



HAL
open science

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

Ana S Luis, Chunsheng Jin, Gabriel Vasconcelos Pereira, Robert W P Glowacki, Sadie R Gugel, Shaleni Singh, Dominic P Byrne, Nicholas A Pudlo, James A London, Arnaud Baslé, et al.

► To cite this version:

Ana S Luis, Chunsheng Jin, Gabriel Vasconcelos Pereira, Robert W P Glowacki, Sadie R Gugel, et al.. A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature*, 2021, 598 (7880), pp.332-337. 10.1038/s41586-021-03967-5 . hal-03758685

HAL Id: hal-03758685

<https://hal.science/hal-03758685v1>

Submitted on 23 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

A single sulfatase is required for metabolism of colonic mucin O-glycans and intestinal colonization by a symbiotic human gut bacterium

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

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

²Department of Medical Biochemistry, University of Gothenburg, Box 440, 405 30 Gothenburg, Sweden

³Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3BX, United Kingdom

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

⁵Centre for synthesis and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.

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

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

34 **Summary**

35 Humans have co-evolved with a dense community of microbial symbionts that
36 inhabit the lower intestine. In the colon, secreted mucus creates a physical barrier that
37 separates these microbes from the intestinal epithelium¹. Some gut bacteria are able to
38 utilize mucin glycoproteins, the main mucus component, as a nutrient source. However,
39 it remains unclear which enzymes initiate the degradation of the highly complex O-
40 glycans found in mucins. In the colon, these glycans are heavily sulfated, but sulfatases
41 active on colonic mucins have not been identified. Here, we show that sulfatases are
42 essential to the utilization of colonic mucin O-linked glycans by the human gut symbiont
43 *Bacteroides thetaiotaomicron*. We characterized the activity of 12 different sulfatases
44 encoded by this species, showing that these enzymes collectively are active on all of the
45 known sulfate linkages in colonic O-glycans and even possess the ability to cleave
46 additional linkages not yet known to occur in host glycans. Crystal structures of 3 enzymes
47 provide mechanistic insight into the molecular basis of substrate-specificity.
48 Unexpectedly, we found that a single sulfatase is essential for utilization of sulfated O-
49 glycans *in vitro* and also plays a major role *in vivo*. Our results provide insight into the
50 mechanisms of mucin degradation by a prominent group of gut bacteria, an important
51 process for both normal microbial gut colonization² and diseases such as inflammatory
52 bowel disease³.

53

54

55 **Introduction**

56 The human gut microbiota (HGM) significantly impacts several aspects of intestinal
57 health and disease, including inflammatory bowel disease (IBD)¹ and colorectal cancer
58 (CRC)². In the colon, secreted mucus creates a physical barrier that separates gut
59 microbes from the intestinal epithelium³ preventing close contact that leads to
60 inflammation and eventual CRC when this barrier is either experimentally eliminated^{4,5} or
61 has reduced glycosylation⁶⁻⁸. A major component of the colonic mucus is mucin 2 (Muc2),
62 a glycoprotein that contains up to 80% glycans by weight and more than 100 different
63 glycan structures that are O-linked to serine or threonine residues⁹. Mucin glycosylation
64 is variable along the gastrointestinal (GI) tract with a marked increase in sulfation in the

65 colon¹⁰. In mucins, O-linked sulfate may be attached to the 6-hydroxyl of *N*-acetyl-D-
66 glucosamine (6S-GlcNAc) and non-reducing end D-galactose (Gal) sugars at hydroxyl
67 positions 3-, 4- or 6- (3S-, 4S- and 6S-Gal, respectively)^{10,11} (**Fig. 1a**). To degrade and
68 utilise colonic mucin O-glycans, members of the HGM need to express appropriate
69 carbohydrate sulfatases to remove these modifications, which often occur as terminal
70 capping sulfates that block enzymatic degradation of these sulfated glycans. *Bacteroides*
71 *thetaiotaomicron* (*Bt*) is a dominant member of the human gut microbiota that is able to
72 utilize O-glycans as a sole nutrient source¹². Underscoring the importance of sulfatases,
73 *Bt* requires active sulfatases for competitive colonization of the wild-type mouse gut¹³ and
74 to induce inflammation in genetically susceptible mice¹⁴. However, the specific sulfatases
75 that mediate these effects remain unknown. Indeed, despite the critical roles of sulfatases
76 in many biological processes, including several human diseases¹⁵, there is a significant
77 knowledge gap regarding the biochemical, structural and functional roles of these
78 enzymes.

79 We hypothesized that specific *Bt* sulfatases play key roles in initiation of O-glycan
80 degradation. To test this, we measured the activities of 23 putative sulfatases,
81 determining that 12 enzymes are active on either model glycan substrates or purified
82 colonic mucin O-glycans. Together with defining the positional activities of these
83 enzymes, we determined the corresponding structures for 3 sulfatases, revealing the
84 basis of substrate specificity. Using molecular genetics, we determined the contributions
85 of these sulfatases to *Bt* fitness *in vitro* and *in vivo*, unexpectedly revealing that a single
86 enzyme is essential for utilisation of sulfated mucin O-glycans and for competitive gut
87 colonization. Identifying specific bacterial sulfatases that are critical for intestinal mucin
88 degradation will provide potential targets to inhibit with the goal of blocking progression
89 of diseases like IBD and possibly other disorders that result from bacterial disruption of
90 the mucus barrier.

91

92 ***Utilization of colonic mucins by HGM species***

93 Several studies have identified HGM members that are able to utilize porcine
94 gastric mucin O-glycans (gMO) as a sole carbon source^{12,16,17}. However, this substrate
95 does not adequately reflect the structural complexity of mucin O-glycans found in the

96 colon, especially those with increased sulfation¹⁰. To identify HGM species that utilize
97 sulfated colonic mucins, we measured the growth of 19 *Bacteroides* type strains, plus
98 *Akkermansia muciniphila*, on highly sulfated porcine colonic mucin oligosaccharides
99 (cMO). We identified six *Bacteroides* strains that utilize cMO (**Fig. 1b**, **Extended Data**
100 **Fig. 1**). Interestingly, two known mucin-degraders, *A. muciniphila*¹⁸ and *B. massiliensis*¹⁹,
101 grew robustly on gMO but failed to utilize sulfated cMO to any detectable level (**Extended**
102 **Data Fig. 1**), highlighting the importance of using colonic mucins as substrates to draw
103 more complete and potentially relevant conclusions about the metabolic potential of these
104 organisms. *Bt*, the bacterium with the highest number of sulfatases (28), was one of the
105 strains with the best growth on cMO (**Fig. 1b**), suggesting that some of these enzymes
106 might play key roles in promoting this ability. Therefore, to understand the role of
107 sulfatases in colonic mucin utilization by HGM bacteria we focused on characterizing *Bt*
108 enzymes. See **Supplementary Discussion 1** for additional details of GI mucins and
109 bacterial growth on them.

110

111 **Substrate Specificity of *B.thetaiotaomicron* sulfatases**

112 Sulfatases are classified into four main families (S1 to S4) in the SulfAtlas
113 database according to sequence similarity, catalytic mechanism and fold²⁰. Family S1,
114 currently divided into 72 subfamilies (designated S1_X), comprises the formylglycine
115 sulfatases which operate via a hydrolytic mechanism that utilises a non-coded
116 formylglycine amino acid as a catalytic residue. In *Bt* and other anaerobic bacteria, this
117 residue is introduced co-translationally by the anaerobic sulfatase maturing enzyme
118 (anSME)¹³, which converts a serine or cysteine, within the consensus sequence **C/S-X-**
119 **P/A-S/X-R**, to formylglycine that then serves as the catalytic nucleophile²¹. The *Bt*
120 genome encodes 28 S1 sulfatases classified into twelve different subfamilies
121 (**Supplementary Table 1**). Four *Bt* sulfatases have been previously characterized and
122 all act on glycosaminoglycans (GAG) that are components of extracellular matrix and
123 glycocalyx (**Fig. 1c**)^{22,23}. Interestingly, several of the uncharacterized S1 sulfatases are
124 encoded within polysaccharide utilization loci (PULs) that are known to be upregulated *in*
125 *vivo* or during growth on gMO¹² and encode other glycoside hydrolases enzymes
126 potentially involved in degrading mucin O-glycans (**Extended Data Fig. 2**).

127 In order to understand the role of sulfatases in mucin metabolism, we cloned and
128 expressed in soluble form 23 of the remaining 24 uncharacterized sulfatases. The
129 recombinant proteins were tested for activity against a panel of commercially available
130 sulfated saccharides (**Supplementary Table 2**). In this initial screen, we identified
131 activities for twelve sulfatases (**Fig. 1c, Extended Data Figs. 3, 4 and Supplementary**
132 **Table 3**). Among these, 5 of the enzymes represent the first activities reported for their
133 respective subfamilies: two S1_20 members (BT1636 and BT1622) were determined to
134 target 3S-Gal, with BT1622 also preferentially cleaving 3S-*N*-acetyl-D-galactosamine (3S-
135 GalNAc), two S1_16 enzymes (BT3796 and BT3057) cleave 4S-Gal/4S-GalNAc and one
136 S1_46 enzyme (BT1918) cleaves 3S-GlcNAc, using the *N*-acetyl group as an absolute
137 specificity determinant. This represents the first reports of a bacterial sulfatase active on
138 3S-GalNAc, indicating that this sulfation could exist as a yet unidentified modification of
139 host glycans. Herein, we refer to these enzymes by their gene/locus tag number with the
140 corresponding activity in superscript (e.g., BT1636^{3S-Gal}).

141 In addition to discovering new catalytic activities associated with three subfamilies
142 that previously lacked defined members, we also identified sulfatases displaying novel
143 activities inside other subfamilies. These include three S1_15 enzymes (BT1624^{6S-}
144 Gal/GalNAc, BT3109^{6S-Gal/GalNAc} and BT4631^{6S-Gal/GalNAc}) that extend this family, previously
145 only known to include 6S-GalNAc sulfatases, to those cleaving 6S-Gal. Two members of
146 S1_4 were active on 3S-Gal (BT4683^{3S-Gal}) or 6S-Gal (BT3487^{6S-Gal}) representing novel
147 activities inside this arylsulfatase subfamily. Finally, consistent with the activity previously
148 described for S1_11 members, two enzymes were 6S-GlcNAc sulfatases (BT1628^{6S-}
149 GlcNAc and BT3177^{6S-GlcNAc}) (**Fig. 1c and Extended Data Fig. 3**).

150 Characterized sulfatases within the same family, with the exception of two S1_4
151 members, cleaved the same sulfonate linkages (**Fig. 1C**). However, despite these
152 enzymes targeting the same linkages, their optimal activity depends on the surrounding
153 glycan context. The activity of the 3S-Gal sulfatases is dependent on the linkage between
154 Gal and GlcNAc with BT4683^{3S-Gal} showing a preference for 3'-sulfate-*N*-acetyl-D-
155 lactosamine (3'S-LacNAc, 3'S-D-Gal-β1,4-D-GlcNAc). In contrast, BT1622^{3S-Gal/GalNAc}
156 demonstrated enhanced activity towards 3'-sulfate-lacto-*N*-biose (3'S-LNB, 3'S-D-Gal-
157 β1,3-D-GlcNAc) (**Extended Data Fig. 4a and Supplementary Table 3**). Furthermore,

158 additional affinity/activity studies revealed that BT1622^{3S-Gal/GalNAc} preferentially targets
159 GalNAc and not Gal (**Extended data Fig. 4b** and **Supplementary Table 3**), suggesting
160 that this sulfatase evolved to optimally target sulfate O3-linked to GalNAc. *Bt* sulfatase
161 activity is also affected by the presence of terminal epitopes such as those that occur in
162 Lewis antigens. Despite BT1636^{3S-Gal} being equally active on 3'S-LacNAc and 3'S-LNB,
163 this protein has a lower affinity and it is 100-fold less active when L-fucose (Fuc) is linked
164 to GlcNAc (3'S-Lewis^{a/x}) (**Extended data Fig. 4a and 4b** and **Supplementary Table 3**).
165 Whilst BT1622^{3S-Gal/GalNAc} is only weakly active on 3'S-Lewis^a antigen and not active at all
166 on 3'S-Lewis^x; the reciprocal is true for BT4683^{3S-Gal} (**Supplementary Table 3**). The
167 subfamily S1_15 enzyme BT1624^{6S-Gal/GalNAc} was only weakly active on 6S-Lewis
168 antigens (**Extended Data Fig. 4a**), suggesting that these enzymes cannot accommodate
169 Fuc linked to GlcNAc and that for both families Fuc needs to be removed prior to sulfate
170 cleavage. Additionally, affinity studies showed that BT3109^{6S-Gal/GalNAc} has a strong affinity
171 for Gal, whilst the previously characterized GAG sulfatase BT3333^{6S-GalNAc} showed a
172 preference for GalNAc, suggesting that optimal activity of S1_15 sulfatases likely
173 depends on the glycan context and BT3333^{6S-GalNAc} evolved to target sulfated linkages in
174 GAGs, a substrate that contains GalNAc but not Gal.

175

176 ***B.thetaiotaomicron* sulfatase activity on cMO**

177 We next tested the activity of *Bt* sulfatases on cMO, which we determined to
178 contain at least 131 different oligosaccharides (**Supplementary Table 5**). Only 4 of the 6
179 sulfatases tested displayed activity on cMOs (**Fig 2** and (**Supplementary Table 5**)).
180 BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc} removed 6-O-sulfate from all GlcNAc structures but
181 only when present at the non-reducing end of O-glycans confirming an exo-mode of
182 action observed in commercial oligosaccharides (**Fig 2** and **Extended Data Fig. 5a**).
183 After incubation with BT1636^{3S-Gal}, we were able to detect 14 new oligosaccharides and
184 an overall increase of non-sulfated structures (**Fig 2, Extended Data Fig. 5c**). Compared
185 to the non-enzyme treated control, 36 oligosaccharides could no longer be detected after
186 incubation with BT1636^{3S-Gal} (**Fig 2a** and **Supplementary Table 5**). We could only
187 determine the structures of 8 of these glycans, however, they all present a terminal 3S-
188 Gal (**Fig. 2**). This sulfatase was active on 3'S-Gal-β1,3-GalNAc (core 1) and more

189 complex sulfated structures built around other common core structures (**Fig. 2c**),
190 indicating that BT1636^{3S-Gal} evolved to accommodate the various linkages and
191 substitutions found in mucin O-glycans. After incubation with BT4683^{3S-Gal} we failed to
192 detect 9 oligosaccharides, indicating that this enzyme is active on a smaller subset of
193 structures (**Fig. 2a and Supplementary Table 5**). We were only able to determine the
194 structure of one of those glycans and, unexpectedly, we found that this enzyme is endo-
195 active on sialylated 3S-Gal (**Fig. 2b**), which is consistent with structural data discussed
196 below. Additionally, the strong activity of BT1636^{3S-Gal} on higher m/z glycans with
197 unknown structures (**Supplementary Table 5**) reveals that porcine cMOs are highly
198 sulfated at the O3 position of galactose. Although previous studies have reported the
199 presence of 3S-Gal in colonic mucins²⁴, the analysis of such complex samples is
200 technically challenging and, until now, no precise enzymatic tools were available to probe
201 these linkages. Together, these results represent the first report of HGM sulfatases active
202 on colonic mucin O-glycans and highlight the utility of gut bacterial sulfatases as analytical
203 tools in structural characterization of mucin glycans. See **Supplementary Discussion 2**
204 for additional details of sulfatase activity on cMO.

205

206 ***Structural characterization of 3S-Gal/GalNAc sulfatases***

207 To further understand the molecular details of carbohydrate recognition by S1
208 sulfatases, we determined the crystal structures of the three different 3S-Gal sulfatases
209 that belong to two different subfamilies. Consistent with previous structures of S_1
210 sulfatases, all 3 enzymes display a $\alpha/\beta/\alpha$ topology with a C-terminal sub-domain, and the
211 active site residues interacting with the sulfate group are fully conserved (**Extended data**
212 **Fig. 6a,b**). The structure of BT1636^{3S-Gal} in complex with the product LacNAc revealed
213 that His177 coordinates with O4 of D-Gal and mutation to alanine ablates enzyme activity,
214 suggesting that this residue is the major specificity determinant for Gal (presenting an
215 axial O4) over glucose (equatorial O4) (**Fig. 3 and Supplementary Table 3**). BT1636^{3S-}
216 ^{Gal} also makes strong interactions with O2 via R353 and E334. The essential His177 in
217 BT1636^{3S-Gal} is highly conserved (92%) within S1_20 sulfatases (**Extended data Figs. 7,**
218 **8**). In BT1622^{3S-Gal/GalNAc}, mutation of the corresponding H176 to alanine causes a ~300-
219 fold reduction in activity (**Supplementary Table 3**), further highlighting the importance of

220 this residue in Gal recognition. In BT1622^{3S-Gal/GalNAc}, R353 and E334 are replaced by
221 C357 and N334, shorter amino acids, that allow the accommodation of a C2-linked N-
222 acetyl group found in GalNAc (**Fig. 3**). This observation is consistent with the ability of
223 BT1622^{3S-Gal/GalNAc} to cleave 3S-GalNAc and preferentially bind GalNAc over Gal, whilst
224 BT1636^{3S-Gal} only recognizes Gal (**Extended Data Fig. 4b,c**). Additionally, consistent with
225 *exo*-activity observed for both enzymes the substrates are buried in a deep pocket and
226 only O1 is solvent exposed (**Fig. 3**).

227 For the S1_4 enzyme BT4683^{3S-Gal}, a structure solved in complex with LacNAc did
228 not reveal any interaction between the protein and the O4 of Gal. In BT4683^{3S-Gal}, the
229 interaction with Gal is driven by the residues R72 and E335, spatially equivalent to R353
230 and E334 in BT1636^{3S-Gal}, that form hydrogen bonds with O2 of D-Gal (**Fig. 3**) and
231 disruption of either of these residues eliminates activity (**Supplementary Table 3**).
232 However, in BT4683^{3S-Gal}, a sulfatase that does not have any affinity for monosaccharides
233 (**Extended data Fig. 4b**), the active site is located in an open cleft (**Fig. 3**) that allows the
234 accommodation of additional substitutions on Gal (**Extended Data Fig. 6d**), a finding that
235 is consistent with the *endo*-activity found in cMO. Together, these structures reveal the
236 key specificity determinants in 3S-Gal/GalNAc sulfatases, highlighting that these
237 enzymes have evolved to target sulfate groups in the different contexts in which they are
238 found in complex host glycans. This is especially true for BT1636^{3S-Gal} which utilises high
239 affinity interactions with both O2 and O4 to drive enhanced activity to remove terminal
240 3S-Gal linkages in cMOs. See **Supplementary Discussion 3** for additional details of
241 sulfatase structural characterization and phylogeny.

242

243 ***Roles of sulfatases in B. thetaiotaomicron O-glycan utilization***

244 *Bt* is able to utilize cMO as a sole carbon source (**Fig. 1b**), but the key enzymes
245 involved in the degradation of these glycans remain unclear. Deletion of the gene
246 encoding the only anaerobic sulfatase maturing enzyme (*anSME*) eliminates activation
247 of the 28 family S_1 sulfatases¹³ and the ability of *Bt* to grow efficiently on cMOs (**Fig.**
248 **4a**), highlighting the importance of sulfatases to this symbiont's physiology. Based on this
249 observation we generated a series of compounded gene deletion strains in which one or
250 more groups of sulfatases were eliminated based on their biochemical activity. Deletion

251 of all 3S-Gal/GalNAc sulfatases ($\Delta bt1636^{3S-Gal} + \Delta bt1622^{3S-Gal/GalNAc} + \Delta bt4683^{3S-Gal}$)
252 resulted in a growth phenotype similar to $\Delta anSME$ (**Extended Data Fig. 8a**). Interestingly,
253 we observed a similar growth defect when just BT1636^{3S-Gal} was deleted, but not the other
254 3S-Gal sulfatases (**Fig. 4a and Extended Data Fig. 8a**), consistent with the prominent
255 activity of the recombinant form of this enzyme on cMOs. In contrast, a strain with
256 compounded deletions of eight other sulfatases ($\Delta bt1622^{3S-Gal/GalNAc} + \Delta bt4683^{3S-Gal} +$
257 $\Delta bt1624^{6S-Gal/GalNAc} + \Delta bt3109^{6S-Gal} + \Delta bt4631^{6S-Gal/GalNAc} + \Delta bt1628^{6S-GlcNAc} + \Delta bt3177^{6S-}$
258 $GlcNAc + \Delta bt3051^{putative_6S-GlcNAc}$) displayed a growth phenotype similar to wild-type (**Fig.**
259 **4a**), indicating that these enzymes are not essential for cMO utilization. However, a $\Delta 10X$
260 *sulf* mutant, which included the deletion of BT1636^{3S-Gal} and the two 4S-Gal/GalNAc
261 sulfatases showed a similar growth defect as $\Delta anSME$ and $\Delta bt1636$. Complementation
262 of this and other loss of function mutants with only *bt1636*^{3S-Gal} restored growth on cMO
263 to levels similar to wild-type (**Fig. 4a, Extended Data Fig. 8a**). Cellular localization
264 experiments revealed that BT1636^{3S-Gal} is located at the cell surface in *Bt* (**Fig. 4c**) and
265 together, these data suggest that a single 3S-Gal sulfatase, BT1636^{3S-Gal}, is a critical cell
266 surface enzyme for the utilization of sulfated O-glycans that are prominent in the colon.

267 To further investigate the role of BT1636^{3S-Gal} in O-glycan utilization, we analysed
268 the oligosaccharides present in the culture supernatant of the wild-type and $\Delta bt1636^{3S-Gal}$
269 strains after growth on cMO. Consistent with a robust ability to degrade diverse colonic
270 O-glycans, no oligosaccharides were detected in wild-type supernatant (**Supplementary**
271 **Table 8**). Compared to the control, the supernatant of $\Delta bt1636^{3S-Gal}$ showed accumulation
272 of terminal 3S-Gal capped glycans (62% vs 3%, respectively). Indeed, the three most
273 common 3S-Gal structures detected in the $\Delta bt1636^{3S-Gal}$ supernatant accounted for 50%
274 of the total oligosaccharides detected in this sample (**Fig. 4b**). These results reveal that
275 deletion of BT1636^{3S-Gal} results in loss of the ability to utilize 3S-Gal O-glycans.
276 Interestingly, 49 of the 72 glycans detected in $\Delta bt1636^{3S-Gal}$ supernatant were not present
277 in the control sample (**Supplementary Table 8**), suggesting that although the mutant
278 does not grow well on cMO it is able to modify these oligosaccharides to create new
279 structures (**Supplementary discussion 5**). These data, combined with the cell surface
280 location of BT1636^{3S-Gal}, suggest that this sulfatase works early in O-glycan catabolism

281 likely by cleaving 3S-Gal from O-glycans prior to importing them into the periplasm where
282 these oligosaccharides will be sequentially degraded by additional enzymes and also
283 serve as cues for activating transcription of the O-glycan PULs (**Extended Data Fig. 2**).
284 Although *Bt* does encode two additional 3S-Gal/GalNAc sulfatases, the low activity of
285 these additional sulfatases on cMOs (**Fig. 2**) and their likely periplasmic location,
286 suggests why these enzymes cannot compensate for loss of BT1636^{3S-Gal}. Interestingly,
287 all of the *Bacteroides* species able to utilize cMO (**Fig. 2a**) have homologues of BT1636^{3S-}
288 Gal sulfatase, suggesting that this activity has a key role in mucin utilization by other HGM
289 members (**Supplementary table 10**).

290 Finally, to investigate the requirement for specific sulfatases *in vivo*, we utilized
291 gnotobiotic mice in which we competed individual mutants against the wild-type strain to
292 evaluate their colonization fitness. It has been reported that mouse colonic Muc2
293 prominently displays 6S-GlcNAc modifications²⁵. However, mutants lacking either just the
294 two active 6S-GlcNAc sulfatases ($\Delta 6S\text{-GlcNAc}$ double mutant), or these two enzymes,
295 another putative 6S-GlcNAc sulfatase and all three 6S-Gal/GalNAc sulfatases ($\Delta 6S\text{-}$
296 *GlcNAc* + $\Delta 6S\text{-Gal/GalNAc}$ hexa mutant), competed equally with wild-type (**Extended**
297 **Data Fig. 8c**), suggesting that neither of these two sulfatase activities are essential
298 determinants *in vivo*. An overall mild defect was observed with a mutant lacking all 3S-
299 Gal/GalNAc sulfatases although this effect was severe in one mouse (**Fig. 4d**). The
300 fitness defect was exacerbated by eliminating 3S-Gal/GalNAc and 6S-GlcNAc sulfatase
301 activities together ($\Delta 3S\text{-Gal/GalNAc}$ + $\Delta 6S\text{-GlcNAc}$) (**Fig. 4d**), suggesting that they
302 synergise *in vivo*. Consistent with its prominent role in cMO utilization *in vitro*, a mutant
303 lacking just BT1636^{Gal-3Sulf} displayed a strong, but variable defect when competed with
304 the wild-type strain (**Fig. 4d**), further suggesting that this enzyme plays an essential role
305 in gut colonization by allowing *Bt* to access 3S-Gal O-glycans. See **Supplementary**
306 **Discussion 4-5** for additional details regarding growth and competition of mutants *in vitro*
307 and *in vivo*.

308

309 **Conclusion**

310 To degrade the complex O-glycans found in mucins some HGM bacteria have
311 evolved complex arsenals of degradative enzymes that include sulfatases. Disarming all

312 of the sulfatases in *Bt* via *anSME* deletion results in drastically reduced competitive
313 colonization (**Extended Data Fig. 8c**)¹³ and an inability to elicit colitis in an animal model
314 of IBD¹⁴. While these findings support a critical role for active sulfatases in both fitness
315 and promoting inflammation, they provide no insight into the complexity of catalytic events
316 carried out by these enzymes. In this study, we reveal that *Bt* has a robust ability to grow
317 on highly sulfated mucin oligosaccharides from colonic tissue and that it has sulfatases
318 to remove sulfate in all of the contexts in which it is known to occur in mucin, and perhaps
319 also unknown linkages. Surprisingly, we found that a single key sulfatase is essential for
320 growth on colonic mucin O-glycans. The critical role of BT1636^{3S-Gal} supports the
321 conclusion that keystone steps exist in the complex pathway of mucin degradation,
322 although, given that mucin glycan structures may vary between mammalian hosts, these
323 critical steps may eventually need to be validated in humans for therapeutic translation.
324 Nevertheless, establishment of these critical steps is a prerequisite to blocking this
325 complex enzyme pathway and potentially inhibiting mucin-degrading activities in bacteria
326 that contribute to diseases such as IBD.

327

328 **Methods**

329

330 **Recombinant Protein Production**

331 Genes were amplified by PCR using the appropriate primers and the amplified DNA
332 cloned in pET28b using NheI/XhoI restriction sites or pETite (ExpressoTM T7 cloning and
333 expression system, Lucigen) generating constructs with either N- or C-terminal His₆ tags
334 (**Supplementary Table 11**). The catalytic serine was mutated to cysteine since
335 *Escherichia coli* is only able to convert cysteine to formylglycine. Recombinant genes
336 were expressed in *Escherichia coli* strains BL21 (DE3) or TUNER (Novagen), containing
337 the appropriate recombinant plasmid, and cultured to mid-exponential phase before
338 induction with 1 mM (BL21) or 0.2 mM (TUNER) of isopropyl β-D-1-
339 thiogalactopyranoside; cells were cultured for another 16 h at 16 °C and 180 rpm.
340 Recombinant protein were purified to >90% electrophoretic purity by immobilized metal
341 ion affinity chromatography using a cobalt-based matrix (Talon, Clontech) and eluted with
342 imidazole as described previously²². For the proteins selected for structural studies,

343 another step of size exclusion chromatography was performed using a Superdex 16/60
344 S200 column (GE Healthcare), with 10 mM HEPES, pH 7.5, and 150 mM NaCl as the
345 eluent, and they were judged to be $\geq 95\%$ pure by SDS-PAGE. Protein concentrations
346 were determined by measuring absorbance at 280 nm using the respective molar
347 extinction coefficient. When necessary, proteins were then concentrated in centrifugal
348 concentrators with a molecular mass cutoff of 30 kDa.

349

350 **Site-Directed Mutagenesis**

351 Site-directed mutagenesis was conducted using the PCR-based QuikChange kit
352 developed by Stratagene and conducted according to the manufacturer's instructions,
353 using the appropriate plasmid as the template, and appropriate primer (**Supplementary**
354 **Table 12**). The mutations were confirmed by DNA sequencing.

355

356 **Sources of Carbohydrates**

357 All carbohydrates were from Sigma, Carbosynth or Dextra Laboratories. All other
358 chemical reagents were purchased from Sigma. The 3S-GalNAc was chemically
359 synthesized as previously described²⁶.

360

361 **Mucin purification**

362 Gastric mucin oligosaccharides (gMO) were purified from porcine gastric mucins (type III,
363 Sigma) as previously described. Colonic mucins oligosaccharides (cMO) were purified
364 from pig distal pig colons and rectum. Briefly, the tissue was open and the fecal contents
365 were carefully removed. The mucosa was scrapped off and mucus was extracted by
366 homogenizing the tissue in at least 5 times volume of extraction buffer (6M guanidine
367 chloride, 5mM EDTA, 10mM NaH₂PO₄, pH 6.5) and slow stirring at 4 °C for 16 h. The
368 solution was spun down at 15 000 rpm and 10 °C for 30 min and supernatant was
369 discharged. The pellets were resuspended in extraction buffer and the process was
370 repeated until the supernatant was clear for at least two extractions. After the extraction
371 the mucins were solubilized by reducing the disulfide bonds. The pellets were
372 resuspended in fresh reduction buffer (6 M guanidine chloride, 0.1M Tris, 5 mM EDTA,
373 pH 8.0) containing 25 mM of 1,4-dithiothreitol and slowly stirred at 37 °C for 5 h. After this

374 incubation, 62.5 mM of iodoacetamide were added and the solution was stirred slowly in
375 the dark at room temperature for 16 °C. The solution was centrifuged at 10 000 rpm at 4
376 °C for 30 min and the supernatant containing the solubilized mucins was extensively
377 dialysed into water. Samples were dissolved into 100 mM Tris-HCl pH 8.0 containing 1
378 mg/ml of trypsin and incubated slowly stirring at 37 °C for 16 h. The glycans were β -
379 eliminated by adding 0.1 M NaOH and 1 M NaBH₄ and incubate the solution at 65 °C for
380 18 h. After cooling the solution to the room temperature the pH was adjusted to 7.0 with
381 concentrated HCl and extensively dialysed in water. The released porcine colonic mucin
382 glycans were recovered by lyophilisation the solution until completely dry and used in
383 further experiments.

384

385 **Sulfatase enzymatic assays**

386 The sulfatase activity screen against commercially available sulfated oligosaccharides
387 (Supplementary Table 2) was performed with 1 μ M of recombinant enzyme and 1 mM of
388 substrate in 10 mM MES pH6.5 with 5 mM CaCl₂ for 16h at 37 °C. Sulfated *N*-acetyl-D-
389 lactosamine and lacto-*N*-biose were generated by incubating the respective sulfated
390 Lewis antigens with 1 μ M of α -1,3/1,4-fucosidase BT1625²⁷ in the same conditions.
391 Reactions were analysed by thin layer chromatography (TLC). Briefly, 2 μ L of each
392 sample was spotted onto silica plates and resolved in butanol:acetic acid:water (2:1:1)
393 running buffer. The TLC plates were dried, and the sugars were visualized using
394 diphenylamine stain (1 ml of 37.5% HCl, 2 ml of aniline, 10 ml of 85% H₃PO₃, 100 ml of
395 ethyl acetate and 2 g diphenylamine) and heated at 100°C for 20 minutes. When relevant,
396 the enzymatic activity was confirmed by high-performance anionic exchange
397 chromatography (HPAEC) with pulsed amperometric detection using standard
398 methodology. The sugars (reaction substrate/products) were bound to a Dionex
399 CarboPac P100 column and eluted with an initial isocratic flow of 10 mM NaOH during 20
400 minutes then a gradient of 10-100 mM of NaOH for 20 minutes at a flow rate of 1.0 ml
401 min⁻¹. The reaction products were identified using the appropriated standards. All
402 experiments were performed in triplicate.

403

404 **Liquid Chromatograph-Electrospray Ionization Tandem Mass Spectrometry**

405 Enzymatic reactions of sulfatases in colonic mucin oligosaccharides and culture
406 supernatant were cleaned up with graphitized. Reactions with sulfated defined
407 saccharides (Lewis antigens and disaccharides) were reduced and desalted. Briefly,
408 reactions were dried in Speed vac, reconstituted in 20 μ L of 50 mM NaOH and 500 mM
409 NaBH₄ and incubated at 50 °C for 3 h. Reactions were cool down on ice, neutralised with
410 1 μ L of glacial acetic acid and desalted using a cation exchange column containing
411 AG[®]50W-X8 resin. All cleaned and desalted reactions were reconstituted in water before
412 analysis by liquid chromatograph-electrospray ionization tandem mass spectrometry (LC-
413 ESI/MS). The oligosaccharides were separated on a column (10 cm \times 250 μ m) packed
414 in-house with 5 μ m porous graphite particles (Hypercarb, Thermo-Hypersil, Runcorn,
415 UK). The oligosaccharides were injected on to the column and eluted with an acetonitrile
416 gradient (Buffer A, 10 mM ammonium bicarbonate; Buffer B, 10 mM ammonium
417 bicarbonate in 80% acetonitrile). The gradient (0-45% Buffer B) was eluted for 46 min,
418 followed by a wash step with 100% Buffer B, and equilibrated with Buffer A in next 24
419 min. A 40 cm \times 50 μ m i.d. fused silica capillary was used as transfer line to the ion source.
420 Samples were analyzed in negative ion mode on a LTQ linear ion trap mass spectrometer
421 (Thermo Electron, San José, CA), with an IonMax standard ESI source equipped with a
422 stainless steel needle kept at -3.5 kV. Compressed air was used as nebulizer gas. The
423 heated capillary was kept at 270°C, and the capillary voltage was -50 kV. Full scan (*m/z*
424 500-1800 for acidic GSLs and *m/z* 380-2000 for neutral GSLs, two microscan, maximum
425 100 ms, target value of 30,000) was performed, followed by data-dependent MS² scans
426 (two microscans, maximum 100 ms, target value of 10,000) with normalized collision
427 energy of 35%, isolation window of 2.5 units, activation *q*=0.25 and activation time 30
428 ms). The threshold for MS² was set to 300 counts. Data acquisition and processing were
429 conducted with Xcalibur software (Version 2.0.7). The schematic glycosidic or cross-ring
430 cleavages were assigned according to the Domon and Costello nomenclature²⁸.

431

432 **Microfluidics desulfation assays**

433 Sulfated carbohydrates were labelled at their reducing end with BODIPY which has a
434 maximal emission absorbance of ~503 nm, which can be detected by the EZ Reader via
435 LED-induced fluorescence. Non-radioactive microfluidic mobility shift carbohydrate

436 sulfation assays were optimised in solution with a 12-sipper chip coated with CR8 reagent
437 and a PerkinElmer EZ Reader II system using EDTA-based separation buffer and real-
438 time kinetic evaluation of substrate de-sulfation. Pressure and voltage settings were
439 adjusted manually (1.8 psi, upstream voltage:2250 V, downstream voltage: 500 V) to
440 afford optimal separation of the sulfated and unsulfated product with a sample (sip) time
441 of 0.2 s, and total assay times appropriate for the experiment. Individual de-sulfation
442 assays were carried out at 28°C and assembled in a 384-well plate in a volume of 80 µl
443 in the presence of substrate concentrations between 0.5 and 20 µM with 100 mM Bis-
444 Tris-Propane or Tris, as pH appropriate, 150 mM NaCl, 0.02% (v/v) Brij-35 and 5 mM
445 CaCl₂. The degree of de-sulfation was directly calculated using the EZ Reader software
446 by measuring the sulfated carbohydrate : unsulfated carbohydrate ratio at each time-point.
447 The activity of sulfatase enzymes was quantified in 'kinetic mode' by monitoring the
448 amount of unsulfated glycan generated over the assay time, relative to control assay with
449 no enzyme; with sulfate loss limited to ~20% to prevent of substrate and to ensure assay
450 linearity. k_{cat}/K_M values, using the equation $V_0=(V_{max}/K_M)/S$, were determined by linear
451 regression analysis with GraphPad Prism software. Substrate concentrations were
452 halved and doubled to check linearity of the rates ensuring substrate concentrations were
453 significantly $<K_M$.

454

455 **NMR desulfation assays**

456 NMR experiments, monitoring the de-sulfation of 6S-D-galactose and 6S-N-acetyl-D-
457 galactosamine, were conducted in D₂O with 50 mM sodium phosphate, pH 7.0,
458 supplemented with 150 mM NaCl at 25°C on a 800MHz Bruker Avance III spectrometer
459 equipped with a TCI CryoProbe and a 600MHz Bruker Avance II+ spectrometer, also
460 fitted with a TCI CryoProbe. 1D and 2D proton and TOCSY spectra (mixing time 80 ms)
461 were measured using standard pulse sequences provided by the manufacturer. Spectra
462 were processed and analysed using TopSpin 3.4A and TopSpin 4.0 software (Bruker).
463 Galactose integrals were recorded directly for the C(6)H₂-OH peak within the region 3.694
464 to 3.721ppm, referenced to the combined C(2) peaks of D-galactose and 6S-D-galactose
465 with in the region 3.415 to 3.475ppm. Similarly, 6S-N-acetyl-D-galactosamine integrals

466 were recorded directly for the C(6)H₂-OH peak within the region 3.674 to 3.747ppm,
467 referenced to the combined C(4) peaks for N-acetyl D-galactosamine and 6S-N-acetyl D-
468 galactosamine in the region 3.925 to 3.968ppm.

469

470 **Differential scanning fluorimetry**

471 Thermal shift/stability assays (TSAs) were performed using a StepOnePlus Real-Time
472 PCR machine (LifeTechnologies) using SYPRO-Orange dye (emission maximum 570
473 nm, Invitrogen), with thermal ramping between 20 and 95°C in 0.3°C step intervals per
474 data point to induce denaturation in the presence or absence of various carbohydrates
475 as appropriate to the sulfatase being analysed. The melting temperature (T_m)
476 corresponding to the midpoint for the protein unfolding transition was calculated by fitting
477 the sigmoidal melt curve to the Boltzmann equation using GraphPad Prism, with R² values
478 of >0.99. Data points after the fluorescence intensity maximum were excluded from fitting.
479 Changes in the unfolding transition temperature compared with the control curve (ΔT_m)
480 were calculated for each ligand. A positive ΔT_m value indicates that the ligand stabilises
481 the protein from denaturation, and therefore binds the protein. All TSA experiments were
482 conducted using a protein concentration of 5μM in 100 mM Bis-Tris-Propane, pH 7.0, and
483 150 mM NaCl supplemented with appropriate ligand. When BT1622^{GalNAc3S} and
484 BT1636^{Gal3S} were assessed against 3S-LacNAc and 3S-LNB 100 mM Hepes was used
485 instead of BTP, no difference in the T_m value of the proteins was observed. Three
486 independent assays were performed for each protein and protein ligand combination.

487

488 **Glycan labelling**

489 Sulfated saccharide samples were labelled according to a modification of the method
490 previously described reporting the formation of N-glycosyl amines for 4,6-O-benzilidene
491 protected D-gluopyranose monosaccharides with aromatic amines²⁹. Briefly, the
492 lyophilised sugar (1 mg) was dissolved in 0.50 ml anhydrous methanol in a 1.5 ml screw-
493 top PTFE microcentrifuge tube. 0.1 mg, BODIPY-FL hydrazide (4,4-difluoro-5,7-
494 dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid hydrazide, ThermoFischer, λ
495 ex./em. 493/503, extinction coefficient 80,000 M⁻¹ cm⁻¹) was added and the mixture
496 vortexed (1 min), then incubated in darkness at 65 °C for 24. The products were then

497 cooled and a portion purified by TLC on silica coated aluminium plates developed with
498 methanol or 1:1 v/v ethyl acetate/methanol. The unreacted BODIPY-FL label (orange on
499 the TLC plate) was identified by reference to a lane containing the starting material
500 (BODIPY-FL hydrazide) on the TLC plate, allowing differentiation from the putative
501 labelled product (also orange). This latter band was scraped from the plates and extracted
502 in fresh methanol (2 x 0.5 ml), spun for 3 min at 13,000g, and the supernatant was
503 recovered and dried (rotary evaporator) to recover the fluorescent-coloured product
504 (bright green when dissolved in aqueous solution), which was then employed in
505 subsequent experiments.

506

507 **Anaerobically Bacterial Strains Culture and Genetic Manipulation**

508 All strains were anaerobic grown at 37 °C in a chamber (10% H₂, 5% CO₂, and 85% N₂;
509 Coy Manufacturing, Grass Lake, MI). *Bacteroides* type strains were culture in either
510 tryptone-yeast extract-glucose medium (TYG), brain heart infusion medium or minimal
511 medium (MM) containing an appropriate carbon source. *Bacteroides massilliensis* and
512 *Akkermansia muciniphila* were culture as described before^{19,30}. *Bt* strains containing
513 specific gene deletions or inactivated versions of enzymes (BT1636^{3S-Gal} S77A) were
514 made by counterselectable allelic exchange as previously described³¹. Complementation
515 of deletion strains was performed using pNBU2 vector as previously described³¹,
516 containing a constitutive promotor used previously³². All primers used to generate the
517 mutants and complementation are listed in **Supplementary table 13**. Growth of the WT
518 and mutants was measured on an automated plate reader by increase in absorbance at
519 600 nm in 96-well plates containing 200 µl of minimal media mixed with the respective
520 filter-sterilised (monosaccharide and gMO) or autoclave-sterilised cMO as described
521 before³⁰. To achieve consistent growth, all carbon sources were used at 5 mg/ml with
522 exception of gMO that was added in a final concentration of 10 mg/ml. All growth curves
523 presented are averages and s.e.m of three technical replicates.

524 **Crystallization of carbohydrate sulfatases**

525 After purification, all proteins were carried forward in the same eluent as used for the size
526 exclusion chromatography (see Recombinant Protein Production). Sparse matrix screens
527 were set up in 96-well sitting drop TTP Labtech plates (400-nL drops). Initial hits crystals

528 for all proteins were obtained between 20 and 35 mg/ml. For BT1622^{3S-Gal/GalNAc} and
529 BT1636^{3S-Gal} wildtype *Bt* variants were used, having a Ser at the catalytic formylglycine
530 position, whilst for BT4683^{3S-Gal} the S73C mutant was used. BT1622^{3S-Gal/GalNAc} with
531 20 mM LNB crystallised in 20 % PEG 3350 and 0.2 M sodium citrate tribasic dihydrate.
532 BT1636^{3S-Gal} with 20 mM LacNAc crystallised in 40 % MPD and 0.2 M sodium cacodylate
533 pH 6.5. for BT4683^{3S-Gal} with 20 mM LacNAc crystallised in 20 % PEG 3350, 0.2 M sodium
534 iodide and BTP pH 8.5. All crystals were cryoprotected with the addition of the ligand they
535 were crystallised with plus 20% PEG 400 and 20% glycerol was used as the
536 cryoprotectant for BT4683^{3S-Gal} and BT1622^{3S-Gal/GalNAc}, respectively. No cryoprotectant
537 was added to BT1636^{3S-Gal} crystals. Data were collected at Diamond Light Source
538 (Oxford) on beamlines I03, I04, I04-1 and I24 at 100 K. The data were integrated with
539 XDS³³, or Xia2³⁴ 3di or 3dii and scaled with Aimless³⁵. Five percent of observations were
540 randomly selected for the R_{free} set. The phase problem was solved by molecular
541 replacement using the automated molecular replacement server Balbes³⁶ for all proteins
542 except BT1622^{3S-Gal/GalNAc}. The phase problem for BT1622^{3S-Gal/GalNAc} was initially solved
543 using Molrep³⁷ and BT1636^{3S-Gal} as the search model. This gave a partial solution, which
544 could not be fully solved due to twinning. An acceptable model of BT1622^{3S-Gal/GalNAc} was
545 constructed to be used to better solve the phase problem and the molecular replacement
546 was re-performed. Models underwent recursive cycles of model building in Coot³⁸ and
547 refinement cycles in Refmac5³⁹. Bespoke ligands were generated using JLigand⁴⁰. The
548 models were validated using Coot³⁸ and MolProbity⁴¹. Structural figures were made using
549 Pymol and all other programs used were from the CCP4⁴² and CCP4i2 suite⁴³. The data
550 processing and refinement statistics are reported in **Supplementary Table 14**.

551

552 **Ligand docking**

553 The docking of 3-sulfate-*N*-Acetyl-D-galactosamine (3S-GalNAc) into BT1622^{GalNAc3S} was
554 performed using the automated Swissdock server with the default settings^{44,45}. A pdb and
555 cif file for the ligand 3S-GalNAc were generated using JLigand⁴⁰ and then the pdb file was
556 converted to a mol2 file using openbabel for use with Swissdock. Clustered solutions were
557 inspected using UCSF chimera⁴⁶ and then the solutions which clustered appropriately to

558 the active site pocket were manually converted to individual pdb files and further
559 inspected using PyMol.

560

561 **Immunolabelling of BT1636 in *Bt* cell surface**

562 For the fluorescence microscopy, *Bt* cells (Wild type (Δtdk) and $\Delta bt1636^{3S-Gal}$) were grown
563 to early exponential phase (Abs_{600nm} 0.25–0.35) in rich TYG medium. One ml of the
564 cultures was collected, centrifuged at 13,000 RPM, and subsequently washed three times
565 in MM with no carbon source. *Bt* cells incubated with cMO for four hours and fixed in 4.5%
566 formalin overnight at 4 °C with gentle rocking. Cells were stained with a polyclonal
567 antibody raised in rabbit against purified recombinant BT1636 (BT1636^{Ab}, Cocalico
568 Biologicals) and detected with an Alexa Fluor® 488-conjugated goat anti-rabbit IgG
569 secondary antibody (Molecular Probes). Images were taken with Zeiss Apotome using
570 the same exposure time between samples.

571

572 **Gnotobiotic Mouse Experiments**

573 All experiments involving animals, including euthanasia via carbon dioxide asphyxiation,
574 were approved by the University Committee on Use and Care of Animals at the University
575 of Michigan (NIH Office of Laboratory Animal Welfare number A3114-01) and overseen
576 by a veterinarian. Groups of 3 to 5, 6-8 week old germfree Swiss Webster mice were
577 randomly assigned to each experiment. 7 days prior gavage the animals diet was
578 switched to a fiber-free diet (Envigo-Teklad TD 130343) that was maintained through all
579 the experiment. At day 0, mice were gavage with equal amount of *Bt* WT strain and mutant
580 and fecal samples were collected at day 2 and every 5 days until day 42. At the end-point
581 of the experiment distal small intestine and cecca contents were also collected. The
582 bacteria gDNA extraction and quantification by qPCR of the relative abundance of each
583 strain on the various samples was carried out as described previously³⁰.

584

585

586 **Phylogenetic analysis**

587 To avoid identical sequences, we selected 800 and 920 representative sequences of
588 subfamily S1_20 (composed of 1356 sequences) and S1_4 (composed of 1895

589 sequences), respectively. The sequences were aligned by MAFFT v.7⁴⁷ using L-INS-i
590 algorithm. The multiple sequence alignment was visualized by Jalview software v.11.0⁴⁸
591 and non-aligned regions were removed. 404 and 364 positions were used for the S1_4
592 and S1_20 phylogeny, respectively. Phylogeny was made using RAxML v. 8.2.4⁴⁹. The
593 phylogenetic tree was build with the Maximum Likelihood method⁵⁰ and the LG matrix as
594 evolutive model⁵¹ using a discrete Gamma distribution to model evolutionary rate
595 differences among sites (4 categories). The rate variation model allowed for some sites
596 to be evolutionarily invariable. The reliability of the trees was tested by bootstrap analysis
597 using 1,000 resamplings of the dataset⁵². All the final global phylogenetic trees were
598 obtained with MEGA v.7⁵³. Fifteen S1_0 sequences from the sulfAtlas database were
599 used as outgroups.

600

601 **Quantification and Statistical Analysis**

602 For in vivo competitions, when three or more fecal samples were collected, Student's t
603 tests were performed for each time point in GraphPad Prism version 8.1 with a paired,
604 two-tailed distribution. When necessary, the statistical analysis for remaining samples in
605 stated in the respective figure legend.

606

607 **Data availability statement**

608 All data for the experiments, along with corresponding statistical test values, where
609 appropriate, are provided within the paper and in its Supplementary Information. The
610 crystal structure datasets generated have been deposited in the PDB under the following
611 accession numbers: 7ANB, 7ANA, 7AN1, 7OQD and 7ALL. The MS raw files have been
612 deposited in the GlycoPOST database under the following IDs: GPST000150 and
613 GPST000196. Glycan structural annotations were deposited to the UniCarb database at
614 <https://unicarb-dr.glycosmos.org/references/462>. There are no restrictions on data or
615 biological resource availability. Data and biological resources can be obtained by
616 contacting the corresponding authors. Source data are provided with this paper.

617 **Code Availability Statement:** No new codes were developed or compiled in this study

618

619 **Competing Interests Statement**

620 The authors declare no competing interests.

621

622 **Acknowledgements**

623 This project has received funding from the European Union's Horizon 2020 research and
624 innovation programme under the Marie Skłodowska-Curie grant agreement N° 748336.

625 The authors acknowledge access to the SOLIEL and Diamond Light sources via both the
626 University of Liverpool and Newcastle university BAGs (proposals mx21970 and
627 mx18598, respectively). We thank the staff of DIAMOND, SOLIEL, and members of the
628 Liverpool's Molecular biophysics group for assistance with data collection. This work was
629 supported by National Institutes of Health grants DK118024 and DK125445 awarded to
630 ECM.

631

632 **Author Contributions.** A.S.L., A.C. and E.C.M. designed experiments and wrote the
633 manuscript. A.S.L. and A.C. cloned, expressed and purified sulfatases and performed the
634 enzymatic assays. A.C., D.P.B., J.A.L. and P.A.E. carried out and analysed the data from
635 kinetic and binding experiments. E.A.Y., M.R. and S.O. performed chemical syntheses.
636 A.C. and A.B. performed structural biology experiments. C.J., A.S.L., G.C.H. and N.G.K.
637 performed and interpreted data from analytical glycobiology experiments. A.S.L., G.V.P.,
638 R.W.P.G., S.G., S.S. and N.A.P. performed bacterial growth experiments and analysed
639 in vivo competition data. M.C., G.M. and T.B. performed sulfatase phylogenetic analyses.
640 All authors read and approved the manuscript.

641

642 **Additional information**

643 Supplementary information The online version contains supplementary material available
644 at <https://doi.org/10.1038/s41586-021-03967-5>.

645 Correspondence and requests for materials should be addressed to Ana S. Luis, Alan
646 Cartmell or Eric C. Martens.

647

648 **Figure Legends**

649

650 **Figure 1. Bacterial growth on colonic mucin and *B. thetaiotaomicron* sulfatases**
651 **activities.** **a**, Schematic representation of mucin O-glycans and relevant terminal
652 epitopes (dashed box). Sugars are shown according to the Symbol Nomenclature for
653 Glycans system⁵⁴. **b**, Growth of *Bacteroides* type strains and *Akkermansia muciniphila*
654 on colonic mucin O-glycans (cMO) and number of respective encoded S1 sulfatases. The
655 bars represent the average of two independent experiments with different batches of
656 cMO. Bacteria species able to utilize gastric mucin glycans are highlighted in blue.
657 Maximum absorbance is the difference of the maximum absorbance value (Abs_{600nm}) for
658 each culture and the initial absorbance at time 0 (T_{0h}). Graphic shows the example of
659 growth curves for *B. fragilis* (Bf), *B. thetaiotaomicron* (Bt), *B. vulgatus* (Bv) and *B.*
660 *massiliensis* (Bm). **c**, Phylogeny of Bt sulfatases with S1 sulfatases and respective
661 substrate. Enzymes colour coded according the respective subfamilies. * indicates
662 sulfatase activity previously characterized and arrows point the substrate preferentially
663 targeted by the respective enzyme. Sulfatases sharing more than 86% and 39-58% of
664 sequence identity are highlighted in blue and green background, respectively. Data from
665 biological replicates n = 3 to 6 and error bars denote s.e.m..

666
667 **Figure 2. Activity of *B. thetaiotaomicron* sulfatases on colonic mucin O-glycans.** **a**,
668 Representation of O-glycans detected by mass spectrometry in cMO (control) and after
669 sulfatase treatment from the lower (top) to the higher m/z (bottom). **b**, Relative abundance
670 and putative structures for the specific m/z shown in panel a. Remaining structures are
671 shown on Extended figure 5a. **c**, Schematic representation of the putative structures that
672 were not detected in reaction after incubation with BT1636^{3S-Gal}.

673
674 **Figure 3. Crystal structures of 3S-Gal/GalNAc sulfatases.** **a**, Schematic
675 representation of the residues interacting with targeted sugars, including the putative
676 catalytic residues (in dark red), the calcium ion (grey sphere) and subsites S, 0 and +1
677 highlighted in red. BT1636^{3S-Gal} and BT4683^{Gal-3Sulf} in complex with LacNAc (D-Gal-β1,4-
678 D-GlcNAc) and BT1622^{3S-Gal/GalNAc} in complex with GalNAc. **b**, Surface representation of
679 the active pocket. The equivalent Gal/GalNAc specificity residues in BT1636^{3S-Gal} and
680 BT4683^{Gal-3Sulf} are highlighted in red and blue. The open active site of BT4683^{Gal-3Sulf} is

681 highlighted in purple. In all structures the amino acids and ligands are represented as
682 stick.

683

684 **Figure 4. BT1636^{3S-Gal} activity is required to the utilization of cMO and competitive**
685 **fitness *in vivo*.** **a**, Growth of *Bt* wild-type Δtdk (WT), different sulfatase gene-deletion
686 mutants (named " $\Delta btXXXX$ ") and strains complemented with *bt1636*^{3S-Gal} on colonic or
687 gastric mucin O-glycans (cMO and gMO, respectively) (line represents the average of
688 biological replicates (n = 3) and error bars denote s.e.m.) **b**, Immunofluorescent and
689 differential interference contrast (DIC) microscopy of *Bt* WT and sulfatase mutant staining
690 with polyclonal antibody (Ab) against BT1636^{3S-Gal} (green) and DNA staining with DAPI
691 (blue). **c**, Relative abundance of different O-glycans structures detected by mass
692 spectrometry in $\Delta bt1636$ ^{3S-Gal} culture supernatant or cMO in minimal media without
693 bacteria (control), after 96h in anaerobic conditions. The m/z and putative structure of the
694 3 more abundant structures in both samples are shown. **d**, *in vivo* competitions in
695 gnotobiotic mice (n = 3-4) fed fiber-free diet and inoculated with WT and mutants. The
696 fecal relative abundance of each strain determined along the experiment time course and
697 in small intestine and cecum at day 42. The relative abundance in each mice is
698 represented in the respective light colour. The error bars denote s.e.m.. Gal, D-Galactose;
699 GalNAc, N-acetyl-D-galactosamine; GlcNAc, N-acetyl-D-glucosamine.

700

701 **Extended Data 1. Growth of Bacteroides type strains and Akkermansia muciniphila**
702 **in different mucin O-glycans.** The graphics show the growth of strains able to utilize
703 O-glycans in minimal media containing the indicated carbon source (biological replicates
704 n =3, error bars denote the s.e.m.). cMO, colonic mucin O-glycans; gMO, gastric mucin
705 O-glycans; GlcNAc, N-acetyl-D-glucosamine.

706

707 **Extended Data 2. Schematic representation of polysaccharide utilization loci**
708 **(PULs) encoding sulfatases (sulf).** Genes are colour coded according with the
709 predicted function of the respective proteins. Glycoside hydrolases (GH) in known families

710 are indicated by GHXX or GH*, where XX and * indicates the respective family number
711 or non-classified, respectively.

712

713 **Extended Data 3. Enzymatic screen of *B. thetaiotaomicron* sulfatases against**
714 **sulfated monosaccharides.** Recombinant enzymes (1 μ M) were incubated with 1 mM
715 of substrate in 10 mM MES pH6.5 with 5 mM CaCl₂ for 16h at 37 °C. Reactions were
716 analysed by thin layer chromatography (left hand side) or HPAEC with pulsed
717 amperometric detection (right hand side). Control reactions without sulfatases were
718 carried in the same conditions. The standards in TLC and HPAEC are labelled on the left
719 hand side and top, respectively. The different panel represent activities found for
720 sulfatases targeting: **(a)** 4S-Gal/GalNAc; **(b)** 6S-Gal/GalNAc; **(c)** 6S-GlcNAc and; **(d)** 3S-
721 GlcNAc. The data shown is a representative from biological replicates (n = 3).

722

723 **Extended Data 4. Activity and affinity of sulfatases to targeted substrates. a,**
724 Recombinant enzymes (1 μ M) were incubated with 1 mM of substrate in 10 mM MES
725 pH6.5 with 5 mM CaCl₂ for 16h at 37 °C. Sulfated disaccharides were generated by
726 adding 1 μ M of an characterized α 1,3/1,4-fucosidase (BT1625) in the enzymatic reaction.
727 Control reactions without sulfatases were carried in the same conditions. Samples were
728 analysed by mass spectrometry and the intensity of the substrate and reaction products
729 was used to determine the relative abundance of these sugars after incubation with the
730 respective enzymes. **b,** Affinity studies looking at the effect of ligand binding on the
731 melting temperature of 3S and 6S-Gal sulfatases. All reactions were performed in 100
732 mM BTP, pH 7.0 with 150 mM NaCl. **c,** Activity of 3S-Gal/GalNAc sulfatases (10 μ M)
733 against 3S-GalNAc (10 mM). Reactions were performed in 10 mM Hepes, pH 7.0, with
734 150 mM NaCl and 5 mM CaCl₂. The data shown is a representative from biological
735 replicates (n = 3).

736

737 **Extended Data 5. Activity of *B. thetaiotaomicron* sulfatases against colonic mucin**
738 **O-glycans (cMO) detected by mass spectrometry. a,** Relative abundance of defined
739 oligosaccharides after incubation of cMO with different sulfatases or no enzyme (control).
740 The putative structure for the different m/z is shown of the right hand side of the graphic.

741 The reactions were performed with 1 μ M of enzyme and 0.5% cMO in 10 mM MES pH6.5
742 with 5 mM CaCl_2 for 16h at 37 °C. **b**, Relative abundance of structures detected in different
743 samples organized by sulfate-linkage (top panel) or presence of one or several sugar
744 substitution such as sulfate, sialic acid and fucose (bottom panel). The colour-coded bars
745 represent the relative abundance and the numbers the total number of the structures
746 containing the specific linkage/substitution. The complete data is shown in Supplementary
747 Table 5.

748
749 **Extended Data 6. Schematic representation of 3S-Gal/GalNAc sulfatases.** **a**, Cartoon
750 representation colour ramped from blue ($\alpha/\beta/\alpha$ N-terminal domain) to red (b-sheet C-
751 terminal domain). **b**, Overlay of the active site S residues of BT1636^{3S-Gal} (green)
752 BT1622^{3S-Gal/GalNAc} (blue) and BT4683^{3S-Gal} (pink). The putative catalytic residues are
753 shown in bold. The calcium ion is represented as a grey sphere and its polar interactions
754 indicated as dashed lines. **c**, Ligand density of maps for LacNAc in BT1636^{Gal-3sulf} and
755 BT4683^{Gal-3sulf}, and GalNAc in BT1622^{Gal-3sulf}, contoured at 1σ (0.33 e/A³, 0.37 e/A³ and
756 0.18e/A³, respectively); **d**, Docking of putative structures O-glycans targeted by
757 BT4683^{3S-Gal} using the LacNAc as reference point showing that this structure can
758 accommodate a sialic acid in -1 subsite and additional sugars in positive subsites (left
759 hand side). The docking sugars are schematic shown as sticks (middle panel) and
760 schematic represented inside the dashed box (right hand side). Using the LacNAc product
761 as an 'anchor' additional sugars were built in manually with Coot 0.9 and regularised to
762 low energy conformations.

763
764 **Extended Data 7. Phylogenetic tree of S1_20 and S1_4 sulfatases.** The radial tree
765 were constructed using the branched trees shown in Supplementary figures 2 and 3. For
766 clarity all labels and sequence accession codes have been omitted. Red filled circles
767 designate sequences from *B. thetaiotaomicron* sulfatases. The residue is written in black
768 without any attributes if present in the sequence, in grey and italics if the residue is
769 mutated to any type in that sequence, or to a specific residue type if given in brackets
770 right after. **a**, Radial representation of the phylogenetic tree constructed with
771 representative sequences of the sulfatase S1_4 subfamily. The colour code is given as a

772 pattern of presence or absence of the residues R72, E335 and W505, which are crucial
773 in substrate recognition by BT4683 (acc-code Q89YP8, coloured red). A grey X in italics
774 specifically designates that the residue W505 is absent in that sequence, and no obvious
775 orthologous residue can be found from the alignment. **b**, Radial representation of the
776 phylogenetic tree constructed with representative sequences of the sulfatase S1_20
777 subfamily. The colour code is given as a pattern of presence or absence of the residues
778 E100, Q173 H177, E334, R353, which are crucial in substrate recognition by BT1636
779 (acc-code Q8A789, coloured red). A grey X in italics specifically designates that the
780 residue E100 is absent in that sequence, and no obvious orthologous residue can be found
781 from the alignment.

782

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

784 **a**, Growth curves of *Bt* wild-type Δtdk (WT), different sulfatase mutants ($\Delta btXXX$) and
785 complemented strains on glucose, colonic or gastric mucin O-glycans (cMO and gMO,
786 respectively). The curves represent the average of biological replicates (n = 3) and the
787 error bars denote s.e.m.. **b**, Relative abundance of oligosaccharides detected by mass
788 spectrometry in culture supernatant of WT and $\Delta bt1636^{3S-Gal}$ after growth in cMO for 96h
789 at anaerobic conditions. The control correspond to cMO incubated in the same conditions
790 without bacterium. The colours represent the relative abundance of structures grouped
791 according the presence of epitopes (sulfate, fucose and sialic acid) and the numbers
792 represent the total number of structures that contain the respective substitution. **c**,
793 Colonization of gnotobiotic mice fed with fiber-free diet by *Bt* WT and mutants lacking the
794 full ($\Delta anSME$) or specific sulfatase activity ($\Delta 6S-GlcNAc$ and $\Delta 6S-GlcNAc+\Delta 6S-$
795 $Gal/GalNAc$). The fecal relative abundance of each strain was determined in regular
796 intervals until day 42. The relative abundance of time 0 represents the abundance in
797 gavaged inoculum. At the experimental endpoint the relative abundance was also
798 determined in small intestine and cecum. The graphics represent the average of n =3 and
799 the error bars denote the s.e.m.. The relative abundance in each animal is represented
800 in a lighter colour in each of the respective graphics.

801

802

803 **References**

- 804
- 805 1 Packey, C. D. & Sartor, R. B. Commensal bacteria, traditional and opportunistic
806 pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Current*
807 *opinion in infectious diseases* **22**, 292-301 (2009).
- 808 2 Sears, C. L. & Garrett, W. S. Microbes, Microbiota, and Colon Cancer. *Cell Host*
809 *Microbe* **15**, 317-328, doi:10.1016/j.chom.2014.02.007 (2014).
- 810 3 Johansson, M. E., Larsson, J. M. & Hansson, G. C. The two mucus layers of colon are
811 organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial
812 interactions. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4659-4665,
813 doi:10.1073/pnas.1006451107 (2011).
- 814 4 Van der Sluis, M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating
815 that Muc2 is critical for colonic protection. *Gastroenterology* **131**, 117-129, doi:Doi
816 10.1053/J.Gastro.2006.04.020 (2006).
- 817 5 Velcich, A. *et al.* Colorectal cancer in mice genetically deficient in the mucin Muc2.
818 *Science* **295**, 1726-1729, doi:10.1126/science.1069094 (2002).
- 819 6 Bergstrom, K. *et al.* Core 1- and 3-derived O-glycans collectively maintain the colonic
820 mucus barrier and protect against spontaneous colitis in mice. *Mucosal immunology*,
821 doi:10.1038/mi.2016.45 (2016).
- 822 7 Bergstrom, K. *et al.* Core 1- and 3-derived O-glycans collectively maintain the colonic
823 mucus barrier and protect against spontaneous colitis in mice. *Mucosal Immunol* **10**, 91-
824 103, doi:10.1038/mi.2016.45 (2017).
- 825 8 Fu, J. *et al.* Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in
826 mice. *J Clin Invest* **121**, 1657-1666, doi:10.1172/JCI45538 (2011).
- 827 9 Arike, L. & Hansson, G. C. The Densely O-Glycosylated MUC2 Mucin Protects the
828 Intestine and Provides Food for the Commensal Bacteria. *J Mol Biol* **428**, 3221-3229,
829 doi:10.1016/j.jmb.2016.02.010 (2016).
- 830 10 Holmen Larsson, J. M., Thomsson, K. A., Rodriguez-Pineiro, A. M., Karlsson, H. &
831 Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. III.
832 Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific
833 distribution. *Am J Physiol Gastrointest Liver Physiol* **305**, G357-363,
834 doi:10.1152/ajpgi.00048.2013 (2013).
- 835 11 Robbe, C. *et al.* Evidence of regio-specific glycosylation in human intestinal mucins:
836 presence of an acidic gradient along the intestinal tract. *J Biol Chem* **278**, 46337-46348,
837 doi:10.1074/jbc.M302529200 (2003).
- 838 12 Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal Glycan Foraging Enhances Fitness
839 and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host Microbe*
840 **4**, 447-457, doi:DOI 10.1016/j.chom.2008.09.007 (2008).
- 841 13 Benjdia, A., Martens, E. C., Gordon, J. I. & Berteau, O. Sulfatases and a radical S-
842 adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the
843 prominent human gut symbiont, *Bacteroides thetaiotaomicron*. *J Biol Chem* **286**, 25973-
844 25982, doi:10.1074/jbc.M111.228841 (2011).
- 845 14 Hickey, C. A. *et al.* Colitogenic *Bacteroides thetaiotaomicron* Antigens Access Host
846 Immune Cells in a Sulfatase-Dependent Manner via Outer Membrane Vesicles. *Cell Host*
847 *Microbe* **17**, 672-680, doi:10.1016/j.chom.2015.04.002 (2015).

- 848 15 Diez-Roux, G. & Ballabio, A. Sulfatases and human disease. *Annu Rev Genomics Hum*
849 *Genet* **6**, 355-379, doi:10.1146/annurev.genom.6.080604.162334 (2005).
- 850 16 Wlodarska, M. *et al.* Indoleacrylic Acid Produced by Commensal Peptostreptococcus
851 Species Suppresses Inflammation. *Cell Host Microbe* **22**, 25-37 e26,
852 doi:10.1016/j.chom.2017.06.007 (2017).
- 853 17 Tramontano, M. *et al.* Nutritional preferences of human gut bacteria reveal their
854 metabolic idiosyncrasies. *Nat Microbiol* **3**, 514-522, doi:10.1038/s41564-018-0123-9
855 (2018).
- 856 18 Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila*
857 gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol*
858 *Microbiol* **54**, 1469-1476 (2004).
- 859 19 Pudlo, N. A. *et al.* Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for
860 Host Mucosal Glycans. *MBio* **6**, e01282-01215, doi:10.1128/mBio.01282-15 (2015).
- 861 20 Barbeyron, T. *et al.* Matching the Diversity of Sulfated Biomolecules: Creation of a
862 Classification Database for Sulfatases Reflecting Their Substrate Specificity. *PLoS One*
863 **11**, e0164846, doi:10.1371/journal.pone.0164846 (2016).
- 864 21 Benjdia, A. *et al.* Anaerobic sulfatase-maturing enzymes, first dual substrate radical S-
865 adenosylmethionine enzymes. *J Biol Chem* **283**, 17815-17826,
866 doi:10.1074/jbc.M710074200 (2008).
- 867 22 Cartmell, A. *et al.* How members of the human gut microbiota overcome the sulfation
868 problem posed by glycosaminoglycans. *Proc Natl Acad Sci U S A* **114**, 7037-7042,
869 doi:10.1073/pnas.1704367114 (2017).
- 870 23 Ndeh, D. *et al.* Metabolism of multiple glycosaminoglycans by *Bacteroides*
871 *thetaiotaomicron* is orchestrated by a versatile core genetic locus. *Nat Commun* **11**, 646,
872 doi:10.1038/s41467-020-14509-4 (2020).
- 873 24 Thomsson, K. A., Karlsson, H. & Hansson, G. C. Sequencing of sulfated
874 oligosaccharides from mucins by liquid chromatography and electrospray ionization
875 tandem mass spectrometry *Analytical Chemistry* **72**, 4543-4549, doi:10.1021/ac000631b
876 (2000).
- 877 25 Tobisawa, Y., Imai, Y., Fukuda, M. & Kawashima, H. Sulfation of colonic mucins by N-
878 acetylglucosamine 6-O-sulfotransferase-2 and its protective function in experimental
879 colitis in mice. *J Biol Chem* **285**, 6750-6760, doi:10.1074/jbc.M109.067082 (2010).
- 880 26 Egan, M., Jiang, H., O'Connell Motherway, M., Oscarson, S. & van Sinderen, D.
881 Glycosulfatase-Encoding Gene Cluster in *Bifidobacterium breve* UCC2003. *Appl*
882 *Environ Microbiol* **82**, 6611-6623, doi:10.1128/AEM.02022-16 (2016).
- 883 27 Briliute, J. *et al.* Complex N-glycan breakdown by gut *Bacteroides* involves an extensive
884 enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol* **4**,
885 1571-1581, doi:10.1038/s41564-019-0466-x (2019).
- 886 28 Domon, B. & Costello, C. E. A systematic nomenclature for carbohydrate fragmentations
887 in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J* **5**, 397-409,
888 doi:10.1007/BF01049915 (1988).
- 889 29 Das, T. M., Rao, C. P. & Kolehmainen, E. Synthesis and characterisation of N-glycosyl
890 amines from the reaction between 4,6-O-benzylidene-D-glucopyranose and substituted
891 aromatic amines and also between 2-(o-aminophenyl)benzimidazole and pentoses or
892 hexoses. *Carbohydr Res* **334**, 261-269, doi:10.1016/s0008-6215(01)00202-6 (2001).

893 30 Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic
894 Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353 e1321,
895 doi:10.1016/j.cell.2016.10.043 (2016).

896 31 Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a
897 prominent human gut symbiont is directed by the recognition of amylose helices.
898 *Structure* **16**, 1105-1115, doi:10.1016/j.str.2008.03.017 (2008).

899 32 Degan, P. H., Barry, N. A., Mok, K. C., Taga, M. E. & Goodman, A. L. Human gut
900 microbes use multiple transporters to distinguish vitamin B(1)(2) analogs and compete in
901 the gut. *Cell Host Microbe* **15**, 47-57, doi:10.1016/j.chom.2013.12.007 (2014).

902 33 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132,
903 doi:10.1107/S0907444909047337 (2010).

904 34 Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr D*
905 *Biol Crystallogr* **69**, 1260-1273, doi:10.1107/S0907444913015308 (2013).

906 35 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta*
907 *Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).

908 36 Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-
909 replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**, 125-132,
910 doi:10.1107/S0907444907050172 (2008).

911 37 Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D*
912 *Biol Crystallogr* **66**, 22-25, doi:10.1107/S0907444909042589 (2010).

913 38 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
914 *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493
915 (2010).

916 39 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal
917 structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367,
918 doi:10.1107/S0907444911001314 (2011).

919 40 Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library.
920 *Acta Crystallogr D Biol Crystallogr* **68**, 431-440, doi:10.1107/S090744491200251X
921 (2012).

922 41 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular
923 crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21,
924 doi:10.1107/S0907444909042073 (2010).

925 42 Collaborative Computational Project, N. The CCP4 suite: programs for protein
926 crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763,
927 doi:10.1107/S0907444994003112 (1994).

928 43 Potterton, L. *et al.* CCP4i2: the new graphical user interface to the CCP4 program suite.
929 *Acta Crystallogr D Struct Biol* **74**, 68-84, doi:10.1107/S2059798317016035 (2018).

930 44 Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking
931 web service based on EADock DSS. *Nucleic Acids Res* **39**, W270-277,
932 doi:10.1093/nar/gkr366 (2011).

933 45 Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field
934 with EADock DSS. *J Comput Chem* **32**, 2149-2159, doi:10.1002/jcc.21797 (2011).

935 46 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and
936 analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

937 47 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid
938 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**,
939 3059-3066, doi:10.1093/nar/gkf436 (2002).

940 48 Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor.
941 *Bioinformatics* **20**, 426-427, doi:10.1093/bioinformatics/btg430 (2004).

942 49 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
943 large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033
944 (2014).

945 50 Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach.
946 *Journal of molecular evolution* **17**, 368-376, doi:10.1007/BF01734359 (1981).

947 51 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol Biol*
948 *Evol* **25**, 1307-1320, doi:10.1093/molbev/msn067 (2008).

949 52 Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap.
950 *Evolution* **39**, 783-791, doi:10.1111/j.1558-5646.1985.tb00420.x (1985).

951 53 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics
952 Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874,
953 doi:10.1093/molbev/msw054 (2016).

954 54 Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines.
955 *Glycobiology* **29**, 620-624, doi:10.1093/glycob/cwz045 (2019).
956

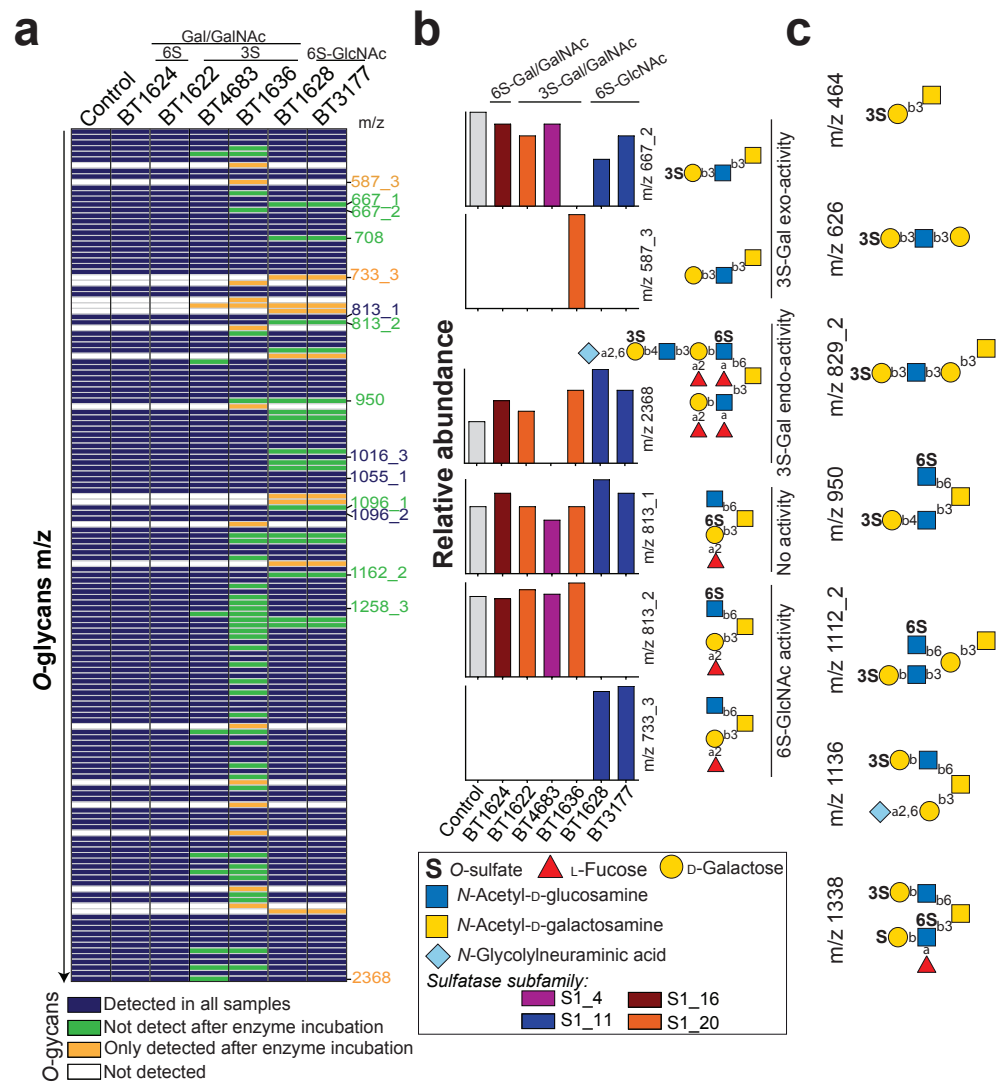


Fig 2.

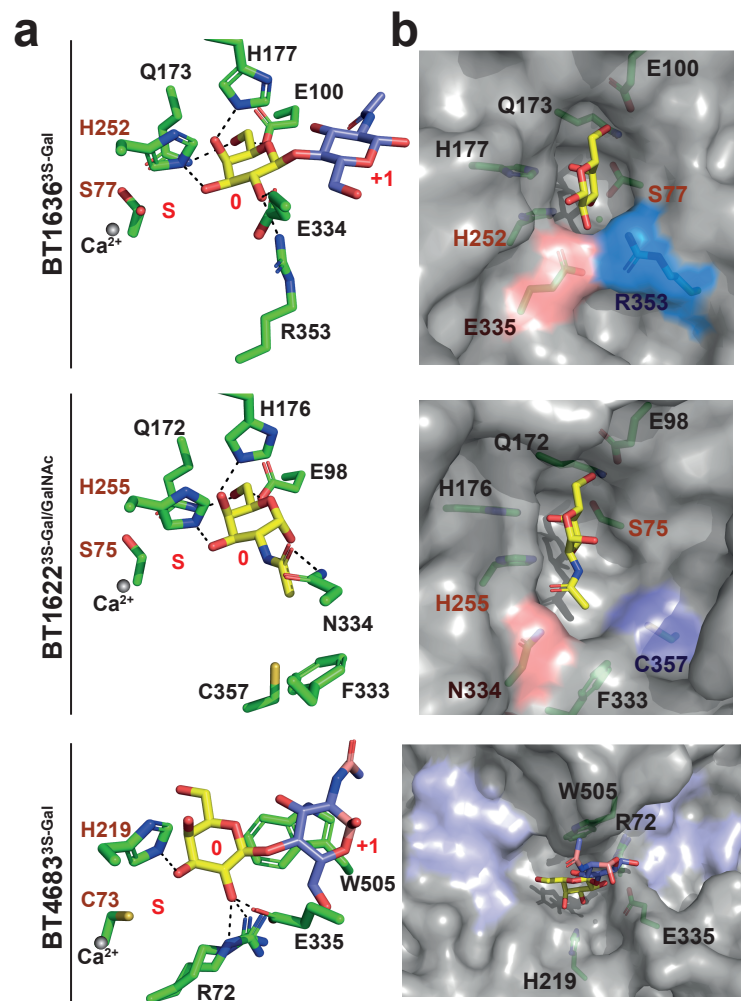


Fig 3.

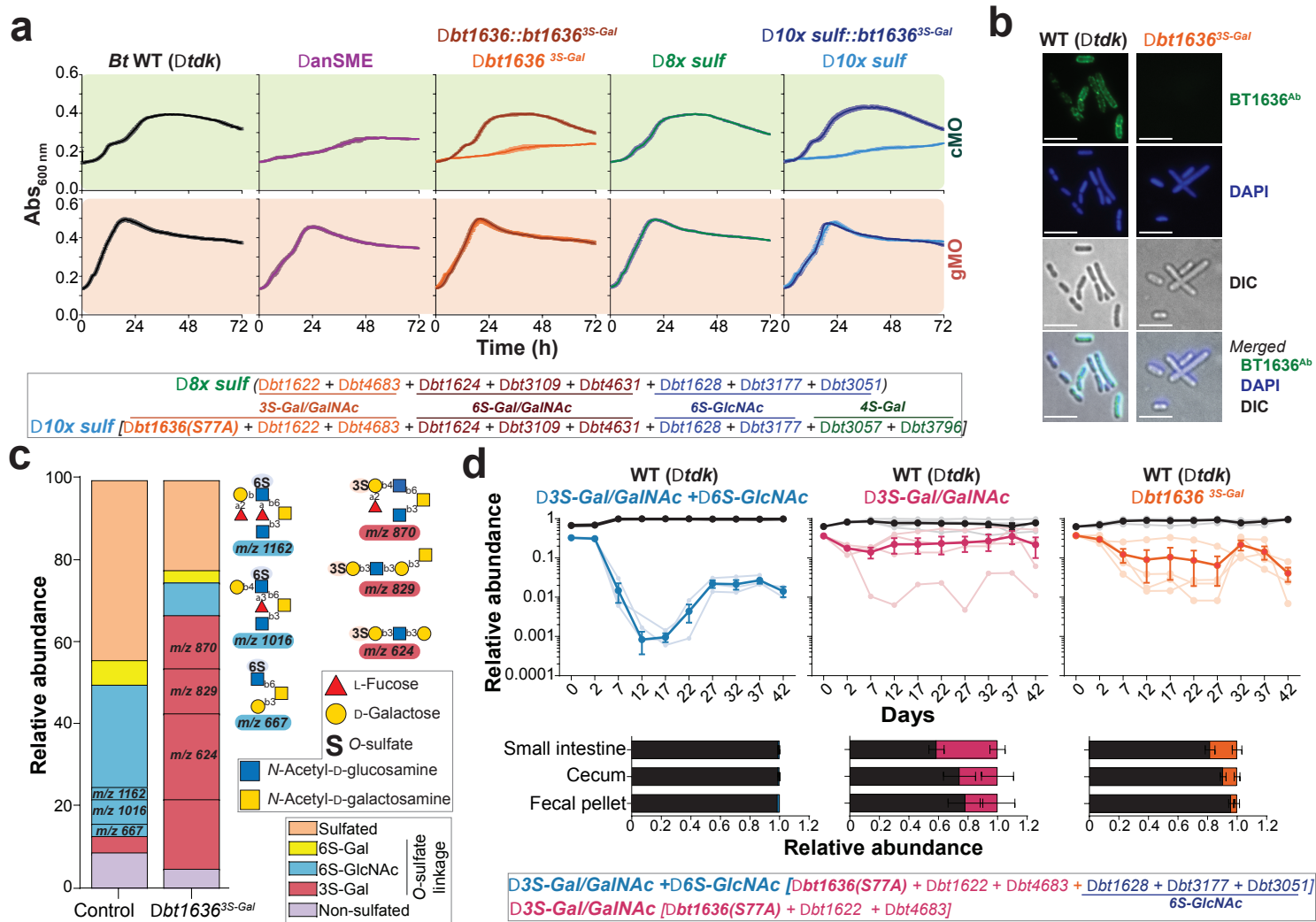
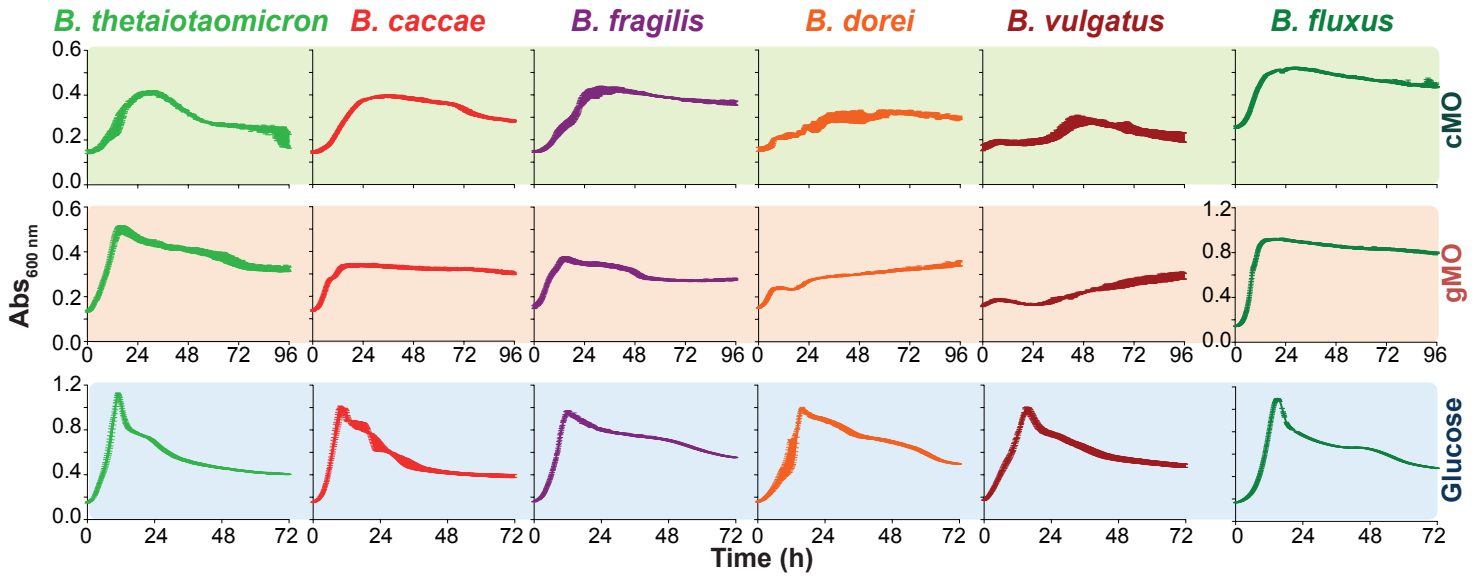
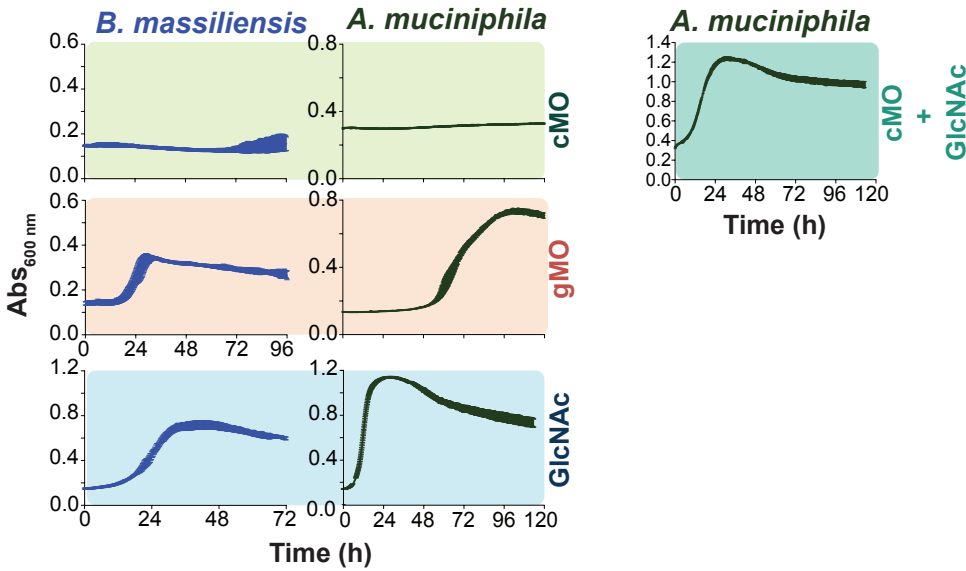


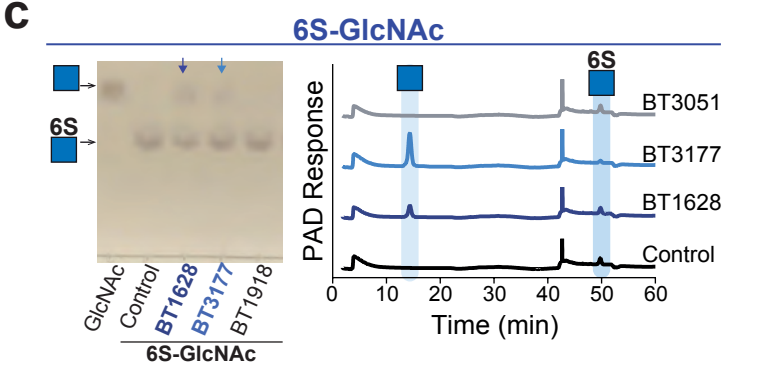
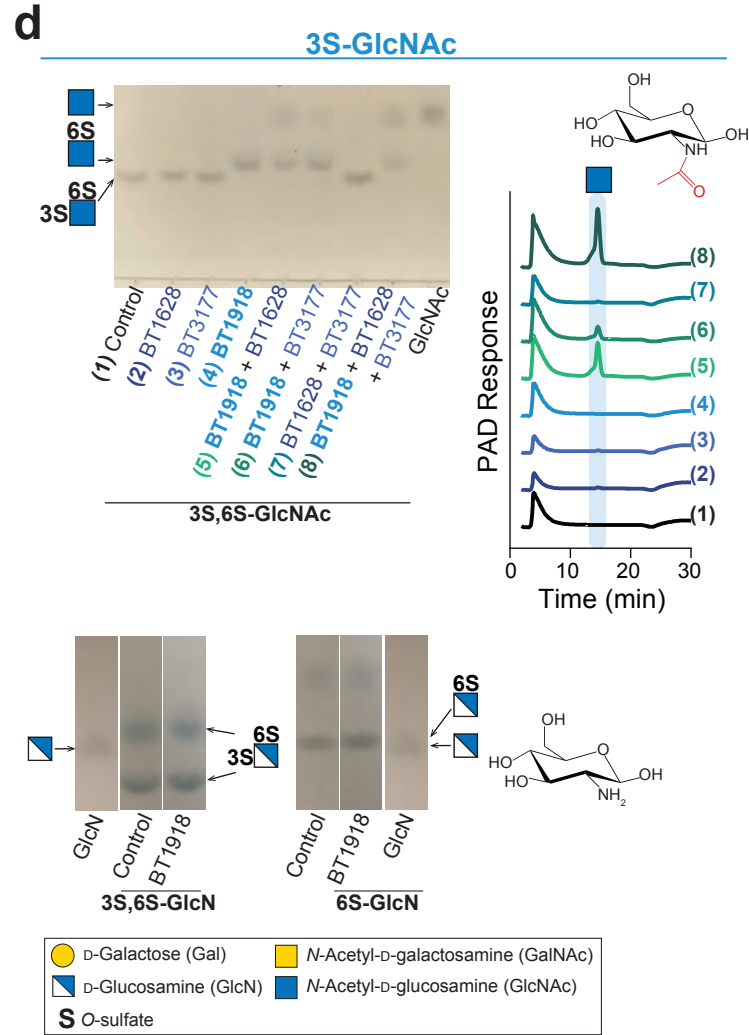
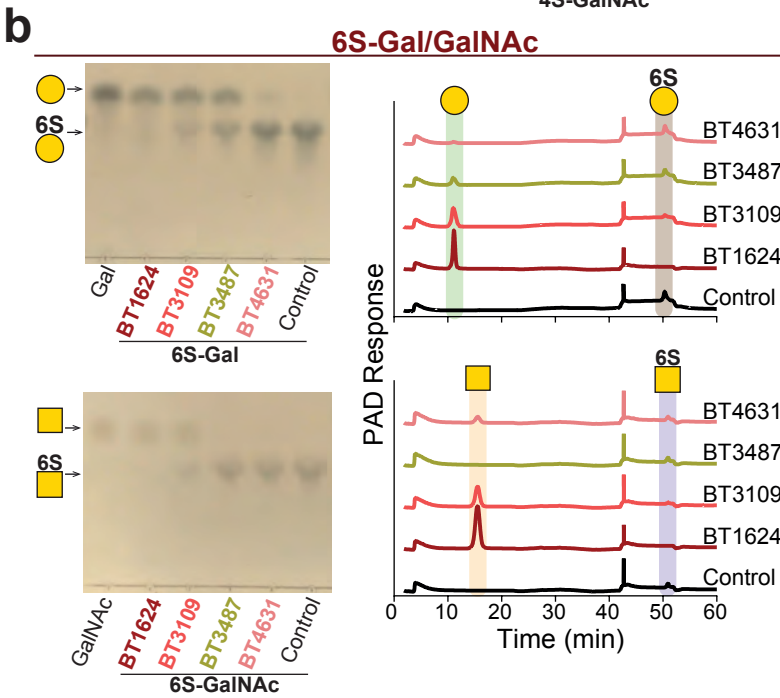
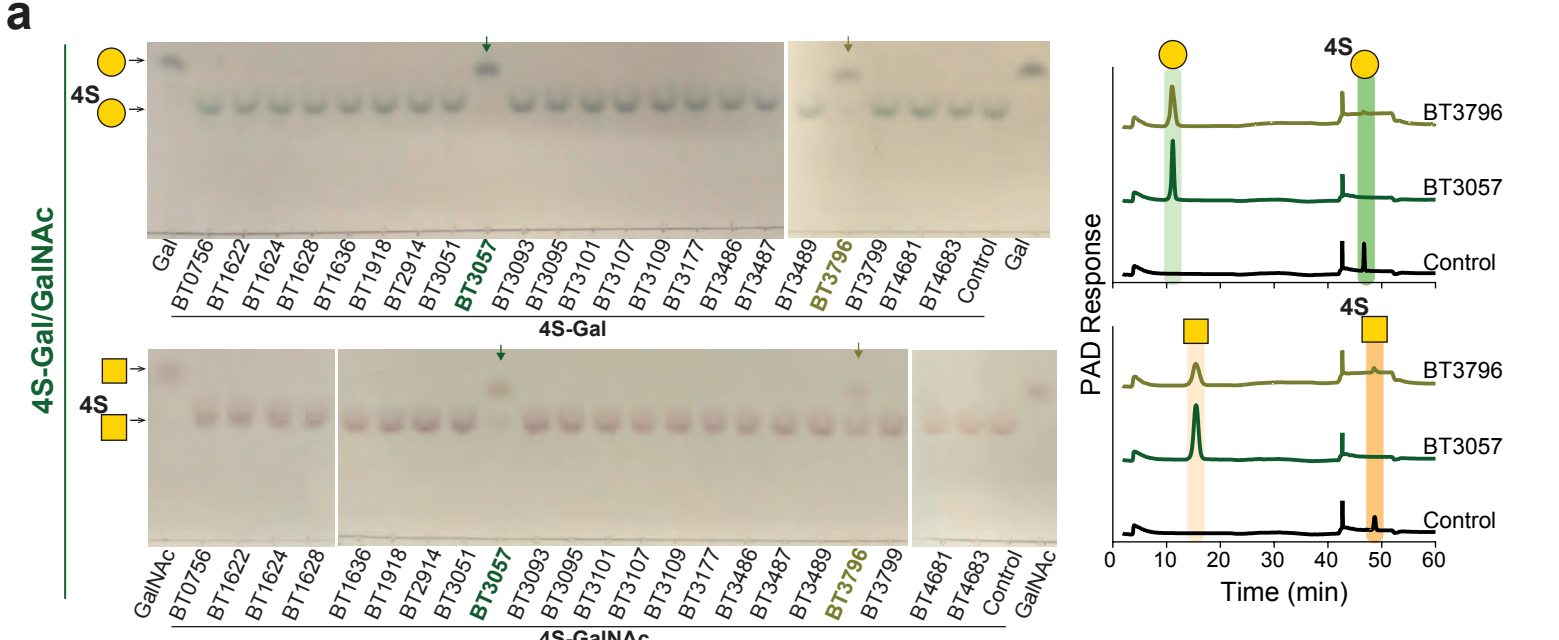
Fig 4.

O-glycans utilizer



Gastric O-glycans utilizer





Extended Data Figure 3

a

3S-Gal/GalNAc

6S-GlcNAc

6S-Gal/GalNAc

Control BT1622 BT1636 BT4683

Control BT1624 BT3109 BT3487 BT4631

Relative abundance

Control BT1622 BT1636 BT4683

Control BT1624 BT3109 BT3487 BT4631

Lewis^x 3'S-Lewis^x Lewis^a 3'S-Lewis^a LacNAc 3'S-LacNAc LNB 3'S-LNB 6S-Lewis^x 6S-Lewis^a 6S-LacNAc 6S-LNB 6'S-Lewis^x 6'S-Lewis^a

3S b₄ a₄ 3S b₄ a₄ 6S b₃ a₄ 6S b₃ a₄ 6S b₃ a₄ 6S b₃ a₄ 6S b₃ a₄ 6S b₃ a₄

b

3S-Gal/GalNAc

BT1622 BT1636 BT4683

3'S-Lewis^x

3'S-Lewis^a

3'S-LacNAc

3'S-LNB

6S-Gal/GalNAc

BT3333 BT1624 BT3109 BT4631

DTm

mM

3'S b₄ a₄ 3'S b₄ a₄ 3'S b₄ a₄ 3'S b₄ a₄

HO OH HO OH HO OH HO OH

HO OH HO OH HO OH HO OH

c

3S-Gal/GalNAc

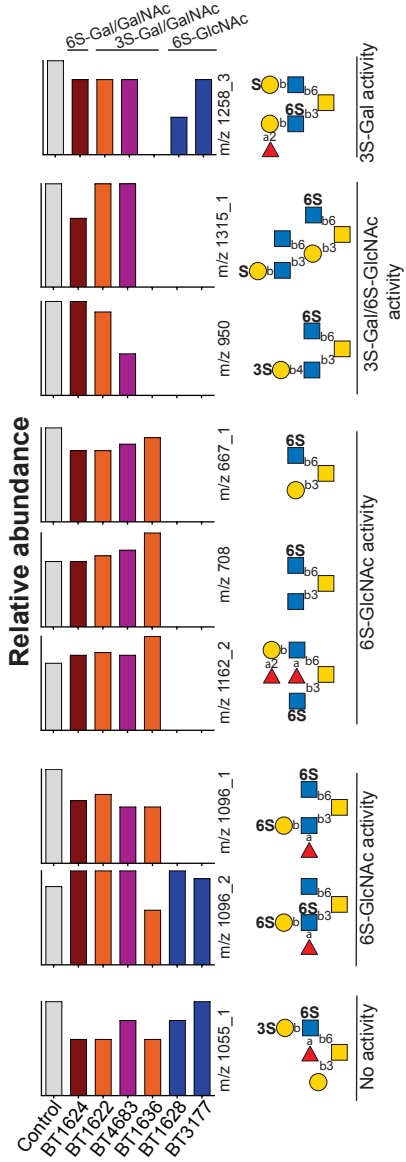
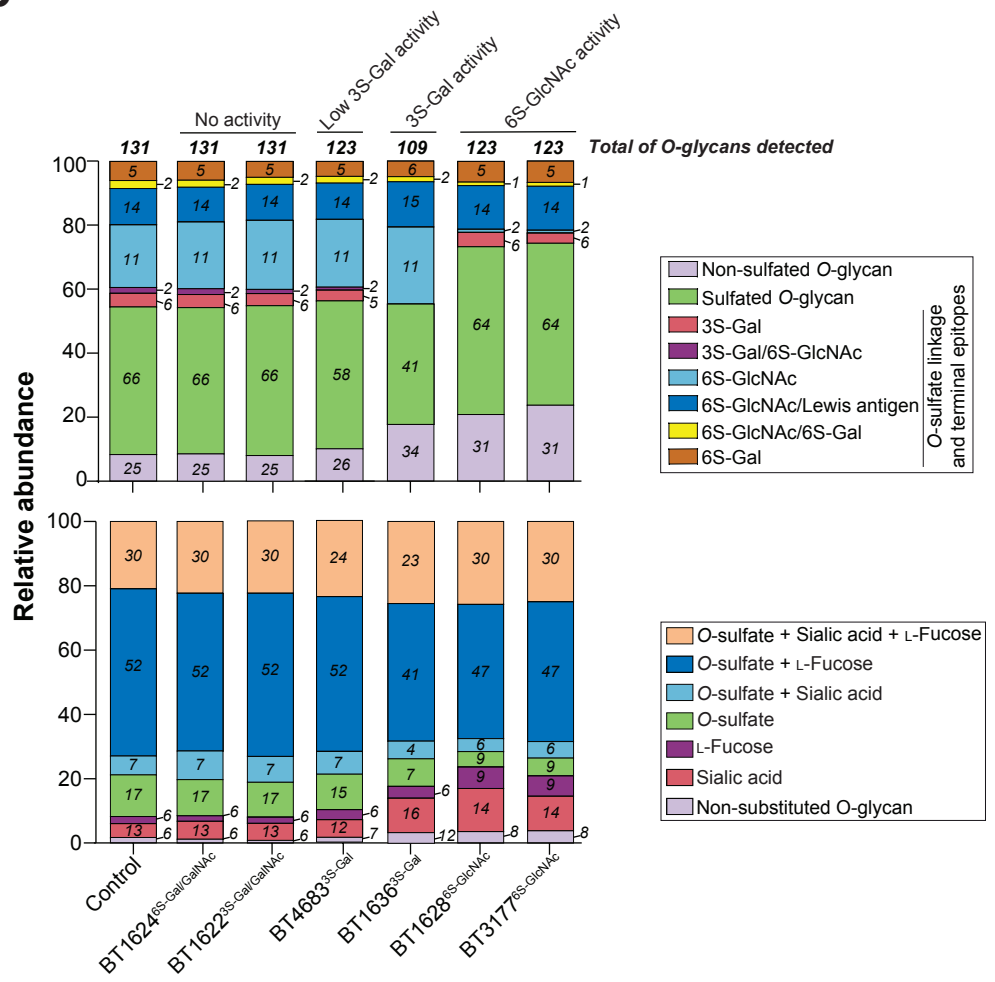
3S

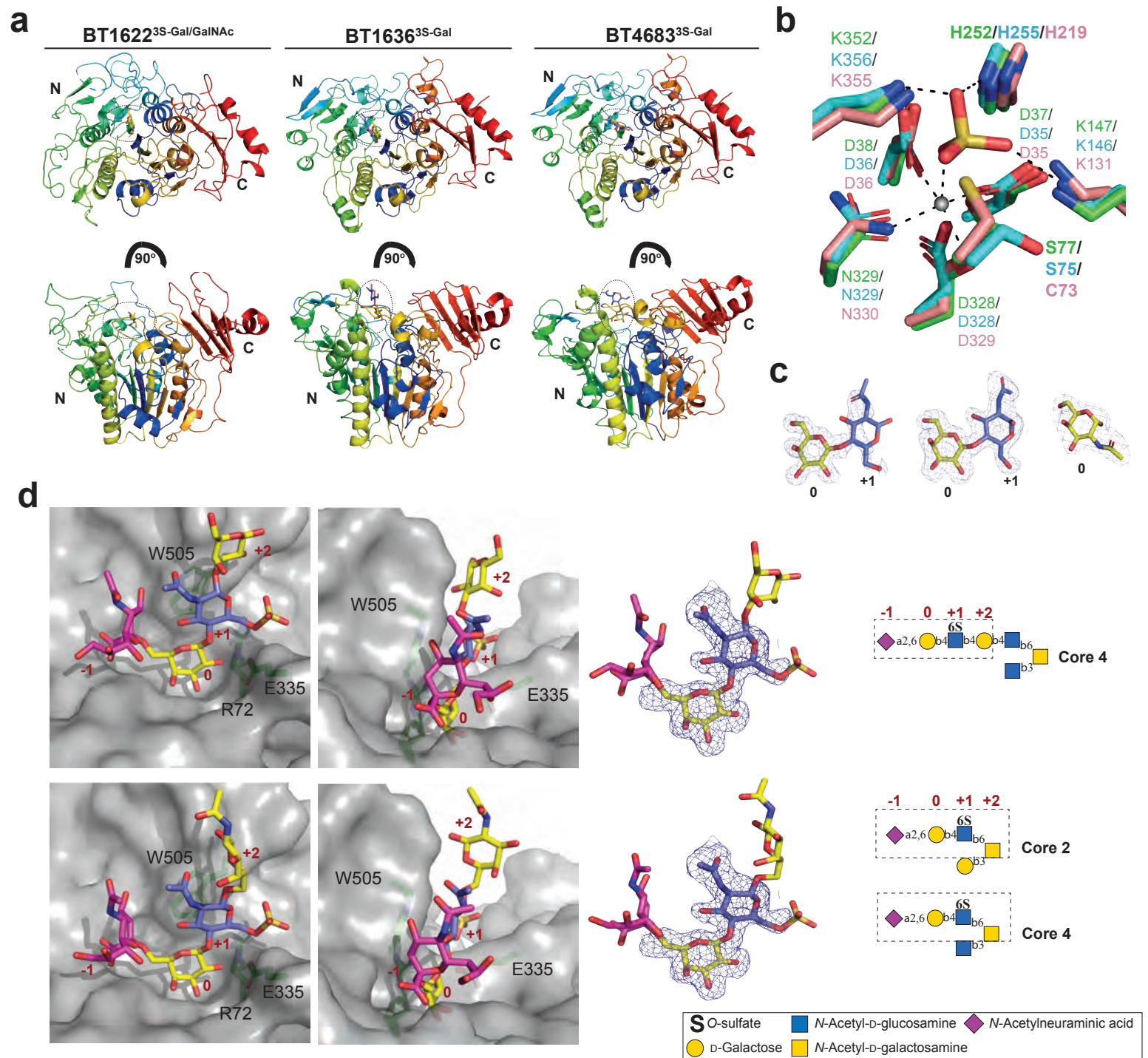
GlcNAc Control BT1622 BT1636

3S-GalNAc

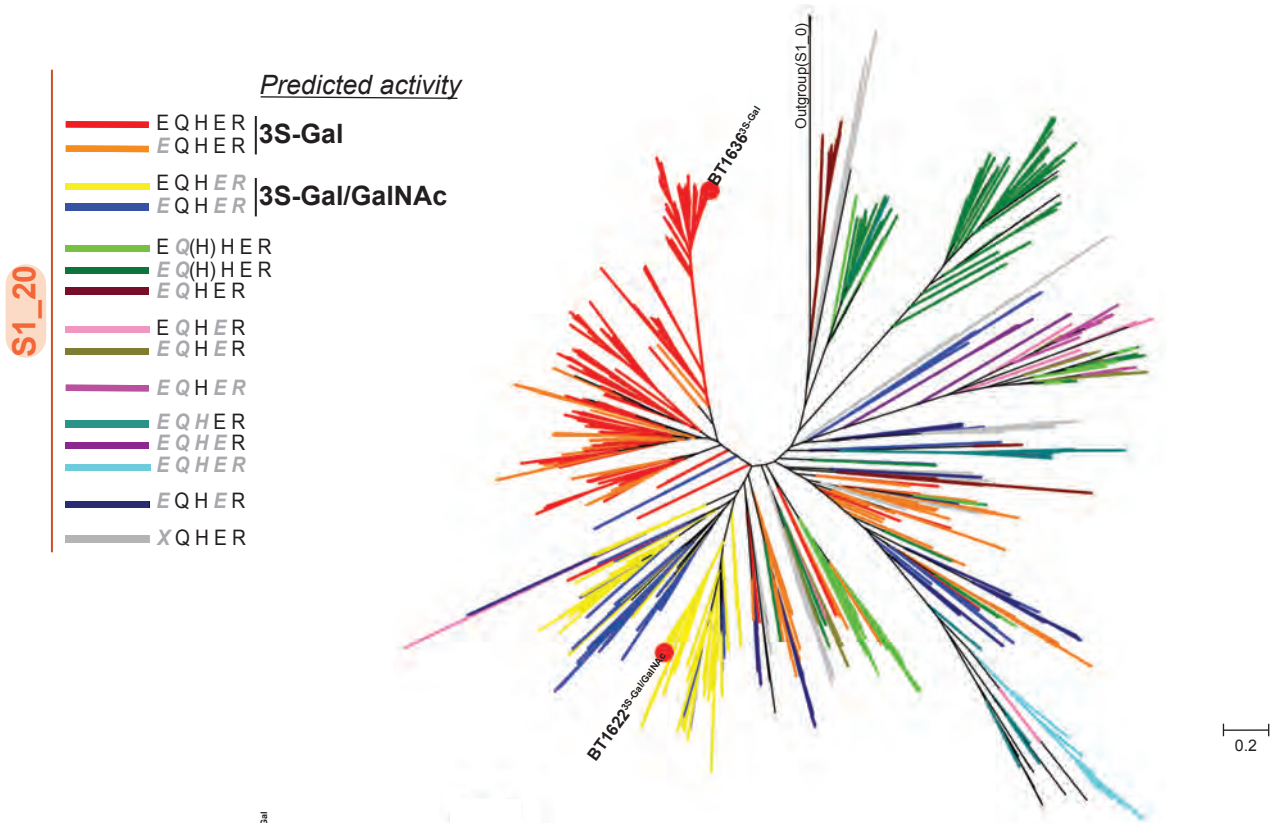
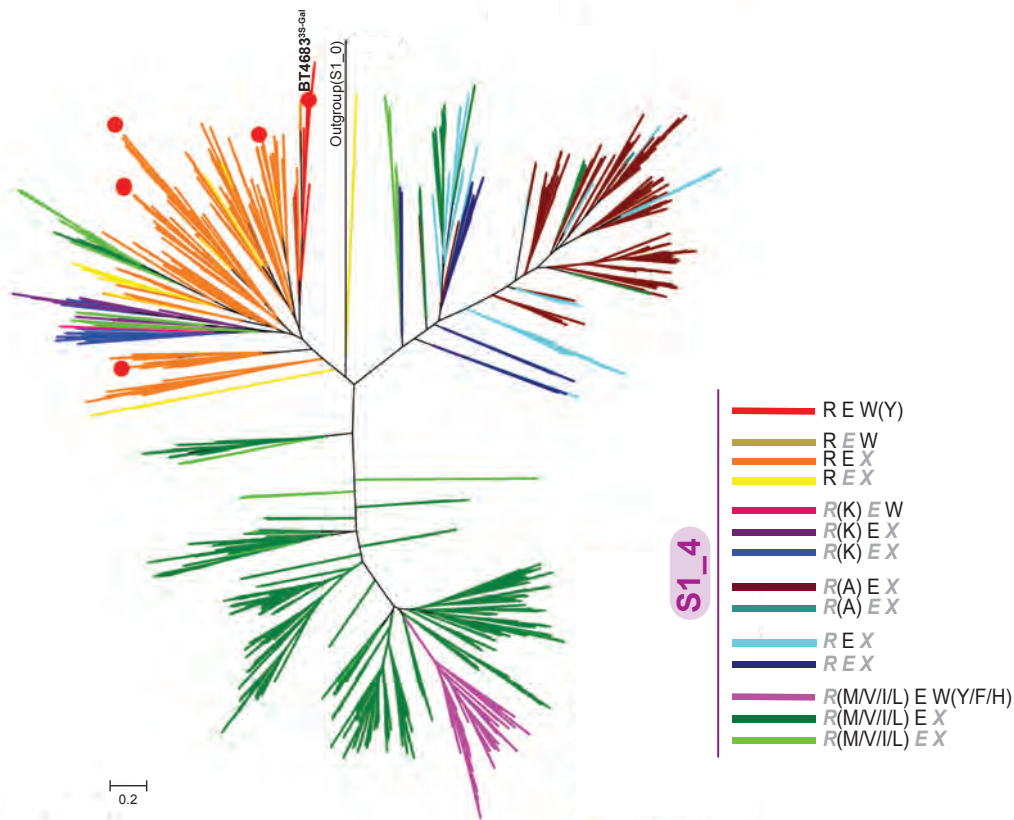
▲ L-Fucose (Fuc)
● D-Galactose (Gal)
■ N-Acetyl-D-glucosamine (GlcNAc)
■ N-Acetyl-D-galactosamine (GalNAc)
S O-sulfate

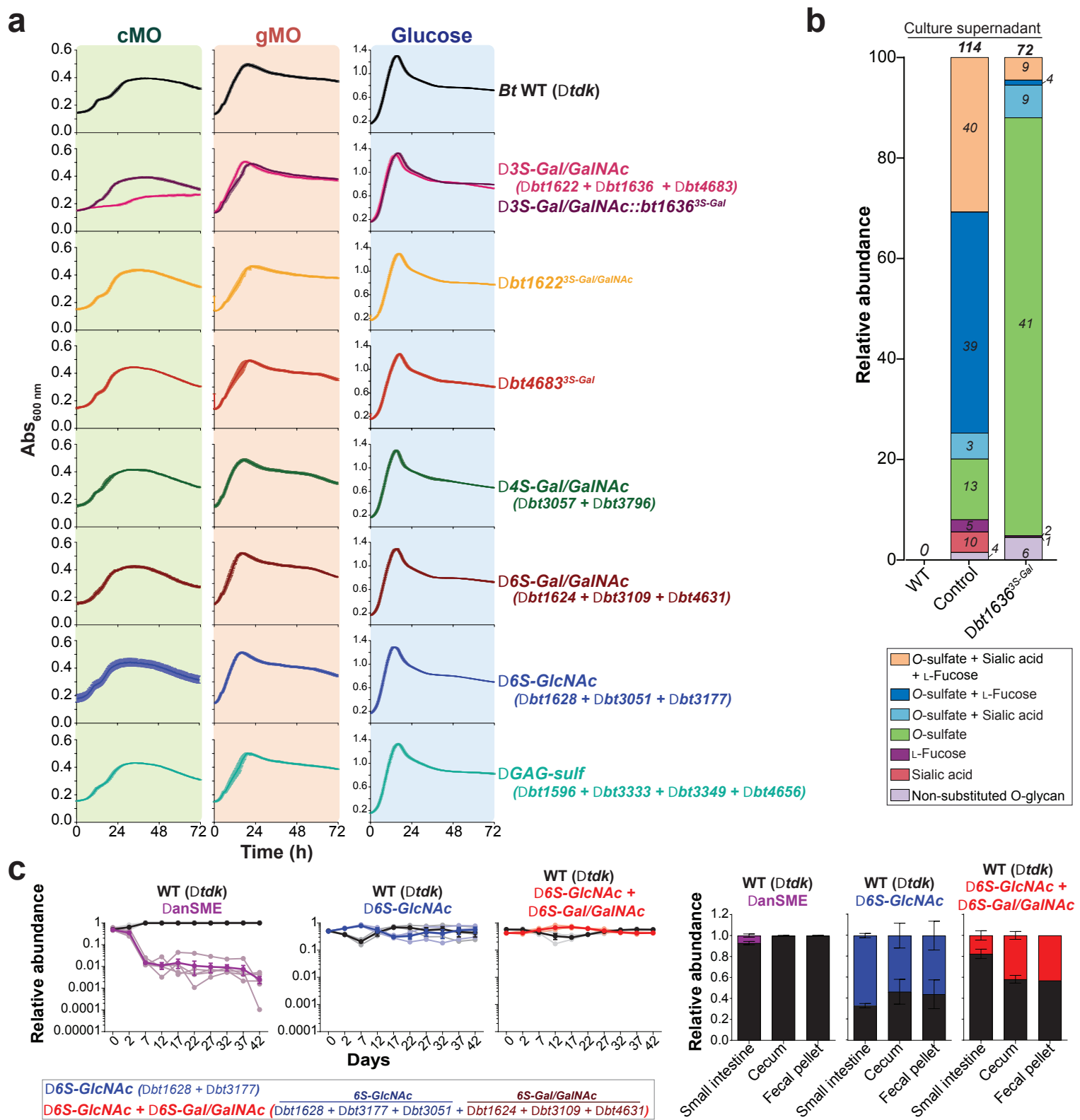
Extended Data Figure 4

a**b**



Extended Data Figure 6

a**b**



Extended Data Figure 8

Supplemental Discussion

1. Utilization of different mucin O-glycans sources by HGM

Mucin composition varies through the gastrointestinal tract (GI), with the stomach having mainly Muc5AC and the colon mainly Muc2¹. The glycosylation of these respective mucins also varies along the GI tract with higher levels of sulfation and sialylated structures observed in the distal colon compared to the upper GI tract². From the 20 bacterial strains tested for growth in this study, 12 failed to grow on gastric mucin O-glycans (gMO) or colonic mucin O-glycans (cMOs) (**Fig. 1**). Only 6 bacteria were able to utilize both O-glycans substrates but growth was variable. In both O-glycans substrates, *Bacteroides thetaiotaomicron* (*Bt*), *B. caccae*, *B. fragilis* and *B. fluxus* grew better than *B. dorei* and *B. vulgatus* (**Fig. 1** and **Extended Data Fig. 1**). The differences observed in the growth profiles were reproducible in two different batches of purified cMOs (**Fig. 1**) Indeed, it is likely that different HGM members evolve to target different (or only a subset) of the available O-glycans and that this fine-tuning of host glycan utilization may have important implications in gut colonization and symbiosis. Additionally, *B. massiliensis* and *Akkermansia mucinipila* (*Amuc*) grew on gMO but failed to utilize cMO (**Fig. 1** and **Extended Data Fig. 1**). Both strains were able to grow on *N*-acetylglucosamine (GlcNAc) and *Amuc* grew on GlcNAc in the presence of cMO suggesting that these O-glycans do not inhibit the growth of this bacterium. Previous studies have determined *B. massiliensis* and *Amuc* to be mucin-degraders by demonstrating growth on gastric mucins^{3,4}. However, the lack of growth in colonic O-glycans suggests that these bacteria are not able to initiate the degradation of more complex and sulfated colonic glycans. This finding highlights the importance of taking in account O-glycosylation differences along the GI tract and the need to utilize colonic mucins to draw conclusions regarding the colonic HGM.

2. Sulfatase activity in cMO

Despite all 12 sulfatases being active on defined oligosaccharides, of those tested on cMO, BT1622^{3S-Gal/GalNAc} and BT1624^{6S-Gal/GalNAc} did not show any activity on this complex substrate (**Fig. 2**, **Extended Data Fig. 5** and **Supplementary Table 5**).

These findings are consistent the results observed in defined commercial oligosaccharides where BT1622^{3S-Gal/GalNAc} showed a preference to for sulfated GalNAc over Gal glycans (**Extended Data Fig. 4**) and BT1624^{6S-Gal/GalNAc} activity is blocked by the presence of additional substitutions (such as Lewis antigens) (**Extended Data Fig. 4a**). Additionally, 2 of 6 detected 6S-Gal structures contained a capping sialic acid and a terminal blood group H type 2 [Fuc- α 1,2-(6S)Gal-] (**Fig. 2**). The lack of activity of BT1624^{6S-Gal/GalNAc} towards such structures confirms the exo-mode of action that we described for this sulfatase using commercial substrates.

Overall, when compared to the non-enzyme treated control, we detected an increase of non-sulfated structures and decrease of sulfated oligosaccharides in all samples with the active enzymes (BT4683^{3S-Gal}, BT1636^{3S-Gal}, BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc}) (**Extended Data Fig. 5b**). BT1636^{3S-Gal} was active towards all detected 3S-Gal structures with the exception of glycan m/z 1055_1 that is a doubly sulfated 3S-Gal/6S-GlcNAc fucosylated structure (**Extended Data Fig. 5a**). As we observed using commercial substrates, the presence of Lewis a/x epitopes leads to a decrease in the activity of this sulfatase (**Extended Data Fig. 4a** and **Supplementary Table 5**) and the presence of a second sulfate group might exacerbate this negative effect leading to the lack of activity towards this complex sulfated O-glycan. The incubation of the 6S-GlcNAc sulfatases BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc} with cMO suggests that these enzymes are redundant, but because they are encoded in different PULs they could be expressed in response to different activating cues (**Fig. 2** and **Supplementary Table 5**). Compared to the non-enzyme treated control, 16 glycans were not detected after incubation with these sulfatases, 14 of these structures have a terminal 6S-GlcNAc (**Fig. 2** and **Supplementary Table 5**). BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc} were active in 6S-GlcNAc core 3 (GlcNAc- β 1,3-GalNAc) and core 4 (GlcNAc- β 1,6-GalNAc) structures (**Extended Data Fig. 5a**), suggesting that these sulfatases are well suited to accommodate the variations in linkages/sugars found in mucin O-glycans. Additionally, we also detect 7 new glycans that are likely to be reaction products of BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc} (**Fig. 2** and **Supplementary Table 5**).

The identification and characterization of the first sulfatases active on mucin O-glycans creates the opportunity to improve our understanding of O-glycan structures by using these enzymes as analytical tools. After the treatment with BT1636^{3S-Gal}

several oligosaccharides predicted to contain a terminal sulfate linked to Gal were not detected. Although we could not determine the specific sulfate linkage by mass spectrometry, the activity of the 3S-sulfatase suggests that these oligos contain a terminal 3S-Gal (**Extended Data Fig. 5a** and **Supplementary Table 5**). The specificity of the 6S-GlcNAc sulfatases for non-fucosylated O-glycans also illuminates their potential use as tools to characterize the structure of these complex structures since it allows the differentiation of different isomers. For example, we detect two oligosaccharides with m/z 1096, however, after incubation with BT1628^{6S-GlcNAc} or BT3177^{6S-GlcNAc}, only the isomer 1096_2 was detected, indicating that the isomer 1096_1 contain a terminal 6S-GlcNAc (**Extended Data Fig. 5a** and **Supplementary Table 5**).

3. **Conserved structural features of the S1 formylglycine family**

Protein fold and subsites nomenclature

S1 sulfatases comprise the most common and largest family of sulfatases, currently encompass 36,816 members in sulfAtlas and are found in all domains of life⁵. S1 sulfatases are part of the alkaline phosphatase superfamily and adopt an alkaline phosphatase-like fold. This is an N-terminal $\alpha/\beta/\alpha$ domain with S1 sulfatases also possessing a smaller C-terminal 'sub domain'. The active site is located in the N-terminal domain that has a large mixed β -sheet composed of ~ 10 β strands, sandwiched between α helices above and below. The C-terminal 'sub-domain' is composed of a 4 stranded antiparallel β -sheet and a single amphipathic terminal helix. This C-terminal domain abuts the N-terminal domain through the antiparallel β -sheet with loops from the β strands sometimes contributing to the active site architecture (**Extended data Fig. 6a**). The subsite nomenclature for carbohydrate sulfatases is such that the invariant sulfate binding site is denoted the S site. The S site sulfate is appended to is the 0 subsite sugar. Subsites then increase in number (i.e. +1, +2, +3) as the sugar moves toward the reducing end (free O1) and decreases in number as the sugar chain moves towards the non-reducing end (i.e. -1, -2, -3)⁶.

S1 formylglycine active site conservation

The sulfate binding site (S site) is invariant across the S1 family and comprises the catalytic residues (nucleophile and catalytic acid) and a calcium binding site (**Extended data Fig. 7b**). An invariant histidine is likely the potential catalytic acid but a lysine has also been suggested to possibly fulfill this role⁷. The pKa of His is ~6.0, whilst Lys has a pKa of >10, making it more chemically feasible that His performs the role of the catalytic acid. Homologues of these residues (H252 and K352 in BT1636^{3S-Gal}) hydrogen bond to the scissile sulfoester linkage (**Extended data Fig. 6b**). Previously published work with BT1596 and BT4656, a 2S-Uronic acid and 6S-GlcNAc sulfatase, respectively, showed that the mutation of either residue to alanine causes inactivation⁷. Consistent with this work, a BT4683^{3S-Gal} H219A mutant was inactive. However, in BT1622^{3S-Gal/GalNAc}, the mutation of H255 to Ala caused only a ~30 fold decrease in activity (**Supplementary Table 3**). Thus, it is possible that in BT1622^{3S-Gal/GalNAc} the loss of H255 is compensated by the invariant residue K356 and interestingly BT1622^{3S-Gal/GalNAc} has a pH optimum ~2 units higher than most sulfatases assayed (**Supplementary Fig. 4**).

The calcium binding site is located at the base of the S site interacting with the sulfate group. This calcium ion is an essential component of the catalytic mechanism helping to stabilise negative charges that occur during the catalysis. All three of the solved structures had occupation for calcium. In BT1636^{3S-Gal} D328 and the sulfate group of the substrate coordinate above and below the calcium with D37, D38, N329 and the formylglycine binding in a plane completing an octahedral coordination (**Extended Data Fig. 6b**). These three Asp and the Asn coordinated with calcium are structurally conserved in all 3S-Gal/GalNAc sulfatases structures (Extended data Fig. 6b).

The solved structures of BT1636^{3S-Gal} and BT1622^{3S-Gal/GalNAc} were native *Bt* proteins having a Ser at the formylglycine position. However, structure of BT4683^{3S-Gal} was obtained with the active protein where S73 was mutated to Cys (as *E. coli* can only convert Cys, not Ser, to formylglycine). The analysis of BT4683^{3S-Gal} reveals that the crystallized protein still has the Cys and not formylglycine indicating poor installation of the formylglycine. This observation means the kinetic data, (although the rates are significant and readily measurable) may be an underestimation of true catalytic performance. This will affect the k_{cat} component of the k_{cat}/K_M measurement and thus the k_{cat}/K_M reported in **Supplementary Table 3** is an underestimate of the true activity.

Additional 3S-Gal/GalNAc specificity determinants based on structures

BT1636^{3S-Gal} was solved in complex with the product LacNAc; the 0 subsite galactose is well ordered and makes extensive interactions, whilst the +1 GlcNAc is highly disordered and appears to make no interactions with the protein (**Fig. 3** and **Extended Data Fig. 6c**). O2 of the 0 Gal hydrogen bonds with Oε1 of E334 and NH2 of R353. Mutation of these residues to Ala causes ~300 and ~60-fold reductions in k_{cat}/K_M , respectively. The O6 group of Gal potentially coordinates with Oε2 of E100 and Nε2 of Q173 and mutations of these residues to Ala cause ~80 and ~50-fold decreases in k_{cat}/K_M , respectively (**Fig. 3** and **Supplementary Table 3**). Comparison of the BT1622^{3S-Gal/GalNAc} structure with BT1636^{3S-Gal} shows that E98 and Q172 (which correspond to E100 and Q173 in BT1636^{3S-Gal}) are conserved (**Fig. 3**) and mutating E98 to Ala caused only a 15-fold decrease in k_{cat}/K_M (**Supplementary Table 3**). Additionally, in BT1622^{3S-Gal/GalNAc} the hydrophobic interactions with the *N*-acetyl group, and the more open pocket, may offset the effects H176A (300-fold loss in activity) when compared to H177A (complete loss in activity) in BT1636^{3S-Gal} (**Fig. 3** and **Supplementary Table 3**).

BT4683^{Gal-3s} also displayed the same 3S Gal activity as the S1_20 enzymes but showed a preference for 3S-LacNAc, reciprocal to BT1622^{3S-Gal}. BT4683^{3S-Gal} bound the O2 Gal of LacNAc via Oε2 of E335 (equivalent to E334 in BT1636^{3S-Gal}) and through either Nε or NH1 of R72 (**Fig. 3**). Although R72 is sequentially distal to R353 in BT1636^{3S-Gal} it is spatially similar and likely contributes in a similar capacity (**Fig. 3**). Despite the mutations R72A and E335A resulting in loss of activity, the Glu and Arg are only conserved in 62 % and 19 % of S1_4 sequences, respectively, suggesting there is a significant but not absolute selection for an equatorial O2 in this subfamily (**Extended Data Fig. 7** and **Supplementary Fig. 2**). Uniquely among the 3S-Gal sulfatases identified, BT4683^{3S-Gal} utilises a hydrophobic stacking interaction through W505 to provide a platform for the +1 GlcNAc and partially the 0 Gal. Mutation of W505 to Ala almost completely abolishes activity on 3S-LacNAc (**Supplementary Table 3**) but surprisingly this residue is not conserved in our phylogenetic analyses of S1_4 being present in only 8 other sequences (**Extended Data 7** and **Supplementary Fig. 3**). It is important to note here, that overall W505 is not well conserved; potential equivalent aromatic residues can be found in some additional clades, coloured light brown (or bronze), pink or dark red, but it is not evident from the alignment that these

are true equivalents. Only 3D structural work will be able to show if these aromatic residues take equivalent positions in those sulfatases. Additionally, the BT4683^{3S-Gal} activity against defined sulfated saccharides was suggestive of an exo-acting enzyme that cleaves terminal 3S-Gal (**Extended Data Fig. 4**). However, a close analysis of this sulfatase structure shows that the active site is located in an open cleft characteristic of an endo-active enzyme⁶. This more open cleft of BT4683^{3S-Gal} may allow additional sugars/sulfates to be accommodated on the O6 of both the 0 Gal and +1 GlcNAc. Indeed, the activity determined in cMO shows that this sulfatase can act on sialylated O-glycan (**Fig. 2**). Further modelling of different O-glycan structures (using the crystallographically solved LacNAc as an 'anchor') indicate that this enzyme can accommodate complex O-glycans with internal sulfation (**Fig. 3 and Extended Data 7**). Together, these results suggest that BT4683^{Gal-3S}, and its close homologues, could be endo 3S sulfatases where the 0 subsite specificity for Gal is driven by glycan context and/or distal subsites such as -1 and +2, rather than an axial O4 as in S1_20.

Additionally, it is unclear why BT1636^{Gal-3S} acts better on LacNAc substrates than BT1622^{3S-Gal/GalNAc}. It is interesting to note, however, that both BT1636^{3S-Gal} and BT4683^{3S-Gal} perform well on LacNAc configured substrates and utilise an Arg and Glu to coordinate O2 whilst BT1622^{3S-Gal} lacks these residues (**Fig. 3**). These residues may lead to the enhanced activity on LacNAc (β 1,4 glycan) vs. LNB (β 1,3 substrate). Another thing to note is that a β 1,4 vs β 1,3 linkage will rotate the GlcNAc $\sim 60^\circ$ but switch the position of the Fuc residue from being on the 'N-acetyl side' of the glycosidic bond to the 'O6 side' of the glycosidic bond, and this may also be the cause of the differential activities on β 1,4 vs β 1,3 linked substrates.

Phylogenetic analyses of S1_20 specificity determinants

The essential His that acts as a key specificity determinant of galacto- over gluco-substrates (H177 and H176 in BT1636^{3S-Gal} and BT1622^{3S-Gal/GalNAc}, respectively) is highly conserved (92% of S1_20 sequences) (**Extended data Fig. 7**). The Gln (Q173 and Q172 in BT1636^{3S-Gal} and BT1622^{3S-Gal/GalNAc}, respectively) is only conserved in 66% of sequences and in 25% of the cases is substituted with a histidine, a residue that can also fulfil the same role of Gln interacting with Gal O6. Indeed, these conserved residues are located in a highly conserved domain with the consensus sequence [CDNS]-[QH]-[RVF]-[QHLD]-[AG]-H-[NRST]-[YHF]-[YF]-P (Prosite syntax).

With H177 targeting the axial O4 of Gal directly, a Q173 may function indirectly to select for an axial O4 and thus these residues may operate as a selectivity 'dyad' for Gal with S1_20. Additionally, the residues implicated in recognition of Gal over GalNAc, E335 and R353 in BT1636^{3S-Gal} are conserved in 64 and 74% of S1_20 sequences, whilst the residue that allows the accommodation of O2 N-acetyl and activity in GalNAc (N334 in BT1622^{3S-Gal/GalNAc}) is only found in 8% of members of this family (**Extended Data Fig. 7** and **Supplementary Fig. 3**). This observation suggests that the majority of the S1_20 sulfatases evolved to target sulfated Gal and only a subset of this subfamily's members can actually also be active on GalNAc. Interestingly, all of the close homologs of BT1636^{3S-Gal} and BT1622^{3S-Gal/GalNAc} that share the critical specificity determinants of these proteins (**Supplementary Tables 6 and 7**) were isolated from mammals at body regions rich in mucins, highlighting the role of these sulfatases in accessing sulfated host glycans.

4. Growth of sulfatase mutants on O-glycans

The deletion strain lacking 4S-Gal/GalNAc sulfatases ($\Delta bt3057 + \Delta bt3796$) did not show any phenotype in cMO (**Extended Data Fig. 9a**), a result that is consistent with the lack of these sulfated linkages in colonic mucins (**Supplementary Table 5**). Unexpectedly, the deletion strains lacking the identified 6S-Gal/GalNAc sulfatases ($\Delta bt1624 + \Delta bt3109 + \Delta bt4631$) and 6S-GlcNAc sulfatases ($\Delta bt1628 + \Delta bt3051 + \Delta bt3177$) also did not show any phenotype on cMOs (**Extended Data Fig. 8a**). Analysis of cMO by MS showed that this substrate contains a low abundance of sulfate linked to O6 of Gal but a relatively high abundance of sulfate linked to O6 of GlcNAc, especially in shorter structures (**Supplementary Table 5**). Although the low abundance of O6-sulfated Gal could explain the lack of phenotype of the 6S-Gal/GalNAc sulfatase deficient strain, the lack of a phenotype of the $\Delta 6S-GlcNAc$ mutant in cMO was unexpected (**Extended Data Fig. 8a**). Due to the limitations of the MS technique it is not possible to analyse sulfation in longer oligos, making it unclear what the real complexity of structures found in colonic mucins are. Indeed, the lack of phenotype of $\Delta 6S-GlcNAc$ mutant in cMO suggests that 6S-GlcNAc might not be a major terminal epitope in colonic mucins. It is also important to note that the mutant $\Delta 6S-GlcNAc$ is the deletion of two characterized 6S-GlcNAc sulfatases active on cMO (BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc}) and a third closely related S1_11 sulfatase

(BT3051^{putative_6S-GlcNAc}) for which no activity was also deleted to avoid possible compensation of function after loss of BT1628^{6S-GlcNAc} and BT3177^{6S-GlcNAc} activities.

The deletion of previously characterized GAG-specific sulfatases⁸ ($\Delta bt1596 + \Delta bt3333 + \Delta bt3349 + \Delta bt4656$) did not result in any observable phenotype in cMO (**Extended Data Fig. 8a**), indicating that this substrate is not contaminated with additional endogenous host glycans. Additionally, despite some mutants exhibiting growth defects on sulfated cMO, all of the mutants grew well on gMO and glucose (**Fig. 4** and **Extended Data Fig. 9a**), suggesting that the phenotypes observed are dependent on the mucin source (colon) and cannot be observed utilizing mucins from other regions of the gastrointestinal tract. Together these results highlight the contribution of sulfatases in utilization of colonic mucins by the HGM.

5. Analysis of $\Delta bt1636^{Gal3Sulf}$ culture supernatant by MS

The analysis of the oligosaccharides present in $\Delta bt1636^{3S-Gal}$ culture supernatant after 96h incubation revealed that this sample is different from the cMO profile in the starting material (**Fig. 4c**, **Extended Data Fig. 8b** and **Supplementary Table 8**). We detected 114 glycans in the cMO sample, of which 39 were substituted with O-sulfate and L-fucose (44% total) (**Extended Data Fig. 8b**) and, the three most common structures (12% total) were 6S-GlcNAc oligosaccharides (**Fig. 4c** and **Supplementary Table 8**). In the control sample, the levels of sulfation, sialylation and fucosylation were 92%, 40% and 77% (**Supplementary Table 8**). In the $\Delta bt1636^{Gal-3Sulf}$ culture supernatant, we detected 72 glycans, of which 41 were substituted only with O-sulfate (84% total) (**Extended Data Fig. 8b**). In the mutant supernatant the levels of sulfation (95%) were similar to cMO, however the levels of sialylation (11%) and fucosylation (5%) decreased substantially (**Supplementary Table 8**), suggesting that the mutant is not able to utilize sulfated structures and these accumulate in culture media.

Additionally, a total of 98 of the 114 structures present in cMO were not detected in $\Delta bt1636^{3S-Gal}$ culture supernatant whereas in mutant supernatant, we detected 49 glycans that were not detected in the initial substrate (**Supplementary Table 8**). This unexpected result suggests that some of the oligosaccharides present in cMO can support the limited growth of $\Delta bt1636^{3S-Gal}$ and, although this mutant is not able to utilize many sulfated cMO structures it can still modify the glycans to create new structures. It remains unclear which enzymes are encoded by the mutant to modify

the O-glycans but the presence of a cell surface sialidase⁹ can explain the decrease of sialylation levels in structures found in *Δbt1636*^{3S-Gal} supernatant. Additionally, the presence of surface endo acting glycoside hydrolases able to cleave O-glycans into shorter oligosaccharides¹⁰ while preserving structures like sulfate at the non-reducing end can also contribute to new glycan structures in the mutant culture supernatant. Together these results show that *Δbt1636*^{3S-Gal} is not able to utilize most sulfated O-glycans explaining the limited growth of this mutant in cMO.

- 1 Arike, L. & Hansson, G. C. The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *J Mol Biol* **428**, 3221-3229, doi:10.1016/j.jmb.2016.02.010 (2016).
- 2 Arike, L., Holmen-Larsson, J. & Hansson, G. C. Intestinal Muc2 mucin O-glycosylation is affected by microbiota and regulated by differential expression of glycosyltransferases. *Glycobiology* **27**, 318-328, doi:10.1093/glycob/cww134 (2017).
- 3 Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **p54**, 1469-1476, doi:10.1099/ijls.0.02873-0 (2004).
- 4 Pudlo, N. A. *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).
- 5 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).
- 6 Hettle, A. G. *et al.* The Molecular Basis of Polysaccharide Sulfatase Activity and a Nomenclature for Catalytic Subsites in this Class of Enzyme. *Structure* **26**, 747-758 e744, doi:10.1016/j.str.2018.03.012 (2018).
- 7 Cartmell, A. *et al.* How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc Natl Acad Sci U S A* **114**, 7037-7042, doi:10.1073/pnas.1704367114 (2017).
- 8 Ndeh, D. *et al.* Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. *Nat Commun* **11**, 646, doi:10.1038/s41467-020-14509-4 (2020).
- 9 Briiliute, J. *et al.* Complex N-glycan breakdown by gut *Bacteroides* involves an_{SI}. *Nat Microbiol* **4**, 1571-1581, doi:10.1038/s41564-019-0466-x (2019).
- 10 Crouch, L. I. *et al.* Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nat Commun* **11**, 4017, doi:10.1038/s41467-020-17847-5 (2020).