




**Please cite the Published Version**

Zhang, Z, Dong, M, Zallot, R , Blackburn, GM, Wang, N, Wang, C, Chen, L, Baumann, P, Wu, Z, Wang, Z , Fan, H , Roth, C, Jin, Y and He, Y (2023) Mechanistic and Structural Insights into the Specificity and Biological Functions of Bacterial Sulfoglycosidases. *ACS Catalysis*, 13 (1). pp. 824-836. ISSN 2155-5435

**DOI:** <https://doi.org/10.1021/acscatal.2c05405>

**Publisher:** American Chemical Society (ACS)

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/636705/>

**Usage rights:**  In Copyright

**Additional Information:** This is an author accepted manuscript of an article published in *ACS Catalysis*, by American Chemical Society.

**Enquiries:**

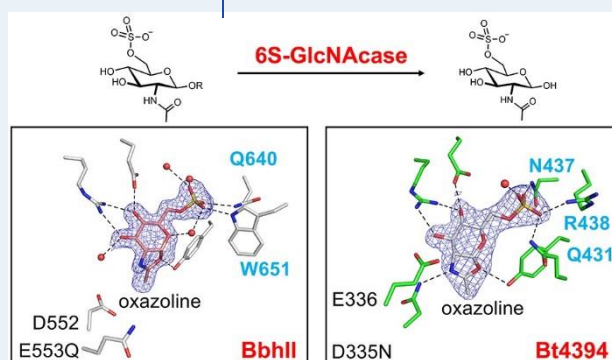
If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

# Mechanistic and Structural Insights into the Specificity and Biological Functions of Bacterial Sulfoglycosidases

Zhen Zhang,<sup>▽</sup> Mochen Dong,<sup>▽</sup> Rémi Zallot,<sup>▽</sup> George Michael Blackburn, Nini Wang, Chengjian Wang, Long Chen, Patrick Baumann, Zuyan Wu, Zhongfu Wang, Haiming Fan, Christian Roth, Yi Jin,<sup>\*</sup> and Yuan He<sup>\*</sup>

**ABSTRACT:** Glycan sulfation is an important modification supporting the functionalities of many proteins in biology. *Exo*-acting 6S-GlcNAcases from human microbiota are glycosidases that participate in the removal of 6-sulfo-GlcNAc from host glycans and thereby play an important role in human health and disease. Nonetheless, mechanisms underlying their ability to recognize the sulfate group remain poorly understood. Using structural and kinetic analyses, we here reveal the catalytically important amino acids directly involved in the recognition and cleavage of 6S-GlcNAc, but not of 6-phospho-GlcNAc, in BbhII from *Bifidobacterium bifidum*, Bt4394 from *Bacteroides thetaiotaomicron*, and SGL from *Prevotella* spp. The defining features of their sulfate recognition motifs underpin a genomic enzymological exploration of 6S-GlcNAcases to identify a wider range of human health-associated bacterial species having 6S-GlcNAcase activity. Our data provide significant insights into distinct molecular mechanisms of sulfated sugar recognition employed by 6S-GlcNAcases from both Gram-positive and Gram-negative bacteria along with valuable information for the exploration of extensive interactions between microbiota and their host glycans.

**KEYWORDS:** human microbiota, glycan sulfation, glycosyl hydrolase, *N*-acetyl-6-*O*-sulfo-*D*-glucosamine, oxazoline intermediate, genomic enzymology

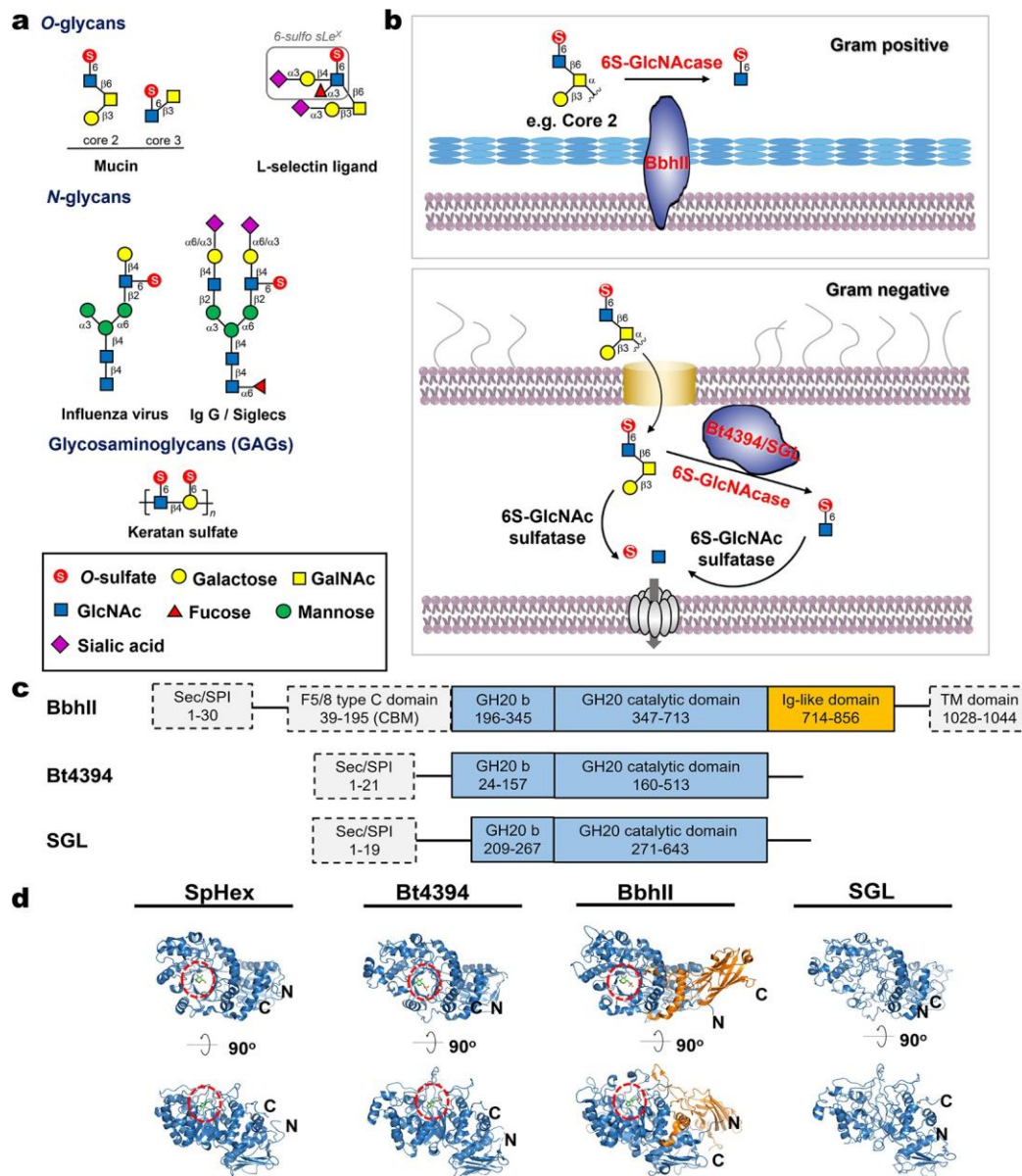


## INTRODUCTION

Glycan sulfation is a major type of enzyme-catalyzed post-glycosylation modification (PGM).<sup>1</sup> Sulfation has been found in *N*- and *O*-glycans as well as in proteoglycans, including glycosaminoglycans (GAGs), decorating *N*-acetylglucosamine (6S-GlcNAc), *N*-acetylgalactosamine (6S-GalNAc and 4S-GalNAc), galactose (3S-Gal, 4S-Gal, and 6S-Gal), and mannose (6S-Man)<sup>2-6</sup> (Figure 1a). Sulfated *N*-glycans on neuraminidase and hemagglutinin play a role in the antigenicity and elicitation of innate immunity of influenza virus.<sup>4</sup> Human coronaviruses bind to the sulfated *N*-glycans from the human lung.<sup>5</sup> Sulfation of  $\alpha$ -2,6-sialyl LacNAc affects its binding affinity to Siglec-2 in the regulation of immune cell function.<sup>7</sup> The 6-sulfo sialyl Lewis X glycan serves as a ligand for L-selectin in the targeting of leukocyte homing to endothelial cells.<sup>8</sup> The negatively charged sulfate group also plays a role in pathogenic/commensal bacterial colonization during the bacterial adhesion to the mucin *O*-glycans at various locations, such as the bronchial airway,<sup>9</sup> lung,<sup>3</sup> and ovary cyst.<sup>10</sup> The level of sulfation can alter the physicochemical properties of mucins<sup>11</sup> which act as a barrier between human microbiota and the epithelium. This is evidenced by the association of sulfation pattern change with a

compromised functionality of the mucus barrier, which is clinically linked to inflammatory bowel disease, colitis, Crohn's disease, carcinoma, and cystic fibrosis.<sup>9,12-14</sup>

Because sulfated glycans have so many implications in various biological processes, the identification and characterization of the bacterial enzymes involved in desulfation are essential to understanding the roles that bacteria play in these events. Human microbiota are known to use either or both of the two strategies to remove the sulfation: desulfation with sulfatases and glycosidase action to remove a non-reducing end-sulfated sugar by sulfoglycosidases. Exclusive sulfatase activity permits a wide variety of human microbiota to acquire uncapped glycans from host sulfated glycans by using digestive carbohydrate-active enzymes (CAZymes).<sup>13-17</sup> By comparison, only a few bacterial



**Figure 1.** (a) Representative *N*-, *O*-, and proteoglycan units that have  $\beta$ -linked 6S-GlcNAc as a building block. (b) Schematic representation of strategies for 6S-GlcNAc extracellular desulfation from a Core 2 oligosaccharide by Gram-positive bacteria and by Gram-negative bacteria in the periplasm using either or both 6S-GlcNAc sulfatases and 6S-GlcNAcases. (c) Domain arrangement of GH20 6S-GlcNAcase enzymes BbhII, Bt4394, and SGL. All the three enzymes have a type I signal peptide (Sec/SPI). BbhII also has an N-terminal F5/8 type C domain which is annotated as a carbohydrate binding domain32 (CBD32), Ig-like domain (in orange), and TM-domain (GH20 catalytic domains in light blue). (d) Comparison of the fold and domain structures of Bt4394, BbhII, and SGL characterized in this study, with SpHex (PDB: 1hp5, active sites highlighted by a red circle).

sulfoglycosidases have yet been reported, and they are predominantly 6S-GlcNAcases, which are responsible for the cleavage of terminal 6S-GlcNAc on oligosaccharides.

*Bifidobacterium bifidum* uses BbhII, a 6S-GlcNAcase,<sup>18,19</sup> and can grow on mucin *O*-glycans as the sole nutrient source<sup>20,21</sup> (Figure 1b). *Bacteroides thetaiotaomicron* has sulfatase-encoding gene clusters<sup>15,17,22</sup> that enable the intake of highly sulfated GAGs as priority nutrients<sup>23</sup> and also mucin *O*-glycans.<sup>17,24</sup> It also has a putative 6S-GlcNAcase, Bt4394<sup>25</sup> (Figure 1b). Similarly, *Prevotella* Strain RS2 encodes sulfatases, such as MdsA,<sup>16</sup> while it also has a periplasmic 6S-GlcNAcase, SGL, that releases 6S-GlcNAc from sulfomucin *O*-glycan (Figure 1b).<sup>25</sup> Recently, *Bacteroides caccae* and *Phocaeicola dorei* have also been found to encode 6S-GlcNAcases that release 6S-GlcNAc from

*N*-glycans.<sup>6</sup> These bacteria are prominent members of human microbiota. In addition, Niako3494 from a soil bacterium, *Niastella koreensis*, has been reported to have the ability to hydrolyze 6S-GlcNAc from high-throughput screening.<sup>26</sup> Despite the ever-increasing interest in characterizing new bacterial 6S-GlcNAcases and in studying their action on the host glycans, this task has been greatly hindered by a lack of information on the recognition mode of the sulfate group employed by 6S-GlcNAcases.

Based on the conserved patterns of residues participating in the so-called “substrate-assisted hydrolysis”, all 6S-GlcNAcases identified to date belong to the glycosyl hydrolase 20 (GH20) CAzy family, which also includes *exo*-acting  $\beta$ -*N*-acetylglucosaminidases,  $\beta$ -*N*-acetyl-galactosaminidase, and *endo*-acting lacto-

Table 1. Kinetic Parameters for the Hydrolysis of 4MU-6S-GlcNAc, 4MU-GlcNAc, and pNP-6P-GlcNAc by Wild-Type BbhII, Bt4394, SGL, and Niako3494<sup>a</sup>

		$K_M$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\text{s}^{-1} \text{M}^{-1}$ )	relative activity (%)	pH		
BbhII	wild-type	16 ± 1	44 ± 1	$(2.7 \pm 0.8) \times 10^6$	100	6.0		
	D552N	2.1 ± 0.2	$(6.1 \pm 0.1) \times 10^{-2}$	$(2.9 \pm 0.5) \times 10^4$	1.1			
	E553Q	1.1 ± 0.1	0.11 ± 0.03	$(1.0 \pm 0.0) \times 10^5$	3.7			
	D463N	16 ± 1	29.0 ± 0.6	$(1.8 \pm 0.5) \times 10^6$	66			
	H467F	5.2 ± 0.6	2.9 ± 0.0	$(5.6 \pm 0.0) \times 10^5$	21			
	Y637F	0.6 ± 0.1	$(7.3 \pm 0.1) \times 10^{-2}$	$(1.2 \pm 0.1) \times 10^5$	4.0			
	Q640D	122 ± 12	8.0 ± 0.2	$(6.6 \pm 1.7) \times 10^4$	2.4			
	Q640K	107 ± 11	11.4 ± 0.3	$(1.1 \pm 0.3) \times 10^5$	4.1			
	W651F	31 ± 3	24.3 ± 0.5	$(7.9 \pm 2.0) \times 10^5$	29			
	C656S	14.0 ± 0.7	11.4 ± 0.1	$(8.1 \pm 1.4) \times 10^5$	30			
	4MU-GlcNAc	907 ± 150	1.04 ± 0.05	$(1.1 \pm 0.3) \times 10^3$	0.04			
pNP-6P-GlcNAc	2489 ± 334	4.1 ± 0.2	$(1.6 \pm 0.6) \times 10^3$	0.06	7.0			
Bt4394	wild-type	39 ± 4	25.8 ± 0.7	$(6.7 \pm 1.6) \times 10^5$	100	5.5		
	D335N	332 ± 38	$(2.4 \pm 0.1) \times 10^{-3}$	7.2 ± 2.7	0.001			
	E336Q	2.3 ± 0.3	0.11 ± 0.01	$(4.8 \pm 4.0) \times 10^4$	7.2			
	D266N	77 ± 5	22.7 ± 0.6	$(2.9 \pm 1.2) \times 10^5$	43.3			
	H270F	552 ± 56	$(1.06 \pm 0.04) \times 10^{-2}$	19.2 ± 7.1	0.003			
	Y435F	1.0 ± 0.1	$(3.0 \pm 0.1) \times 10^{-2}$	$(2.9 \pm 0.8) \times 10^4$	4.3			
	Q431E	271.1 ± 30.9	1.30 ± 0.04	$(4.8 \pm 1.3) \times 10^3$	0.72			
	N437D	708 ± 86	12.6 ± 0.6	$(1.8 \pm 0.7) \times 10^4$	2.7			
	R438A	662 ± 93	9.4 ± 0.5	$(1.4 \pm 0.5) \times 10^4$	2.1			
	4MU-GlcNAc	2183 ± 189	2.9 ± 0.1	$(1.3 \pm 0.5) \times 10^3$	0.19			
	pNP-6P-GlcNAc	2207 ± 270	2.7 ± 0.1	$(1.2 \pm 0.4) \times 10^3$	0.18		7.0	
	SGL	wild-type	29 ± 3	38 ± 1	$(1.3 \pm 0.4) \times 10^6$		100	6.0
		4MU-GlcNAc	2232 ± 258	2.5 ± 0.1	$(1.1 \pm 0.4) \times 10^3$		0.08	
pNP-6P-GlcNAc		2591 ± 234	15.9 ± 0.7	$(6.1 \pm 3.0) \times 10^3$	0.5	7.0		
Niako3494	wild-type	756 ± 72	$(2.8 \pm 0.1) \times 10^{-2}$	37 ± 14	100	5.0		
	D196N	848 ± 131	$(3.7 \pm 0.2) \times 10^{-5}$	$(4.4 \pm 1.5) \times 10^{-2}$	0.12			
	E197Q	438 ± 46	$(9.8 \pm 0.3) \times 10^{-5}$	0.2 ± 0.1	0.54			
	4MU-GlcNAc	1233 ± 126	$(4.6 \pm 0.2) \times 10^{-2}$	37 ± 16	100			

<sup>a</sup>Activities of their variants were measured using 4MU-6S-GlcNAc as a substrate.

*N*-biosidases. However, given the high holistic sequence similarity among these enzymes, there is a great challenge in the identification and prediction of new 6S-GlcNAcases solely *via* a bioinformatics approach, independent of the knowledge of the sequence motifs for sulfate recognition that are distant in sequence from the key catalytic residues.

In this work, we have used a structural enzymology approach to explore the molecular mechanism of 6S-GlcNAcase catalysis with particular attention to the recognition of the 6'-sulfate, to understand whether there are any common features of sulfate recognition among the so far identified 6S-GlcNAcases, and to facilitate a rationale exploration of this sub-family of enzymes using a genomic enzymology approach. We selected an extracellular 6S-GlcNAcase, BbhII, from the Gram-positive bacterium<sup>19</sup> to compare with putative 6S-GlcNAcases from the Gram-negative bacteria, Bt4394,<sup>25</sup> SGL,<sup>25</sup> and Niako3494.<sup>26</sup> The crystal structures and mutagenesis have revealed three distinct sulfate binding motifs in BbhII, Bt4394, and SGL, while Niako3494 exhibited no 6S-GlcNAcase-specific activity, and no evident sulfate binding residues were found in the crystal structure. By exploring the specificity toward the sulfated substrate 4-methylumbelliferyl 6-sulfo-2-acetamido-2-deoxy- $\beta$ -D-glucopyranoside (4MU-6S-GlcNAc) and the phosphorylated substrate *p*-nitrophenyl 6-phospho-2-acetamido-2-deoxy- $\beta$ -D-

glucopyranoside (*p*NP-6P-GlcNAc) using kinetic analysis, it became clear that the ability of these GH20 6S-GlcNAcases to recognize the 6'-sulfate is specific, although the 6'-phosphate might also appear to have the same tetrahedral geometry and anionic charge at physiological pH. Independently, bioinformatics investigation with Enzyme Function Initiative (EFI) web tools (<https://efi.igb.illinois.edu/>) has mapped conserved sulfate-recognizing sequence motifs within the GH20 family and identified other members of human microbiota that have 6S-GlcNAcase capability. In light of the ever-increasing important roles played by sulfated glycans in human health and disease, a thorough mutagenic, structural, and bioinformatic analysis provides here a systematic approach to study the structural features that define this family of 6S-GlcNAc-acting sulfoglycosidases and guides the search for novel 6S-GlcNAcases. Knowledge of this family of enzymes is not only valuable in advancing our understanding of host-microbiota interactions but could also allow these enzymes to be used or evolved for practical applications, such as engineering natural glycans (in collaboration with other GH enzymes) and aiding glycomic identification of sulfo-sugars.<sup>27</sup>

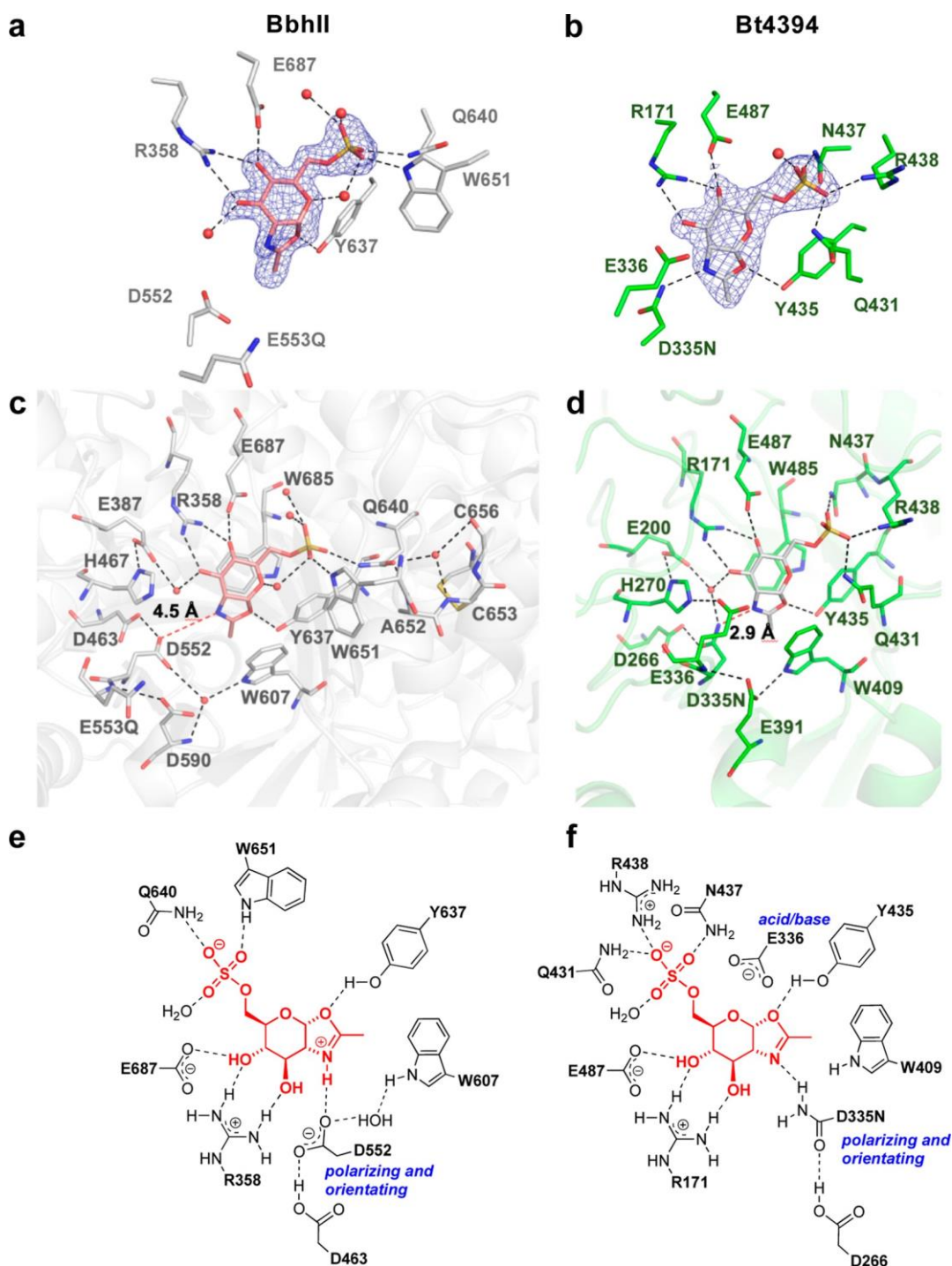
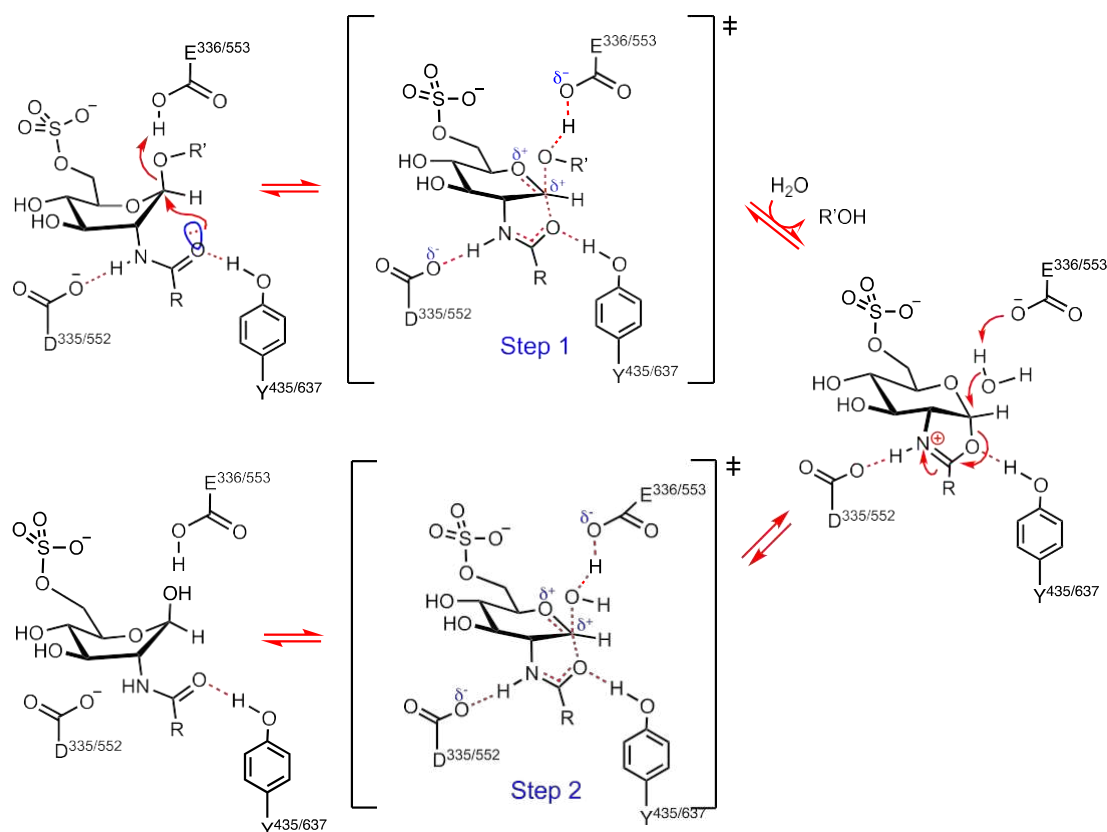


Figure 2. Active sites of BbhII and Bt4394 with 6S-NAG-oxazoline.  $2F_o - F_c$  map contoured at  $1.0\sigma$  ( $0.21$  and  $0.27 e^{-}/\text{\AA}^3$ , respectively) for the 6S-NAG-oxazoline intermediate in complex with (a) BbhII<sub>E553Q</sub> and (b) Bt4394<sub>D335N</sub>. (c) H-bonding network around sugar and sulfate binding sites in BbhII. 4'-OH could be recognized by R358 and E687 in both glucose and galactose. The catalytic loop is in an "open" conformation that is not competent for catalysis (as D552 is  $4.5 \text{ \AA}$  distant from the oxazoline nitrogen), while Y637 is only  $2.9 \text{ \AA}$  distant from the oxazoline oxygen. (d) H-bonding network around sugar and sulfate binding sites in Bt4394. The catalytic loop is in a catalytically competent conformation. N335 is  $2.9 \text{ \AA}$  distant from the oxazoline nitrogen, and Y435 is  $2.7 \text{ \AA}$  distant from the oxazoline oxygen (H-bonds between the ligand and protein in black dashes). (e,f) Schematic representation of the H-bonding network within the active site between BbhII<sub>E553Q</sub> and 6S-NAG-oxazoline (e) and between Bt4394<sub>D335N</sub> and 6S-NAG-oxazoline (f).

## RESULTS

**Sulfate Recognition by BbhII from Gram-Positive *B. bifidum*.** BbhII orthologues from *B. bifidum* JCM7004 and JCM1254 have been proposed as extracellular membrane-bound 6S-GlcNAcases with a type I signal peptide (Figure

1c).<sup>19,28</sup> We grew *B. bifidum* JCM1254 on an anaerobic agar supplemented with 1% (w/v) a type II porcine gastric mucin and showed that *B. bifidum* JCM1254 whole cells hydrolyze 4MU-6S-GlcNAc (Figure S1), as well as liberate 6S-GlcNAc directly



**Figure 3.** Schematic presentation of the catalytic mechanism of 6S-GlcNAcase in both BbhII and Bt4394. Step 1: 2-acetamido- $\beta$ -D-glucopyranose has its NH coordinated by aspartate, its amide oxygen aligned for C1 attack with inversion of stereochemistry, and the leaving oxygen of the glycosidic bond coordinated to glutamic acid, leading to the formation of the *cis*-fused oxazoline ring. Following replacement of the leaving group by water, the reaction proceeds in step 2 *via* deprotonation of an incoming water and general acid catalysis of the amide nitrogen to open the intermediate oxazoline with overall C1-retention of configuration.

from gastric mucin *O*-glycans by LC-ESI-MS with a  $m/z$  of 300.08 (Figure S2c).

To enable detailed characterization, the BbhII gene from *B. bifidum* JCM1254 was cloned and overexpressed in *E. coli* as a His<sub>6</sub>-tagged protein. Recombinant BbhII shows 6S-GlcNAcase activity on a porcine gastric mucin glycan (Figure S2d) and displays maximum activity at pH 6.0 toward 4MU-6S-GlcNAc in a fluorometric kinetics assay (Figure S3r), which is appropriate for an extracellular enzyme from *B. bifidum* that colonizes the human intestine where the pH is 6–7.5.<sup>29</sup> Michaelis–Menten parameters show that BbhII has a substrate preference (using  $k_{\text{cat}}/K_M$  values) of 2500-fold for sulfated 4MU-6S-GlcNAc over that for 4MU-GlcNAc (Figures S4 and S5 and Table 1), substrates that differ only in the presence or absence of a 6-sulfate motif.<sup>30</sup> Compared to the wild-type BbhII, the polarizing residue D552N variant exhibits a 93-fold reduction in  $k_{\text{cat}}/K_M$  and a ~7.6-fold decrease in  $K_M$ , whereas the general acid/base (GAGB) residue E553Q variant only gives a 27-fold  $k_{\text{cat}}/K_M$  reduction with a 14-fold  $K_M$  decrease (Table 1 and Figure S5).

We crystallized full-length BbhII<sub>WT</sub> (1.75 Å, Table S2) and the GAGB variant BbhII<sub>E553Q</sub> in their apo form and solved their structures by molecular replacement using *Streptomyces plicatus* hexosaminidase (SpHEX, PDB: 1hp5) as the model. The electron densities for the GH20 catalytic b domain and the GH20 catalytic domain are clearly defined and show a ( $\beta/\alpha$ )<sub>8</sub>-barrel fold almost identical to that of SpHEX, a prototypical GH20 enzyme (Figure 1d). The Ig-like domain (residues 714–856) is also well ordered. In addition, we soaked the 4MU-6S-

GlcNAc substrate into BbhII<sub>E553Q</sub> apo form crystals to give a complex (BbhII<sub>E553Q</sub>-6S-GlcNAc-oxazoline) that diffracted well (1.67 Å, Table S2). This structure showed a clear electron density for the trapped 6S-NAG-oxazoline intermediate and revealed its pyranose ring in a <sup>4</sup>C<sub>1</sub> conformation (Figure 2a). The bidentate coordination to the 3'- and 4'-OH groups of the 6S-GlcNAc glycoside by R358 and E687 (Figure 2a,c,e) would also allow 4'-OH in an axial configuration, which explains why BbhII also hydrolyzes 6S-GalNAc.<sup>24</sup> Alignment of the 6S-NAG-oxazoline structure with the thiazoline-containing structure of SpHex (Figure S6a) shows the catalytic loop, containing GAGB and polarizing residues, a catalytic diad, which is in a structurally incompetent conformation, with E553Q 10.5 Å distant from the anomeric carbon, while D552 is 4.5 Å distant from the oxazoline ring nitrogen. This conformation has a water molecule bridging the side chains of W607 and D590, which is H-bonded to the backbone amide of E553Q (Figure 2c). This structure shows that the reaction can proceed in *crystallo* for step 1 (Figure 3) as a result of soaking the BbhII<sub>E553Q</sub> crystal in a synthetic substrate that has a good C-1 leaving group (4MU pK<sub>a</sub> 7.8) and does not depend on general acid catalysis.<sup>31</sup> Equally, the hydrolysis of this bicyclic intermediate required in step 2 calls for both general base and polarizing activities to deliver the final product of the glycolytic reaction.

In the sulfate binding site within the sequence Q(X<sub>10</sub>)WAC (sulfate binding residues in bold), two sulfonyl oxygens are strongly coordinated directly, each by one H-bond each from the side chains of Q640 and W651 (Figure 2a,c,e). The third

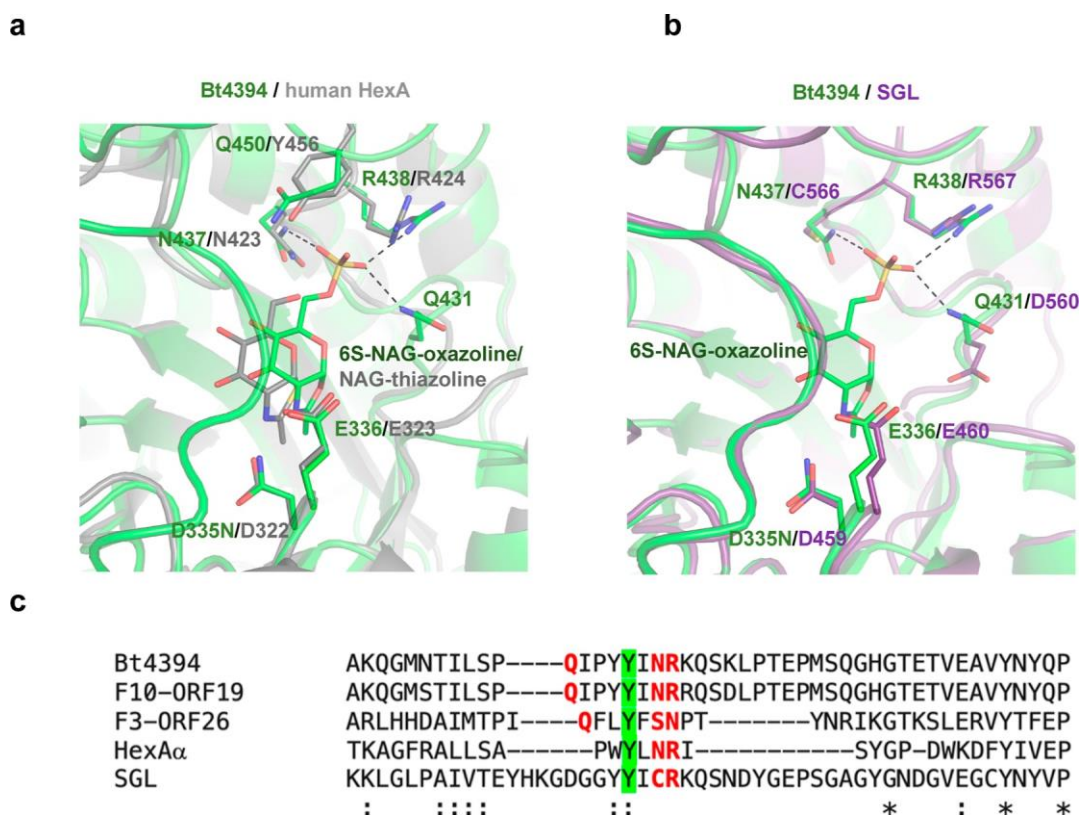


Figure 4. (a) Overlay of active site structures of Bt4394<sub>D335N</sub>-6S-NAG-oxazoline intermediate complex and human HexA-NAG-thiazoline complex (PDB: 2gk1) showing good alignment of N423 and R424 in the sulfate recognition motif of human HexA subunit- $\alpha$  with the sulfate recognition residues N437 and R438 in Bt4394. (b) Overlay of structures of Bt4394<sub>D335N</sub>-6S-NAG-oxazoline intermediate complex and apo SGL showing the sulfate binding residues in SGL, C566, and R567, align well with N437 and R438 in Bt4394, and are within H-bonding distance to the sulfate group. (c) MSA by T-coffee with the default setting for Bt4394, human HexA  $\alpha$ -subunit, F10-ORF19 and F3-ORF26, and SGL showing the predicted sulfate binding residues (red) and the catalytic tyrosine that coordinates with the oxazoline ring oxygen (green).

sulfonyl oxygen is coordinated only indirectly *via* water molecules to E687. To examine the contribution of these residues in the first and second shells of the sulfate binding site, we explored a range of amino acid mutations. Compared to BbHII<sub>WT</sub>, variants Q640D, Q640K, and W651F show increases in  $K_M$  of 8-, 7-, and 2-fold, with  $k_{cat}/K_M$  reduced by 41-, 24-, and 3.4-fold, respectively (Table 1). This identifies them as significant components of sulfate coordination and explains why sulfation increases the substrate affinity ( $K_M$ ) by 57-fold and specificity by 2500-fold ( $k_{cat}/K_M$ ). Clearly, Q640 has a more important role in sulfate recognition than W651. The observed disulfide bond between C653 and C656 in the second shell have an impact to a lesser extent on the binding of the sulfate as the variant C656S shows only *ca.* 3-fold reduction in catalytic efficiency.

**Sulfate Recognition by Bt4394 from Gram-Negative *B. thetaiotaomicron*.** Bt4394 has been proposed as being responsible for 6S-GlcNAcase activity from the cell extract of the Gram-negative bacterium *B. thetaiotaomicron*,<sup>25</sup> with an increase in Bt4394 gene expression in response to host gut mucin glycans.<sup>24</sup> While its N-terminal type I signal peptide has been used to predict Bt4394 as a periplasmic enzyme<sup>32</sup> (Figure 1c), no detailed characterization has yet been delivered. Because some *B. thetaiotaomicron* proteins, originally thought to be periplasmic, are in fact located on the cell surface<sup>17</sup> and the cellular localization is directly linked to facilitate access of their substrates to these enzymes, we tested the hydrolytic activity of 6S-GlcNAc by a fluorometric assay with whole cell *B.*

*thetaitaomicron*. We found that there was no 6S-GlcNAcase activity toward 4MU-6S-GlcNAc (Figure S7). This provides strong evidence that Bt4394 is indeed a periplasmic enzyme.

We next expressed the recombinant His<sub>6</sub>-tagged Bt4394 in *E. coli*. It displays a specificity toward 4MU-6S-GlcNAc that is 515-fold higher than that with 4MU-GlcNAc (Table 1), validating it being a 6S-GlcNAcase. As no extra inhibitory effect on enzyme activity results from the addition of 200 mM Na<sub>2</sub>SO<sub>4</sub> or Na<sub>2</sub>SO<sub>3</sub>, there appears to be no cooperative binding between NAG-thiazoline and sulfate or sulfite (Figure S8a).<sup>33</sup> Recombinant Bt4394 with 4MU-6S-GlcNAc as a substrate has an optimal pH 5.5 (Figure S3s). Bt4394 has  $k_{cat} = 25.8 \text{ s}^{-1}$ ,  $K_M = 39 \text{ }\mu\text{M}$ , and  $k_{cat}/K_M = 6.7 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$  at pH 5.5 (Table 1 and Figure S9). The catalytic residues are predicted to be D335 and E336. Our data show an 8.5-fold increase in  $K_M$  for the D335N variant and a 17-fold decrease in  $K_M$  for the E336Q variant (Table 1).

To gain insights into the reactivity and recognition of the sulfate motif, we first crystallized Bt4394<sub>WT</sub> and Bt4394<sub>D335N</sub> variant proteins and solved their apo form structures at high resolution (>1.6 Å) by molecular replacement (Bt4394<sub>WT</sub> apo; Table S2). After soaking the apo Bt4394<sub>D335N</sub> crystals in the substrate 4MU-6S-GlcNAc (Figure 1d), the clear electron density defines a <sup>4</sup>C<sub>1</sub> conformation for a trapped 6S-NAG-oxazoline intermediate (2.05 Å; Bt4394<sub>D335N</sub>-6S-GlcNAc-oxazoline; Figure 2b and Table S2). Although the rotameric orientation of the 6S-GlcNAc-oxazoline residue is the same in both BbHII and Bt4394, it is evident that the Bt4394<sub>D335N</sub>-6S-NAG-oxazoline complex has loop 332–339 in a catalytic

conformation which places the D335N side chain amide just 2.9 Å from N1 in the oxazoline ring (Figure 2d), as has been seen in GH20  $\beta$ -*N*-acetylhexosaminidase from *Streptomyces coelicolor* A3 (SchHex, PDB: 4c7g, Figure S10a) and other GH20s (Figure S10b).<sup>34</sup> The polarizing residue variant D335N carboxamide-NH<sub>2</sub> group donates a H-bond to the oxazoline nitrogen in its unprotonated form ( $pK_a$  5.5),<sup>35</sup> and this clearly rationalizes the formation and stability of the 6S-NAG-oxazoline complex (Figure 2d,f). The carboxylate of the GAGB residue E336 forms a H-bond with H270 (Figure 2d,f). We also obtained a Bt4394<sub>WT</sub>-6S-GlcNAc product complex structure (Table S2), which has 6S-GlcNAc in a <sup>1,4</sup>B conformation. Taken as a whole, these data are consistent with a mechanism of “substrate-assisted catalysis” with a conformational itinerary of <sup>4</sup>C<sub>1</sub> → <sup>4</sup>E<sup>‡</sup> → <sup>1,4</sup>B for the intermediate hydrolysis half-reaction. A loop region of residues 278–307 in Bt4394 structures forms a cleft and brings Y228 closer to H452. In contrast, the same loop (478–511) in BbhII has a more open position (Figure S11). This suggests that the leaving group glycosides in the substrates of these two enzymes may well be of significantly different sizes.

In the sulfate binding site within the sequence of Q<sub>431</sub>IPYYIN<sub>437</sub>R<sub>438</sub>, one sulfonyl oxygen is coordinated directly by two H-bonds from Q431 and R438 and a second oxygen accepts a H-bond from N437 (Figure 2b,d,f). The third sulfonyl oxygen is not protein-coordinated but only a water molecule. The variants Q431E, N437D, and R438A show diminished affinity by 7-, 18-, and 17-fold with corresponding reductions in  $k_{cat}/K_M$  of 139-, 37-, and 47-fold, respectively, thereby demonstrating the importance of these three residues delivering sulfate recognition by both H-bonding and electrostatic interactions (Table 1). At long last, this result explains the 6S-GlcNAcase activity of Tay-Sachs and Sandhoff disease-causing human HexA subunit- $\alpha$  observed in an early study.<sup>36</sup> Its N423 and R424 residues align perfectly with the sulfate binding residues N437 and R438 in Bt4394 structure (Figure 4a). Our results also rationalize a recently reported 6S-GlcNAcase activity of F3-ORF26 from *P. dorei*, with the sulfate binding sequence QFLYFSNP identified by multiple sequence alignment (MSA) (Figure S12).<sup>6</sup> It is noteworthy that this recognition pattern for Bt4394 contrasts directly with the requirement for an indispensably conserved arginine in the sulfoquinovose hydrolase from the GH31 family for the recognition of the sulfonate monoanion of sulfoquinovose diacylglycerol (SQDG).<sup>37,38</sup>

**Sulfate Binding Site of SGL from Gram-Negative *Prevotella* RS2.** SGL, another 6S-GlcNAcase, shows a 1200-fold higher specificity toward 4MU-6S-GlcNAc over 4MU-GlcNAc at optimal pH 6.0 (Table 1 and Figures S4t and S13) but has a low similarity sequence GGY<sub>564</sub>IC<sub>566</sub>R<sub>567</sub> (Figure 4c) at the sulfate binding region compared to Bt4394 and BbhII according to sequence alignment. Therefore, to strengthen our identification of the key sulfate binding residues, we expressed the *sgl* gene recombinantly and purified the protein, leading to a 2.7 Å resolution crystal structure (Table S2). When overlaid with the 6S-NAG-oxazoline complex structures of Bt4394 and BbhII, the apo-SGL structure showed that only residues C566 and R567 were able to donate two H-bonds to the sulfate. D560 aligns well with Q431 in Bt4394, but it is unlikely to be protonated or to be within the H-bonding distance to participate in sulfate binding (Figure 4b).

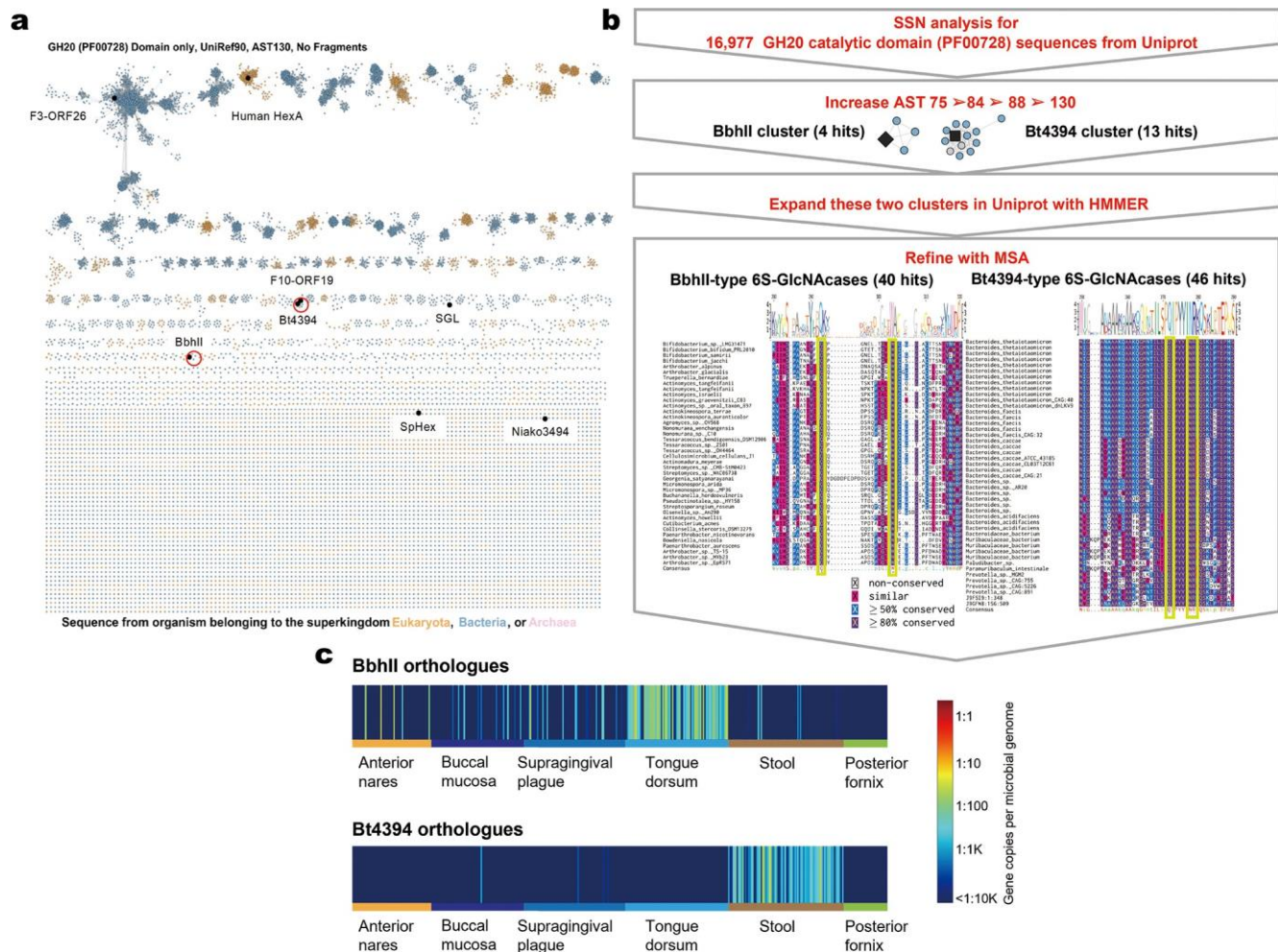
**Niako3494 Is Not a Sulfoglycosidase.** Niako3494 has been reported to have 6S-GlcNAcase activity in a high-throughput assay without detailed characterization,<sup>26</sup> although it is only defined as a “distant” member of the GH20 family. To

unlock the mystery of its sulfate binding ability, we cloned and expressed the *niako3494* gene. Our kinetic analysis identified only weak activity with  $k_{cat} = 0.028 \text{ s}^{-1}$ ,  $K_M = 756 \mu\text{M}$ , and  $k_{cat}/K_M = 37 \text{ s}^{-1} \text{ M}^{-1}$  at pH 5.0 toward 4MU-6S-GlcNAc and a similar weak specificity for 4MU-GlcNAc (Table 1 and Figure S14). This is 105 orders of magnitude lower than other 6S-GlcNAcases we have studied. This conflict with the published data led us to crystalize and obtain a structure of Niako3494 at 2.27 Å resolution (Table S2). Although Niako3494 has conserved residues H122, D196, E197, W239, W277, and W342 seen in the active site as other GH20 enzymes (Figure S15), the “back-up” aspartic acid (D463 for BbhII and D266 for Bt4394) has been replaced by N118 to abut onto the carboxyl group of the polarizing aspartate (D196 for Niako3494, D552 for BbhII, and D335 for Bt4394). Given that the asparagine variants BbhII<sub>D463N</sub> and Bt4394<sub>D266N</sub> both retain half of their activity relative to the wild-type enzymes (Table 1), the complete inactivity of Niako3494 to 4MU-6S-GlcNAc and 4MU-GlcNAc must have a different origin. We see that the active site of Niako3494 is more open and lacks residues for the recognition of 3'- and 4'-OH of NAG (Figure S15a,b). This suggests that it may well have *endo*-glycosidase activity for larger substrates and explains why *niako3494* has poor *exo*-hexaminidase activity toward 4MU-GlcNAc. We note that all other 6S-GlcNAcases, including BbhII and HexA, have the conserved oxazoline-coordinating tyrosine near the sequence of the sulfate binding motif, but in Niako3494, the equivalent tyrosine Y279 comes into the active site from the opposite side (Figure S15c). This is because it is located in the same loop as the oxazoline-flanking W277 and may clash with the *N*-acetyl group of NAG. More importantly, when compared to the Bt4394<sub>D335N</sub>-6S-NAG-oxazoline complex structure, only two residues C301 and S302 are near the sulfate binding site, though they are still beyond the H-bonding range to engage sulfate (Figure S15a,d). Taken together, these variations all explain why Niako3494 is only a distant member of the GH20 family and lacks sulfoglycosidase activity.

**Sulfoglycosidases Are Not Phosphoglycosidases.** Phosphorylation on the 6'-position of GlcNAc (6P-GlcNAc) has been found to have biological relevance but rarely reported. Thus, *O*-linked 6P-GlcNAc has been found on the synapse-specific clathrin assembly protein AP180 as a PGM to increase the protein net negative charge and hydrophilicity in synaptic vesicle endocytosis.<sup>39,40</sup> Also, 6P-GlcNAc has also been found in bovine colostrum and milk saccharides.<sup>41,42</sup> While the tetrahedral geometry and bond lengths are similar for phosphate and sulfate monoesters, they differ in basicity. The GlcNAc 6-phosphate monoester has  $pK_{a2} \sim 6^{43}$  and thus can be either dianionic or monoanionic with an acidic-OH group to donate a H-bond. A 6'-sulfate monoester is only monoanionic above pH 1 and can only be a H-bond acceptor at physiological pH.

We therefore explored whether the 6S-GlcNAcases in the present study may also be 6P-GlcNAcases. To this end, the activities of BbhII, Bt4394, and SGL for a 6'-phosphate substrate were tested at pH 7.0 using *p*NP-6P-GlcNAc synthesized chemoenzymatically (Scheme S1). We selected pH 7.0 to ensure that the 6'-phosphate group was predominantly dianionic and would not act as a H-bond donor. Our data delivered specificity constants  $k_{cat}/K_M$  with 6P-GlcNAc of the same order of magnitude as for GlcNAc and 3 orders of magnitude lower than for 6S-GlcNAc (Table 1 and Figure S16), thereby showing that binding to a 6'-phosphate monoester is not effective, notwithstanding its greater negative charges.





**Figure 5.** (a) Mapping the known 6S-GlcNAcases (Bt4394, BbhII, SGL, human HexA, F3-ORF26, and F10-ORF19) and other GH20 enzymes (SpHex and Niako3494) (in •) on the UniRef90 SSN for just the GH20 catalytic domain (Protein Family PF00728). (b) Flowchart to show, with 16977 Uniprot GH20 catalytic domain sequences, an increased AST from 75 to 130 is needed to separate the subclusters as shown in (a), whereby the five subclusters represented by the five characterized 6S-GlcNAcases can be positively assigned, guided by the structurally confirmed sulfate binding motif sequences. This action initially identified 4 hits for the BbhII cluster and 13 hits for the Bt4394 cluster. Hence, a comprehensive search of unannotated genes belonging to GH20 in Uniprot was made with HMMER, which expanded the number of sequences of BbhII and Bt4394 clusters to 42 and 48, respectively. MSA refinement for the expanded BbhII and Bt4394 clusters was performed to ensure that all the sequences included were intact and satisfied the criteria identified by our mutagenic and structural analysis. This reduced the final numbers of sequences of BbhII and Bt4394 clusters to 40 and 46, respectively. Organism taxonomy (class level) is also depicted. (c) CGFP for the clusters that embrace organisms encoding Bt4394 and BbhII orthologues: heatmaps representing the quantification of sequences from SSN clusters per metagenome are available. The metagenomes are grouped according to body sites so that trends/consensus can be easily discerned across the six body sites. The trend is clear: while BbhII itself is isolated from human gut bacteria *Bifidobacterium bifidum*, BbhII orthologues are found more abundant in the tongue dorsum and other oral cavity locations, implying that these organisms take advantage of the sulfoglycans in the human diet or participate in oral homeostasis.

**Identification of More 6S-GlcNAcases Using the Genomic Enzymology Approach.** GH20 proteins have a wide variety of module and domain organizations for fulfilling their cellular functions.<sup>44</sup> The high sequence similarity between the GH20 catalytic domain-containing enzymes presents a practical challenge to identify putative 6S-GlcNAcase without the knowledge of the sequence motifs for sulfate recognition which are not close to the catalytic residues in the sequence. In our study, using 3D structure and mutagenesis analysis, we identified the sulfate binding motif of BbhII to be Q(X<sub>10</sub>)WAC and that of Bt4394 to be QIPYYINR (SGL with YYICR) (Figure 4). This information allows us to search for additional 6S-GlcNAcases with the sulfate binding motifs within the GH20 family across species using the genomic enzymology approach.<sup>45</sup> By using the EFI web tools, we first generated sequence

similarity networks (SSNs) of GH20 family proteins for PF00728 using the Protein families database Pfam.<sup>46</sup> The GH20 catalytic domain is often found to be associated with other domains, whose functions are not always understood.<sup>47</sup> Color nodes on the SSN based on the length of the polypeptides reveal the huge length variability within the family and illustrate high domain fusion occurrences with varying identities (Figure S17).

It is noteworthy that sequences with high similarity over a longer sequence portion tend to have a higher BLAST score. Thus, the high sequence similarity in GH20 catalytic domains and the diversity of the domain fusion pattern of the GH20 family both affect the BLAST score as well as the SSN patterns, making it difficult to identify GH20 enzymes with short sulfate binding motif sequences. To deconstrain the effect of pairwise

sequence similarity calculation from those fused domains, we generated a refined SSN for PF00728 with only the GH20 Pfam-defined domain for each sequence. This refined SSN focused only on the primary sequence elements that contain GH20 catalytic domains (Figure S18). A subsequent increase of the alignment score threshold (AST) from low (AST value 57 corresponding to  $\pm 35\%$  ID, Figure S17b) to high (AST 75,  $\pm 40\%$  ID, Figure S18a) separated clusters of Bt4394 and SGL<sup>25</sup> from clusters of BbhII and human lysosomal HexA.<sup>36</sup> To segregate more similar sequences of SGL and Bt4394, we had to apply an increase in AST from 84 to 130 (Figures 5a and S18b-d). Subsequently, a MSA for UniRef90 representative sequences from the Bt4394 and BbhII clusters identified the absolute conservation of residues for catalysis (Figure 5b). These two clusters were then used to search against the Uniprot database using the HMMER server (<https://www.ebi.ac.uk/Tools/hmmer/>) for additional sequences that may not have been initially annotated as GH20 protein family members in Pfam.<sup>48</sup> This approach expanded the two putative 6S-GlcNAcase orthologous groups to 46 and 40 sequences using the respective structural guided criteria with high confidence after excluding truncated and experimentally unsupported sequences (Figures 5b and S19 and S20). For example, this approach identified a recently discovered 6S-GlcNAcase F10-ORF19 from *B. caccae* with an identical sulfate binding sequence to Bt4394 (QIPYYINR).<sup>6</sup> Both orthologous groups also include 6S-GlcNAcases from previously unidentified species that are associated with human health. Moreover, this approach has demonstrated that it can be used to identify more putative enzymes for other GH families with “high domain fusion” but with short motifs, such as sulfate binding motif sequences in BbhII and Bt4394 in this study. SGL, however, is a unique sequence in our network.

Subsequently, we used the clusters of BbhII and Bt4394 as an entry point for a chemically guided functional profiling (CGFP) search<sup>45</sup> in order to evaluate the occurrence of members of the defined orthologous group in the Human Microbiome Project (HMP) metagenome. Interestingly, both orthologue groups are found to be present in the buccal mucosa, supragingival plaque, and gut-associated metagenomes (Figure 5c), in accordance with the identified known habitat of the species to which they belong to. More interestingly, Gram-positive bacteria that encode BbhII orthologues, for example, *Actinomyces* sp. Oral and *Actinomyces graevenitzi*, are found to be more abundant in the tongue dorsum and other oral cavity locations, which could be an implication of these organisms taking advantage of the sulfoglycans in the human diet or participating the oral homeostasis by changing the 6S-GlcNAc-containing glycans

from sulfomucins that coat the oral surface, such as MUC5B<sup>49,50</sup> and MUC1.

## DISCUSSION

The 6-sulfate adjunct on GlcNAc is contributed by three types of human GlcNAc-6-*O*-sulfotransferases (GlcNAc6STs-CHT2, CHT4, and CHT5) and it serves diverse roles in biological processes.<sup>53,54</sup> The characterization of *exo*-acting 6S-GlcNAcases expressed by prominent human microbes is an important step forward to advance our understanding of the complex relationship between these bacteria and their human host. We present here the first structural enzymology study on three GH20 family 6S-GlcNAcases that hydrolyze 6S-GlcNAc: BbhII from a Gram-positive *B. bifidum* JCM1254, Bt4394 from a

*Prevotella* RS2. The crystal structures of these three enzymes show different structural features that correlate with their different cellular localizations: BbhII can access larger mucin complexes extracellularly, whereas Bt4394 and SGL are more competent on oligosaccharide substrates as periplasmic enzymes. The visible Ig-like domain in BbhII may fulfill a structural anchor by providing the correct spatial organization of the associated GH20 catalytic module.<sup>55,56</sup> This potentially enables the membrane-anchored BbhII to extend beyond the cell wall peptidoglycan to within the optimal distance for engaging target macromolecular substrates such as the mucin glycans.<sup>55</sup> For *B. thetaiotaomicron*, it is known that numerous sulfatases in its periplasm can facilitate desulfation under operonic regulation.<sup>15</sup> Previously identified periplasmic sulfatases include those that can desulfate a monosaccharide 6S-GlcNAc, such as Bt4656 ( $k_{cat}/K_M = 1.1 \times 10^{-3} \text{ s}^{-1} \mu\text{M}^{-1}$  at pH 6.0),<sup>57</sup> or 6S-GlcNAc at the non-reducing end of a glycan, for example, Bt1628 and Bt3177 ( $k_{cat}/K_M = 0.43 \times 10^{-3} \text{ s}^{-1} \mu\text{M}^{-1}$  at pH 8.0).<sup>17,22</sup> However, these rates are 3 orders of magnitude slower than that of the glycosidase Bt4394 identified in our study ( $k_{cat}/K_M = 0.67 \text{ s}^{-1} \mu\text{M}^{-1}$  at pH 5.5). Such a huge activity difference indicates that *B. thetaiotaomicron* may utilize operonic regulation induced by different 6S-GlcNAc-containing oligosaccharides to process the glycan substrates they access. *B. thetaiotaomicron* may also rely more effectively on the 6S-GlcNAcase activity of Bt4394 rather than on other sulfatases to desulfate 6S-GlcNAc-containing mucin oligosaccharides.

The well-established catalytic mechanism of GH20 enzymes involves a substrate-assisted mechanism (Figure 3).<sup>58-62</sup> A polarizing aspartate is always anionic and steers and promotes the nucleophilic attack of the carbonyl oxygen of the 2-acetamido group<sup>61</sup> while a neighboring glutamic acid (catalytic diad) provides general acid catalysis for the leaving group leaving to form an intermediate oxazoline. In a second step, a water is activated for attack on C1 using general base catalysis provided by the same glutamate, resulting in net retention of configuration at C1. For 6S-GlcNAcases, the specific catalysis originates from the tight binding and hexose orientation of the sulfated substrate in the recognition site. The catalytic loop containing the Glu-Asp diad has been observed to adapt flexible conformations (Figures S6 and S10). We are now able, for the first time, to compare directly the intermediate structures formed for these two GH20 6S-GlcNAcases, namely, BbhII in a catalytically incompetent conformation and Bt4394 in an active conformation. The direct W409-E391 interaction seen in the Bt4394<sub>D335N</sub>-6S-NAG-oxazoline intermediate structure delivers the important closure of the essential catalytic loop through the formation of a direct H-bond from the E336 backbone amide to

E391, which is not the case in the BbhII<sub>E553Q</sub>-6S-NAG-oxazoline

Gram-negative bacterium *B. thetaiotaomicron*, and SGL from

intermediate structure. Moreover, the conserved catalytically important histidine (H467 from BbhII and H270 from Bt4394) has been proposed, quantum mechanically, to modulate the  $pK_a$ s of the catalytic Glu-Asp diad by oscillating between them in a GH20 family *endo* Lacto-*N*-biosidase LnbB.<sup>63</sup> The unaltered position of histidine with different catalytic loop conformations observed in our two 6S-oxazoline intermediate structures has provided another possibility for this histidine to achieve the same interaction and modulation of the catalytic diad by dynamic movements of the catalytic loop (Figure 2c,d). This is further exemplified by the 3000-fold difference in the activities of His-to-Phe variants relative to the wild-type enzymes. BbhII has a more flexible catalytic loop, so that BbhII<sub>H467F</sub> maintains 21% of wild-type activity, whereas Bt4394 has a more rigid and

ordered catalytic loop, thus its H270 has a more important role in modulating the catalytic diad during turnover. This analysis is supported by the residual 0.003% activity for Bt4394<sub>H270F</sub> (Table 1 and Figure S21).

Moreover, in seeking to co-crystallize NAG-thiazoline-bound Bt4394 (the IC<sub>50</sub> value for NAG-thiazoline is 166.6 μM, Figure S8) with sodium sulfite, we obtained a crystal diffracting at 1.47 Å with a clear electron density of a pyramidal SO<sub>2</sub> species coordinated to O6' in the sulfate binding site, partially overlapping with that of the NAG-thiazoline 6'-OH group. Our refined model showed an elongated covalent bond with an S–O6' distance of 1.69 Å, which is 0.15 Å longer than the S–O6' bond in Bt4394-6S-GlcNAc and Bt4394<sub>D335N</sub>-6S-NAG-oxazoline structures of comparable resolution. This serendipitously observed sulfinylation reaction may result from condensation of 6'-OH onto a proximate sulfite in the well-ordered sulfate binding site (Figure S22). Such *in situ* sulfinylation may prove to be a useful structural biology tool for the future identification of potential sulfate binding sites in sulfoglycosidases.

We have identified here the distinct sulfate binding motifs from Bt4394, SGL, and BbhII by our use of structural analysis. Based on our experimental results, SSN and CGFP analyses have led us to a novel identification of species with potential 6S-GlcNAcase activities on *N*- and *O*-glycans from Gram-negative and Gram-positive bacteria, including oral cavity microbes. The clear distinction between Gram-negative and Gram-positive bacteria will provide a means of securely annotating 6S-GlcNAcases. Assignment of the function of these enzymes in glycan desulfation also aids in the identification of new glycan-desulfating microbes and can lead to the discovery of new sulfoglycan-degrading pathways.

It is noteworthy that the structural definition of the sulfate binding motif is necessary for a reliable assignment. We surveyed the general features of molecular recognition of a 6'-sulfate group for 19 different CAZy protein structures (Table S3). Of these, only nine use a cationic amino acid to coordinate the sulfate (five *via* arginine and two by lysine). This is because the sulfuryl oxygens in R-OSO<sub>3</sub><sup>-</sup> have relatively small net negative electron density compared to that on phosphoryl monoesters, with the result that enzymes that bind sulfate-containing carbohydrate substrates do not have a high demand for cationic, H-bond donating residues such as arginine or lysine. They rather recruit “soft”, neutral H-bond donors (cysteine, tyrosine, asparagine, glutamine, tryptophan, and water) to coordinate the sulfuryl oxygens (Table S3). This contrasts with the recognition of passive phosphate-containing carbohydrate substrates, for example, glucose 6-phosphate and 1-phosphate, that is regularly achieved by “hard” H-bond donors (arginine and lysine cations) to match the much higher charge density on the phosphoryl oxygens<sup>64,65</sup> Our result in demonstrating the ultimate specificity of 6S-GlcNAcases over 6P-GlcNAcases is a clear and powerful manifestation of this analysis.

The conserved tyrosine in the active site of GH20 enzymes orientates the nucleophilic oxygen for stereoelectronic formation of the oxazoline and, equally, assists C–O bond breaking in the second, hydrolytic step.<sup>59,66</sup> Mutation of this tyrosine to phenylalanine in BbhII (Y637F) and Bt4394 (Y435F) leads to *ca.* 25-fold reduction of activity (Table 1). Furthermore, it is striking that in all the three 6S-GlcNAcases characterized here, this key tyrosine (Y637 in BbhII, Y435 in Bt4394, and Y564 in SGL) is located in the sulfate binding loop. As this tyrosine is indispensable for substrate-assisted catalysis in GH20,<sup>67</sup> the

sulfate binding motif in the 6S-GlcNAcases must have emerged around this tyrosine to satisfy both the need for including residues with long H-bonding donating side chains and/or positive charges to recognize the sulfate, while not causing steric hindrance with the tyrosine. Clearly, our analysis calls for a more detailed investigation by further bioinformatics and phylogenetic studies.

Independent of mucin, sulfated GlcNAc species are present in human milk oligosaccharides (HMO) and high molecular mass GAGs, such as heparin/heparan sulfate, keratan sulfate, and chondroitin/dermatan sulfate, all of which are important in animal tissues.<sup>12,68–70</sup> Although we have tested Bt4394 and BbhII activity only with commercially available mucin *O*-glycans, we can expect that these enzymes play a role in the catabolism of other sulfo-sugars to facilitate access of human microbiota to a wide variety of glycan sources. This is based on the observation that these enzymes hydrolyze 6S-GlcNAc from both the synthetic and natural substrates and their structures show no recognition in their +1 binding site, thus offering an advantage for these enzymes to be used as engineered biosynthesis catalysts for glycosyl transfer, as demonstrated by others.<sup>71,72</sup> In light of the important roles of sulfo-sugars in human health, we believe that the findings reported here not only deliver structural and mechanistic insights into three representatives of 6S-GlcNAcase but also provide valuable information to explore the extensive interactions between microbiota and the host glycans and their involvements in the development of human disease.

## EXPERIMENTAL SECTION

Full details of all experimental procedures used are described in Supporting Information Materials and Methods. Coordinates have been deposited in Protein Data Bank with accession codes 6YXZ and 6Z14 for the BbhII native enzyme and 6S-GlcNAc-oxazoline complex, respectively; 7DUP, 7DVB, and 7DVA for the Bt4394 native enzyme, 6S-GlcNAc-oxazoline complex and 6S-GlcNAc product complex, respectively; and 8BAL and 8BBL for the Niako3494 and SGL native apo-proteins, respectively.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.2c05405>.

Experimental procedures; oligonucleotides, data collection and refinement statistics, and complex structures; identification of cellular localizations, detection of released 6S-GlcNAc, standard curves of 4MU, Michaelis–Menten plots, structures of the 6S-NAG-oxazoline intermediate complexes, inhibition of recombinant Bt4394, structures of the oxazoline intermediate complexes, structure comparison of BbhII with Bt4394, MSA, active site of Niako3494, sequences with 6S-GlcNAcase activity, GH20 domain sequences, comparison of the flexibility, observed NAG thiazoline-SO<sub>2</sub><sup>-</sup> species, MS spectrum, <sup>1</sup>H NMR spectrum, <sup>13</sup>C NMR spectrum, and <sup>31</sup>P NMR spectrum; and MS and NMR spectra (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Yi Jin – School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.; Present Address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess

Street, Manchester, M1 7DN, United Kingdom;  
 ● [orcid.org/0000-0002-6927-4371](https://orcid.org/0000-0002-6927-4371); Email: [yi.jin@manchester.ac.uk](mailto:yi.jin@manchester.ac.uk)

YuanHe – Key Laboratory of Synthetic and Natural Functional Molecule, College of Chemistry and Materials Science, Northwest University, Xi'an 710127, P. R. China;  
 ● [orcid.org/0000-0002-1712-0776](https://orcid.org/0000-0002-1712-0776); Email: [yuanhe@nwu.edu.cn](mailto:yuanhe@nwu.edu.cn)

## Authors

Zhen Zhang – Key Laboratory of Synthetic and Natural Functional Molecule, College of Chemistry and Materials Science, Northwest University, Xi'an 710127, P. R. China; School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.

Mochen Dong – School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.

Rémi Zallot – Institute of Life Sciences, Swansea University Medical School, Swansea SA2 8PP, U.K.; Present Address: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom; ● [orcid.org/0000-0002-7317-1578](https://orcid.org/0000-0002-7317-1578)

George Michael Blackburn – School of Biosciences, University of Sheffield, Sheffield S10 2TN, U.K.

Nini Wang – Key Laboratory of Synthetic and Natural Functional Molecule, College of Chemistry and Materials Science, Northwest University, Xi'an 710127, P. R. China

Chengjian Wang – Glycobiology and Glycotechnology Research Center, College of Food Science and Technology, Northwest University, Xi'an 710069, P. R. China

Long Chen – School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.

Patrick Baumann – School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.

Zuyan Wu – School of Chemistry, Cardiff University, Cardiff CF10 3AT, U.K.

Zhongfu Wang – Glycobiology and Glycotechnology Research Center, College of Food Science and Technology, Northwest University, Xi'an 710069, P. R. China; ● [orcid.org/0000-0003-1616-5056](https://orcid.org/0000-0003-1616-5056)

Haiming Fan – Key Laboratory of Synthetic and Natural Functional Molecule, College of Chemistry and Materials Science, Northwest University, Xi'an 710127, P. R. China;  
 ● [orcid.org/0000-0002-0091-772X](https://orcid.org/0000-0002-0091-772X)

Christian Roth – Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14195 Berlin, German

## Author Contributions

<sup>†</sup>Z.Z., M.D., and R.Z. contributed equally to this work.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. Pierre Rizkallah from Cardiff University Medical School for assistance with the X-ray data collection, the Diamond Light Source for access to beamline I04 under proposal number mx18812, and Shanghai Synchrotron Radiation Facility for access to beamline BL17B1 at the National Facility for Protein Science in Shanghai (NFPS). We

thank Northwest University for a visiting studentship for Z.Z. We thank China Scholarship Council (202006280011) for a PhD studentship for M.D. We also thank Wellcome Trust Seed Award (209057/Z/17/Z) and The Academy of Medical Sciences Springboard Award (SBE003\1154) for supporting Y.J., who is a Wellcome Trust Sir Henry Dale Fellow (218568/Z/19/Z). R.Z. was supported by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Individual Fellowships Grant agreement H2020-MSCA-IF-2018 839116 deCrYPtion. G.M.B. thanks the Leverhulme Trust for an Emeritus Fellowship (EM-2022-057). C.R. thanks the Max Planck Society for funding. Y.H. was supported by the National Natural Science Foundation of China (31400663) and Shaanxi Provincial Department of Education Funds (grant no. 22JP081).

## REFERENCES

- (1) Muthana, S. M.; Campbell, C. T.; Gildersleeve, J. C. Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chem. Biol.* **2012**, *7*, 31–43.
- (2) Holmén Larsson, J. M.; Thomsson, K. A.; Rodríguez-Piñero, A. M.; Karlsson, H.; Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucinO-glycan patterns reveal a regiospecific distribution. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2013**, *305*, G357–G363.
- (3) Lo-Guidice, J. M.; Wieruszkeski, J. M.; Lemoine, J.; Verbert, A.; Roussel, P.; Lamblin, G. Sialylation and sulfation of the carbohydrate chains in respiratory mucins from a patient with cystic fibrosis. *J. Biol. Chem.* **1994**, *269*, 18794–18813.
- (4) She, Y.-M.; Li, X.; Cyr, T. D. Remarkable structural diversity of N-glycan sulfation on influenza vaccines. *Anal. Chem.* **2019**, *91*, 5083–5090.
- (5) Byrd-Leotis, L.; Lasanajak, Y.; Bowen, T.; Baker, K.; Song, X.; Suthar, M. S.; Cummings, R. D.; Steinhauer, D. A. SARS-CoV-2 and other coronaviruses bind to phosphorylated glycans from the human lung. *Virology* **2021**, *562*, 142–148.
- (6) Chuzel, L.; Fossa, S. L.; Boisvert, M. L.; Cajic, S.; Hennig, R.; Ganatra, M. B.; Reichl, U.; Rapp, E.; Taron, C. H. Combining functional metagenomics and glycoanalytics to identify enzymes that facilitate structural characterization of sulfated N-glycans. *Microbial. Cell Factories* **2021**, *20*, 162.
- (7) Macauley, M. S.; Crocker, P. R.; Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nat. Rev. Immunol.* **2014**, *14*, 653–666.
- (8) Bowman, K. G.; Hemmerich, S.; Bhakta, S.; Singer, M. S.; Bistrup, A.; Rosen, S. D.; Bertozzi, C. R. Identification of an N-acetylglucosamine-6-O-sulfotransferase activity specific to lymphoid tissue: an enzyme with a possible role in lymphocyte homing. *Chem. Biol.* **1998**, *5*, 447–460.
- (9) Xia, B.; Royall, J. A.; Damera, G.; Sachdev, G. P.; Cummings, R. D. Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiology* **2005**, *15*, 747–775.
- (10) Kenny, D.; Hayes, C. A.; Jin, C.; Karlsson, N. G. Perspective and review of mass spectrometric based sulfoglycomics of N-linked and O-linked oligosaccharides. *Curr. Proteomics* **2011**, *8*, 278–296.
- (11) Carlson, T. L.; Lock, J. Y.; Carrier, R. L. Engineering the mucus barrier. *Annu. Rev. Biomed. Eng.* **2018**, *20*, 197–220.
- (12) Boltin, D.; Perets, T. T.; Vilkin, A.; Niv, Y. Mucin Function in Inflammatory Bowel Disease. *J. Clin. Gastroenterol.* **2013**, *47*, 106–111.
- (13) Tsai, H. H.; Sunderland, D.; Gibson, G. R.; Hart, C. A.; Rhodes, J. M. A novel mucin sulphatase from human faeces: its identification, purification and characterization. *Clin. Sci.* **1992**, *82*, 447–454.
- (14) Robinson, C. V.; Elkins, M. R.; Bialkowski, K. M.; Thornton, D. J.; Kertesz, M. A. Desulfurization of mucin by *Pseudomonas aeruginosa*: influence of sulfate in the lungs of cystic fibrosis patients. *J. Med. Microbiol.* **2012**, *61*, 1644–1653.

- (15) Cartmell, A.; Lowe, E. C.; Baslé, A.; Firbank, S. J.; Ndeh, D. A.; Murray, H.; Terrapon, N.; Lombard, V.; Henrissat, B.; Turnbull, J. E.; Czjzek, M.; Gilbert, H. J.; Bolam, D. N. How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 7037–7042.
- (16) Wright, D. P.; Knight, C. G.; Parkar, S. G.; Christie, D. L.; Robertson, A. M. Cloning of a mucin-desulfating sulfatase gene from *Prevotella* Strain RS2 and Its expression using a *Bacteroides* recombinant system. *J. Bacteriol.* **2000**, *182*, 3002–3007.
- (17) Luis, A. S.; Jin, C.; Pereira, G. V.; Glowacki, R. W. P.; Gugel, S. R.; Singh, S.; Byrne, D. P.; Pudlo, N. A.; London, J. A.; Baslé, A.; Reihill, M.; Oscarson, S.; Eysers, P. A.; Czjzek, M.; Michel, G.; Barbeyron, T.; Yates, E. A.; Hansson, G. C.; Karlsson, N. G.; Cartmell, A.; Martens, E. C. A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **2021**, *598*, 332–337.
- (18) Milani, C.; Lugli, G. A.; Duranti, S.; Turrioni, F.; Mancabelli, L.; Ferrario, C.; Mangifesta, M.; Hevia, A.; Viappiani, A.; Scholz, M.; Arioli, S.; Sanchez, B.; Lane, J.; Ward, D. V.; Hickey, R.; Mora, D.; Segata, N.; Margolles, A.; van Sinderen, D.; Ventura, M. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci. Rep.* **2015**, *5*, 15782.
- (19) Katoh, T.; Maeshibu, T.; Kikkawa, K.-i.; Gotoh, A.; Tomabechi, Y.; Nakamura, M.; Liao, W.-H.; Yamaguchi, M.; Ashida, H.; Yamamoto, K.; Katayama, T. Identification and characterization of a sulfoglycosidase from *Bifidobacterium bifidum* implicated in mucin glycan utilization. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 2018–2027.
- (20) Turrioni, F.; Bottacini, F.; Foroni, E.; Mulder, I.; Kim, J.-H.; Zomer, A.; Sánchez, B.; Bidossi, A.; Ferrarini, A.; Giubellini, V.; Delledonne, M.; Henrissat, B.; Coutinho, P.; Oggioni, M.; Fitzgerald, G. F.; Mills, D.; Margolles, A.; Kelly, D.; van Sinderen, D.; Ventura, M. Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19514–19519.
- (21) Ruas-Madiedo, P.; Gueimonde, M.; Fernández-García, M.; de los Reyes-Gavilán, C. G.; Margolles, A. Mucin degradation by *Bifidobacterium* strains isolated from the human intestinal microbiota. *Appl. Environ. Microbiol.* **2008**, *74*, 1936–1940.
- (22) Ulmer, J. E.; Vilén, E. M.; Namburi, R. B.; Benjdia, A.; Beneteau, J.; Malleron, A.; Bonnaffé, D.; Driguez, P. A.; Descroix, K.; Lassalle, G.; Le Narvor, C.; Sandström, C.; Spillmann, D.; Berteau, O. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont *Bacteroides thetaiotaomicron* reveals the first GAG-specific bacterial endosulfatase. *J. Biol. Chem.* **2014**, *289*, 24289–24303.
- (23) Pudlo, N. A.; Urs, K.; Kumar, S. S.; German, J. B.; Mills, D. A.; Martens, E. C. Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *mBio* **2015**, *6*, No. e01282.
- (24) Martens, E. C.; Chiang, H. C.; Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **2008**, *4*, 447–457.
- (25) Rho, J.-h.; Wright, D. P.; Christie, D. L.; Clinch, K.; Furneaux, R. H.; Robertson, A. M. A novel mechanism for desulfation of mucin: identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *J. Bacteriol.* **2005**, *187*, 1543–1551.
- (26) Helbert, W.; Poulet, L.; Drouillard, S.; Mathieu, S.; Loidice, M.; Couturier, M.; Lombard, V.; Terrapon, N.; Turchetto, J.; Vincentelli, R.; Henrissat, B. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 6063–6068.
- (27) Chen, J.-Y.; Huang, H.-H.; Yu, S.-Y.; Wu, S.-J.; Kannagi, R.; Khoo, K.-H. Concerted mass spectrometry-based glycomic approach for precision mapping of sulfo sialylated N-glycans on human peripheral blood mononuclear cells and lymphocytes. *Glycobiology* **2017**, *28*, 9–20.
- (28) Miwa, M.; Horimoto, T.; Kiyohara, M.; Katayama, T.; Kitaoka, M.; Ashida, H.; Yamamoto, K. Cooperation of  $\beta$ -galactosidase and  $\beta$ -N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. *Glycobiology* **2010**, *20*, 1402–1409.
- (29) Evans, D. F.; Pye, G.; Bramley, R.; Clark, A. G.; Dyson, T. J.; Hardcastle, J. D. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* **1988**, *29*, 1035–1041.
- (30) Ray, W. J.; Long, J. W.; Owens, J. D. An analysis of the substrate-induced rate effect in the phosphoglucomutase system. *Biochemistry* **1976**, *15*, 4006–4017.
- (31) Greig, I. R.; Macauley, M. S.; Williams, I. H.; Vocadlo, D. J. Probing synergy between two catalytic strategies in the glycoside hydrolase O-GlcNAcase using multiple linear free energy relationships. *J. Am. Chem. Soc.* **2009**, *131*, 13415–13422.
- (32) Almagro Armenteros, J. J.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423.
- (33) Tsang, W. Y.; Amyes, T. L.; Richard, J. P. A Substrate in Pieces: Allosteric Activation of Glycerol 3-Phosphate Dehydrogenase (NAD<sup>+</sup>) by Phosphite Dianion. *Biochemistry* **2008**, *47*, 4575–4582.
- (34) Thi, N. N.; Offen, W. A.; Shareck, F.; Davies, G. J.; Doucet, N. Structure and Activity of the *Streptomyces coelicolor* A3(2)  $\beta$ -N-Acetylhexosaminidase Provides Further Insight into GH20 Family Catalysis and Inhibition. *Biochemistry* **2014**, *53*, 1789–1800.
- (35) Greenhalgh, R.; Heggie, R. M.; Weinberger, M. A. THE DECOMPOSITION OF 2-METHYL- $\Delta^2$ -OXAZOLINE IN AQUEOUS SOLUTION. *Can. J. Chem.* **1963**, *41*, 1662–1670.
- (36) Kresse, H.; Fuchs, W.; Glössl, J.; Holtfrech, D.; Gilberg, W. Liberation of N-acetylglucosamine-6-sulfate by human  $\beta$ -N-acetylhexosaminidase A. *J. Biol. Chem.* **1981**, *256*, 12926–12932.
- (37) Speciale, G.; Jin, Y.; Davies, G. J.; Williams, S. J.; Goddard-Borger, E. D. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **2016**, *12*, 215–217.
- (38) Abayakoon, P.; Jin, Y.; Lingford, J. P.; Petricevic, M.; John, A.; Ryan, E.; Wai-Ying Mui, J.; Pires, D. E. V.; Ascher, D. B.; Davies, G. J.; Goddard-Borger, E. D.; Williams, S. J. Structural and biochemical insights into the function and evolution of sulfoquinovosidases. *ACS. Cent. Sci.* **2018**, *4*, 1266–1273.
- (39) Graham, M. E.; Thaysen-Andersen, M.; Bache, N.; Craft, G. E.; Larsen, M. R.; Packer, N. H.; Robinson, P. J. A novel post-translational modification in nerve terminals: O-linked N-acetylglucosamine phosphorylation. *J. Proteome Res.* **2011**, *10*, 2725–2733.
- (40) Graham, M. E.; Stone, R. S.; Robinson, P. J.; Payne, R. J. Synthesis and protein binding studies of a peptide fragment of clathrin assembly protein AP180 bearing an O-linked  $\beta$ -N-acetylglucosaminyl-6-phosphate modification. *Org. Biomol. Chem.* **2012**, *10*, 2545–2551.
- (41) Parkkinen, J.; Finne, J. Occurrence of N-acetylglucosamine 6-phosphate in complex carbohydrates. Characterization of a phosphorylated sialyl oligosaccharide from bovine colostrum. *J. Biol. Chem.* **1985**, *260*, 10971–10975.
- (42) Vicaretti, S. D.; Mohtarudin, N. A.; Garner, A. M.; Zandberg, W. F. Capillary electrophoresis analysis of bovine milk oligosaccharides permits an assessment of the influence of diet and the discovery of nine abundant sulfated analogues. *J. Agric. Food Chem.* **2018**, *66*, 8574–8583.
- (43) Alberty, R. A. Thermodynamic properties of weak acids involved in enzyme-catalyzed reactions. *J. Phys. Chem. B* **2006**, *110*, 5012–5016.
- (44) Val-Cid, C.; Biarnés, X.; Faijes, M.; Planas, A. Structural-functional analysis reveals a specific domain organization in family GH20 hexosaminidases. *PLoS One* **2015**, *10*, No. e0128075.
- (45) Zallot, R.; Oberg, N. O.; Gerlt, J. A. “Democratized” genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77–85.
- (46) Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G. A.; Sonnhammer, E. L. L.; Tosatto, S. C. E.; Paladin, L.; Raj, S.; Richardson, L. J.; Finn, R. D.; Bateman, A. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2020**, *49*, D412–D419.
- (47) Henrissat, B.; Davies, G. J. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol.* **2000**, *124*, 1515–1519.

- (48) Potter, S. C.; Luciani, A.; Eddy, S. R.; Park, Y.; Lopez, R.; Finn, R. D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204.
- (49) Nielsen, P. A.; Bennett, E. P.; Wandall, H. H.; Therkildsen, M. H.; Hannibal, J.; Clausen, H. Identification of a major human high molecular weight salivary mucin (MG1) as tracheobronchial mucin MUC5B. *Glycobiology* **1997**, *7*, 413–419.
- (50) Van Klinken, B. J.-W.; Dekker, J.; Gool, S. A. V.; Marle, J. V.; Büller, H. A.; Einerhand, A. W. C. MUC5B is the prominent mucin in human gallbladder and is also expressed in a subset of colonic goblet cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **1998**, *274*, G871–G878.
- (51) Seko, A.; Ohkura, T.; Ideo, H.; Yamashita, K. Novel O-linked glycans containing 6'-sulfo-Gal/GalNAc of MUC1 secreted from human breast cancer YMB-S cells: Possible carbohydrate epitopes of KL-6(MUC1) monoclonal antibody. *Glycobiology* **2011**, *22*, 181–195.
- (52) Sun, L.; Konstantinidi, A.; Ye, Z.; Nason, R.; Zhang, Y.; Büll, C.; Kahl-Knutson, B.; Hansen, L.; Leffler, H.; Vakhrushev, S. Y.; Yang, Z.; Clausen, H.; Narimatsu, Y. Installation of O-glycan sulfation capacities in human HEK293 cells for display of sulfated mucins. *J. Biol. Chem.* **2022**, *298*, 101382.
- (53) Tobisawa, Y.; Imai, Y.; Fukuda, M.; Kawashima, H. Sulfation of colonic mucins by N-acetylglucosamine 6-O-Sulfotransferase-2 and its protective function in experimental colitis in mice. *J. Biol. Chem.* **2010**, *285*, 6750–6760.
- (54) Yu, S. Y.; Hsiao, C. T.; Izawa, M.; Yusa, A.; Ishida, H.; Nakamura, S.; Yagi, H.; Kannagi, R.; Khoo, K. H. Distinct substrate specificities of human GlcNAc-6-sulfotransferases revealed by mass spectrometry-based sulfoglycomic analysis. *J. Biol. Chem.* **2018**, *293*, 15163–15177.
- (55) Venditto, I.; Najmudin, S.; Luís, A. S.; Ferreira, L. M.; Sakka, K.; Knox, J. P.; Gilbert, H. J.; Fontes, C. M. Family 46 Carbohydrate-binding Modules Contribute to the Enzymatic Hydrolysis of Xyloglucan and  $\beta$ -1,3-1,4-Glucans through Distinct Mechanisms. *J. Biol. Chem.* **2015**, *290*, 10572–10586.
- (56) Itoh, T.; Hibi, T.; Suzuki, F.; Sugimoto, I.; Fujiwara, A.; Inaka, K.; Tanaka, H.; Ohta, K.; Fujii, Y.; Taketo, A.; Kimoto, H. Crystal structure of chitinase ChiW from *Paenibacillus* sp. str. FPU-7 reveals a novel type of bacterial cell-surface-expressed multi-modular enzyme machinery. *PLoS One* **2016**, *11*, No. e0167310.
- (57) Byrne, D. P.; London, J. A.; Eyers, P. A.; Yates, E. A.; Cartmell, A. Mobility shift-based electrophoresis coupled with fluorescent detection enables real-time enzyme analysis of carbohydrate sulfatase activity. *Biochem. J.* **2021**, *478*, 735–748.
- (58) Hou, Y.; Vocadlo, D. J.; Leung, A.; Withers, S. G.; Mahuran, D. Characterization of the Glu and Asp Residues in the Active Site of Human  $\beta$ -Hexosaminidase B. *Biochemistry* **2001**, *40*, 2201–2209.
- (59) Mark, B. L.; Vocadlo, D. J.; Knapp, S.; Triggs-Raine, B. L.; Withers, S. G.; James, M. N. Crystallographic Evidence for Substrate-assisted Catalysis in a Bacterial  $\beta$ -Hexosaminidase. *J. Biol. Chem.* **2001**, *276*, 10330–10337.
- (60) Mark, B. L.; Vocadlo, D. J.; Zhao, D.; Knapp, S.; Withers, S. G.; James, M. N. Biochemical and Structural Assessment of the 1-N-Azasugar GalNAc-isofagomine as a Potent Family 20  $\beta$ -N-Acetylhexosaminidase Inhibitor. *J. Biol. Chem.* **2001**, *276*, 42131–42137.
- (61) Williams, S. J.; Mark, B. L.; Vocadlo, D. J.; James, M. N. G.; Withers, S. G. Aspartate 313 in the *Streptomyces plicatus* Hexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. *J. Biol. Chem.* **2002**, *277*, 40055–40065.
- (62) Alteen, M. G.; Oehler, V.; Nemcovicová, I.; Wilson, I. B.; Vocadlo, D. J.; Gloster, T. M. Mechanism of human nucleocytoplasmic hexosaminidase D. *Biochemistry* **2016**, *55*, 2735–2747.
- (63) Cuxart, I.; Coines, J.; Esquivias, O.; Fajjes, M.; Planas, A.; Biarnés, X.; Rovira, C. Enzymatic hydrolysis of human milk oligosaccharides. The molecular mechanism of *Bifidobacterium bifidum* Lacto-N-biosidase. *ACS Catal.* **2022**, *12*, 4737–4743.
- (64) Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; MacKerell, A. D., Jr. CHARMM Additive all-atom force field for phosphate and sulfate linked to carbohydrates. *Chem. Theor. Comput.* **2012**, *8*, 759–776.
- (65) LoPachin, R. M.; Gavin, T.; DeCaprio, A.; Barber, D. S. Application of the Hard and Soft, Acids and Bases (HSAB) Theory to Toxicant-Target Interactions. *Chem. Res. Toxicol.* **2012**, *25*, 239–251.
- (66) Tews, I.; Terwisscha van Scheltinga, A. C.; Perrakis, A.; Wilson, K. S.; Dijkstra, B. W. Substrate-assisted catalysis unifies two families of chitinolytic enzymes. *J. Am. Chem. Soc.* **1997**, *119*, 7954–7959.
- (67) Intra, J.; Pavesi, G.; Horner, D. S. Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family. *BMC Evol. Biol.* **2008**, *8*, 214.
- (68) Bradshaw, D. J.; Homer, K. A.; Marsh, P. D.; Beighton, D. Metabolic cooperation in oral microbial communities during growth on mucin. *Microbiology* **1994**, *140*, 3407–3412.
- (69) Ridley, C.; Thornton, D. J. Mucins: the frontline defence of the lung. *Biochem. Soc. Trans.* **2018**, *46*, 1099–1106.
- (70) Amerongen, A. V. N.; Bolscher, J. G. M.; Bloemena, E.; Veerman, E. C. Sulfomucins in the human body. *Biol. Chem.* **1998**, *379*, 1–18.
- (71) Mészáros, Z.; Nekvasilová, P.; Bojarová, P.; Kren, V.; Slámová, K. Advanced glycosidases as ingenious biosynthetic instruments. *Bio-technol. Adv.* **2021**, *49*, 107733.
- (72) Tegl, G.; Hanson, J.; Chen, H. M.; Kwan, D. H.; Santana, A. G.; Withers, S. G. Facile Formation of  $\beta$ -thioGlcNAc Linkages to Thiol-Containing Sugars, Peptides, and Proteins using a Mutant GH20 Hexosaminidase. *Angew. Chem., Int. Ed. Engl.* **2019**, *58*, 1632–1637.