microbial networks that engage in ruminal fibre digestion.

Key words: Carbohydrate active enzymes, CAZymes, Cellulosomes, Polysaccharide utilization loci, Outer membrane vesicles, Multi-modular CAZymes, Type 9 secretion system
1. Introduction

Herbivorous ruminants rely on a close symbiosis with their ruminal microbiota for the proficient conversion of plant biomass to microbial cell protein and volatile fatty acids (McCann et al., 2014). By breaking down the complex matrix of polymers that constitute the cell walls of lignocellulosic feedstuffs, the microorganisms fulfill their host’s nutritional demands while thriving in a suitable environment where they are provided a constant influx of energy and relative environmental stability. These microbiomes utilize a prodigious catalogue of carbohydrate active enzymes (CAZymes) to deconstruct the complex carbohydrates of the plant cell wall, and access the large amounts of inherently stored energy that is otherwise extremely difficult to access. Cellulose in itself is highly recalcitrant to degradation, and in the plant cell wall, it is embedded in a matrix of complex hemicelluloses and lignin, making enzymatic access difficult. To overcome this challenge, the bacteria, ciliates and fungi colonizing the rumen have developed powerful strategies in which they utilize the CAZymes in various mechanisms to liberate and utilize the monomers from the lignocellulose.

Until recently, the study of cellulose degradation in the cow rumen was mostly attributed to a few cultured isolates from the phyla Firmicutes and Fibrobacteres, namely Ruminococcus albus, R. flavefaciens and Fibrobacter succinogenes (Hungate, 1950, Hungate, 1960, Russell et al., 2009). The advent of culture-independent techniques, has revealed that these well-studied isolates are often found in low abundance in situ, with the rumen dominated by uncultured Firmicutes and Bacteroidetes populations (Stevenson and Weimer, 2007, Konietzny et al., 2014). These uncharacterized phylotypes are a potential source of novel CAZymes and new knowledge into the saccharolytic mechanisms that are employed, which builds on classical views of biomass and cellulose degradation in the rumen. The known CAZyme configurations and mechanisms that have been described from well-known cultured microbes include cellulosomes, secreted free cellulases, and polysaccharide utilization loci. In addition, the means of cellulose degradation in certain species is still elusive.

Considering that ruminants obtain most of their energy from their symbiotic microbiota, the efficiency of feed conversion and end-product meat and milk quality in bovines is tightly linked to the dynamics and function of the rumen microbiome. In this chapter, we will give an overview of the role of the microbiota
in ruminal lignocellulose degradation, who is doing what, and cover the mechanisms they utilize in the
decomposition of biomass.

2. Lignocellulosic biomass

Via their anatomical structure and dietary actions, ruminants consume a plethora of plant glycans. In order
to discuss the mechanisms of plant fibre degradation in the rumen, we will first briefly discuss the general
composition of the plant cell wall (Fig. 1).

Plant cell walls comprise of a complex extracellular matrix of polysaccharides, lignin, lipids, minerals, and
glycoproteins, which provide both mechanical strength and protection for the cell. The cell wall is
composed of distinct layers; the middle lamella, the primary cell wall, and the secondary cell wall, which
are all sequentially secreted by the cells’ proplasts (Gibson, 2012). The primary cell wall is composed of
cellulose fibres that are embedded in a matrix of pectin and hemicelluloses such as xyloglucan, xylans and
 glucomannans (Scheller and Ulvskov, 2010). This layer is flexible, and allows for cell growth. When growth
subsides, the secondary cell wall is produced either by modifying the primary cell wall, or by deposition of
an additional second layer. The secondary cell wall is distinct from the primary by the incorporation of
lignin, a matrix of cross-linked phenolic compounds which provides additional mechanical strength to the
cell wall (Zhong and Ye, 2015). Vascular plants use cells with lignified cell-wall as mechanical tissues,
enabling them to grow tall and compete for sunlight. Although not technically part of the structural
polysaccharides that make up the cell wall, starch is an important storage polysaccharide that consists of
\(\alpha\)-D-glucose units and is commonly ingested by ruminants.

2.1 Cellulose

The main component of the cell wall is cellulose, comprising 25-51.4% of typical lignocellulosic biomasses
(Isikgor and Becer, 2015). Cellulose consists of chains of D-glucopyranoside (glucose) molecules linked by
\(\beta\)-1,4-glycosidic bonds (Fig. 1A). The individual glucose molecules are rotated 180° relative to each other,
making the repeating unit cellobiose. The innate molecular structure of cellobiose allows for a great
hydrogen bonding potential in longer oligomers, and with degrees of polymerization (DP) greater than
seven, the affinity for other oligomers is so high that the molecules aggregate and cannot be solubilized in aqueous solvents (Brown, 2004). Native cellulose in the plant cell wall takes on a crystalline form called cellulose I\textsubscript{β}, where 18-24 cellulose chains produced by the cellulose synthase complex on the cell’s plasma membrane, forming micro-fibrils that are stabilized by extensive hydrogen bonding between parallel aligned chains (Schneider et al., 2016). Elementary fibrils aggregate into larger structures called microfibrils, with crystalline and less crystalline (amorphous) regions, which again are interconnected to each other in the plant cell wall by hemicellulose and lignin.

**Figure 1.** Structural and compositional variety of plant polysaccharides, including cellulose (A). hemicellulose (B-E) and pectin (F). Specific substrates include A: cellulose, B: xyloglucan, C: mixed linkage glucans, D: arabinoxylan, E: glucomannan, F: pectin (including rhamnogalacturon).

2.2 Hemicellulose
Hemicellulose is a broad term that describes different heteropolymers that are embedded in the cell walls of plants, of which the detailed structure and abundance can vary depending on the plant species. Traditionally, after extracting pectins and lignin from the biomass, extraction with alkaline treatment was used to separate “the rest” of the polysaccharides from cellulose, which was subsequently termed “hemicellulose” (Scheller and Ulvskov, 2010). More recently, the term hemicellulose has been suggested by Scheller & Ulvskov (2010) to be redefined to include the cell wall polysaccharides that share a common
equatorial β-1,4-glycosidic bond in their backbone. These include xyloglucans, mixed-linkage β-(1,3, 1,4)-
glucans, xylans, mannans, glucomannans and galactomannans.

Xyloglucan is found in all plant cell walls except Charophytes (Scheller and Ulvskov, 2010), and consists of a β-1,4-linked backbone of glucose with or without xylose substitutions (termed monomer X and G, respectively). The xylose residues can in turn be substituted with galactose (termed L) or in grasses, with fucosylated galactose (F) (Fig. 1B). The ratios and distribution of X and G in the backbone and the branching patterns vary between plant species, with common repeating units of XXXG for dicots, or XXGG for Solanaceous species, where the Xs can be un-substituted (X), or substituted (L or F) (Attia and Brumer, 2016). Another hemicellulose with a glucose backbone is mixed linkage β-glucan. It is found in the cell walls of grasses, and consist of trimers and tetramers of β-1,4-linked glucose, linked together via β-1,3-glycosidic bonds (Fig. 1C).

Xylans are hemicelluloses with β-1,4-linked xylose units in the backbone. They are often the dominating non-cellulose polysaccharide in secondary cell walls in dicots, and are usually found with α-1,2-linked glucoronosyl and 4-O-methyl glucuronosyl substitutions, and are called glucuronoxylans. In commelinid monocots, arabinoxylans are dominating in the primary cell wall, more heavily substituted with arabinose (Scheller and Ulvskov, 2010) (Fig. 1D).

The third type of hemicelluloses contains a backbone comprised of mannose linked by β-1,4-glycosidic bonds, and is found in variable amounts in all cell walls. In mannans and galactomannans, the backbone contains only mannose, variously substituted with galactose units. Glucomannans have both mannose and glucose in their backbone, linked by β-1,4-bonds in nonrepeating patterns (Fig. 1E).

3. Carbohydrate active enzymes

Lignocellulosic biomass is the most abundant biomass on Earth, and its complex structure makes it highly resistant to microbial attack. The structure of carbohydrates is enormously varied in nature, with one reducing hexameric sugar yielding 1012 possible linear and branched isomers (Laine, 1994). However,
without recycling of fixated carbon within the carbohydrates of plant cell walls, heterotrophic organisms would not be able to acquire energy. Therefore, Nature has evolved a vast array of tools to overcome the complexity of carbohydrate breakdown.

Enzymes and accessory proteins that act on carbohydrates for both assembly and breakdown of polysaccharides are collectively designated carbohydrate active enzymes, otherwise referred to as CAZymes (Lombard et al., 2014). In 1999, the CAZy Database [www.cazy.org] was launched to act as a central repository for CAZyme information, including sequence, 3D structures and biochemical data. The database currently holds protein families divided in the six classes based on their mode of action: glycoside hydrolases (GH), carbohydrate esterases (CE) and polysaccharide lyases (PL) for deconstruction: glycosyltransferases (GT) for synthesis: carbohydrate binding modules (CBMs) that help targeting enzymes to their substrates, and auxiliary activity (AA) enzymes that cover redox-enzymes that act in concert with other CAZymes.

CAZymes are classified into families based on their amino acid sequence, which in turn reflects their three-dimensional structure and fold (Henrissat, 1991, Cantarel et al., 2009). As the number of carbohydrate substrates greatly exceed the number of folds, the enzymes have evolved from common folds and thus several enzyme specificities can exist within the same family. Likewise, the same enzyme specificities can be found in different families, exemplifying convergent evolution. CAZymes are often multi-modular and can contain several domains from different families, which allows one protein sequence to be classified into several families. The rest of this section will provide an overview of the functions of the various CAZyme classes.

3.1 Glycoside hydrolases

The largest class in terms of both the number of sequences and the number of families in the CAZy database is the class of glycoside hydrolases, with 156 families and over 600 000 sequences, reflecting the enormous variation in available carbohydrate substrates (Lombard et al., 2014). Glycoside hydrolases catalyze the hydrolysis of glycosidic bonds and the reaction can occur with two different mechanisms, where the anomeric configuration of the glycosidic bond is either retained or inverted (Koshland, 1953).
The reaction is catalyzed by two conserved amino acid residues in the enzyme, normally glutamic acid or aspartic acid (McCarter and Withers, 1994). These act as a general acid (proton donor) and a base, and the spatial position of their side chains correlates with type of mechanism.

Glycoside hydrolases have a wide variety of specificities, attacking the backbone of linear polysaccharides, targeting crystalline substrates or acting as debranching enzymes attacking only specific substitutions of a particular hemicellulose. GHs acting on polymers can be either endo- or exo-acting, referring to whether the enzyme attacks glycosidic bonds within the polysaccharide or at the chain ends, respectively. Often, exo-acting enzymes processively perform several hydrolytic events without dissociation from their substrate, and they have specificities towards either the reducing- or non-reducing end of the polysaccharide (Davies and Henrissat, 1995, Barr et al., 1996). In cellulose degradation, endocellulases (currently found in 14 GH families, but typically belonging to GH5, GH9 or GH45) and exocellulases (Typically GH6, GH7, GH48, but some GH9s have been reported as processive endocellulases, releasing cellobiose) work synergistically to degrade the crystalline substrate (Wood and McCrae, 1979, Kostylev and Wilson, 2012). In this process, endocellulases cleave cellulose chains internally in amorphous regions, creating chain-ends for the processive exocellulases (cellobiohydrolases) that release cellobiose.

3.2 Carbohydrate binding modules

Carbohydrate binding modules are non-catalytic modules of CAZymes that help target the enzyme towards a carbohydrate (Boraston et al., 2004). The first characterized CBMs were cellulose binding modules, found to facilitate binding to cellulose by Trichoderma reesei cellobiohydrolase I and II, and in two cellulases from Cellulomonas fini (Van Tilbeurgh et al., 1986, Tomme et al., 1988, Gilkes et al., 1988). Subsequently, the carbohydrate targets of CBMs have been shown to cover almost all known carbohydrates, including cellulose and hemicelluloses found in the plant cell wall (McCartney et al., 2004, Lombard et al., 2014). CBMs are classified into families based on their amino acid sequence, with 84 different families to date, and can be divided into three functional classes (Gilbert et al., 2013). Type A targets surfaces of crystalline polysaccharides, type B targets sites internally on carbohydrate chains, and type C targets the termini of glycan chains.
CBMs are thought to contribute to the efficiency of the appended catalytic domain by increasing the concentration of the enzyme near the substrate (McCartney et al., 2004). By keeping the catalytic domain in closer proximity to its substrate, there is a greater probability for catalytic events to occur. This is more important in low substrate concentration conditions, as demonstrated by the observation that the presence of a cellulose-binding CBM in a cellulase had a lesser, or even negative effect on enzyme efficiency at high substrate concentrations (Várnai et al., 2013). Even though the CBM normally targets the substrate of the appended catalytic domain, there are examples of CBMs that bind to differing substrates enabling the catalysis of target polysaccharides located in the proximity of the bound glycan (Hervé et al., 2010).

### 3.3 Other CAZyme families

The auxiliary activity group encompasses laccases, cellobiose dehydrogenases, copper radical oxygenases and various enzymes utilizing oxidative mechanisms on carbohydrates (Lombard et al., 2014). Lytic polysaccharide monooxygenases (LPMOs) were previously thought to be CBMs that helped disrupt the crystalline structure of chitin, but the *Serratia marcesens* Cbp21 enzyme was shown to introduce chain-breaks, generating oxidized chain-ends on the crystalline chitin surface (Vaaje-Kolstad et al., 2010). Subsequent discovery of cellulose-active LPMOs from bacteria and fungi spurred the need for a establishing a separate class in the CAZy database, namely the auxiliary activities (Forsberg et al., 2011, Quinlan et al., 2011, Levasseur et al., 2013). LPMOs are now classified into AA families 9, 10, 11 and 13, out of 13 AA families in total. These enzymes are currently receiving massive attention because they boost the activities of classical GHs and thus contribute to the overall efficiency of enzyme cocktails. Importantly, LPMOs may be crucial in solving the accessibility challenge discussed above, because they are capable of breaking glycosidic bonds that are in a crystalline context, thus generating access for classical GHs. Despite a variety of rumen isolate genomes encoding predicted AA10 representatives (Seshadri et al., 2018), LPMO activity has not been detected in the rumen, and thus far has only been reported in aerobic microorganisms.

Carbohydrate esterases function as debranching enzymes, and de-O-, or de-N-acylate ester-based modifications of complex polysaccharides. In removing these ester-based modifications, they allow GHs easier access to their targets in complex polysaccharides (Cantarel et al., 2009). Polysaccharide lyases...
utilize β-elimination to cleave the glycosidic bonds of uronic acid-containing polysaccharides. This leaves the sugar on the new non-reducing end unsaturated with a double bond between C4 and C5, while the new reducing end is saturated (Garron and Cygler, 2010). As the only anabolic members of the CAZy database, glycosyl transferases utilize sugar phosphates to form glycosidic linkages between the “activated” sugar and other saccharides, lipids or proteins (Lairson et al., 2008). The glycosyl group is transferred to a nucleophilic group on the substrate in a retaining or inverting fashion.

4. Prokaryotic strategies for fibre digestion in the rumen

Decades of research on various digestive ecosystems (i.e. soil, marine, host-associated) has shown that all saccharolytic microbes rely on the actions of CAZymes, however how these CAZymes are employed by their microbial host can vary considerably. In this section, we will describe the various mechanisms employed by the bacteria and anaerobic fungi of the rumen.

4.1 Cellulosomes

The cellulosome is the best known cellulolytic mechanism and is one of the two main paradigms of microbial cellulose degradation, along with secreted free enzymes by aerobic fungi (Wilson, 2011). The cellulosome organization of cellulases is found in several anaerobic bacteria and anaerobic fungi, but was first described in Clostridium thermocellum, an anaerobic thermophilic soil bacterium (Lamed et al., 1983, Bayer et al., 2008). The cellulosome is a multi-modular enzyme complex that enables the cell to adhere to crystalline cellulose, which is degraded by cellulases in the ultrastructure (Bayer and Lamed, 1986). The endo- and exo-cellulases along with hemicellulases of the cellulosome contain dockerin domains in addition to their catalytic domains, which facilitate docking to cohesin domains on the large non-catalytic scaffoldin subunit (Bayer et al., 1994, Yaron et al., 1995). The primary scaffoldin subunit of C. thermocellum (cipA) contains nine cohesin domains for the binding of dockerin-linked enzyme subunits, a CBM3 module that binds to crystalline cellulose, and a C-terminal dockerin domain that binds to type II cohesin in the cell-wall anchoring scaffoldins. The anchoring scaffoldins contain S-layer homology (SLH) domains that fix the cellulosome to the cell surface, enabling the cell to be in close proximity to the solubilized cello dextrans released by the cellulases (Bayer et al., 2008). Three different anchoring scaffoldins allow up to 63 different cellulosome components to be attached in a single complex in C.
thermocellum. The modular structure of the cellulosome brings the endo- and exocellulases close together and bound to the substrate, allowing for synergy in cellulose degradation (Krauss et al., 2012).

Figure 2. Schematic representation of cellulosome (A) and free-enzyme (B) strategies that are employed by rumen microorganisms. A specific example of the cellulosome characterized from Ruminococcus flavefaciens strain FD-1 is illustrated in (A). OM: outer membrane, GH: glycoside hydrolase, CBM: carbohydrate binding module, SLH: S-layer homology.

C. thermocellum is not found in ruminal environments, however the known cellulose degraders R. flavefaciens and R. albus both partly utilize cellulosomes to contribute to the plant biomass degradation of the rumen (Ohara et al., 2000, Ding et al., 2001). These differ from C. thermocellum in their modular composition of the cellulosome where R. flavefaciens has a particularly elaborate system encoding a large
amount of dockerin encoding proteins, including novel CBMs (Dassa et al., 2014, Venditto et al., 2016). The cellulosomal proteins from \textit{R. flavefaciens} have been shown to vary within strains, with genome analysis demonstrating that the number of dockerins can vary between 53-223 (Seshadri et al., 2018). Furthermore, the cohesin-dockerin interactions are for the most part strain specific (Israeli-Ruimy et al., 2017), while it has been demonstrated that up to 14 enzyme subunits from the \textit{R. flavefaciens} strain FD-1 cellulosome are assembled across four distinct scaffoldins (Fig. 2A). In contrast to \textit{R. flavefaciens}, \textit{R. albus} contains a lower abundance of dockerin encoding genes, and two of three sequenced strains contained only one cohesin encoding gene, whereas the third contained no cohesin counterparts. This suggests the existence of a so-far undiscovered type of scaffoldin with a novel cohesin-like domain, or that the cellulolytic bacterium does not utilize a “full” cellulosome mechanism (Dassa et al., 2014).

The cellulosomes found in anaerobic fungal genomes greatly differ from those found in anaerobic bacteria (Haitjema et al., 2017). Orthologues of a large scaffoldin protein (ScaA) with no sequence similarity to bacterial scaffoldins were found in all five sequenced genomes of anaerobic fungi. Interestingly, the dockerin domains from three genera of gut fungi were able to bind to all combinations of ScaA fragments, containing cohesin-motifs. The authors therefore speculate that in their native environment, fungal cellulosomes may actually exist as composites of enzymes from different fungal species, unlike bacterial dockerin-cohesin interactions which are highly species-specific.

**4.2 Secreted enzymes**

The second most renowned paradigm of microbial cellulose degradation is that of secreted free cellulases, primarily in aerobic fungi and bacteria (Fig. 2B). This mechanism has been well studied in the mesophilic filamentous fungus \textit{Trichoderma reesei}, which was originally isolated from rotting US Army equipment in the Solomon Islands during World War II, and is the dominant industrial cellulase-producing organism (Bischof et al., 2016, Reese, 1955).

\textit{T. reesei} secretes large amounts of cellulase enzymes to the culture broth when grown on cellulose, enabling rapid degradation of cellulose to glucose (Sheir-Neiss and Montenecourt, 1984). The secreted enzymes include endoglucanases (EG, Cel5A, Cel5B, Cel12A, Cel45A), non-reducing- and reducing end
cellobiohydrolases (CBHI/GH6 and CBHII/Cel7A, respectively), β-glucosidase I (GH3), and AA9 LPMOs (formerly GH61) (Saloheimo et al., 1997, Wilson, 2009, Westereng et al., 2011, Li et al., 2016). All these enzymes work in concert, where endoglucanases (with or without CBM1 domains) attack β-1,4 glycosidic bonds in the amorphous regions of cellulose, creating chain-ends for the processive CBHs that attack from both ends of the chains, disrupting the crystalline structure. Cellobiose released from the CBHs is degrades to glucose monomers which are taken up by the cell. The AA9 type LPMOs introduce oxidative breaks in the crystalline region of cellulose, creating more chain-ends for CBH, possibly acquiring reducing power from non-enzymatic donors such as lignin (Westereng et al., 2015). Other fungal cellulase systems utilize secreted enzymes as well, but these have not been studied to the same extent (Wilson, 2008). Several bacteria also use the secreted free enzyme mechanism for cellulose degradation, in a similar manner to the fungi (Wilson, 2011). Thermobifida fusca is one well-studied example, utilizing GH5, GH6 and GH9 endo-cellulases, GH6 and GH48 exo-cellulases containing CBM2 cellulose binding domains, and two AA10 type LPMOs (Gomez Del Pulgar and Saadeddin, 2013, Forsberg et al., 2014).

The role of secreted free-enzymes in the rumen of herbivore is not well understood. For example, based on sequence prediction, it is believed that R. flavefaciens and rumen fungi also utilize secreted enzymes in addition to cellulosomes (Dassa et al., 2014). Several cellulose-degrading Firmicutes are predicted to use secreted exo- (GH48) and endo-cellulases (GH5, GH9) such as selected species affiliated to the Lachnoclostrium, Cellulosilyticum (Cai et al., 2010), Ruminoclostridium and Ruminococcus genera (Seshadri et al., 2018). Similarly, Cellulomonas sp. affiliated to the Actinobacteria phylum is also predicted to degrade cellulose via the actions of free exo- and endo-cellulases. Many populations are also predicted to degrade hemicellulose and starch using free enzymes such as the predominant rumen microbe Butyrivibrio fibrisolvens (Seshadri et al., 2018).

4.3 Polysaccharide Utilization Loci

A large fraction of the bacteria colonizing gut microbial ecosystems are affiliated to the Bacteroidetes phylum (Tajima et al., 1999, Hold et al., 2002, Flint et al., 2008). These gram-negative bacteria are known to have broad carbohydrate degradation capabilities derived from numerous gene clusters termed polysaccharide utilization loci (PULs), first characterized in the human gut bacterium Bacteroidetes thetaiotaomicron (Bjursell et al., 2006, Martens et al., 2008). B. thetaiotaomicron encodes 101 susC/D
pairs, enabling it to grow on a wide array of polysaccharides and host associated mucins. Moreover, transcriptomic analysis has demonstrated up-regulation of PULs specific to available substrates (Martens et al., 2011). Not limited to human gut bacteria, PULs have been shown to be prevalent in Bacteroidetes found in environmental and herbivorous gut microbiomes (including ruminants) (McBride et al., 2009, Pope et al., 2010, Pope et al., 2012, Naas et al., 2014, Dodd et al., 2010, Terrapon et al., 2014, Accetto and Avguštin, 2015, Mackenzie et al., 2015, Güllert et al., 2016, Rosewarne et al., 2014). PULs typically target starches, hemicelluloses and pectins, and are encoded by a large number of the *Prevotellaceae* in the cow rumen (Stewart et al., 2018b, Solden et al., 2018), where they likely contribute to the liberation of cellulose from the hemicellulose matrix. In particular, a xylan-degrading PUL from *Prevotella bryantii* was one of the first to be characterized in detail outside of the human gut, using both culture and omic based methodologies (Dodd et al., 2010).

PULs are defined by the co-occurrence of SusC- and SusD-like genes, along with glycoside hydrolases and sugar transporters (Fig. 3A). SusC and SusD-like proteins are named after their first description in the starch utilization system (Reeves et al., 1996, Reeves et al., 1997). The sus gene cluster contains eight genes, named *susRABCDEFG*, and contains all the enzymes required for the cell to degrade and import starch (Foley et al., 2016). SusR, an inner membrane trans-membrane protein, is the transcriptional regulator of the gene cluster, and can recognize the starch degradation product maltose (D'Elia and Salyers, 1996b). This up-regulates the expression of the rest of the gene cluster, enabling the cell to respond to the presence of starch. The outer membrane lipoproteins SusD, E, and F, facilitate binding of starch to the cell surface, bringing the outer-membrane lipo-anchored α-amylase SusG in close proximity of its substrate (Shipman et al., 1999, Shipman et al., 2000, Koropatkin and Smith, 2010). SusD is essential for growth on maltooligosaccharides greater than DP4, and also greatly enhances the sensing of maltose compared to mutants where its single starch binding domain was disrupted (Cameron et al., 2014). The binding sites and expression of SusE and SusF were dispensable for activation of transcription, but enhanced the growth rate in a substrate-dependent manner. Maltooligosaccharides released by SusG are imported through the TonB-dependent outer-membrane porin SusC into the periplasmic space, where they are further hydrolyzed by SusB α-glucosidase and SusA neopullunase (D'Elia and Salyers, 1996a, Reeves et al., 1996). The resulting glucose is further transported into the cell for fermentation.
Individual PULs are varied in the number of co-located CAZymes, which is correlated to the complexity of their predicted substrate. They can range from two enzymes in the predicted pectic galactan PUL of *B. thetaiotaomicron*, to 32 in the pectic rhamnogalacturonan II PUL (Koropatkin, Cameron and Martens, 2012). Several PULs have been studied extensively through biochemical analyses of individual genes along with growth experiments, revealing complex enzyme interplays with sequential polysaccharide deconstruction activities on substrates including fructan, porphyran, xyloglucan, xylan and α-mannan (Sonnenburg and Zheng, 2010; Hehemann *et al.*, 2012; Larsbrink *et al.*, 2014; Cuskin *et al.*, 2015; Rogowski *et al.*, 2015). PULs usually contain enzymes to target one specific polysaccharide, but a PUL from an uncultured reindeer rumen Bacteroidetes species was shown to degrade mannans, xylans, xyloglucan and β-glucans (Mackenzie *et al.*, 2015).

**Figure 3.** Schematic representation of polysaccharide utilization loci (PULs) that are found in gram-negative (A) and gram-positive (B) rumen microorganisms. Gram-negative PULs are typified by the genomic co-localization of SusC- and SusD-like lipoproteins and glycoside hydrolases, whereas in gram-positive microbiota, a carbohydrate binding protein (CBP) and permeases (MPP) of an ABC transporter are encoded. Both strategies include extracellular (membrane attached) and intracellular CAZymes to take complex polysaccharides to monomeric sugars. OM: outer membrane, IM: inner membrane, reg.: transcriptional regulator, trans.: sugar transporter.
Until recently, Bacteroidetes PULs were thought to only target soluble glycans (Koropatkin et al., 2012). However, the soil Bacteroidetes Flavobacterium johnsoniae can degrade the crystalline cellulose homologue chitin efficiently via a PUL in cooperation with a secreted multi-modular chitinase (McBride et al., 2009, Larsbrink et al., 2016). In addition, PULs have also been linked with putative cellulases in several herbivore gut metagenomes (Pope et al., 2010, Pope et al., 2012, Dai et al., 2012, Naas et al., 2014), however no isolated Bacteroidetes representative has been shown to degrade crystalline cellulose via the actions of a PUL.

4.4 PULs are in the gram positives as well

While gram-negative bacteria, in particular the Bacteroidetes, have long been associated with PULs, recent studies in Firmicutes have shown that similar cell-wall enveloped plant biomass-degrading strategies exist in gram-positives as well (La Rosa et al., 2019) (Fig. 3B). Detailed genomic, transcriptomic and proteomic analysis of PULs from human gut Roseburia species illustrated that multi-gene loci encode and express the necessary cell-wall anchored CAZymes, binding proteins, ABC transporters, transcriptional regulators and cytoplasmic oligosaccharide-degrading CAZymes to convert complex hemicellulosic substrates into monomeric sugars. Moreover, detailed biochemical characterization of each component of the aforementioned PUL validated each of their predicted metabolic functions (La Rosa et al., 2019). To the best of our knowledge, such detailed examples of elucidated gram-positive PULs have not been described in ruminant microorganisms, however perusal of the Hungate1000 genome collection quickly identifies potential examples of gram positive PULs in various phyla including the Firmicutes (e.g. Blautia schinkii DSM 10518) and Actinobacteria (e.g. Bifidobacterium longum AGR2137) (Seshadri et al., 2018).

4.5 Fibrobacter succinogenes and outer membrane vesicles (OMVs)

One of the classically recognized cellulose degrading species from the cow rumen, Fibrobacter succinogenes, utilizes a cellulolytic mechanism that does not conform to the classical views of cellulose degradation (Suen et al., 2011). It is regarded as one of the key fibrolytic populations in the rumen, and has demonstrated a greater ability to digest cellulose from forages than other species of rumen bacteria (Dehority and Scott, 1967). Despite this, its genome is devoid of the cellulosome components dockerin and cohesins, and it does not appear to encode any exo-cellulases, known to be required for both the
secreted free enzymes and cellulose mechanisms. *F. succinogenes* is equipped with a surprisingly high diversity of CAZymes and abilities, considering that the bacterium only utilizes cellulose as a carbon source. The 31 encoded endo-cellulases of *F. succinogenes* also do not contain CBMs associated with binding to cellulose, further discrediting secreted enzymes as a possible mechanism.

**Figure 4.** Hypothetical strategies predicted to occur in *Fibrobacter succinogenes* (A) and as-yet uncultured members of the Bacteroidetes family “*Candidatus MH11*” (B). *F. succinogenes* produces outer membrane vesicles that contain multiple CAZymes and a fibro-slime protein complex that target various cellulose, hemicellulose and pectin substrates. The Ca. MH11-affiliated “*Candidatus Paraporphyrromonas polyenzyomogenes*” encodes components of the Type IX secretion system (T9SS), which is predicted to facilitate secretion of multi-modular CAZymes. Whether or not TPSS CAZymes in Ca. MH11-affiliated populations are free or cell-wall attached is not known. OM: outer membrane, IM: inner membrane.

Several models of its cellulose degradation have been proposed, and some studies point towards cell attachment to cellulose using fibro-slime proteins and pili, bringing the substrate close to outer
membrane-bound endo-cellulases (Burnet et al., 2015). The bacterium also produces outer-membrane vesicles (OMVs) containing carbohydrate-active enzymes, which are released from the cell to target plant biomass (Arntzen et al., 2017) (Fig. 4A). OMVs have been observed in a wide range of gram-negative species found in various ecosystems (Kulp and Kuehn, 2010), and have been suggested to play wide-ranging roles including horizontal gene transfer, biofilm formation, communication and biomolecule delivery (Elhenawy et al., 2014, Roier et al., 2016). OMVs from F. sugginogenes were found to contain fibro-slime proteins, cellulases and hemicellulases, and were able to degrade cellulose, pectin and hemicelluloses. The role of these OMVs in fibre-digestion is debated, with studies claiming their production is due aging of the cells (Gaudet and Gaillard, 1987). However, there is support they may have a biological role in cellulose degradation (Forsberg et al., 1981), with pre-treatment of switchgrass using the F. sugginogenes OMVs showing a 2.4-fold increase in subsequent saccharification by a commercial cellulase cocktail (Arntzen et al., 2017). It has been illustrated that other fibre-digesting Bacteroidetes species from the human gut also produce fibrolytic OMVs (Elhenawy et al., 2014) and it is hypothesized that F. sugginogenes OMVs could facilitate increased accessibility for the cell to the substrate by disrupting the complex structure of the lignocellulose (Arntzen et al., 2017).

4.6 Bacteroidetes, the Type IX secretion system and multi modular CAZymes

A different non-classical mechanism is predicted to be utilized by deeply branched, as-yet uncultured members of the Bacteroidales order that are inherent to the rumen (Naas et al., 2018). A novel family referred to as “Candidatus MH11” consists of populations that do not encode cellulosomes or PUL-like systems, and instead utilize multi-modular cellulases and hemicellulases that are secreted via the Type IX secretion system (T9SS). The T9SS is essential for crystalline substrate degradation in both soil Bacteroidetes Cytophaga hutchinsonii (cellulose) and F. johnsoniae (chitin), which utilize T9SS secreted large cellulases and multi-modular chitinases, respectively. The importance of the T9SS for fibre-digestion was first described as part of the aerobic gliding mechanism in C. hutchinsonii, which is cellulolytic but lacks dockerin and cohesin modules, as well as predicted exo-cellulases (Xie et al., 2007). The cells attach to, and move across cellulose fibres utilizing its gliding motility, which is speculated to aid the cellulose deconstruction by locating more easily accessible substrate. The gliding motility of Bacteroidetes has been well studied in the distant relative F. johnsoniae, and has been functionally linked to both crystalline substrate degradation and secretion through the T9SS (Sato et al., 2010, Ji et al., 2012, Zhu and McBride, 2014, McBride and Nakane, 2015).
In addition to the T9SS being important, multi-modular CAZymes have recently been shown to be widespread in bacterial genomes, and many putatively target cellulose and the cellulose homologue chitin (Talamantes et al., 2016). For example, a multi-modular cellulase (CbCelA: GH9/CBM3c/CBM3b/CBM3b/GH48) secreted by the thermophilic *Caldicellulosiruptor* species, is predicted to play an important part in cellulose degradation due to its putative endo/exo synergistic activity, and its deletion was shown to greatly reduce growth on-, and degradation of cellulose (Yi et al., 2013, Young et al., 2014). A study by Brunecky et. al. showed that purified CbCelA from culture supernatant could outperform mixtures of commercial exo- and endocellulases, likely due to its inter-domain synergy (Brunecky et al., 2013). Within the Bacteoidetes phylum, the main chitinase of the chitin utilization locus (ChiUL) of *F. johnsoniae*, FjChiA, is similar to CbCelA in that it has a flanking endo/exo acting pair of GH18 chitinases, with chitin/cellulose binding domains in the middle region (Larsbrink et al., 2016). The high activities of these two similar enzymes on recalcitrant substrates suggests that multi-modularity in glycoside hydrolases is a well-functioning strategy that could possibly be found in other biomass degrading organisms. *Fj*ChiA has also been shown to be secreted via the T9SS, whereas gene knock-out mutagenesis has demonstrated that the enzyme is vital for chitin metabolism and cell growth (Kharade and McBride, 2014).

Within the cow rumen, the Ca. MH11-affiliated, provisionally named “*Candidatus* Paraporphyromonas polyenzymogenes”, is predicted to have cellulose-degrading capabilities, with over 100 CAZyme domains including 17 putative endo-cellulases and members of GH3 β-glucosidases and GH94 cellobiose phosphorylases (Naas et al., 2018). Notably, many of the encoded CAZymes were multi-modular and found to include a specific C-terminal domain (CTD) known to target proteins for secretion through the T9SS (Fig. 4B). Similar cellulase-gene organizations were found in the six other representatives in Ca. MH11 affiliates, all native to the sheep rumen, suggesting that T9SS secreted multi-modular enzymes play a part in ruminal cellulose and hemicellulose degradation. Moreover, biochemical and structural analysis of selected Ca. P. polyenzymogenes CAZymes has further supported the predicted cellulolytic phenotype of this hitherto uncultured bacterium, and has demonstrated activity towards linear polymers, such as amorphous and crystalline cellulose as well as mixed linkage β-glucans (Naas et al., 2018).
5. What are we missing?

5.1 The eukaryotes

While prokaryotes numerically dominate the rumen microbiome, it is also widely acknowledged that eukaryotic populations make important contributions towards fibre digestion. Rumen protozoa make up approximately 20% of the microbial biomass within the rumen due to their comparably large cell volume (Huws et al., 2018). Anaerobic rumen fungi were first reported back in the 1970’s (Bauchop, 1979, Orpin, 1975), and have long been known to harness CAZymes to deconstruct the plant cell walls of ingested grasses (Borneman et al., 1989). However for both rumen protozoa and fungi, there remains a dearth of genomic information, largely due to the difficulty with growing them axenically, and the ability to purify, sequence and annotate the genomes of both cultured and uncultured representatives. The first draft macronuclear genome sequence of a ruminal protozoa (Entodinium caudatum MZG-1,) was released in 2018 (Park et al., 2018), and as of writing only representatives fungal genomes from the genera Anaeromyces, Neocallimastix, Orpinomyces and Piromyces are available (Haitjema et al., 2017). New developments in culture-independent metagenomic data generation and the bioinformatic processing of this data is creating new possibilities to reconstruct representative genomes from uncultured fungal and protozoal populations. For example, long-read sequencing technologies (Oxford Nanopore, PacBio) is improving the assembly of eukaryotic genomes (Díaz-Viraqué et al., 2019), whereas genome binning software (EukRep) has recently been designed to enable reconstruction of eukaryotic genomes from complex microbial communities (West et al., 2018). Looking closer at eukaryotic function, the exact mechanisms by which rumen protozoa and fungi degrade plant fibre have not been detailed in the same manner as the rumen prokaryotes. Protozoa are believed to use multi-modular cellulases and hemicellulases, whilst cellulosomes with similar structure to those produced by bacteria have been observed in rumen fungi (Steenbakkars et al., 2001, Fanutti et al., 1995, Ljungdahl, 2009), albeit with dockerin domains that share no sequence homology (Haitjema et al., 2017).

5.2 How much CAZyme and glycan diversity is there in the rumen?

Plant glycans present within animal feed are often broadly characterized as cellulose, hemicellulose, starch and/pectin. However, within several of these categories (i.e. hemicelluloses, pectins) there exists an extraordinary diversity of glycan structures, many of which remain unchartered. At a higher compositional level, common feed sources for pen-fed ruminants such as grasses, grains and legumes are
dominated by cellulose, which accounts for between 15-50% depending on the plant cellular location (primary, secondary etc). Matrix polysaccharides, such as hemicellulose and pectin are more structurally complex, and can be branched and/or substituted with methyl esterifications or acetylations. Grasses (and grains) are more commonly dominated by xylans (20-40%) and mixed-linkage glucans (10-30%), whereas many dicots (such as legumes) contain higher levels of xyloglucan (20-25%), mannans (5-10%) and pectins (20-35%) (Mertens, 2003, Vogel, 2008, Pattathil et al., 2015). In grazing animals, our understanding of the dietary fibre composition is much more deficient, however plant microarray methods that attempt to estimate and/or map the “carbon landscape” within a given plant sample have recently been used to characterize the diets of wild ruminants such as moose and reindeer (Mackenzie et al., 2015, Solden et al., 2018). These methods, which rely on monoclonal antibodies or CBMs (Moller et al., 2007), have reaffirmed that complex fibres dominate their cosmopolitan diets, and include varyingly structured mannans, xylans, xyloglucans, galactans, arabinans and pectins (Mackenzie et al., 2015, Solden et al., 2018). While many diverse fibres have been identified using these and other methods (Voiniciuc et al., 2018, Wood et al., 2017), many uncharacterized structural layers still exists within ruminant diets, requiring greater efforts to map this inherent glycan diversity.

Indigenous fibre-degraders in the rumen have evolved to match this glycan diversity with an equally enormous inventory of CAZymes, which includes endo- and exo-acting enzymes that deconstruct the polysaccharide backbone as well as the auxiliary enzymes (carbohydrate esterases, polysaccharide lyases and other debranching enzymes). Just looking alone at the CAZyme profile of 964 available Bacteroidetes genomes (predominated by the human gut), over 13,500 PULs have been grouped and used to estimate that there exists approximately several thousand enzyme combinations for the breakdown of the various glycan structures found in nature (Lapébie et al., 2019). Analysis of Individual rumen populations, such as R. flavefaciens, has illustrated an extraordinary level of glycan recognition (Venditto et al., 2016), however community wide estimations have not yet been calculated for the 100’s of rumen fibrolytic species that exist, many of which are uncultured. It is thus apparent that much cataloguing of both glycan and CAZyme diversity remains for rumen microbiologists.

### 5.3 What do functional omic studies tell us?

Coinciding with rapidly improving DNA-based sequencing and bioinformatics techniques that are enabling rumen microbiologists to create 1000’s of microbial genomes from the rumen, functional RNA- and
peptide-based expression studies have also experienced a transformation. Several chip-based microarrays have been developed over the years to quantify transcript levels from both bacterial as well as protozoan and fungal populations (Abot et al., 2016, Comtet-Marre et al., 2018). Overall, these methods have illustrated the high expression values of various cellulases and hemicellulases from both prokaryotic and eukaryotic origin, although the design of such chips on previously available CAZyme data, prevent the detection of specific activities from individual populations that have not yet been genomically sampled.

Quantitative metatranscriptomic methods have been used to analyze gene expression patterns across to the entire rumen microbiome, and have shown that similar taxa (*Prevotellaceae*, *Succinivibrionaceae*, and *Fibrobacteraceae*) and CAZymes are prevalent in multiple studies (Söllinger et al., 2018, Comtet-Marre et al., 2017). Moreover, these studies have highlighted that anaerobic fungi and ciliates contribute an unexpectedly large share of transcripts for cellulose- and hemicellulose-degrading enzymes (Söllinger et al., 2018, Comtet-Marre et al., 2017, Qi et al., 2011). New multi-omic approaches that combine large metagenomic and metatranscriptomic or metaproteomic datasets, are further improving resolution to specifically identify individual populations and the mechanisms they use to actively degrade fibre (Solden et al., 2018). For example, we used multi-omics to show that Ca. P. polyenzymogenes was detectable in metaproteomic data and was enriched in samples recovered from rumen-incubated plant biomass, thus indicating that active digestion of complex carbohydrates could be assigned to members of the novel Ca. MH11 family, which uses a non-conventional T9SS-based saccharolytic mechanism (Fig. 4B) (Naas et al., 2018). Multi-omics have also reiterated that Bacteroidetes-affiliated PULs are critical for rumen fibre digestion, with MAG-centric metaproteomic data recovered from Alaskan moose revealing greater than 90% of the detected CAZymes were expressed from PULs (Solden et al., 2018). While these initial studies illustrate that such methods can be used to create deeper understanding regarding microbial plant fibre degradation, they have thus far only observed a fraction of the larger community dynamics that are in play, which is constantly varying in response to time (i.e. such as before and after eating) as well as individuality factors of the animal host.
6. How can this accumulated knowledge be used to improve the fibre digestion process?

The rumen microbiome has long been viewed as a potential target for manipulation to increase fibre digestion, improve animal productivity and wellbeing, as well as to reduce methane emissions. For the most part, attempts to directly target specific fibrolytic populations via supplementation of actual live cultures (Chiquette et al., 2007, Præsteng et al., 2013, Krause et al., 2001), chemical agents (Chalupa, 1977) or exogenous CAZymes (Beauchemin et al., 2003) to boost ruminal fibre digestion have been unsuccessful (Morais and Mizrahi, 2019). While diet has long-been assumed to play the main driver in shaping the gut microbiota of bilaterians (Spor et al., 2011), and especially for ruminants (Henderson et al., 2015), new evidence is validating that host genetics is also important. It has been highlighted that individual variation of rumen microbiota exists in both beef (Li and Guan, 2017) and dairy cattle (Jami and Mizrahi, 2012), even when animals were fed the same diet and managed under the same environment. More recently, genome-wide association studies (GWAS) have identified heritable rumen bacteria (Sasson et al., 2017, Li et al., 2019), and it has also been demonstrated that genetic variation in cows can lead to differences in microbial gene/taxa abundance, host feed efficiency and methane production (Roehe et al., 2016, Difford et al., 2018). Moreover, Li et al. recently showed that heritable microbiota are additionally associated with single nucleotide polymorphisms located in the bovine genome that are known quantitative trait loci for feed efficiency in cattle, further highlighting that perhaps breeding strategies can be used to manipulate or select for beneficial microbiota (Li et al., 2019).

This exciting new line of research thus proposes the tantalizing question: could we improve fibre digestion in ruminants by linking host genetics to microbiome function and beyond to specific glycan profiles in their diets? The ultimate idea being that we could theoretically customize diets for specific cattle breeds with specific glycan structures, which would match the enzymatic/mechanistic capabilities of their host-linked microbiota. Thus far, previous ruminant GWAS have yet to elucidate at a profound functional level how the expressed metabolic enzymes or pathways within (multiple) microbial populations are linked to host genotypes as well as specific glycan structures. Moreover, the majority of heritable populations identified are assigned to taxa for which no cultured isolate, genome or metabolic information is available. Therefore, we lack a deeper understanding of “holobiont phenotypes” i.e. how important interactions among cow and microbial genomes, and their expressed enzymes / metabolic pathways, affect variation...
in fibre digestion. Tackling such a challenge has historically been “out of bounds” both technically and economically, however today’s molecular toolkits are increasing the feasibility to create 1000’s of rumen microbial genomes (Stewart et al., 2018a, Seshadri et al., 2018), map the glycan structures consumed in animal feed and disentangle extremely complex interactions between feed, the “gut microbiome”, and host genetics. Thus, we hypothesize that it is becoming possible to study the high-dimensional multispecies molecular phenotype of animals and their residential microbiomes, i.e. the holobiont. This includes the genomes, which genes are expressed, and what these genes ‘produce’ in terms of enzymes and interacting biochemical reactions. Ultimately, such knowledge could be incorporated into commercial feed design and breeding programs to optimize fibre digestion and animal production in general.

7. Summary, future trends, and where to look for further information.

The symbiotic microbiota of the cow rumen are the backbone of the “world’s largest bioreactor” (Weimer et al., 2009), enabling the animal to efficiently utilize fibre-rich plant material for its energy needs. In a short space of time, advancements in culture-dependent and independent techniques have rapidly advanced our understanding and appreciation of the genetic and metabolic diversity that exists within the rumen microbiome. Since 2010, our knowledge of the different strategies that rumen microbiota employ for fibre-digestion has expanded from free-enzymes and cellulosomes to polysaccharide utilization loci in both gram negative and positive populations, outer membrane vesicles and T9SS-secreted multi-modular enzymes. Furthermore, new resources have recently been released that will enable further development in this discipline. For example, the sequencing and annotation of 410 rumen isolate genomes as part of the Hungate1000 has created an invaluable resource whereby one can quickly examine the CAZyme profiles of each entry and predict their fibrolytic potential and strategies (Seshadri et al., 2018). Concurrently, in the space of two years, metagenomic approaches used to reconstruct population genomes from rumen microbiota has gone from 100’s (Stewart et al., 2018b, Solden et al., 2018) of recovered genomes to 1000’s (Stewart et al., 2018a), which releases an immense amount of genomic data and CAZymes to be mined. It is highly likely that closer examination of these resources will uncover CAZymes and fibrolytic strategies that vary from those outlined in this chapter.
Despite the potential value of these genomic resources, they do not currently address an elemental shortcoming that rumen microbiology still needs to overcome, which is that our knowledge is built from well documented bacteria and archaea, whereas virtually nothing is known about the eukaryotic and viral populations. These under-represented facets of rumen microbiology are believed to contribute to digestion and enteric gas formation, but are poorly understood due to their unculturability and/or genome complexity. It is hoped that new technologies will help address the eukaryotic/viral challenge, such as long-read sequencing technology (Oxford nanopore (Stewart et al., 2018a)) and specific binning software VirSorter (Roux et al., 2015) and EukRep (West et al., 2018) that target uncultured virus (Emerson et al., 2018) and eukaryotes, respectively. In addition, there needs to be concerted efforts to incorporate axenic isolation and both biochemical and enzymological approaches into future efforts to deconvolute new fibrolytic mechanisms, as without hard biochemical evidence we cannot proceed beyond prediction and elucidate true metabolic function.

8. References


Krause, D. O., Bunch, R. J., Conlan, L. L., Kennedy, P. M., Smith, W. J., et al. 2001. Repeated ruminal dosing of *Ruminococcus* spp. does not result in persistence, but changes in other microbial populations occur that can be measured with quantitative 16S-rRNA-based probes. *Microbiology.*, 147, 1719-1729.


