Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown

Lucy I. Crouch1,12✉, Marcelo V. Liberato2, Paulina A. Urbanowicz3, Arnaud Baslé1, Christopher A. Lamb4, Christopher J. Stewart5, Katie Cooke5, Mary Doona4, Stephanie Needham6, Richard R. Brady7, Janet E. Berrington8, Katarina Madunic9, Manfred Wuhrer9, Peter Chater1, Jeffery P. Pearson1, Robert Glowacki10, Eric C. Martens10, Fuming Zhang11, Robert J. Linhardt11, Daniel I. R. Spencer3 & David N. Bolam1

The thick mucus layer of the gut provides a barrier to infiltration of the underlying epithelia by both the normal microbiota and enteric pathogens. Some members of the microbiota utilise mucin glycoproteins as a nutrient source, but a detailed understanding of the mechanisms used to breakdown these complex macromolecules is lacking. Here we describe the discovery and characterisation of endo-acting enzymes from prominent mucin-degrading bacteria that target the polyLacNAc structures within oligosaccharide side chains of both animal and human mucins. These O-glycanases are part of the large and diverse glycoside hydrolase 16 (GH16) family and are often lipoproteins, indicating that they are surface located and thus likely involved in the initial step in mucin breakdown. These data provide a significant advance in our knowledge of the mechanism of mucin breakdown by the normal microbiota. Furthermore, we also demonstrate the potential use of these enzymes as tools to explore changes in O-glycan structure in a number of intestinal disease states.
The human gastrointestinal (GI) tract is home to a large and complex community of microbes known as the human gut microbiota (HGM), with the greatest densities assembling in the large intestine where numbers of bacterial cells are estimated to be ~100 trillion\(^1\). The mucus layer shields the host epithelial cells of the GI tract from both the normal microbiota and enteric pathogens. Mucus is predominantly composed of gel-forming mucins, which are complex glycoproteins secreted by the epithelial cells\(^2\). Different mucin genes are expressed in different mucosal surfaces throughout the body and mucins are at least 50% O-glycan by mass\(^3\). In the colon, MUC2 is the most abundant gel-forming mucin and is composed of ~80% glycan\(^1\). While the number of different monosaccharides and types of sulphate decoration making up mucin oligosaccharide side chains are relatively limited, the order in which they can be assembled is hugely variable (Fig. 1a). The heterogeneity between individual O-glycan chains leads to a highly complex macromolecule and it is this complexity that provides some resistance to microbial degradation and contributes to the mucus layers’ protective role\(^4\). Despite this heterogeneity, some prominent bacterial members of the microbiota have developed the capacity to graze on mucins, including certain Bacteroides spp. and Akkermanis\(^a\) muciniphila\(^5\)–\(^9\). This ability is thought to be critical to the initial colonisation by the microbiota in a new-born and therefore to the development of a healthy adult microbiota\(^10\). Mucin grazing also enables survival during the absence of diet-derived glycans\(^11\) and non-mucin degrading species have been shown to be cross-fed by mucin degraders, contributing to the long-term survival and

![Fig. 1 Mucin structure and genomic context of the loci encoding the mucin-associated GH16 enzymes. a Left: the main structural features of a model mucin O-glycan chain. All mucin oligosaccharides are linked via an α-GalNAc to serine and threonine residues in the peptide backbone. A number of different core structures are then attached, with core 3 (shown) being the most common in the large intestinal MUC2. The cores are then often extended with polyLacNAc repeats of varying lengths which are decorated along their length by sulfation and fucosylation and capped at the non-reducing end by a variety of α-linked monosaccharides. Right: a model of an intestinal mucin glycoprotein showing complexity and variability of glycan chains attached to peptide backbone. b Genetic context of the GH16 encoding genes identified as being upregulated in the four species shown during growth on mucin (see Supplementary Figs. 1–3). In Bacteroides spp. the GH16 genes (highlighted red) are part of discrete polysaccharide utilisation loci (PULs), cluster of co-regulated genes encoding glycan degradation and uptake apparatus (SusC-like and SusD-like outer membrane proteins, additional CAZymes and putative surface glycan binding proteins (pSGBPs), often adjacent to a hybrid two component system (HTCS) sensor-regulators that likely control expression of the associated PUL. Glycan utilisation genes are not organised into PULs in the A. muciniphila genome.](image-url)
stability of the microbiota\textsuperscript{12,13}. In contrast, aberrant or excess degradation of the mucosal layer by the normal microbiota has been linked to enhanced pathogen susceptibility, inflammatory bowel disease (IBD) and even colorectal cancer\textsuperscript{11,14}.

Despite the importance to gut health of mucin breakdown by the microbiota, little is known about the molecular details of this process. Current models of mucin degradation propose extracellular sequential trimming of terminal sugars from the O-glycan side chains by extracellular exo-acting glycosidases to eventually expose the peptide backbone for proteolysis\textsuperscript{15}. However, this extracellular ‘exo-trimming’ model is based only on the activity of currently characterised mucin active enzymes, such as sialidases and fucosidases. The exo-trimming model is also in direct contrast with the degradation pathways used by Gram-negative Bacteroidetes, known as the Sus paradigm\textsuperscript{16}. Sus-like systems derive their name (Starch utilisation system) after the first such system characterised, but each Sus-like apparatus targets a distinct glycan, and many Bacteroides spp. contain tens to hundreds of these systems. In general, in Sus-like systems, a surface endo-acting glycanase cleaves the substrate (polysaccharide or glyco-conjugate) into smaller oligosaccharides for uptake by SusC/D outer membrane complexes\textsuperscript{17–19}.

Here we describe the discovery and characterisation of endo-acting glycoside hydrolases expressed by mucin-degrading members of the HGM that are able to cleave the O-glycan chains of a range of different animal and human mucins. These O-glycans are enzymes that cluster together (Supplementary Fig. 6). The enzymes that cluster with these GH16 enzymes were identified as involved degradation of O-glycosidases found in Metazoa, Fungi, Archaea and Bacteria\textsuperscript{29}. A total of nine GH16 enzymes were identified from the four species (Figs. 1b, 2, and Supplementary Figs. 1–3). Five of the nine GH16 enzymes are predicted lipoproteins and therefore likely cell surface associated (Supplementary Table 1). Interestingly, these GH16 family members generally had a relatively low sequence identity between 24–34%. The exceptions to this were two pairs of B. fragilis and B. cacaeces enzymes: BF4058 and BACCAC\_02679 (red asterisk) display 87% identity, while BF4060 and BACCAC\_02680 (black asterisk) display 79% identity.

Specifically, the GH16 enzymes identified are a part of subfamily 3, which is a large and sequence-diverse subfamily characterised predominantly as β1,3/4-glucosidases found in Metazoa, Fungi, Archaea and Bacteria\textsuperscript{29}. Phylogenetic analysis of mucin-associated GH16 enzymes. The protein sequences of the nine GH16 family members identified as upregulated during growth on mucins were compared with the characterised GH16 family members from the CAZY database (Supplementary Fig. 5 and Supplementary Table 3). The phylogenetic tree indicates that the mucin-associated GH16 enzymes are most closely related to the β-glucanase GH16 family members rather than those with activities on β-galactans, xyloglucan or chitin- β1,6-glucanotransferases. Another analysis of the GH16 subfamily 3 protein sequences indicate that seven of the mucin-associated GH16 enzymes cluster together (Supplementary Fig. 6). The branch where they sit is composed of proteins only from mucosal-associated organisms, including known pathogens. Two of the A. muciniphila enzymes (Amuc\_0724 and Amuc\_0875) cluster in a different branch and are relatively close to GH16 family members that have been characterised as having endo β-1,3-galactanase activity (Supplementary Fig. 6). The two separate clusters indicate that the mucin-associated GH16 enzymes have evolved twice from different β-glucanase ancestors (Supplementary Fig. 6). The enzymes that cluster with these sequences potentially have similar activities. While the mucin-associated GH16 enzymes are present in prominent mucin degrading Gram-negative bacteria such as Bacteroides spp., inspection of the genomes of mucin-degrading Gram positive members of the gut microbiota, including some Ruminococcus and Bifidobacteria spp., revealed no evidence for these enzymes\textsuperscript{29}.

The mucin-associated GH16 genes encode endo-acting O-glycanases. To explore the activity of the nine mucin-upregulated GH16 family members, the recombinant forms of the enzymes
were screened against porcine small intestinal (SI) mucin and porcine gastric mucins (PGM type II and III; Supplementary Fig. 7). Initial analysis by thin layer chromatography (TLC) suggested that all nine enzymes were active against both SI and gastric mucins and released a range of products from these glycoproteins that are larger than monosaccharides, suggesting endo-like cleavage of the O-glycan chains.

To investigate the identity of these products in more detail, the glycans were labelled at their reducing end with the fluorophore procainamide and analysed by liquid chromatography-fluorescence-detection-electrospray-mass spectrometry (LC-FLD-ESI-MS) and the glycan structures determined by MS/MS (Fig.3). The data show that all the GH16 enzymes produce oligosaccharides of alternating hexose and HexNAc sugars with a variety of lengths. The reducing ends were all hexoses, indicating hydrolysis occurred at β-galactose (α-galactose only occurs in mucins as a terminal sugar in blood group B structures and GH16 enzymes only target β-linked sugars) and the products also had a range of fucose and sulphate decorations, revealing these can be accommodated by the GH16 enzymes. Overall, these data indicate that the nine GH16 enzymes are all endo-acting β-galactosidases that are active on the O-glycan side chains of mucin (O-glycanases).

Notably, sialic acid (SA) was never observed as a decoration on any products released by the enzymes, even though SA is present on mucin glycans (Fig.3). These data suggest this terminal sugar decoration cannot be accommodated by the GH16 O-glycanases and, as a result, the broad acting sialidase BT0455GH33 was included in all assays to maximise access of the GH16 enzymes to the mucin chains.

To investigate the composition of the oligosaccharides that are released by the GH16 O-glycanases in greater detail we digested the GH16 products with a series of exo-acting glycosidases of known specificity. For example, digestion with glycosidases...
Partial activity
Inactive
of de
linkages by the GH16 mucinases. The middle bond is the side chains and was hydrolysed at two of the Gal
Figs. 9 hydrolysed further to produce GlcNAc and GlcNAc be cleaved to produce two trisaccharides, one of these is then
Investigating the speci
identi
in combination against different mucins could be a powerful tool (Supplementary Fig. 8). The use of endo- and exo-acting enzymes
these non-reducing end decorations on some GH16 products (Supplementary Fig. 8). The use of endo- and exo-acting enzymes in combination against different mucins could be a powerful tool in exploring different structures in O-glycan research and identification of disease biomarkers.

**Fig. 4 Heat map showing the activity of the GH16 O-glycanases against different oligosaccharides.** The data summarises the specificity of the GH16 O-glycanases described in this report. From left to right the glycans are TetraLacNAc, TriLacNAc, paraLacto-N-neohexaose, Lacto-N-neotetraose, Lacto-N-tetraose, Lacto-N-triose, Galβ1,3GalNAc β1,3Galβ1,4Glc tetrasaccharide Blood group A hexasaccharide, Blood group B hexasaccharide, Blood group H pentasaccharide, LacNAc, Blood group H tetrasaccharide II, 3-sialylactose and P1 antigen. The linkages are β unless otherwise labelled and the bonds cleaved are indicated by the black arrows. Partial and trace activity are the estimation of greater than or less than 50% degradation, respectively, under the assay conditions used. A more detailed summary can be found in Supplementary Table 4. The predicted cellular locations of each enzyme is indicated on the far right of each row.

specific to either the α-GalNAc or α-galactose found on blood group A or B structures, respectively, enabled identification of these non-reducing end decorations on some GH16 products (Supplementary Fig. 8). The use of endo- and exo-acting enzymes in combination against different mucins could be a powerful tool in exploring different structures in O-glycan research and identification of disease biomarkers.

**Endo O-glycanase activity on the cell surface.** To assess if there is endo O-glycanase activity on the cell surface we used whole cell assays (Fig. 5 and Supplementary Fig. 13). Bacterial cultures grown on PGM III were harvested, washed and exposed to either TriLacNAc or TetraLacNAc. The data revealed the same pattern of degradation as that produced by the recombinant GH16 O-glycanases against TriLacNAc. The identity of the products was confirmed using diagnostic assays using exo-acting enzymes of known specificity for two of the time points (Fig. 5b and Supplementary Fig. 13). These data indicate that GH16 O-glycanase-like activity is present on the surface of all four of the bacterial species studied here.

**Activity against polyLacNAc structures in non-mucin host glycan.** Keratan sulphate (KS) chains are anchored to the protein through N-linkages, O-linkages and O-mannosylation are termed KS-I, KS-II and KS-III, respectively (Supplementary Fig. 4). Examples of areas of the body enriched in KS-I, KS-II and KS-III include the cornea, skeletal and brain, respectively. Keratan sulphate is present in the GI tract from sloughed off epithelial cells and also dietary sources.

Keratan sulphate has a similar structure to O-glycans and is also composed of a repeating polyLacNAc structure with 6 S sulfation possible on both the galactose and GlcNAc, but with less fucosylation and sialylation than most mucins. The nine GH16 O-glycanases were found to be active against both egg and bovine corneal keratan sulphate and the released products indicate that a significant number of sulphate groups can be tolerated by the enzymes (Supplementary Fig. 14). Degradation of KS also supports the previous finding that fucose decorations are not required for activity of the GH16 O-glycanases. Activity against this KS substrate also demonstrates that the O-glycanases target polyLacNAc chains in a range of glycans, likely including keratan sulphate in the gut, although we could not test growth on KS as the glycan was only available in small amounts.

Activity of the GH16 O-glycanases against classical GH16 substrates. The activity of GH16 enzymes against host glycans is unusual in comparison to the specificities displayed for previously characterised members of the family. The activities of the GH16
O-glycans were tested against polysaccharides previously shown to be substrates for GH16 family members to assess their level of O-glycan specificity (Supplementary Fig. 15). No activity could be detected for any of the enzymes against agarose, κ-carrageenan, porphyran, pectic galactan, xyloligucan or chitin. However, Amuc_0724 displayed significant endo-like activity against laminarin and weak activity against barley β-glucan and lichenan. BF4060, BACCAC_02680 and BACCAC_03717 also displayed some very weak activity against laminarin. The possible rationale for the activity of Amuc_0724 against Glc configured substrates is discussed below. Other non-mucin host polysaccharides are also present in significant amounts in mucosal surfaces, including chondroitin sulphate (CS), heparin (Hep) and hyaluronic acid (HA). The O-glycan active GH16 enzymes were also tested against these polysaccharides and no significant activity could be found, except for a small amount of low molecular weight product released from Hep (Supplementary Fig. 15).

Overall these data reveal that the nine GH16 enzymes analysed are endo acting β-1,4-galactosidases that display a preference for polyLacNAc structures found in mucins and other similar glycans such as KS.

Crystal structures of O-glycan active GH16 family members. To investigate the structural basis for O-glycan specificity displayed by the GH16 O-glycanases we attempted to obtain crystal structures of these enzymes in complex with substrate. Crystal trials were set up with both wild-type protein and mutants of the catalytic nucleophile. We also co-crystallised the different proteins with TriLacNAc to obtain substrate or product complexes where possible. Six separate data sets were collected from four of the enzymes. The apo structures of BACCAC_02680, BACCA-C_02680E143Q, BACCAC_03717 and Amuc_0724 were obtained to 2.0, 2.1, 2.1 and 2.7 Å, respectively. Structures of BF4060 and BACCAC_02680E143Q were also obtained with the Galβ1,4GlcNAcβ1,3Gal present in the negative subsites (despite the latter enzyme being a catalytic mutant; Supplementary Fig. 16) to 3.3 and 2.0 Å (Fig. 6, Supplementary Tables 5–7, and Supplementary Figs. 16–20). The electron density of the trisaccharide product allowed us to model in the sugars, the conformations were checked using Privateer, and these conformations correlated with what had been seen previously in other GH16 structures (Supplementary Fig. 17 and Supplementary Table 7).

All of the GH16 enzymes comprise a β-jellyroll fold, characteristic of the family, consisting of two β-sheets composed of β-stands that form the core fold, which were superimposable with other GH16 structures previously published. A cleft running along the concave surface of the enzymes contains the active site and where the trisaccharide product was bound in the cases of BF4060 and BACCAC_02680E143Q (Fig. 6a). While the location of the substrate binding site is conserved in the GH16 family, the structures of these clefts vary depending on substrate specificity (Supplementary Fig. 18a). Some form a tight tunnel for linear undecorated glycans (e.g., agarase from Zobellia galactanivorans33), others are much more open to accommodate decorations (e.g., xyloligucanase from Tropaeolum majus34), while some GH16 enzymes have substrate binding clefts that are curved to optimise binding to highly curved glycans such as laminarin35.

There is also a single example of a GH16 family member that has evolved a pocket-like active site to recognise and cleave a specific disaccharide from the terminus of glycan chains in gastric mucin36 (Supplementary Fig. 18b). The key structural features that modulate the shape of different clefts are the surrounding loops and short α-helices extending from the β-strands of the core fold (exemplified in Supplementary Fig. 19). These extensions have been likened to fingers that interact with substrate, thereby dictating specificity, and that nomenclature is used herein37. BF4060 and BACCAC_02680E143Q have four fingers and BACCAC_03717 and Amuc_0724 have five out of six possible fingers that have been observed previously in other GH16 structures37.

Inspection of the BACCAC_02680E143Q and BF4060 structures with product reveal most of the interactions between enzyme and sugar are with the Gal at −1 and GlcNAc at −2. The −1 subsite in BACCAC_02680E143Q is composed of a number of aromatics, which are also a common feature of the GH16 structures available (Fig. 6b). This enzyme possesses four fingers (numbers 1, 3, 5 and 6) that extend towards the cleft, with fingers 1 and 3 sandwiching the negative subsites and fingers 5 and 6 sandwiching the positive subsites. Finger 3 contains the sequence motif for GH16 subfamily 3, which consists of three
tryptophans interspaced by other residues. BF4060 and BACCAC_02680 display 79% identity and unsurprisingly the structures of these two enzymes are almost identical in the cleft region. In contrast, BACCAC_03717 and Amuc_0724 both possess a finger 2 (in addition to 1, 3, 5 and 6) and this has a more variable topology than the other fingers (Fig. 6a). For Amuc_0724, finger 2 sits over the top of finger 1, but in the BACCAC_03717 structure it points away from the cleft. This could reflect the flexibility of finger 2 in this enzyme and could potentially come down over loop 1 in solution like in the Amuc_0724 structure. The B-factor putty projections of the GH16 crystal structures show finger 2 is dynamic in the BACCAC_03717 (Supplementary Fig. 19) and alternative conformations of individual fingers from of other GH16 family members has been observed previously, a finding which is indicative of flexibility.

Structural basis for specificity of the GH16 O-glycanases. Previously characterised GH16 family enzymes target a variety of β-glucan and galactan substrates (Supplementary Fig. 4). Glucose and galactose differ only in the hydroxyl group at C4 being equatorial or axial, respectively. Anhydrogalactose is also present in agarose and carrageenan and sulfation in porphyran and carrageenan (Fig. 6e). Porphyran and carrageenan are 6 S and 4 S sulfated, respectively, and these decorations would therefore point into the GH16 binding cleft at subsites −2 and −1, respectively.
Structural features characteristic to O-glycans include alternating Glc and Gal configured sugars and additionally the presence of GlcNAc, which is not found in other GH16 substrates (Fig. 6c). Furthermore, 6 S is found on both Gal and GlcNAc and 3 S is possible on the galactose at the non-reducing ends of O-glycan chains.

The new GH16 structures reported here were systematically compared to the GH16 structures available for glucanases, laminarinases, porphyranases, carrageenases, xyloglucanases and other activities. Each subsite was analysed in comparison to the structures of GH16 enzymes with other specificities to understand the structural basis for O-glycan specificity. This detailed analysis highlighted four structural features of the mucin active GH16 family members that tailor these enzymes towards O-glycans.

Firstly, in the −1 subsites of the structures from Bacteroides spp., the closed space around the O4 hydroxyl explains why only Gal configured sugars can be recognised as the equatorial O4 of glucose would not be accommodated (Fig. 6c). The structure of Amuc_0724 in this area is much more open and is a likely explanation for this enzymes additional activity against laminarin (Supplementary Fig. 15). Furthermore, the open space at the O4 in Amuc_0724 is a potential pocket for sulfation that the Bacteroides spp. enzymes would not be able to accommodate (Supplementary Fig. 20). Phylogenetic analysis reveals the mucin active GH16 enzymes are likely to have derived from β-glucanase GH16 family members that drive specificity for a β,3 between the −1 and −2 sugars are shown (W129, W138 and W131, W140; See Supplementary Fig. 20 for active sites of BACCAC_03717, and Amuc_0724). c A surface representation of the regions surrounding the −1 subsite showing the selection for the axial O4 of Gal in the three Bacteroides enzymes, while Amuc_0724 has a more open ‘tunnel’ like space that appears to also allow accommodation of the equatorial O4 of Glc. The product from BACCAC_02680E143Q was overlaid in the BACCAC_03717, and Amuc_0724 structures. Colours represent the different ‘fingers’. d A view of the predicted −1 subsite of BF4060 and BACCAC_03717 overlaid with the glucose from the +1 subsite of a laminarinase from Phaeonerochaete chrysosporium. The +1 subsites are much more closed for BF4060 and BACCAC_02680E143Q compared to BACCAC_03717, and Amuc_0724. e, An overview of the monosaccharides occupying the different subsites in GH16 family members with different activities. Linkages also shown. It should be noted that the sulfation will be variable along the O-glycan chain and there will also be fucose decorations. This situation is similar to porphyran, where the polysaccharide can have a variable composition, but the subsite occupancy shown here reflects the observations of structures of enzyme-glycan complexes currently available for porphyranases.

Generating O-glycan profiles from human tissue samples. O-linked glycans are common modifications to proteins and lipids in addition to being the major component of mucins. Changes in O-glycosylation patterns in mucins and other glycoproteins have previously been detected in a variety of disease states and these changes can contribute to disease progression and severity, for example, by facilitating the metastasis of cancer. Therefore, these alterations represent promising biomarkers for the screening and prognosis of different diseases, especially in combination with other exo-acting enzymes of known specificity with activity against O-glycans. Indeed, sialyl Lewis X, sialyl lewis A and
Forssman antigens are already used as biomarkers for different types of cancers.

To demonstrate that the GH16 enzymes could be used to generate O-glycan profiles for such screening purposes, we analysed the activity of one of the O-glycanases against mucins from three different types of human tissue/cells. Tissues from two adults suffering from ulcerative colitis (UC) were obtained as well as samples from preterm infants with necrotising enterocolitis (NEC); 4 infants of gestations 26, 27, 28 and 35 weeks. The mucus layer was scraped from the small amount of tissue (approximately 1 cm²) that could be spared by the pathologist and we used this to carry out assays. We also tested mucins produced by a number of cultured colorectal cancer (CRC) cell lines originating from different patients.

The human mucin samples were incubated with Amuc_0724, due to its broad activity, in combination with the sialidase BT0455GH33 and the products were labelled with procainamide, and analysed using LC-FLD-ESI-MS. The data reveal that Amuc_0724 was able to release a range of Gal terminated and variably fucosylated and sulfated oligosaccharides from all of the samples tested, with similar structures to those identified from porcine SI mucin (Fig. 7 and Supplementary Fig. 21).

Fig. 7 Examples of the O-glycan profiles that can be produced from a variety different human mucins by the Amuc_0724 O-glycanase. Mucins from different samples were pre-treated with the broad acting sialidase BT0455GH33 and then digested with the GH16 and products analysed by LC-FLD-ESI-MS. a Inflamed colonic tissue removed during a laproscopic panproctocolectomy from a patient with UC, b Bowel tissue from neonates with necrotising enterocolitis. c Colorectal cancer cell lines. Small amounts of Neu5Gc are seen in some of the UC samples (e.g., panel a) suggesting either the presence of contaminating dietary animal O-glycans remaining in the mucus layer or that this xenobiotic sugar has been incorporated into human mucins from dietary sources. The O-glycan profiles of all the different samples analysed are shown in Supplementary Fig. 21.

Fig. 7 Examples of the O-glycan profiles that can be produced from a variety different human mucins by the Amuc_0724 O-glycanase. Mucins from different samples were pre-treated with the broad acting sialidase BT0455GH33 and then digested with the GH16 and products analysed by LC-FLD-ESI-MS. a Inflamed colonic tissue removed during a laproscopic panproctocolectomy from a patient with UC, b Bowel tissue from neonates with necrotising enterocolitis. c Colorectal cancer cell lines. Small amounts of Neu5Gc are seen in some of the UC samples (e.g., panel a) suggesting either the presence of contaminating dietary animal O-glycans remaining in the mucus layer or that this xenobiotic sugar has been incorporated into human mucins from dietary sources. The O-glycan profiles of all the different samples analysed are shown in Supplementary Fig. 21.

Discussion

Here we describe the characterisation of members of the GH16 family that have a specificity for the Galβ1,4GlcNAc linkage in the polyLacNAc chains found in mucins and other O-glycans. This discovery of endo-acting O-glycanases expressed by prominent mucin degrading gut bacteria is an important step forward in furthering our understanding of the complex relationship between the HGM and host.

We observed activity of these O-glycanases on both animal and human mucins from a range of tissues including the stomach and the small and large intestine. While much of the O-glycan that colonic microbiota will be exposed to will be from MUC2, as the major mucin expressed in the distal intestine, it is worth noting that these bacteria will also come into contact with significant amounts of MUC5AC, MUC5B and MUC6 mucins that have moved down the digestive tract from the saliva, oesophagus and stomach where they originated. In addition to these gel-forming mucins, gut microbes will be exposed to membrane-associated mucins that are a part of the apical surface glyocalyx of epithelial cells, especially when dead cells are sloughed off the epithelium throughout the GI tract and these include MUC3, MUC12 and MUC17. Furthermore, greater than 80% of secreted proteins are O-glycosylated and the gut microbiota will come into contact with these from both host and dietary proteins.
The activity of the GH16 enzymes against the broad range of substrates reported here indicate that these microbes can access O-glycans from the different types of mucins moving through the GI tract.

The discovery of this endo O-glycanase activity is in contrast to the previously proposed ‘exo-trimming’ model of mucin degradation and is more similar to pathways for degradation of other glycans seen in Bacteroides spp. In all of the Bacteroides species studied here, at least one of the GH16 endo-mucinases is predicted to be a lipoprotein and whole cell assays support GH16 O-glycanases being localised to the cell surface. On the cell surface, the endo-mucinase and sialidase activities are likely the initial steps in O-glycan breakdown and other exo-acting CAZymes that trim capping sugars are also likely present (Fig. 8).

Although targeting of polyLacNAc structures by the GH16 enzymes is likely an initial step in mucin breakdown, further processing of the remaining mucin would be required. The polyLacNAc side chains are attached to different core glycan structures, which are in turn linked to the peptide backbone. This is most likely dealt with through a combination of extracellular or surface exo-acting glycosidases and peptidases. Indeed, it has recently been shown that gut microbes, including B. thetaiotaomicron, express glyco-peptidases that specifically target polypeptides with O-glycan core structures still attached.

While this model applies to Bacteroides spp, it is currently unknown how A. muciniphila cells access complex glycans. However, there is direct experimental evidence that at least one of the GH16 mucinases expressed by A. muciniphila (Amuc_2108) is localised to the outer membrane during growth on mucin, and A. muciniphila cells also display O-glycanase-like activity on the cell surface, supporting a similar role for these enzymes in initiating mucin breakdown.

Structures of several O-glycanase enzymes revealed the characteristics of the subsites driving the specificity towards polyLacNAc chains. In particular, alternating Gal and GlcNAc sugars, a requirement for a β1,3 linkage between the −2 and −1 subsites, and potential pockets for sulfation or fucosylation. There is a relatively low number of required subsites for catalysis, in contrast to that seen for GH16 enzymes with other activities, likely reflecting the highly heterogeneous nature of mucin O-glycans.

The combination of these new GH16 endo O-glycanase activities with other exo-acting CAZymes and the sensitive analytical techniques applied to human mucin samples described here demonstrates potential applications in both fundamental research and medicine. We hope this study facilitates partnerships between basic researchers and clinicians to explore the structures of O-glycans in a range of diseases on a larger scale in order to develop more effective biomarkers (e.g. for earlier detection of disease). This could lead to less invasive and more rapid techniques for diagnosis and monitoring remission in prevalent diseases like the ones explored here.

Overall, the findings reported here contribute significantly towards our understanding of the molecular mechanisms of mucin breakdown by the microbiota, a key process in maintaining the host-microbe symbiosis in the gut. These findings also open up the exciting possibility of exploiting this activity for characterisation and detection of biomarkers to allow more effective and earlier diagnosis of intestinal diseases such as IBD and CRC.

**Methods**

**Sources of glycans and glycoproteins.** TriLacNAc was purchased from Elicityl and the rest of the defined oligosaccharides were from Carboynth. PGM II and III (Sigma) was produced by dissolving in DI water at 50 mg ml−1 and the precipitate removed by centrifugation before assays were carried out (leaving 35–40 mg ml−1). Porcine small intestinal mucin was prepared as previously described with the only modification being a double CsCl gradient without Sepharose separation or SDS-PAGE in between. Keratan was prepared as described previously.

**Fig. 8 Model for the role of GH16 O-glycanases in mucin breakdown.** A model of the initial steps of mucin degradation on the surface of Bacteroides spp. and A. muciniphila. Sialic acid is removed by surface-localised sialidases and the GH16 enzymes remove oligosaccharides for import into the periplasm for further degradation by other CAZymes, including periplasmic GH16 O-glycanases. In Bacteroides species, glycan import is via energy dependent SusCD-like complexes, but in A. muciniphila the mechanism of glycan import across the outer membrane is unknown. The remaining mucin glycoprotein is likely further degraded by other extracellular CAZymes and glycopeptidases.
Bacterial strains. The Bacteroides strains used were: B. thetaiotaomicron VPI-5482, B. fragilis NCTC9343, B. caccae ATCC43185, B. celularisolicus DSM1438, B. fenvielloi DSM14756, R. vulgai ATCC8483, R. ovattii ATCC8482, B. intestinalis DSM17393, and Akkermansia muciniphila ATCC BAA835/DSM22959.

Cloning, expression and purification of recombinant proteins. The DNA encoding the enzymes described in this report were amplified from genomic DNA and cloned into pET28b (Novagen) excluding the signal sequences, which were washed, and increasing amounts of imidazole used to elute the recombinant chromatography using Talon resin (Clontech). The protein was bound to the resin, Labtech) and incubated at 20 °C. BACCAC_02680E143Q was incubated with 5 mM 0.1 AKTA Pure FPLC system (GE Healthcare Life Sciences). SDS-PAGE gels were used to

Human sample collection. IBD tissue samples were from two subjects. Matched ileal and colonic samples were obtained from one panproctocolectomy and one ileorectal resection. Samples were transferred on wet ice directly to the laboratory for mechanical isolation of the mucus layer by gently scraping using a pipette tip. For NEC samples, fresh tissue was collected from surgically resected specimens when a clinically necessary procedure was taking part, stored briefly in sterile phosphate buffered saline and transported to the laboratory on ice.

Procainamide labelling. Reducing ends of GH16 products were labelled by reductive amination using a procainamide labelling kit containing sodium cyanoborohydride as reductant (Ludger). Before and after labelling the O-glycan samples were cleaned up using PBM plates and S-cartridges, respectively (Ludger).

LC-FLD-ESI-MS of procainamide labelled glycans. Procainamide-labelled samples were analysed by LC-FLD-ESI-MS. 25 μl of sample was injected to a Waters ACQUITY UPLC Glycan BEH Amide column (2.1 × 150 mm, 1.7 μm particle size, 130 A pore size) at 40 °C on a Dionex Ultimate 3000 UHPLC instrument with a flow rate of 0.4 ml min−1; capillary temperature was 370 °C; detector at a Bruker Amazon Speed ETD. Mobile phase A was a 50 mM ammonium formate solution (pH 4.4) and mobile phase B was neat acetonitrile. Analyte separation was accomplished by gradients running at a flow rate of 0.4 ml min−1 from 85 to 57% mobile phase B over 115 min and from 85 to 62% over 95 min for mucin and keratan samples, respectively. The Ammon speed was operated in the positive sensitivity mode using the following settings: source temperature, 180 °C; gas flow, 41 min−1; capillary voltage, 4500 V; ICC target, 200,000; maximum accumulation time, 50.00 ms; rolling average, 2; number of precursor ions selected, 3; scan mode, enhanced resolution; mass range scanned, 400 to 1700. HyStar v3.2 was used for data collection of chromatography and mass spectrometry. GraphPad Prism was used to produce the figures.

Analysis of mass spectrometry data. Mass spectrometry of procainamide-labelled glycans was analysed using Bruker Compass Data Analysis Software and GlycoWorkbench65. Glycan compositions were elucidated on the basis of MS2 fragmentation and previously published data.

Bioinformatics. Putative signal sequences were identified using SignalP 5.053. Sequence identities were determined using Clustal Omega using full sequences. The IMG database (https://img.jgi.doe.gov/) was used to analyse synteny between different species75. The CAZY database (www.cazy.org) was used as the main reference for CAZymes71. To determine the boundaries between different modules in a protein Pfam52 and SMART73,74 were used.

Alignments and phylogenetic trees were completed in SeaView79. Sequences were aligned in SeaView using Clustal. Gblocks were applied to: allow smaller final blocks, allow gap positions within the final blocks, and allow less strict flank positioning. Gblocks were built using the ARTEMIS platform, model-given amino acid equilibrium frequencies, no invariable sites, optimised across site rate variation and the best of NNI and SPR tree searching operations.
Received: 8 January 2020; Accepted: 21 July 2020; Published online: 11 August 2020

References

1. Johansson, M. E., Sjovall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. Nat. Rev. Gastroenterol. Hepatol. 10, 352–361 (2013).
2. Lang, T., Hansson, G. C. & Samuelsson, T. Gel-forming mucins appeared early in metazoan evolution. Proc. Natl. Acad. Sci. USA 104, 16209–16214 (2007).
3. Johansson, M. E. & Hansson, G. C. Immunological aspects of intestinal mucin and mucus. Nat. Rev. Immunol. 16, 639–649 (2016).
4. Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. A complex, but uniform 0-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. Glycobiology 19, 756–766 (2009).
5. Forster, S. C. et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat. Biotechnol. 37, 186–192 (2019).
6. Derrien, M., Vaughan, E. E., Pluge, C. M. & de Vos, W. M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. 54, 1469–1476 (2004).
7. Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordination regulation of glyc can degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. J. Biol. Chem. 284, 18445–18457 (2009).
8. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. 9, e1001221 (2011).
9. Marcelob, A. et al. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe 10, 507–514 (2011).
10. Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell 167, 1339–1353. e1321 (2016).
11. Egan, M. et al. Cross-feeding by Bifidobacterium breve UCC2003 during co-cultivation with Previobacter PRL2010 in a mucus-based medium. BMC Microbiol. 14, 282 (2014).
12. Schroeder, B. O. et al. Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. Cell host microbe 23, 27–40.e27 (2018).
13. Corfield, A. P. The interaction of the gut microbiota with the mucus barrier in health and disease in human. Microorganisms 6, 78 (2018).
14. Marcelob, A., Southwick, A. M., Earle, K. A. & Sonnenburg, J. L. A refined palate: bacterial consumption of host glycans in the gut. Glycobiology 23, 1038–1046 (2013).
15. Reeves, A. R., Wang, G. R. & Salyers, A. Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J. Bacteriol. 179, 643–649 (1997).
16. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16, 1105–1115 (2008).
17. Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. Nat. Commun. 6, 7481 (2015).
18. Brulé, J. et al. Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. Nat. Microbiol. 4, 1571–1581 (2019).
19. Pudlo, N. A. et al. Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. mBio 6, e01282–01215 (2015).
20. Ottman, N. et al. Genome-scale model and omics analysis of metabolic capacities of Akkermansia muciniphila reveal a preferential mucin-degrading lifestyle. Appl. Environ. Microbiol. 83, e01144–17 (2017).
21. Ottman, N. et al. Characterization of outer membrane proteome of Akkermansia muciniphila reveals sets of novel proteins exposed to the human intestine. Front. Microbiol. 7, 1157 (2016).
22. Shin, J. et al. Elucidation of Akkermansia muciniphila probiotic traits driven by mucin depletion. Front. Microbiol. 10, 1137 (2019).
Acknowledgements

We thank Carl Moxland (Newcastle University, UK) for his expert technical assistance. We thank Dr. Mirjam Cajzek for her expert advice on the structural data and Prof Robert Hirt for his insightful conversations about phylogenetics. We would like to thank Diamond Light Source (Oxfordshire, UK) for beamtime (proposal mx18598) and staff of beamline I03, I04-1 and I24. We are grateful to Newcastle Biobank and NIHR Newcastle Biomedical Research Centre. Dr. Jose Muñoz-Muñoz kindly gifted the arabinogalactan.

R.G. designed the experiments, analysed the data, and wrote the manuscript. R.R.B., M.D., and S.N. were responsible for ethical approval, governance, patient identification, and sample collection for IBD tissues. R.R.B. performed surgery where adult intestinal samples were collected. K.C. and C.A.L. were responsible for lab preparation of IBD tissue. C.J.S. and J.E.B. were responsible for ethical approval and provision of samples for the preterm neonate NEC samples. J.E.B. performed the neonate surgery. K.M. prepared the colorectal cancer cells and M.W. provided resources. L.I.C. and D.N.B. designed the experiments, analysed the data, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Ethics declaration

Adult IBD tissue was collected as part of the Newcastle Biobank following written consent according to approval from Newcastle and North Tyneside Research Ethics Committee 1 (REC:17/NE/0361). Necrotising enterocolitis samples were collected with written consent obtained from parents and ethical permission provided through the SERVS study (approvals 10/H0908/39 and 15-NE-0334).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-17847-5.

Correspondence and requests for materials should be addressed to L.I.C. or D.N.B.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020