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

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

^These authors contributed equally
*To whom correspondence should, be addressed: david.bolam@ncl.ac.uk or lucy.crouch@ncl.ac.uk.

[1]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. [2]Ludger Ltd, Culham Science Centre, Oxfordshire, OX14 3EB, UK. [3]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA. [4]Diamond Light Source, Didcot, Oxfordshire, OX11 0DE, UK.

Supplementary Results and Discussion

Growth of *Bacteroides spp*. on different CNGs

Growth on α_1AGp as the sole carbon source was initially explored. Nine *Bacteroides* species were grown on α_1AGp and *Bt*, *B. fragilis* and *B. xylanisolvans* were found to grow well (Fig. 1). Other species did not grow to a high OD, started to grow later on or not at all. No growth of *Bt* on deglycosylated α_1AGp was observed (Fig. 1c). *Bt* also showed growth on human serum IgG, (Supplementary Fig. 1f) and characterisation of the glycans present on this substrate show that they are solely biantennary structures, similar to α_1AGp. When fetuin was used as the sole carbon source, *B*t did not grow and only *B. fragilis* and *B. xylanisolvans* grew well. As fetuin comprises predominantly triantennary CNGs (Supplementary Fig. 10), these data suggest that *Bt* has a preference for biantennary CNGs (Supplementary Fig. 11). *B. fragilis* and *B. xylanisolvans* failed to grow on human serum IgG, suggesting that they cannot access either core-fucosylated or bisecting structures or both. *B. ovatus* only grew to half the maximum cell density than *Bt* on α_1AGp, but the same OD with the IgG substrates. Observations of the variable growth phenotypes of the *Bacteroides* spp. on different CNGs highlight possible different nutrient niches within the gut.

The specificity different species have for different CNG substrates could be used to look for further enzyme activities. For instance, growth on biantennary or triantennary CNG substrates would be interesting avenues for further exploring activity of other GH18 family members encoded by the different species. Previously characterised N-glycan active GH18 family members (catalogued in Supplementary Table 3) and homologues to those upregulated in this work from other *Bacteroides* species were subject to phylogenetic analysis (Supplementary Fig. 19 and Supplementary Table 13). The clustering seen correlates well to the known activities. For example, BT3987^{GH18} that is specific to HMNG cleavage clusters with other characterised GH18 enzymes with this activity from divergent organisms rather than with the CNG-active GH18 from the same organism, BT1044^{GH18}. As another example, the CNG-active GH18 from *C. canimosus* clusters with the *Bt* GH18 enzymes highlighted in this study. The accuracy of the correlation between sequence and activity could also be used alongside growth data to predict the N-glycan specificities of other GH18 family members.

Identification of CNG upregulated CAZYme genes from the RNAseq data

The RNAseq data is provided in Supplementary Table 1. Loci that contained predicted CAZymes and were upregulated >4-fold on α_1AGp vs glucose were selected for analysis. The genes spanning BT0455-0461 are in an operon and on average are upregulated <4 fold, but several factors led to us identifying the CAZymes encoded as being involved in CNG breakdown. From the initial analysis of the spent growth media it was clear that SA was being released by *Bt* and BT0455^{GH33} is the only enzyme in this organism with sialidase activity[1]. The importance of BT0455^{GH33} in CNG breakdown was later confirmed with the deletion strain (Supplementary Fig. 1g). Furthermore, this locus also contained the previously characterised BT0458^{GH2} β-mannosidase previously shown to hydrolyse the core GlcNAc-β1,4-man linkage in N-glycans[2]. The other CAZymes in the 0455-0461 operon were also potentially involved in CNG breakdown based on their predicted activities and this prediction was confirmed by our biochemical data.

In total six discrete loci in *Bt* were identified from the RNAseq data as potentially being involved in CNG breakdown, with only two of these being classified as classical PULs (ie. containing SusC/D gene pairs as well as genes encoding predicted or known CAZymes; Fig 2). The larger of the two PULs (BT1032-BT1053) has three SusC/D pairs, five putative surface glycan binding proteins (SGBPs) and at least six genes encoding CAZymes. The CAZymes included the previously characterised GH130 (BT1033), which is a β1,4-mannosyl-N-acetyl-glucosamine phosphorylase that cleaves the Man-β1,4-GlcNAc present in the core of all N-glycans ([3] and Fig. 1a). The other CAZymes in this locus are all uncharacterised; a GH92 (BT1032), three GH18 enzyme family members (BT1038, BT1044 and BT1048), a GH with no currently assigned family (BT1035) and a GH20 (BT1051). The

second PUL is much smaller, comprising only a SusC/D pair, a predicted SGBP and a GH18 (BT4406). Of the other four non-PUL loci, one locus (BT0455-BT0461) comprises only CAZymes, including two previously characterised enzymes; BT0455 a GH33 broad-acting α-sialidase[4] and the only one present in the *Bt* genome, and BT0458, a GH2 β1,4-mannosidase[2]. The other CAZymes of this locus include three GH20 family members (BT0456, BT0459 and BT0460), a putative carbohydrate esterase (BT0457) and another GH2 (BT0461). The other non-PUL loci comprise either a single CAZyme each; a GH20 (BT0506) in one and a GH97 (BT0683) in the other, or a pair of genes encoding a putative sulfatase (BT1624) and a GH29 (BT1625) that are part of a larger PUL that has previously been shown to be expressed during growth on mucin O-glycans[5,6].

Homologous PUL structures in other *Bacteroides* species

The genes upregulated in this study were traced throughout seven other *Bacteroides* species by searching for homologues for each enzyme (see Materials and Methods). This analysis revealed that certain loci and parts of loci were well conserved and others were much more variable. The most variable loci in terms of structure, gene content, and level of identity between predicted enzymes were those loci encoding the GH18 enzymes (Supplementary Fig. 16). The equivalent PUL in *B. ovatus* retains the synteny with the BT1032-BT1053 PUL. Between the different species analysed, the loci containing GH18 family members often contain predicted GH130, GH97, GH92 and GH163 homologous to those in *Bt*, suggesting the loci are also involved in N-glycan breakdown in these species. The homologous genes to the BT0455-461 locus found in other species were quite well conserved in comparison to the GH18-containing loci in terms of structure, content and enzyme homology (Supplementary Fig. 20). For example, the sialidase BT0455 is highly conserved in all species, attesting to the importance of the removal of sialic acid for gut survival. In some species, there were also additional ORFs in the sialidase containing locus whose predicted activities are associated with sialic acid degradation. Interestingly, the four *Bt* GH20 family members identified here were usually present, even in species that showed no direct growth on CNG glycoproteins (Supplementary Fig. 20). The conservation of genes encoding CNG-specific enzymes in species that don't directly use CNG is unexpected and is currently unclear why this would be, but may be related to cross-feeding on partially degraded N-glycan structures released by primary degraders such as *Bt*.
BT1625^{GH29} and a putative sulfatase (BT1624) are two CNG upregulated genes that are part of a larger PUL, spanning BT1623-36 and previously reported to be associated with O-glycan degradation[5], suggesting these enzymes have roles in degrading both N- and O-glycans. The activity of the predicted sulphatase BT1624 was not investigated during this study due to lack of relevant substrate, but sulphated CNG structures have been reported in humans[7] and other eukaryotes[8].

Previously described association between the core CNG PUL BT1032-53 and IgA secreted in the gut

Regarding the finding that *Bt* is able to deglycoslyate IgA it is interesting to note that previous data has shown that transposon mutants in BT1032, BT1033, BT1034 and BT1035 prevent recognition of the mutant strains by IgA produced by immune cells that had been primed against wild-type *Bt*[9]. Which this finding directly links components of the core CNG PUL with IgA function, the mechanism of this phenomenon is not known and may not be IgA deglycosylation.

Localisation of the sialidase BT0455^{GH33} and fate of released sialic acid

A range of experimental techniques were used to explore the cellular location of BT0455^{GH33} sialidase. Activity of the spent media alone (cells removed at mid-exponential phase) against α_1AGp showed no further release of SA or CNG, indicating BT0455^{GH33} or other enzymes are not secreted or present on outer membrane vesicles (Supplementary Fig. 4d). Whole cell assays, which report on surface enzyme activity, show the production of free SA from α_1AGp over time indicating BT0455^{GH33} is present on the surface of *Bt* (Supplementary

Fig. 4a). A parallel experiment using the same cells showed no cell lysis during the course of the experiment, confirming activity could only be coming from surface localised enzymes (Supplementary Fig. 4c). When the cells were exposed to proteinase K prior to the whole cell assay no SA production could be seen (Supplementary Fig. 4e), intriguingly, however, when the proteinase K treated cells were lysed, sialidase activity was again observed (Supplementary Fig. 4f). Western blot analysis of proteinase K treated and untreated cells antibodies against BT0455^{GH33} revealed that the enzyme was not degraded by the protease, suggesting the GH33 is located inside the cell (Supplementary Fig. 7c). Overall, these data indicate that BT0455^{GH33} is localised both inside the cell (likely the periplasm due to the presence of a signal peptide) and on the cell surface. The mechanism by which this dual partitioning occurs is not known as the apparatus used by Bacteroidetes spp. to present lipoproteins on the cell surface has not been identified, but notably a related sialidase from *Capnocytophaga canimorsus* has a very similar signal sequence to BT0455^{GH33} and was shown to be a periplasmic facing lipoprotein[10]. This finding suggests that BT0455^{GH33} may be also be an OM lipoprotein that is simply not fully partitioned to the cell surface. The biological rationale for the dual location of BT0455^{GH33} is possibly due to the predicted periplasmic location of the associated SA esterase, BT0457 (see below for details).
SA released by BT0455^{GH33}, either on the cell surface or secreted from the periplasm, is then available to other members of the gut community as *Bt* is unable to metabolise this sugar [16]. For example, SA can be utilised by other *Bacteroides* spp. and this requires the *nanLET* loci, as shown for *B. fragilis*[11]. The effect of SA released by non-users like *Bt* on the composition of the gut community has previously been explored[12,13]. In the healthy gut the released SA cross-feeds with other mutualists, some of which cannot release SA themselves. SA is also frequently used by opportunistic pathogens and in healthy individuals they can be effectively outcompeted by the mutualists. However, when the normal bacterial community is depleted, by for example antibiotics or inflammation, the released SA can be used by organisms such as *Clostridium difficile* that are resistant to antibiotics, leading to unregulated pathogen expansion and deleterious consequences for gut health[12]. SA can also incorporated into the capsules and lipopolysaccharides of some pathogenic bacteria as a form of camouflage from the host immune system[14-17].

The sialic acid esterase BT0457^{Est}

BT0457 is homologous to the *Tanneralla forsythia* SA esterase, NanS [18]. NanS has been shown to improve access of its cognate sialidase, NanH, to sialylated glycoconjugates including CNGs, as the 7-, 8- and 9-O-acetylated forms of SA are not a substrate for NanH. Assays of BT0457^{est} against a variably acetylated SA panel confirmed the protein was an esterase like NanS (Supplementary Fig. 8) and therefore likely plays the same role in *Bt* as NanS does in *Tannerella, i.e.* BT0457^{Est} activity enables BT0455 sialidase to remove 7-, 8- and 9-acetylated SA from certain glycoconjugates. Interestingly, BT0457^{Est} has a strong SPI prediction, suggesting the protein is periplasmic, and thus that cleavage of 7-, 8- and 9Ac forms of SA from glycans likely occurs after import, providing a plausible explanation for the dual surface and periplasmic localisation of the BT0455^{GH33}.

Antennary fucose is removed by BT1625^{GH29}

Many CNG structures are fucosylated via α1,3 and 1,4 linkages to the antennary GlcNAcs as well as α1,6 fucosylation the core GlcNAc. Inspection of the upregulated CAZymes revealed a single GH29, BT1629, that has previously been shown to be active against PNP-fucose (Fig. 2)[19]. To determine the specificity of this fucosidase we assayed BT1625^{GH29} against a range variably fucosylated glycans, including those found in CNGs (Supplementary Table 7). The data revealed that BT1625^{GH29} displays specificity for α1,3 and 1,4 linked fucose substrates, but could not remove α1,2 linked fucose[19]. Incubation of BT1625^{GH29} with antennary and core fucosylated CNGs showed the enzyme was able to remove all of the fucose from the antenna only and not the core (Supplementary Fig. 11). The ability of BT1625^{GH29} to remove fucose from the galactosylated structures indicates that the fucosidase acts prior to the BT0461^{GH2} β-galactosidase in the *Bt* CNG breakdown.

CNG active GH18 family members

Potential role of the three CNG upregulated GH18 family members that appear inactive in recombinant form

It is unexpected that four GH18 genes are upregulated during growth of *Bt* on α_1AGp (Fig 2), but only one enzyme, BT1044^{GH18}, has any detectable activity in recombinant form against N-glycans. One possibility for this is that the other three GH18 enzymes display different pH optima to BT1044^{GH18} enabling *Bt* to access CNGs in distinct niches in the gut (e.g ascending vs descending colon) where pH varies - as has been observed with other microbiota CAZymes[13]. To explore this possibility, assays with all four GH18 enzymes were repeated with α_1AGp and RNaseB at pH 4.0 and pH 9.5, but no additional activity was observed to that seen at pH 7.0 (Supplementary Fig. 12b). Enzymes from the GH18 family can also often display activity against the β1,4 GlcNAc polysaccharide chitin, but none of the four enzymes were active against α- or β-forms of chitin or chito-oligosaccharides with a degree of polymerisation of 2-5 (Supplementary Fig. 9c)

Indeed, there is more extensive deglycosylation of α_1AGp by *Bt in vivo* than the recombinant BT1044^{GH18} and BT0455^{GH33} can achieve *in vitro* (Supplementary Fig. 1a&c), suggesting that one or more of the other three GH18 family members expressed are active on the cell surface.

Comparisons between the structures of BT1044^{GH18} and EndoF3 in terms of substrate specificity

The only available GH18-CNG complex structure is EndoF3 from *E. meningoseptica* which is bound to an octasaccharide product (Fig. 4e). The complex shos that direct contacts between the glycan and protein only occurred for the core GlcNAc at +1 and two of the core mannoses, but not the antenna sugars. Comparison of EndoF3 and BT1044^{GH18} revealed that all the amino acids involved in glycan binding in EndoF3 are conserved in BT1044^{GH18} (Supplementary Fig. 12c). Point mutations to alanine of all of these predicted substrate binding residues in BT1044^{GH18} resulted in mutant enzymes that were inactive against α_1AGp, supporting their proposed roles in substrate recognition in the *Bt* enzyme (Supplementary Fig. 12d).

EndoF3 is active on both bi- and tri-antennary CNGs, but cannot accommodate bisecting GlcNAc structures and also requires core fucosylation for activity (Supplementary Table 3), whereas BT1044^{GH18} removes bi-antennary structures, can accommodate bisecting GlcNAcs and does not require core fucose for activity. In the negative subsite region, Endo F3 has a distinct pocket comprising aromatic residues that could make specific interactions with a core α1,6-fucose, suggesting that recognition of this sugar plays a key role in providing enough binding energy for catalysis (Fig. 4e). In comparison, BT1044^{GH18} has a more open channel in the equivalent position, providing an explanation for why it can accommodate core fucose, but it is not a requirement for substrate recognition (Fig. 4e). For EndoF3, at the point where a bisecting β1,4-GlcNAc would be, there is a methionine pointing into this space from loop 3 in EndoF3, whereas the equivalent position in BT1044^{GH18} is much more open with a less extensive loop 3 (Fig. 4e & Supplementary Fig. 21).

Substrate binding surface comparisons between CNG and HMNG active GH18 family members

The available crystal structures of GH18 family members with N-glycan activities are presented for comparison in Supplementary Fig. 21. It is not currently clear from these how characteristics of the protein surfaces around the active sites dictate specificity, such as CNG vs. HMNG or bi- vs. tri-antennary CNG. For instance, it is not clear from a comparison of the EndoF3 and BT1044^{GH18} structures why the former can act on full tri-antennary CNG and the latter is more limited (Fig. 4). A lack of direct contacts between the EndoF3 and the antenna sugars led to the proposal that this enzyme operates by excluding certain N-glycans, where particular antenna structures would not be accommodated rather than a specificity for specific CNG structures[20].

We examined if the characteristics of the binding surfaces of available GH18 enzyme structures could be correlated with functions (Supplementary Fig.21). A clear difference between the different structures involved a β-hairpin structure (loop 3; orange) in those GH18 enzymes with activity towards HMNG. For the three enzymes specific for CNG, loop 3 is not an extended β-sheet, but a much more compact loop occupying less space with only EndoF3 showing a more ordered structure with two short α-helices[21]. The extended β-hairpin loop in HMNG-active enzymes provides a much larger surface to potentially act as a binding platform for the α-mannose extensions of these N-glycans. Evidence for this comes from the observation that EndoF1 and EndoH (both acting on HMNG) require the α1,3-mannose extension on the core α1,6-mannose for catalysis, which would involve the region with the β-hairpin structure[22]. In comparison, those enzymes missing this β-hairpin loop might exclude HMNG activity through steric clashes with this N-glycan branch or lack of necessary contacts. This β-hairpin pushes against loop 2 (green) to form this site, whereas, in the structures where it is missing (CNG-active enzymes) loop 2 tends to be longer with α-helical structures. Notably, this β-hairpin structure is not present in the GH18 with HMNG specificity from the fungus *Hypocrea jecorina*, but loops 2 and 3 still occupy approximately the same space[23].

GH20 substrate specificity and structure

GH20 specificities and activity on disaccharides and chito-oligosaccharides

Four GH20 enzymes in two discrete loci were expressed during growth of *Bt* on CNG (Fig. 2). All of these enzymes possess the HxGG(DE) catalytic-motif characteristic of this family with the exception of BT0456^{GH20} where the second glycine is replaced by a threonine. All recombinant GH20 enzymes displayed activity against PNP-GlcNAc, but BT0459^{GH20} had the fastest kinetics and was also the only enzyme to have a measurable rate on N-glycan derived disaccharide GlcNAc-β1,2-Man (Supplementary Table 5). BT0459^{GH20} also had comparable kinetics against GlcNAc-β1,3-Man, although this is not found in N-glycan structures, and also GlcNAc-β1,3-Gal (LNB), which is a common O-glycan linkage. This highlights the broad specificity displayed by BT0459^{GH20} for both the linkage and the +1 sugar and suggests the enzyme may have roles in degrading other classes of glycan as proposed for some of the other CNG active enzymes. In contrast, the other three GH20 enzymes had much slower rates against these disaccharides than BT0459^{GH20} (Supplementary Fig. 22a) indicating they may require more specific positive subsites interactions for activity.

Activity against PNP-GalNAc was also seen for most of the enzymes; dual GlcNAc and GalNAc activity is common in GH20 family (Supplementary Table 5). Notably, β-linked GalNAc is known to occur on the antenna of certain CNGs (either as well as, or instead of, GlcNAC).

Chitooligosaccharides (β1,4 linked GlcNAc) between chitobiose to chitopentaose were also tested and BT0459^{GH20} was able to rapidly degrade all of these down to GlcNAc, supporting the broad specificity of this enzyme (Supplementary Fig. 22b). BT0456^{GH20} and BT0460^{GH20} produced GlcNAc and chitobiose for all chitooligosaccharides tested. These enzymes were, however, slow compared to BT0459GH20, indicating that both enzymes require recognition of a +2 sugar for significant activity and that chito-oligosaccharides are not their preferred substrates.

There was some degradation of chitobiose with BT0456^{GH20} and BT0460^{GH20}, but the rate was extremely slow. This is an interesting observation considering both BT0456^{GH20} and BT0460^{GH20} are active against the bisecting β1,4 GlcNAc and implies that these enzymes either have specificity for the core mannose in CNGs at the +1 position and/or require interactions with the antennary mannose(s) and/or the core GlcNAcs for full activity. Finally, BT0506^{GH20}, which appears to be specific against antennary GlcNAc on CNG structures, had very poor activity on all defined oligosaccharides tested here indicating this enzyme requires recognition of specific features of the antennary GlcNAc presentation for activity (Supplementary Table 5 & Supplementary Fig. 22).

Comparison of activities of GH20 family members described in this study with previously published CNG-active GH20 enzymes

Some of the preferences seen for the four GH20 enzymes described in this study in the main discussion are similar to what has been observed previously for other members of the same family. For example, *Sp*GH20A from *Streptococcus pneumoniae* can remove both β1,2-linked GlcNAcs from the α1,3 or α1,6 arms of CNGs, however, it cannot cleave bisecting GlcNAc or the α1,3-arm GlcNAc when Gal and SA is present on the α1,6 arm [26]. This is similar to the activity seen with BT0459^{GH20} and BT0506^{GH20}. In contrast, *Sp*GH20B also from *Streptococcus pneumoniae* can only cleave the GlcNAc from the α1,3 arm, but can hydrolyse bisecting structures[24], very similar to BT0456^{GH20}.

Structures of these two *S. pneumoniae* GH20 enzymes either in complex with CNG substrates or with CNGs modelled in from a previous study, revealed the likely structural basis for specificity[24]. The positive subsites of *Sp*GH20A and B have an aromatic "clamp" and "ramp", respectively, providing a surface for certain CNG structures to stack against, but sterically hindering others (Fig. 6c and d). In the case of the *Sp*GH20B aromatic "ramp", a CNG structure with an antenna α1,6-arm GlcNAc could not be modelled in without clashes (reflecting its activity against the α1,3 arm only), whereas the *Sp*GH20A aromatic "clamp" allowed binding of either antennary structure, but precluded recognition of a bisecting GlcNAc (reflecting its broader activity against the antenna compared to *Sp*GH20B, but lack of activity against bisecting GlcNAc). As discussed below, overlay of the CNG glycan product provided molecular insight into why BT0459^{GH20} had difficulty cleaving GlcNAc from some antenna when a bisecting GlcNAc was present.

These different specificities have also been observed in GH20 enzymes from eukaryotes. For example, two characterised GH20 family members from *Arabidopsis thaliana* have the same activity as *Sp*GH20A in being able to remove GlcNAc from both arms of CNGs[25], whereas the enzymes from *Caenorhabditis elegans*[25] and *Drosophila melanogaster*[26] have activities only on the α1,3 arms as described for *Sp*GH20B.

Structural comparison of the active site of BT0459^{GH20} to other GH20 family members

The active site of BT0459^{GH20} was compared to the two *S. pneumoniae* GH20 structures described above and also a GH20 enzyme from *O. furnacalis* active on chito-oligosaccharides (Fig. 6). The loops surrounding the active sites in the *S. pneumoniae* and *O. furnacalis* structures are much more extensive than those found in BT0459^{GH20} with no obvious potential interactions with the +1 sugar. This is supported by an overlay of the GlcNAc-β1,2-Man disaccharide from the *Sp*GH20A structure with BT0459^{GH20} showing no possible interactions with the mannose in the +1 subsite. The open active site of BT0459^{GH20} provides a ready explanation for broad specificity of this enzyme. One aromatic residue (Y433) in BT0459^{GH20} may explain the difficulty this enzyme has in removing an antenna structure in the presence of a bisecting β1,4-GlcNAc. When the bisecting substrate structure from *Sp*GH20B was overlaid into the BT0459^{GH20} active site, Y433 clashed with the bisecting GlcNAc (Fig. 6b). The openness of the BT0459^{GH20} active site is further explored below considering the possible role of the other modules in the enzymes and in comparison to a previously described GH20 from *Serratia marcescens* (1C7S) with similar ancillary modules (Fig. 6g and Supplementary Fig. 16)[27].

Putative role of the F5/F8 Type C domain of BT0459^{GH20}

The C-terminal F5/F8 Type C domain of BT0459^{GH20} is a jelly roll β-sandwich and was found to be structurally most similar to members of CBM family 32 (Supplementary Fig. 16). The β-sheet core of the CBM32s and the BT0459$^{F5/F8TypeC}$ were very similar whereas the loops extending out of the β-sheets forming the ligand binding site of CBM32s are the most variable (Fig. 6h and Supplementary Fig. 16). BT0459$^{F5/F8TypeC}$ had relatively extensive loop structures in this region with numerous solvent exposed aromatic residues that could potentially be involved in glycan binding (Fig. 6h).

Interestingly, there was one structural match with BT0459$^{F5/F8TypeC}$ to the N-terminal domain from a GH20 chitobiase from *Serratia marcescens* (1C7S). In the *Serratia* structure, the N-

terminal CBM-like domain sits on top of the catalytic subunit over the active site (Fig. 6g). The part of the jelly roll module facing the active site overlays with the ligand binding region on known CBMs (Supplementary Fig. 16). It is tempting to speculate that the N-term *Serratia* module and the equivalent C-term BT0459 module are CBMs and possibly provide subsites for their respective substrates. Comparison of these two multi-modular protein structures suggest that they could be snapshots of different conformations that are common to both enzymes in solution. In the case of BT0459^{GH20}, the C-term CBM like domain may provide more specificity to the relatively non-specific open active site described above. Notably, the other Bt GH20 enzymes described in this study also have similar modular structures to BT0459^{GH20}, albeit with a different family of putative CBM-like domain at the C-term in some cases (e.g. PA14 in BT0506; Supplementary Fig. 2).

Cellular localisation of BT1035^{GH163}

To understand where BT1035 acts in the *Bt* CNG degradation pathway, analysis of spent growth media and whole cell assays were performed using both wild type and a *BT1035* deletion strain (ΔBT1035). Spent media from growth of *Bt* on α_1AGp (Fig. 1e and Supplementary Fig. 3) and assays of cell surface enzyme activity against CNG (Supplementary Fig. 4-6) were analysed. The data revealed glycans that appeared to be products of BT1035^{GH163} activity, specifically the sialylated trisaccharide (sialyl-LacNAc; predominantly Neu5Gc). The build-up of the Neu5Gc form of the trisaccharide suggests this glycan is a poor substrate for the surface sialidase. The ΔBT1035 mutant strain displayed no defect during growth on α_1AGp, but the trisaccharide was no longer produced, either in the spent media or the cell surface assays, confirming that this glycan is a product of BT1035^{GH163} activity and that the enzyme is localised to the cell surface (Supplementary Fig. 4-6 and Supplementary Fig. 7b).

Activity of BT1035^{GH163} against defined oligosaccharides

To further explore the subsite specificity of BT1035^{GH163}, the enzyme was tested against a range of β-GlcNAc containing di- and oligo-saccharides (Supplementary Fig. 17). BT1035^{GH163} displayed no activity against these any of these sugars, including lacto-N-tetraose and lacto-N-neotetraose, suggesting either the enzyme cannot cleave the β1,3 GlcNAc linkage or cannot accommodate Gal at the +1 subsite, or both, supporting the proposed role of BT1035^{GH163} in primarily CNG breakdown.

Processing of the core tetrasaccharide in *Bt*

Following removal of the antennae glycans, the core tetrasaccharide structure common to all N-glycans is likely degraded in the periplasm by a number of previously characterised exo-acting enzymes that are upregulated on α_1AGp. These include GH92 α-mannosidases and the BT0458^{GH2} β1,4-mannosidase[2,28]. Notably, the core Man-β1,4-GlcNAc disaccharide can also be cleaved by the cytoplasmic β1,4-mannosyl-GlcNAc phosphorylase (BT1033^{GH130}) that is part of the large CNG PUL (Fig. 2), suggesting at least some of this disaccharide is imported into the cytoplasm for degradation[3].

Potential role of a GH97 α-glucosidase in N-glycan degradation.

Interestingly, there is an orphan gene BT0683 that is upregulated during growth of *Bt* on CNG and belongs GH97, a family that to date comprises only either α-glucosidases or α-galactosidases. Phylogenetic analysis placed BT0683^{GH97} within the GH97a subfamily and the enzyme displayed activity against PNP-α-Glucose[29].

The expression of this enzyme is puzzling as there is no α-glucose in mature N-glycans (or O-glycans) that we know of. However, during the synthesis of N-glycosylated proteins in the endoplasmic reticulum, Glc3Man9GlcNAc2 (where the capping Glc is α1,2 and 1,3 linked) is the initial glycan transferred to the protein as it is synthesised[30]. The N-glycans are then trimmed and built back up to differentiate them to their mature high-mannose, hybrid or complex structures. The gut microbiota would be exposed to the contents of the whole cells from dietary meat and plants, as well as sloughed host epithelial cells, suggesting non-

mature α-glucose containing N-glycans could be targets for BT0683^{GH97}. Homologues of the BT0683^{GH97} are present in six out of eight of the other *Bacteroides* species from the *in silico* analysis (Supplementary Fig. 20) and the BT0683 homologue in *B. fragilis* and *B. vulgatus* are part of loci which include GH18 and GH130 family members and in the latter species also a homologue to BT1035^{GH163}, supporting a role for the GH97 α-glucosidases in N- and/or O-glycan degradation. Further studies are required to explore this possibility. BT0683 was also shown to be upregulated in a previous study when the BT1053 anti sigma regulator was deleted, further supporting that BT0683^{GH97} is involved in degradation of substrates that activate the BT1053 controlled apparatus such as CNGs (see discussion above).

Supplementary Table 1 | RNAseq data table showing average normalised base counts in Glucose and bovine α_1AGP samples. The information is sorted according to the fold change. Base mean= mean expression level across three replicates. Full data can be found in a separate Excel Spreadsheet and submitted to https://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE129572.

Supplementary Table 2 | A list of the predicted and previously characterised CAZymes upregulated in *Bt* when grown on α_1AGp. The enzymes in bold have been characterised in previous studies, predicted localisation is listed (determined using LipoP) and experimentally determined localisation is listed where carried out.

Locus Tag	Cazy Classification	Predicted or characterised function	Predicted localisation	Determined in this study
BT0455[4]	**Sialidase**	**Broad acting exo sialylidase**	**SPI**	**Extra & intra**
BT0456	GH20	β-hexosaminidase	SPI	Intra
BT0457	Esterase	Sialic acid esterase	SPI	Intra
BT0458[2]	**GH2**	**β-mannosidase**	**SPI**	
BT0459	GH20	β-hexosaminidase	SPII	Intra
BT0460	GH20	β-hexosaminidase	SPI	Intra
BT0461	GH2	β-galactosidase	SPI	Intra
BT0506	GH20	β-hexosaminidase	SPII	Intra
BT0683	GH97	α-glucosidase	SPII	
BT1032	GH92	α-mannosidase	SPI	
BT1033[3]	**GH130**	**β-mannosyl-GlcNAc phosphorylase**	**Cytoplasmic**	
BT1035	Putative GH	Endo-β-GlcNAcase	SPI	Extra
BT1038	GH18	Endo-β-hexosaminidase	SPII	
BT1044	GH18	Endo-β-hexosaminidase	SPII	Extra
BT1048	GH18	Endo-β-hexosaminidase	SPII	
BT1051	GH20	β-hexosaminidase	SPII	
BT1624	Sulfatase	Sulfatase	SPI	
BT1625[19]	**GH29**	**α-fucosidase**	**SPII**	
BT4406	GH18	Endo-β-hexosaminidase	Cytoplasmic	

Supplementary Table 3 | Specificity of the characterised GH18 endo-β-N-acetylglucosaminidases for different N-glycans.
Information available from the published literature on N-glycan active enzymes from the GH18 family are summarised.

Organism	Locus Tag	Pathogen/commensal and environment	Direct evidence for growth on N-glycan	Activity	Substrates tested	Modules	Structure	Reference
Streptococcus plicatus	EndoH	Human pathogen.	No	HMNG Some hybrid	Carboxypeptidase Y, Mung bean nuclease I and Penicillium nuclease I	GH18	1C3F 1EDT	31,32
Elizabethkingia meningoseptica	EndoF1 (EK1)	Wide variety of habitats – soil, water and animals. Opportunistic pathogen causing meningitis and can be multidrug resistance.	No	HMNG Hybrid (core fucosylation reduces activity)		GH18	1EBN	20,22,33
	EndoF2 (EK2)		No	CNG (Bi-antennary) HMNG		GH18		
	EndoF3 (EK3)		No	CNG (Bi- and tri-antennary, α1,6-fucose specificity)		GH18	1EOK 1EOM (ligand)	
Steptococcus oralis	Not identified	Commensal human oral-pharyngeal tract, but potential pathogen in immunocompromised patients.	Yes	Exo sugar trimming of CNGs	α1-acid glycoprotein	Unknown		34
Enterococcus faecalis V583	EndoE	Opportunistic pathogen - clinical vancomycin resistant isolate.	Yes	HMNG CNG (Bi-antennatry)	RNaseB IgG Human lactoferrin	GH18-GH20		35-38
Streptococcus pyogenes	EndoS	Human pathogen.	Yes	CNG (Bi-antennary)	IgG	GH18-Unknown	4NUY 4NUZ	39,40
Flavobacterium sp.	Endo-Fsp	Soil isolate.	No	HMNG Hybrid	Ovalbumin	GH18		41
Hypocrea jecorina (Trichoderma reesei)	EndoT	Mesophilic soft-rot fungus.	No	HMNG	RNaseB	GH18	4AC1	23,42
Flammulina velutipes	EndoFV	Fungus occurring on rotting wood.	No	HMNG	Ovalbumin	GH18		43
Capnocytophaga canimorsus	Ccan_08720	Mammalian oral cavity commensal capnophile. Human pathogen.	Yes	CNG (Bi- and tri-antennary)	IgG and fetuin	GH18		10
Enterococcus faecalis V583	*Ef*Endo18A EF2863	Opportunistic pathogen - clinical vancomycin resistant isolate.	No	HMNG	RNaseB Ovalbumin	GH18		44
Bifidobacterium longum subsp. Infantis	Blon_2468 (EndoBI-1)	Commensal species important in human newborns.	Yes on yeast α-mannan Less well on th e lactoferrin substrates	HMNG CNG	RNaseB, bovine and human lactoferrin	GH18		45
Bifidobacterium longum subsp. Infantis	BLIF_1310 (EndoBI-2)	Commensal species important in human newborns.		HMNG CNG	RNaseB, bovine and human lactoferrin	GH18-Unknown		45
Streptococcus pyogenes	$EndoS_2$	Human pathogen.	No	CNG (Bi-antennary, tolerates sialic acid)	IgG and α1 acid glycoprotein	GH18-Unknown		46
Bacteroides fragilis	BF638R_3441 (DonE)	Usually commensal, but opportunistic human pathogen.	Yes	CNG (Bi- and tri-antennary)	Transferrin from rat serous fluid. Human transferrin. Not IgG or IgA.	GH18		47
Trichoderma atroviride	Eng18A Eng18B	Present in soil and is an opportunistic avirulent plant symbiont.	No	HMNG	RNaseB	GH18 GH18		48
Bacteroides thetaiotaomicron	BT3987	Commensal - human large intestine.	Yes	HMNG	RNaseB	DUF1735-GH18	3POH	28
Xanthomonase campestris	NixF	Plant pathogen.	No	Predicted plant N-glycan		GH18		49
Bacteroide thetaiotaomicron	BT1044	Commensal - human large intestine.	Yes	CNG (Bi-antennary, tolerates sialic acid and core and antennae fucose)	bovineα_1AGp, human serum IgG and IgA, human colostrum IgA.	GH18	6Q64	This study

Supplementary Table 4 | Activity of BT0461^{GH2} β-galactosidase against different oligosaccharides Kinetics was carried out using either monitoring PNP release or galactose release using a linked assay detection kit. The results are representative of at least three technical replicates and at least one biological replicate. The values are the mean and standard deviation of the data obtained.

Substrate	kcat (min^{-1})	Km (mM)	kcat/Km (mM^{-1}/ min^{-1})
PNP-β-Gal	3118 (± 90.7)	0.87 (± 0.13)	3584
Gal-β1,4-GlcNAc (LacNAc)	2109 (± 48.8)	1.77 (± 0.18)	1191
Gal-β1,3-GlcNAc (LNB)	NA	NA	NA
Gal-β1,4-Glc (Lactose)	138 (± 5.7)	2.6 (± 0.15)	54
Gal-β1,4-Gal	409 (± 18.9)	0.8 (± 0.08)	506

NA = no activity detectable

Supplementary Table 5 | Activity of upregulated GH20 family members against different oligosaccharides Kinetics was carried out using either monitoring PNP release or galactose or mannose release using a linked assay detection kit. The results are representative of at least three technical replicates and at least one biological replicate. The values are the mean and standard deviation of the data obtained.

Substrate	Kinetic parameter	BT0456	BT0459	BT0460	BT0506
PNP-β-GlcNAc	**kcat** (min^{-1})	382 (± 9.5)	20200 (± 300)	8690 (± 357)	3990 (± 78.1)
	Km (mM)	0.172 (± 0.01)	0.48 (± 0.03)	1.58 (± 0.23)	0.47 (± 0.05)
	kcat/Km (mM^{-1}/ min^{-1})	2180 (± 217)	42100 (± 3250)	5500 (± 1030)	8490 (± 1070)
PNP-β-GalNAc	**kcat** (min^{-1})	105 (± 7.6)	2870 (± 63.6)	1220 (± 39.3)	378 (± 14.0)
	Km (mM)	1.15 (± 0.17)	0.06 (± 0.003)	0.72 (± 0.07)	0.7 (± 0.03)
	kcat/Km (mM^{-1}/ min^{-1})	92 (± 20)	47800 (± 3450)	1690 (± 219)	7560
GlcNAc-β1,2-Man	**kcat** (min^{-1})	TLTQ	888 (± 20.6)	-	-
	Km (mM)		0.78 (± 0.05)	-	-
	kcat/Km (mM^{-1}/ min^{-1})		1140 (± 104)	103 (± 3.26)	2.5 (± 0.06)
GlcNAc-β1,3-Man	**kcat** (min^{-1})	TLTQ	1520 (± 95.6)	-	TLTQ
	Km (mM)		0.72 (±0.13)	-	
	kcat/Km (mM^{-1}/ min^{-1})		21120 (± 506)	60 (± 0.21)	
GlcNAc-β1,3-Gal	**kcat** (min^{-1})	TLTQ	597 (± 48.5)	TLTQ	TLTQ
	Km (mM)		1.76 (± 0.29)		
	kcat/Km (mM^{-1}/ min^{-1})		339 (± 83.5)		
GalNAc-β1,3-Gal	**kcat** (min^{-1})	TLTQ	Active	TLTQ	TLTQ
	Km (mM)				
	kcat/Km (mM^{-1}/ min^{-1})				

TLTQ = activity too low to quantify
NT = not tested
o/n = overnight (~16h)
$Vmax=kcat \times [E]_{conc}$

Supplementary Table 6 | Activity of *Bt* GH20 family members against specific CNG structures This summarises the enzyme activities of the GH20 family members analysed in this study. The information is derived from Figure 5 and Supplementary Figures 11 and 13.

	BT0456	BT0459	BT0460	BT0506
Antenna β1,2 GlcNAc (either with or without bisecting GlcNAc)	Yes, but preference for one arm	Yes, but bisecting inhibits	Yes, but preference for one arm	Yes
Antenna β1,2 GlcNAc removal when the other arm has a Gal (Supplementary Fig 13, glycan 4&6 to 22&23)	No	Yes, but trace remaining	Yes	Yes
Antenna β1,2 GlcNAc removal when the other arm has a Gal and SA (Supplementary Fig 13, glycan 11 to 28)	No	No	Yes	Yes
Antenna β1,4 GlcNAc (triantennary) removal (Supplementary Fig. 11, glycan 10 and 11)	Yes	Yes	Yes	Yes
Bisecting GlcNAc in presence of uncapped antennary GlcNAc (Supplementary Fig. 11; bisecting GlcNAc structures shown with asterisk inside sugar symbol)	Yes, but leaves trace in all samples	Trace activity	Yes	No
Bisecting GlcNAc removal when one arm has Gal (Supplementary Fig 13, glycan 7 digestion to 6, 23 or 24)	No (some 6 and 7 remaining; no 23 or 24)	Not all (some 7 still remaining and 23 product)	Yes (23 is product of antennary and bisecting digestion)	No (24 is the product of antennary digestion)
Bisecting GlcNAc removal when both arms have Gal (Supplementary Fig 13, glycan 10 digestion to 9)	No	No	Yes	No
Bisecting GlcNAc removal when both arms have Gal and one had SA (Supplementary Fig 13, glycan 13&15 to 12&14)	No	No	Yes	No

Supplementary Table 7 | Kinetics of BT1625^{GH29} against fucosylated oligosaccharides Kinetics was carried out using either monitoring PNP release or fucose release using a linked assay detection kit. The results are representative of at least three technical replicates and at least one biological replicate. The values are the mean and standard deviation of the data obtained.

Substrate	$kcat/K_M$ (min^{-1} M^{-1})
Blood group A	Not active
Blood group B	Not active
Blood group B type 1	Not active
Blood group B type 2	Not active
2'-*O*-fucosyllactose	Not active
3-*O*-fucosyllactose	4.7×10^5 ($\pm 4.6 \times 10^4$)
Lewis a	7.5×10^5 ($\pm 3.8 \times 10^4$)
Lewis b	3.5×10^4 ($\pm 1.9 \times 10^3$)
Lewis x	1.6×10^6 ($\pm 4.0 \times 10^4$)
Lewis y	8.0×10^5 ($\pm 8.5 \times 10^4$)

Supplementary Table 8 | Data collection and refinement statistics

Data collection	BT1044 (SeMet) peak	BT1044 (SeMet) inflection	BT1044 (SeMet) high energy remote	BT1044 (SeMet) E190Q	BT0459
Date	05/02/18	15/04/18	12/03/17		
Source	I03	I04-1	I03		
Wavelength (Å)	0.9794	0.9796	0.9794	0.9163	0.9762
Space group	$P3_221$	$P3_221$	I222		
Cell dimensions					
a, b, c (Å)	120.9, 120.9, 83.2	119.6, 119.6, 83.1	175.8, 186.9, 242.3		
α, β, γ (°)	90, 90, 120	90 ,90, 120	90, 90, 90		
No. of measured reflections	396796 (53297)	396260 (53067)	396277 (52925)	302290 (30560)	1088213 (54227)
No. of independent reflections	19562 (2569)	19546 (2569)	19546 (2567)	27124 (2797)	147443 (7253)
Resolution (Å)	48.91 – 2.70 (2.83 – 2.70)	48.90 – 2.70 (2.83 – 2.70)	48.90 – 2.70 (2.83 – 2.70)	48.50 – 2.40 (2.49 – 2.40)	49.75 – 2.44 (2.48 – 2.44)
$CC_{1/2}$	0.993 (0.603)	0.994 (0.582)	0.994 (0.622)	0.999 (0.835)	0.998 (0.551)
$I/\sigma I$	7.0 (0.7)	7.0 (0.7)	7.3 (0.8)	12.6 (1.3)	12.2 (1.6)
Completeness (%)	99.7 (60.3)	99.7 (99.6)	99.7 (99.6)	99.8 (99.5)	100.0 (100.0)
Redundancy	20.3 (20.7)	20.3 (20.7)	20.3 (20.6)	11.1 (10.9)	7.4 (7.5)
Anomalous completeness (%)	99.9 (99.8)	99.9 (99.8)	99.9 (99.8)		
Anomalous multiplicity	10.6 (10.6)	10.6 (10.6)	10.6 (10.6)		
Refinement					
R_{work}/ R_{free}				19.7 / 24.5	21.1 / 23.5
No. atoms					
Protein				2638	17880
Ligand/Ions					6
Water					141
B-factors					
Protein				89.9	63.9
Ligand/Ions					45.5
Water					44.0
R.m.s deviations					
Bond lengths (Å)				0.010	0.006
Bond angles (°)				1.94	1.28
Ramachandran plot (%) Favoured/Outliers				89.8 / 0.6	96.5 / 0.3
PDB				6Q64	6Q63

Values in parenthesis are for the highest resolution shell. Rfree was calculated using a set (5%) of randomly selected reflections that were excluded from refinement.

Supplementary Table 9 | Percentage identity between all members of the GH2 family included in this study

	BT0458	BT0461
BT0458	100 %	19.5 %
BT0461		100 %

Supplementary Table 10 | Percentage identity between all members of the GH18 family included in this study

	BT1038	BT1044	BT1048	BT4406
BT1038	100 %	38.1 %	35.7 %	30.7 %
BT1044		100 %	38.0 %	32.1 %
BT1048			100 %	36.9 %
BT4406				100 %

Supplementary Table 11 | Percentage identity between all members of the GH20 family included in this study

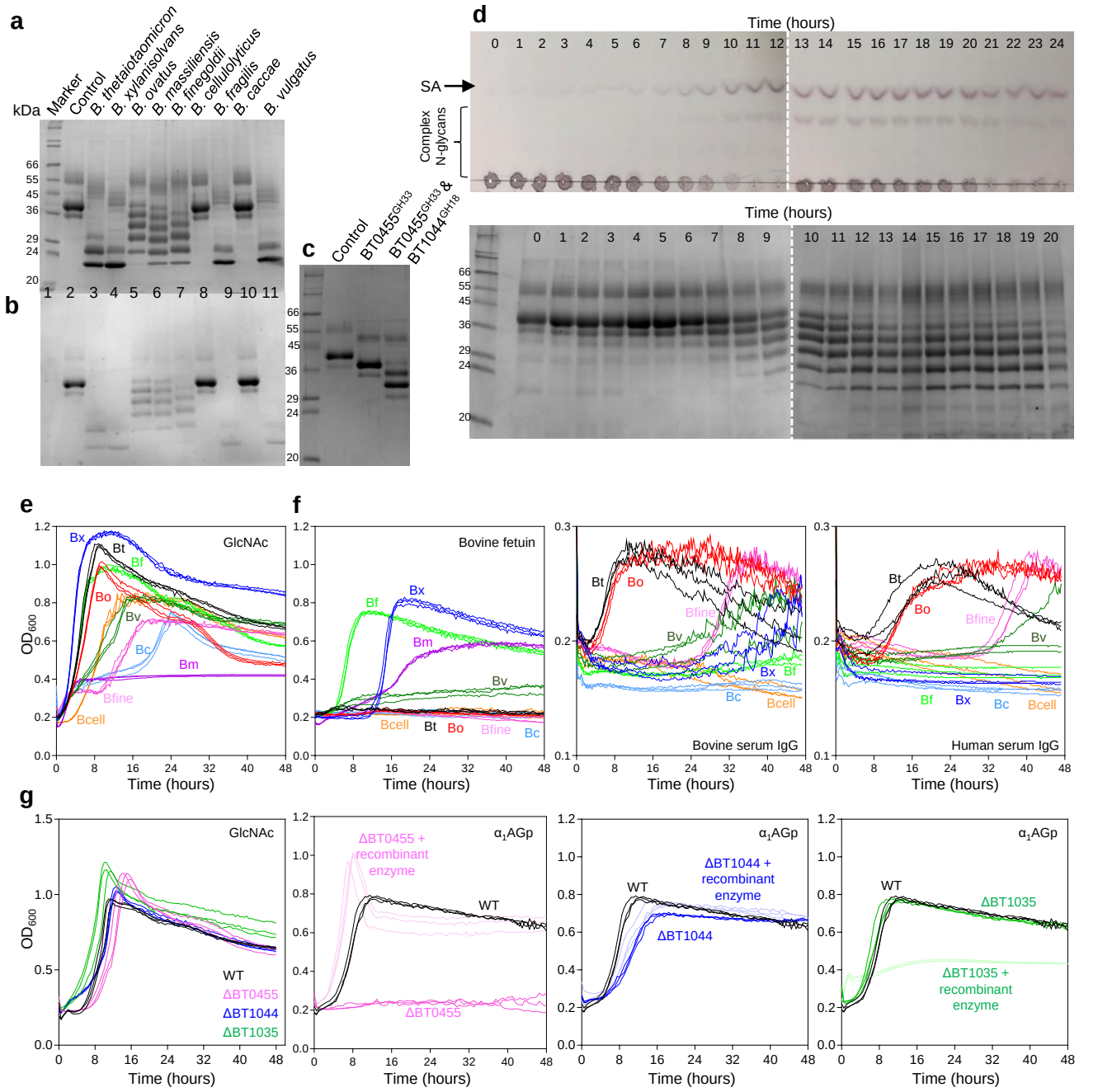
	BT0456	BT0459	BT0460	BT0506	BT1051
BT0456	100 %	24.0 %	24.6 %	17.1 %	24.0 %
BT0459		100 %	41.1 %	19.6 %	61.1 %
BT0460			100 %	19.6 %	39.6 %
BT0506				100 %	21.6 %
BT1051					100 %

Supplementary Table 12 | Percentage identity between all members of the putative surface binding proteins included in this study

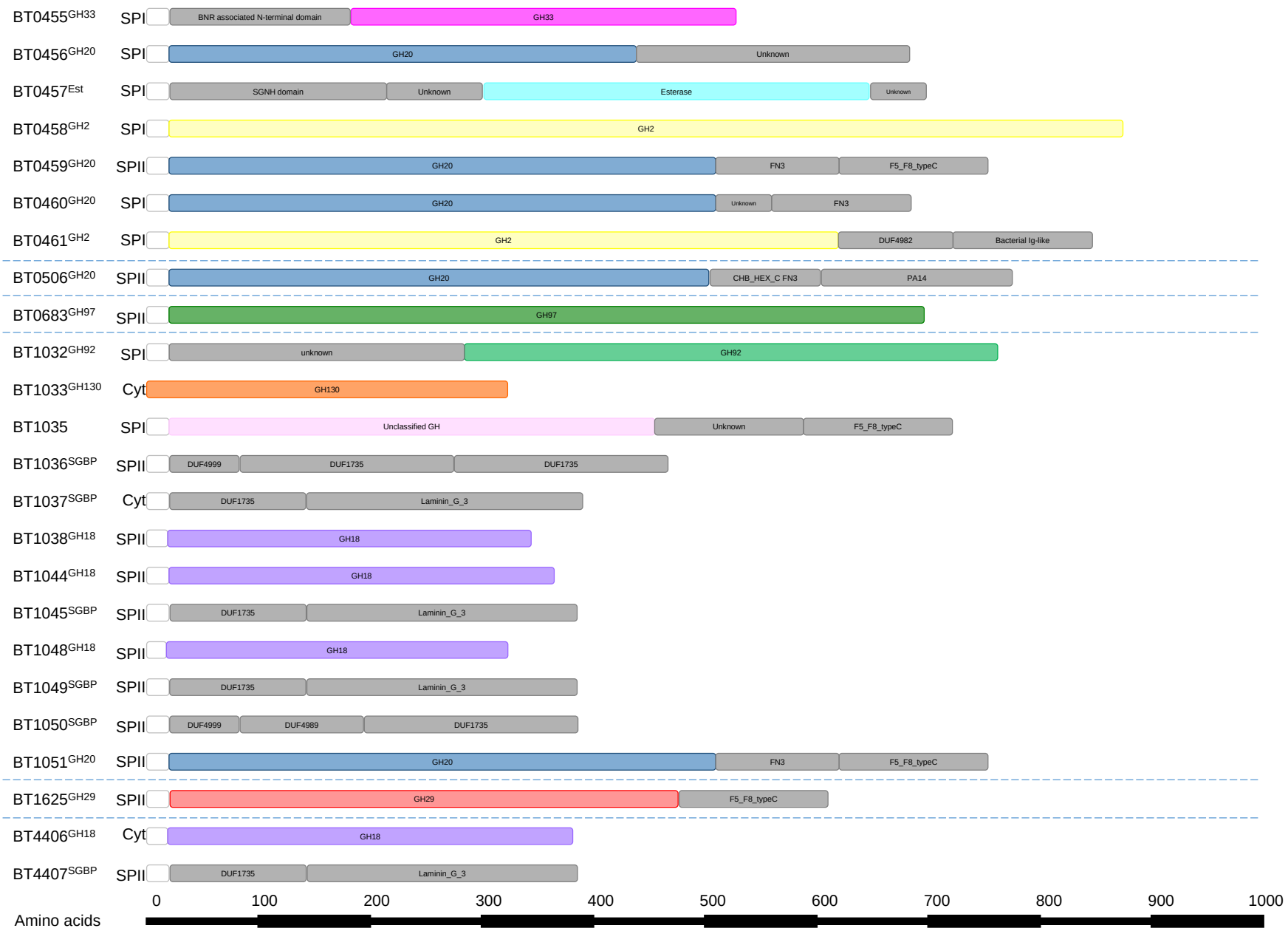
	BT1036	BT1037	BT1045	BT1049	BT1050	BT4407
BT1036	100 %	20.0 %	17.7 %	19.4%	19.0 %	20.4 %
BT1037		100 %	26.4 %	26.6 %	14.5 %	23.8 %
BT1045			100 %	22.1 %	16.5 %	22.1 %
BT1049				100 %	12.0 %	26.4 %
BT1050					100 %	14.6 %
BT4407						100 %

Supplementary Table 13 | List of strains used in this study, the locus tag prefixes and also shortened versions used in Supplementary Figure 20

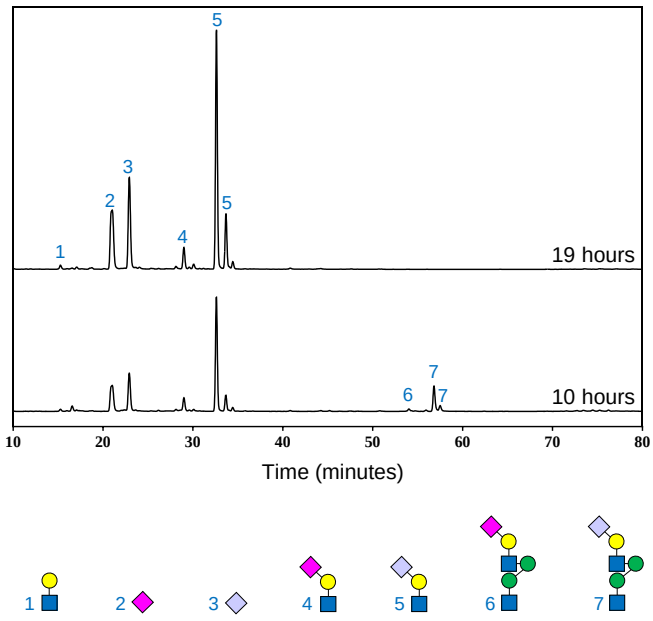
Species	Strain analysed	True locus tag prefix	Prefix used for brevity in Supplementary Figure 20
B. caccae	ATCC43185	BACCAC	BC
B. cellulosilyticus	DSM 14838	BACCELL	BAC
B. fragilis	NCTC 9343	BF	BF
B. massiliensis	DSM 17679	B035DRAFT	BM
B. finegoldii	DSM 17565	BACFIN	BACFIN
B. vulatus	ATCC 8483	BVU	BVU
B. thetaiotaomicron	VPI-5482	BT	BT
B. ovatus	ATCC 8482	BACOVA	BO
B. xylanisolvans	XB1A	BXY	BXY



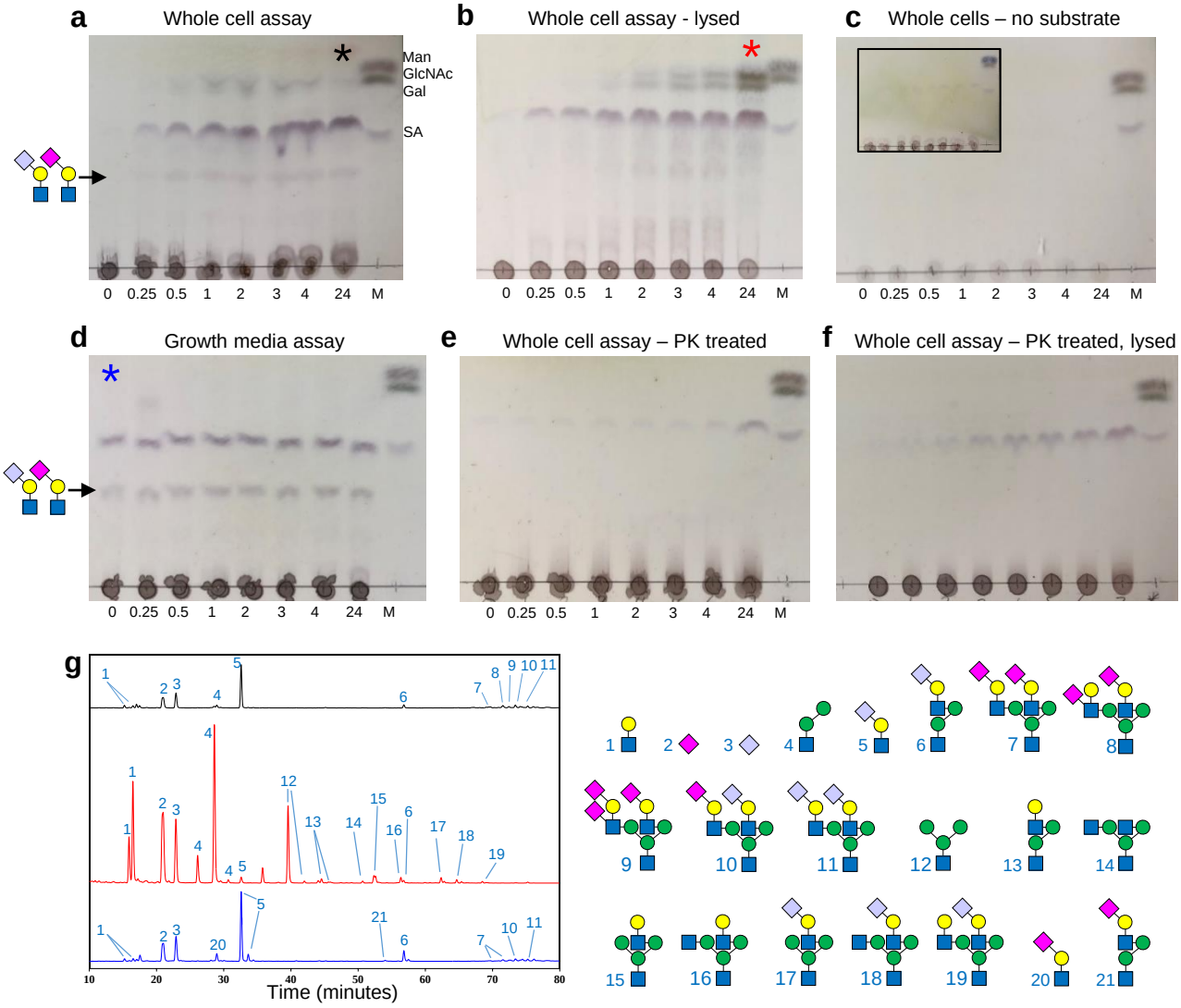
Supplementary Figure 1 | Growth of *Bacteroides spp.* on CNG Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE gels) of supernatants from the 48 hour growths of different *Bacteroides* on α_1AGp shown in Fig. 1(b) were stained for protein with Coomassie brilliant blue **a**, and stained for glycan with Pierce™ Glycoprotein staining kit **b**. Both gels were loaded with 4 µg of α_1AGp per lane. The control stained well for both protein and glycan (Lane 2) and for those species that grow on the glycan there is a range of degradation. As the glycoprotein decreases in size there is less glycan staining. **c**, An SDS-PAGE gel of α_1AGp after treatment with recombinant enzymes. **d**, The full versions of the TLC (top) and SDS-PAGE gels (bottom) shown in Fig. 1d&e. The dotted lines represent where two gels or TLCs have been merged. Full versions of gels and TLCs are presented in Supplementary Fig 23. **e**, Growth of the different *Bacteroides spp.* on 5 mg/ml GlcNAc and **f**, 20 mg/ml of different glycoproteins in minimal media (except for *B. massiliensis*, where this was replaced with CMB (see Materials and Methods). *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*)*. The differences in final OD is likely to relate to differences in the glycan:protein ratio in different substrates. IgA could not be obtained in high enough quantities to assess *in vivo*. **g**, Growth of *Bt* wild-type (black) and gene deletion strains (coloured) on 5 mg/ml GlcNAc and 20 mg/ml α_1AGp. The recombinant enzymes BT0455^{GH33}, BT1044^{GH18} and BT1035^{GH163} were also added to the media of the deletion strains at 5 µM (paler version of particular colour). The results are representative of at least three independent replicates.



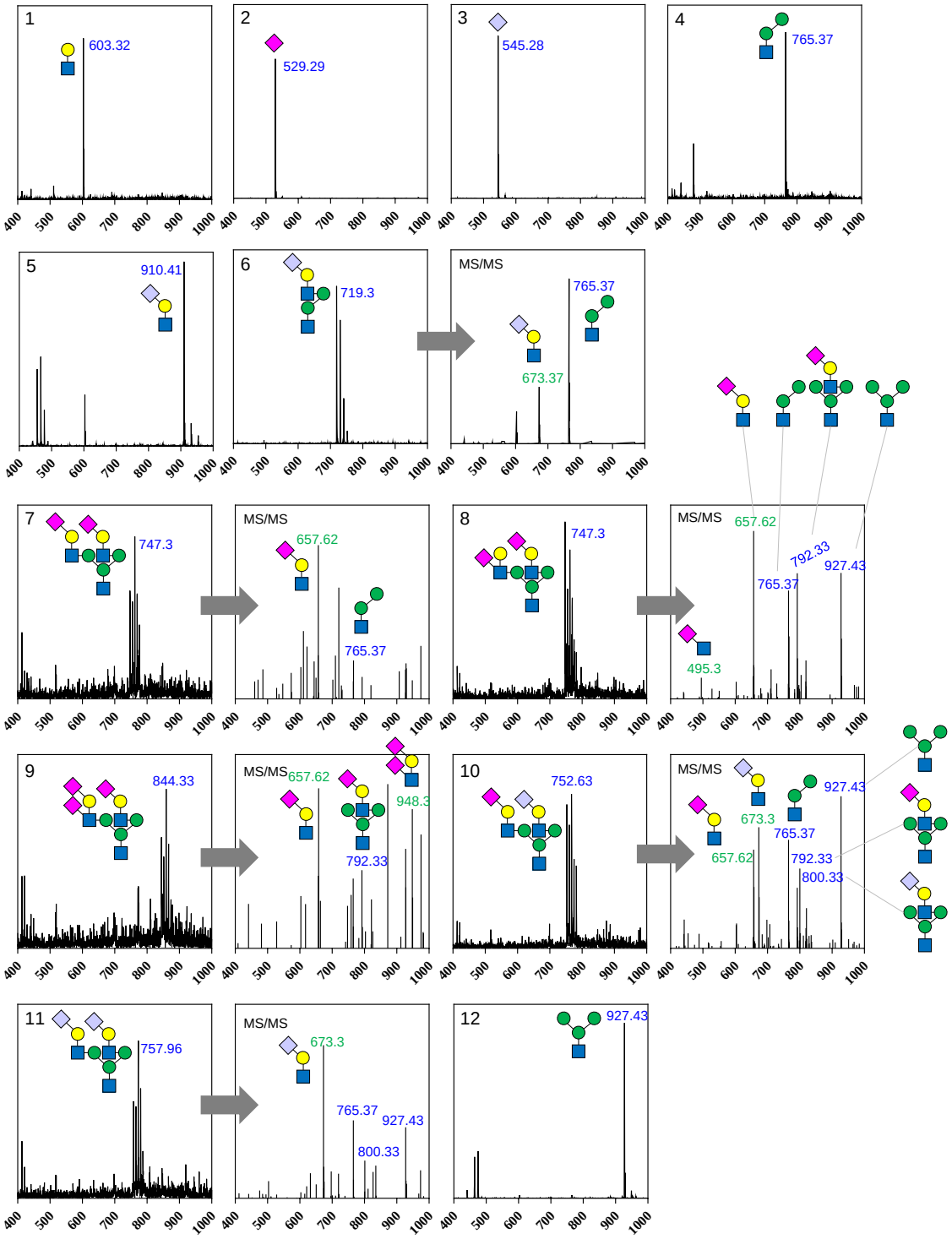
Supplementary Figure 2 | Predicted domains of putative CAZymes and SGBPs from *Bt* involved in CNG utilisation The domain structure and approximate lengths of the ORFs from *Bt* expressed during growth on α_1AGp. The domains were defined as described in the Materials and Methods. The white boxes at the start indicate the predicted signal sequences and the predicted cellular localisation according to these sequences are indicated (determined by the LipoP server, see Materials and Methods). The blue dotted lines separate out the different loci. The sequences identity between CAZymes from the same family and also the predicted SGBPs are shown in Supplementary Tables 9-12.



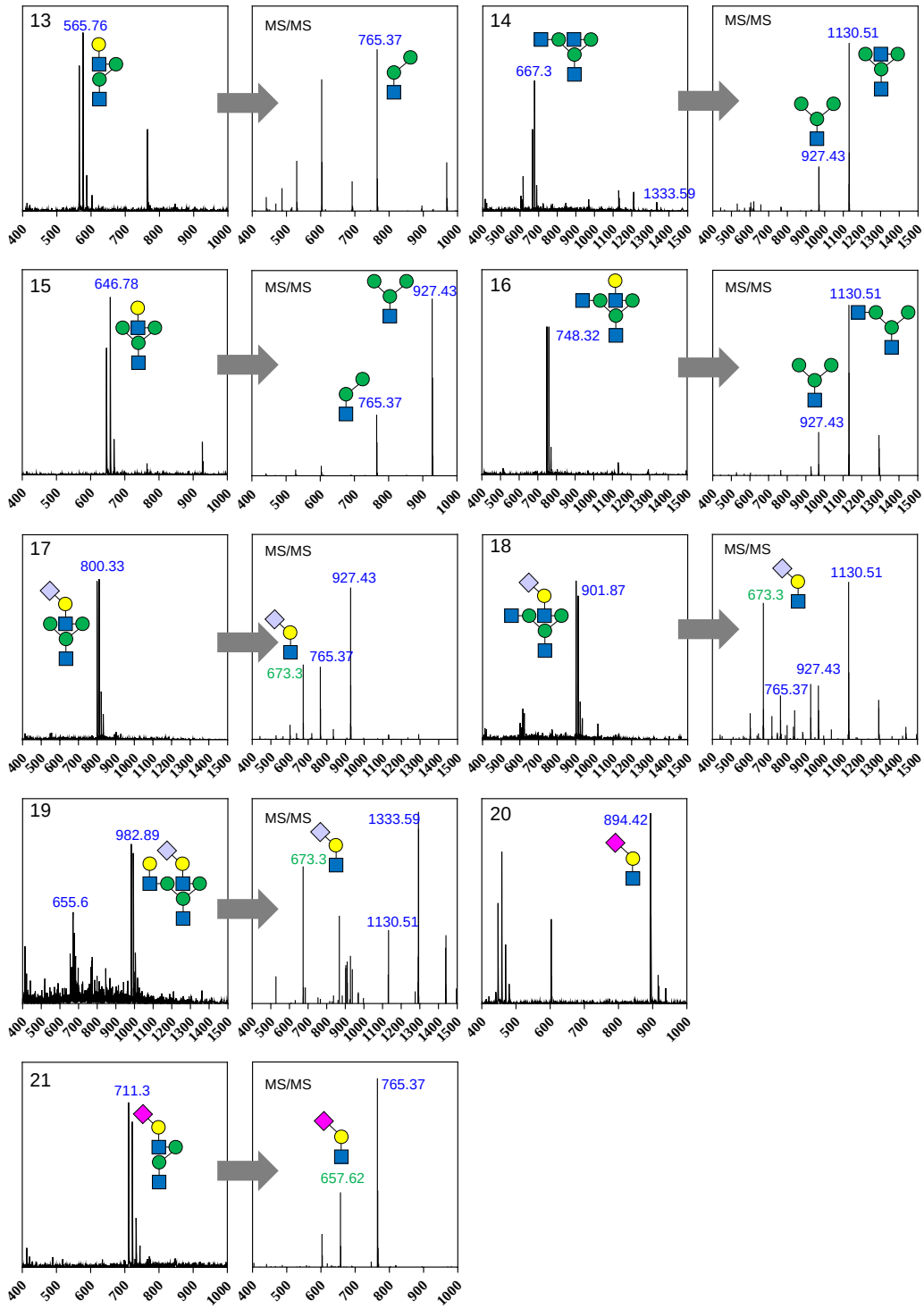
Supplementary Figure 3 | Glycans present at different time points when *Bt* is grown on α_1AGp Samples were taken from the growth presented in Fig. 1c at 10 h and 19 h and labelled with procainamide and analysed by LC-FLR-ESI-MS. The cells and the supernatant were included in the purification, so the glycans detected represent what is both inside the cell and in the supernatant. Other monosaccharides elute close to the dead volume of the column and have been mostly excluded throughout the figures for simplicity. The results are representative of at least three independent replicates.



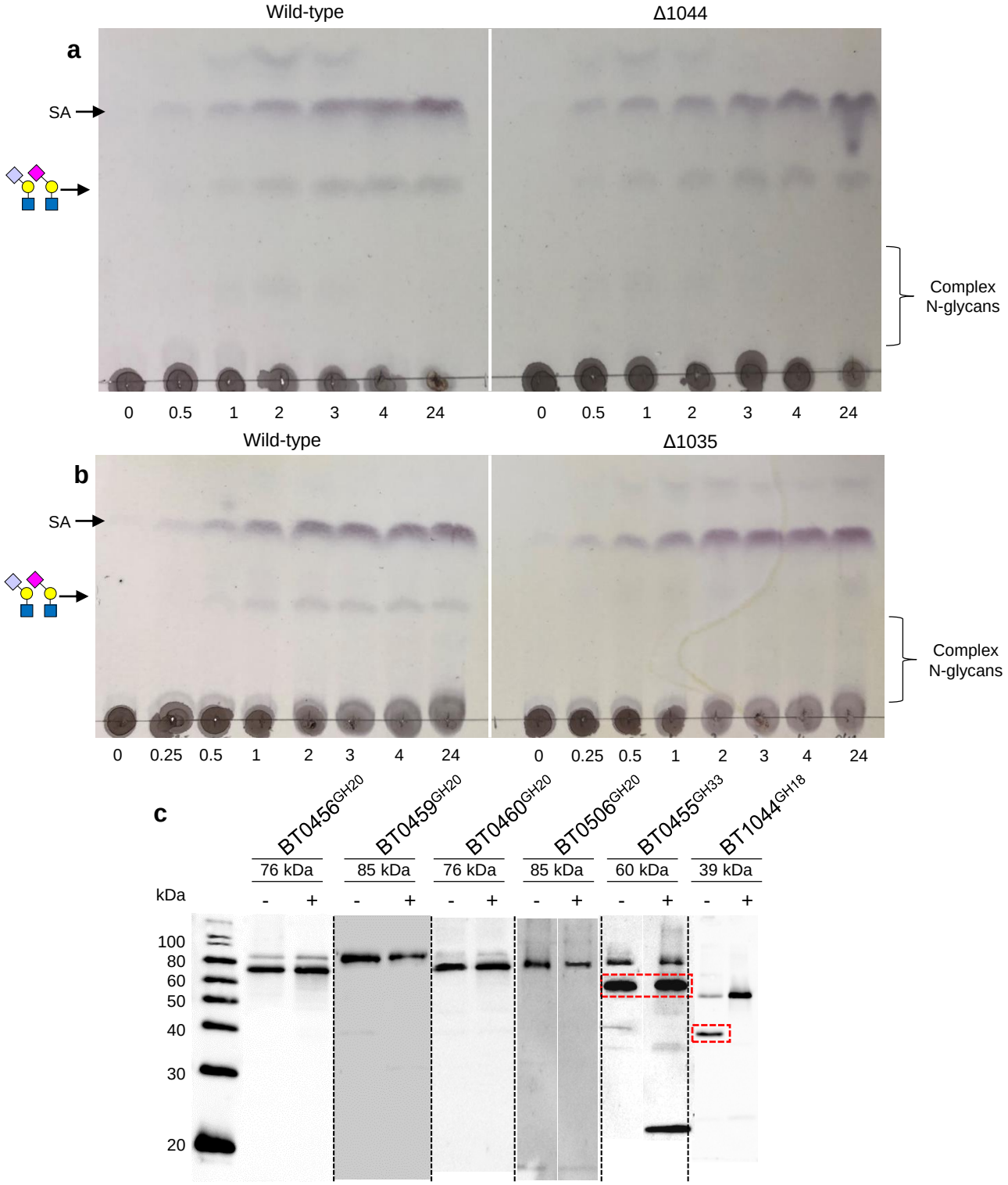
Supplementary Figure 4 | Cellular localisation of CNG active enzymes in *Bt* - Whole cell assays with *Bt* on α_1AGp Cells were grown on 10 mg/ml α_1AGp were harvested at mid-exponential, washed several times and split into different assays. The numbers along the bottom of the TLCs indicate the time course (h) and M is the marker, which contains NeuAc (pink), Galactose (blue), GlcNAc (yellow) and Mannose (blue). The latter three monosaccharides run close together and can be hard to distinguish between, but see Supplementary Figure 17 for a more insightful TLC. The same sugar(s) can also run slightly differently depending on buffer or other conditions. **a**, Whole cells, **b**, lysed cells and **d**, the growth media from the growth were incubated with 10 mg/ml fresh α_1AGp. **c**, Whole cells were also incubated without substrate. These cells were then removed at the end of this assay and the supernatant re-checked for activity against fresh substrate to make sure the cells were not lysing over the course of the experiment. See Supplementary Figure 23 for larger version (**c, inset**). **e**, Whole cells were also treated with proteinase K to remove the surface enzymes (2 mg/ml final, 16 hours), washed well and then the same assay carried out and **f**, also with lysed cells. The proteinase K treated cells were incubated with BSA for the same time course to make sure all the proteinase K had been removed. **g**, Three samples from this assay (as indicated by the coloured asterisks) were labelled with procainamide and analysed by LC-FLR-ESI-MS. The HPLC chromatograms have colours corresponding to the asterisks and are annotated with the different N-glycan structures that were deduced by mass spectrometry. Mass spectra of the different peaks in **g** are presented in Supplementary Fig. 5 and 6.



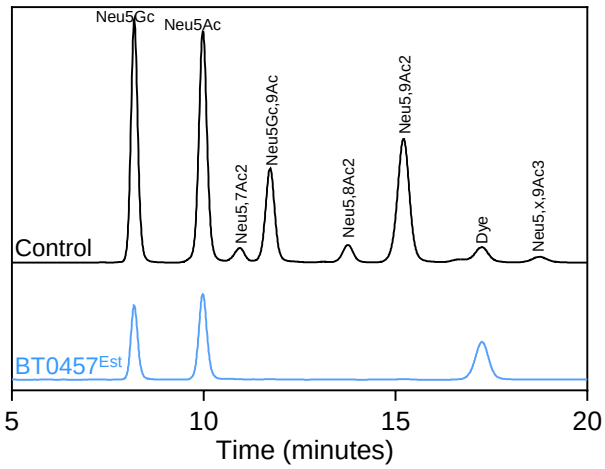
Supplementary Figure 5 | Cellular localisation of CNG active enzymes in *Bt* - Whole cell assays with *Bt* on α_1AGp Examples of mass spectra for each of the different structures identified in Supplementary Fig. 4. Also included are the MS/MS fragment data where required and indicated to by grey arrows. The numbering corresponds to the numbering in Supplementary Fig. 4g. The blue and green numbers indicate the Y and B fragments, respectively, depending on how the species fragmented.



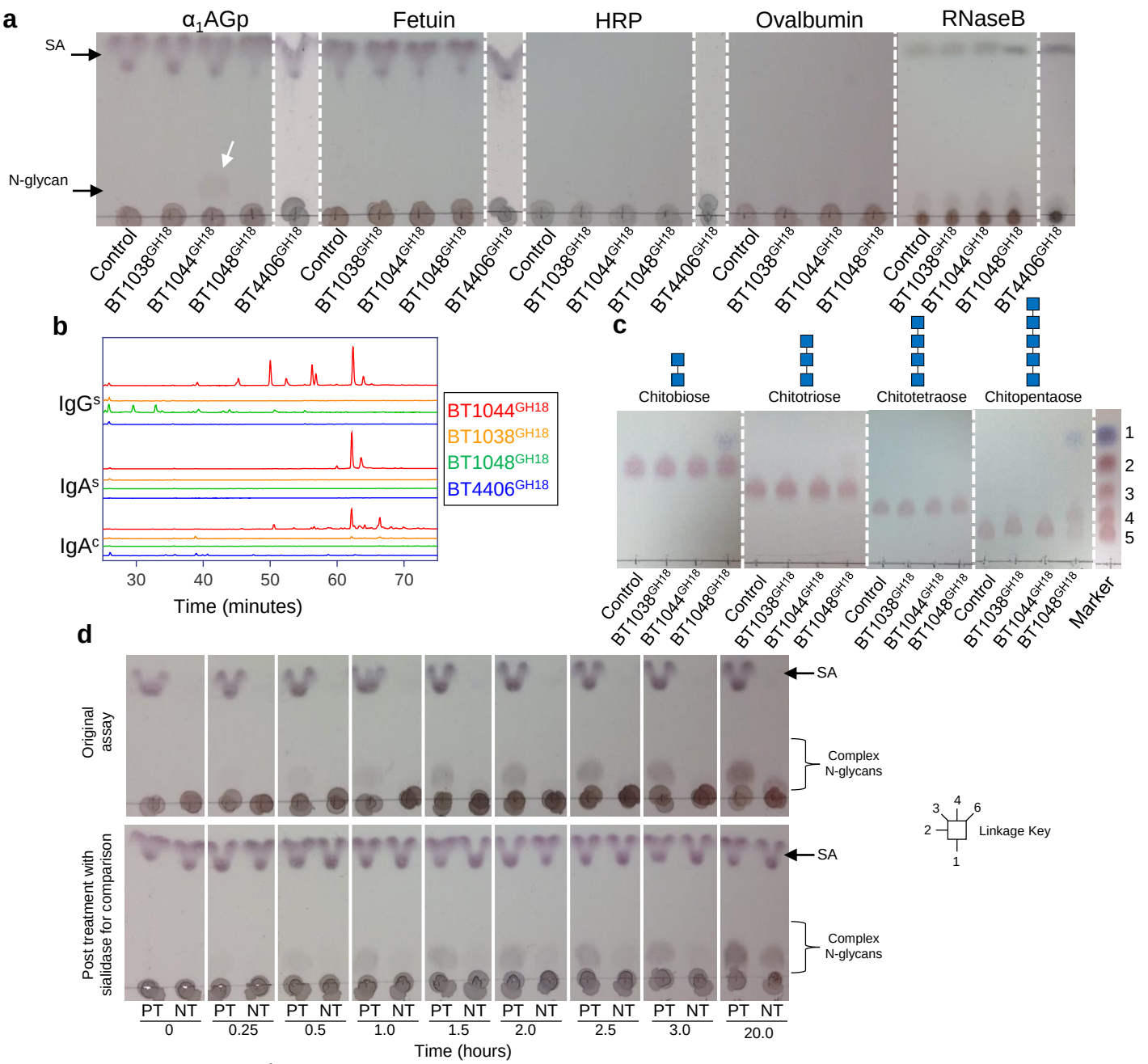
Supplementary Figure 6 | Cellular localisation of CNG active enzymes in *Bt* - Whole cell assays with *Bt* on α_1AGp Examples of mass spectra for each of the different structures identified in Supplementary Fig. 4. Also included are the MS/MS fragment data where required and indicated to by grey arrows. The numbering corresponds to the numbering in Supplementary Fig. 4g. The blue and green numbers indicate the Y and B fragments, respectively, depending on how the species fragmented.



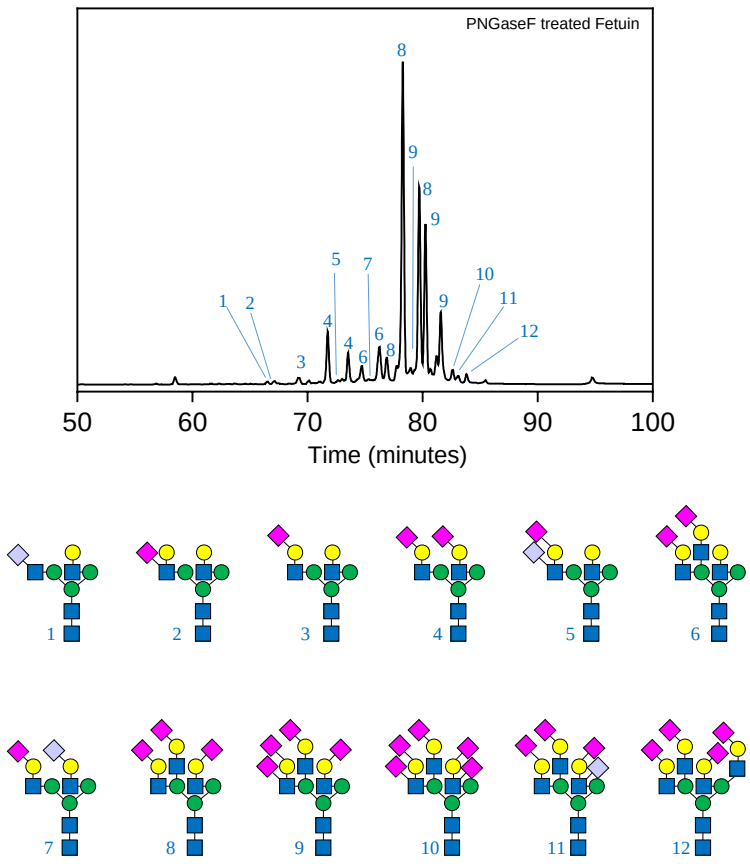
Supplementary Figure 7 | Cellular localisation of CNG active enzymes in *Bt* - Whole cell assays with *Bt* wild type and gene deletion mutants on α_1AGp a, Whole cell assays of Δ1044 compared to wild type *Bt*. **b,** Whole cell assays of Δ1035 compared to wild type *Bt*. The results are representative of two independent replicates. The white lines provide a distinction between the different assays. Full TLCs are included in Supplementary Figure 23. **Cellular localisation of CNG active enzymes in *Bt* - Western blots c**, *Bt* cells grown on 10 mg/ml α_1AGp were harvested at mid-exponential and either untreated (-) or treated with 2 mg/ml proteinase K for 16 h (+). The protein was precipitated with trichloroacetic acid, washed with acetone and the pellets resuspended in PBS for running on SDS-PAGE. The protein was subsequently transferred to nitrocellulose membrane and probed with antibodies raised against the different recombinant enzymes (see materials and Methods). The black dotted lines indicate where gels have been put together. Red dotted boxes indicate the protein band in question. Full gels are included in Supplementary Figure 23 and independent Western blots were produced at least twice for each protein. The results are representative of two independent replicates



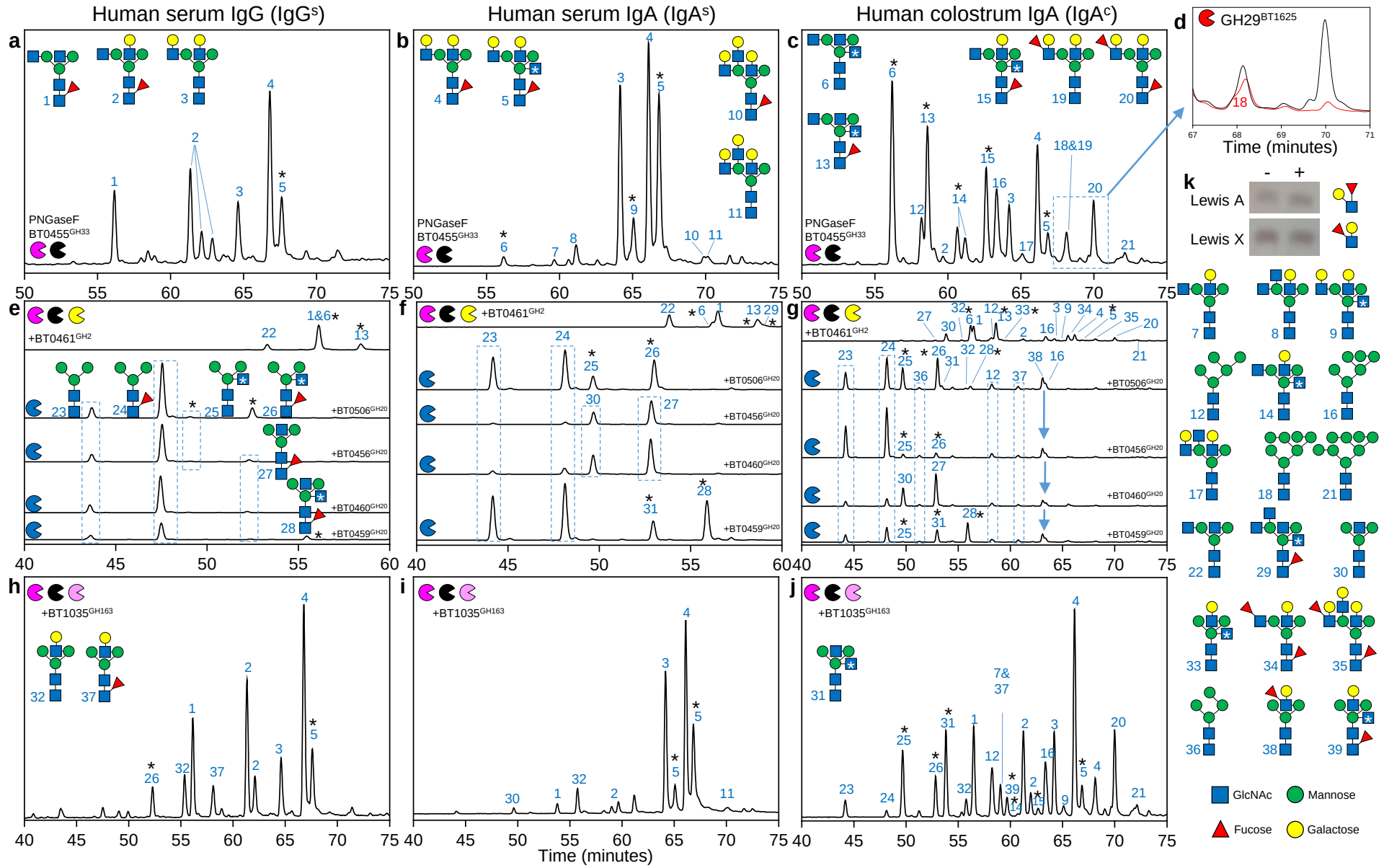
Supplementary Figure 8 | Activity of BT0457^{est} on a sialic acid reference panel Incubation was carried out overnight at 37 °C on substrates with 1 μM of BT0457 and only non-esterified species were showed to be remaining (light blue). The samples DMB-labelled post-reaction and analysed by HPLC. A control with no enzyme is included (black). The results are representative of two independent replicates.



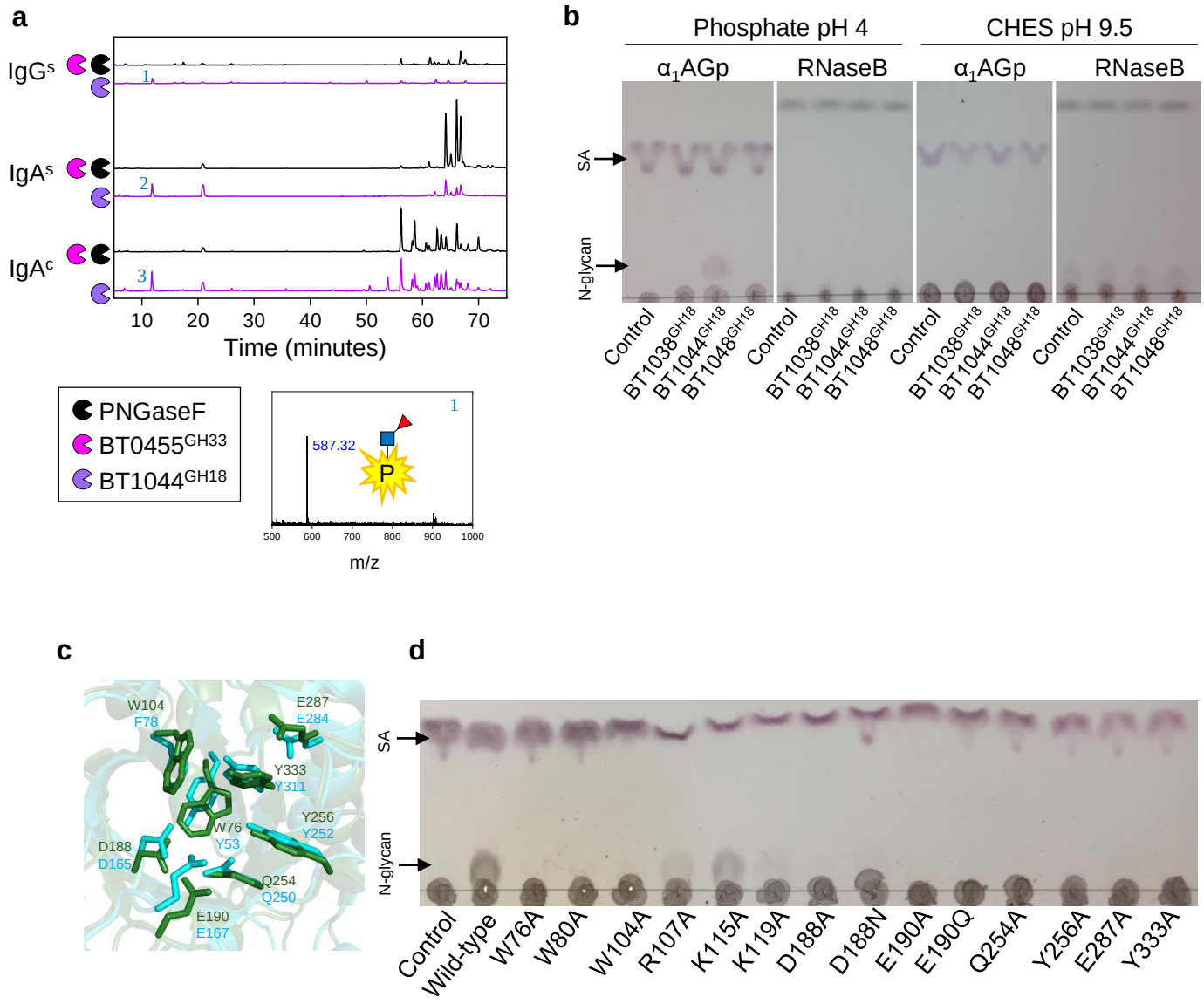
Supplementary Figure 9 | GH18 activity a, Activity of the four *Bt* GH18 enzymes expressed during growth on CNG against different glycoprotein substrates. α_1AGp, Fetuin, HRP, ovalbumin and RNaseB have bi-antennary, tri-antennary, plant-type, hybrid and high-mannose and only high-mannose decorations, respectively. Sialic acid was removed from the CNG substrates using BT0455^{GH33} prior to testing for activity. Only BT1044^{GH18} had activity on α_1AGp (white arrow). **b,** Activity of the four GH18 enzymes against different human Ig substrates. The peaks seen in one of the BT1048^{GH18} traces were not glycans, just contamination. **c**, Activity of the GH18 enzymes against different chitooligosaccharides. A ladder of chitooligosaccharides is shown with degrees of polymerisation from 1-5. **d**, The activity of BT1044^{GH18} was tested against α_1AGp with the sialic acid present (NT=no treatment) and also when it had been removed with BT0455^{GH33} (PT=pre-treatment). The top TLC shows the original assay, where for each time point there is a sample with sialic acid removed (PT; monosaccharide stained pink) and a sample where sialic has been left on the glycan (NT; no pink staining). At T=20 h the de-sialylated N-glycans are further up the TLC than the sialylated N-glycans. For a direct comparison, the NT samples were post-assay treated with BT0455^{GH33} and the results shown in the bottom TLC. The N-glycans from the two different assays now run to the same position with N-glycans from PT appearing earlier and present at a higher concentration at the end of the assay compared to those in the NT condition. This result shows BT1044^{GH18} has a preference for de-sialylated CNGs. The white dotted lines indicate where different TLCs have been merged together and the solid white lines provide a distinction between the different time points of the assays. Full TLCs are included in Supplementary Figure 23.



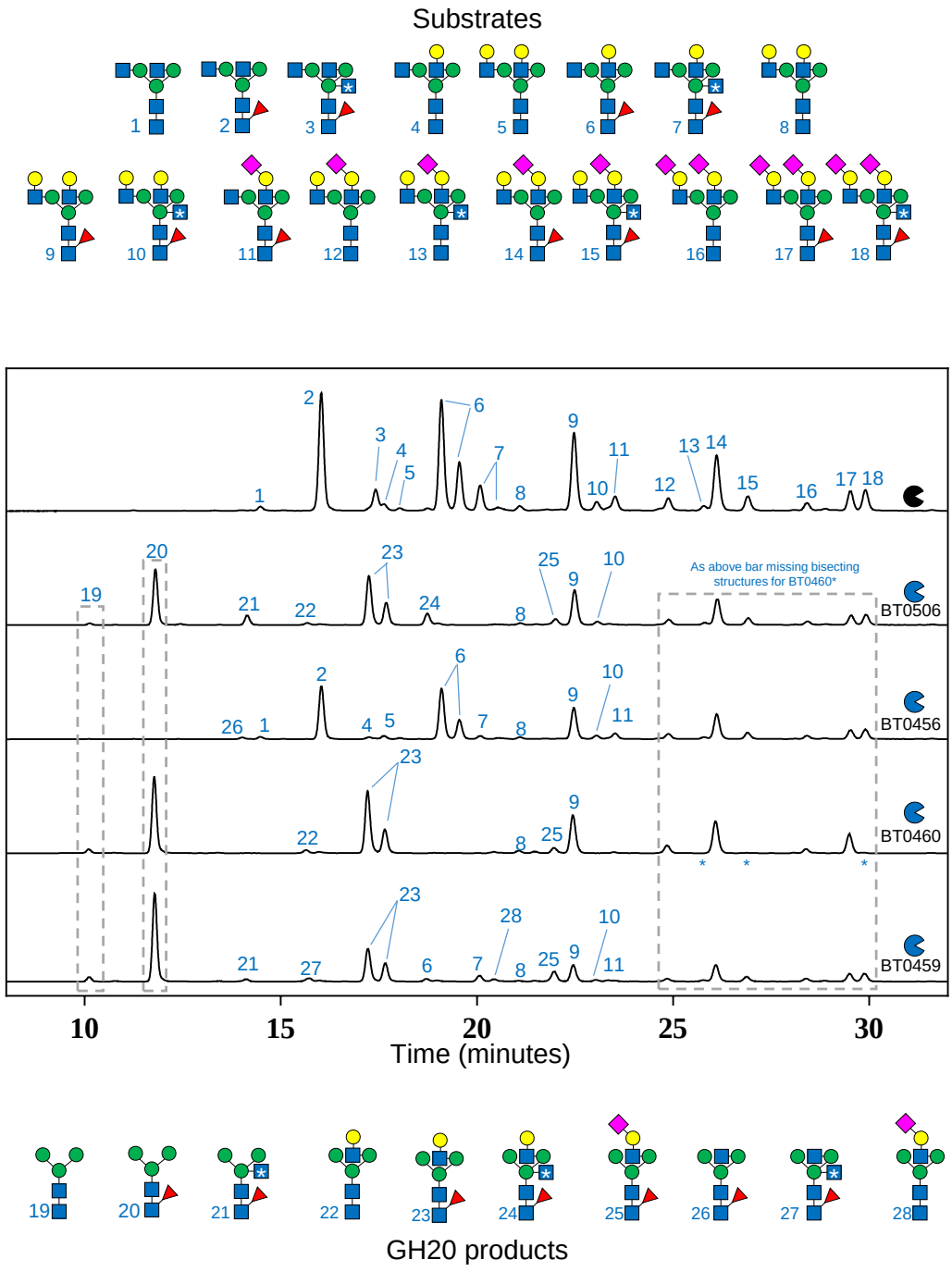
Supplementary Figure 10 | The CNGs that commercially available PNGaseF can remove from bovine fetuin. The results are representative of at least three independent replicates.



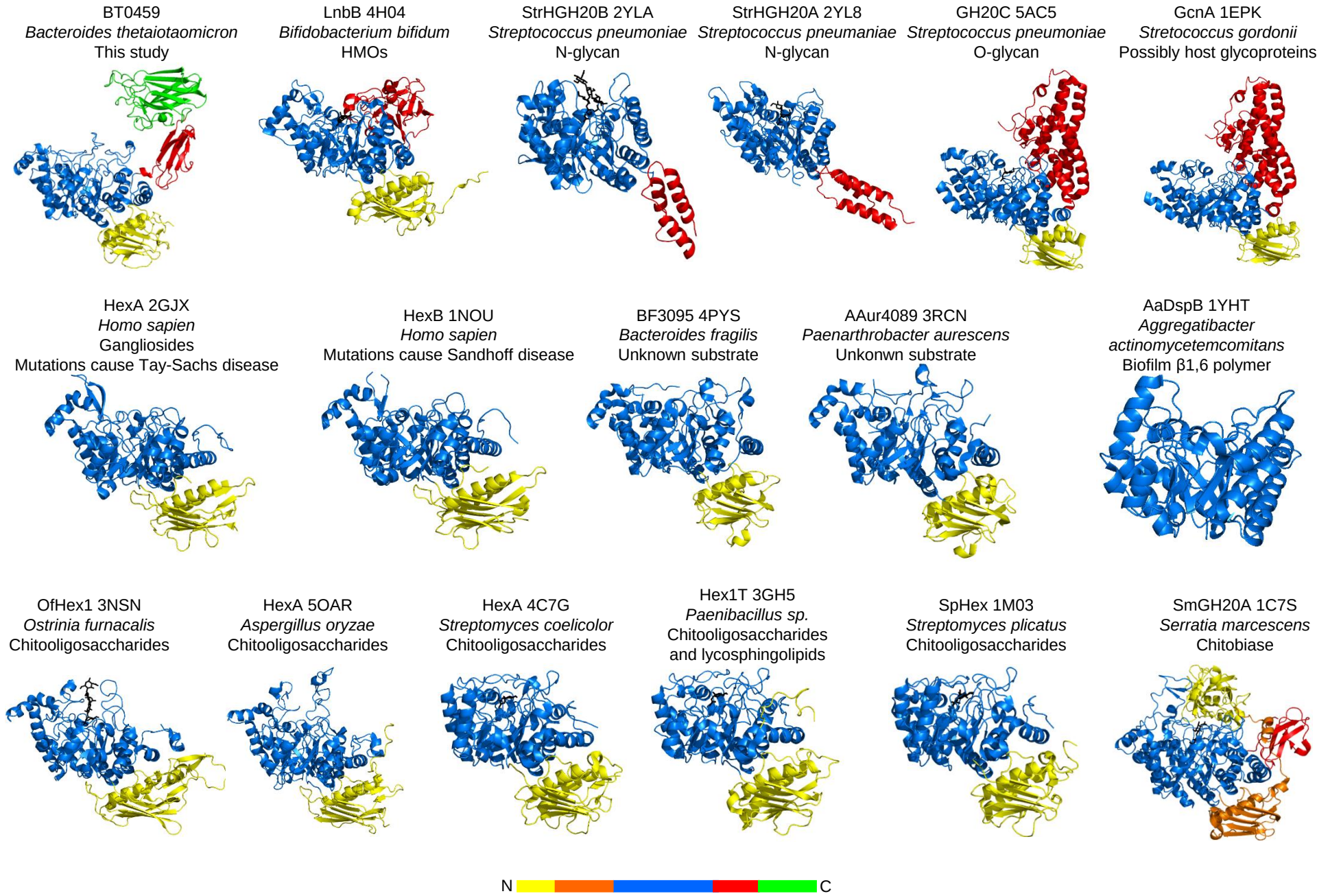
Supplementary Figure 11 | Activity of *Bt* CAZymes on human immunoglobulin derived CNGs a-c, CNGs from three human sources were treated with commercially available PNGaseF (black) alongside BT0455^{GH33} (pink). The products are predominantly biantennary, but there were also significant amounts of antenna structures missing galactose. Core fucosylation and bisecting GlcNAc (asterisks) are common to all samples, but triantennary species and antenna fucosylation were only detected in serum IgA (IgA^{s}) and colostrum IgA (IgA^{c}), respectively. Bisecting structures are indicated throughout with an asterisk. **d**, Addition of BT1625^{GH29} (red) to colostrum IgA CNG resulted in antenna fucosylation not being detected (species #19 & 20), but core fucosylation remained. The CNG in (**a-c**) were treated with BT0461^{GH2} (yellow)(top chromatogram, e-g). All the galactose could be removed in the IgG^{s} and IgA^{s} substrates, but there were small amounts of galactose of some structures remaining in IgA^{c}. The activity of the four GH20 enzymes were tested individually against the BT0461^{GH2}-treated samples (remaining chromatograms, e-g). **h-j**, The activity of BT1035^{GH163} (link pink) was tested against the CNG produced by PNGaseF and BT0455^{GH33} (a-c). **k**, BT0461^{GH2} has no activity against Lewis A and X substrates. The results are representative of at least two independent replicates.



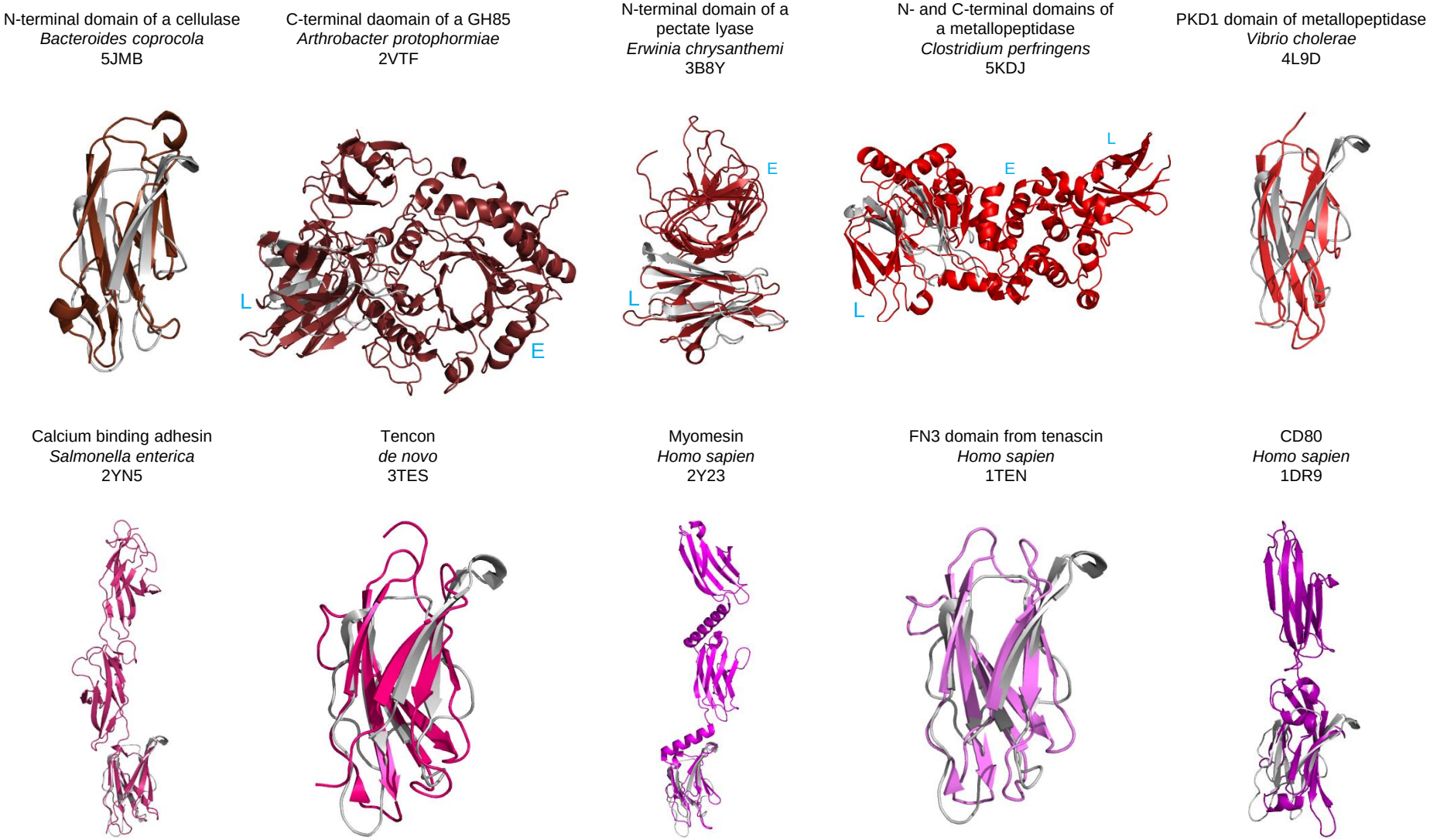
Supplementary Figure 12 | GH18 activity a, PNGase-liberated and BT0455^{GH33}-treated Ig CNGs (black lines) were treated with BT1044^{GH18} (pink lines). The assay was carried out for IgG^{s}, IgA^{s} and IgA^{c} (top, middle and bottom traces, respectively). The release of a GlcNAc-Fucose disaccharide could be seen in all samples, indicating BT1044^{GH18} can access core fucosylated CNGs, although potential substrate did remain. A mass spectrum of one of these peaks is given as an example. **b**, Activity of GH18 enzymes at different change in pH. Glycoproteins with CNG and HMNG were used to test these parameters. The solid lines differentiate between the different conditions and substrates. **c**, The crystal structures of the active sites of BT1044^{GH18} (dark green) and Endo F3 (1EOM; cyan) are overlaid and the +1 subsite residues are shown as sticks. **d**, Activity of various point mutant of active site residues in BT1044^{GH18} against α_1AGp. Many of the mutations result in loss of or significant decrease in activity compared wild type. The results are representative of at least two independent replicates. Full versions of the TLCs can be found in Supplementary Fig. 23.



Supplementary Figure 13 | Activity of the *Bt* GH20 enzymes against human IgG CNGs and defined oligosaccharides PNGase-released CNG was hydrolysed with GH20 enzymes. No sialidase or galactosidase were used in this assay. The blue asterisks indicate to the peaks of bisecting structures that have been hydrolysed by BT0460^{GH20}. The results are representative of two independent replicates.

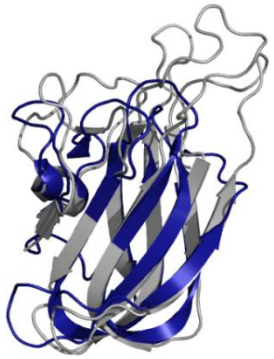


Supplementary Figure 14 | Currently available crystal structures of GH20 enzymes The catalytic modules are in blue. The N-terminal domains are coloured yellow and a second N-terminal domain is coloured orange in the case of 1C7S. C-terminal domains are coloured red except for BT0459^{GH20}, where the FN3 linker domain is coloured red and the C-terminal F5/F8 Type C domain is in green. Any bound glycans are shown as black sticks.

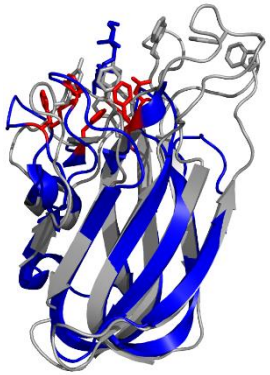


Supplementary Figure 15 | Structural homologues of the FN3 linker domain from BT0459^{GH20} The FN3 linker domain from BT0459^{GH20} is shown in grey and overlaid with the closest structural hits found using the PDBeF and DALI search engines. The identity of each structural homologue is stated, along with the name of the organism the protein is from and the PDB code. For those structures where it is appropriate the catalyticand linker modules are indicated to with an E or L, respectively.

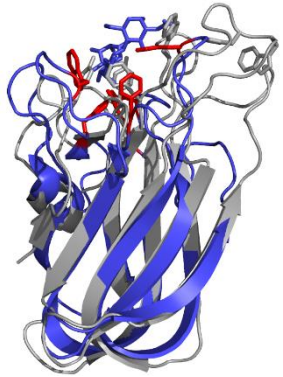
CBM32
Paenibacillus fukuinensis
4ZZ5



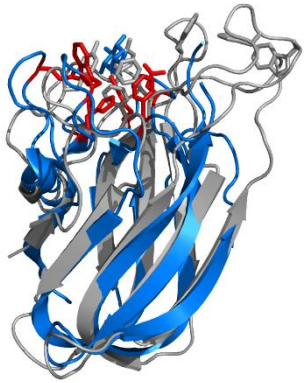
CBM32 appended to a GH89
Clostridium perfringens
4A45
α-D-GalNAc-β-D-Gal



CBM32 appended to a GH84
Clostridium perfringens
2J7M
α -L-Fuc-(1,2)-β-D-Gal-(1,4)-α-D-GlcNAc-

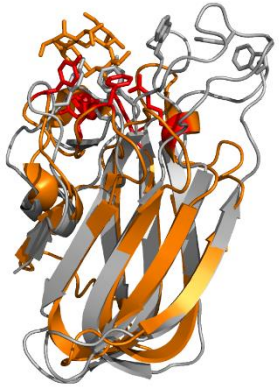


CBM32 appended to a GH84
Clostridium perfringens
2J1A
β-D-Gal

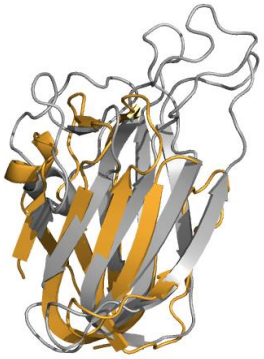


Putative CBM appended to GH20
Serratia marcescens
1C7S
β-D-GlcNAc-(1,4)-β-D-GlcNAc

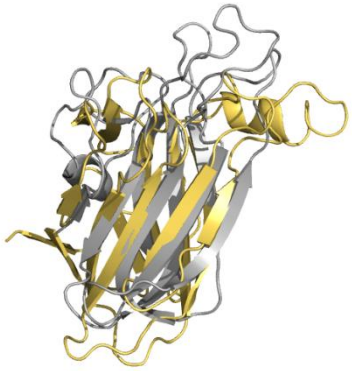
CBM47
Streptococcus mitis
3LEK
Lewis B



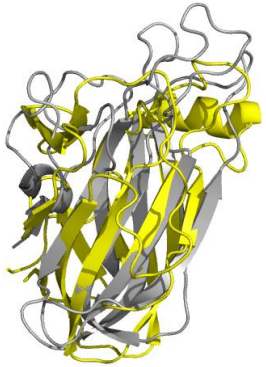
Putative CBM
appended to a GH84
Bacteroides thetaiotaomicron
5FL1



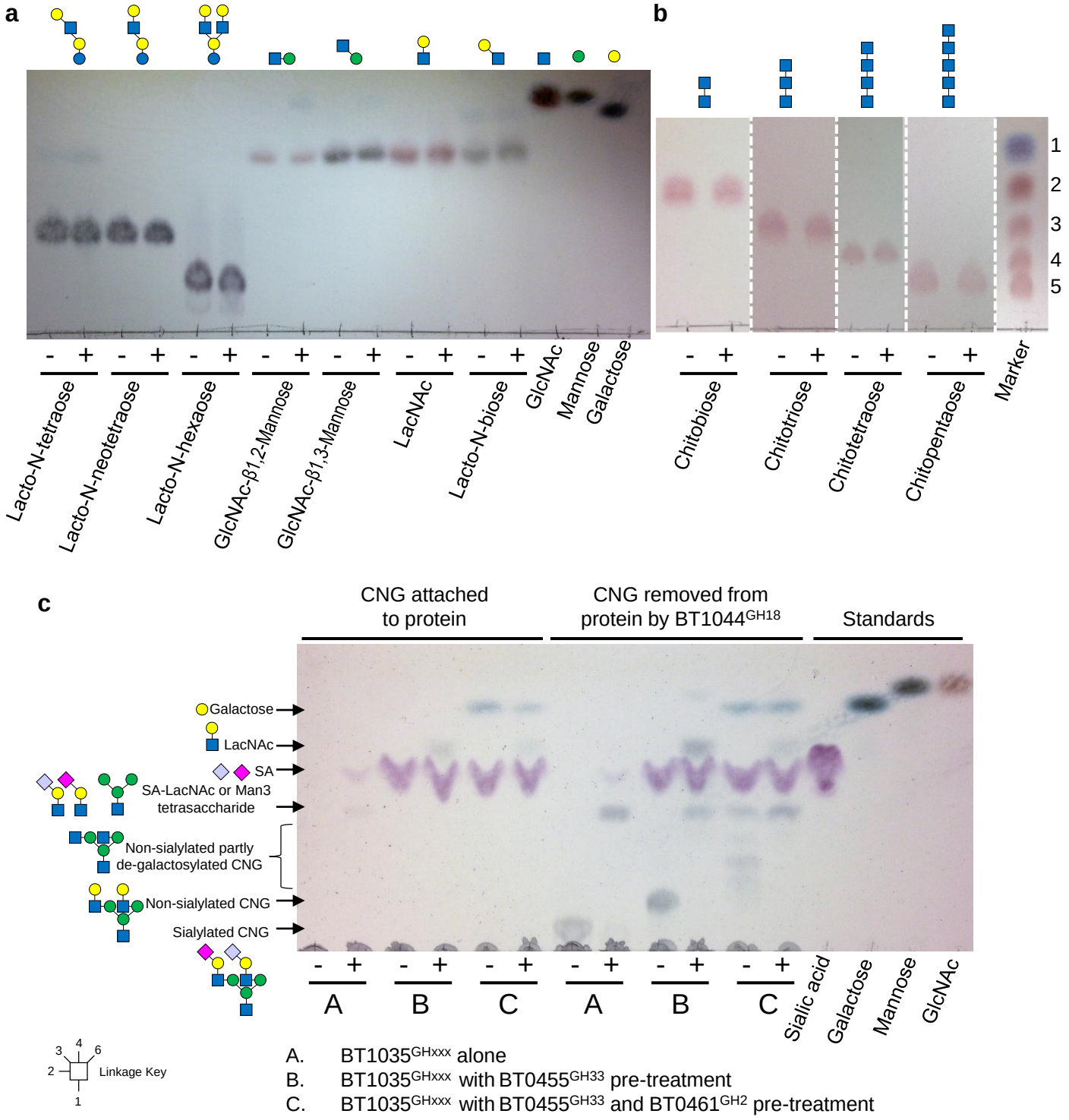
Putative CBM
appended to GH115
Bacteroides ovatus
4C90



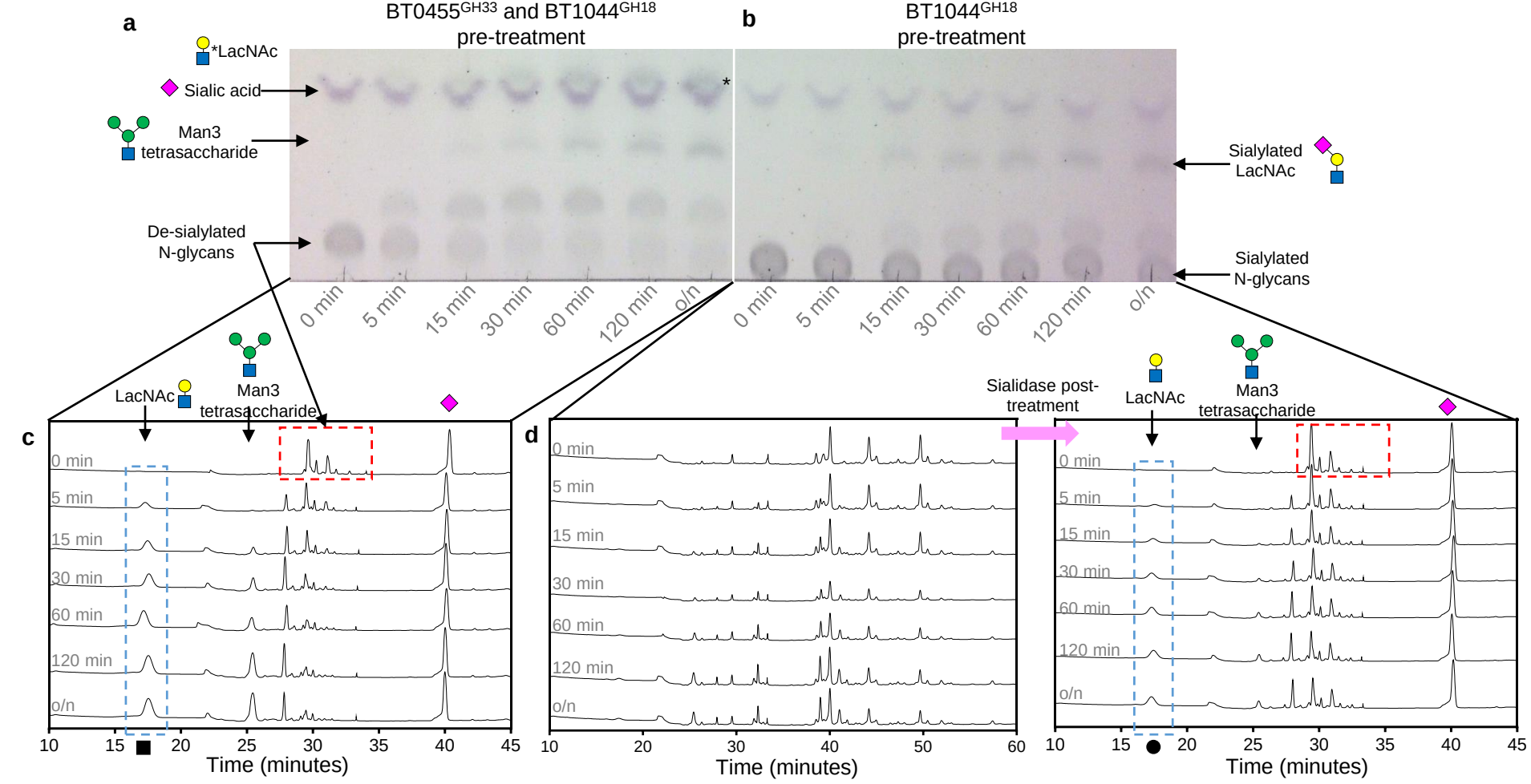
Putative CBM
appended to GH115
Sacchrophagus degradans
4ZMH



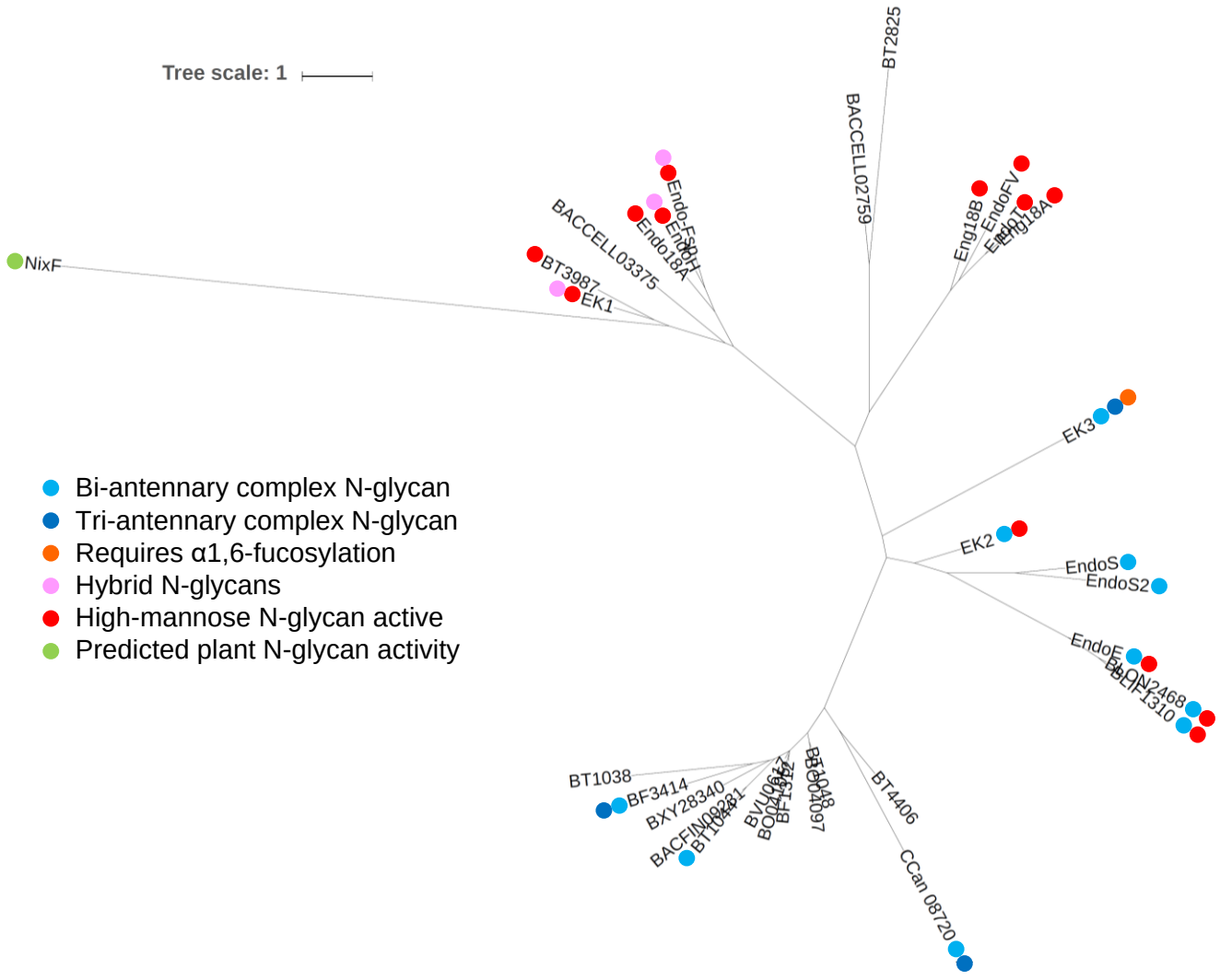
Supplementary Figure 16 | Structural homologues of the C-terminal domain from BT0459^{GH20} The C-terminal F5/F8 Type C domain from BT0459^{GH18} is shown in grey and overlaid with the closest structural hits found using PDBeF and DALI search engines. In the BT0459^{GH20} structures, solvent exposed aromatic residues that could potentially be involved in substrate binding (also Main Figure 6h) are shown as sticks. Those CBMs with ligand also have the interacting residues as read sticks for comparison. The family of CBM or appended enzyme is stated, along with the name of the organism, the PDB code and the bound ligand.



Supplementary Figure 17 | Activity of BT1035^{GH163} against β-GlcNAc containing di- and oligosaccharides and CNGs a, disaccharides with no added enzyme (-) and BT1035^{GH163} added (+). Standard monosaccharides are shown. **b**, Chitooligosaccharides with no added enzyme (-) and BT1035^{GH163} added (+). A ladder of chitooligosaccharides is shown with degrees of polymerisation from 1-5 and the dotted lines indicate where TLCs have been merged together. **c**, Activity of BT1035^{GH163} on α_1AGp. Pre-treatments of no additional enzymes, BT0455^{GH33}, and BT0455^{GH33} plus BT0461^{GH2} were compared when the CNG was still attached to the native glycoprotein (left section) and when it had been removed by BT1044^{GH18} (centre section). Standards for the different monosaccharides are shown. The results show that BT1035^{GH163} has better activity when the CNG has been removed from the protein, but some activity is seen under all conditions. The results are representative of at least two independent replicates. The dotted white lines indicated where different TLCs have been merged together and full versions of all TLCs can be found in Supplementary Fig. 23

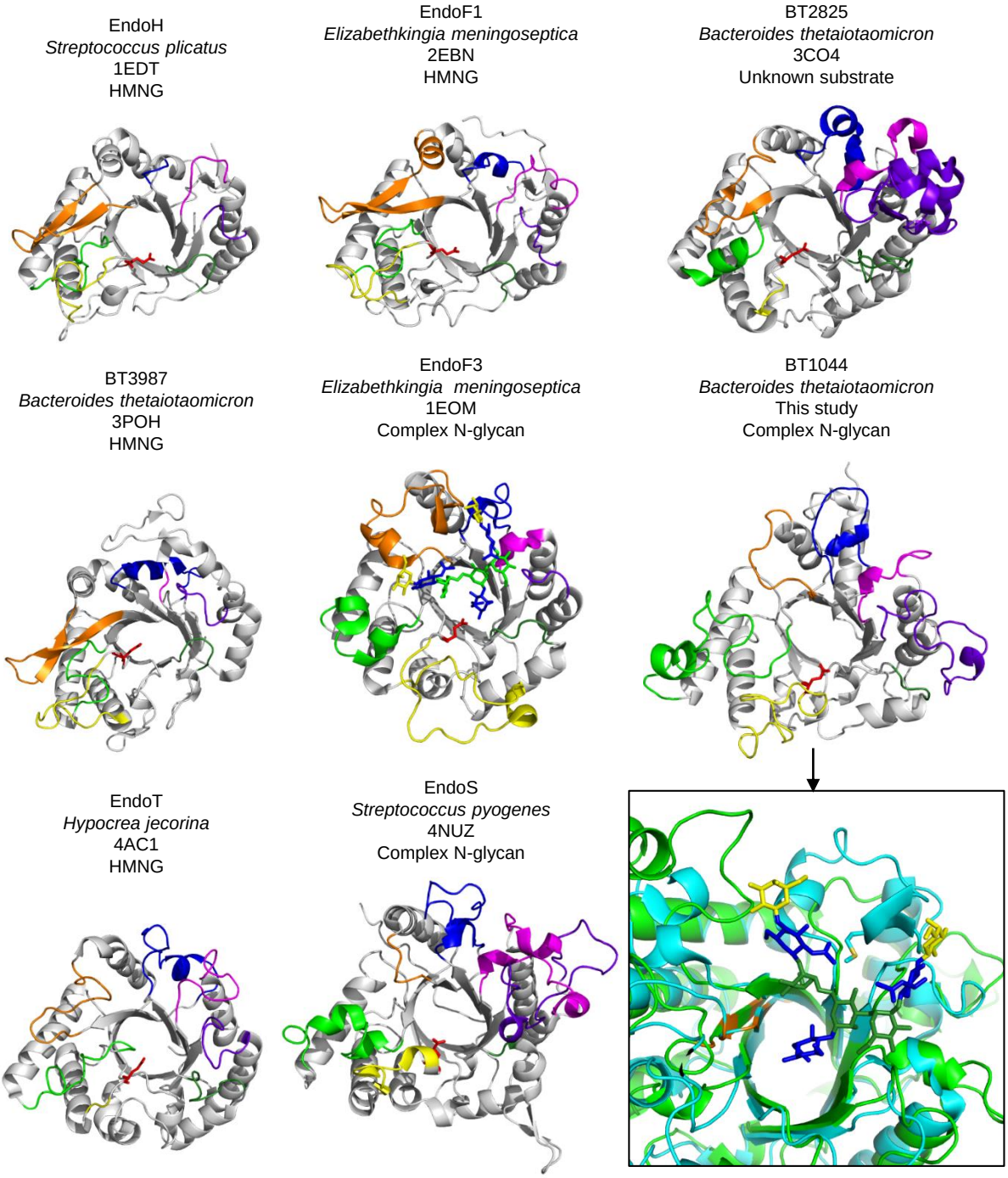


Supplementary Figure 18 | BT1035^{GH163} activity on sialylated and de-sialylated CNG Sialylated and de-sialylated CNGs were prepared from α_1AGp using BT1044^{GH18} and BT0455^{GH33}. The activity of 50 nM BT1035^{GH163} was tested against these two substrates at a final concentration of 6 mg/ml. **a**, A TLC of the assay on desialylated CNGs showing sialic acid and de-sialylated N-glycans at T=0 (labelled for clarity). The production of LacNAc (asterisk) and Man3 tetrasaccharide (labelled) can be seen over time, which run directly above and below sialic acid (stained pink and labelled), respectively. **b**, A TLC of the assay on sialylated CNGs, where sialylated LacNAc and Man3 run in the same position (labelled). For a direct comparison the two assays the originally sialylated CNGs were treated post-reaction with sialidase and LacNAc released over time quantified using HPAEC-PAD (**c**,**d**). The blue dotted boxes highlight the increasing LacNAc over time and the red dotted boxes highlight the de-sialylated CNG substrate at the beginning of the reaction. **e**, The area of the LacNAc peaks was used to determine concentration using a standard curve. The results are representative of two independent replicates.

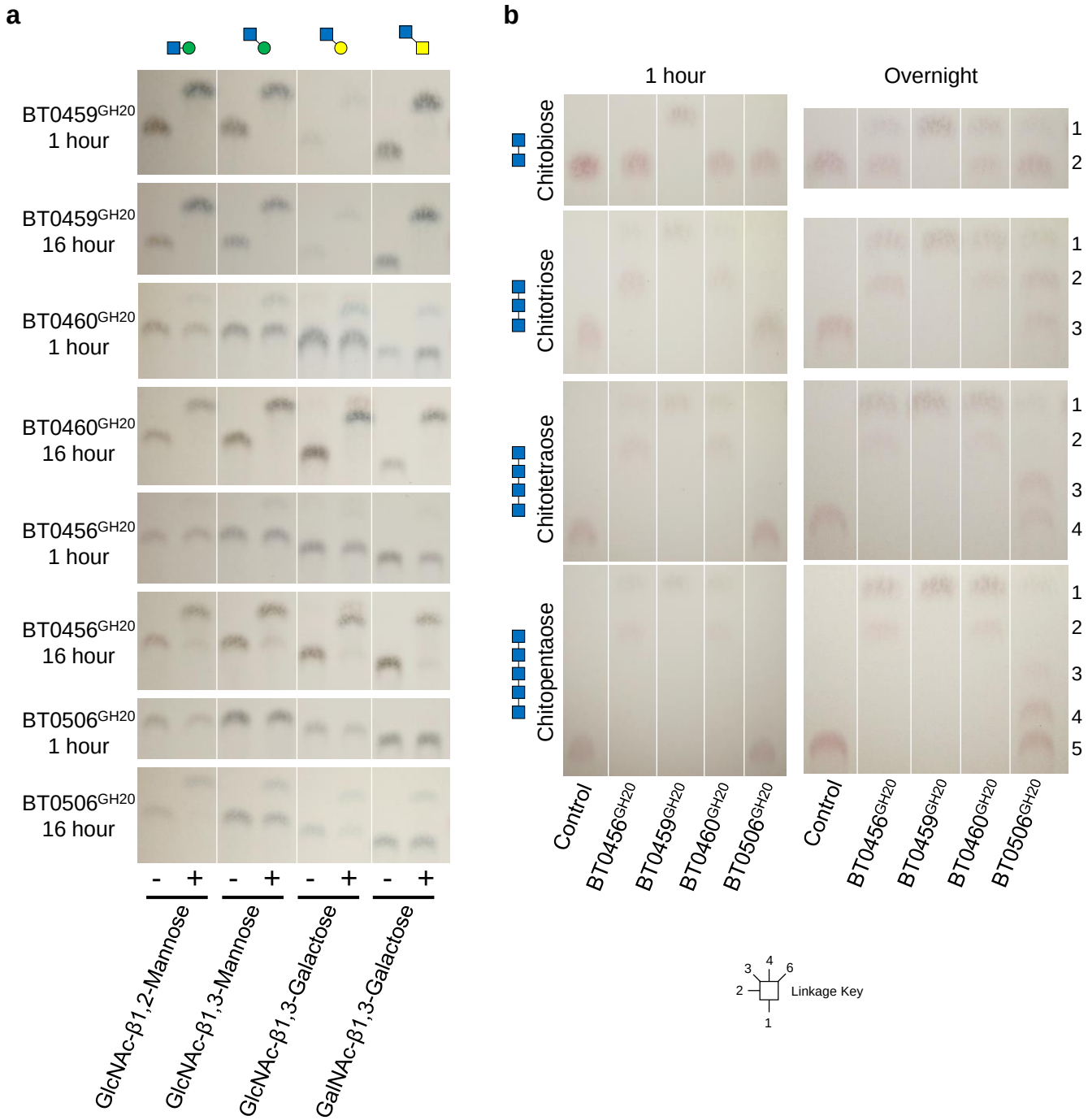


Supplementary Figure 19 | Phylogenetic tree of GH18 family members The sequences of GH18 family members with reported N-glycan activities (Supplementary Table 3) and those highlighted by the *in silico* analysis (Supplementary Fig. 20 and Supplementary Table 13) were compared as described in Materials and Methods. 32 different enzymes are included from 20 different species. Sub-activities can be seen branching off together and could be useful in making predictions about the activities of enzymes with unknown function.

Supplementary Figure 20 | Structure of CNG upregulated loci in *Bt* and an *in silico* analysis of their conservation in other gut *Bacteroides* species. *Bt* ORFs were used to search for homologues in other *Bacteroides* species using the Bt ones to search with (see Materials and Methods). Predicted transcriptional regulators have mostly been left out for clarity. The lines connect the homologues. The red lines connect the GH2 members that are predicted to be β-mannosidases, whereas those predicted to be β-galactosidases are connected with blue lines. GH and other protein families are colour-coded: GH2 (blue), GH18 (purple), GH20 (blue), GH29 (red) GH92 (green), GH97 (dark green), GH33 (pink), GH130 (orange), esterase (cyan), sulfatases (black, yellow writing), mutarotase (black), SusC/D-like pairs (grey) and the family GH163 (light pink). Abbreviations include: MFS - major facilitator superfamily, TM – transmembrane, R - regulator.



Supplementary Figure 21 | Comparison of structures of the catalytic modules of GH18 family members with N-glycan activity. The left and middle columns are enzymes with HMNG and CNG activity, respectively. The right column contains a GH18 of unknown activity (3CO4) and BT1044^{GH18} reported in this study. Loops 1-7 for all structures are coloured yellow, green, orange, blue, magenta, purple and dark green, respectively. The glutamate in red sticks is a part of the characteristic GH18 motif and the proton donor for hydrolysis. EndoF3 has an octasaccharide bound, where GlcNAc, mannose and galactose are blue, green and yellow, respectively. This enzyme-product (cyan) complex is overlaid with BT1044^{GH18} (green) in the bottom right-hand box and the catalytic motifs are coloured orange. This overlay highlights the area where a bisecting GlcNAc would be and shows the BT1044^{GH18} appears capable of accommodating a bisecting sugar, but EndoF3 a methionine (sticks) protruding into this space that would clash with the GlcNAc of a bisecting structure. See Main Figure 4e for surface representation of this.

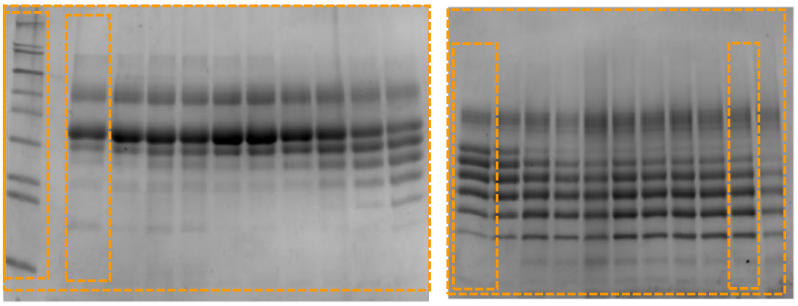


Supplementary Figure 22 | Activity of the *Bt* GH20 enzymes against defined oligosaccharides a, Activity of GH20 enzymes against β-GlcNAc containing disaccharides. **b**, Activity of GH20 enzymes against chitooligosaccharides. The results are representative of two independent replicates. The solid white lines differentiate between different assays and full versions of all the TLCs can be found in Supplementary Fig. 23.

Supplementary Figure 23 | Full image disclosure of TLCs, Western blots and SDS-PAGE gels. a, SDS-PAGE from Figure 1d and Supplementary Figure 1d. **b**, TLC from Figure 1e and Supplementary Figure 1d. **c**, SDS-PAGE from Supplementary Figure 1a-c. **d,** TLC from Supplementary Figure 4a-f. **e,** TLC from Supplementary Figure 7a. **f,** TLC from Supplementary Figure 7b. **f,** Western blots from Supplementary Figure 7c. **g,** TLC from Supplementary Figure 9a. **h,** TLC from Supplementary Figure 9c and 17b. **i,** TLC from Supplementary Figure 9d. **j,** TLC from Supplementary Figure 12b. **k,** TLC from Supplementary Figure 12d. **l,** TLC from Supplementary Figure 11k. **m,** TLC from Supplementary Figure 22a. **n,** TLC from Supplementary Figure 22b. **o,** TLC from Supplementary Figure 18. **p,** TLC from Supplementary Figure 17a and c. Orange dotted boxes indicate the regions used in the figure.

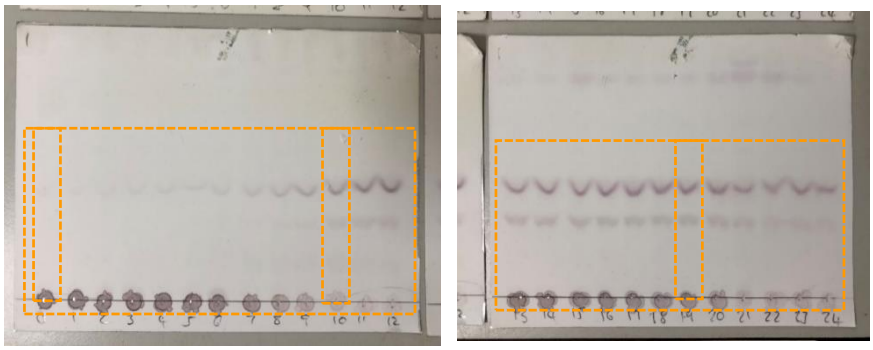
a

SDS-PAGE from Figure 1d and Supplementary Figure 1d



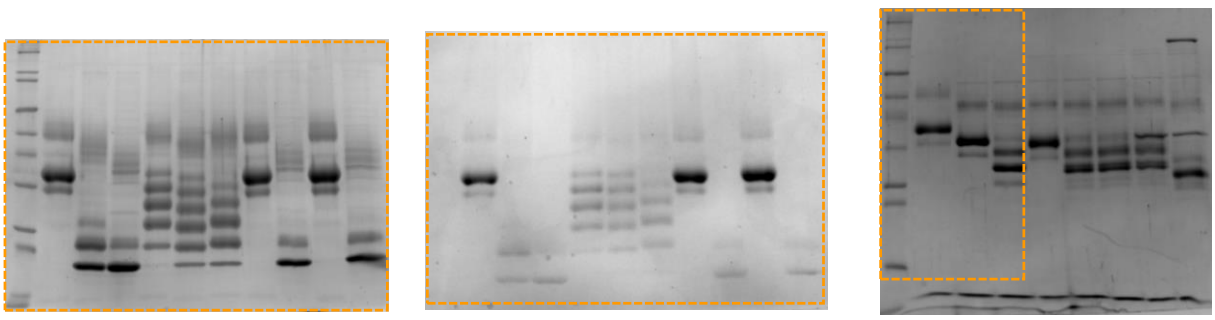
TLC from Figure 1e and Supplementary Figure 1d

b



c

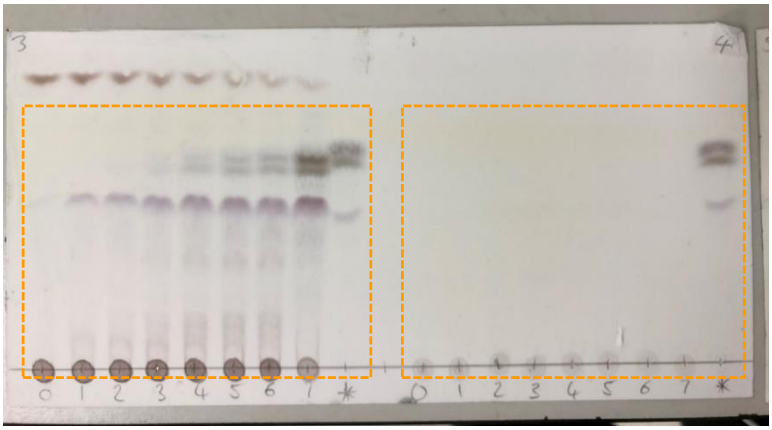
SDS-PAGE from Supplementary Figure 1a-c



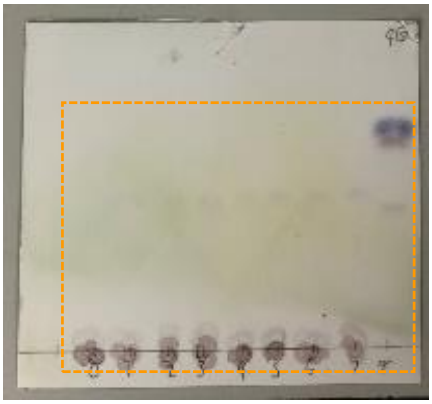
d TLC from Supplementary Figure 4a-f

SF4b

SF4c

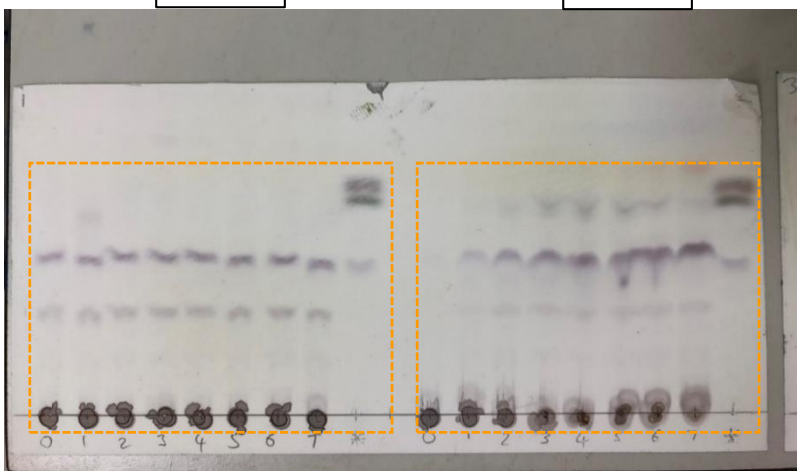


SF4c inset



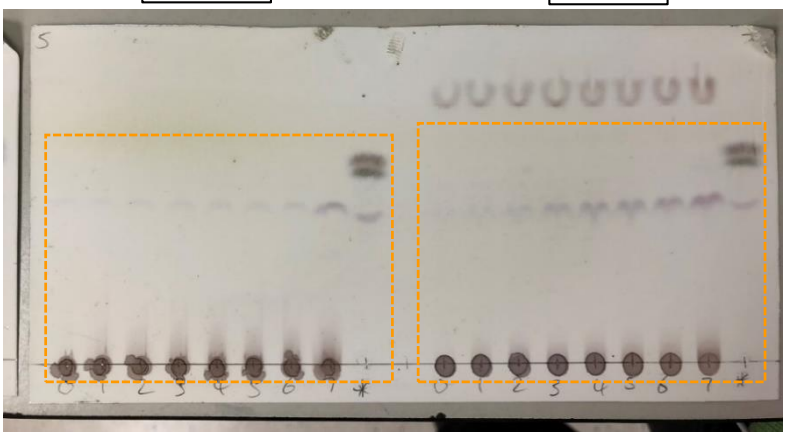
SF4d

SF4a



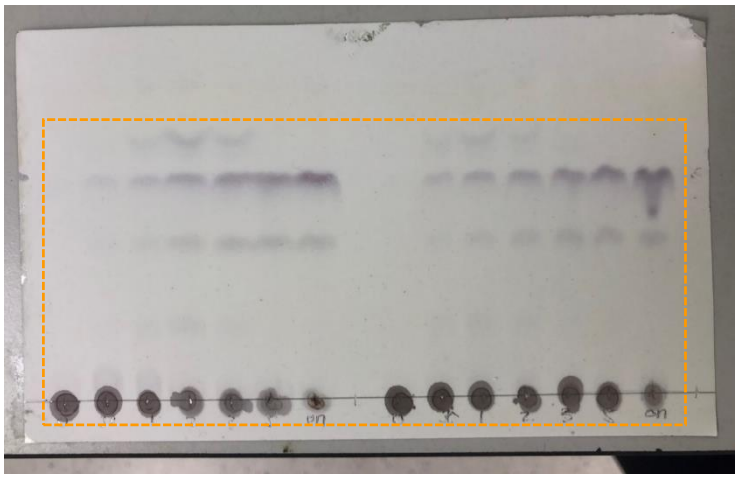
SF4e

SF4f

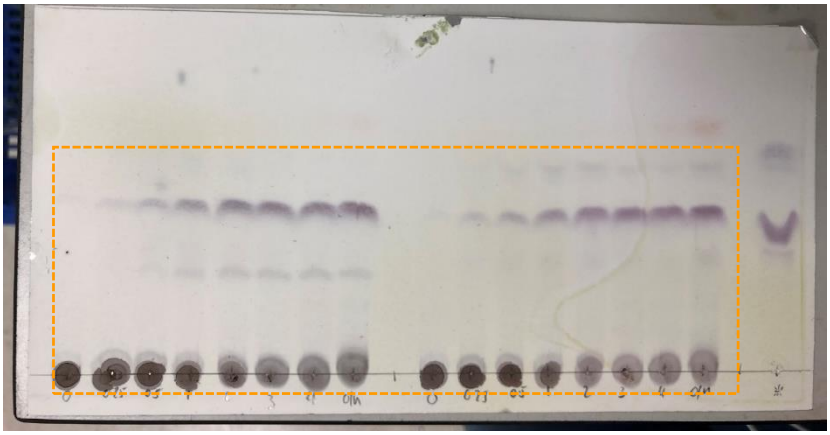


e

TLC from Supplementary Figure 7a

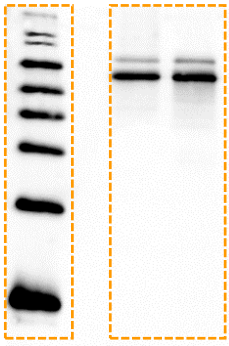


TLC from Supplementary Figure 7b

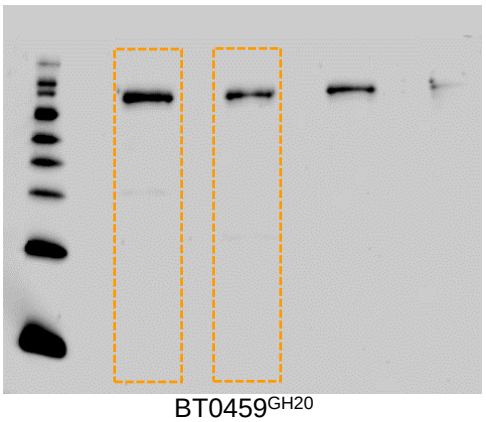


f

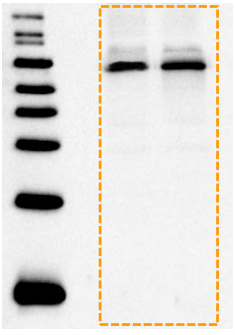
Western blots used to make Supplementary Figure 7c



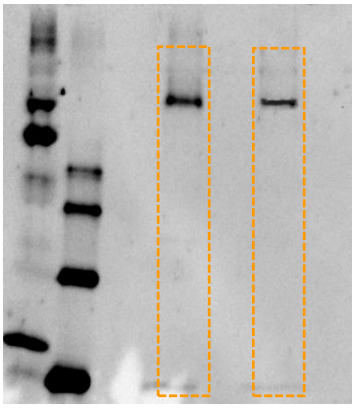
BT0456^{GH20}



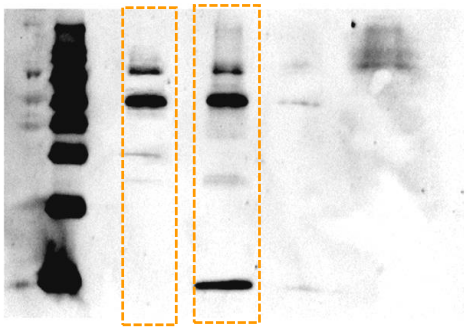
BT0459^{GH20}



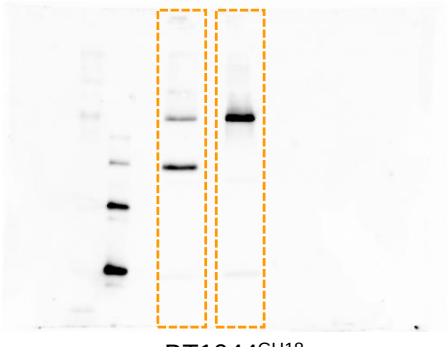
BT0460^{GH20}



BT0506^{GH20}



BT0455^{GH33}

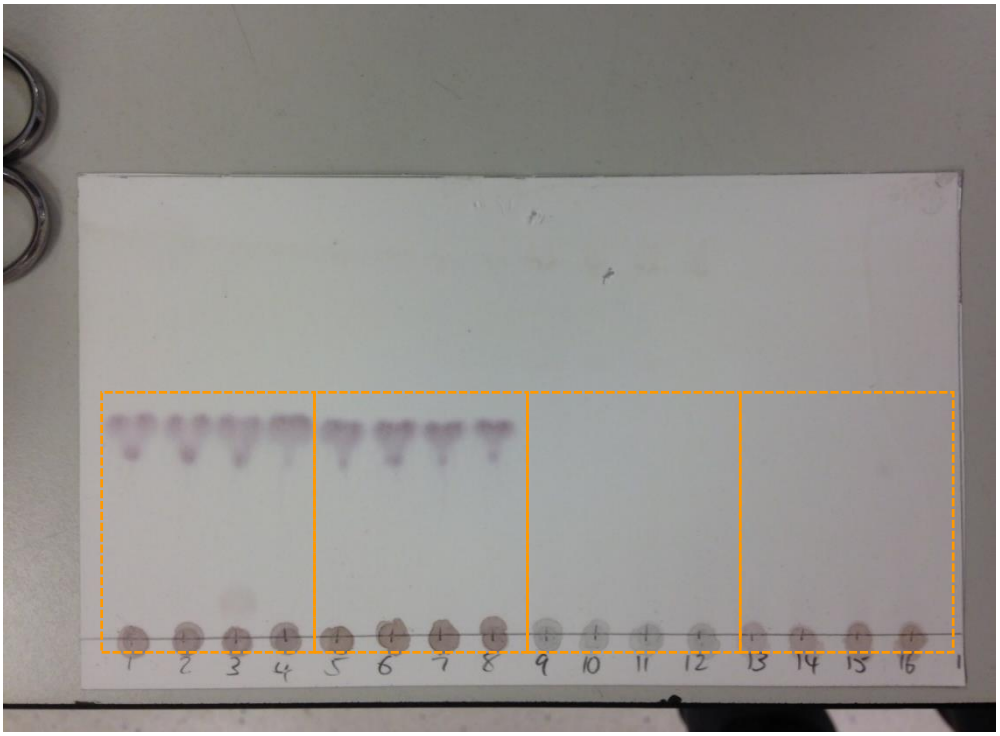


BT1044^{GH18}

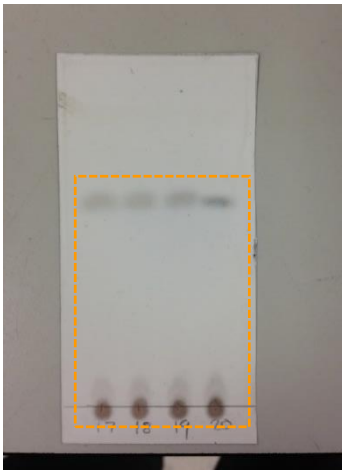
g

TLCs used to make Supplementary Figure 9a

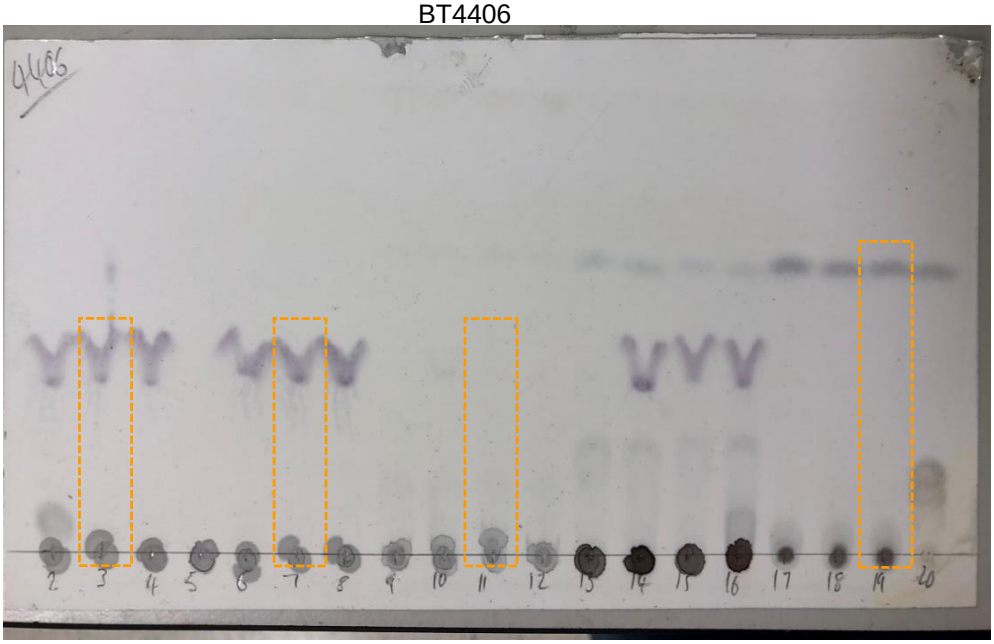
BT1038, BT1044 and BT1048



BT1038, BT1044 and BT1048



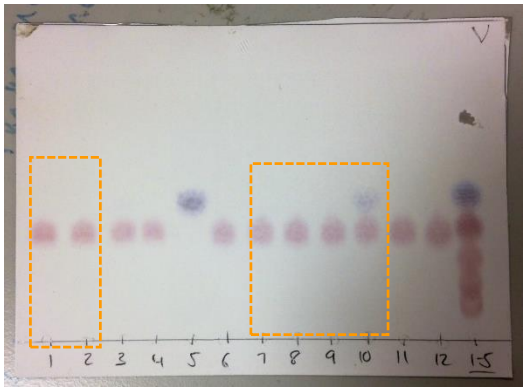
BT4406



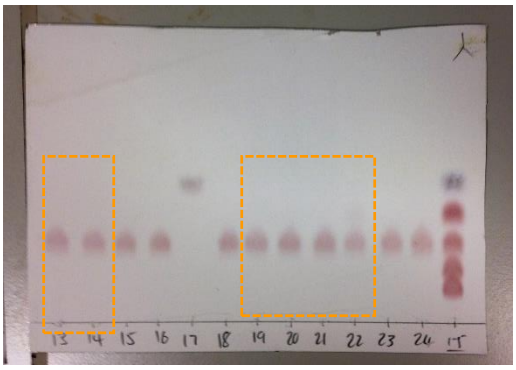
h

TLCs used to make Supplementary Figure 9c and 17b

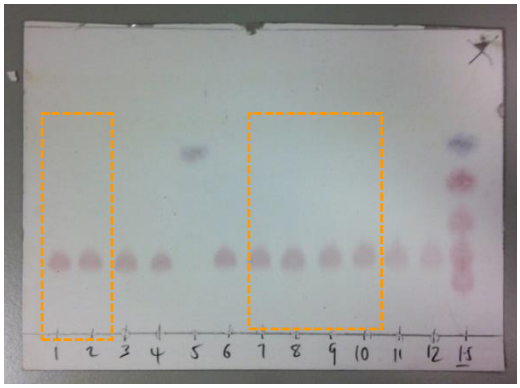
Chitobiose



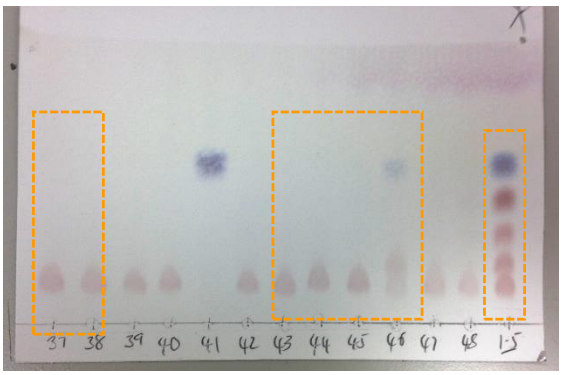
Chitotriose



Chitotetraose

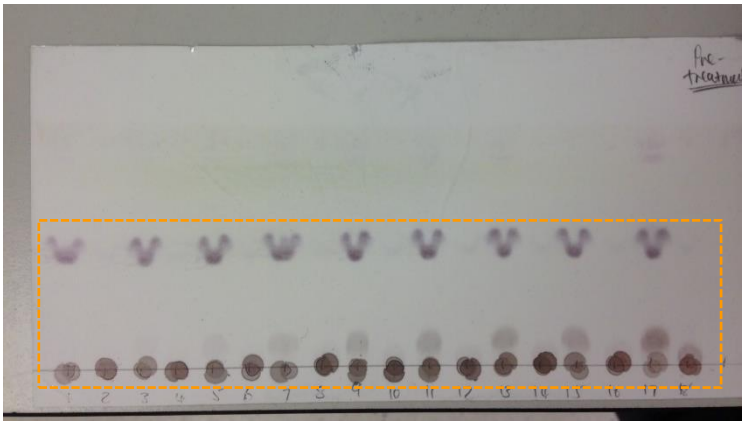


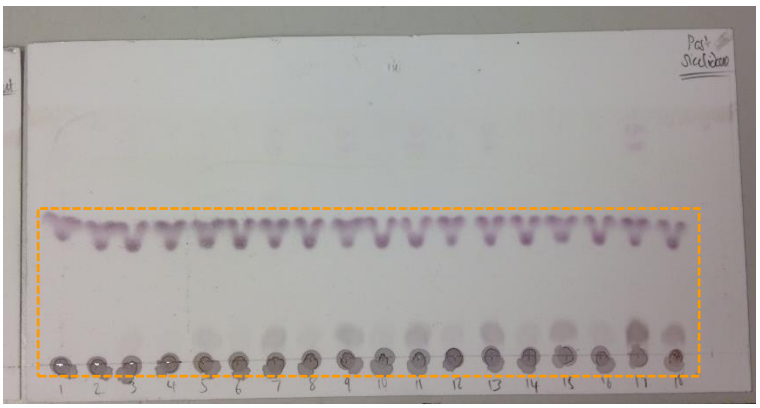
Chitopentaose



i

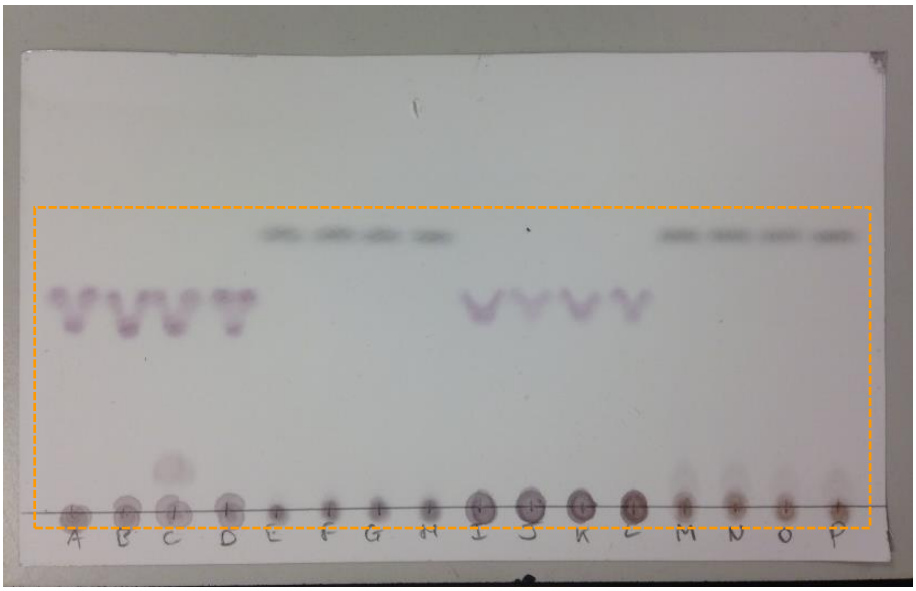
TLCs used to make Supplementary Figure 9d





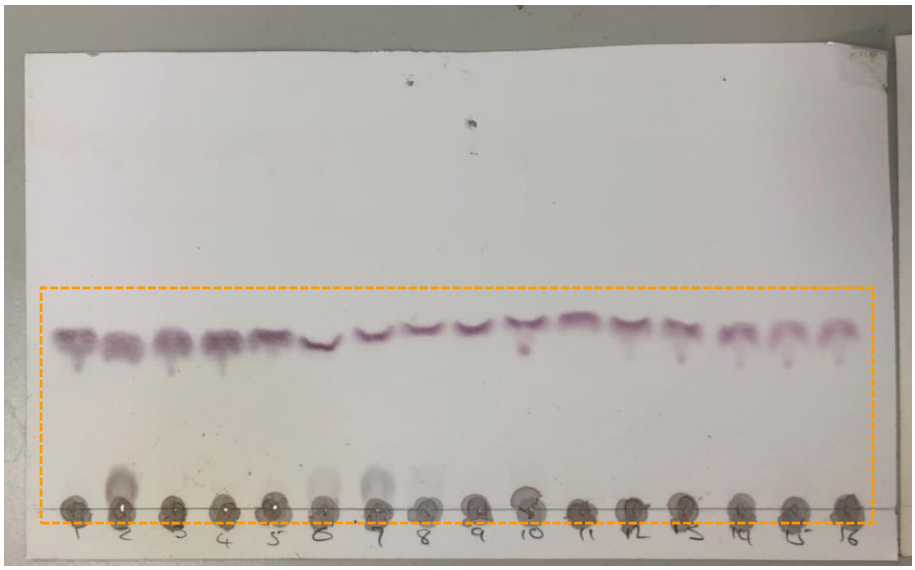
j

TLCs used to make Supplementary Figure 12b



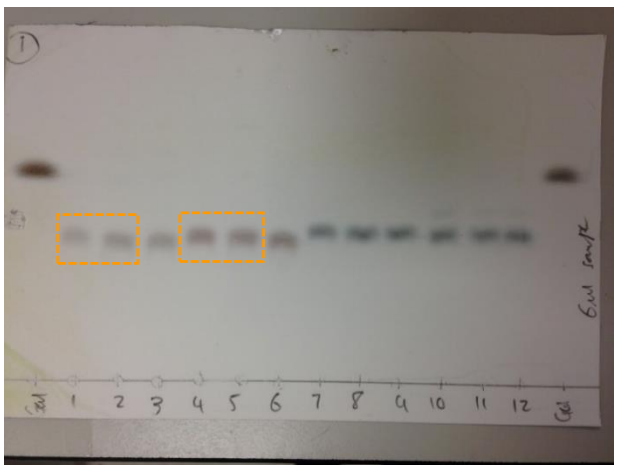
k

TLCs used to make Supplementary Figure 12d



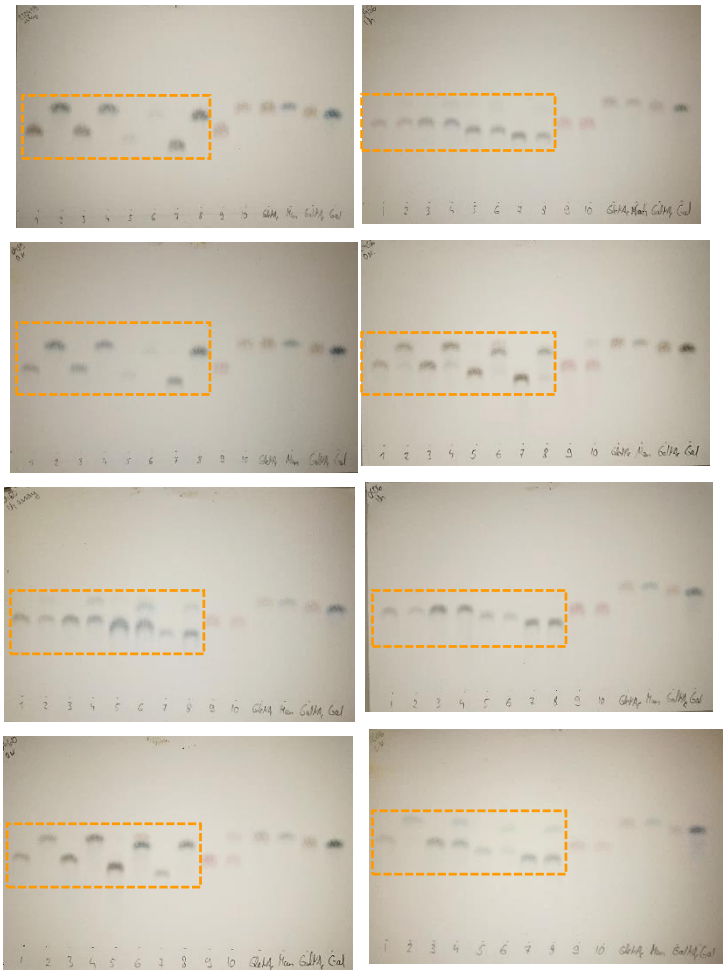
l

TLCs used to make Supplementary Figure 11k



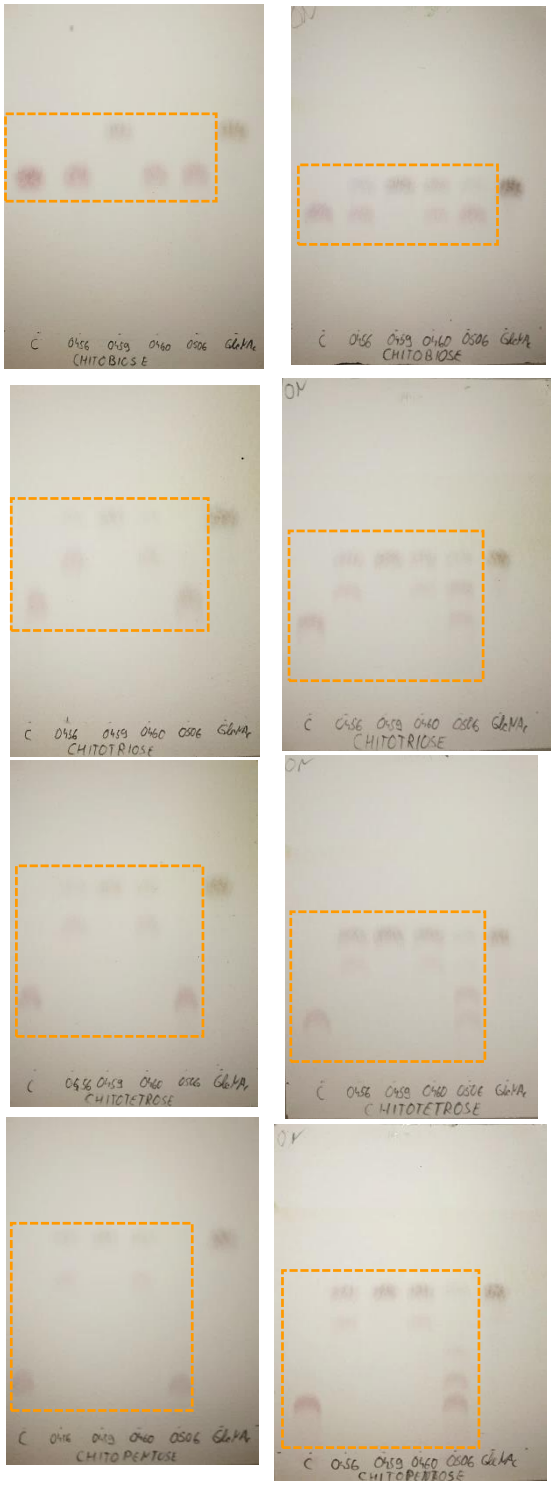
m

TLCs used to make Supplementary Figure 22a



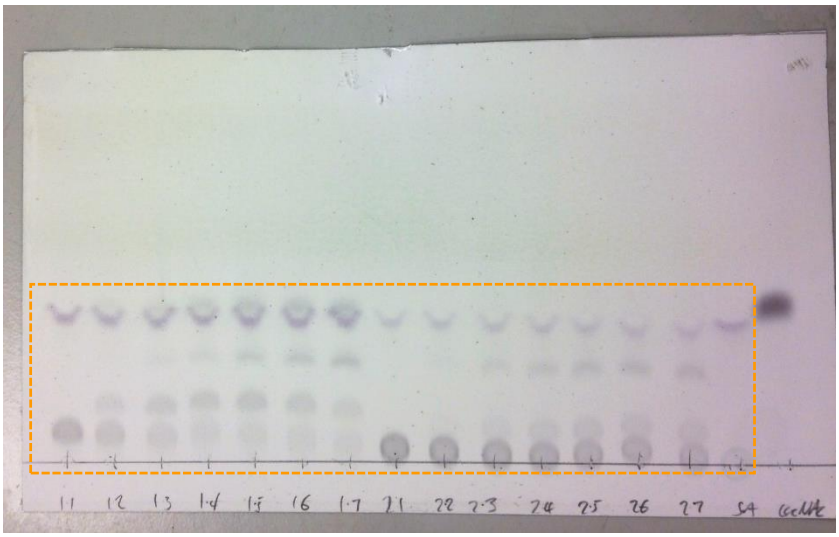
n

TLCs used to make Supplementary Figure 22b



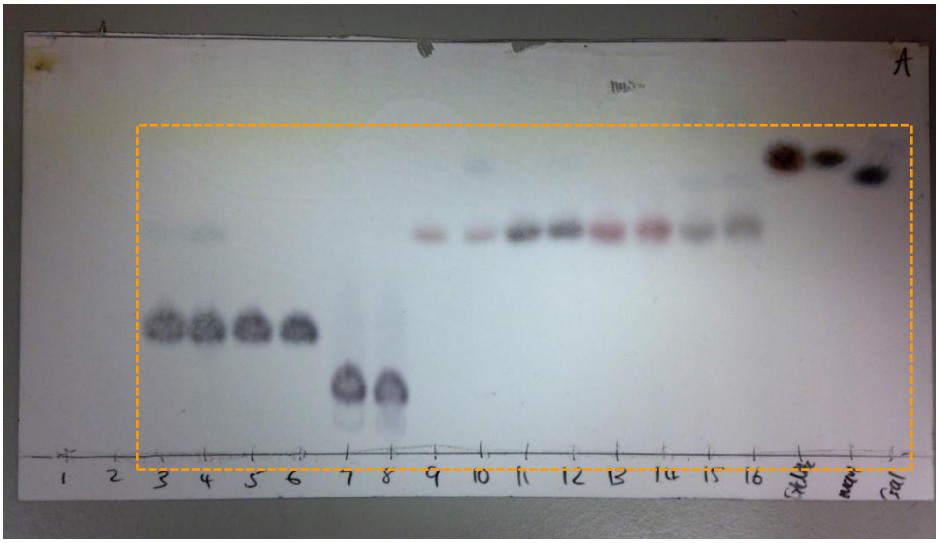
o

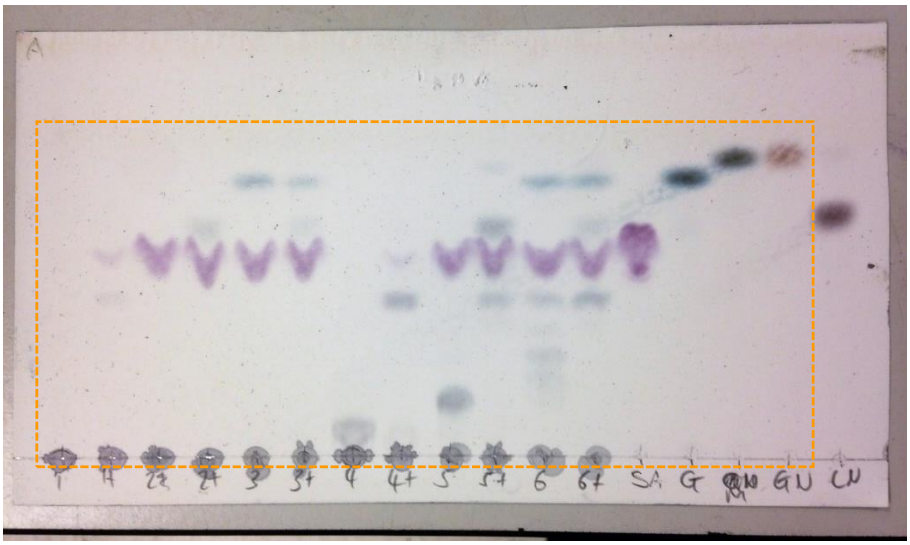
TLCs used to make Supplementary Figure 18



p

TLCs used to make Supplementary Figure 17a and c





References

1 Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99, doi:10.1038/nature12503 (2013).

2 Tailford, L. E. *et al.* Mannose foraging by Bacteroides thetaiotaomicron: structure and specificity of the beta-mannosidase, BtMan2A. *The Journal of biological chemistry* **282**, 11291-11299, doi:10.1074/jbc.M610964200 (2007).

3 Nihira, T. *et al.* Discovery of beta-1,4-D-mannosyl-N-acetyl-D-glucosamine phosphorylase involved in the metabolism of N-glycans. *The Journal of biological chemistry* **288**, 27366-27374, doi:10.1074/jbc.M113.469080 (2013).

4 Park, K. H. *et al.* Structural and biochemical characterization of the broad substrate specificity of Bacteroides thetaiotaomicron commensal sialidase. *Biochimica et biophysica acta* **1834**, 1510-1519, doi:10.1016/j.bbapap.2013.04.028 (2013).

5 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, doi:10.1016/j.chom.2008.09.007 (2008).

6 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS biology* **9**, e1001221, doi:10.1371/journal.pbio.1001221 (2011).

7 Chen, J. Y. *et al.* Concerted mass spectrometry-based glycomic approach for precision mapping of sulfo sialylated N-glycans on human peripheral blood mononuclear cells and lymphocytes. *Glycobiology* **28**, 9-20, doi:10.1093/glycob/cwx091 (2018).

8 Eckmair, B., Jin, C., Abed-Navandi, D. & Paschinger, K. Multistep Fractionation and Mass Spectrometry Reveal Zwitterionic and Anionic Modifications of the N- and O-glycans of a Marine Snail. *Molecular & cellular proteomics : MCP* **15**, 573-597, doi:10.1074/mcp.M115.051573 (2016).

9 Peterson, D. A., McNulty, N. P., Guruge, J. L. & Gordon, J. I. IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell host & microbe* **2**, 328-339, doi:10.1016/j.chom.2007.09.013 (2007).

10 Renzi, F. *et al.* The N-glycan glycoprotein deglycosylation complex (Gpd) from Capnocytophaga canimorsus deglycosylates human IgG. *PLoS pathogens* **7**, e1002118, doi:10.1371/journal.ppat.1002118 (2011).

11 Brigham, C. *et al.* Sialic acid (N-acetyl neuraminic acid) utilization by Bacteroides fragilis requires a novel N-acetyl mannosamine epimerase. *Journal of bacteriology* **191**, 3629-3638, doi:10.1128/jb.00811-08 (2009).

12 Sebaihia, M. *et al.* The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics* **38**, 779-786, doi:10.1038/ng1830 (2006).

13 Huang, K. *et al.* Biochemical characterisation of the neuraminidase pool of the human gut symbiont Akkermansia muciniphila. *Carbohydrate research* **415**, 60-65, doi:10.1016/j.carres.2015.08.001 (2015).

14 Barry, G. T. Detection of sialic acid in various Escherichia coli strains and in other species of bacteria. *Nature* **183**, 117-118 (1959).

15 Wessels, M. R., Rubens, C. E., Benedi, V. J. & Kasper, D. L. Definition of a bacterial virulence factor: sialylation of the group B streptococcal capsule. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 8983-8987 (1989).

16 Bhattacharjee, A. K., Jennings, H. J., Kenny, C. P., Martin, A. & Smith, I. C. Structural determination of the sialic acid polysaccharide antigens of Neisseria meningitidis serogroups B and C with carbon 13 nuclear magnetic resonance. *The Journal of biological chemistry* **250**, 1926-1932 (1975).

17 Mandrell, R. E. & Apicella, M. A. Lipo-oligosaccharides (LOS) of mucosal pathogens: molecular mimicry and host-modification of LOS. *Immunobiology* **187**, 382-402, doi:10.1016/s0171-2985(11)80352-9 (1993).

18 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. *The Biochemical journal* **472**, 157-167, doi:10.1042/bj20150388 (2015).
19 Shaikh, F. A., Lammerts van Bueren, A., Davies, G. J. & Withers, S. G. Identifying the catalytic acid/base in GH29 alpha-L-fucosidase subfamilies. *Biochemistry* **52**, 5857-5864, doi:10.1021/bi400183q (2013).
20 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).
21 Van Roey, P., Rao, V., Plummer, T. H., Jr. & Tarentino, A. L. Crystal structure of endo-beta-N-acetylglucosaminidase F1, an alpha/beta-barrel enzyme adapted for a complex substrate. *Biochemistry* **33**, 13989-13996 (1994).
22 Trimble, R. B. & Tarentino, A. L. Identification of distinct endoglycosidase (endo) activities in Flavobacterium meningosepticum: endo F1, endo F2, and endo F3. Endo F1 and endo H hydrolyze only high mannose and hybrid glycans. *The Journal of biological chemistry* **266**, 1646-1651 (1991).
23 Stals, I. *et al.* High resolution crystal structure of the endo-N-Acetyl-beta-D-glucosaminidase responsible for the deglycosylation of Hypocrea jecorina cellulases. *PloS one* **7**, e40854, doi:10.1371/journal.pone.0040854 (2012).
24 Pluvinage, B. *et al.* Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. *Structure (London, England : 1993)* **19**, 1603-1614, doi:10.1016/j.str.2011.08.011 (2011).
25 Gutternigg, M. *et al.* Biosynthesis of truncated N-linked oligosaccharides results from non-orthologous hexosaminidase-mediated mechanisms in nematodes, plants, and insects. *The Journal of biological chemistry* **282**, 27825-27840, doi:10.1074/jbc.M704235200 (2007).
26 Leonard, R. *et al.* The Drosophila fused lobes gene encodes an N-acetylglucosaminidase involved in N-glycan processing. *The Journal of biological chemistry* **281**, 4867-4875, doi:10.1074/jbc.M511023200 (2006).
27 Prag, G. *et al.* Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. *Journal of molecular biology* **300**, 611-617, doi:10.1006/jmbi.2000.3906 (2000).
28 Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169, doi:10.1038/nature13995 (2015).
29 Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B. & Davies, G. J. Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chemistry & biology* **15**, 1058-1067, doi:10.1016/j.chembiol.2008.09.005 (2008).
30 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.e251, doi:10.1016/j.cell.2017.09.014 (2017).
31 Trimble, R. B. & Maley, F. The use of endo-beta-N-acetylglucosaminidase H in characterizing the structure and function of glycoproteins. *Biochemical and biophysical research communications* **78**, 935-944 (1977).
32 Rao, V., Guan, C. & Van Roey, P. Crystal structure of endo-beta-N-acetylglucosaminidase H at 1.9 A resolution: active-site geometry and substrate recognition. *Structure (London, England : 1993)* **3**, 449-457 (1995).
33 Plummer, T. H., Jr. & Tarentino, A. L. Purification of the oligosaccharide-cleaving enzymes of Flavobacterium meningosepticum. *Glycobiology* **1**, 257-263 (1991).
34 Byers, H. L., Tarelli, E., Homer, K. A. & Beighton, D. Sequential deglycosylation and utilization of the N-linked, complex-type glycans of human alpha1-acid glycoprotein mediates growth of Streptococcus oralis. *Glycobiology* **9**, 469-479 (1999).

35 Roberts, G. *et al.* Distribution of endo-beta-N-acetylglucosaminidase amongst enterococci. *Journal of medical microbiology* **50**, 620-626, doi:10.1099/0022-1317-50-7-620 (2001).
36 Collin, M. & Fischetti, V. A. A novel secreted endoglycosidase from Enterococcus faecalis with activity on human immunoglobulin G and ribonuclease B. *The Journal of biological chemistry* **279**, 22558-22570, doi:10.1074/jbc.M402156200 (2004).
37 Garbe, J. *et al.* EndoE from Enterococcus faecalis hydrolyzes the glycans of the biofilm inhibiting protein lactoferrin and mediates growth. *PloS one* **9**, e91035, doi:10.1371/journal.pone.0091035 (2014).
38 Roberts, G., Tarelli, E., Homer, K. A., Philpott-Howard, J. & Beighton, D. Production of an endo-beta-N-acetylglucosaminidase activity mediates growth of Enterococcus faecalis on a high-mannose-type glycoprotein. *Journal of bacteriology* **182**, 882-890 (2000).
39 Collin, M. & Olsen, A. EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG. *The EMBO journal* **20**, 3046-3055, doi:10.1093/emboj/20.12.3046 (2001).
40 Trastoy, B. *et al.* Crystal structure of Streptococcus pyogenes EndoS, an immunomodulatory endoglycosidase specific for human IgG antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6714-6719, doi:10.1073/pnas.1322908111 (2014).
41 Fujita, K. *et al.* Identification of amino acid residues essential for the substrate specificity of Flavobacterium sp. endo-beta-N-acetylglucosaminidase. *Bioscience, biotechnology, and biochemistry* **65**, 1542-1548, doi:10.1271/bbb.65.1542 (2001).
42 Stals, I. *et al.* Identification of a gene coding for a deglycosylating enzyme in Hypocrea jecorina. *FEMS microbiology letters* **303**, 9-17, doi:10.1111/j.1574-6968.2009.01849.x (2010).
43 Hamaguchi, T. *et al.* Purification, characterization and molecular cloning of a novel endo-β-N-acetylglucosaminidase from the basidiomycete, Flammulina velutipes. *Glycobiology* **20**, 420-432, doi:10.1093/glycob/cwp188 (2010).
44 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 microbiology letters* **325**, 123-129, doi:10.1111/j.1574-6968.2011.02419.x (2011).
45 Garrido, D. *et al.* Endo-beta-N-acetylglucosaminidases from infant gut-associated bifidobacteria release complex N-glycans from human milk glycoproteins. *Molecular & cellular proteomics : MCP* **11**, 775-785, doi:10.1074/mcp.M112.018119 (2012).
46 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. *The Biochemical journal* **455**, 107-118, doi:10.1042/bj20130126 (2013).
47 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. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 12901-12906, doi:10.1073/pnas.1407344111 (2014).
48 Tzelepis, G. D., Melin, P., Jensen, D. F., Stenlid, J. & Karlsson, M. Functional analysis of glycoside hydrolase family 18 and 20 genes in Neurospora crassa. *Fungal genetics and biology : FG & B* **49**, 717-730, doi:10.1016/j.fgb.2012.06.013 (2012).
49 Dupoiron, S. *et al.* The N-Glycan cluster from Xanthomonas campestris pv. campestris: a toolbox for sequential plant N-glycan processing. *The Journal of biological chemistry* **290**, 6022-6036, doi:10.1074/jbc.M114.624593 (2015).