HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2022 May 24.

Published in final edited form as:

Nature. 2021 October ; 598(7880): 332–337. doi:10.1038/s41586-021-03967-5.

A single sulfatase is required to access colonic mucin by a gut bacterium

Ana S. Luis[*,1,2], **Chunsheng Jin**[2], **Gabriel Vasconcelos Pereira**[1], **Robert W. P. Glowacki**[1], **Sadie Gugel**[1], **Shaleni Singh**[1], **Dominic P. Byrne**[3], **Nicholas Pudlo**[1], **James A London**[3], **Arnaud Baslé**[4], **Mark Reihill**[5], **Stefan Oscarson**[5], **Patrick A. Eyers**[3], **Mirjam Czjzek**[6], **Gurvan Michel**[6], **Tristan Barbeyron**[6], **Edwin A Yates**[3], **Gunnar C. Hansson**[2], **Niclas G. Karlsson**[2], **Alan Cartmell**[*,3], **Eric C. Martens**[*,1]

[1]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

[2]Department of Medical Biochemistry, Institute for Biomedicine, Sahlgrenska Academy, University of Gothenburg, Box 440, 405 30 Gothenburg, Sweden

[3]Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3BX, United Kingdom

[4]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, United Kingdom

[5]Centre for Synthesis and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.

[6]Sorbonne Université, Univ Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, Roscoff, Bretagne, France.

Summary

Humans have co-evolved with a dense community of microbial symbionts that inhabit the lower intestine. In the colon, secreted mucus creates a barrier that separates these microbes from the intestinal epithelium[1]. Some gut bacteria are able to utilize mucin glycoproteins, the main mucus component, as a nutrient source. However, it remains unclear which bacterial enzymes initiate degradation of the complex *O*-glycans found in mucins. In the colon, these glycans are heavily sulfated, but specific sulfatases that are active on colonic mucins have not been identified. Here we show that sulfatases are essential to the utilization of distal colonic mucin *O*-glycans by the human gut symbiont *Bacteroides thetaiotaomicron*. We characterized the activity of 12 different

[*]Correspondence to: emartens@umich.edu, Alan.Cartmell@liverpool.ac.uk, ana.luis@medkem.gu.se.

Author contributions

ASL, AC, and ECM designed experiments and wrote the manuscript.
ASL and AC cloned, expressed, purified sulfatases and performed the enzymatic assays.
AC, DPB, JAL and PAE carried out and analysed kinetic and binding experiments EY,MR, SO performed chemical syntheses
AC and AB performed structural biology experiments.
CJ, ASL GCH and NGK performed and interpreted analytical glycobiology experiments.
ASL, GP, RWPG, SG, SS and NAP performed bacterial growth experiments and analysed in vivo competition data.
MC, GM and TB performed sulfatase phylogenetic analyses.
All authors read and approved the manuscript.

Code availability statement

No new codes were developed or compiled in this study

Competing interests statement

The authors declare no competing interests.

sulfatases produced by this species showing that they are collectively active on all of the known sulfate linkages in colonic *O*-glycans. Crystal structures of 3 enzymes provide mechanistic insight into the molecular basis of substrate-specificity. Unexpectedly, we found that a single sulfatase is essential for utilization of sulfated *O*-glycans *in vitro* and also plays a major role *in vivo*. Our results provide insight into the mechanisms of mucin degradation by a prominent group of gut bacteria, an important process for both normal microbial gut colonization[2] and diseases such as inflammatory bowel disease[3].

The human gut microbiota (HGM) significantly impacts several aspects of intestinal health and disease, including inflammatory bowel disease (IBD)[4] and colorectal cancer (CRC)[5]. In the colon, secreted mucus creates a physical barrier that separates gut microbes from the intestinal epithelium[1] preventing close contact that can lead to inflammation and eventual CRC if this barrier is either experimentally eliminated[6,7] or has reduced glycosylation[8–11]. A major component of colonic mucus is mucin 2 (MUC2), a glycoprotein that contains up to 80% glycans by mass and more than 100 different glycan structures that are *O*-linked to serine or threonine residues[12]. Mucin glycosylation is variable along the gastrointestinal (GI) tract with a marked increase in sulfation in the colon, especially the distal colon[13]. In mucins, *O*-linked sulfate may be attached to the 6-hydroxyl of *N*-acetyl-D-glucosamine (6S-GlcNAc) and terminal D-galactose (Gal) sugars at hydroxyl positions 3-, 4- or 6- (3S-, 4S- and 6S-Gal, respectively)[13–15] (Fig. 1a). Sulfation often occurs as terminal caps that typically block enzymatic degradation of oligosaccharides. This requires members of the HGM to express appropriate carbohydrate sulfatases to remove these modifications and access *O*-glycan sugars, although enzymatic mechanisms to remove sugars with attached sulfate present an alternate strategy and has been demonstrated in a human gut *Bifidobacterium sp.*[16]. *Bacteroides thetaiotaomicron* (*Bt*) is a dominant member of the human gut microbiota that can utilize *O*-glycans as a sole nutrient source[17]. *Bt* requires active sulfatases for competitive colonization of the wild-type mouse gut[2] and to induce inflammation in genetically susceptible mice[3]. However, the specific sulfatases that mediate these effects remain unknown. Indeed, despite the critical roles of sulfatases in many biological processes, including several diseases[18], a significant knowledge gap exists regarding the roles of these enzymes.

We hypothesized that specific *Bt* sulfatases play essential roles in initiation of *O*-glycan degradation. To test this, we determined the specificities of 12 *Bt* sulfatases on either model glycan substrates or purified colonic mucin *O*-glycans. We solved the structures for 3 of these enzymes, revealing the basis of substrate specificity. Using molecular genetics, we assessed the contributions of these sulfatases to *Bt* fitness *in vitro* and *in vivo*, unexpectedly revealing that a single enzyme is essential for utilization of sulfated mucin *O*-glycans and plays a large role in competitive gut colonization.

Utilization of colonic mucins by HGM species

Several studies have identified HGM members that are able to utilize porcine gastric mucin *O*-glycans (gMO) as a sole carbon source[17,19,20]. However, this substrate does not adequately reflect the structural complexity of mucin *O*-glycans found in the colon,

especially those with increased sulfation that are lacking in gMO[13]. To identify HGM species that utilize sulfated colonic mucins, we measured the growth of 16 *Bacteroides* type strains, plus 3 *Phocaeicola* strains (previously classified as *Bacteroides*) and *Akkermansia muciniphil*a, on highly sulfated porcine colonic mucin oligosaccharides (cMO) extracted from the distal colon. We identified six strains that utilize cMO (Fig. 1b, Extended Data Fig. 1a). Interestingly, two known mucin-degraders, *A. muciniphila*[21] and *P. massiliensis*[22], grew on gMO but failed to utilize sulfated cMO (Extended Data Fig. 1b), highlighting the importance of employing colonic mucins as more relevant substrates. *Bt*, the bacterium with the highest number of sulfatases (28)[2], was one of the strains with the best growth on cMO (Fig. 1b), suggesting that some of these enzymes might play key roles in promoting this ability. Therefore, to understand the role of sulfatases in colonic mucin utilization by HGM bacteria we focused on biochemical and genetic characterization of the *Bt* enzymes. See Supplementary Discussion 1 for additional details of GI mucins and bacterial growth on them.

Substrate specificity of Bt sulfatases

Sulfatases are classified into four main families (S1 to S4) in the SulfAtlas database according to sequence similarity, catalytic mechanism and fold[23]. Family S1 is currently divided into 72 subfamilies (designated S1_X) and comprises sulfatases which operate via a hydrolytic mechanism and utilize a non-genetically coded formylglycine amino acid as its catalytic nucleophile. In *Bt* and other anaerobic bacteria, this residue is introduced co-translationally by the anaerobic sulfatase maturating enzyme (anSME)[2], which converts a serine or cysteine to formylglycine[24]. The *Bt* genome encodes 28 S1 sulfatases classified into twelve different subfamilies (Supplementary Table 1). Four *Bt* sulfatases have been previously characterized and all act on glycosaminoglycan (GAG) polysaccharides in the extracellular matrix[25–27]. Several of the uncharacterized S1 sulfatases are encoded within polysaccharide utilization loci (PULs) that are known to be upregulated *in vivo* or during growth on gMO[17] and encode other glycoside hydrolase enzymes potentially involved in degrading mucin *O*-glycans (Extended Data Fig. 2).

To understand the role of sulfatases in mucin degradation, we tested 23 of the remaining 24 uncharacterized sulfatases for activity against a panel of sulfated saccharides (Supplementary Table 2). These experiments identified activities for 12 sulfatases, revealing that *Bt* has the capacity to remove sulfate from all of the positions that it occurs in *O*-glycans (Fig. 1c, Extended Data Figs. 3, 4 and Supplementary Table 3). Among the active enzymes, 5 represent the first activities reported for their respective subfamilies. Two S1_20 members (BT1636 and BT1622) were determined to target 3S-Gal (Extended Data Fig. 3a,b), with BT1622 preferentially binding *N*-acetyl-D-galactosamine (GalNAc) (Extended Data Fig. 3b) and cleaving 3S-GalNAc (Extended Data Fig. 3c). This represents the first report of a bacterial sulfatase active on 3S-GalNAc, indicating that this sulfation could exist as a yet unidentified modification of host glycans. Two S1_16 enzymes (BT3796 and BT3057) cleave 4S-Gal/4S-GalNAc (Extended Data Fig. 4a). One S1_46 enzyme (BT1918) cleaves 3S-GlcNAc, using the *N*-acetyl group as an absolute specificity determinant (Extended Data Fig. 4b). Subsequently, we refer to these enzymes by their gene/locus tag number with the corresponding activity in superscript (*e.g.*, $BT1636^{3S\text{-}Gal}$).

We also identified sulfatases displaying new activities inside previously characterized subfamilies. These include three S1_15 enzymes (BT1624$^{6S\text{-}Gal/GalNAc}$, BT3109$^{6S\text{-}Gal/GalNAc}$ and BT4631$^{6S\text{-}Gal/GalNAc}$) that extend this family previously only known to include 6S-GalNAc sulfatases (Extended Data Fig. 3b and 4c). Two members of S1_4 were active on either 3S-Gal (BT4683$^{3S\text{-}Gal}$) (Extended Data Fig. 3a,b) or 6S-Gal (BT3487$^{6S\text{-}Gal}$) (Extended Data Fig. 4c). Finally, consistent with the activity previously described for S1_11 members, two enzymes were 6S-GlcNAc sulfatases (BT1628$^{6S\text{-}GlcNAc}$ and BT3177$^{6S\text{-}GlcNAc}$) (Extended Data Fig. 4d).

The activities of some sulfatases were inhibited by the presence of reducing end substitutions, such as fucose, or exhibited preferences for different glycosidic linkages between sugars (Extended Data Fig. 3a,b and Supplementary Table 3). This suggests that *Bt* sulfatases are part of more complex multi-enzyme pathways involving glycoside hydrolases and their genomic expansion may reflect adaptations to cleave sulfate that occurs in many different glycan-specific contexts in host glycans. See Supplementary Discussion 2 for additional details of sulfatase activity based on glycan context.

Bt sulfatase activity on cMO

We next tested the activity of *Bt* sulfatases on a cMO fraction that we determined to contain at least 131 different oligosaccharides (Supplementary Table 4). Only 4 of the 6 sulfatases tested displayed activity on cMOs (Fig. 2 and Supplementary Table 4). BT1628$^{6S\text{-}GlcNAc}$ and BT3177$^{6S\text{-}GlcNAc}$ removed 6-*O*-sulfate from all GlcNAc structures that presented this modification at the non-reducing end of *O*-glycans, confirming an exo-mode of action (Fig. 2 and Extended Data Fig. 5a). BT4683$^{3S\text{-}Gal}$ is active on a smaller subset of structures (Fig. 2a and Supplementary Table 4). We were only able to determine the structure of one of those glycans, unexpectedly revealing that this enzyme is endo-active on sialylated 3S-Gal (Fig. 2b, Supplementary Fig. 1), which is consistent with structural data discussed below. After incubation with BT1636$^{3S\text{-}Gal}$, we detected 14 new oligosaccharides (Fig. 2a) and an overall increase of non-sulfated glycans (Extended Data Fig. 5a). Compared to the non-enzyme treated reference, 36 oligosaccharides could no longer be detected after incubation with BT1636$^{3S\text{-}Gal}$ (Fig. 2a and Supplementary Table 4). We determined the structures of 8 of these glycans and all present a terminal 3S-Gal (Fig. 2c). BT1636$^{3S\text{-}Gal}$ was active on 3'S-Gal-β1,3-GalNAc (core 1) and more complex sulfated glycans built around other core structures, indicating that it evolved to accommodate various linkages and substitutions found in *O*-glycans. Repetition of the cMO degradation experiments with a different batch of cMO supported the key enzymatic findings described above (Extended Data Fig. 5 and Supplementary Table 5). See Supplementary Discussion 3 for additional details of activity on cMO.

Structures of 3S-Gal/GalNAc sulfatases

To understand the molecular details of carbohydrate recognition by S1 sulfatases, we determined the crystal structures, with ligand, of the three different 3S-Gal sulfatases from two subfamilies. Consistent with previous structures[25,27], all 3 enzymes display a α/β/α topology with a C-terminal sub-domain, and the active site residues interacting

with the sulfate group are fully conserved (Extended data Fig. 6a,b). Consistent with *exo*-activity observed for both S1_20 enzymes (BT1636$^{3S\text{-}Gal}$, BT1622$^{3S\text{-}Gal/GalNAc}$) the substrates are buried in a deep pocket and only *O1* is solvent exposed (Fig. 3). The structure of BT1636$^{3S\text{-}Gal}$ and BT1622$^{3S\text{-}Gal/GalNAc}$ revealed that H177 and H176, respectively, coordinate with *O*4 of Gal (Fig. 3) and mutation to alanine had a major impact on enzyme activity (Supplementary Table 3), suggesting that this histidine is the major specificity determinant for galacto- configurations (presenting an axial *O*4) over gluco- configurations (equatorial *O*4). This essential residue is highly conserved (92%) within S1_20 sulfatases (Extended data Fig. 7a). BT1636$^{3S\text{-}Gal}$ also makes strong interactions with *O*2 of Gal via R353 and E334. In BT1622$^{3S\text{-}Gal/GalNAc}$, these residues are replaced by C357 and N334, amino acids that present shorter side chains that accommodate the *C*2-linked *N*-acetyl group found in GalNAc (Fig. 3). This is consistent with the ability of BT1622$^{3S\text{-}Gal/GalNAc}$ to cleave 3S-GalNAc and preferentially bind GalNAc over Gal, while BT1636$^{3S\text{-}Gal}$ only recognizes Gal (Extended Data Fig. 3b,c).

In the S1_4 enzyme BT4683$^{3S\text{-}Gal}$, the interaction with Gal is driven by the residues R72 and E335, conserved in 19% and 62% of analysed sequences, respectively (Extended Data Fig. 7b). These residues are spatially equivalent to R353 and E334 in BT1636$^{3S\text{-}Gal}$ and form hydrogen bonds with *O2* of D-Gal (Fig. 3). Disruption of either of these residues eliminates activity (Supplementary Table 3). In BT4683$^{3S\text{-}Gal}$, a sulfatase that does not have any affinity for monosaccharides, the active site is located in an open cleft (Fig. 3) that allows accommodation of additional substitutions on Gal (Extended Data Fig. 6c). This finding is consistent with the apparent endo-activity found using cMO. Together, these structures reveal the key specificity determinants in 3S-Gal/GalNAc sulfatases, highlighting that these enzymes have evolved to target sulfate groups in different contexts in which they are found in complex host glycans. This is especially true for BT1636$^{3S\text{-}Gal}$ which utilizes high affinity interactions with both *O*2 and *O*4 to drive enhanced activity to remove terminal 3S-Gal linkages in cMOs. See Supplementary Discussion 4 for additional details of sulfatase structural characterization and phylogeny.

Roles of sulfatases in Bt O-glycan utilization

Bt is able to utilize cMO as a sole carbon source (Fig. 1b), but the key enzymes involved in the degradation of these glycans remain unclear. Deletion of the gene encoding the only anaerobic sulfatase maturating enzyme (*anSME*) eliminates activation of all 28 S1 sulfatases[2] and as expected decreased the ability of *Bt* to grow efficiently on cMOs (Fig. 4a). Based on this, we generated a series of strains with compounded gene deletions in which one or several groups of sulfatases were eliminated based on their activity. Deletion of all 3S-Gal/GalNAc sulfatases resulted in a growth phenotype similar to Δ*anSME* (Extended Data Fig. 8a). Interestingly, we observed a similar growth defect when just BT1636$^{3S\text{-}Gal}$ was deleted, but not the other 3S-Gal sulfatases (Fig. 4a and Extended Data Fig. 8a), consistent with the prominent activity of the recombinant form of this enzyme on cMOs. In contrast, a strain with compounded deletions of eight other sulfatases besides BT1636 displayed a growth phenotype similar to wild-type (Fig. 4a), indicating that these enzymes are not essential for cMO utilization. However, a Δ*10X sulf* mutant, which included the deletion of BT1636$^{3S\text{-}Gal}$ and the two 4S-Gal/GalNAc sulfatases, showed a similar growth

defect as Δ*anSME* and Δ*bt1636*. Complementation of this and other loss of function mutants with only *bt1636*$^{3S\text{-}Gal}$ restored growth on cMO to levels similar to wild-type (Fig. 4a, Extended Data Fig. 8a). Cellular localization experiments revealed that BT1636$^{3S\text{-}Gal}$, predicted to be a periplasmic protein (Supplementary Table 6), is actually located at the cell surface (Fig. 4b).

To further investigate the role of BT1636$^{3S\text{-}Gal}$, we analysed the oligosaccharides present in the culture supernatant of the wild-type and Δ*bt1636*$^{3S\text{-}Gal}$ strains after growth on cMO. Consistent with a robust ability of *Bt* to degrade diverse colonic *O*-glycans, no oligosaccharides were detected in wild-type supernatant (Supplementary Table 7 and Extended Data Fig. 8b). Compared to the cMO used as substrate, the supernatant of Δ*bt1636*$^{3S\text{-}Gal}$ showed a 20-fold accumulation of terminal 3S-Gal capped glycans, suggesting they could not be degraded (Fig. 4c and Supplementary Table 7). These data, combined with the cell surface location of BT1636$^{3S\text{-}Gal}$, suggest that this critical sulfatase is required early in *O*-glycan catabolism by cleaving 3S-Gal from *O*-glycans prior to importing them into the periplasm, where these oligosaccharides will be degraded by additional enzymes and serve as cues for activating *O*-glycan PULs. Interestingly, all of the *Bacteroides* species able to utilize cMO (Fig. 1a) have homologues of BT1636$^{3S\text{-}Gal}$, suggesting that this activity plays a key role in mucin utilization by other HGM members (Supplementary Table 8 and Supplementary Fig. 2).

Finally, to investigate the requirement for specific sulfatases *in vivo*, we utilized gnotobiotic mice in which we competed individual mutants against the wild-type strain to evaluate their colonization fitness. It has been reported that mouse colonic Muc2 prominently displays 6S-GlcNAc modifications[28]. However, three mutant strains that either lacked both characterised 6S-GlcNAc sulfatases or these same two enzymes, plus loss of another putative 6S-GlcNAc sulfatase and all three 6S-Gal/GalNAc sulfatases competed equally with wild-type (Extended Data Fig. 8c), suggesting that neither of these two sulfatase activities is essential *in vivo*. An overall mild defect was observed with a mutant lacking all 3S-Gal/GalNAc sulfatases (Fig. 4d). The fitness defect was exacerbated by eliminating 3S-Gal/GalNAc and 6S-GlcNAc sulfatases together (Fig. 4d), suggesting that they synergise *in vivo*. Consistent with its prominent role *in vitro*, a mutant lacking just BT1636$^{3S\text{-}Gal}$ displayed a strong defect when competed with the wild-type strain (Fig. 4d), further suggesting that this enzyme plays an essential role in gut colonization by allowing *Bt* to access 3S-Gal *O*-glycans (see Supplementary Discussion 5–7 for additional details regarding growth, prediction of cellular localization of proteins and competition of mutants *in vivo*).

Conclusion

To degrade the complex *O*-glycans found in mucins some HGM bacteria have evolved complex arsenals of degradative enzymes which include sulfatases. Disarming all of the sulfatases in *Bt* via *anSME* deletion results in drastically reduced competitive colonization (Extended Data Fig. 8c)[2] and an inability to elicit colitis in an animal model of IBD[3]. While these findings support a critical role for active sulfatases in both normal colonization and inflammation, they do not provide insight into the complexity of catalytic events carried out by these enzymes. In this study, we reveal that *Bt* has a robust ability to grow on

highly sulfated mucin oligosaccharides and that it possesses sulfatases capable of removing sulfate groups in all contexts in which they are known to occur in mucin, including novel specificities. Surprisingly, we found that a single key sulfatase is disproportionately important for growth on colonic mucin *O*-glycans. The critical role of BT1636$^{3S\text{-}Gal}$ supports the conclusion that keystone steps exist in the complex pathway of mucin degradation, although given that mucin glycan structures may vary between mammalian hosts these critical steps may eventually need to be validated in humans for therapeutic translation. Nevertheless, establishment of these critical steps is a prerequisite to blocking this complex enzyme pathway and potentially inhibiting mucin-degrading activities in bacteria that contribute to diseases such as IBD.

Methods

Recombinant Protein Production

Genes were amplified by PCR using the appropriate primers and the amplified DNA cloned in pET28b using *NheI/XhoI* restriction sites or pETite (Expresso™ T7 cloning and expression system, Lucigen) generating constructs with either N- or C-terminal His_6 tags (Supplementary Table 9). The catalytic serine was mutated to cysteine since *Escherichia coli* is only able to convert cysteine to formylglycine. Recombinant genes were expressed in *Escherichia coli* strains BL21 (DE3) or TUNER (Novagen), containing the appropriate recombinant plasmid, and cultured to mid-exponential phase before induction with 1 mM [BL21(DE3)] or 0.2 mM (TUNER) of isopropyl β-D-1-thiogalactopyranoside; cells were cultured for another 16 h at 16°C and 180 rpm. Recombinant proteins were purified to >90% electrophoretic purity by immobilized metal ion affinity chromatography using a cobalt-based matrix (Talon, Clontech) and eluted with imidazole as described previously[25]. For the proteins selected for structural studies, another step of size exclusion chromatography was performed using a Superdex 16/60 S200 column (GE Healthcare), with 10 mM HEPES, pH 7.5, and 150 mM NaCl as the eluent, and they were judged to be ≥95% pure by SDS-PAGE. Protein concentrations were determined by measuring absorbance at 280 nm using the respective molar extinction coefficient. When necessary, proteins were then concentrated by centrifugaton using a molecular mass cutoff of 30 kDa. Since only 12 of the 23 recombinantly-expressed proteins showed activity, we cannot rule out that purification of enzymes from an anaerobic bacterium in oxic conditions or poor installation of catalytic formylglycine resulted in inactive enzymes. However, SDS-PAGE analysis suggested that most enzymes were produced in relatively pure and soluble forms (Supplementary Figure 5).

Site-Directed Mutagenesis

Site-directed mutagenesis was conducted using the PCR-based QuikChange kit (Stratagene) and conducted according to the manufacturer's instructions, using the appropriate plasmid as the template and primers (Supplementary Table 10). All mutations were confirmed by DNA sequencing.

Sources of purified carbohydrates

All carbohydrates were from Sigma, Carbosynth or Dextra Laboratories. All other chemical reagents were purchased from Sigma. The 3S-GalNAc was chemically synthesized as previously described[29].

Mucin purification

Gastric mucin oligosaccharides (gMO) were purified from commercial available porcine gastric mucins (type III, Sigma) as previously described[17]. Colonic mucins oligosaccharides (cMO) were purified from pig distal pig colons and rectum. Briefly, the tissue was open and the fecal contents were carefully removed. The mucosa was scrapped off and mucus was extracted by homogenizing the tissue in at least 5 times volume of extraction buffer (6 M guanidine chloride, 5 mM EDTA, 10 mM NaH_2PO_4, pH 6.5) and slow stirring at 4°C for 16 h. The solution was spun down at 15,000 rpm and 10°C for 30 min and supernatant was discharged. The pellets were resuspended in extraction buffer and the process was repeated until the supernatant was clear for at least two extractions. After the extraction the mucins were solubilized by reducing the disulfide bonds. The pellets were resuspended in fresh reduction buffer (6 M guanidine chloride, 0.1 M Tris, 5 mM EDTA, pH 8.0) containing 25 mM of 1,4-dithiothreitol and slowly stirred at 37°C for 5 h. After this incubation, 62.5 mM of iodoacetamide were added and the solution was stirred slowly in the dark at room temperature for 16 h. The solution was centrifuged at 10,000 rpm at 4°C for 30 min and the supernatant containing the solubilized mucins was extensively dialysed into water. Samples were dissolved into 100 mM Tris-HCl pH 8.0 containing 1 mg/ml of trypsin and incubated slowly stirring at 37°C for 16 h. The glycans were β-eliminated by adding 0.1 M NaOH and 1 M $NaBH_4$ and incubate the solution at 65°C for 18 h. After cooling the solution to room temperature, the pH was adjusted to 7.0 with concentrated HCl and extensively dialysed in water. The released porcine colonic mucin glycans were recovered by lyophilization the solution until completely dry and used in further experiments.

HPLC and TLC sulfatase enzymatic assays

The sulfatase activity screen against commercially available sulfated oligosaccharides (Supplementary Table 2) was performed with 1 μM of recombinant enzyme and 1 mM of substrate in 10 mM MES pH6.5 with 5 mM CaCl2 for 16h at 37°C (pH optima were determined for each enzyme, Supplementary Figure 6, and this pH is compatible with all enzymes). Sulfated *N*-acetyl-D-lactosamine and lacto-*N*-biose were generated by incubating the respective sulfated Lewis antigens with 1 μM of α−1,3/1,4-fucosidase BT1625[30] in the same conditions. Reactions were analysed by thin layer chromatography (TLC). Briefly, 2 μL of each sample was spotted onto silica plates and resolved in butanol:acetic acid:water (2:1:1) running buffer. The TLC plates were dried, and the sugars were visualized using diphenylamine stain (1 ml of 37.5% HCl, 2 ml of aniline, 10 ml of 85% H3PO3, 100 ml of ethyl acetate and 2 g diphenylamine) and heated at 100°C for 20 min. When relevant, the enzymatic activity was confirmed by high-performance anionic exchange chromatography (HPAEC) with pulsed amperometric detection using standard methodology. The sugars (reaction substrate/products) were bound to a Dionex CarboPac P100 column and eluted with an initial isocratic flow of 10 mM NaOH during 20 min then a gradient of 10–100 mM

of NaOH for 20 min at a flow rate of 1.0 ml min^{-1}. The reaction products were identified using the appropriated standards. All experiments were performed in triplicate.

Liquid Chromatograph-Electrospray Ionization Tandem Mass Spectrometry

Enzymatic reactions of sulfatases in colonic mucin oligosaccharides and culture supernatant were desalted with graphitized carbon[31]. Reactions with sulfated defined saccharides were reduced and desalted. Briefly, reactions were dried in Speed vac, reconstituted in 20 μL of 50 mM NaOH and 500 mM $NaBH_4$ and incubated at 50°C for 3 h. Reactions were cooled on ice, neutralized with 1 μL of glacial acetic acid and desalted using a cation exchange column containing AG®50W-X8 resin. All cleaned and desalted reactions were reconstituted in water before analysis by liquid chromatograph-electrospray ionization tandem mass spectrometry (LC-ESI/MS). The oligosaccharides were separated on a column (10 cm × 250 μm) packed in-house with 5 μm porous graphite particles (Hypercarb, Thermo-Hypersil, Runcorn, UK). The oligosaccharides were injected on to the column and eluted with a 0–40 % acetonitrile gradient in 10 mM ammonium bicarbonate over 46 min at a flow rate of 10 μl/min. A 40 cm × 50 μm i.d. fused silica capillary was used as transfer line to the ion source. Samples were analyzed in negative ion mode on a LTQ linear ion trap mass spectrometer (Thermo Electron, San José, CA), with an IonMax standard ESI source equipped with a stainless-steel needle kept at −3.5 kV. Compressed air was used as nebulizer gas. The heated capillary was kept at 300°C, and the capillary voltage was −33 kV. Full scan (*m/z* 380–2,000, two microscan, maximum 100 ms, target value of 30,000) was performed, followed by data dependent MS^2 scans (two microscans, maximum 100 ms, target value of 10,000) with normalized collision energy of 35%, isolation window of 2.5 units, activation q=0.25 and activation time 30 ms). The threshold for MS^2 was set to 300 counts. Data acquisition and processing were conducted with Xcalibur software (Version 2.0.7). Glycans were identified from their MS/MS spectra by manual annotation and validated by available structures stored in Unicarb-DB database (2020–01 version)[32]. *O*-Glycan structural characterization was based on diagnostic fragment ions[33].The schematic glycosidic or cross-ring cleavages were assigned according to the Domon and Costello nomenclacture[34]. For comparison of glycan abundance between samples, the individual glycan structures were quantified relative to the total content by integration of the extracted ion chromatogram peak area using Progenesis QI. The area under the curve (AUC) of each structure was normalized to the total AUC and expressed as a percentage. The LC-MS/MS raw files and annotated structures are submitted to the Glycopost and Unicarb-DB database, respectively.

Microfuidic-based enzymatic desulfation assays

Sulfated carbohydrates were labelled at their reducing end with BODIPY which has a maximal emission absorbance of ~503 nm, which can be detected by the EZ Reader via LED-induced fluorescence. Non-radioactive mobility shift carbohydrate sulfation assays were optimised in solution with a 12-sipper chip coated with CR8 reagent using a PerkinElmer EZ Reader II system using EDTA-based separation buffer. This approach allows real-time kinetic evaluation of substrate de-sulfation[35]. Pressure and voltage settings were adjusted manually (1.8 psi, upstream voltage:2250 V, downstream voltage: 500 V) to afford optimal separation of the sulfated product and unsulfated substrate with a sample (sip)

time of 0.2 s, and total assay times appropriate for the experiment. Individual de-sulfation assays were carried out at 28°C after assembly in a 384-well plate in a final volume of 80 µl in the presence of substrate concentrations between 0.5 and 20 µM with 100 mM Bis-Tris-Propane, 150 mM NaCl, 0.02% (v/v) Brij-35 and 5 mM $CaCl_2$. The degree of de-sulfation was calculated by peak integration using EZ Reader software, which measures the sulfated carbohydrate : unsulfated carbohydrate ratio at each individual time-point. The activity of sulfatase enzymes was quantified in 'kinetic mode' by monitoring the amount of unsulfated glycan generated over the assay time, relative to control assay with no enzyme; with sulfate loss limited to ~20% to prevent loss of substrate and to ensure assay linearity. k_{cat}/K_M values, using the equation $V_0=(k_{cat}/K_M)[E][S]$, were determined by linear regression analysis with GraphPad Prism v8.3.0 software. Substrate concentrations were varied to ensure assay linearity, and substrate concentrations present were significantly $<K_M$.

NMR desulfation assays

NMR experiments, monitoring the de-sulfation of 6S-D-galactose and 6S-*N*-acetyl-D-galactosamine, were conducted in D_2O with 50 mM sodium phosphate, pH 7.0, supplemented with 150 mM NaCl at 25°C on a 800MHz Bruker Avance III spectrometer equipped with a TCI CryoProbe and a 600MHz Bruker Avance II+ spectrometer, also fitted with a TCI CryoProbe. 1D and 2D proton and TOCSY spectra (mixing time 80 ms) were measured using standard pulse sequences provided by the manufacturer. Spectra were processed and analysed using TopSpin 3.4A and TopSpin 4.0 software (Bruker). Galactose integrals were recorded directly for the $C(6)H_2$-OH peak within the region 3.694 to 3.721ppm, referenced to the combined C(2) peaks of D-galactose and 6S-D-galactose with in the region 3.415 to 3.475ppm. Similarly, 6S-*N*-acetyl-D-galactosamine integrals were recorded directly for the $C(6)H_2$-OH peak within the region 3.674 to 3.747ppm, referenced to the combined C(4) peaks for *N*-acetyl-D-galactosamine and 6S-*N*-acetyl-D-galactosamine in the region 3.925 to 3.968ppm.

Differential scanning fluorimetry

Thermal shift/stability assays (TSAs) were performed using a StepOnePlus Real-Time PCR machine (LifeTechnologies) and SYPRO-Orange dye (emission maximum 570 nm, Invitrogen) as previously described[36] with thermal ramping between 20 and 95°C in 0.3°C step intervals per data point to induce denaturation in the presence or absence of various carbohydrates as appropriate to the sulfatase being analysed. The melting temperature (Tm) corresponding to the midpoint for the protein unfolding transition was calculated by fitting the sigmoidal melt curve to the Boltzmann equation using GraphPad Prism v8.3.0, with R^2 values of >0.99. Data points after the fluorescence intensity maximum were excluded from the fitting. Changes in the unfolding transition temperature compared with the control curve (ΔT_m) were calculated for each ligand. A positive ΔT_m value indicates that the ligand stabilises the protein from thermal denaturation, and confirms binding to the protein. All TSA experiments were conducted using a final protein concentration of 5µM in 100 mM Bis-Tris-Propane (BTP), pH 7.0, and 150 mM NaCl supplemented with the appropriate ligand. When BT1622$^{3S\text{-}Gal/GalNAc}$ and BT1636$^{3S\text{-}Gal}$ were assessed against 3'S-LacNAc and 3'S-LNB 100 mM Hepes (pH 7.0) was employed instead of BTP, although no difference in the T_m value of the proteins was observed. Three independent assays were performed for

each protein and protein ligand combination (Supplementary table 11). Separate DSF assays were conducted to validate the structural integrity of selected inactive sulfatase mutants (Supplementary Figure 7).

Glycan labelling

Sulfated saccharide samples were labelled according to a modification of the method previously described reporting the formation of *N*-glycosyl amines for 4,6-O-benzilidene protected D-gluopyranose monosaccharides with aromatic amines[37]. Briefly, the lyophilised sugar (1 mg) was dissolved in 0.50 ml anhydrous methanol in a 1.5 ml screw-top PTFE microcentrifuge tube. 0.1 mg, BODIPY-FL hydrazide (4,4–difluoro-5,7–dimethyl-4-bora-3a,4a–diaza-s-indacene-3-propionic acid hydrazide, $\lambda_{ex./em.\ 505/513}$, extinction coefficient 80,000 M^{-1} cm^{-1}) was added and the mixture vortexed (1 min), then incubated in darkness at 65°C for 24 h. The products were then cooled and a portion purified by TLC on silica coated aluminium plates and developed with methanol or 1:1 v/v ethyl acetate/methanol to provide R_f values suitable to allow separation of unreacted label from labelled glycan product. The unreacted BODIPY-FL hydrazide label (orange on the TLC plate) was identified by reference to a lane containing the starting material, allowing differentiation from the putative labelled product (also orange). This latter band was scraped from the plates and extracted in fresh methanol (2 × 0.5 ml), spun for 3 min at 13,000 × *g* and the supernatant was recovered and dried (rotary evaporator) to recover the fluorescent-coloured product (bright green when dissolved in aqueous solution), which was then employed in subsequent experiments.

Crystallization of carbohydrate sulfatases

After purification, all proteins were carried forward in the same eluent as used for the size exclusion chromatography (see Recombinant Protein Production). Sparse matrix screens were set up in 96-well sitting drop TTP Labtech plates (400-nL drops) at 20°C. Initial hits crystals for all proteins were obtained between 20 and 35 mg/ml. For BT1622$^{3S\text{-}Gal/GalNAc}$ and BT1636$^{3S\text{-}Gal}$ wildtype *Bt* variants were used, having a Ser at the catalytic formylglycine position, whilst for for BT4683$^{3S\text{-}Gal}$ the S73C mutant was used. BT1622$^{3S\text{-}Gal/GalNAc}$ with 20 mM LNB or 100 mM GalNAc crystallised in 20% PEG 3350 and 0.2 M sodium citrate tribasic dihydrate. BT1636$^{3S\text{-}Gal}$ with 20 mM LacNAc, or 20 mM 3'S-Lewisa, crystallized in 40% MPD, 5 % PEG 8000, and 0.1 M sodium cacodylate pH 6.5. for BT4683$^{3S\text{-}Gal}$ with 20 mM LacNAc crystallised in 20% PEG 3350, 0.2 M sodium iodide and BTP pH 8.5. All crystals were cryoprotected with the addition of the ligand they were crystallised with plus 20% PEG 400 and 20% glycerol was used as the cryoprotectant for BT4683$^{3S\text{-}Gal}$ and BT1622$^{3S\text{-}Gal/GalNAc}$, respectively. No cryoprotectant was added to BT1636$^{3S\text{-}Gal}$ crystals. Data were collected at Diamond Light Source (Oxford) on beamlines I03, I04, I04–1 and I24 at 100 K, except for BT1636$^{3S\text{-}Gal}$ with 3'S-Lewis-a which was collected 'in house' on a MetalJet D2 X-ray source at Newcastle University. The data were integrated with XDS[38], or Xia2[39] 3di or 3dii and scaled with Aimless v0.7.7[40]. Five percent of observations were randomly selected for the R_{free} set. The phase problem for BT1636$^{3S\text{-}Gal}$ and BT4683$^{3S\text{-}Gal}$ was solved by molecular replacement using the automated molecular replacement server Balbes[41]; for BT1636$^{3S\text{-}Gal}$ Balbes used the PDB 1FSU with 28 % identity, whilst for BT4683$^{3S\text{-}Gal}$ Balbes used the PDB 1HDH with 34 % identity.

The phase problem for BT1622$^{3S\text{-}Gal/GalNAc}$ was initially solved using Molrep v11.7.03[42] and BT1636$^{3S\text{-}Gal}$ (44% identity) as the search model. This gave a partial solution, which could not be fully solved due to twinning. An acceptable model of BT1622$^{3S\text{-}Gal/GalNAc}$ was constructed to be used to better solve the phase problem and the molecular replacement was re-performed. Models underwent recursive cycles of model building in Coot v0.9.5[43] and refinement cycles in Refmac v5.8.0267[44]. Bespoke ligands were generated using JLigand (in CCP4 v7.1.013)[45]. The protein models were validated using Coot v0.9.5[43] and MolProbity v4.5.1[46], whilst carbohydrate structures were validated using privateer MKIII[47] (Supplementary Table 12). Structural figures were made using Pymol v2.2.2, simulated annealing composite omit maps were generated with Phenix v1.19.2[48]; all other programs used were from the CCP4 v7.1.013[49] and CCP4i2 v7.1.013 suite[50]. The data processing and refinement statistics are reported in Supplementary Table 13.

Circular dichroism of purified proteins

Circular dichroism (CD) spectra were recorded using a J-1100 Jasco CD spectrometer, Spectral Manager II software and a 0.2 mm path length quartz cuvette (Hellma, USA) scanning at 100 nm.min-1 with 1 nm resolution throughout the wavelength range 190 – 260 nm. Reported spectra were the mean of 9 independent scans following calibration against a buffer baseline. Collected data were analysed with Spectral Manager II software prior to processing with GraphPad Prism v8.3.0. Secondary structural prediction was calculated through the BeStSel analysis server[51]. For all samples, we used 10 μM of protein and the buffer was 20 mM sodium phosphate buffer at pH 7.0. See Supplemental Figure 6.

Anaerobic bacterial culture and genetic manipulation

All strains were anaerobic grown at 37 °C in a chamber (10% H_2, 5% CO_2, and 85% N_2; Coy Manufacturing, Grass Lake, MI). *Bacteroides* type strains were culture in either tryptone-yeast extract-glucose medium (TYG), brain heart infusion medium or minimal medium (MM)[17] containing an appropriate carbon source. *Bacteroides massilliensis* and *Akkermansia muciniphila* were cultured as described before[22,52]. *Bt* strains containing specific gene deletions or inactivated versions of enzymes (BT1636$^{3S\text{-}Gal}$ S77A) were made by counterselectable allelic exchange as previously described[53]. Complemention of deletion strains was performed using pNBU2 vector as previously described[53], containing a constitutive promotor used previously[54]. All primers used to generate the mutants and complementation are listed in Supplementary Table 14. Growth of the WT and mutants was measured on an automated plate reader (Biotek) by increase in absorbance at 600 nm in 96-well plates containing 200 μl of minimal media mixed with the respective filter-sterilised (monosaccharide and gMO) or autoclave-sterilised cMO as described before[52]. To achieve consistent growth, all carbon sources were used at 5 mg/ml with exception of gMO that was added in a final concentration of 10 mg/ml. Data was collected and analyzed in Gen5 v1.08 (Biotek) or Prism v8.3.0. All growth curves presented are averages and s.e.m of three technical replicates.

Immunolabelling of BT1636 in *Bt* cell surface

To empirically test cellular location of BT1636 empirically using fluorescence microscopy, *Bt* cells (Wild type (Δ*tdk*) and Δ*bt1636*$^{3S\text{-}Gal}$) were grown to early exponential phase

(Abs_{600nm} 0.25–0.35) in rich TYG medium. One ml of the cultures was collected, centrifuged at 13,000 × *g*, and subsequently washed three times in MM with no carbon source. *Bt* cells incubated with cMO for four hours and fixed in 4.5% formalin overnight at 4°C with gentle rocking. Cells were stained with a polyclonal antibody raised in rabbit against purified recombinant BT1636 ($BT1636^{Ab}$, Cocalico Biologicals) and detected with an Alexa Fluor® 488-conjugated goat anti-rabbit IgG secondary antibody (Molecular Probes). Images were taken with Zeiss Apotome using the same exposure time between samples.

Gnotobiotic Mouse Experiments

All experiments involving animals, including euthanasia via carbon dioxide asphyxiation, were approved by the University Committee on Use and Care of Animals at the University of Michigan (NIH Office of Laboratory Animal Welfare number A3114–01) and overseen by a veterinarian. Groups of 3 to 5, 6–8 week old male and female germfree Swiss Webster mice were randomly assigned to each experiment. 7 days prior gavage the animals diet was switched to a fiber-free diet (Envigo-Teklad TD 130343) that was maintained through all the experiment. At day 0, mice were gavage with equal amount of *Bt* WT strain and mutant and fecal samples were collected at day 2 and every 5 days until day 42. At the end-point of the experiment distal small intestine and cecal contents were also collected. The bacteria gDNA extraction and quantification by qPCR of the relative abundance of each strain on the various samples was carried out as described previously[52].

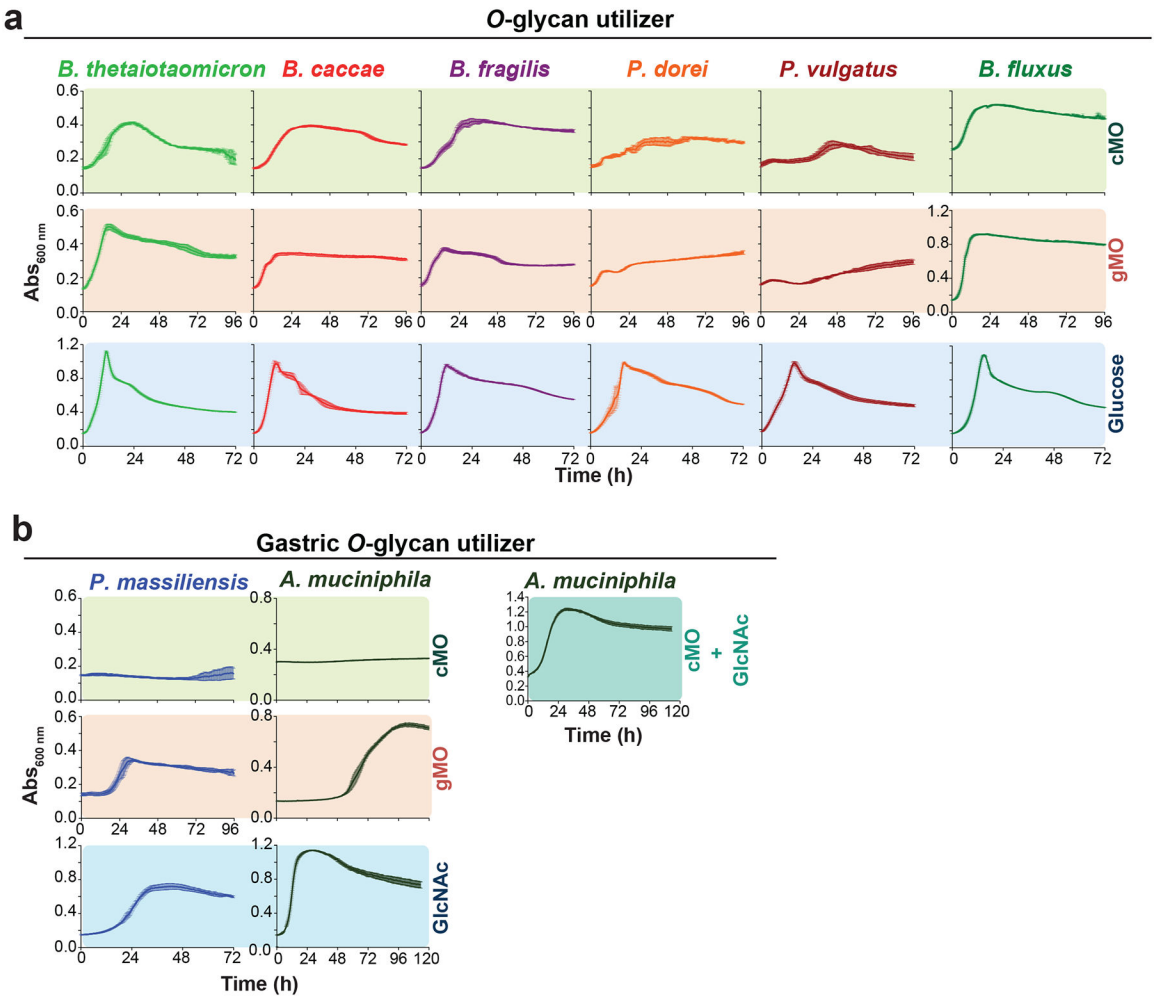
Phylogenetic analysis

To maximise sequence coverage, and avoid repetition, we selected 800 and 920 representative sequences of subfamily S1_20 (composed of 1356 sequences) and S1_4 (composed of 1895 sequences), respectively. To generate the phylogenetic tree of sulfatases from type strains (Supplementary Table 1) we selected all the sulfatases reported in SulfAtlas database (328 sequences). The sequences were aligned by MAFFT v.7[55] using L-INS-i algorithm. The multiple sequence alignment was visualized by Jalview software v.11.0[56] and non-aligned regions were removed. For phylogeny, we used 404, 364 and 294 positions for the S1_4, S1_20 and type strains, respectively. Phylogeny was made using RAxML v. 8.2.4[57]. The phylogenetic tree was build with the Maximum Likelihood method[58] and the LG matrix as evolutive model[59] using a discrete Gamma distribution to model evolutionary rate differences among sites (4 categories). The rate variation model allowed for some sites to be evolutionarily invariable. The reliability of the trees was tested by bootstrap analysis using 1,000 (S1_4 and S1_20) and 100 (type-strains phylogeny) resamplings of the dataset[60]. All the final global phylogenetic trees were obtained with MEGA v.7[61]. Fifteen S1_0 sequences from the sulfAtlas database were used as an outgroup in S1_4 and S20 phylogeny. The S1_0 sequence (Phosphonate monoester hydrolase / phosphodiesterase) of *Rhizobium leguminosrum viciae* 3841 was used as outgroup in the phylogenetic tree of the type strains. Homologs to $BT1636^{3S\text{-}Gal}$ and BT16223S-Gal are listed in Supplementary Tables 15 and 16, respectively. Input fasta and output (.nwk) alignment files are provides as Source Data 2.

Quantification and statistical analysis

For *in vivo* competitions, when three or more fecal samples were collected, Student's t tests were performed for each time point in GraphPad Prism v8.3.0 with a paired, two-tailed distribution. When necessary, the statistical analysis for remaining samples in stated in the respective figure legend.

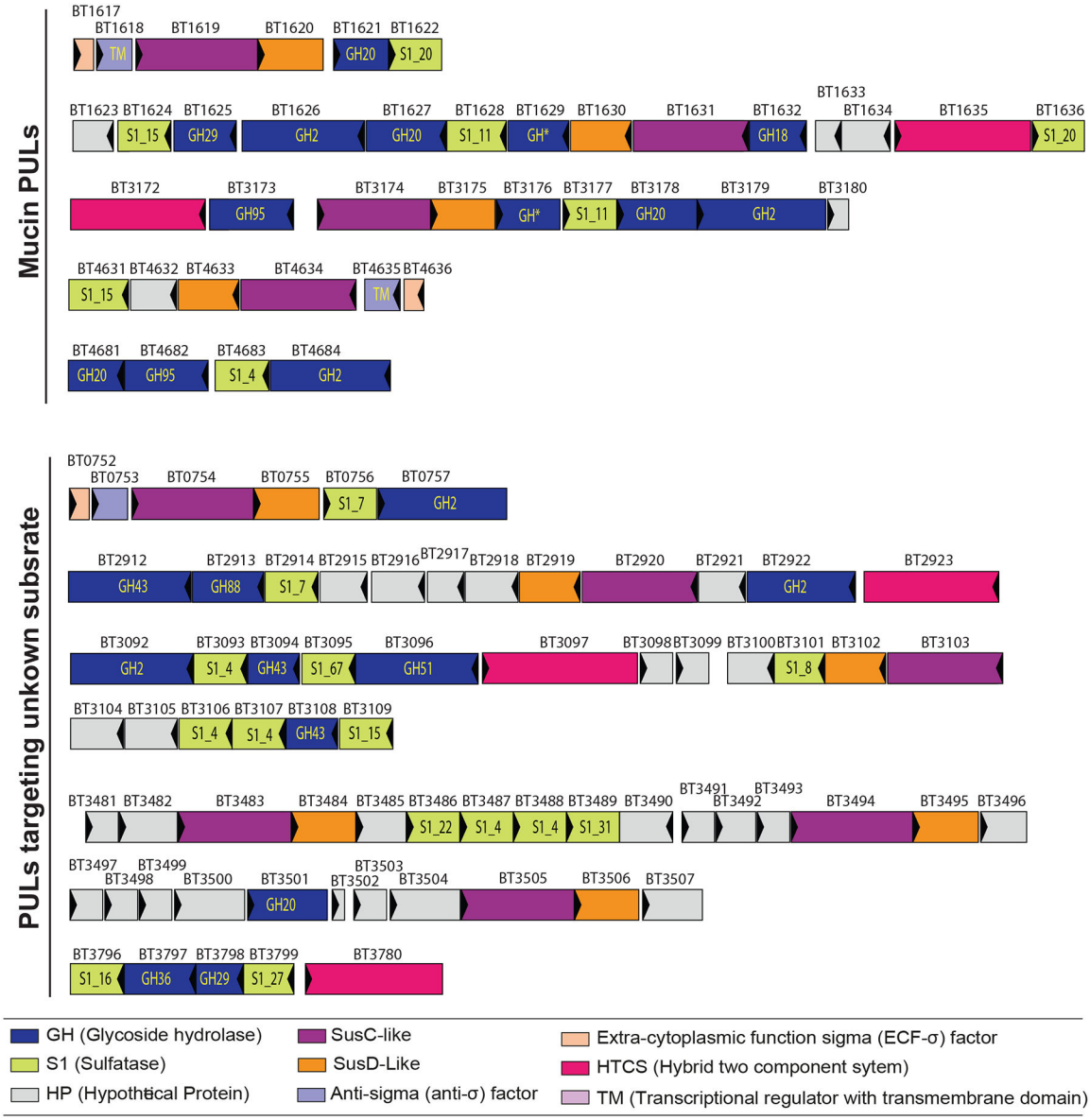
Extended Data



Extended Data 1. Growth of *Bacteroides* and *Phocaeicola* type strains and A*kkermansia muciniphila* in different mucin *O*-glycans.

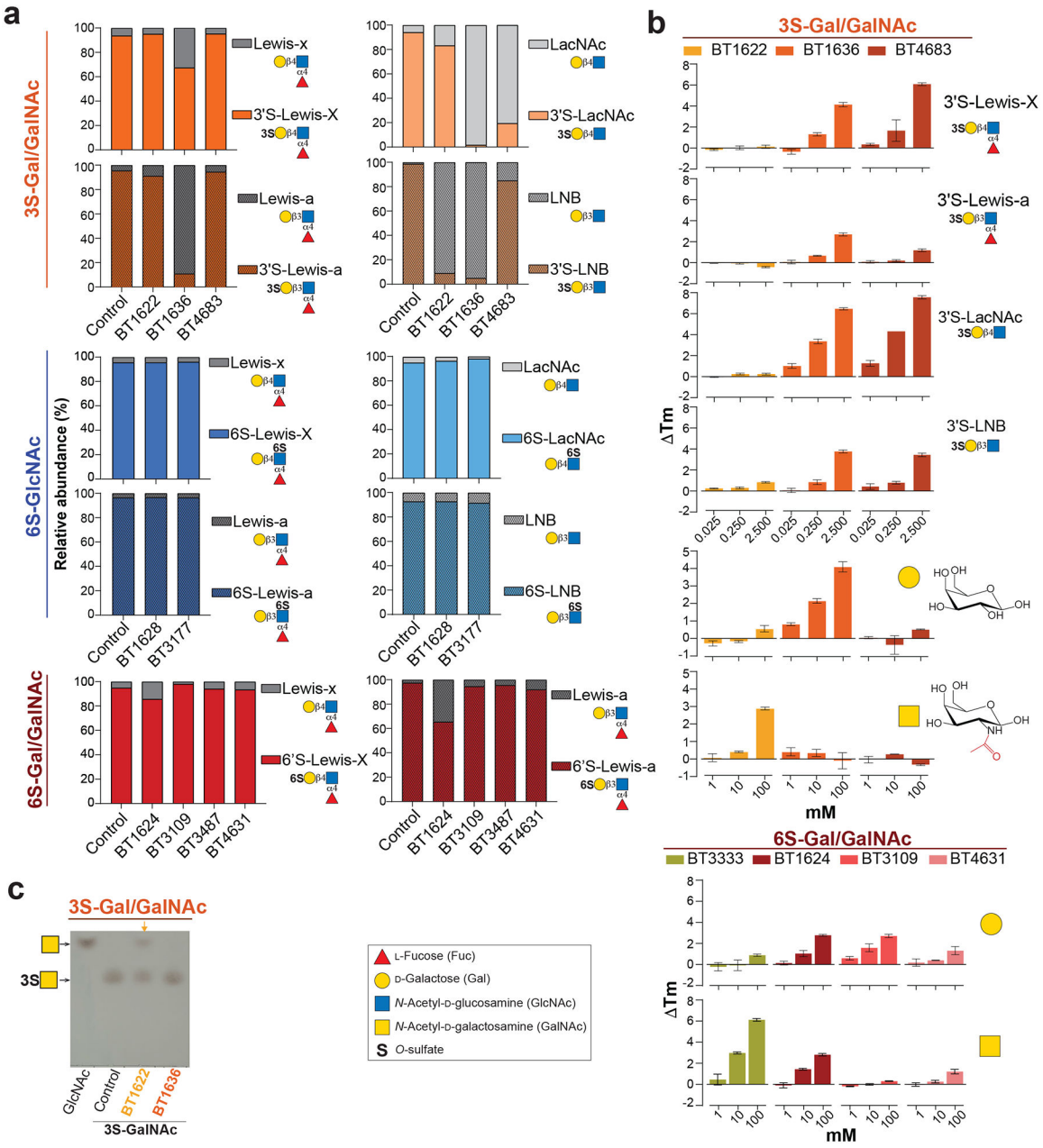
a, Graphs showing the growth of strains that are able to utilize colonic or gastric *O*-glycans. Growths were performed in minimal media containing the indicated carbon source. **b**, Growth experiments performed identically to panel **a**, but with two species, *P. massiliensis* and *A. muciniphila*, that grow on gMO but not cMO. A control experiment was performed with *A. muciniphila* grown on cMO plus added GlcNAC to verify that cMO does not contain material that is inhibitory to this species (biological replicates n =3 for both panels, error bars denote the s.e.m. for each time point). Note that gMO were used at 10mg/ml final concentration, while cMO were used at 5mg/ml due to background turbidity. This reduced

concentration and the higher amount of sulfate in cMO account for the lower growth on this substrate. cMO, colonic mucin *O*-glycans; gMO, gastric mucin *O*-glycans, GlcNAc, *N*-acetyl-D-glucosamine.



Extended Data 2. Schematic representation of polysaccharide utilization loci (PULs) encoding sulfatases (sulf).

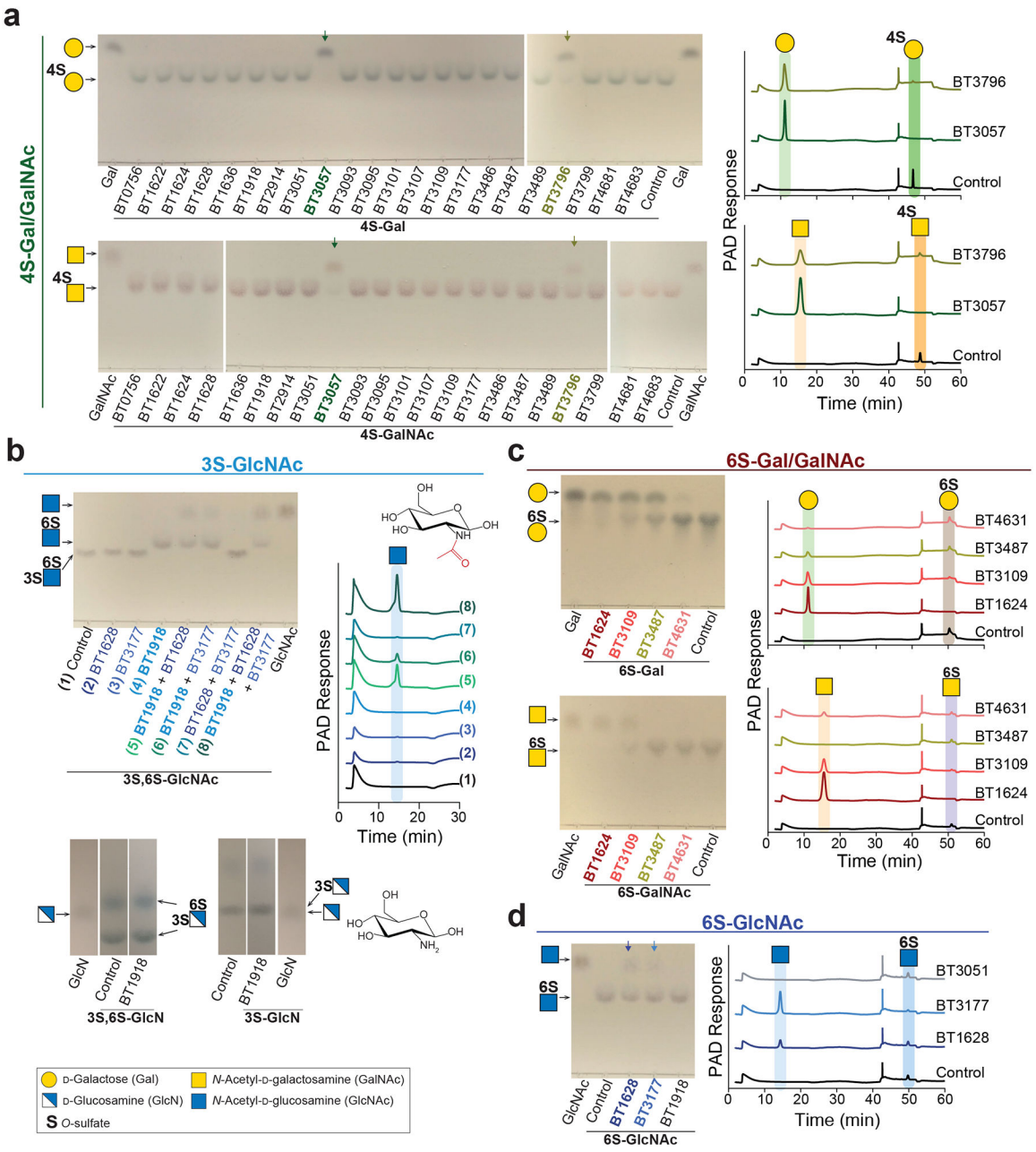
Genes are colour coded according to the predicted function of the respective proteins. Glycoside hydrolases (GH) in known families are indicated by GHXX or GH*, where XX and * indicates the respective family number or non-classified, respectively.



Extended Data 3. Activity and affinity of sulfatases to targeted substrates.

a, Recombinant enzymes (1 μM) were incubated with 1 mM of substrate in 10 mM MES pH6.5 with 5 mM $CaCl_2$ for 16h at 37 °C. Sulfated disaccharides were generated by adding 1 μM of a characterized α1,3/1,4-fucosidase (BT1625) in the enzymatic reaction. Control reactions without sulfatases were carried in the same conditions. Samples were analysed by mass spectrometry and the intensity of the substrate and reaction products was used for comparison of the relative abundance of these sugars after incubation with the respective enzymes. **b**, Affinity studies looking at the effect of ligand binding on the melting temperature of 3S and 6S-Gal sulfatases. All reactions were performed in 100 mM BTP, pH 7.0 with 150 mM NaCl. For sample melting temperatures see Supplementary Table 11. **c**,

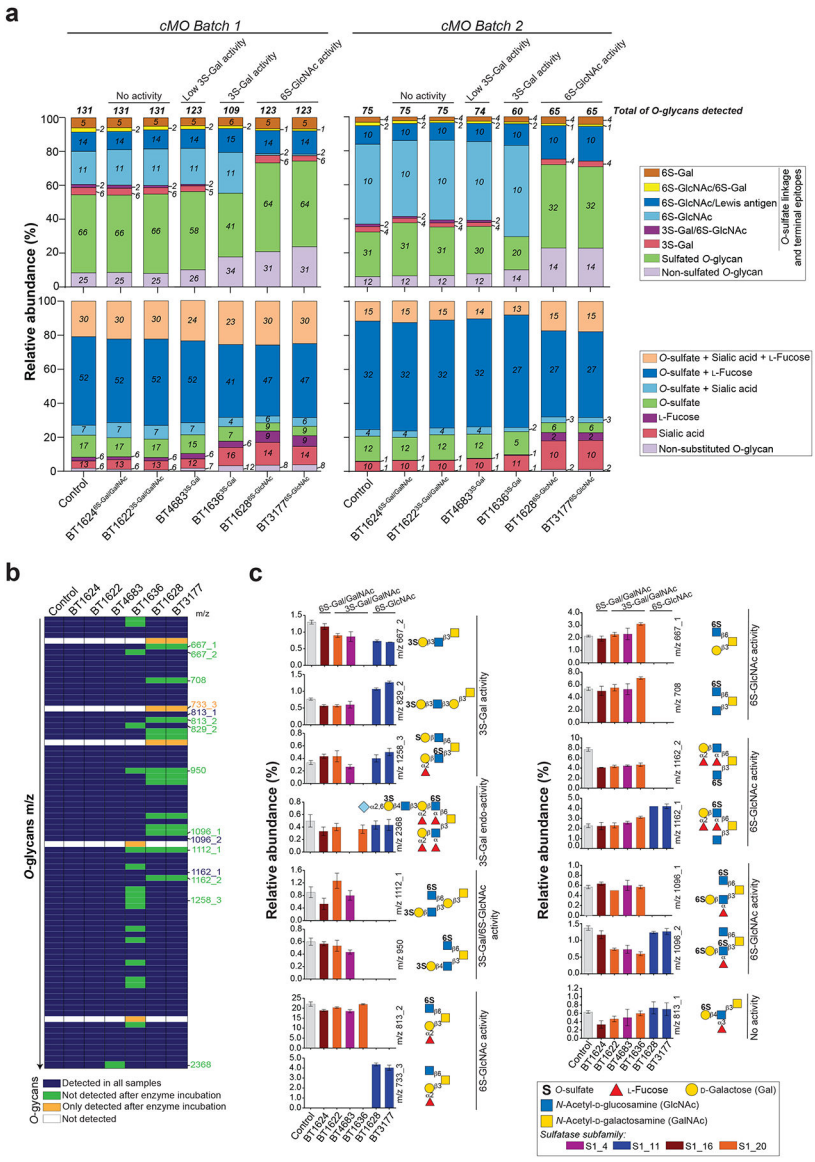
Activity of 3S-Gal/GalNAc sulfatases (10 μM) against 3S-GalNAc (10 mM). Reactions were performed in 10 mM Hepes, pH 7.0, with 150 mM NaCl and 5 mM $CaCl_2$. The data shown are one representative from the biological replicates conducted (n = 3).



Extended Data 4. Enzymatic screen of *Bt* sulfatases using sulfated monosaccharides.

Recombinant enzymes (1 μM) were incubated with 1 mM of substrate in 10 mM MES pH6.5 with 5 mM $CaCl_2$ for 16 h at 37°C. Reactions were analyzed by thin layer chromatography (left side) or HPAEC with pulsed amperometric detection (right side). Control reactions without sulfatases were carried out in the same conditions. The standards in TLC and HPAEC-PAD are labelled on the left side and top, respectively. The different

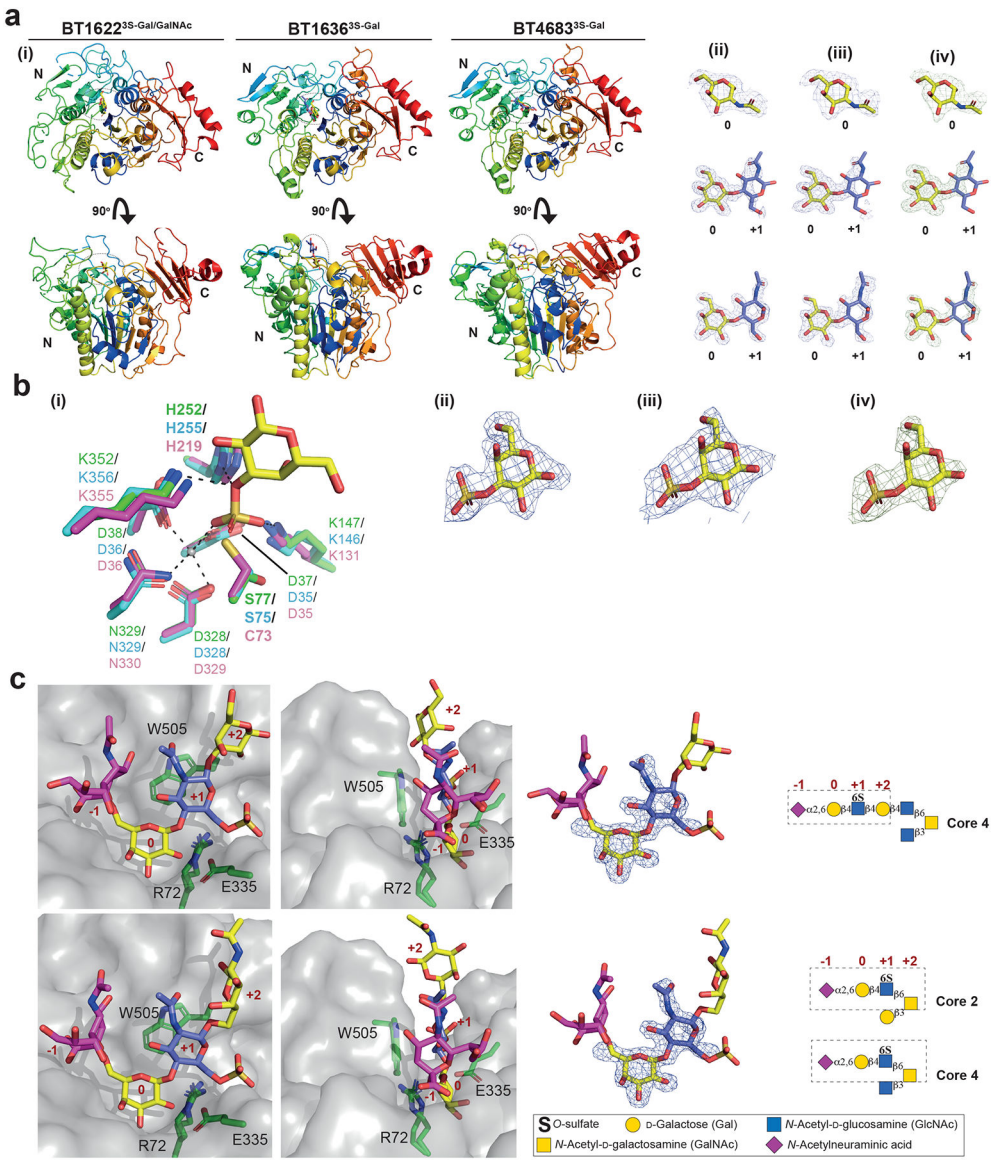
panel represent activities found for sulfatases targeting: **(a)** 4S-Gal/GalNAc; **(b)** 3S-GlcNAc **(c)** 6S-Gal/GalNAc; **(d)** 6S-GlcNAc. The data shown are representative from biological replicates (n = 3).



Extended Data 5. Activity of *Bt* sulfatases against colonic mucin *O*-glycans (cMO) analysed by mass spectrometry.

a, Relative abundance of structures detected in different samples organized by sulfate-linkage (top panel) or presence of one or several sugar substitutions such as sulfate, sialic acid and fucose (bottom panel). The colour-coded bars represent the relative abundance and the total number of the structures containing the specific linkage/substitution; **b**, Representation of *O*-glycans detected by mass spectrometry in cMO batch 2 (control) and after sulfatase treatment from the lower (top) to the higher (bottom) mass range; **c**, Relative abundance and putative structures for the specific m/z shown in panel b. The putative structure for the different mass is shown on the right side of the graphic. The reactions were

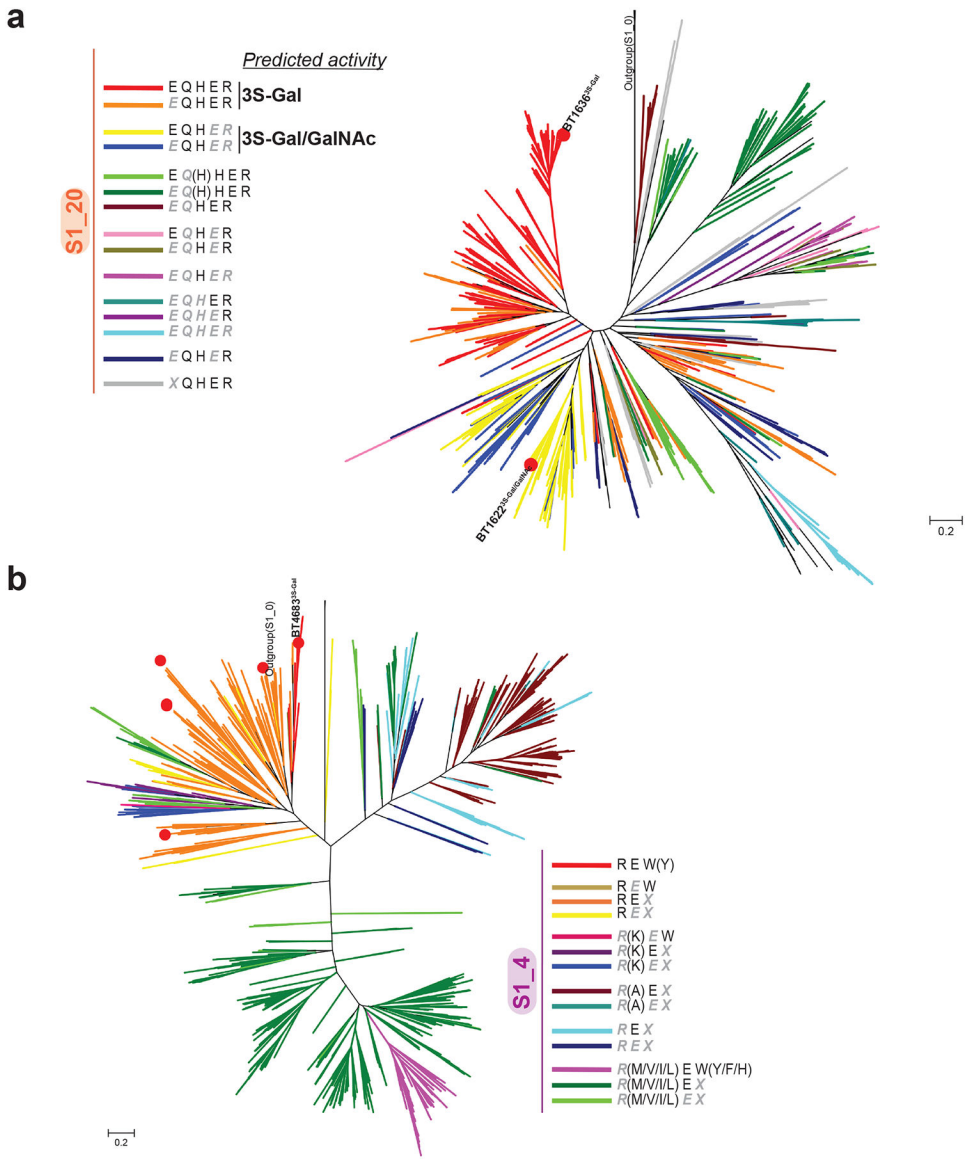
performed with 1 μM of enzyme and 0.5% cMO in 10 mM MES pH 6.5 with 5 mM $CaCl_2$ for 16 h at 37°C. The complete dataset is provided in Supplementary Table 4 and 5 for cMO batch 1 and 2, respectively.



Extended Data 6. Schematic representation of 3S-Gal/GalNAc sulfatases.

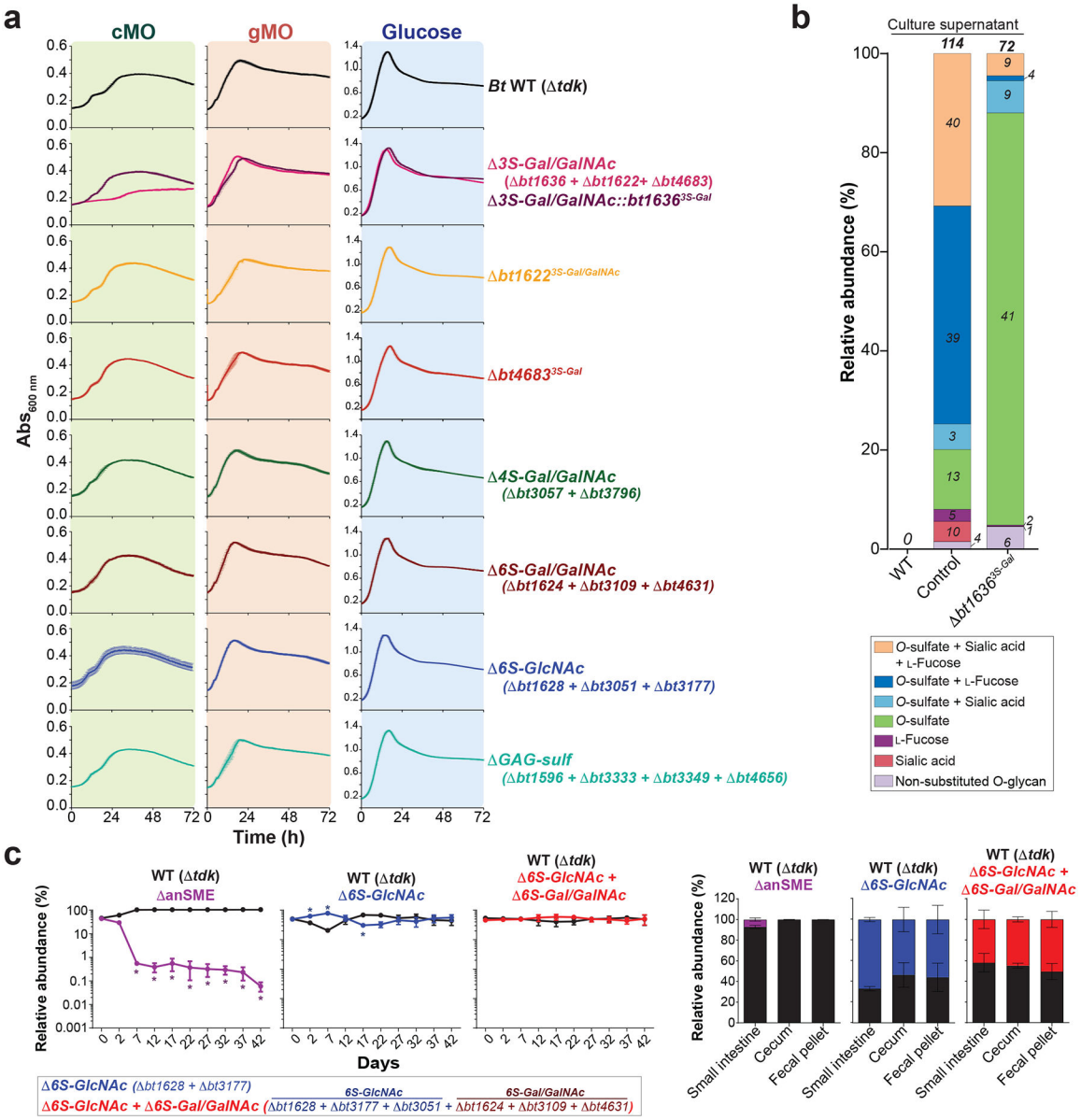
a,(i) Cartoon representation colour ramped from blue (α/β/α N-terminal domain) to red (β-sheet C-terminal domain); (ii) the final $2mF_{obs}$-DF_{calc} maps contoured at 1σ for GalNAc in BT1622$^{3S\text{-}Gal/GalNAc}$ (Top) LacNAc in BT1636$^{3S\text{-}Gal}$ (middle) and BT4683$^{3S\text{-}Gal}$ (bottom); (iii) represents the simulated annealed composite omit $2mF_{obs}$-DF_{calc} maps contoured at 1σ and (iv) represents the mF_{obs}-DF_{calc} maps, prior to building of the ligand contoured at 3σ;. **b**,(i) Overlay of the active site S residues of BT1636$^{3S\text{-}Gal}$ (green) BT1622$^{3S\text{-}Gal/GalNAc}$ (blue) and BT4683$^{3S\text{-}Gal}$ (pink). The putative catalytic residues are shown in bold. The calcium ion is represented as a grey sphere and its polar interactions indicated as

dashed lines. The 3S-Gal substrate is from the BT1636$^{3S\text{-}Gal}$ 3'S-Lewis-a complex, and BT1622$^{3S\text{-}Gal/GalNAc}$ and BT4683$^{3S\text{-}Gal}$ structures have been overlaid, (ii) the final $2mF_{obs}$-DF_{calc} maps of the observed 3'S-Lewis-a substrate contoured at 1σ, (iii) represents the simulated annealed composite omit $2mF_{obs}$-DF_{calc} maps contoured at 1σ, and (iv) represents the mF_{obs}-DF_{calc} maps of the observed 3'S-Lewis-a substrate, prior to building of the ligand, contoured at 3σ; **c**, Docking of putative structures of *O*-glycans targeted by BT4683$^{3S\text{-}Gal}$ using the LacNAc as reference point showing that this structure can accommodate a sialic acid in −1 subsite and additional sugars in positive subsites (left hand side). The docking sugars are shown as sticks (middle panel) and a schematic is represented inside the dashed box (right hand side). Using the LacNAc product as an 'anchor' additional sugars were built in manually with Coot 0.9 and regularized to low energy conformations.



Extended Data 7. Phylogenetic tree of S1_20 and S1_4 sulfatases.

The radial trees were constructed using the branched trees shown in Supplementary Figs. 3 and 4. For clarity, all labels and sequence accession codes have been omitted. Red filled circles designate sequences from *B. thetaiotaomicron* sulfatases. The residue is written in black without any attributes if present in the sequence, in grey and italics if the residue is mutated to any type in that sequence, or to a specific residue type if given in brackets. **a**, Radial representation of the phylogenetic tree constructed with representative sequences of the sulfatase S1_20 subfamily. The colour code is given as a pattern of presence or absence of the residues E100, Q173 H177, E334, R353, which are crucial in substrate recognition by BT1636 (acc-code Q8A789, coloured red). A grey X in italics specifically designates that the residue E100 is absent in that sequence, and no obvious orthologous residue can be found from the alignment. **b**, Radial representation of the phylogenetic tree constructed with representative sequences of the sulfatase S1_4 subfamily. The colour code is given as a pattern of presence or absence of the residues R72, E335 and W505, which are crucial in substrate recognition by BT4683 (acc-code Q89YP8, coloured red). A grey X in italics specifically designates that the residue W505 is absent in that sequence, and no obvious orthologous residue can be found from the alignment.



Extended Data 8. Sulfatase activity is required for growth in cMO and *in vivo* fitness.

a, Growth curves of *Bt* wild-type Δ*tdk* (WT), different sulfatase mutants (Δ*btXXX*) and complemented strains on glucose, colonic or gastric mucin *O*-glycans (cMO and gMO, respectively). The curves represent the average of biological replicates (n = 3) and the error bars denote s.e.m. **b**, Relative abundance of oligosaccharides detected by mass spectrometry in culture supernatant of WT and Δ*bt1636*$^{3S\text{-}Gal}$ after growth in cMO for 96h at anaerobic conditions. The control corresponds to cMO incubated in the same conditions without bacterium. The colours represent the relative abundance of structures grouped according to the presence of epitopes (sulfate, fucose and sialic acid) and the numbers represent the total number of structures that contain the respective substitution. **c**, Colonization of gnotobiotic mice fed a fiber-free diet by *Bt* WT and mutants lacking the full (Δ*anSME*, no S1 sulfatases active) or specific sulfatase activity (Δ*6S-GlcNAc* and Δ*6S-GlcNAc*+Δ*6S-Gal/GalNAc*). The

fecal relative abundance of each strain was determined at regular intervals until day 42. The relative abundance of time 0 represents the abundance in the gavaged inoculum. At the experimental endpoint the relative abundance was also determined in small intestine and cecum. The graphs represent the average of n=3–7 and the error bars denote the s.e.m. The relative abundance in each individual animal is represented in a lighter colour in each of the respective graphics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 748336. This work was supported by National Institutes of Health grants (DK118024 and DK125445 awarded to ECM, U01AI095473 awarded to GCH), the European Research Council ERC (694181), The Knut and Alice Wallenberg Foundation (2017.0028), Swedish Research Council (2017-00958), Wilhelm och Martina Lundgrens Vetenskapsfond (2020.3597, awarded to ASL) and the Academy of Medical Sciences/Wellcome Trust through the Springboard Grant SBF005\1065 163470 awarded to AC. The authors acknowledge access to the SOLEIL and Diamond Light sources via both the University of Liverpool and Newcastle university BAGs (proposals mx21970 and mx18598, respectively). We thank the staff of DIAMOND, SOLEIL, and members of the Liverpool's Molecular biophysics group for assistance with data collection. We thank members of the University of Michigan Mouse Facility and acknowledge the University of Michigan Center for Gastrointestinal Research (UMCGR, NIDDK 5P30DK034933) for support. Mass spectrometry of glycans was performed in the Swedish infrastructure for biologic mass spectrometry (BioMS) supported by the Swedish Research Council. We are also grateful for Dr. Erwan Corre's help regarding bioinformatics analyses (ABIMS platform, Station Biologique de Roscoff, France).

Data availability statement

Source Data for all experiments, along with corresponding statistical test values, where appropriate, are provided within the paper and in Supplementary Information. The crystal structure dataset generates have been deposit in the in the Protein Data Bank (PDB) under the following accession numbers: 7ANB, 7ANA, 7AN1, 7OQD, and 7ALL. The MS raw files were deposit in the GlycoPOST database under the following ID: GPST000150 and GPST000196. Glycan structural annotations were deposited to the Unicarb database as accession (https://unicarb-dr.glycosmos.org/references/462). There are no restrictions on data or biological resource availability. Data and biological resources can be obtained by contacting the corresponding authors.

Main References

1. Johansson ME et al. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proceedings of the National Academy of Sciences of the United States of America 105, 15064–15069, doi:10.1073/pnas.0803124105 (2008). [PubMed: 18806221]
2. Benjdia A, Martens EC, Gordon JI & Berteau O Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, Bacteroides thetaiotaomicron. J Biol Chem 286, 25973–25982, doi:10.1074/jbc.M111.228841 (2011). [PubMed: 21507958]
3. Hickey CA et al. Colitogenic Bacteroides thetaiotaomicron Antigens Access Host Immune Cells in a Sulfatase-Dependent Manner via Outer Membrane Vesicles. Cell Host Microbe 17, 672–680, doi:10.1016/j.chom.2015.04.002 (2015). [PubMed: 25974305]

4. Packey CD & Sartor RB Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. Current opinion in infectious diseases 22, 292–301 (2009). [PubMed: 19352175]
5. Sears CL & Garrett WS Microbes, Microbiota, and Colon Cancer. Cell Host Microbe 15, 317–328, doi:10.1016/j.chom.2014.02.007 (2014). [PubMed: 24629338]
6. Van der Sluis M et al. Muc2-deficient mice spontaneously develop colitis, indicating that Muc2 is critical for colonic protection. Gastroenterology 131, 117–129, doi:Doi 10.1053/J.Gastro.2006.04.020 (2006). [PubMed: 16831596]
7. Velcich A et al. Colorectal cancer in mice genetically deficient in the mucin Muc2. Science 295, 1726–1729, doi:10.1126/science.1069094 (2002). [PubMed: 11872843]
8. Bergstrom K et al. Core 1- and 3-derived O-glycans collectively maintain the colonic mucus barrier and protect against spontaneous colitis in mice. Mucosal immunology, doi:10.1038/mi.2016.45 (2016).
9. Larsson JM et al. Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. Inflamm Bowel Dis 17, 2299–2307, doi:10.1002/ibd.21625 (2011). [PubMed: 21290483]
10. Fu J et al. Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. J Clin Invest 121, 1657–1666, doi:10.1172/JCI45538 (2011). [PubMed: 21383503]
11. Kudelka MR et al. Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. Proceedings of the National Academy of Sciences of the United States of America 113, 14787–14792, doi:10.1073/pnas.1612158114 (2016). [PubMed: 27930307]
12. Larsson JM, Karlsson H, Sjovall H & Hansson GC A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. Glycobiology 19, 756–766, doi:10.1093/glycob/cwp048 (2009). [PubMed: 19321523]
13. Holmen Larsson JM, Thomsson KA, Rodriguez-Pineiro AM, Karlsson H & Hansson GC Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. American journal of physiology. Gastrointestinal and liver physiology 305, G357–363, doi:10.1152/ajpgi.00048.2013 (2013). [PubMed: 23832516]
14. Thomsson KA et al. Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. Glycobiology 22, 1128–1139, doi:10.1093/glycob/cws083 (2012). [PubMed: 22581805]
15. Robbe C et al. Evidence of regio-specific glycosylation in human intestinal mucins - Presence of an acidic gradient along the intestinal tract. Journal of Biological Chemistry 278, 46337–46348, doi:DOI 10.1074/jbc.M302529200 (2003). [PubMed: 12952970]
16. Katoh T et al. Identification and characterization of a sulfoglycosidase from Bifidobacterium bifidum implicated in mucin glycan utilization. Bioscience, biotechnology, and biochemistry 81, 2018–2027, doi:10.1080/09168451.2017.1361810 (2017). [PubMed: 28814130]
17. Martens EC, Chiang HC & Gordon JI Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host Microbe 4, 447–457, doi:DOI 10.1016/j.chom.2008.09.007 (2008). [PubMed: 18996345]
18. Diez-Roux G & Ballabio A Sulfatases and human disease. Annu Rev Genomics Hum Genet 6, 355–379, doi:10.1146/annurev.genom.6.080604.162334 (2005). [PubMed: 16124866]
19. Wlodarska M et al. Indoleacrylic Acid Produced by Commensal Peptostreptococcus Species Suppresses Inflammation. Cell Host Microbe 22, 25–37 e26, doi:10.1016/j.chom.2017.06.007 (2017). [PubMed: 28704649]
20. Tramontano M et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat Microbiol 3, 514–522, doi:10.1038/s41564-018-0123-9 (2018). [PubMed: 29556107]
21. Derrien M, Vaughan EE, Plugge CM & de Vos WM Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54, 1469–1476 (2004). [PubMed: 15388697]

22. Pudlo NA et al. Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. mBio 6, e01282–01215, doi:10.1128/mBio.01282-15 (2015). [PubMed: 26556271]
23. Barbeyron T et al. Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. PLoS ONE 11, e0164846, doi:10.1371/journal.pone.0164846 (2016). [PubMed: 27749924]
24. Benjdia A et al. Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. J Biol Chem 283, 17815–17826, doi:10.1074/jbc.M710074200 (2008). [PubMed: 18408004]
25. Cartmell A et al. How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. Proceedings of the National Academy of Sciences of the United States of America 114, 7037–7042, doi:10.1073/pnas.1704367114 (2017). [PubMed: 28630303]
26. Ndeh D et al. Metabolism of multiple glycosaminoglycans by Bacteroides thetaiotaomicron is orchestrated by a versatile core genetic locus. Nature communications 11, 646, doi:10.1038/s41467-020-14509-4 (2020).
27. Ulmer JE et al. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. J Biol Chem 289, 24289–24303, doi:10.1074/jbc.M114.573303 (2014). [PubMed: 25002587]
28. Tobisawa Y, Imai Y, Fukuda M & Kawashima H Sulfation of colonic mucins by N-acetylglucosamine 6-O-sulfotransferase-2 and its protective function in experimental colitis in mice. J Biol Chem 285, 6750–6760, doi:10.1074/jbc.M109.067082 (2010). [PubMed: 20018871]
29. Egan M, Jiang H, O'Connell Motherway M, Oscarson S & van Sinderen D Glycosulfatase-Encoding Gene Cluster in Bifidobacterium breve UCC2003. Appl Environ Microbiol 82, 6611–6623, doi:10.1128/AEM.02022-16 (2016). [PubMed: 27590817]
30. Briliute J et al. Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. Nat Microbiol 4, 1571–1581, doi:10.1038/s41564-019-0466-x (2019). [PubMed: 31160824]
31. Packer NH, Lawson MA, Jardine DR & Redmond JW A general approach to desalting oligosaccharides released from glycoproteins. Glycoconjugate journal 15, 737–747, doi:10.1023/a:1006983125913 (1998). [PubMed: 9870349]
32. Hayes CA et al. UniCarb-DB: a database resource for glycomic discovery. Bioinformatics 27, 1343–1344, doi:10.1093/bioinformatics/btr137 (2011). [PubMed: 21398669]
33. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH & Campbell MP Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. J Am Soc Mass Spectrom 24, 895–906, doi:10.1007/s13361-013-0610-4 (2013). [PubMed: 23605685]
34. Domon B & Costello CE Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. Biochemistry 27, 1534–1543, doi:10.1021/bi00405a021 (1988). [PubMed: 3365408]
35. Byrne DP et al. New tools for carbohydrate sulfation analysis: heparan sulfate 2-O-sulfotransferase (HS2ST) is a target for small-molecule protein kinase inhibitors. Biochem J 475, 2417–2433, doi:10.1042/BCJ20180265 (2018). [PubMed: 29934491]
36. Byrne DP et al. cAMP-dependent protein kinase (PKA) complexes probed by complementary differential scanning fluorimetry and ion mobility-mass spectrometry. Biochem J 473, 3159–3175, doi:10.1042/BCJ20160648 (2016). [PubMed: 27444646]
37. Das TM, Rao CP & Kolehmainen E Synthesis and characterisation of N-glycosyl amines from the reaction between 4,6-O-benzylidene-D-glucopyranose and substituted aromatic amines and also between 2-(o-aminophenyl)benzimidazole and pentoses or hexoses. Carbohydrate research 334, 261–269, doi:10.1016/s0008-6215(01)00202-6 (2001). [PubMed: 11527527]
38. Kabsch W Xds. Acta Crystallogr D Biol Crystallogr 66, 125–132, doi:10.1107/S0907444909047337 (2010). [PubMed: 20124692]
39. Winter SE, Lopez CA & Baumler AJ The dynamics of gut-associated microbial communities during inflammation. EMBO Rep 14, 319–327, doi:10.1038/embor.2013.27 (2013). [PubMed: 23478337]

40. Evans PR & Murshudov GN How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69, 1204–1214, doi:10.1107/S0907444913000061 (2013). [PubMed: 23793146]
41. Long F, Vagin AA, Young P & Murshudov GN BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 64, 125–132, doi:10.1107/S0907444907050172 (2008). [PubMed: 18094476]
42. Vagin A & Teplyakov A Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66, 22–25, doi:10.1107/S0907444909042589 (2010). [PubMed: 20057045]
43. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501, doi:10.1107/S0907444910007493 (2010). [PubMed: 20383002]
44. Murshudov GN et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67, 355–367, doi:10.1107/S0907444911001314 (2011). [PubMed: 21460454]
45. Lebedev AA et al. JLigand: a graphical tool for the CCP4 template-restraint library. Acta Crystallogr D Biol Crystallogr 68, 431–440, doi:10.1107/S090744491200251X (2012). [PubMed: 22505263]
46. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12–21, doi:10.1107/S0907444909042073 (2010). [PubMed: 20057044]
47. Agirre J et al. Privateer: software for the conformational validation of carbohydrate structures. Nat Struct Mol Biol 22, 833–834, doi:10.1038/nsmb.3115 (2015). [PubMed: 26581513]
48. Terwilliger TC et al. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. Acta Crystallogr D Biol Crystallogr 64, 515–524, doi:10.1107/S0907444908004319 (2008). [PubMed: 18453687]
49. Collaborative computational project, n. The CCP4 suite: Programs for protein crystallography. Acta Crystallographica D50, 760–763 (1994).
50. Potterton L et al. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallogr D Struct Biol 74, 68–84, doi:10.1107/S2059798317016035 (2018). [PubMed: 29533233]
51. Micsonai A et al. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Res 46, W315–W322, doi:10.1093/nar/gky497 (2018). [PubMed: 29893907]
52. Desai MS et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell 167, 1339–1353 e1321, doi:10.1016/j.cell.2016.10.043 (2016). [PubMed: 27863247]
53. Koropatkin NM, Martens EC, Gordon JI & Smith TJ Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16, 1105–1115, doi:10.1016/j.str.2008.03.017 (2008). [PubMed: 18611383]
54. Degnan PH, Barry NA, Mok KC, Taga ME & Goodman AL Human gut microbes use multiple transporters to distinguish vitamin B(1)(2) analogs and compete in the gut. Cell Host Microbe 15, 47–57, doi:10.1016/j.chom.2013.12.007 (2014). [PubMed: 24439897]
55. Katoh K, Misawa K, Kuma K & Miyata T MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059–3066, doi:10.1093/nar/gkf436 (2002). [PubMed: 12136088]
56. Clamp M, Cuff J, Searle SM & Barton GJ The Jalview Java alignment editor. Bioinformatics 20, 426–427, doi:10.1093/bioinformatics/btg430 (2004). [PubMed: 14960472]
57. Stamatakis A RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313, doi:10.1093/bioinformatics/btu033 (2014). [PubMed: 24451623]
58. Felsenstein J Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of molecular evolution 17, 368–376, doi:10.1007/BF01734359 (1981). [PubMed: 7288891]
59. Le SQ & Gascuel O An improved general amino acid replacement matrix. Mol Biol Evol 25, 1307–1320, doi:10.1093/molbev/msn067 (2008). [PubMed: 18367465]

60. Felsenstein J Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution 39, 783–791, doi:10.1111/j.1558-5646.1985.tb00420.x (1985). [PubMed: 28561359]

61. Kumar S, Stecher G & Tamura K MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870–1874, doi:10.1093/molbev/msw054 (2016). [PubMed: 27004904]

62. Neelamegham S et al. Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology 29, 620–624, doi:10.1093/glycob/cwz045 (2019). [PubMed: 31184695]

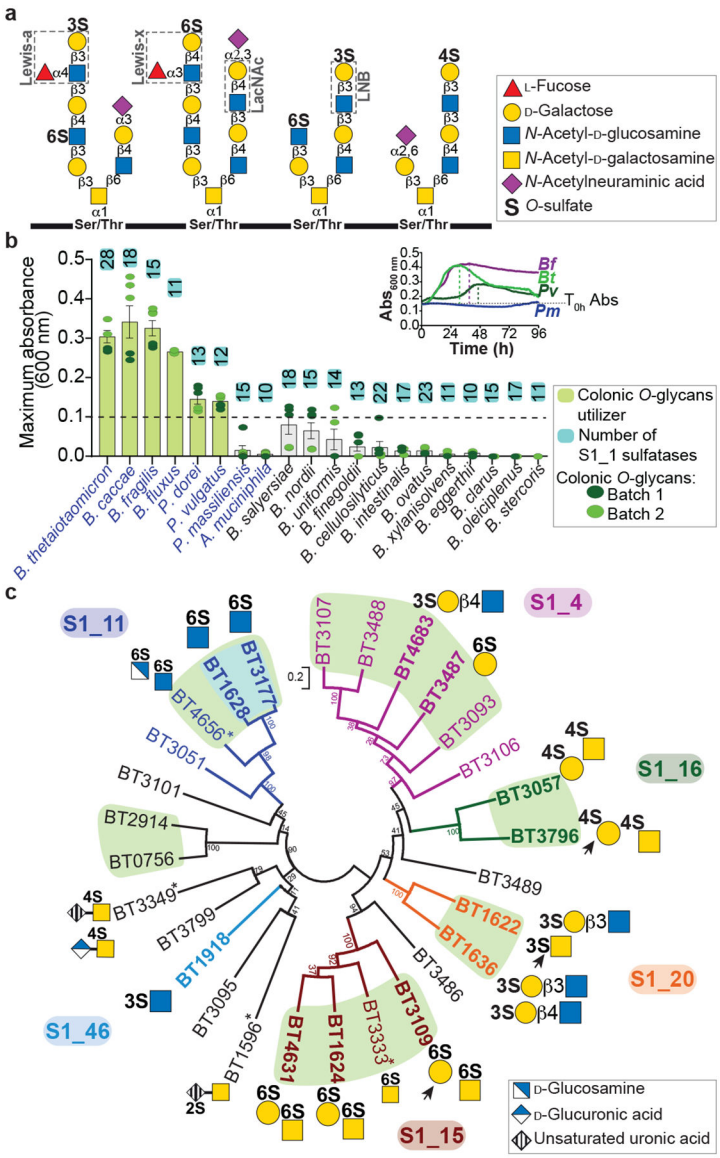


Figure 1. Bacterial growth on colonic mucin and *Bt* sulfatase activities.

a, Example mucin *O*-glycan structures with relevant terminal epitopes highlighted (dashed boxes). Sugars are shown according to the Symbol Nomenclature for Glycan system[62]. **b**, Growth of *Bacteroides*/*Phocaeicola* strains and A*kkermansia muciniphila* on colonic mucin *O*-glycans (cMO) and number of respective encoded S1 sulfatases. The bars represent the average of two independent experiments with different batches of cMO (total biological replicates n = 3 to 6, error bars denote s.e.m.). Bacterial species able to utilize gastric mucin glycans are highlighted in blue. Maximum absorbance is the maximum A_{600nm} for each culture minus the initial absorbance at time 0. The inset to panel **b** shows example growth curves for *B. fragilis* (*Bf*), *B. thetaiotaomicron* (*Bt*), *P. vulgatus* (*Pv*) and *P. massilliensis* (*Pm*). **c**, Phylogeny of *Bt* sulfatases showing the 28 S1 sulfatases and their respective substrates where known. Enzymes are color coded by subfamiy with those characterized in this study in bold. * indicates previously characterized activity and arrows point the

sulfate preferentially targeted by the respective enzyme. Sulfatases on a shared branch that share more than 86% and 39–58% of sequence identity are highlighted in blue and green background, respectively.

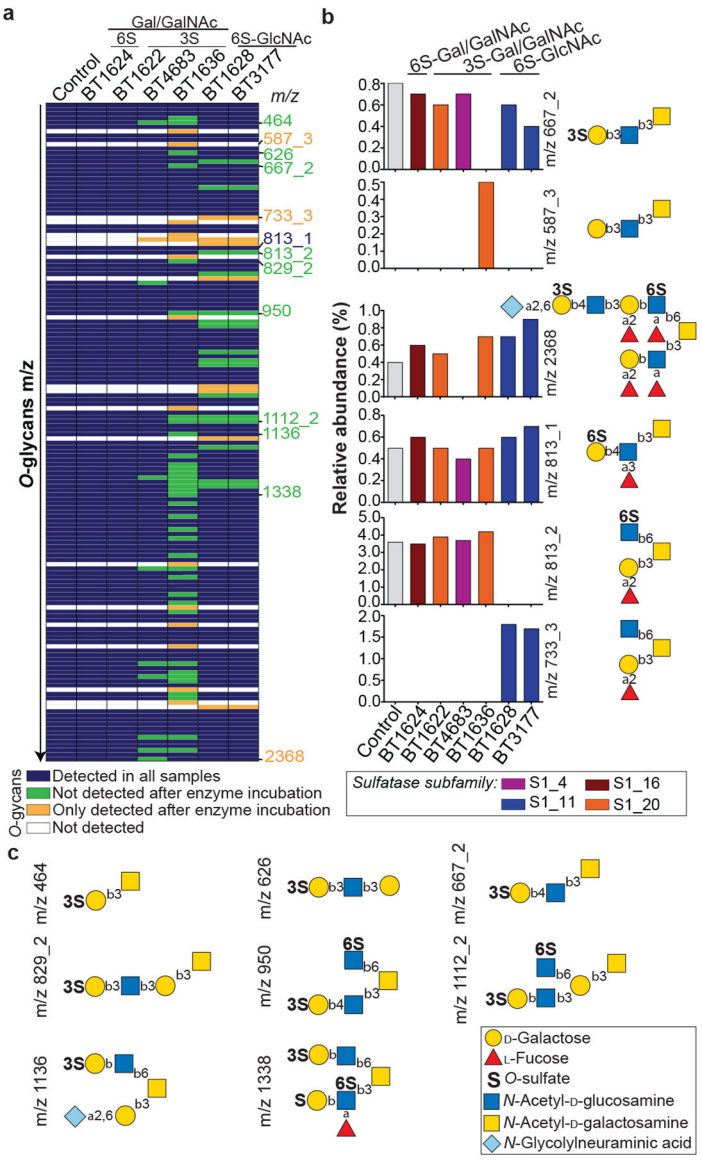


Figure 2. Activity of *Bt* sulfatases on colonic mucin *O*-glycans.

a, Representation of *O*-glycans detected by mass spectrometry in cMO (control) and after sulfatase treatment from the lower (top) to the higher (bottom) mass range. **b**, Relative abundance and putative structures for the specific m/z shown in panel a. **c**, Schematic representation of the putative structures that were not detected after treatment with BT1636$^{3S\text{-}Gal}$.

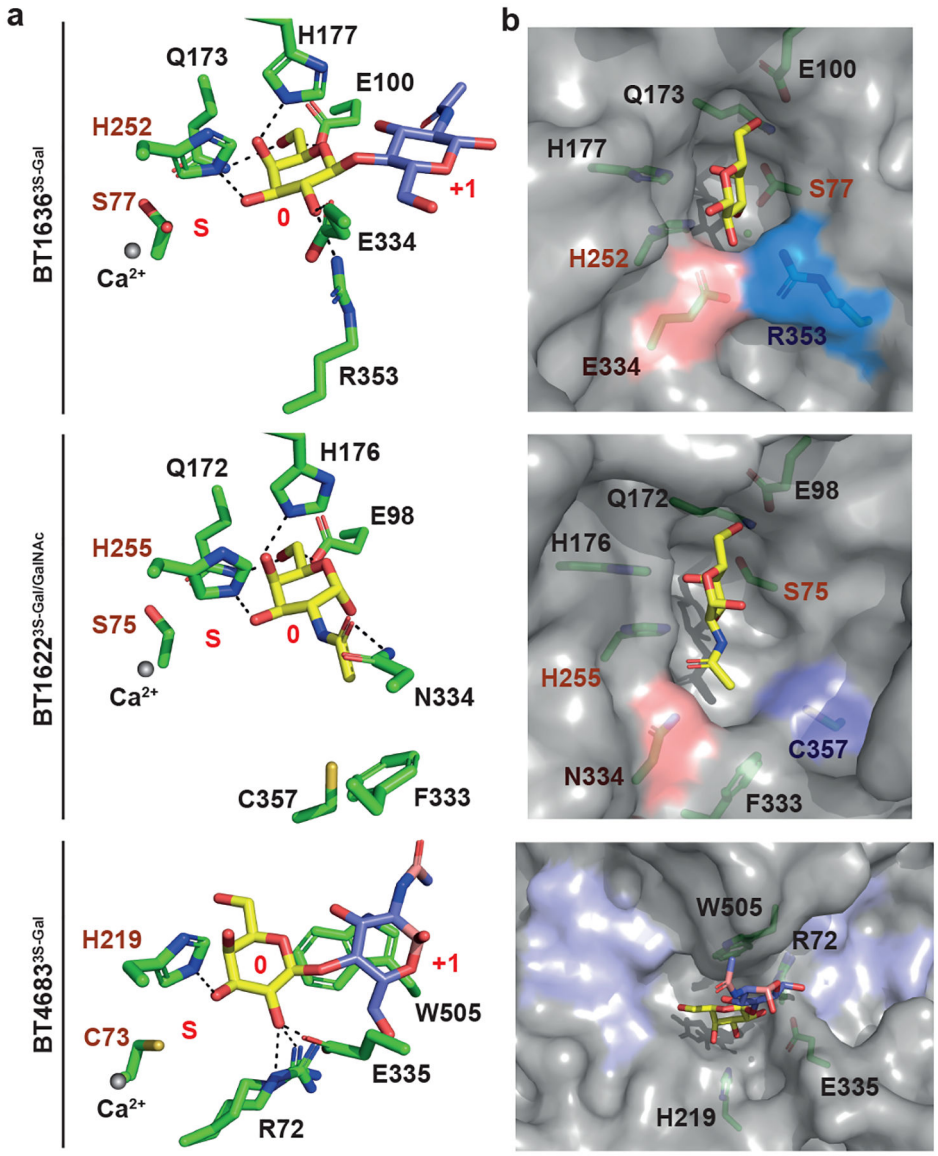


Figure 3. Crystal structures of 3S-Gal/GalNAc sulfatases.

a, Stick representation of the residues interacting with targeted sugars, including the putative catalytic residues (in dark red), the calcium ion (grey sphere) and subsites S, 0 and +1 highlighted in red. BT1636$^{3S\text{-}Gal}$ and BT4683$^{3S\text{-}Gal}$ in complex with LacNAc (D-Gal-β1,4-D-GlcNAc) and BT1622$^{3S\text{-}Gal/GalNAc}$ in complex with GalNAc. Note that in BT1636 the S77 occurs in two conformations. Since this in an inactive enzyme, we do not believe these conformations play a role in activity. **b**, Surface representation of the active site pocket. The equivalent Gal/GalNAc specificity residues in BT1636$^{3S\text{-}Gal}$ and BT4683$^{3S\text{-}Gal}$ are highlighted in red and blue. The open active site of BT4683$^{3S\text{-}Gal}$ is highlighted in purple. In all structures the amino acids and ligands are represented as stick. Note that the GlcNAc has been removed in panel b for clarity.

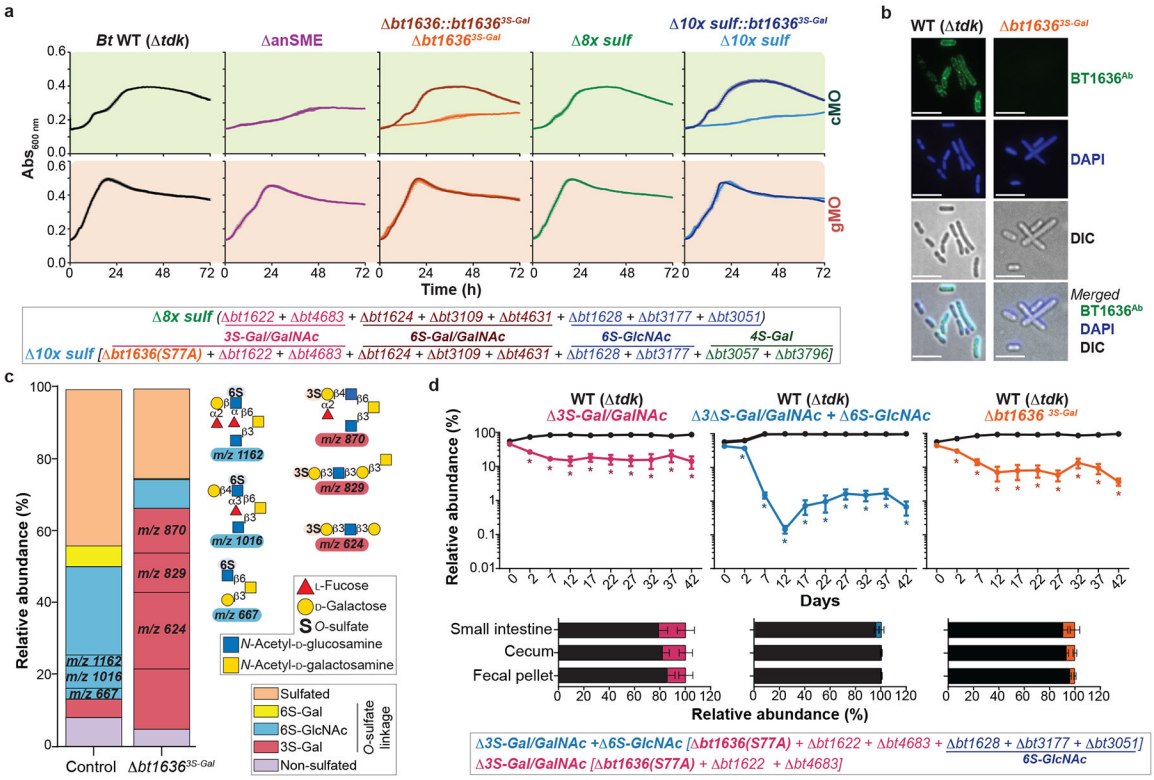


Figure 4. BT1636$^{3S\text{-}Gal}$ activity is required for the utilization of cMO and competitive fitness *in vivo*.

a, Growth of *Bt* wild-type Δ*tdk* (WT), different sulfatase gene-deletion mutants (named "Δ*btXXXX*") and strains complemented with *bt1636*$^{3S\text{-}Gal}$ on cMO and gMO. Note that gMO were used at 10mg/ml final concentration, while cMO were used at 5mg/ml due to background turbidity. This reduced concentration and the higher amount of sulfate in cMO account for the lower growth. Line represents the average of biological replicates (n = 3) and error bars denote s.e.m. **b**, Immunofluorescent and differential interference contrast (DIC) microscopy of *Bt* WT and sulfatase mutant staining with polyclonal antibody (Ab) against BT1636$^{3S\text{-}Gal}$ (green) and DNA staining with DAPI (blue). **c**, Relative abundance of different *O*-glycans detected by mass spectrometry in Δ*bt1636*$^{3S\text{-}Gal}$ culture supernatant or cMO in minimal media without bacteria (control), after 96h in anaerobic conditions. The mass and associated structures of the 3 most abundant glycans in both samples are shown. **d**, *in vivo* competitions in gnotobiotic mice (n = 7–9) fed fiber-free diet and inoculated with WT and mutants. The fecal relative abundance of each strain was determined at regular intervals until day 42. The relative abundance of time 0 represents the abundance in the gavaged inoculum. At the experimental endpoint the relative abundance was also determined in small intestine and cecum (represented in the histogram bars at the bottom). The error bars denote s.e.m.