Outer membrane vesicles from *Fibrobacter succinogenes* S85 contain an array of Carbohydrate-Active Enzymes with versatile polysaccharide-degrading capacity

Magnus Ø. Arntzen[a*], Anikó Várnai[a], Roderick I. Mackie[b], Vincent G. H. Eijsink[a], Phillip B. Pope[a]

a. Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway.
b. Institute for Genomic Biology, and Department of Animal Sciences, University of Illinois at Urbana-Champaign, Illinois, USA

*Address correspondence to Magnus Ø. Arntzen, magnus.arntzen@nmbu.no

Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, NORWAY. Tel: +47 67 23 24 46, Fax: +47 64 96 59 01

Running title: *F. succinogenes* outer membrane vesicles

Keywords: *Fibrobacter succinogenes* S85, outer membrane vesicle, proteomics, protein complex, polysaccharide degradation

Originality-Significance Statement

Outer membrane vesicles (OMVs) are gaining increasing attention for their role in pathogenesis and microbial ecology. OMVs provide a means to increase bacterial outreach since they allow delivery of degradation-protected biomolecules to the environment, at high local concentrations. Indeed OMVs are known to exert influences on eco-systems via horizontal gene transfer, biofilm formation, intra- and interspecies communication, and biomass degradation. Here we report that OMVs produced by *Fibrobacter succinogenes* are equipped with a diverse suite of enzymes able to depolymerize most common plant polysaccharides, including cellulose. Our data indicate that OMVs assist the metabolism of the host cell by deconstructing non-essential polysaccharides that restrict access to the host's target carbon source, cellulose. We also demonstrate that previously identified cellulose binding proteins are arranged in novel putative complexes in OMVs. Thus, *F. succinogenes* degrades biomass using means that differ fundamentally from well-known degradative machineries in Nature.

Summary

Fibrobacter succinogenes is an anaerobic bacterium naturally colonizing the rumen and cecum of herbivores where it utilizes an enigmatic mechanism to deconstruct cellulose into cellobiose and glucose, which serve as carbon sources for growth. Here, we illustrate that outer membrane vesicles (OMVs) released by *F. succinogenes* are enriched with carbohydrate-active enzymes and that intact OMVs were able to depolymerize a broad range of linear and branched hemicelluloses and pectin, despite the inability of *F. succinogenes* to utilize non-cellulosic (pentose) sugars for growth. We hypothesize that the degradative versatility of *F. succinogenes* OMVs is used to prime hydrolysis by destabilizing the tight networks of polysaccharides intertwining cellulose in the plant cell wall, thus increasing accessibility of the target substrate for the host cell. This is supported by observations that OMV-pretreatment of the natural complex substrate switchgrass increased the catalytic efficiency of a commercial cellulose-degrading enzyme cocktail by 2.4-fold. We also show that the OMVs contain a putative multiprotein complex, including the fibro-slime protein previously found to be important in binding to crystalline cellulose. We hypothesize that this complex has a function in plant cell wall degradation, either by catalyzing polysaccharide degradation itself, or by targeting the vesicles to plant biomass.

Introduction

Cellulose and hemicellulose are the most abundant components of plant biomass. These polysaccharides, although recalcitrant, do not accumulate on our planet due to their removal by the concerted action of highly specialized (hemi)-cellulose degrading microbes, including fungi and bacteria. These microorganisms exploit sophisticated enzyme systems to degrade plant material, and the enzymes involved in plant cell wall degradation have potential in biotechnological applications, such as in biofuel production (Himmel et al., 2010). In aerobic cellulolytic microorganisms, cellulose degradation is catalyzed by a consortium of mostly secreted enzymes including cellobiohydrolases, endoglucanases, β-glucosidases and lytic polysaccharide monooxygenases (LPMOs) (Horn et al., 2012; Mba Medie et al., 2012). The polysaccharide-degrading enzymes release soluble oligosaccharides and sugars that are transported into the cell and further metabolized. In contrast, some anaerobic cellulolytic bacteria form large multi-enzyme complexes referred to as cellulosomes, which often are bound to the outer surface of the cells (Bayer et al., 2004; Bayer et al., 2008). These complexes contain a backbone scaffoldin protein onto which several types of cellulases are docked via dockerin domains. The scaffoldin binds to cellulose primarily through family-3 carbohydrate-binding modules (CBMs), whereas substrate-affinity may be additionally tuned by CBMs attached to the cellulosomal enzymes. Recently, a third enzyme system, the Bacteroidetes-affiliated Polysaccharide Utilization Loci (PULs), has been described, which entails physically-linked genes organized around a signature SusCD-encoding gene pair (representing an outer membrane porin and a carbohydrate-binding protein, respectively). PULs seem to predominantly target soluble glycans, but PUL-based conversion of crystalline chitin has been shown (Larsbrink et al., 2016), and there are indications that uncultured rumen populations utilize PULs to degrade cellulose (Naas et al., 2014). In addition to these

strategies, there are examples of cellulolytic enzymes being attached directly to the peptidoglycan layer (such as in *Clostridium thermocellum* (Zhao et al., 2006)) or to cell surface polysaccharides (such as in *Ruminococcus albus* (Ezer et al., 2008)) of biomass-degrading bacteria.

One of the most highly specialized cellulose-degrading bacteria is *Fibrobacter succinogenes*, a strictly anaerobic, Gram-negative, rod-shaped bacterium. It is considered one of the major cellulolytic bacteria within the herbivore gut (Krause et al., 2003; Kobayashi et al., 2008) and has been the subject of extensive research due to its ability to adhere to and efficiently degrade plant cell walls. *F. succinogenes* does not produce cellulosomes, does not secrete high titers of cellulolytic enzymes, and its genome seems devoid of genes encoding known cellobiohydrolases and PULs (Suen et al., 2011). These observations suggest that *F. succinogenes* employs an alternative strategy for cellulose degradation. To understand why *F. succinogenes* is such a powerful biomass degrader, a number of endoglucanases, xylanases and cellulose-binding proteins have been cloned and characterized (see summary in (Toyoda et al., 2009)), without revealing particularly powerful enzymes. It has been suggested that outer membrane (OM) proteins are involved in cellulose degradation (Jun et al., 2007; Raut et al., 2015), but details remain ambiguous.

In 1981 it was discovered that *F. succinogenes* releases sedimentable membranous fragments into the culture fluid, which are able to hydrolyze carboxymethylcellulose (CMC) (Groleau and Forsberg, 1981). Subsequently, it was demonstrated that the membrane fragments are in fact vesicles originating from the outer membrane (OMV: outer membrane vesicle) that are produced during growth on cellulose (Forsberg et al., 1981). The OMVs showed a distinct and complex protein composition (Groleau and Forsberg, 1983) and were shown to exhibit both endoglucanase, xylanase and acetylesterase activity (Gong and Forsberg, 1993). These studies also showed that

96 the OMVs adhere to cellulose and are not produced during growth on glucose (Forsberg et al.,
97 1981; Gong and Forsberg, 1993; Burnet et al., 2015). The role of these OMVs in *F. succinogenes*
98 is currently debated; some claim that their production merely reflects aging of the cells i.e. a
99 stationary phase phenomenon (Gaudet and Gaillard, 1987), while others speculate that they have
100 a biological function in cellulose degradation (Forsberg et al., 1981). Interestingly, it was recently
101 shown that OMVs from *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* are equipped with
102 hydrolytic enzymes and are important in polysaccharide degradation (Elhenawy et al., 2014).
103 OMVs are spherical, bi-layered, membranous structures that are released naturally from the OM
104 of Gram-negative bacteria (Beveridge, 1999). They are typically between 10-300 nm in diameter
105 and contain phospholipids, liposaccharides, OM proteins and proteins from the periplasmic space.
106 OMVs have been observed in a wide range of Gram-negative species grown in different
107 environments and under various growth conditions (see references in (Kulp and Kuehn, 2010)).
108 They have been suggested to play wide-ranging roles in microbial ecology (e.g. horizontal gene
109 transfer, biofilm formation, communication and biomolecule delivery) and can be numerically far
110 more abundant than the organisms themselves (Elhenawy et al., 2014; Roier et al., 2016)).

111 In this study, we have isolated and studied the content of OMVs produced by *F. succinogenes*
112 during growth on crystalline cellulose. We used proteomics to identify the proteins in the OMVs
113 and show they are enriched in polysaccharide-degrading enzymes. Importantly, we demonstrate
114 the presence of a novel putative multiprotein complex, comprising several proteins known to be
115 involved in interactions with cellulose, that could be a driver of polysaccharide degradation.
116 Activity assays showed that the OMVs are able to depolymerize a broad range of hemicelluloses
117 in addition to cellulose, and use of OMVs as pretreatment of a natural grass substrate (switchgrass)

enabled a 2.4-fold increase in downstream saccharification. The results add support to the hypothesis that *F. succinogenes* actively uses OMVs to convert biomass.

Results

OMVs produced by *F. succinogenes* S85 vary in size and are equipped with carbohydrate-active enzymes

Similar to other Gram-negative bacteria and according to previous reports, *F. succinogenes* S85 produces OMVs (Forsberg et al., 1981; Burnet et al., 2015), but currently little is known about their specific nature and enzyme contents. To isolate OMVs, we employed a series of microfiltration and ultra-centrifugation steps and vesicles were obtained as a broad, strong band in a sucrose gradient, with an average density of 1.13 g/mL. The band broadness suggested a heterogeneous size distribution, which was confirmed by dynamic light scattering experiments that indicated a population ranging from 8-136 nm in radius, with an average of 49 nm (Figure S1A). Transmission electron microscopy (TEM) confirmed that the OMV preparation contained vesicles (Figure S1B).

Using quantitative proteomics, we detected 347 proteins in the OMVs covering a range in abundance of four orders of magnitude (Table S1) and with high reproducibility between biological replicates (Pearson correlation R = 0.805) (Figure S2). Using an algorithm for predicting signal peptides, lipoprotein signal peptides and transmembrane helices (LipoP; see Supplementary Text S2), 79% of the detected proteins were predicted to be associated with the extracellular milieu. In particular, 50% harbored a SpI signal peptide, 28% an SpII lipoprotein signal peptide, and 1% contained a transmembrane helix. The remaining 21% were predicted to be cytosolic

proteins. We performed functional annotation of the complete proteome of *F. succinogenes* (2871 protein sequences) via protein searches and categorical classification using the NCBI Conserved Domain Database (NCBI Web-CD) and the database of Clusters of Orthologous Groups (COG) of proteins (see Supplementary Text S2). This analysis revealed that the OMVs showed a higher proportion of proteins in the COG-category 'carbohydrate transport and metabolism', which covered 12% of the OMV proteome, compared to 4% in the complete proteome (Figure S3).

Analysis of the OMV proteins using dbCAN, a specialized database for prediction of carbohydrate-active enzymes (CAZymes; (Yin et al., 2012)), showed that 21% of the OMV proteins (i.e. 74 of the 347 proteins) had predicted carbohydrate-active functions (Figure 1A). Comparing these numbers with predicted extracellular proteins in *Fibrobacter* (992 proteins with either SpI or SpII cleavage sites or containing a TMH, according to LipoP, of which 116 are CAZymes), suggests an enrichment of carbohydrate-active enzymes in the OMVs (Fisher's Exact p-value 1.36E-5). Forty-eight were classified as glycoside hydrolases (GHs), two as glycosyl transferases (GTs), five as polysaccharide lyases (PLs) and 13 as carbohydrate esterases (CEs), while no auxiliary activities (AAs) were identified. In addition, we found six proteins that contain a carbohydrate-binding module (CBM), but lack a catalytic domain with a known carbohydrate-active function. Figure 1A shows these 74 proteins plotted against their relative abundance in the OMV proteome. The most abundant protein (FSU_2303) belongs to the GH family 8 and could be responsible for hydrolyzing the backbone of cellulose and xylan. Amongst the 50 most abundant proteins in the OMVs, there are seven CAZymes (one GH8, two GH9 and four GH5; Table S1). The six CBM-only proteins show similar abundances as the catalytic CAZymes, and one of these, a CBM11 (FSU_2007), is highly abundant. These proteins could be interesting to investigate further for the presence of hitherto unknown carbohydrate-active catalytic domains.

To look further into the enrichment of certain proteins in the OMVs, we performed an enrichment analysis using Pfam, a tool for predicting functional domains in proteins. First, we counted the occurrence of all the Pfam domains in the complete proteome of *F. succinogenes* and then compared these values to similar values for the OMVs. Using Fisher's Exact test to calculate the significance of enrichment, we detected 18 domains to be overrepresented in the OMVs, half of which were CAZyme-domains (Table 1, Figure S4). The most frequent of the enriched Pfam domains was the family-6 CBM (PF03422), which is known to target amorphous cellulose or xylan. In the OMV proteins, this module is found associated with GH5 endoglucanases (PF00150), GH43 (PF04616) and GH30 (PF17189) hemicellulases, and a sialic acid-specific acetylesterase (PF03629), indicating involvement in degradation of both cellulose and hemicellulose. Another enriched CBM, the family-11 CBM (PF03425), is known to target amorphous cellulose and is appended to a GH51 endoglucanase domain (e.g. in FSU_0382) or to a GH5 endoglucanase domain (e.g. in FSU_2914) or occurs as a single domain protein (FSU_2007). In addition to CBMs, several endoglucanases (GH5 and GH9: PF00759) and hemicellulase (GH16: PF00722, GH30 and GH43) domains were enriched, indicating a potential role of vesicles in delivering carbohydrate-active enzymes to the substrate. Notably, the analysis of Pfam domains revealed the enrichment of several non-carbohydrate-active domains, some of which are potentially involved in carbohydrate-binding or metabolism, as discussed below.

OMVs are active on a wide range of plant-derived substrates

To explore the actual enzymatic activity present in the OMVs, we incubated the purified vesicles with nine different substrates: phosphoric acid swollen cellulose (PASC) made from Avicel, tamarind xyloglucan, cabbage pectin, wheat arabinoxylan, birchwood xylan, aspen xylan, ivory nut mannan, carob galactomannan and konjac glucomannan. The products formed by substrate

hydrolysis were identified by LC-MS using a library of *m/z*-time tags (combination of measured mass and retention time) established on a high-sensitive mass spectrometer connected to a HPLC. The dbCAN analysis (Figure 1A) predicted OMV proteins that target these abovementioned substrates, namely: endoglucanases (e.g. GH5s), xyloglucanases (e.g. GH74s), pectin lyases (e.g. PL1s), endo-xylanases (e.g. GH11s and GH43s linked to xylan-binding CBM6s) and mannanases (e.g. GH26s). In accordance with the prediction, we detected formation of oligosaccharide products from each substrate (Figure 1B:I-IX, details in Table S2), indicating that the OMVs are able to degrade the plant cell wall polysaccharides tested.

Fresh forages, including green leaves and stems, are commonly found in the rumen of pasture fed ruminants, the natural habitat for *F. succinogenes*. These are rich in primary cell walls which are mainly composed of cellulose, xyloglucan and pectin, where the two latter polysaccharides cross-link cellulose microfibrils (Park and Cosgrove, 2015). The activity assays with tamarind xyloglucan and cabbage pectin revealed that the OMVs contain enzymes that are able to break these polymers, which theoretically would yield improved access to cellulose, the breakdown products of which serve as the main carbon source for growth of *F. succinogenes*. The OMV proteins cleaved xyloglucan not only into its repeating units (cellotetraose backbone with three xylosyl substitution, e.g. $Hex_{4-5}Pen_3$; Hex: hexose, Pen: pentose) but also into fragments with a shorter backbone (e.g. $Hex_{1-4}Pen_1$, and $Hex_{2-3}Pen_2$; Figure 1B:IX). The occurrence of xyloglucan oligosaccharides carrying less than three pentose units (most likely xylosyl substitutions) indicates cleavage of the xyloglucan backbone between two substituted glucosyl units. This unique cleavage pattern has only been shown for a handful of enzymes belonging to the GH74 and AA9 families so far (Desmet et al., 2007; Feng et al., 2014; Kojima et al., 2016; Nekiunaite et al., 2016) and

could potentially be attributed to FSU_2866, an OMV protein annotated as a BNR repeat protein and predicted to harbor four GH74 modules (Table S1).

Incubation of cabbage pectin, a mixture of homogalacturonan (partly methyl esterified polygalacturonic acid) and rhamnogalacturonan type I (a rhamnose-galacturonic acid copolymer substituted with arabinogalactan side chains), with the OMVs led to fragmentation of various structural elements of pectin (Figure 1B:VII). The formation of galacturonic acid oligosaccharides containing an unsaturated galacturonic acid revealed the cleavage of homogalacturonan by β-elimination with a pectate lyase. In the OMVs, five proteins with polysaccharide lyase domains were identified (belonging to PL families 1, 9 and 22), of which one has been identified as being potentially active on pectin (FSU_0577, putative pectate lyase) (Table S1). OMVs could also depolymerize the arabinogalactan side chains of rhamnogalacturonan moieties. While we did not identify rhamnose-containing oligosaccharides (indicative of cleavage of the rhamnogalacturonan backbone), oligosaccharides that are likely to originate from the arabinogalactan side chains were observed (Hex_{2-4}, Pen_{2-14}, $Hex_{3-5}Pen_2$, corresponding to Gal_{2-4}, Ara_{2-14}, $Gal_{3-5}Ara_2$, respectively; Gal: galactose, Ara: arabinose). The OMV proteins performing this action could be FSU_3024 (identified as a GH53 arabinogalactan endo-β-1,4-galactanase), FSU_0145 (a GH43 arabinosidase) and FSU_2288 (a GH2 β-1,4-galactosidase).

The OMVs were also active on cellulose releasing cellobiose, cellotriose and cellotetraose from PASC (Figure 1B:VIII). The OMVs were able to depolymerize close to 70% of the PASC within 24 hours at a reasonable enzyme loading (2.4 mg OMV proteins with predicted carbohydrate-active function was loaded per g cellulose). The initial depolymerization rate was 1.8 U/mg/min (i.e. one mg enzyme releases 1.8 µmol reducing end sugars during one minute incubation); using the same conditions, the activity of the commercial enzyme cocktail Celluclast by Novozymes

(Bagsvaerd, Denmark) was determined to be 2.5 U/mg/min. For more details, see Experimental Procedures.

The OMVs were also capable of depolymerizing substituted hemicelluloses. The most common hemicelluloses in grasses, commonly fed to ruminants, are branched xylans. In reactions with arabinoxylan (Figure 1B:VI), where the β-1,4-xylan backbone is 3-*O*-mono- or 2,3-*O*-disubstituted with α-L-arabinose, we observed a range of oligosaccharides with a degree of polymerization (DP) up to 10. Although elution times indicated hydrolysis products were not linear oligosaccharides, we were unable to identify arabinosylation patterns of the released oligosaccharides, because arabinose and xylose have the exact same mass and are undistinguishable by mass spectrometry. The OMVs were also active on xylans with different substituting groups that are more common in woody plant cell walls. In the reactions with birchwood xylan (Figure 1B:IV), we detected three types of xylo-oligosaccharides: linear, substituted with 4-*O*-methyl-glucuronic acid and substituted with glucuronic acid. Reactions with aspen xylan (Figure 1B:V) showed release of xylooligosaccharides carrying methyl-glucuronyl and/or acetyl groups.

The OMVs were also shown to target mannans with various backbone and substitution patterns. The fact that the OMVs were able to depolymerize ivory nut mannan (Figure 1B:I) (a linear mannose homopolymer) to mannooligosaccharides shows the presence of true mannanases that can cleave β-1,4-linkages between two mannose units in the polymer backbone. Galactosylation (as in carob galactomannan) of the mannan backbone did not prevent depolymerization (Figure 1B:II) and yielded both linear (nongalactosylated) mannooligosaccharides and galactosylated oligosaccharides. From konjac glucomannan (acetylated glucomannan), the OMVs released both

253 cello- and mannooligosaccharides and a range of glucomannan oligosaccharides (Figure 1B:III).
254 In addition, mono- and diacetylated oligosaccharides were detected.

255 Finally, as *Fibrobacter succinogenes* grows exclusively on cellulose, which in plant cell walls is
256 embedded in a hemicellulose and pectin network, we hypothesized that a potential function for
257 OMVs (carrying hemicellulose- and pectin-degrading enzymes) could be to increase the
258 accessibility of cellulose in grasses in the rumen by loosening up the pectin-hemicellulose matrix
259 localized around the cellulose fibers. To test this, we selected a milled and washed switchgrass
260 substrate and compared its degradability with a commercial enzyme cocktail with and without
261 pretreatment with OMVs. Pretreatment with OMVs had a significant effect on saccharification of
262 switchgrass by the commercial enzyme cocktail, leading to a 2.4-fold increase in the solubilized
263 sugar yield as compared to when the commercial enzymes were acting alone (Figure 2A). MS
264 analysis of the products formed during switchgrass degradation revealed that even though the
265 commercial enzyme mixture was able to depolymerize both pentose and hexose-containing sugars,
266 auxiliary enzyme activities were present in the OMVs leading to additional products formed
267 (Figure 2B). Perhaps the most important difference is the formation of uGalA-GalA2 (GalA:
268 galacturonic acid; u: unsaturated) and acetylated oligosaccharides containing both hexose and
269 pentose units. The latter most likely originate from plant xyloglucan since only arabinogalactan of
270 the other hemicelluloses contain both sugar types and arabinogalactan has not been shown to carry
271 any acetyl groups. The appearance of these compounds suggests that the OMVs were able to
272 hydrolyze the homogalacturonan backbone in pectin using lyase activities (hence the unsaturated
273 galacturonic acid) as well as heavily substituted (acetylated and probably also fucosylated)
274 xyloglucans, and that they thus likely open up the intertwined pectin-hemicellulose-cellulose
275 network. These observations may explain why OMV-pretreatment increases the saccharification

of switchgrass by the commercial enzyme cocktail and strengthen the hypothesis that the primary role of OMVs may be to provide *F. succinogenes* better access to cellulose.

OMVs also contain putative multiprotein complexes

To detect potential protein complexes amongst the OMV proteins, we utilized high-resolution clear native electrophoresis (hrCNE). This technique, which takes advantage of mixed micelles to stabilize proteins and convey a negative net charge, has proven to separate equally well compared to blue native electrophoresis, while being superior for downstream catalytic activity assays (Wittig et al., 2007). In combination with SDS-PAGE, it is possible to generate two-dimensional gels in which the protein complexes separated in the first dimension (hrCNE) are separated into single protein spots in the second dimension. Proteins originating from the same complex will fall on a straight vertical line. Figure 3A shows such a 2D-hrCN-SDS-PAGE separation of 40 µg OMV proteins using a 6.5% native gel and 10% SDS-gel. 15 protein spots were selected for proteomics analysis (Table 2). Three putative protein complexes can be seen: complex C1: spot number 3, 4 and 5, complex C2: spot number 6, 7, 8 and 9, and complex C3: spot number 11, 12, 13 and 14. The C2 and C3 putative complexes seem to contain at least some identical proteins (Figure 3) as was indeed confirmed by the proteomic analysis (Table 2). The main difference between the putative complexes is the lack of spot number 9 in C3. This may indicate that the complex could exist in two variants, with or without the protein(s) in spot 9, or that a part of the complex was lost during sample preparation. Considering only the most abundant proteins in each spot, complex C1 consists of two proteins with no predicted functional domains (FSU_1029, FSU_2008) and one OmpA family protein (FSU_2078) harboring a C-terminal OmpA-like domain and five thrombospondin type 3-like repeats, which are known to bind calcium (Kvansakul et al., 2004). Complex C2 consists of four proteins, two OmpA family proteins (FSU_2396, FSU_2078), a

299 tetratricopeptide repeat (TPR) domain protein (FSU_2397) and a fibro-slime domain protein
300 (FSU_2502). Spot 9 was broad and dense, and found to contain many proteins (Table S3),
301 including several endoglucanases. It is not possible to judge whether all these spot 9 proteins are
302 part of the C2 complex. Regardless, the emPAI values clearly show that the fibro-slime domain
303 protein is the dominating protein in spot 9. Notably, we have consistently observed spot 9 to co-
304 occur with spots 6, 7 and 8, independent of the acrylamide percentage in the first dimension (data
305 not shown); this indicates a true association of the proteins in these spots. Complex C3 seems to
306 be a fragment of C2, containing only two of the proteins, the OmpA family protein (FSU_2396)
307 and the TPR domain protein (FSU_2397). Strikingly, these two proteins, which are partners in
308 both complex C2 and C3, are neighboring genes located in an operon, according to the Database
309 of prokaryotic operons (DOOR; (Mao et al., 2009)), and show co-expression with high abundance
310 in the OMV total data set (Table S1).

311 To assess the carbohydrate degrading capabilities of these putative complexes, we used another
312 lane from the native gel, identical to the one used for the SDS-PAGE separation, and divided it
313 into seven fractions as indicated on the top of Figure 3A. The gel pieces were ground using a pestle
314 and mortar and then incubated with PASC for detection of enzyme activity. The products were
315 analyzed using PGC-MS, and the amounts of the different oligosaccharide products were
316 determined (Figure 3B). All fractions, except fraction VII gave release of cello- and
317 xylooligosaccharides from PASC. The first two fractions (I and II) released oligosaccharides to a
318 low extent, suggesting that complex C1 has a limited role in cellulose degradation. Fractions III -
319 VI, including complexes C2 and C3, all produced high amounts of oligosaccharides. Notably,
320 separation is not optimal due to horizontal streaking in the first dimension, meaning that it is
321 impossible to assign activities to particular protein complexes or individual proteins. It is

interesting to note that Fraction V, lacking the fibro-slime protein seems less active on cellulose. No products were detected in fraction VII, indicating that this protein, *F. succinogenes* major paralogous domain protein (FSU_2794), is not able to degrade PASC under these conditions.

Discussion

OMVs are formed by membrane blebbing, followed by release of spherical outer membrane vesicles, which enclose a fraction of the periplasmic space. Vesiculation appears to be a common phenomenon for Gram-negative bacteria (Beveridge, 1999; Roier et al., 2016), suggesting an important physiological role for this process (Kulp and Kuehn, 2010). OMVs represent a confined transportable environment where enzymes, virulence factors or other molecules are protected from inhibitors and can be present in high concentrations (Biller et al., 2014). Recent reports have shown that OMVs from *B. fragilis* and *B. thetaiotaomicron* are selectively packaged with acidic hydrolases and proteases compared with the outer membrane, which contains more alkaline proteins (Elhenawy et al., 2014). This suggests a sorting mechanism that could be pI related. Interestingly, the vast majority (79%) of the 347 OMV proteins detected in *F. succinogenes* were also acidic (pI < 7; for the whole proteome this fraction was 64%). Further, we used Pfam-based analysis to detect domains that were overrepresented in the OMVs. In this analysis, we compared the domains present in the OMV proteome to the whole cell's proteome and found that half of the domains that were statistically enriched (Fisher's Exact p-value < 0.05) were CAZyme-domains. It is important to note however, that comparing the OMV proteome to the whole cell's proteome is not necessarily a fair comparison as OMV proteins are expected to be biased to the outer membrane, and enriching for extracellular proteins *de facto* does select for CAZymes (as enzymes taking part in polysaccharide-degradation are almost exclusively extracellular). To account for this bias, we compared the fraction of CAZymes present in the OMVs (21%) with that in the predicted

345 extracellular proteome of *F. succinogenes* (12%). This supported our hypothesis that CAZymes
346 are enriched in the OMVs (Fisher's Exact p-value 1.36E-5) and suggests a selective packaging of
347 carbohydrate-active enzymes into OMVs.

348 Reports have shown that *F. succinogenes* releases OMVs during growth on cellulose, but not
349 during growth on glucose (Forsberg et al., 1981; Burnet et al., 2015). Producing OMVs is an
350 energy-demanding task for the bacteria, and given the ubiquitous presence of OMVs across Gram-
351 negative species, the selective sorting of acidic proteins and the enrichment of carbohydrate-
352 metabolizing proteins observed here, it is reasonable to assume that the OMVs have important
353 biological functions.

354 *F. succinogenes* is widely known to efficiently hydrolyze the variety of plant polysaccharides it
355 encounters with in the rumen. The current data shows that OMVs produced by *F. succinogenes*
356 are equipped with enzymes targeting these polysaccharides, in accordance with previous
357 observations (Gong and Forsberg, 1993). OMVs were capable of hydrolyzing nine different
358 isolated plant polysaccharides, but also showed activity on a more complex and natural substrate,
359 switchgrass, whereby OMV-pretreatment increased the efficiency of a commercial cellulase
360 cocktail 2.4-fold. We speculate this is due to complementary enzyme activities present in the
361 OMVs that enhance cellulose accessibility. The promiscuous activity of the OMVs towards plant
362 polysaccharides that are embedded with cellulose (the sole carbon source of the host), suggest that
363 a primary role of OMVs could be to provide *F. succinogenes* better access to cellulose.

364 An analysis of the most enriched protein families in the OMVs revealed several without a CAZyme
365 annotation, yet with high abundance in the OMVs (Table 1). Some of these domains have
366 properties that suggest potential involvement in carbohydrate binding or metabolism. This includes
367 the PA14 domain, a hypothesized carbohydrate-binding module found in a wide variety of

enzymes including glycosidases, and the sulfatase-modifying factor enzyme, which belongs to the lectin-like superfamily. Furthermore, type IV pilin proteins and cadherins were highly abundant in the OMV proteome. A detailed discussion on these domains and their potential contribution to carbohydrate binding or metabolism is provided in Supplementary Text S1.

It has been well documented that *F. succinogenes* does not utilize any of the known polysaccharide-degrading assemblages (i.e. cellulosomes or PULs) (Suen et al., 2011). In this study, we observed high levels of TPR domain proteins in the OMVs, a protein class also observed by others in the outer membrane (Jun et al., 2007; Raut et al., 2015). TPR proteins are commonly found in protein complexes, where multiple TPR domains (three in FSU_2397) have been shown to form a super-helix exposing several binding surfaces that promote formation of multiprotein complexes (Zeytuni and Zarivach, 2012). TPR proteins are consequently believed to act as scaffold proteins (Blatch and Lassle, 1999). This led us to investigate if multiprotein complexes were present in the OMVs. Our analyses revealed the presence of at least three putative multiprotein complexes in the OMVs, two of which, C2 and C3, seemingly degraded PASC. The four main components of these two complexes (FSU_2078, FSU_2502, FSU_2396 and FSU_2397) are all predicted to be secreted. Both putative complexes lack known glycoside hydrolases among their main "highly-detectable" components, although proteomic analysis detected hydrolytic enzymes in the samples, either as "contaminations" or as less abundant parts of the complexes. Interestingly, all four main proteins identified in these putative complexes have previously been detected on the outer membrane of *F. succinogenes*, and accumulating data indicate that they play a role cellulose binding (Gong et al., 1996; Jun et al., 2007; Raut et al., 2015). The abundantly present fibro-slime domain protein (FSU_2502), previously referred to as the 180-kDa cellulose-binding protein, is known to have an important role in cellulose binding (Gong et al., 1996; Suen et al., 2011). Hence,

391 it is likely that this protein helps targeting the vesicles to plant biomass. Notably, the four main
392 proteins in C2 and C3 together contain hypothetical regions summing up to approximately 3000
393 amino acids with unknown functions, which could include hitherto unknown hydrolytic enzymes.
394 In particular, the FSU_2396 OmpA protein contains a beta-helix domain similar to that seen in
395 pectate lyases. In *F. succinogenes*, this domain (Pfam PF13229) is found in only one other protein
396 (FSU_2273), a pectate lyase with a family-6 CBM, also detected in the OMVs. Figure 4A shows
397 the domain organizations for the four proteins involved in complexes C2 and C3, whereas Figure
398 4B depicts an artist impression of a putative OMV-associated complex acting on the substrate.

399 Interestingly, in 2009, Toyoda and colleagues identified cellulose-binding proteins in rumen fluid
400 from sheep through enrichment with crystalline cellulose (Toyoda et al., 2009). The authors
401 detected four proteins belonging to *F. succinogenes*: a TPR domain protein (FSU_2397), a fibro-
402 slime domain protein (FSU_2502), an OmpA family protein (FSU_2396) and cellulose binding
403 protein (FSU_0382). Except from the latter (which we did detect in the OMVs), these proteins are
404 part of complex C2. These observations considered collectively with earlier reports of the
405 importance of these proteins for cellulose binding (Gong et al., 1996; Jun et al., 2007) and the
406 operon structure of the genes encoding FSU_2396 and FSU_2397 indicate that C2 is a real
407 complex with an important role in biomass conversion.

408 In conclusion, *F. succinogenes* is equipped with a surprisingly high diversity of polysaccharide-
409 degrading enzymes and abilities, considering that the bacterium only utilizes one such
410 polysaccharide, cellulose, as a carbon source. The observation that *F. succinogenes* packages many
411 of these enzymes into OMVs that are released as “degrading drones” makes the bacterium even
412 more peculiar. The exact role of OMV formation needs to be further explored for several key
413 purposes, including the identification of signals that trigger OMV biogenesis, to understand the

impact of OMV formation on *F. succinogenes* fitness, and to explore syntrophic OMV interactions with other microbes. To this end, it is interesting to note previously observed OMV-related syntrophic interactions in the human intestine, specifically between OMV-producing polysaccharide-degrading bacteria and bacteria unable to grow on the specific polysaccharide alone (Rakoff-Nahoum et al., 2014). We hypothesize that *F. succinogenes* secretes the OMVs when grown on cellulose to degrade the surrounding hemicellulose, thus making cellulose more accessible. In electron micrographs presented by Burnet and colleagues (Figure 7D in (Burnet et al., 2015)), the OMVs are not found between the cells and the cellulose fibers, but rather distant from the cells, supporting the idea that the vesicles are paving the way for the bacterium. This idea is strengthened by our observation that the OMVs make switchgrass more susceptible to degradation by a commercial cellulose cocktail. Perhaps OMVs are useful tools for industrial biomass saccharification and/or as agents in mild biological biomass pretreatment methods. We predict that further studies on the OMVs from *F. succinogenes* will improve our understanding of the lifestyle of this enigmatic microbe, including its ability to efficiently degrade plant cell walls.

Experimental Procedures

Culture conditions and isolation of OMVs

Fibrobacter succinogenes S85 (ATCC 19169) cultures were grown statically at 37°C under anaerobic conditions, in the medium recommended by ATCC (ATCC medium 1943). Details of the medium can be found in Supplementary Text S2.

For isolation of OMVs, 800 mL cultures were grown. After 24 hours, the cultures were harvested by centrifugation at 9,000 × *g* for 15 minutes. The supernatant was filtered (0.45µm) and concentrated to 100 mL using a Vivaflow 200 cartridge with 10 kDa cut off (Sartorius AG,

Goettingen, Germany) and further down to 4 mL using a centrifugal concentrator with 100 kDa cut off (Pall Life Sciences, Ann Arbor, MI, USA). The retentate was centrifuged at 16,600 × *g* for 20 minutes to remove any debris. After a second filtration (0.45 µm), the supernatant was layered on top of a sucrose gradient and centrifuged at 200,000 × *g* for 3 hours. The brown, strong band, containing the OMVs, was extracted using a needle and syringe, diluted to 12 mL with 10 mM sodium acetate buffer (pH 6.0) containing 100 mM NaCl and re-centrifuged at 100,000 × *g* for 1 hour. The supernatant was discarded and the pellet (containing the OMVs) was collected and resuspended in 10 mM sodium acetate buffer (pH 6.0). The protein concentration in the OMV preparation was measured using Bradford protein assay and the OMVs were analyzed for size and purity using dynamic light scattering (DLS) and transmission electron microscopy (TEM). For details, see Supplementary Text S2.

Native and SDS-PAGE gels

For native gel electrophoresis, we prepared a 6.5% resolving (Tris/HCl pH 8.8, polyacrylamide) gel with a 5% stacking (Tris/HCl pH 6.8, polyacrylamide) gel. The anode buffer consisted of 25 mM Tris/HCl buffer (pH 8.3) containing 192 mM glycine, while the cathode buffer contained in addition 0.02% *n*-dodecyl-β-D-maltoside (DDM) and 0.05% sodium deoxycholate (DOC). The mixed micelles formed by the non-ionic detergent DDM and the anionic detergent DOC has been shown to stabilize membrane proteins while also providing a negative charge on the proteins (hence the anionic detergent), resulting in high-resolution clear native electrophoresis at pH 8.3, even for alkaline proteins (Wittig et al., 2007). Samples were prepared in a sample buffer (pH 8.3) containing 10% glycerol, 0.001% ponceau S, 50 mM NaCl, 25 mM Tris/HCl, and 40 µg OMV proteins were loaded per lane. Electrophoresis was performed at 4 °C and 200 V for 50 minutes. For 2D-hrCN-SDS-PAGE, a homemade 10% resolving SDS-gel were prepared and a lane already

459 separated under native conditions (above) were excised and placed 10 mm above the SDS-gel. A
460 5% stacking gel were poured around the native lane so this would be embedded into the stack.
461 Electrophoresis was performed at 240 V for 20 minutes and the gels were then stained with
462 Coomassie Brilliant Blue R250. 15 spots (gel pieces; see Figure 3A) were excised and destained
463 twice using 25 mM ammonium bicarbonate in 50% acetonitrile. The proteins entrapped in the gel
464 pieces were reduced and carbamidomethylated using 10 mM DTT and 55 mM iodacetamide,
465 respectively, prior to in-gel digestion with trypsin as described previously (Arntzen et al., 2015).
466 Prior to mass spectrometry, peptides were desalted using C_{18} ZipTips (Merck Millipore,
467 Darmstadt, Germany), according to manufacturer's instructions.

468 For proteomic analysis of total OMVs, two biological replicates were used. 50 µg of protein were
469 dissolved in SDS sample buffer, separated by SDS-PAGE using an AnyKD Mini-PROTEAN gel
470 (Bio-Rad Laboratories, Hercules, CA, USA) and stained using Coomassie Brilliant Blue R250.
471 The gel was cut into eight slices and the slices were processed as described above.

472 Proteomics and bioinformatics analysis

473 Peptides were analyzed using a nanoLC-MS/MS system (Dionex Ultimate 3000 UHPLC; Thermo
474 Scientific, Bremen, Germany) connected to a Q-Exactive mass spectrometer (Thermo Scientific,
475 Bremen, Germany) and operated in data-dependent mode to switch automatically between
476 orbitrap-MS and higher-energy collisional dissociation (HCD) orbitrap-MS/MS acquisition. MS
477 raw files were analyzed using MaxQuant (Cox and Mann, 2008) and identifications were filtered
478 in order to achieve a protein false discovery rate (FDR) of 1%. Only proteins identified in both
479 biological replicates were considered true OMV proteins. For analysis of gel spots, we used the
480 Mascot search engine (Perkins et al., 1999) to provide protein identifications. For further details

481 on the proteomics methods and for bioinformatics (LipoP prediction, COG and Pfam analysis),
482 see Supplementary Text S2.

483 Enzymatic assays and PGC-MS analysis

484 To estimate the efficiency of depolymerization of the OMV preparation, 10 mg OMV proteins
485 (corresponding to 2.4 mg carbohydrate-active enzymes based on the proteomics abundance
486 measurements) were loaded per g of PASC. The reaction was carried out in 50 mM sodium acetate
487 buffer, pH 6.0, for 48 hours in triplicates; samples were taken after 1, 4, 24 and 48 hours. After
488 sampling, H_2SO_4 was added (4% final concentration), and the samples were autoclaved for 60 min
489 at 121 °C to hydrolyze the oligosaccharides to monosugars (Sluiter et al., 2006). The sugar yield
490 was measured as reducing sugars using 3,5-dinitrosalicylic acid (Miller, 1959). The activity (i.e.
491 initial rate) was calculated based on the total reducing sugars at 1 hour and expressed as U/mg/min.

492 Enzymatic assays with OMVs were done using 20 µg OMV proteins and 1% (w/v) substrate in 10
493 mM sodium acetate buffer (pH 6.0). Nine different substrates were used: phosphoric acid swollen
494 cellulose (PASC), birchwood xylan, wheat arabinoxylan, aspen xylan, ivory nut mannan, carob
495 galactomannan, konjac glucomannan, tamarind xyloglucan and pectin. PASC from Avicel, pectin
496 from white cabbage and aspen xylan (isolated under mild conditions to avoid autohydrolysis of
497 the acetyl groups during the isolation process (Biely et al., 2013)) were prepared as described
498 earlier (Wood, **1988;** Westereng et al., 2009; Biely et al., 2013), birchwood xylan was purchased
499 from Roth (Karlsruhe, Germany) and all other substrates were purchased from Megazyme
500 (Wicklow, Ireland). Enzyme reactions were performed overnight at 40 °C and supernatants
501 containing soluble products were collected by centrifugation at 16,600 × *g* for three minutes.
502 Control reactions showed that no substrate depolymerization occurred upon incubation of the
503 substrates in buffer, at 40 °C, in the absence of OMVs, except for ivory nut mannan, carob

galactomannan and konjac glucomannan. In these cases, the signal obtained in control reactions were used for background subtraction of the samples.

The products were analyzed using a HPLC system (Dionex Ultimate 3000RS UHPLC; Thermo Scientific, Bremen, Germany) equipped with a porous graphitic carbon (PGC) column (Hypercarb) and connected to an LTQ-Velos Pro ion trap mass spectrometer (Thermo Scientific, Bremen, Germany). Product identification was achieved using *m/z*-values provided by the Velos Pro mass spectrometer or, in ambiguous cases, a mixture of retention time and *m/z*-values. For details on the analysis of products, see Supplementary Text S2.

Assessment of OMVs for biomass pretreatment

Switchgrass, obtained from The Noble Foundation, Ardmore OK, USA, was ball milled at 350 rpm in consecutive series of 10 minutes on and 15 minutes off to keep the temperature <50 °C. After 1.5 hours, the ball milled switchgrass was washed two times with water to remove background color and any soluble sugars prior to usage. Pretreatment assays were done using this ball milled, washed switchgrass at 0.2% (w/v) with 20 µg OMVs for 17.5 hours in 50 mM sodium acetate buffer (pH 6.0). Celluclast (mainly cellulase activity) and Novozym 188 (mainly β-glucosidase activity), both purchased from Novozymes (Bagsvaerd, Denmark), were prepared as a mixture in the ratio 4:1 (w/w) for enzymatic degradation of switchgrass. Enzyme reactions (after pretreatment) were performed by adding 20 µg enzyme cocktail to the above conditions and further incubate for four hours at 40 °C. Supernatants containing soluble products were collected by centrifugation at 16,600 × *g* for three minutes. The products were analyzed as reducing sugars using 3,5-dinitrosalicylic acid as reagent (Miller, 1959).

Funding information

MØA and PBP were supported by the European Research Council through grant 336355 ("MicroDE").

Acknowledgements

The authors would like to thank Dr. Roger Scherrers at Wyatt (Dernbach, Germany) for analysis of vesicles using dynamic light scattering and Dr. Bjørge Westereng (NMBU, Norway) for helpful discussions. The imaging was performed at the Imaging Centre Campus Ås, Department of Plant Sciences, NMBU, Norway. The proteomics data has been deposited to the ProteomeXchange consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al., 2013) with the dataset identifier PXD005442.

References

Arntzen, M.O., Karlskas, I.L., Skaugen, M., Eijsink, V.G., and Mathiesen, G. (2015) Proteomic Investigation of the Response of *Enterococcus faecalis* V583 when Cultivated in Urine. *PLoS One* **10**: e0126694.

Bayer, E.A., Belaich, J.P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* **58**: 521-554.

Bayer, E.A., Lamed, R., White, B.A., and Flint, H.J. (2008) From cellulosomes to cellulosomics. *Chem Rec* **8**: 364-377.

Beveridge, T.J. (1999) Structures of gram-negative cell walls and their derived membrane vesicles. *J Bacteriol* **181**: 4725-4733.

Biely, P., Cziszarova, M., Uhliarikova, I., Agger, J.W., Li, X.L., Eijsink, V.G., and Westereng, B. (2013) Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. *Biochim Biophys Acta* **1830**: 5075-5086.

Biller, S.J., Schubotz, F., Roggensack, S.E., Thompson, A.W., Summons, R.E., and Chisholm, S.W. (2014) Bacterial vesicles in marine ecosystems. *Science* **343**: 183-186.

Blatch, G.L., and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* **21**: 932-939.

Burnet, M.C., Dohnalkova, A.C., Neumann, A.P., Lipton, M.S., Smith, R.D., Suen, G., and Callister, S.J. (2015) Evaluating Models of Cellulose Degradation by *Fibrobacter succinogenes* S85. *PLoS One* **10**: e0143809.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367-1372.

Desmet, T., Cantaert, T., Gualfetti, P., Nerinckx, W., Gross, L., Mitchinson, C., and Piens, K. (2007) An investigation of the substrate specificity of the xyloglucanase Cel74A from *Hypocrea jecorina*. *FEBS J* **274**: 356-363.

Elhenawy, W., Debelyy, M.O., and Feldman, M.F. (2014) Preferential packing of acidic glycosidases and proteases into *Bacteroides* outer membrane vesicles. *MBio* **5**: e00909-00914.

Ezer, A., Matalon, E., Jindou, S., Borovok, I., Atamna, N., Yu, Z. et al. (2008) Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. *J Bacteriol* **190**: 8220-8222.

Feng, T., Yan, K.P., Mikkelsen, M.D., Meyer, A.S., Schols, H.A., Westereng, B., and Mikkelsen, J.D. (2014) Characterisation of a novel endo-xyloglucanase (XcXGHA) from *Xanthomonas* that accommodates a xylosyl-substituted glucose at subsite -1. *Appl Microbiol Biotechnol* **98**: 9667-9679.

Forsberg, C.W., Beveridge, T.J., and Hellstrom, A. (1981) Cellulase and Xylanase Release from *Bacteroides succinogenes* and Its Importance in the Rumen Environment. *Appl Environ Microbiol* **42**: 886-896.

Gaudet, G., and Gaillard, B. (1987) Vesicle formation and cellulose degradation in *Bacteroides succinogenes* cultures: ultrastructural aspects. *Arch Microbiol* **148**: 150-154.

Gong, J., and Forsberg, C.W. (1993) Separation of outer and cytoplasmic membranes of *Fibrobacter succinogenes* and membrane and glycogen granule locations of glycanases and cellobiase. *J Bacteriol* **175**: 6810-6821.

Gong, J., Egbosimba, E.E., and Forsberg, C.W. (1996) Cellulose-binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. *Can J Microbiol* **42**: 453-460.

581 Groleau, D., and Forsberg, C.W. (1981) Cellulolytic activity of the rumen bacterium *Bacteroides*
582 *succinogenes*. *Can J Microbiol* **27**: 517-530.
583 Groleau, D., and Forsberg, C.W. (1983) Partial characterization of the extracellular
584 carboxymethylcellulase activity produced by the rumen bacterium *Bacteroides succinogenes*.
585 *Can J Microbiol* **29**: 504-517.
586 Himmel, M.E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R., and Bayer, E.A. (2010) Microbial
587 enzyme systems for biomass conversion: emerging paradigms. *Biofuels* **1**: 323-341.
588 Horn, S.J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V.G. (2012) Novel enzymes for the
589 degradation of cellulose. *Biotechnol Biofuels* **5**: 45.
590 Jun, H.S., Qi, M., Gong, J., Egbosimba, E.E., and Forsberg, C.W. (2007) Outer membrane
591 proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in
592 cellulose digestion. *J Bacteriol* **189**: 6806-6815.
593 Kobayashi, Y., Shinkai, T., and Koike, S. (2008) Ecological and physiological characterization
594 shows that *Fibrobacter succinogenes* is important in rumen fiber digestion — Review. *Folia*
595 *Microbiologica* **53**: 195-200.
596 Kojima, Y., Varnai, A., Ishida, T., Sunagawa, N., Petrovic, D.M., Igarashi, K. et al. (2016)
597 Characterization of an LPMO from the brown-rot fungus *Gloeophyllum trabeum* with broad
598 xyloglucan specificity, and its action on cellulose-xyloglucan complexes. *Appl Environ*
599 *Microbiol*.
600 Krause, D.O., Denman, S.E., Mackie, R.I., Morrison, M., Rae, A.L., Attwood, G.T., and
601 McSweeney, C.S. (2003) Opportunities to improve fiber degradation in the rumen:
602 microbiology, ecology, and genomics. *FEMS Microbiol Rev* **27**: 663-693.

Kulp, A., and Kuehn, M.J. (2010) Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu Rev Microbiol* **64**: 163-184.

Kvansakul, M., Adams, J.C., and Hohenester, E. (2004) Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J* **23**: 1223-1233.

Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G., Koropatkin, N.M. et al. (2016) A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnol Biofuels* **9**: 260.

Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009) DOOR: a database for prokaryotic operons. *Nucleic Acids Res* **37**: D459-463.

Mba Medie, F., Davies, G.J., Drancourt, M., and Henrissat, B. (2012) Genome analyses highlight the different biological roles of cellulases. *Nat Rev Microbiol* **10**: 227-234.

Miller, G.L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal Chem* **31**: 426-428.

Naas, A.E., Mackenzie, A.K., Mravec, J., Schuckel, J., Willats, W.G., Eijsink, V.G., and Pope, P.B. (2014) Do rumen *Bacteroidetes* utilize an alternative mechanism for cellulose degradation? *MBio* **5**: e01401-01414.

Nekiunaite, L., Petrovic, D.M., Westereng, B., Vaaje-Kolstad, G., Hachem, M.A., Varnai, A., and Eijsink, V.G. (2016) FgLPMO9A from *Fusarium graminearum* cleaves xyloglucan independently of the backbone substitution pattern. *FEBS Lett* **590**: 3346-3356.

Park, Y.B., and Cosgrove, D.J. (2015) Xyloglucan and its interactions with other components of the growing cell wall. *Plant Cell Physiol* **56**: 180-194.

624 Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999) Probability-based protein
625 identification by searching sequence databases using mass spectrometry data. *Electrophoresis*
626 **20**: 3551-3567.
627 Rakoff-Nahoum, S., Coyne, M.J., and Comstock, L.E. (2014) An ecological network of
628 polysaccharide utilization among human intestinal symbionts. *Curr Biol* **24**: 40-49.
629 Raut, M.P., Karunakaran, E., Mukherjee, J., Biggs, C.A., and Wright, P.C. (2015) Influence of
630 Substrates on the Surface Characteristics and Membrane Proteome of *Fibrobacter succinogenes*
631 S85. *PLoS One* **10**: e0141197.
632 Roier, S., Zingl, F.G., Cakar, F., Durakovic, S., Kohl, P., Eichmann, T.O. et al. (2016) A novel
633 mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. *Nat*
634 *Commun* **7**: 10515.
635 Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., and Templeton, D. (2006)
636 Determination of sugars, byproducts, and degradation products in liquid fraction process
637 samples. *Golden: National Renewable Energy Laboratory*.
638 Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J. et al. (2011) The
639 complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic
640 specialist. *PLoS One* **6**: e18814.
641 Toyoda, A., Iio, W., Mitsumori, M., and Minato, H. (2009) Isolation and identification of
642 cellulose-binding proteins from sheep rumen contents. *Appl Environ Microbiol* **75**: 1667-1673.
643 Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M. et al. (2013) The
644 PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic*
645 *Acids Res* **41**: D1063-1069.

Westereng, B., Coenen, G.J., Michaelsen, T.E., Voragen, A.G., Samuelsen, A.B., Schols, H.A., and Knutsen, S.H. (2009) Release and characterization of single side chains of white cabbage pectin and their complement-fixing activity. *Mol Nutr Food Res* **53**: 780-789.

Wittig, I., Karas, M., and Schagger, H. (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* **6**: 1215-1225.

Wood, T.M. (**1988**) Preparation of crystalline, amorphous, and dyed cellulase substrates. In *Methods Enzymol*: Academic Press, pp. 19-25.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**: W445-451.

Zeytuni, N., and Zarivach, R. (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20**: 397-405.

Zhao, G., Ali, E., Sakka, M., Kimura, T., and Sakka, K. (2006) Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. *Appl Microbiol Biotechnol* **70**: 464-469.

TABLES

Table 1. Enriched Pfam domains in OMVs. The table shows Pfam domains, which were found to be overrepresented (Fisher's Exact $p < 0.05$) in the OMVs compared to the complete genome. A more detailed display of these domains, and the proteins harboring them, can be found in Figure S4.

Pfam Accession	Pfam Name	Pfam Description	CAZy Family	Genome sequences	OMV sequences	Enrichment	Average Log_2(LFQ intensity)	Fisher's Exact p-value
PF03422	CBM_6	Carbohydrate binding module (family 6)	CBM6	24	16	67%	23.8	0.000001
PF03781	FGE-sulfatase	Sulfatase-modifying factor enzyme 1		20	11	55%	24.1	0.000231
PF00150	Cellulase	Cellulase (glycosyl hydrolase family 5)	GH5	12	8	67%	26.5	0.000673
PF03425	CBM_11	Carbohydrate binding domain (family 11)	CBM11	5	5	100%	27.8	0.002282
PF04616	Glyco_hydro_43	Glycosyl hydrolases family 43	GH43	13	7	54%	24.0	0.003610
PF00759	Glyco_hydro_9	Glycosyl hydrolase family 9	GH9	7	5	71%	26.2	0.005971
PF09479	Flg_new	Listeria-Bacteroides repeat domain (List_Bact_rpt)		4	4	100%	24.5	0.006596
PF00691	OmpA	OmpA family		13	6	46%	28.0	0.012202
PF07691	PA14	PA14 domain		6	4	67%	25.4	0.016598
PF17189	Glyco_hydro_30C	Glycosyl hydrolase family 30 beta sandwich domain	GH30	4	3	75%	24.8	0.031468
PF00722	Glyco_hydro_16	Glycosyl hydrolases family 16	GH16	4	3	75%	22.2	0.031468
PF02321	OEP	Outer membrane efflux protein		4	3	75%	23.9	0.031468
PF09603	Fib_succ_major	Fibrobacter succinogenes major domain		52	12	23%	25.5	0.038435
PF13568	OMP_b-brl_2	Outer membrane protein beta-barrel domain		5	3	60%	27.9	0.046392
PF13174	TPR_6	Tetratricopeptide repeat		5	3	60%	26.0	0.046392
PF02927	CelD_N[a]	Cellulase N-terminal ig-like domain	GH9	5	3	60%	26.4	0.046392
PF03629	SASA	Carbohydrate esterase, sialic acid-specific acetylesterase	CE	5	3	60%	22.8	0.046392
PF01103	Bac_surface_Ag	Surface antigen		5	3	60%	25.2	0.046392

[a]The N-terminal ig-like domain of GH9

Table 2. Identification of proteins from the 2D-hrCN-SDS-PAGE gel. The table shows the proteomic detection of proteins in spots numbered 1 to 15 in Figure 3A. Only the most abundant protein in each spot is shown. More details, including the proteins present at lower abundance, can be found in Table S3.

Complex	Spot no.	UniProt	Locus	Name	Mascot Score
	1	A7UG45	FSU_2404	Membrane protein	7320
	2	A7UG46	FSU_1029	Membrane protein	2730
C1	3	A7UG46	FSU_1029	Membrane protein	8773
	4	A7UG37	FSU_2078	OmpA family protein	5372
	5	C9RRD7	FSU_2008	Uncharacterized protein	3832
C2	6	A7UG61	FSU_2396[a]	OmpA family protein	4747
	7	A7UG37	FSU_2078	OmpA family protein	4389
	8	A7UG62	FSU_2397[a]	TPR domain protein	5100
	9	A7UG66	FSU_2502[a]	Fibro-slime domain protein	22351
	10	C9RL47	FSU_0797	Uncharacterized protein	3784
C3	11	A7UG61	FSU_2396[a]	OmpA family protein	6853
	12	A7UG61	FSU_2396[a]	OmpA family protein	4207
	13	A7UG62	FSU_2397[a]	TPR domain protein	6181
	14	A7UG62	FSU_2397[a]	TPR domain protein	4282
	15	C9RKA2	FSU_2794	F. succinogenes major paralogous domain protein	51056

[a]Proteins known to be involved in cellulose binding according to Jun HS, Qi M, Gong J, Egbosimba EE, Forsberg CW. 2007. J Bacteriol 189:6806-15.

FIGURE LEGENDS

Figure 1: CAZymes detected in the OMVs and their predicted activities on various plant-derived substrates. A) The figure shows the 74 proteins that could be annotated as carbohydrate-active enzymes (CAZymes) plotted against their relative abundances in the OMVs. Proteins carrying a carbohydrate-binding module (CBM) are colored as indicated in the figure. **B)** The figure shows nine different substrates susceptible to hydrolysis by the OMVs. Each substrate is annotated with CAZymes found in the OMVs, predicted cleavage sites that are based on literature and the detected products. The products detected by PGC-MS after overnight incubation of intact OMVs with the substrates are illustrated in miniature beneath the substrates, with numbers indicating the degree of polymerization (DP). A detailed list of products can be found in Table S2. Signals obtained by incubation of the substrates in buffer at 40 °C in the absence of OMVs, were used for background subtraction. The shown substrates are: I: Ivory nut mannan, II: Carob galactomannan, III: Konjac glucomannan, IV: Birchwood xylan, V: Aspen xylan, VI: Wheat arabinoxylan, VII: Cabbage pectin (including RG, rhamnogalacturonan), VIII: PASC from Avicel cellulose, IX: Tamarind xyloglucan.

Figure 2: Effect of OMV treatment on the enzymatic digestibility of switchgrass. A) Ball milled and washed switchgrass at 0.2% (w/v) was incubated without (OMV –) or with 20 µg OMVs (OMV +) for 17.5 hours at 40 °C, then subjected to enzymatic degradation by 20 µg 4:1 (w/w) mixture of Celluclast and Novozym 188. **B)** Products detected by PGC-MS after overnight incubation of switchgrass either with a mixture of Celluclast and Novozym 188 or with intact OMVs are shown with numbers indicating the degree of polymerization (DP). An explanation of

699 symbols can be found in Figure 1B and a detailed list of products can be found in Table S2. The
700 sugars were assigned as hexoses and pentoses when it was not possible to obtain a clear
701 identification using the combination of retention time and *m/z*-value.

702

703 **Figure 3: 2-dimensional electrophoresis to analyze protein complexes. A**) 2D-hrCN-SDS-
704 PAGE gel; protein complexes are thought to migrate as an intact complex in the first dimension.
705 In the second dimension, the complex disintegrates due to SDS, and the subunits are separated and
706 appear on a vertical line. The gel suggests the presence of three major complexes, C1, C2 and C3.
707 Note that the region of the gel marked by I, is the stacking region of the 1st dimension and will not
708 separate proteins. Spots 1 and 2, which also show divergent electrophoretic elution pattern, are
709 therefore not likely in a complex. **B)** Enzymatic activity on PASC. Another native gel lane,
710 identical to the one used in panel A, was divided into seven pieces (Fractions I – VII) and grinded
711 prior to measurement of enzyme activity. Linear cello- and xylo-oligosaccharides were separated,
712 identified and quantified by PGC-MS, and the amounts of the different products are displayed in
713 the form of a heat map. Note that the amounts are relative per compound due to differences in the
714 efficiency of ionization among the compounds. The scale corresponds to the integrated peak area
715 as reported by the Xcalibur software. Glc: glucose, Xyl: xylose.

716

717 **Figure 4: Visualization of the possible mode of action of complex C2. A)** The four proteins
718 identified as complex C2, two OmpA family proteins (FSU_2078, FSU_2396), a TPR domain
719 protein (FSU_2397) and a fibro-slime protein (FSU_2502), likely form a putative OMV-associated
720 complex (**B**), where the fibro-slime protein mediates binding to cellulose (Gong et al., 1996) and

721 the TPR acts as a protein scaffold for complex assembly (Blatch and Lassle, 1999). The beta helix
722 domain in FSU_2396 OmpA, similar to that seen in pectate lyases, could be responsible for
723 hydrolyzing the substrate.

SUPPLEMENTARY MATERIAL LEGENDS

Figure S1: Outer membrane vesicles isolated from *F. succinogenes* after growth on Avicel cellulose. **A**) Data from dynamic light scattering, revealing a broad population of OMVs with radii ranging from 8-136 nm, with an average of 49 nm. The small peak observed at 3.4 nm represents micelles formed by the *n*-dodecyl-β-D-maltoside detergent present in the sample at 0.05% (w/v). **B**) Visualization of OMVs by negatively stained transmission electron microscopy; the picture shows vesicles with a diameter of around 100 nm. Scale bar: each white/black subsection is 40 nm.

Figure S2: Proteomics reproducibility. The figure shows the comparison of the two replicates analysed by quantitative proteomics, showing high reproducibility with Pearson correlation R=0.805. The axes represent log2(LFQ) values obtained in each replicate and the colors represent the number of peptides associated with each protein.

Figure S3: Clusters of orthologous groups (COG) analysis. The figure shows the mapping of COG functions to proteins in the complete proteome of *F. succinogenes* as well as to the proteins detected in the OMVs. The figure shows a higher proportion of carbohydrate-active enzymes in OMVs compared to the complete proteome.

Figure S4: Pfam enrichment analysis. The figure shows the proteins associated with each of the 18 enriched Pfam modules listed in Table 2, and the relative abundance of these proteins

[log2(LFQ) values ranging from 20.6 (low abundant; light green) to 33.4 (high abundant; light red)]. The data is shown as a hierarchically clustered heat map where shades of red indicate the presence of a given Pfam module in a protein (light red: one occurrence, dark red: four occurrences; note that there are no proteins with three occurrences). Grey indicates the absence of listed Pfam modules.

Table S1. Proteins identified in *F. succinogenes* OMVs. The table shows the MaxQuant output for identified proteins and the LFQ intensities used for quantification. Protein annotations were done using the LipoP server, the peptidase database MEROPS and Pfam, while CAZy predictions were done using dbCAN. PEP: Posterior error probability.

Table S2. Products detected after OMV-mediated hydrolysis of various substrates. The table shows oligosaccharides identified by PGC-MS after overnight incubation of intact OMVs with nine different plant-derived substrates and one natural substrate (switchgrass). Product assignments were based on *m/z* values from the mass spectrometer and retention times on the PGC column. The compositions of the substrates are also provided. Hex: hexose, Pen: pentose, Gal: galactose, GalA: galacturonic acid, uGalA: unsaturated galacturonic acid, Xyl: xylose, Glc: glucose, Man: mannose, Ara: arabinose: GlcA: glucuronic acid, Me: methylated, Ac: acetylated, DP: degree of polymerization.

Table S3. Proteins identified in the 2D-hrCN-SDS-PAGE gel. The table shows the proteomic detection of proteins in spots numbered 1 to 15 in Figure 3A, providing additional information

767 next to the list of most abundant proteins in Table 2. emPAI values provided by the Mascot search
768 engine were used to obtain quantitative estimates within each protein spot, and were used to filter
769 out low abundant hits. Only hits with emPAI > 5 are shown; the most abundant protein in each
770 spot is printed in bold face. Note that emPAI values are not normalized, meaning that abundance
771 levels can only be compared within single spots.

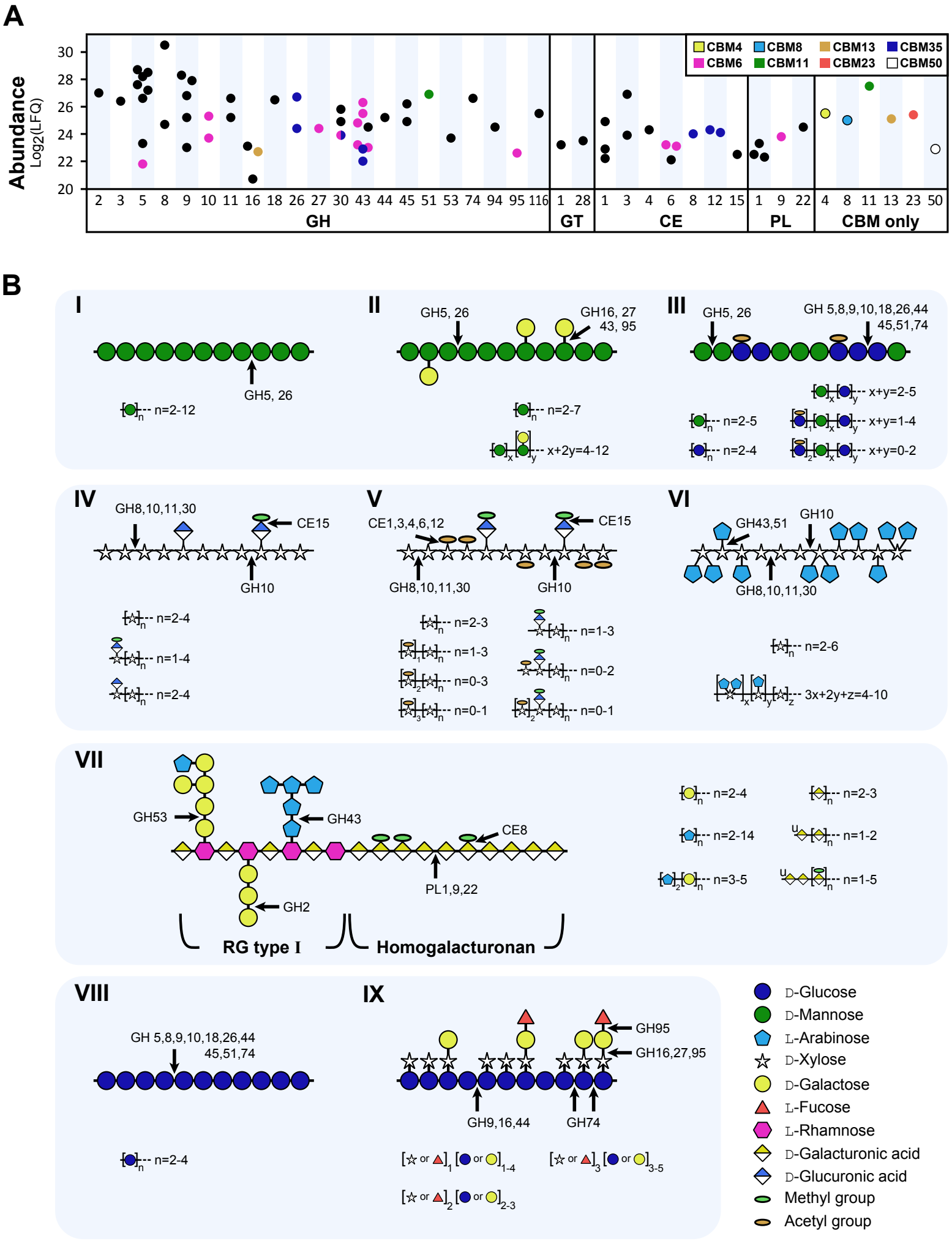
772

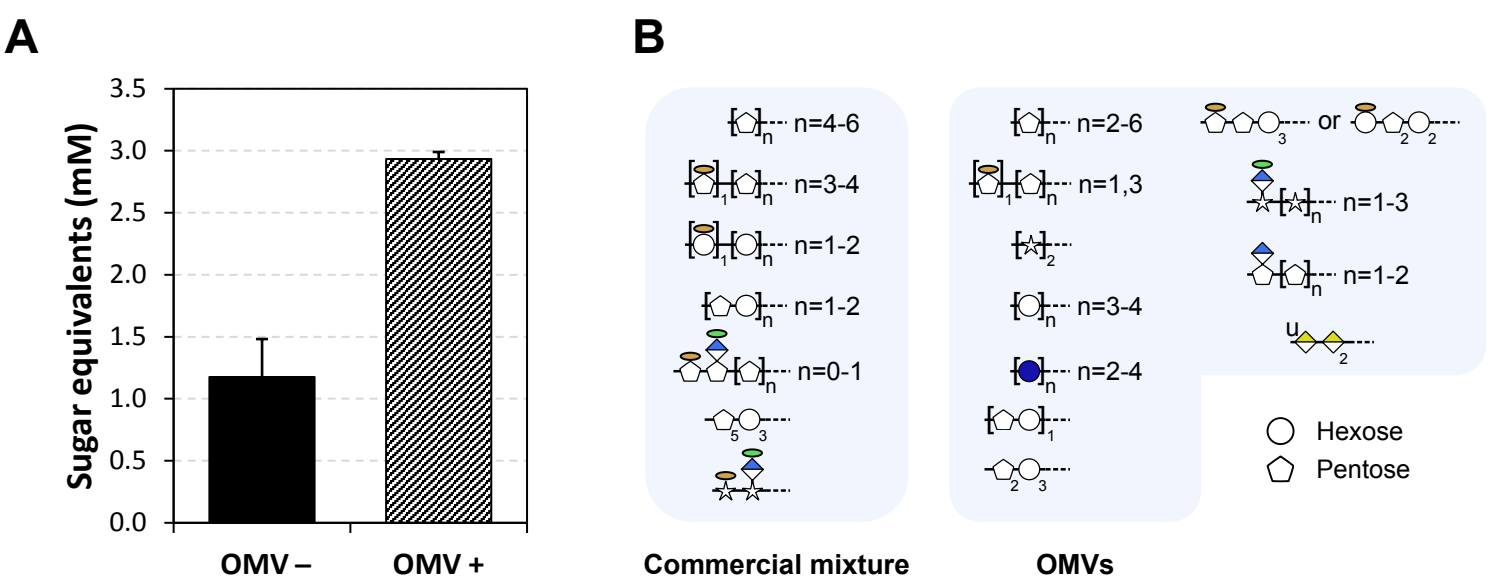
773 **Supplementary Text S1: Enriched non-CAZy domains and other OMV proteins with**
774 **potential involvement in carbohydrate binding or metabolism.** This text extends the
775 Discussion by highlighting how the PA14 domain and the sulfatase-modifying factor enzyme
776 could be involved in carbohydrate binding and/or metabolism. Type IV pilin proteins and
777 cadherins are also discussed.

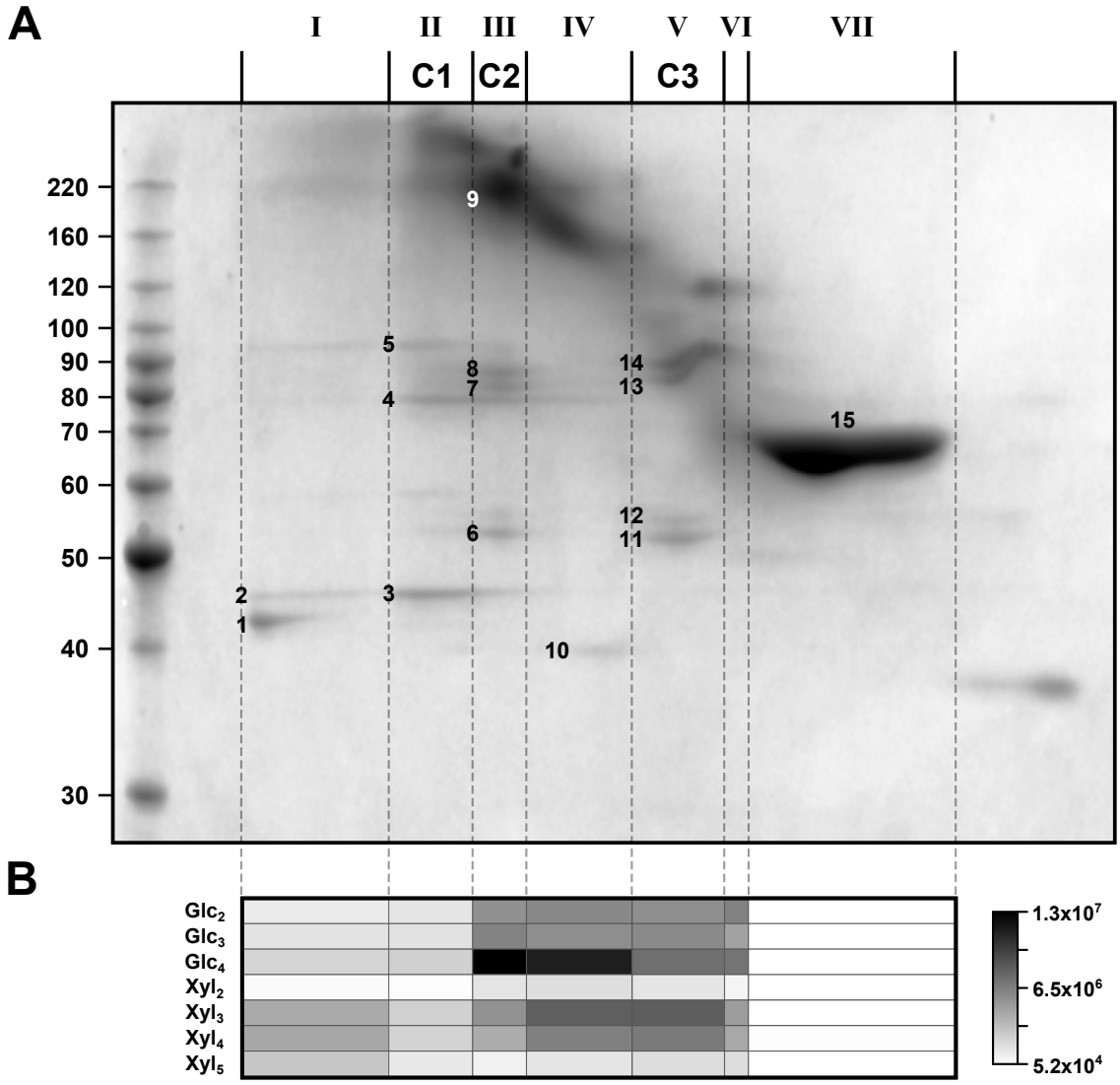
778

779 **Supplementary Text S2. Methodological details.** This text extends the Experimental Procedures
780 section with details regarding the growth medium, dynamic light scattering, transmission electron
781 microscopy, the proteomics and bioinformatics analysis as well as the analysis of products from
782 enzymatic assays. In particular, the use of HPLC and mass spectrometry setups are explained in
783 detail.

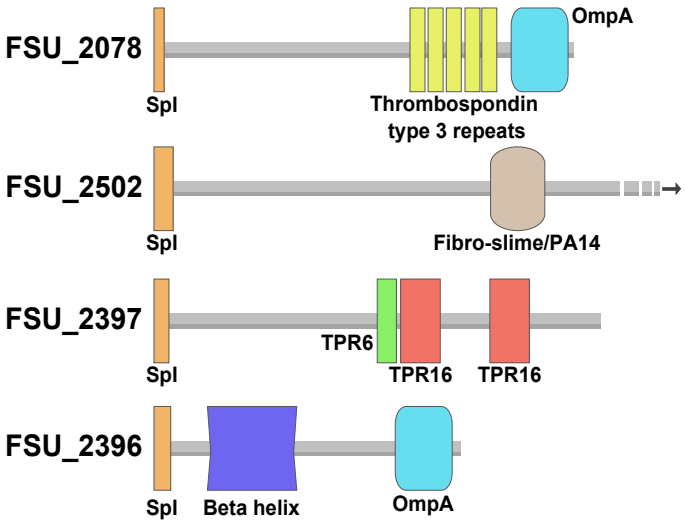
784


A
Abundance
Log2(LFQ)
30
28
26
24
22
20
CBM4
CBM8
CBM13
CBM35
CBM6
CBM11
CBM23
CBM50
2 3 5 8 9 10 11 16 18 26 27 30 43 44 45 51 53 74 94 95 116
GH
1 28
GT
1 3 4 6 8 12 15
CE
1 9 22
PL
4 8 11 13 23 50
CBM only
B
I
GH5, 26
n=2-12
II
GH5, 26
GH16, 27
43, 95
n=2-7
x+2y=4-12
III
GH5, 26
GH 5,8,9,10,18,26,44
45,51,74
x+y=2-5
n=2-5
x+y=1-4
n=2-4
x+y=0-2
IV
GH8,10,11,30
CE15
GH10
n=2-4
n=1-4
n=2-4
V
CE1,3,4,6,12
CE15
GH8,10,11,30
GH10
n=2-3
n=1-3
n=0-3
n=0-1
n=1-3
n=0-2
n=0-1
VI
GH43,51
GH10
GH8,10,11,30
n=2-6
3x+2y+z=4-10
VII
GH53
GH43
CE8
PL1,9,22
GH2
RG type I
Homogalacturonan
n=2-4
n=2-3
n=2-14
n=1-2
n=3-5
n=1-5
VIII
GH 5,8,9,10,18,26,44
45,51,74
n=2-4
IX
GH95
GH16,27,95
GH9,16,44
GH74
D-Glucose
D-Mannose
L-Arabinose
D-Xylose
D-Galactose
L-Fucose
L-Rhamnose
D-Galacturonic acid
D-Glucuronic acid
Methyl group
Acetyl group

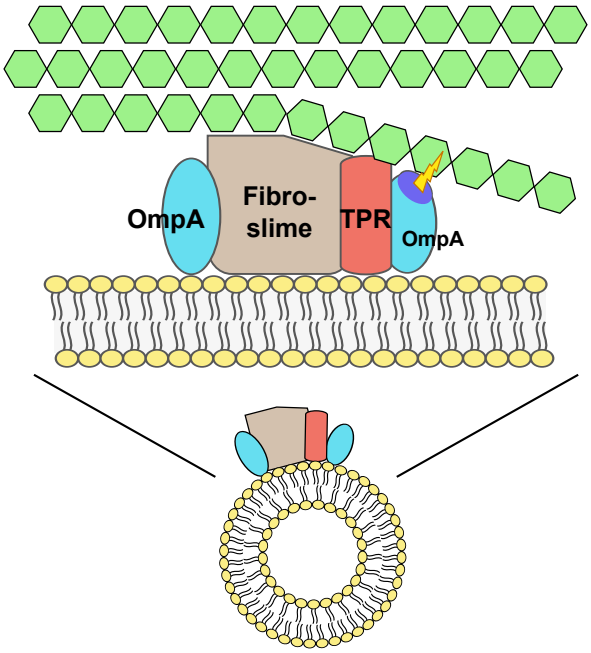

A
Sugar equivalents (mM)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
OMV –
OMV +
B
n=4-6
n=3-4
n=1-2
n=1-2
n=0-1
n=2-6
n=1,3
n=3-4
n=2-4
or
n=1-3
n=1-2
Hexose
Pentose
Commercial mixture
OMVs


A
I
II
III
IV
V
VI
VII
C1
C2
C3
220
160
120
100
90
80
70
60
50
40
30
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
B
Glc2
Glc3
Glc4
Xyl2
Xyl3
Xyl4
Xyl5
1.3x10^7
6.5x10^6
5.2x10^4

A


FSU_2078
OmpA
SpI
Thrombospondin
type 3 repeats
FSU_2502
SpI
Fibro-slime/PA14
FSU_2397
TPR6
SpI
TPR16
TPR16
FSU_2396
SpI
Beta helix
OmpA

B


OmpA
Fibro-
slime
TPR
OmpA