nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0466-x

Complex *N*-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci

Justina Briliūtė[1,5], Paulina A. Urbanowicz[2,5], Ana S. Luis[3], Arnaud Baslé[1], Neil Paterson[4], Osmond Rebello[2], Jenifer Hendel[2], Didier A. Ndeh[1], Elisabeth C. Lowe[1], Eric C. Martens[3], Daniel I. R. Spencer[2], David N. Bolam[1*] and Lucy I. Crouch[1*]

Glycans are the major carbon sources available to the human colonic microbiota. Numerous *N*-glycosylated proteins are found in the human gut, from both dietary and host sources, including immunoglobulins such as IgA that are secreted into the intestine at high levels. Here, we show that many mutualistic gut *Bacteroides* spp. have the capacity to utilize complex *N*-glycans (CNGs) as nutrients, including those from immunoglobulins. Detailed mechanistic studies using transcriptomic, biochemical, structural and genetic techniques reveal the pathway employed by *Bacteroides thetaiotaomicron* (*Bt*) for CNG degradation. The breakdown process involves an extensive enzymatic apparatus encoded by multiple non-adjacent loci and comprises 19 different carbohydrate-active enzymes from different families, including a CNG-specific endo-glycosidase activity. Furthermore, CNG degradation involves the activity of carbohydrate-active enzymes that have previously been implicated in the degradation of other classes of glycan. This complex and diverse apparatus provides *Bt* with the capacity to access the myriad different structural variants of CNGs likely to be found in the intestinal niche.

The composition of the human gut microbiota is closely associated with many aspects of health, development and disease[1–4]. The major nutrients available to the colonic microbiota are complex carbohydrates and glycan preferences shape the composition of this microbial community[5]. While dietary plant polysaccharides probably make up the bulk of glycans reaching the large intestine, host glycans are also a significant nutrient source and access to these host molecules appears to be important for gut colonization and survival[6–9].

N-linked glycans are commonly present within the gut as almost all secreted eukaryotic proteins are *N*-glycosylated to some extent[10]. *N*-glycans are composed of a conserved pentasaccharide core capped with a diverse range of different monosaccharides to give three main types: complex, high mannose and hybrid, although there is significant heterogeneity within these structures (Fig. 1a)[10]. *N*-glycans found in the gut come from a range of sources, including dietary animal products and host molecules such as mucins and antibodies[11,12]. For example, secretory IgA, the major immunoglobulin produced in the intestine, is heavily decorated with mainly complex *N*-glycan (CNG) structures[13,14].

Despite the prominence of *N*-glycans in the human intestine, there are only a few studies describing their use as nutrients by the microbiota and the pathways of *N*-glycan breakdown employed by mutualistic microbes. Some strains of Bifidobacteria have been found to utilize *N*-glycans on lactoferrin from milk[15], while *Bt* has been shown to express a single polysaccharide utilization locus (PUL) to access high-mannose *N*-glycans (HMNG)[8]. While there is a paucity of data on *N*-glycan use by the normal gut microbiota, especially of the complex type, these molecules are well known to be key nutrients for some pathogens, as the ability to access *N*-glycans facilitates survival of these microbes within the extra-intestinal niche[16–21]. Although several studies describe aspects of *N*-glycan degradation by enteric pathogens, most of these focus on only a single part of the degradation pathway or the role of this process in virulence[21–24].

Bacteroides are one of the dominant genera of the human gut[25] and are prominent glycan degraders, with many species encoding hundreds of different carbohydrate-active enzymes (CAZymes) to facilitate access to the diverse array of glycans available[6–8,26–32]. In *Bacteroides*, the genes encoding the glycan-degrading apparatus are usually co-localized in discrete clusters known as PULs[33]. PULs are typically defined by their composition as containing SusC-like and SusD-like gene pairs that encode the outer-membrane glycan-import machinery[34], as well as a sensor-regulator and multiple CAZymes. Often a single PUL encodes all of the apparatus required to access a specific glycan structure, although multiple PULs have been shown to be involved in the degradation of some highly complex glycans[5,35].

In this study we set out to further our understanding of CNG breakdown by the normal gut microbiota, focusing on prominent *Bacteroides* spp. The data reveal that the ability to utilize CNGs is commonplace among gut *Bacteroides* and that growth of *Bt* on CNG activates an extensive enzymatic apparatus encoded by multiple loci, including a single core PUL that likely controls expression of the other non-adjacent loci. Detailed characterization of the *Bt* CNG-breakdown pathway provides significant insights into the mechanism of *N*-glycan degradation by key members of the microbiota and provides knowledge for future studies examining the role of this process in gut survival.

[1]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, UK. [2]Ludger Ltd, Culham Science Centre, Abingdon, UK. [3]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA. [4]Diamond Light Source, Didcot, UK. [5]These authors contributed equally: Justina Briliūtė, Paulina A. Urbanowicz. *e-mail: david.bolam@ncl.ac.uk; lucy.crouch@ncl.ac.uk

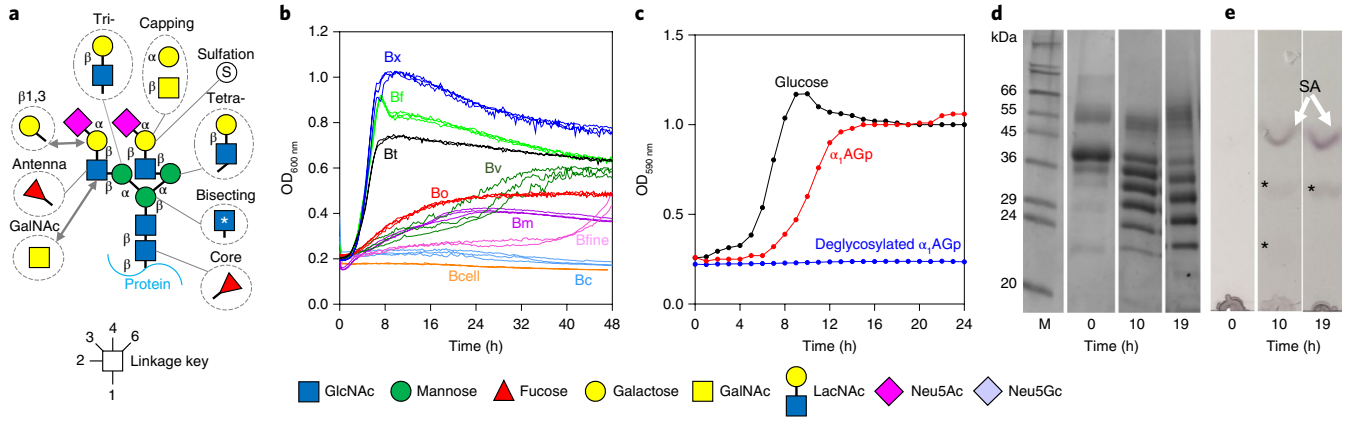

Fig. 1 | Complex *N*-glycans as a nutrient source for *Bacteroides* species. a, The structure of a bi-antennary CNG with other possible decorations and variations. The linkage is indicated by the key; however, the sialic acid sugars can be attached through α2,3 or α2,6 linkages to the galactose and also to the antennary GlcNAc. Bi-antennary CNGs have both antennae linked through β1,2 bonds and additional antennae have either a β1,4 linkage (on the α1,3 arm) or a β1,6 linkage (on the α1,6 arm) to produce tri- and tetra-antennary CNGs, respectively. Common modifications to this core model are shown in dotted circles. **b**, Growth on native α_1AGp of different *Bacteroides* spp. (*B. thetaiotaomicron* (*Bt*, black), *B. xylanisolvans* (*Bx*, dark blue), *B. ovatus* (*Bo*, red), *B. vulgatus* (*Bv*, dark green), *B. finegoldii* (*Bfine*, pink), *B. massiliensis* (*Bm*, magneta), *B. fragilis* (*Bf*, light green), *B. caccae* (*Bc*, light blue) and *B. cellulolyticus* (*Bcell*, orange)). Growth curves were carried out at least twice for each species. **c**, Growth of *Bt* on glucose (5 mg ml^{-1}), α_1AGp and deglycosylated α_1AGp (both 20 mg ml^{-1}) as the sole carbon source. **d**, Supernatant samples were taken at the start, 10 h and 19 h time points from the growth of *Bt* on α_1AGp shown in **c** and were analysed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE). Fully glycosylated α_1AGp is shown at T_0 and fully deglycosylated α_1AGp is the bottom 23.5 kDa protein band in the 10 h and 19 h time point lanes. **e**, Thin-layer chromatography of the cell-free spent media at the 10 h and 19 h time points to analyse glycans present. The top band is sialic acid (SA, white arrow) and two other faint glycan bands are marked by asterisks. The results are representative of at least three independent replicates. Full versions of the gels and thin-layer chromatograms can be found in Supplementary Figs. 1 and 23.

Results

Growth of gut *Bacteroides* on CNGs. Prominent gut *Bacteroides* spp. were tested for their ability to grow on native bovine α_1-acid glycoprotein (α_1AGp) as the sole carbon source (Fig. 1a). Of the 9 species tested, most could grow to some extent on α_1AGp, with only *B. caccae* and *B. cellulosolyticus* showing no growth over 48 h (Fig. 1b). Analysis of supernatant after growth revealed that no protein degradation had occurred and the protein remaining was almost completely deglycosylated, indicating that growth was supported by the glycan component of α_1AGp alone (Fig. 1c–e and Supplementary Fig. 1a,b). Growth on other types of CNG-containing glycoprotein was also assessed (Supplementary Fig. 1f) and is described in Supplementary Results and Discussion.

RNA-Seq to identify CNG degrading apparatus. RNA-Seq was used to identify the predicted CAZyme and associated genes upregulated during growth of *Bt* on α_1AGp. Six discrete loci were identified from this analysis that were composed of, in total, 19 predicted CAZymes, 4 SusC/D pairs and several open reading frames (ORFs) of unknown function (Fig. 2, Supplementary Tables 1 and 2, Supplementary Results and Discussion and Supplementary Fig. 2). Only two of the loci are classed as PULs in that they encode predicted CAZymes, SusC/D homologues and sensor-regulators[33]. These data indicate that breakdown of CNGs by *Bt* requires the cooperative action of a complex multi-locus-encoded machinery. To understand the degradative pathway for CNGs in *Bt* we carried out biochemical characterization of the CAZymes upregulated on α1AGp.

BT0455^{GH33} sialidase is required for growth on CNGs and is localized both on the cell surface and inside the cell. Mammalian CNGs are often capped with sialic acid (SA), with *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) being the dominant forms; however, only Neu5Ac is found in humans (Fig. 1a)[36]. Neu5Gc would therefore be available to the microbiota only from eukaryotic dietary sources, while Neu5Ac would be present in both dietary and host glycans. BT0455^{GH33} is the only sialidase encoded by the *Bt* genome and has been previously characterized as having broad specificity, acting on a range of SA-capped structures including CNGs[37].

Bt cannot metabolize SA, but requires its removal to enable access to underlying glycan[38] and as expected SA build-up can be seen in the supernatant over the course of growth of *Bt* on α_1AGp (Fig. 1e and Supplementary Figs. 1 and 3). BT0455^{GH33} was found to be critical for the utilization of CNGs by *Bt* as a BT0455-deletion strain (ΔBT0455) was unable to grow and the defect could be rescued by addition of recombinant BT0455 to the media (Supplementary Fig. 1).

We next examined the cellular location of BT0455^{GH33} using whole-cell activity assays and protease protection assays (Supplementary Figs. 4–7 and Supplementary Results and Discussion). Unexpectedly, the data revealed that BT0455^{GH33} is localized both inside the cell and on the cell surface. In the same locus as BT0455^{GH33} is an ORF (BT0457) that shares 70% identity with the *Tannerella forsythia* SA esterase NanS[39]. BT0457 was shown to have similar activity to NanS against differentially acetylated forms of SA (Supplementary Fig. 8) and is predicted to be periplasmic, providing a likely explanation for the unusual dual surface and periplasmic localization of the sialidase BT0455^{GH33}. This is discussed in more detail in Supplementary Results and Discussion.

CNGs are cleaved from native glycoproteins at the cell surface by GH18 endo-glucosaminidases. Previously characterized members of the GH18 family include endo-β-*N*-acetylglucosaminidases with a range of different specificities against different types of *N*-glycan (Supplementary Table 3). The RNA-Seq data revealed four GH18 family members upregulated during growth on α_1AGp (Fig. 2), which were tested for activity against a variety of *N*-glycosylated proteins including those with different CNG, high-mannose, plant-type and hybrid structures (Supplementary Fig. 9a,b). Surprisingly,

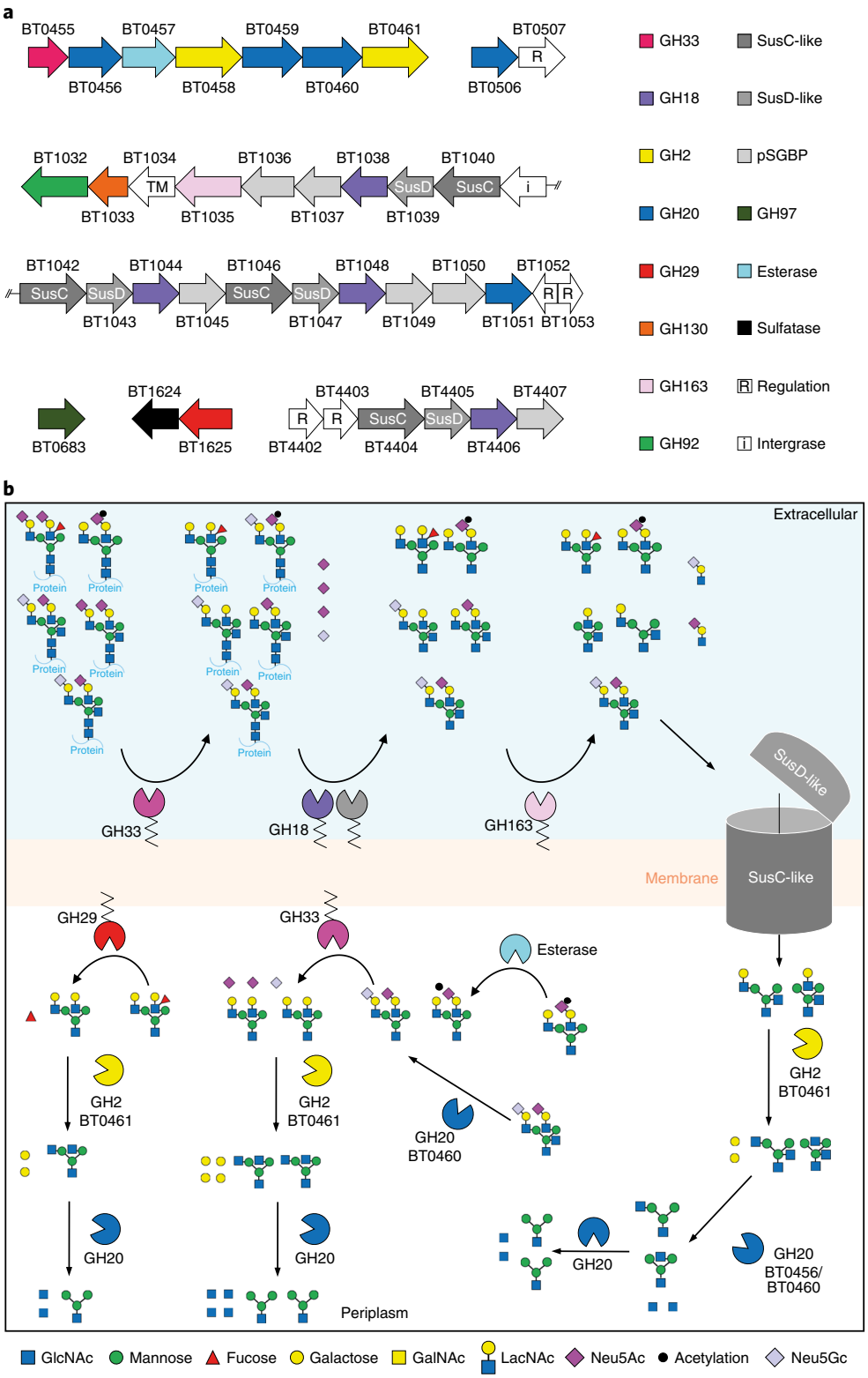

Fig. 2 | Genes upregulated in *Bt* during growth on CNG. a, A schematic of the six CAZyme-containing loci upregulated in *Bt* during growth on α_1AGp. TM, transmembrane protein; pSGBP, putative surface glycan-binding protein. **b**, A model of the degradation of CNGs by *Bt*. BT1625^{GH29} is placed in the periplasm but could in fact be facing the outside of the cell. The degradation of the common core tetrasaccharide likely occurs through the activity of previously characterized enzymes located in the periplasm and cytoplasm (Supplementary Table 2, and Supplementary Results and Discussion).

only BT1044^{GH18} displayed activity against any of the glycoproteins tested, and of the substrates tested this was limited to release of glycans from α_1AGp, IgA and IgG from human serum, and IgA from human colostrum (IgA^s, IgG^s and IgA^c, respectively). None of the four GH18 family members were active on chitin or chito-oligosaccharides (Supplementary Fig. 9c).

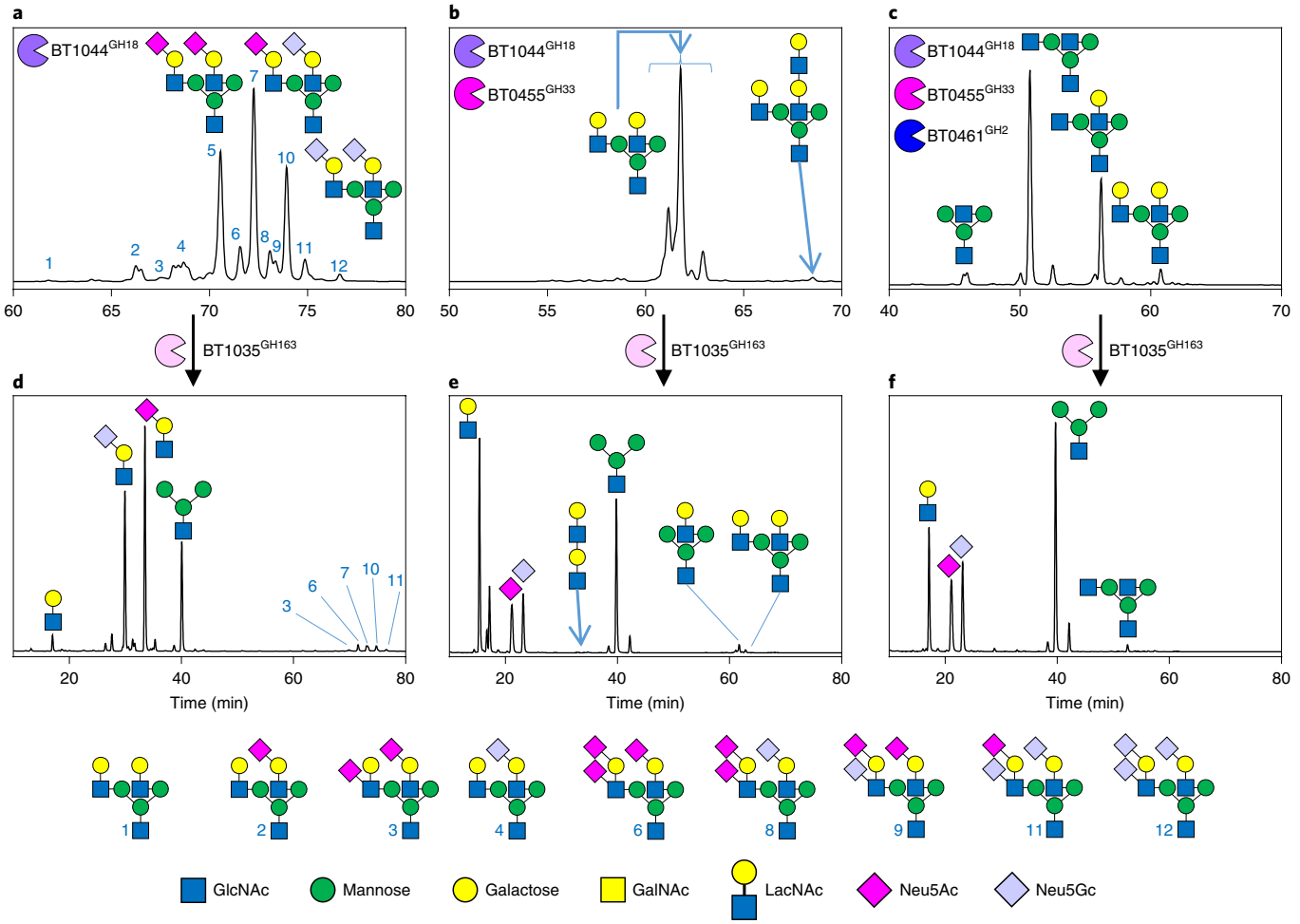

Fig. 3 | The degradation of bi-antennary CNG by recombinant enzymes from *Bt*. a–c, α_1AGp digested with BT1044^{GH18} endo-GlcNAcase alone (**a**) and then in combination with BT0455^{GH33} sialidase (**b**) and then also BT0461^{GH2} β-galactosidase (**c**). **d–f**, Each of the assays shown in **a–c** was stopped after a 24 h incubation using heat denaturation and BT1035^{GH163} was then added. The time shown for the different chromatograms varies between panels to provide clarity of the main peaks. The glycan products were labelled with procainamide and analysed by LC–FLD–ESI–MS (see Methods). The most abundant glycan products are annotated on the chromatograms and the minor products are shown in the key at the bottom. The same glycan species detected in multiple peaks is probably due to different linkages and glycosylation on different arms, which cannot be determined by the analytical methods employed. Neu5Ac and Neu5Gc are pink and light blue, respectively, and the linkages they are attached through were not determined during this study. The results are representative of at least three independent replicates.

To investigate the specificity of BT1044^{GH18} in more detail, products from digested glycoproteins were labelled with the fluorophore procainamide and analysed by liquid chromatography-fluorescence detection-electrospray-mass spectrometry (LC–FLD–ESI–MS) (Figs. 3a,b and 4a–c). The structures of the various CNGs were determined using MS/MS and showed predominantly bi-antennary CNG structures released, including species with bisecting GlcNAcs, antenna fucosylation and polyLacNAc antenna structures. Tri-antennary *N*-glycan can also be removed, but only if the third antenna is limited to a single GlcNAc (for example, Fig. 4, glycan 5). In comparison, PNGaseF treatment (removes all *N*-glycan structures) revealed that tri-antennary CNG structures were present on IgA^s and bovine fetuin, but these were not removed by BT1044^{GH18} (Supplementary Figs. 10 and 11a–c).

Assays against sialylated and desialylated forms of α_1AGp with BT1044^{GH18} showed that the enzyme can act on both, but with a preference for the latter, suggesting that removal of CNGs from the protein mainly occurs after the action of the surface sialidase (Supplementary Fig. 9d). The GH18 EndoE from *Enterococcus faecalis* also has a preference for desialylated *N*-glycans[40], whereas the GH18 EndoS$_2$ from *Streptococcus pyogenes* can remove either type without preference[22] (Supplementary Table 3). BT1044^{GH18} was also shown to tolerate fucosylation of the core GlcNAc (Supplementary Fig. 12a).

A ΔBT1044 strain displayed only a minor growth defect on α_1AGp compared to the wild-type strain (Supplementary Fig. 1g). Furthermore, analysis of surface enzyme activity of ΔBT1044 versus wild-type cells against α_1AGp showed similar product profiles (Supplementary Fig. 7a) and treatment of α_1AGp using both BT0455^{GH33} and BT1044^{GH18} revealed that the level of deglycosylation appeared less than after growth on this glycoprotein (Supplementary Fig. 1a,c). Together, these data suggest that in addition to BT1044^{GH18}, one or more of the other three upregulated GH18 members, or a currently unknown enzyme, are also involved in deglycosylating CNGs in vivo. This apparent redundancy in GH18 activity has been observed previously in *Bacteroides* spp; a *B. fragilis* mutant strain, lacking the main GH18 upregulated during growth on rat transferrin, retained the ability to grow on the serum glycoprotein, albeit at a slower rate[20]. Deletion of the other three *Bt* GH18 genes upregulated on CNGs was

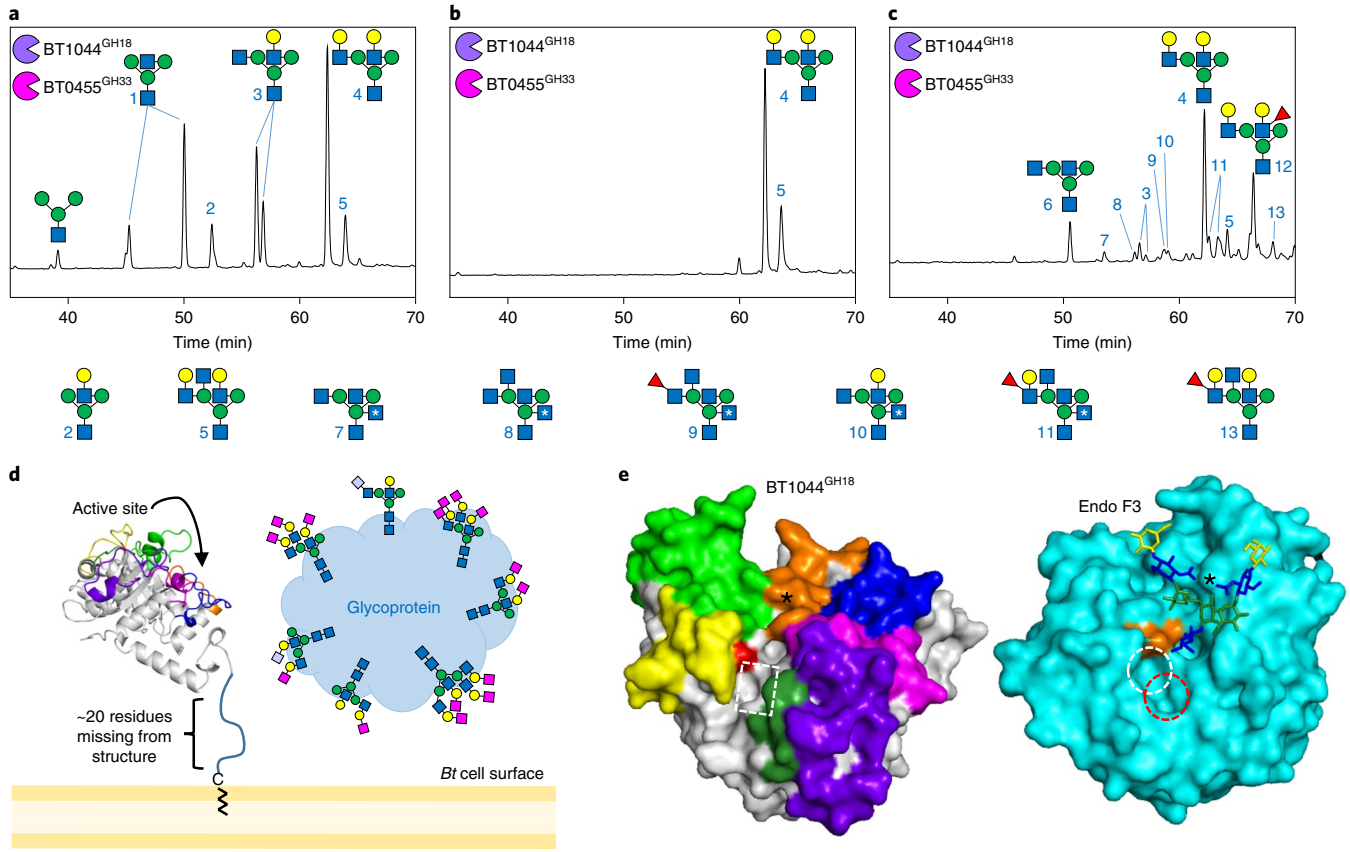

Fig. 4 | Activity of BT1044^{GH18} endo β-GlcNAcase on immunoglobulin substrates and structure of the enzyme. a–c, Activity of BT1044^{GH18} against human serum IgG (**a**), human serum IgA (**b**) and human colostrum IgA (**c**) in the presence of BT0455^{GH33} sialidase. The products with a β1,4 bisecting GlcNAc are indicated by a white asterisk. The results are representative of at least three independent replicates. **d**, A schematic of the predicted orientation of BT1044^{GH18} on the cell surface of *Bt* showing the attachment to the cell surface through a lipid anchor. Loops 1–7 on the surface of the enzyme with the active site are coloured yellow, green, orange, blue, magenta, purple and dark green, respectively. **e**, The surface of the BT1044^{GH18} crystal structure is shown again with the same loop colouring and the two catalytic residues are in red. For comparison, the previously published crystal structure of the CNG-active enzyme EndoF3 (1EOM) from *E. meningoseptica* is shown (cyan). The conserved residues from the DxxDxDxE motif have been coloured orange and the approximate +1 (GlcNAc) and +1′ (fucose) subsites are indicated by white and red dashed lines, respectively. BT1044^{GH18} has more of a groove (white dashed box) in the equivalent section. The area where a bisecting GlcNAc would be predicted to sit is indicated by an asterisk. This space is occupied in Endo F3, whereas in BT1044^{GH18} there is more space to accommodate this sugar, reflecting their respective activities (see Supplementary Fig. 21 for more details).

attempted to explore their role further, but we were unable to obtain these mutants.

The structure of BT1044^{GH18} endo-β-*N*-acetylglucosaminidase. To investigate the structural basis for the substrate specificity displayed by BT1044^{GH18}, the crystal structure of the enzyme was solved to 2.4 Å resolution (Fig. 4d–e). BT1044^{GH18} adopts the canonical $(\beta/\alpha)_8$ barrel (TIM barrel) fold common to many GH structures.

Notably, the amino-terminal region forms a discrete extended structure that would orientate the lipoprotein on the cell surface with the active site facing the extracellular environment (Fig. 4d and Supplementary Results and Discussion). This is probably necessary to prevent spatial restrictions in accessing bulky native glycoprotein substrates. Furthermore, there are ~20 amino acids after the N-terminal lipid anchoring cysteine that are not visible in the structure. It appears likely therefore that this region acts as a relatively unstructured linker between the lipid anchor and the folded N-terminal domain to further extend the distance between the cell surface and the enzyme's active site (Fig. 4d). To understand the substrate specificity in more detail, the structure of BT1044^{GH18} was compared to that of *Elizabethkingia meningoseptica* GH18 (EndoF3) in complex with a bi-antennary CNG product (Fig. 4e and see Supplementary Results and Discussion)[41].

CNG structures are predominantly degalactosylated in the periplasm by BT0461^{GH2}. SA removal from CNGs exposes β1,4-linked galactose (although occasionally this can be β1,3-linked) (Fig. 1)[42]. Inspection of the CAZymes upregulated on α1AGp revealed only two possible β-galactosidase candidates; BT0458^{GH2} and BT0461^{GH2}. As BT0458 has previously been shown to be a β-mannosidase[43], BT0461^{GH2} alone was screened against a variety of defined β-galactose substrates and found to have a preference for the structures and linkages found in CNGs (Supplementary Table 4).

BT0461^{GH2} was tested on BT1044^{GH18}-liberated CNG from α_1AGp pre-treated with BT0455^{GH33} and the products were predominantly fully degalactosylated (Fig. 3c). PNGaseF-liberated and desialylated CNGs from IgA^s, IgG^s and IgA^c substrates were also tested. All of the galactose was removed from the IgG^s and IgA^s glycans, but there was a minor population of galactose-containing structures remaining from the IgA^c sample (Supplementary Fig. 11d–f). The IgA^c glycans with antenna fucosylation on the +1 sugar (GlcNAc) were not degalactosylated, indicating that the fucose was blocking

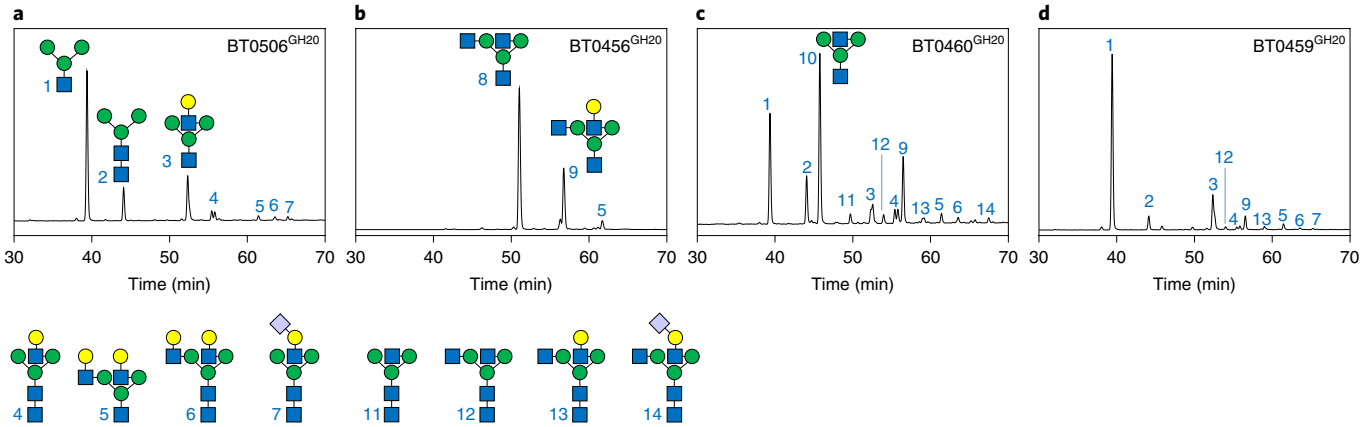

Fig. 5 | GH20 β-hexosaminidase activity against α_1AGp. a–d, Activity was assessed for BT0506^{GH20} (**a**), BT0456^{GH20} (**b**), BT0460^{GH20} (**c**) and BT0459^{GH20} (**d**). Procainamide-labelled products were analysed as described previously for other assays and the samples were pre-digested with BT0455^{GH33}, BT1044^{GH18} and BT0461^{GH2}. The results are representative of at least three independent replicates.

access of BT0461^{GH2}. A lack of BT0461^{GH2} activity against fucosylated LacNAc and LNB (Lewis X and A, respectively) supported the need for antennary fucose removal before galactose release (Fig. 1, Supplementary Fig. 11k and Supplementary Results and Discussion). The antennary fucose, which often decorates CNG structures, is removed by BT1625^{GH29} fucosidase (Supplementary Fig. 11d and Supplementary Table 7). Details of BT1625^{GH29} activity are described in Supplementary Results and Discussion.

The multiple GH20 genes upregulated on α_1AGp display complementary but overlapping activities against CNG. Degalactosylation of CNGs reveals β-linked GlcNAcs that are β1,2-linked in bi-antennary, but β1,4-linked (on the α1,3 arm) and β1,6-linked (on the α1,6 arm) in tri- and tetra-antennary CNGs, respectively (Fig. 1). The core mannose can also have a β1,4-GlcNAc (termed 'bisecting'). GH20 family members are known to be predominantly exo-acting β-hexosaminidases and four of these were expressed during growth on α_1AGp (Fig. 2). Localization studies indicate that all were most likely periplasmic (Supplementary Fig. 7c). Notably, two of the enzymes contain a type II signal sequence suggesting that they are membrane associated but are not trafficked to the cell surface (Supplementary Table 2).

To define the specificity of the four GH20 enzymes, they were tested against CNGs from α_1AGp, IgG^s, IgA^s and IgA^c liberated with PNGaseF and pre-treated with BT0455^{GH33} and BT0461^{GH2} (Fig. 5 and Supplementary Table 6 and Supplementary Fig. 11e–g). Activity of the GH20 family members was also tested against CNGs released by PNGase from IgG^s with no pre-treatment with BT0455^{GH33} and BT0461^{GH2} (Supplementary Fig. 13).

BT0506^{GH20} is specific and highly active against the antenna GlcNAcs, but failed to have any activity against the bisecting β1,4-GlcNAc. BT0459^{GH20} is also predominantly specific for a range of antenna GlcNAcs; however, trace activity on bisecting structures could be detected. The difference between these enzymes is that BT0459^{GH20} struggles to remove both antenna GlcNAcs, usually when a bisecting GlcNAc is present, indicative of a steric block (Supplementary Fig. 11, for example, glycans 28 and 31).

In contrast, BT0456^{GH20} and BT0460^{GH20} were both able to remove the bisecting GlcNAc when samples had SA and galactose removed. Any further antenna decoration meant that BT0456^{GH20} was no longer active on bisecting structures. BT0460^{GH20}, however, could remove bisecting GlcNAc even when both antennae had galactose and SA present (Supplementary Fig. 13 and Supplementary Table 6).

In terms of antennary deglycosylation, BT0460^{GH20} was able to remove all GlcNAc in the substrates tested, while products from the BT0456^{GH20} had a high proportion with only one uncapped GlcNAc remaining, indicating a preference for one arm. Furthermore, if the one antenna has a galactose (or SA as well) cap, then BT0456^{GH20} cannot not access the GlcNAc on the other antenna, again indicating a preference for one arm or a steric block (Supplementary Fig. 13, glycans 4, 6 and 11 remaining). Overall, these data reveal that the four *Bt* GH20 family members characterized here display complementary but overlapping activities that facilitate access to the diverse GlcNAc linkages present on CNGs.

Crystal structure of BT0459^{GH20} β-hexosaminidase. The structure of BT0459^{GH20} was solved to 2.4 Å with GlcNAc product present in the −1 subsite. The enzyme is made up of four domains—a small N-terminal domain, a $(\beta/\alpha)_8$ barrel catalytic domain, an FN3 domain and a carboxy-terminal F5/F8 type C domain that are oriented overall to form a hook-like shape with the C-terminal domain facing the active site (Fig. 6). The potential role of the C-terminal F5/F8 type C domain in substrate binding is discussed in Supplementary Results and Discussion.

Compared to other known GH20 structures, BT0459^{GH20} has a very open active site structure with only the −1 site, indicating little interaction with the +1 sugar and potentially providing a rationale for the broad substrate specificity displayed by this enzyme (Fig. 6a,b and Supplementary Fig. 14–16). A comparison of this active site with CNG-active *Sp*GH20A and *Sp*GH20B from *Streptococcus pneumoniae* (Fig. 6c,d)[44] and also a GH20 from *Ostrinia furnacalis* (Fig. 6e) specific for chitooligosaccharides[45] highlights the unusual openness of the BT0459^{GH20} active site (Supplementary Results and Discussion).

BT1035^{GH163} is a surface-located enzyme with CNG-specific endo-GlcNAcase activity. Within the large CNG-upregulated locus is an ORF (BT1035) of unknown function that sequence analysis suggests is a glycoside hydrolase. It is predicted to have an N-terminal DUF4838 domain that has distant homology to GH20 and a C-terminal F5/F8 type C domain. Incubation of recombinant BT1035^{GH163} against BT1044^{GH18}-liberated CNG from α_1AGp showed that it was able to release sialylated LacNAc (both Neu5Ac and Neu5Gc) (Fig. 3a,d). BT1035^{GH163} was able to act on the CNG while still attached to the protein, but displayed a preference for the GH18-released glycan (Supplementary Fig. 17c). The enzyme was also capable of hydrolysing the same GlcNAc-β1,2-mannose linkage in desialylated α_1AGp (Fig. 3b,e). Comparison of BT1035^{GH163} activity against sialylated and desialylated CNGs showed that desialylated CNG was a more favourable substrate, possibly indicating

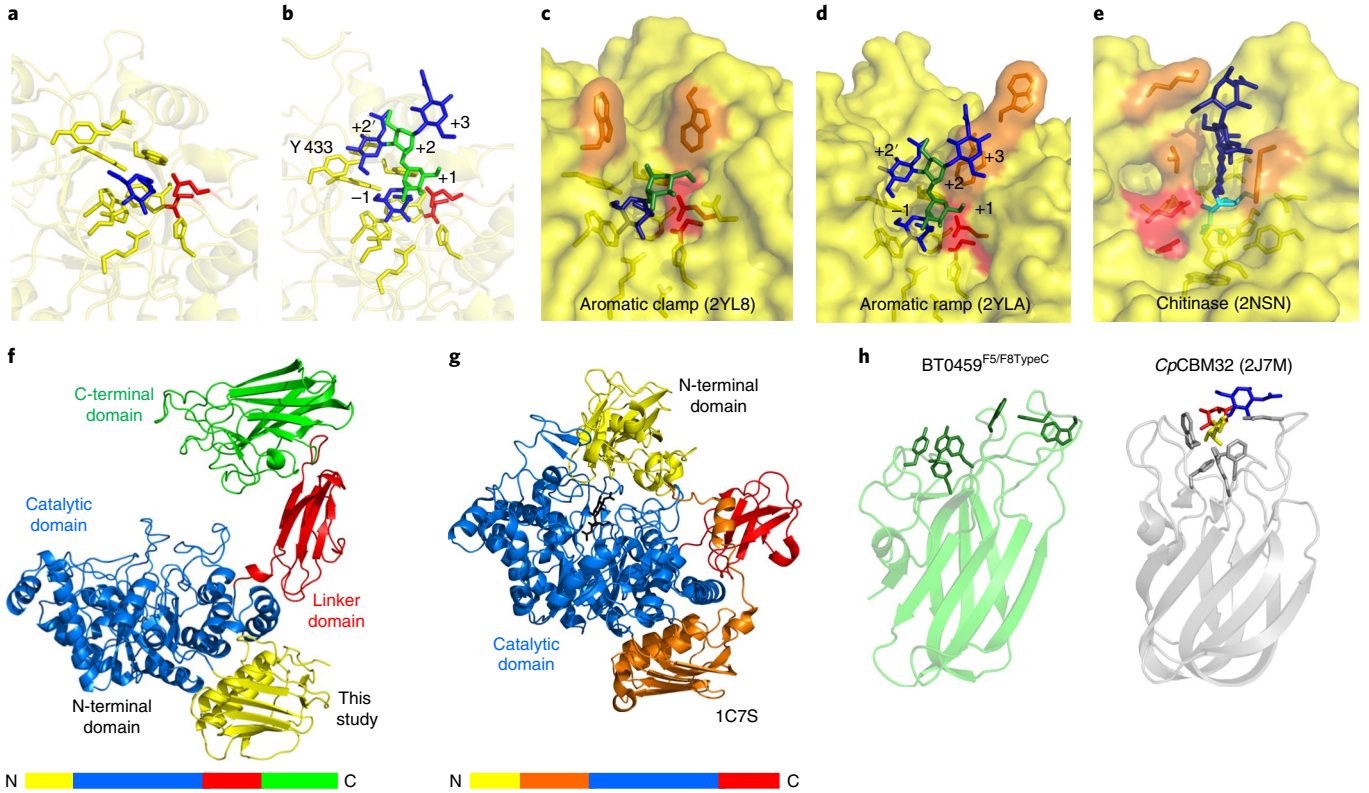

Fig. 6 | Crystal structure of BT0459^{GH20} a, A close-up of the active site of BT0459^{GH20} with the residues forming the −1 subsite shown as sticks and a GlcNAc (blue) product in the active site. **b**, A close-up of the active site of BT0459^{GH20} overlaid with a CNG structure (from 2YLA). Mannose and GlcNAc are green and blue, respectively, with the antennary GlcNAc in the active site. The α1,3 mannose arm is in the +1 position and the core mannose and GlcNAc occupy the +2 and +3 subsites, respectively. A clash can be seen between the Tyr 433 residue and the bisecting GlcNAc in the +2′ position. **c**–**e**, The active sites of the *S. pneumoniae* *Sp*GH20A (**c**) and *Sp*GH20B (**d**) and a GH20 active on chitooligosaccharides from *O. furnacalis* (**e**). The inhibitor in **e** is TMG-chitotriomycin (TMG and GlcNAcs are shown in cyan and dark blue, respectively). Catalytic residues are shown in red and those residues interacting with sugars in the positive subsites are in orange. **f**,**g**, The full-length structure of BT0459^{GH20} (**f**) and the *S. marcescens* GH20 (1C7S) (**g**). The order of the modules is shown as coloured bars. **h**, The C-terminal F5/F8 type C domain of BT0459^{GH20} (BT0459$^{F5/F8TypeC}$, left); aromatic residues on the potential glycan-binding surface are shown as sticks. A structural homologue of BT0459$^{F5/F8TypeC}$ is shown for comparison (right) and is a CBM32 from a *Clostridium perfringens* GH84 (*Cp*CBM32). This has a trisaccharide ligand of fucose, galactose and GlcNAc (red, yellow and blue, respectively) bound and the aromatic residues involved in binding are shown as sticks. The core folds between BT0459$^{F5/F8TypeC}$ and *Cp*CBM32 are very similar but the potential glycan-binding surfaces vary greatly (see Supplementary Fig. 12 for further structural homologues).

that SA attached through particular linkages cannot be accommodated by the enzyme (Supplementary Fig. 18). CNG structures with SA linked to the GlcNAc were not hydrolysed by BT1035^{GH163}, indicating that the enzyme has no −1′ subsite (Fig. 3d, no SA-GlcNAc products). BT1035^{GH163} was also able to remove larger diLacNAc structures from CNG antenna; however the enzyme could not degrade the diLacNAc product further, suggesting that either the +1 subsite cannot tolerate galactose, or has a requirement for mannose (that is, is CNG-specific), or the enzyme is linkage-specific (Fig. 3b,e).

Using CNGs pre-treated with BT0455^{GH33} and BT0461^{GH2} revealed that BT1035^{GH163} was also able to hydrolyse the GlcNAc-β1,2-mannose linkage to almost totally remove both the remaining GlcNAc and LacNAc, showing that the galactose in the −2 subsite is not an absolute requirement for activity (Fig. 3c,f). Notably, however, BT1035^{GH163} was unable to cleave the GlcNAc-β1,2-mannose disaccharide (Supplementary Fig. 17a), indicating that interactions with a +2 subsite are required for activity (that is, with the core mannose), although the context of this sugar would vary depending on the specific arm targeted. Activity against IgA and IgG showed that bisecting GlcNAc does not interfere with BT1035^{GH163} hydrolysis (Supplementary Fig. 11h–j).

The capacity of BT1035^{GH163} to remove LacNAc, sialyl LacNAc and diLacNAc from CNGs and the lack of activity against the GlcNAc–mannose disaccharide indicate that BT1035^{GH163} displays an endo-like activity that targets the β-GlcNAc–mannose found in these structures.

Analysis of the cellular location of BT1035^{GH163} revealed that the enzyme is on the cell surface, providing an explanation for the release of a sialylated-LacNAc structure (predominantly Neu5Gc) outside the cell during growth on α_1AGp (Supplementary Figs. 1, 3 and 4). Growth data suggest that BT0455^{GH33} and BT1035^{GH163} are both active in vivo from early to mid-exponential onwards, with SA released slightly earlier (Supplementary Fig. 1). Whole-cell assay data also show that BT0455^{GH33} and BT1035^{GH163} both act rapidly on a_1AGp in vitro, but the sialylated-LacNAc structures produced by BT1035^{GH163} remain sialylated. This suggests that BT1035^{GH163} is removing glycan structures that BT0455^{GH33} cannot in vivo and a biological rationale for this activity on the cell surface of BT1035^{GH163} may be to increase the proportion of CNG structures that can be accessed. It may also be that CNGs with particular structures are poor substrates for the SusCD glycan-import apparatus and thus removal of some trisaccharides by BT1035^{GH163} facilitates more efficient uptake of these glycans (see Supplementary Results and Discussion).

Discussion

This study characterizes the pathway for CNG breakdown by a prominent member of the normal human gut microbiota. CNGs are highly variable in terms of sugar composition and linkages, displaying extensive heterogeneity even within a single glycoprotein[46]. For example, human IgG is decorated with 18 different *N*-glycan structures, few of which are shared with the other glycoproteins tested here.

Here we describe the extensive enzyme apparatus *Bt* uses to deal with this heterogeneity. We show that significant processing of predominantly bi-antennary forms of CNG occurs at the cell surface before import, including removal of glycan from the protein, desialylation and the action of BT1035^{GH163}. The exact biological advantage that BT1035^{GH163} provides could not be fully elucidated in this study, but one role may be to increase the variety of CNGs that *Bt* can access. After removal of galactose and antennary fucose, the complementary but overlapping activities of the four different GH20 enzymes allow access to a diverse range of antennary and bisecting GlcNAc structures.

Notably, the broad specificity of the exo-acting enzymes could be a reflection of their role in the degradation of multiple types of glycan with conserved structural features in common. Thus, many of the enzymes involved in CNG breakdown in *Bt* also appear to play a role in the degradation of *O*-glycans; both the BT0455–BT0461 locus and BT1624–BT1625 have been shown to be expressed during growth of *Bt* on mucin[26,47]. It is possible that the shared structures contained within *O*-glycans and CNGs (that is, β-linked galactose and GlcNAcs, as well as capping SA and fucose) targeted by these enzymes and the consistent presence of these glycans from host sources in the gut (that is, less variable that dietary glycans) have driven the overlapping regulation of these enzymes in *Bt*.

Related to this, a previous study indicates that the anti-sigma regulator of the core BT1032–BT1053 CNG PUL, BT1053, directly controls expression of the whole multi-locus CNG-degradation apparatus[48]. Thus, when BT1053 was deleted to derepress the genes under its control, the same loci were upregulated as were seen here to be activated by CNGs. As the sensor-regulator for the core BT1032–BT1053 PUL is an ECF-sigma/anti-sigma system, the activating signal for expression of the whole *Bt* multi-locus CNG apparatus will be import of discrete CNG structures by the core-PUL-encoded SusCD.

The ability of *Bacteroides* spp. to deglycosylate secretory IgA is an interesting finding of this work that raises the possibility that this process plays a role in modulating the function of this important gut immune molecule and is an intriguing avenue of research to pursue in the future[11,49,50].

Methods

Sources of glycans and glycoproteins. The glycoproteins bovine α_1AGp, bovine fetuin, bovine RNaseB, chicken egg white ovalbumin, horseradish peroxidase, human serum IgG, human serum IgA, human colostrum IgA and *p*-nitrophenyl monosaccharides were obtained from Sigma. Purified di- and oligosaccharides were obtained from Carbosynth. The sialic acid reference panel was obtained from Ludger. Squid chitin was a gift from G. Davies (University of York, UK).

Bacterial strains. The *Bacteroides* strains used were as follows: *B. thetaiotaomicron* VPI-5482, *B. fragilis* NCTC9342, *B. caccae* ATCC43185, *B. cellulosilyticus* DSM14838, *B. massiliensis* DSM17679, *B. finegoldii* DSM17565, *B. vulgatus* ATCC8483, *B. ovatus* ATCC8482 and *B xylanisolvans* XBA1.

Cloning, expression and purification of recombinant proteins. DNA encoding the appropriate genes (excluding the signal sequences) was amplified from genomic DNA using appropriate primers and cloned into pET28b (Novagen) using NheI–XhoI restriction sites. His-tags were located at the N terminus. Recombinant plasmids were transformed into TUNER (Novagen) cells in LB broth containing 10 μg ml^{-1} kanamycin at 37 °C with shaking at 180 r.p.m. One-litre cultures were grown to mid-exponential phase in 2 l baffled flasks, cooled to 16 °C and isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.2 mM. These cells were then incubated for 16 h at 16 °C in an orbital shaker at 150 r.p.m. Recombinant His-tagged protein was purified from cell-free extracts using immobilized metal-affinity chromatography (using Talon resin; Clontech) as described previously[51].

The purity and size of the proteins were checked using SDS–PAGE and their concentrations were determined using absorbance at 280 nm (NanoDrop 2000c; Thermo Scientific) and their molar extinction coefficients[52].

Purification of proteins for crystallization. Proteins purified by immobilized metal-affinity chromatography were further purified by size-exclusion chromatography using a HiLoad Superdex 200 pg on an AKTA Pure FPLC system (GE Healthcare Life Sciences). The purity of the fractions was determined using SDS–PAGE and those of high enough purity were pooled and concentrated to ~10 mg ml^{-1}.

Crystallization. We used the vapour-diffusion sitting-drop method to screen crystallization conditions. Seleno-methionine (SeMet) crystals for BT1044^{GH18} were obtained in 100 mM imidazole pH 8.0 and 10% polyethylene glycol (PEG) 8000. The samples were cryo-protected with Paratone-N oil. Additional SeMet BT1044^{GH18} crystals were obtained in 200 mM sodium fluoride, 100 mM bis-tris propane pH 6.5 and 20% PEG 3350. These samples were cryo-protected with the addition of 20% PEG 400 to the reservoir solution. BT0459^{GH20} crystals were obtained in 100 mM sodium acetate, 100 mM bis-tris propane pH 7.5 and 10% PEG 3350.

Data collection, structure solution, model building, refinement and validation. A three-wavelength interleaved multi-wavelength anomalous diffraction experiment using SeMet derivatives cryo-protected with oil was collected at beamline I03, Diamond Light Source (UK). The data were indexed and integrated with XDS[53]. Space-group determination was confirmed with Pointless and the data were scaled with Aimless[54]. The phase problem for BT1044^{GH18} was solved by autoSHARP[55]. ShelxD[56] found the requested 8 selenium sites with a final correlation coefficient (*E*) of 0.59. Subsequent density modification with Parrot[57] and automated model building with Buccaneer[58] placed 292 sequenced residues out of 364. Once the phase problem was solved, higher-resolution datasets were obtained on the BT1044^{GH18} SeMet samples cryo-protected with PEG 400 (see above). The data were processed as above and the initial experimental phasing model was refined using Refmac5[59] and manually built using Coot[60]. The data processing and refinement statistics are reported in Supplementary Table 8; PDB accession 6Q64. The phase problem for BT0459^{GH20} was solved by using MrBump[61]. Briefly, the search model 3RCN was prepared with Molrep[62] and used in Phaser[63] all automated through MrBump[61]. The initial phases obtained from MrBump were improved by density modification using Parrot[57] and the model was automatically built using Buccaneer[58]. For all models, iterative cycles of model building with Coot[60] and refinement with Refmac5[59] were stopped when the validation with Coot[60] and Molprobity[64] gave acceptable values. Five percent of the data were randomly selected for Rfree calculation. Structural figures were made using Pymol[65] and all other programs used were from the CCP4 suite[66]. The data processing and refinement statistics are reported in Supplementary Table 8; PDB accession 6Q63.

Growth of *Bacteroides* species. Starter cultures for most *Bacteroides* spp. were grown in tryptone–yeast-extract–glucose medium with the addition of haematin and inoculated from glycerol stocks[32]. *Bacteroides massiliensis* and *xylanisolvans* were grown on chopped meat broth as described previously[9,67]. Cells were typically grown in 5 ml cultures in glass test tubes in an anaerobic cabinet (Whitley A35 Workstation; Don Whitley) and were monitored at $OD_{600\,nm}$ using a Biochrom WPA cell density meter. Growth tests using 20 mg ml^{-1} glycoprotein as the sole carbon source were carried out in minimal medium[33] in either 5 ml cultures (to allow for taking samples) or 600 μl cultures in a 96-well plate using a Biotek Epoch plate reader. The removal of glycan from peptide was monitored by SDS–PAGE with either a Coomassie staining or Pierce Glycoprotein staining kit (ThermoFisher Scientific), to detect just protein or glycoprotein, respectively. Replicates are presented individually in the figures and not as averages.

RNA extraction, sequencing and data processing. A starter culture was used to inoculate minimal medium (5 ml) of either 5 mg ml^{-1} glucose or 20 mg ml^{-1} α_1AGp, which was then grown to OD_{580nm} 0.4–0.6. Cells were collected and stored in RNAProtect (QIAGEN) and the RNA was extracted using the RNeasy Mini Kit (Qiagen). Sample processing, library preparation and sequencing took place at Oxford Genomics Centre (University of Oxford, Oxford, UK). Data processing was done by The Bioinformatics Support Unit at Newcastle University using Bowtie2 for alignments[68] and genome annotation completed using the Ensembl database[69]. The read counts aligning to the genomic features were obtained and differential expression analysis was done using R packages[70]. This data presented in Supplementary Table 1 have been submitted to https://www.ncbi.nlm.nih.gov/geo/ and have the accession number GSE129572.

Genetic manipulation. Gene deletion mutations were produced by allelic exchange using the pExchange vector[6].

Cellular localization. Cultures (15 ml) were grown on minimal medium with α_1AGp (20 mg ml^{-1} final) to mid-exponential growth, collected by centrifugation, washed in PBS and resuspended in 2 ml PBS. The cells were split into two aliquots. One half had Proteinase K (2 mg ml^{-1} final) added and was incubated at 37 °C for 2 h and one half was the control. The Proteinase K sample was collected by centrifugation, washed in PBS and resuspended in 1 ml. Trichloroacetic acid (200 µl) was added to all samples to precipitate out the protein, incubated for 30 min on ice, washed four times with ice-cold acetone (1 ml) and cell pellets were resuspended in 0.5 ml PBS. Samples (15 µl) were run on SDS–PAGE gels with MagicMark XP Western Protein Standard (Thermo Fisher Scientific) and then transferred to Whatman Protran BA 85 nitrocellulose membrane. Specific proteins were detected using anti-sera raised against recombinant versions in rabbit (Eurogentec). The secondary antibody used was a goat anti-rabbit HRP conjugate (SC-2004, Santa Cruz). Antibodies were detected by chemiluminescence using Biorad Clarity Western ECL Substrate.

Whole-cell assays to analyse cell surface enzyme activity. Cultures (15 ml) were grown on minimal medium with α_1AGp (20 mg ml^{-1} final) to mid-exponential and collected by centrifugation. Cells were washed twice in PBS and resuspended in 1.75 ml PBS. To assess the cell surface activity, 250 µl of the resuspended cells were mixed with 250 µl of fresh α_1AGp added (20 mg ml^{-1} final) and a 50 µl aliquot was taken at $T = 0$, 0.25, 0.5, 1, 2, 3, 4 and ~16 h. One aliquot was pelleted again and resuspended in 250 µl of BugBuster (Merck Millipore) to lyse the cells and then mixed with 250 µl of fresh α_1AGp (20 mg ml^{-1} final). Several controls were also included. Cells (250 µl) were mixed with 250 µl PBS, 250 µl of the spent medium from the growth was also mixed with 250 µl of fresh α_1AGp (20 mg ml^{-1} final) and a 500 µl aliquot of cells was mixed with 500 µl of PBS. For the last control, two 50 µl aliquots were taken at every time point and one was checked for activity and one had the cells removed by centrifugation and the supernatant kept on ice. The supernatant was tested for overnight activity against fresh α_1AGp (20 mg ml^{-1} final) to check there was no cell lysis during the time course. Cells (750 µl) were then treated with Proteinase K (2 mg ml^{-1} final) for 2 h and again collected by centrifugation, washed twice and resuspended back to 750 µl. A 250 µl aliquot of these cells was mixed with α_1AGp (20 mg ml^{-1} final) and the same time course was run. To check that all of the Proteinase K had been removed, a 250 µl aliquot of the treated cells was mixed with 250 µl of bovine serum albumin (20 mg ml^{-1} final). The treated cells were also lysed using BugBuster and assayed against α_1AGp (20 mg ml^{-1} final).

Thin-layer chromatography. Between 3 and 12 µl of reactions were spotted in 3 µl aliquots on to silica plates and resolved in butanol/acetic acid/water (2:1:1). Sugars were visualized using diphenylamine–aniline–phosphoric acid stain[71].

Recombinant enzyme assays. Kinetic data was obtained using continuous assays with either *p*-nitrophenyl-linked sugars (monitored at 400 nm) or galactose, mannose or fucose detection kits (Megazyme International), as described previously[8]. To assay the enzymes against CNG, reactions were left overnight (~16 h) unless otherwise stated. The activities of the recombinant enzymes (1 µM final unless otherwise stated) were typically assayed in 20 mM MOPS pH 7, at 37 °C, with a final substrate concentration of 20 mg ml^{-1} (for glycoproteins) or 1 mM (for di- and oligo-saccharides). Exceptions to this were the final concentrations of IgG, serum IgA and colostrum IgA, which were 5 mg ml^{-1}, 0.5 mg ml^{-1} and 2.5 mg ml^{-1}, respectively. All kinetic assays were performed at least in triplicate on multiple occasions and data presented are means and standard deviations. Total *N*-glycans were released from glycoproteins using PNGaseF (Sigma), purified by centrifugation through a 10 kDa protein concentrator, freeze-dried and resuspended in the required volume.

Esterase activity. A sialic acid reference panel containing differently acetylated sialic acid species was digested with 1 µM of BT0457 overnight at 37 °C, derivatized with DMB (Ludger) and analysed using reversed-phase high-performance liquid chromatography (HPLC). Here, 25 µl of sample was injected into the LudgerSep R1 column (4.6 × 150 mm, 3 µm particle size) at 30 °C on a Waters Alliance HPLC instrument with fluorescence detection ($\lambda_{ex} = 373$ nm, $\lambda_{em} = 448$ nm). Mobile phase A was a methanol/acetonitrile/water solution (7:9:84) and mobile phase B was acetonitrile. Analytes were eluted using an isocratic flow at 100% mobile phase A running at a flow rate of 0.5 ml min^{-1} for 19 min. A column wash was accomplished by increasing the proportion of mobile phase B to 90%.

High-performance anion exchange chromatography with pulsed amperometric detection. High-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was used to analyse BT1035^{GH163} activity against sialylated and desialylated CNGs. Samples were separated using a CARBOPAC PA-100 anion exchange column with a CARBOPAC PA-100 guard column (Thermo Fisher Scientific). Flow was 1 ml min^{-1} and elution conditions were 0–20 min, 20 mM NaOH; 20–80 min, 100 mM NaOH with a 0–500 mM sodium acetate gradient. LacNAc was quantified using a standard curve generated using a range of LacNAc standards between 0.1 and 0.01 mM.

Procainamide labelling. *N*-glycans were fluorescently labelled by reductive amination using a procainamide labelling kit containing sodium cyanoborohydride as reductant (Ludger). Before analysis, derivative glycans were cleaned-up from excess reagents using SPE clean-up plates (Ludger).

Analysis of procainamide-labelled glycans. Procainamide-labelled glycans were analysed by LC–FLD–ESI–MS. Here, 25 µl of each sample was injected into a Waters ACQUITY UPLC Glycan BEH Amide column (2.1 × 150 mm, 1.7 µm particle size, 130 Å pore size) at 40 °C on a Dionex Ultimate 3000 UHPLC instrument with a fluorescence detector ($\lambda_{ex} = 310$ nm, $\lambda_{em} = 370$ nm) attached to 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 a gradient running from 85 to 57% mobile phase B over 105 min at a flow rate of 0.4 ml min^{-1}, with the exception of IgG CNG used for screening of GH20 enzymes, where separation was accomplished by a gradient running from 76 to 58% mobile phase B over 70 min. The Amazon Speed was operated in the positive sensitivity mode using the following settings: source temperature, 180 °C; gas flow, 4 l min^{-1}; capillary voltage, 4,500 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 1,700.

Analysis of mass spectrometry data. Procainamide-labelled glycans were analysed using Bruker Compass Data Analysis software and GlycoWorkbench[72]. Glycan compositions were elucidated on the basis of MS^2 fragmentation and previously published data.

Bioinformatics. Putative signal sequences were identified using LipoP[73]. Alignments and sequence identities were determined using Clustal Omega using the full-length protein sequences, not individual modules[74]. The IMG database was used to analyse genomic locations and synteny[75]. The CAZy database (www.cazy.org) was used as the main reference for carbohydrate-active enzyme activity[76]. Dali[77] and PDBefold[78] were used to carry out structural homology searches of protein modules. Phylogenetic trees were constructed using Phylogeny.fr[79,80]. Pfam[81] and SMART[82,83] were used to determine module boundaries and look at the prevalence of modules in other proteins.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The full RNA-Seq data are provided in Supplementary Table 1 and have also been submitted to https://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE129572. The crystal structures are deposited in the Protein Data Bank under the accession numbers 6Q63 and 6Q64. The other data that support the findings in this paper are available upon request from the corresponding authors.

Received: 21 December 2018; Accepted: 24 April 2019;
Published online: 3 June 2019

References

1. McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39**, 338–342 (1984).
2. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
3. Khoruts, A., Dicksved, J., Jansson, J. K. & Sadowsky, M. J. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent *Clostridium difficile*-associated diarrhea. *J. Clin. Gastroenterol.* **44**, 354–360 (2010).
4. O'Keefe, S. J. et al. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. *J. Nutr.* **139**, 2044–2048 (2009).
5. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
6. 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).
7. Luis, A. S. et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat. Microbiol.* **3**, 210–219 (2018).
8. Cuskin, F. et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).
9. Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
10. Chung, C. Y., Majewska, N. I., Wang, Q., Paul, J. T. & Betenbaugh, M. J. SnapShot: N-glycosylation processing pathways across kingdoms. *Cell* **171**, 258–258 (2017).
11. Mathias, A. & Corthesy, B. *N*-Glycans on secretory component: mediators of the interaction between secretory IgA and gram-positive commensals sustaining intestinal homeostasis. *Gut Microbes* **2**, 287–293 (2011).

12. Corfield, A. P. The interaction of the gut microbiota with the mucus barrier in health and disease in human. *Microorganisms* **6**, 78 (2018).
13. Mestecky, J., Russell, M. W., Jackson, S. & Brown, T. A. The human IgA system: a reassessment. *Clin. Immunol. Immunopathol.* **40**, 105–114 (1986).
14. Hughes, G. J., Reason, A. J., Savoy, L., Jaton, J. & Frutiger-Hughes, S. Carbohydrate moieties in human secretory component. *Biochim. Biophys. Acta* **1434**, 86–93 (1999).
15. Garrido, D. et al. Endo-beta-*N*-acetylglucosaminidases from infant gut-associated bifidobacteria release complex *N*-glycans from human milk glycoproteins. *Mol. Cell. Proteomics* **11**, 775–785 (2012).
16. Sebaihia, M. et al. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**, 779–786 (2006).
17. Bohle, L. A., Mathiesen, G., Vaaje-Kolstad, G. & Eijsink, V. G. An endo-beta-*N*-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type *N*-linked glycans. *FEMS Microbiol. Lett.* **325**, 123–129 (2011).
18. Renzi, F. et al. The *N*-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus* deglycosylates human IgG. *PLoS Pathog.* **7**, e1002118 (2011).
19. Collin, M. & Fischetti, V. A. A novel secreted endoglycosidase from *Enterococcus faecalis* with activity on human immunoglobulin G and ribonuclease B. *J. Biol. Chem.* **279**, 22558–22570 (2004).
20. Cao, Y., Rocha, E. R. & Smith, C. J. Efficient utilization of complex *N*-linked glycans is a selective advantage for *Bacteroides fragilis* in extraintestinal infections. *Proc. Natl Acad. Sci. USA* **111**, 12901–12906 (2014).
21. Robb, M. et al. Molecular characterization of *N*-glycan degradation and transport in *Streptococcus pneumoniae* and its contribution to virulence. *PLoS Pathog.* **13**, e1006090 (2017).
22. Sjogren, J. et al. EndoS2 is a unique and conserved enzyme of serotype M49 group A *Streptococcus* that hydrolyses *N*-linked glycans on IgG and alpha1-acid glycoprotein. *Biochem. J.* **455**, 107–118 (2013).
23. Collin, M. & Olsen, A. EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. *EMBO J.* **20**, 3046–3055 (2001).
24. Dupoiron, S. et al. The *N*-glycan cluster from *Xanthomonas campestris* pv. *campestris*: a toolbox for sequential plant *N*-glycan processing. *J. Biol. Chem.* **290**, 6022–6036 (2015).
25. Forster, S. C. et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* **37**, 186–192 (2019).
26. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
27. Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
28. Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
29. Bagenholm, V. et al. Galactomannan catabolism onferred by a polysaccharide utilization locus of *Bacteroides ovatus*: enzyme synergy and crystal structure of a beta-mannanase. *J. Biol. Chem.* **292**, 229–243 (2017).
30. Tamura, K. et al. Molecular mechanism by which prominent human gut bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides. *Cell Rep.* **21**, 417–430 (2017).
31. Temple, M. J. et al. A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan degradation: unique substrate conformation drives specificity of the key endo-1,6-beta-glucanase. *J. Biol. Chem.* **292**, 10639–10650 (2017).
32. Larsbrink, J. et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502 (2014).
33. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
34. Reeves, A. R., Wang, G. R. & Salyers, A. A. Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **179**, 643–649 (1997).
35. Thomas, F. et al. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut *Bacteroides*. *Environ. Microbiol.* **14**, 2379–2394 (2012).
36. Muchmore, E. A., Diaz, S. & Varki, A. A structural difference between the cell surfaces of humans and the great apes. *Am. J. Phys. Anthropol.* **107**, 187–198 (1998).
37. Park, K. H. et al. Structural and biochemical characterization of the broad substrate specificity of *Bacteroides thetaiotaomicron* commensal sialidase. *Biochim. Biophys. Acta* **1834**, 1510–1519 (2013).
38. Almagro-Moreno, S. & Boyd, E. F. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evolut. Biol.* **9**, 118 (2009).
39. Phansopa, C. et al. Characterization of a sialate-*O*-acetylesterase (NanS) from the oral pathogen *Tannerella forsythia* that enhances sialic acid release by NanH, its cognate sialidase. *Biochem. J.* **472**, 157–167 (2015).
40. Garbe, J. et al. EndoE from *Enterococcus faecalis* hydrolyzes the glycans of the biofilm inhibiting protein lactoferrin and mediates growth. *PloS One* **9**, e91035 (2014).
41. Waddling, C. A., Plummer, T. H. Jr., Tarentino, A. L. & Van Roey, P. Structural basis for the substrate specificity of endo-beta-*N*-acetylglucosaminidase F(3). *Biochemistry* **39**, 7878–7885 (2000).
42. Green, E. D., Adelt, G., Baenziger, J. U., Wilson, S. & Van Halbeek, H. The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy. *J. Biol. Chem.* **263**, 18253–18268 (1988).
43. Tailford, L. E. et al. Mannose foraging by *Bacteroides thetaiotaomicron*: structure and specificity of the beta-mannosidase, BtMan2A. *J. Biol. Chem.* **282**, 11291–11299 (2007).
44. Pluvinage, B. et al. Inhibition of the pneumococcal virulence factor StrH and molecular insights into *N*-glycan recognition and hydrolysis. *Structure* **19**, 1603–1614 (2011).
45. Liu, T. et al. Structural determinants of an insect beta-*N*-acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. *J. Biol. Chem.* **286**, 4049–4058 (2011).
46. Theodoratou, E. et al. Glycosylation of plasma IgG in colorectal cancer prognosis. *Sci. Rep.* **6**, 28098 (2016).
47. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
48. Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J. Biol. Chem.* **284**, 18445–18457 (2009).
49. Kubinak, J. L. et al. MyD88 signaling in T cells directs IgA-mediated control of the microbiota to promote health. *Cell Host Microbe* **17**, 153–163 (2015).
50. Lawrence, R. M. & Pane, C. A. Human breast milk: current concepts of immunology and infectious diseases. *Curr. Probl. Pediatr. Adolesc. Health Care* **37**, 7–36 (2007).
51. Charnock, S. J. et al. Key residues in subsite F play a critical role in the activity of *Pseudomonas fluorescens* subspecies cellulosa xylanase A against xylooligosaccharides but not against highly polymeric substrates such as xylan. *J. Biol. Chem.* **272**, 2942–2951 (1997).
52. Gasteiger E. et al. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2015).
53. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
54. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
55. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
56. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
57. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
58. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
59. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
60. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
61. Keegan, R. M. & Winn, M. D. Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. D* **63**, 447–457 (2007).
62. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
63. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
64. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
65. The PyMOL Molecular Graphics System v.2.0 (Schrödinger LLC, 2017).
66. Collaborative Computational Project, Number 4 The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
67. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19786–19791 (2012).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
69. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–d761 (2018).
70. R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, GBIF, 2018).
71. Zhang, Z., Xie, J., Zhang, F. & Linhardt, R. J. Thin-layer chromatography for the analysis of glycosaminoglycan oligosaccharides. *Anal. Biochem.* **371**, 118–120 (2007).
72. Ceroni, A. et al. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659 (2008).

73. Juncker, A. S. et al. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652–1662 (2003).
74. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
75. Markowitz, V. M. et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
76. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
77. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
78. Krissinal, E. & Hanrick, K. *PDBeFold* (European Bioinformatics Institute, 2009).
79. Dereeper, A., Audic, S., Claverie, J. M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evolut. Biol.* **10**, 8 (2010).
80. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
81. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
82. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).
83. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43**, D257–D260 (2015).

Acknowledgements

We thank C. Morland (Newcastle University, UK) for his expert technical assistance, and F. Cuskin (Newcastle University, UK) and L. Royle (Ludger, UK) for insightful conversations about the data. We would like to thank Diamond Light Source (Oxfordshire, UK) for beamtime (proposal mx13587 and mx18598) and the staff of beamline I03 and 104-1 for assistance with crystal testing and data collection. We thank J. Casement (Bioinformatics Support Unit, Newcastle University, UK) for analysing the raw RNA-Seq data. We thank J. Sonnenberg (Stanford, USA) for the ΔBT0455 *Bt* strain, R. Lewis (Newcastle University, UK) for his guidance and advice in looking for structural homologues and R. Hirt (Newcastle University, UK) for his advice on phylogenetics. The work was funded by the BBSRC/Innovate UK IB catalyst award to D.N.B. 'Glycoenzymes for Bioindustries' (BB/M029018/1).

Author contributions

J.B., A.S.L. and L.I.C. carried out enzyme kinetics. J.B., P.A.U., A.S.L., O.R. and L.I.C. carried out enzymes assays. J.H. made substrates. J.B. and L.I.C. carried out *Bacteroides* growth. L.I.C. carried out the whole-cell assays. P.A.U. and O.R. carried out the LC–MS. P.A.U. and L.I.C. analysed the LC–MS data. J.B. and L.I.C. purified proteins and set up crystal trays. A.B. collected crystals and data. A.B. and N.P. solved crystal structures. J.B., D.N. and L.I.C. produced *Bacteroides* gene deletion strains. L.I.C. carried out the bioinformatic analysis. L.I.C., D.N.B., P.A.U., J.B., A.S.L., D.I.R.S. and E.C.M. designed experiments. L.I.C., E.C.L. and D.N.B wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0466-x.

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

nature research

Corresponding author(s): Dr Lucy Crouch
Dr David Bolam

Last updated by author(s): Apr 16, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	XDS (structural data collection), Chromeleon Chromatography Data System Software, HyStar v3.2 (Both for HPLC and mass spec data collection).
Data analysis	Graphpad Prism, Bruker Compass Data Analysis, GlycoWorkBench, ClustalOmega, IMG database, CAZy, Dali, PDBefold, Phylogeny.fr, Pfam, SMART, Pymol, Pointless, Aimless, autoSHARP, ShelxD, Parrot, Buccaneer, Refmac5, Coot, MrBump, Molprobity, Rfree, CCP4 suite

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings in this paper are available upon request from the corresponding authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Enzyme kinetic data are averages and SDs of technical triplicates and repeated at least once with with different enzyme preparations. Non-kinetic enzyme assays were repeated at least once for each substrate tested with different enzyme preparations. Bacterial growths are averages of triplicate cultures and were repeated at least once.
Data exclusions	No data were excluded.
Replication	Bacterial growth experiments on different glycans were carried out in triplicate and the experiments repeated at least once. The RNAseq data are the average reads from triplicate cultures. Enzyme kinetic data are from technical triplicates and repeated at least once with with different enzyme preparations. The same enzyme preparation used over the course of several days was tested daily using a known substrate to ensure the sample was not losing activity. Whole cell and cell lysate assays were repeated at least once. Westerns and SDS-PAGE gels were repeated at least once. The non-kinetic assays on recombinant enzymes were reproduced multiple times with different enzyme preparations on each of the different substrates tested. Where possible, positive and negative controls for enzyme assays were used throughout to be sure of the accuracy of data and experimental conditions.
Randomization	Randomization was not necessary for any of the experiments carried out during this work.
Blinding	Randomization was not necessary for any of the experiments carried out during this work.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Custom primary polyclonal antibodies against purified recombinant Bacteroides enzymes were raised in rabbits (Eurogentec speedy 1 month protcol). The secondary antibody used was Goat anti-rabbit IgG antibody HRP conjugate from Sigma, catalogue number AP307P.
Validation	Primary polyclonals were tested by Western against the cell lysate from E. coli containing the recombinant protein they were raised against to ensure specific recognition of the target. Specificity of the secondarys for primary antibody was validated by running Westerns against Bacteroides thetaiotaomicron lysates with secondary antibodies only to ensure no cross reactivity with Bacteroides proteins.