JBC RESEARCH ARTICLE

Characterization of a new family of 6-sulfo-N-acetylglucosaminidases

Received for publication, June 15, 2023, and in revised form, August 12, 2023 Published, Papers in Press, September 1, 2023,
https://doi.org/10.1016/j.jbc.2023.105214

Rajneesh K. Bains[1,2], Seyed A. Nasseri[1,2], Feng Liu[1,2], Jacob F. Wardman[1,2,3], Peter Rahfeld[1,2], and Stephen G. Withers[1,2,3,*]

From the [1]Department of Chemistry, [2]Michael Smith Laboratories, and [3]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, Canada

Reviewed by members of the JBC Editorial Board. Edited by Robert Haltiwanger

Sulfation is widespread in nature and plays an important role in modulating biological function. Among the strategies developed by microbes to access sulfated oligosaccharides as a nutrient source is the production of 6-sulfoGlcNAcases to selectively release 6-sulfoGlcNAc from target oligosaccharides. Thus far, all 6-sulfoGlcNAcases identified have belonged to the large GH20 family of β-hexosaminidases. Here, we identify and characterize a new, highly specific non-GH20 6-sulfoGlcNAcase from *Streptococcus pneumoniae* TIGR4, Sp_0475 with a greater than 110,000-fold preference toward N-acetyl-β-D-glucosamine-6-sulfate substrates over the non-sulfated version. Sp_0475 shares distant sequence homology with enzymes of GH20 and with the newly formed GH163 family. However, the sequence similarity between them is sufficiently low that Sp_0475 has been assigned as the founding member of a new glycoside hydrolase family, GH185. By combining results from site-directed mutagenesis with mechanistic studies and bioinformatics we provide insight into the substrate specificity, mechanism, and key active site residues of Sp_0475. Enzymes of the GH185 family follow a substrate-assisted mechanism, consistent with their distant homology to the GH20 family, but the catalytic residues involved are quite different. Taken together, our results highlight in more detail how microbes can degrade sulfated oligosaccharides for nutrients.

Most secreted proteins in eukaryotes are glycosylated with either *N*-glycans, or *O*-glycans or a mixture of both. *N*-linked glycoproteins can be classified into three groups: high mannose, complex, or hybrid type (Fig. 1). All *N*-glycans share a core structure comprising three mannose (Man) residues attached to a chitobiose moiety. The core structure is linked to the protein through a β-configured *N*-glycosyl linkage to an asparagine residue present in a defined sequon at the protein surface (1). *O*-Glycoproteins, of which the mucin glycoproteins are the extracellular dominant species, all feature a common core N-acetylgalactosamine (GalNAc) α-linked to a serine or threonine, though no consensus peptide sequences are yet apparent. This foundational GalNAc can then be elaborated in a variety of ways to yield eight core structures of which four are common (Fig. 1). Each of these can be further elaborated to give linear or branched structures that are similar to those on complex *N*-glycans (2). The non-templated nature of glycan synthesis along with subsequent postglycosylational modifications such as phosphorylation, sulfation, and acetylation generate substantial heterogeneity, making the elucidation of glycan structures difficult (3).

Deciphering the structures of these glycoforms has been greatly facilitated by mass spectrometric methodologies. In most cases, these rely upon the selective release of these glycans from their proteins. Although release of glycans can be achieved through chemical means, enzymes remain the reagents of choice (4). For example, *N*-glycans are released using an amidase, PNGaseF, which cleaves nearly all *N*-linked glycans from eukaryotic glycoproteins but cannot be used for analyzing bacterial *N*-glycans (1). Endo-β-N-acetylglucosaminidases (ENGases) from the GH18 and GH85 families have proven to be highly useful in characterizing *N*-glycan glycoforms, even though these ENGases typically have a narrower substrate range than PNGaseF. This specificity can be leveraged when selective removal of a specific glycoform is required. For instance, Tang *et al.*(5), demonstrated that by using an ENGase with a particular specificity, they could achieve *N*-glycan subtype selective editing on the surface of living cells. An equivalent enzyme for the release of eukaryotic *O*-glycans is lacking, and most enzymatic approaches, beyond the release of the Core 1 (and sialyl core 1 to a certain extent) by GH101 endo-β-N-acetylgalactosaminidases (6), rely on *O*-glycoproteases, which have distinct peptide sequence and glycan specificities (7–9).

Postglycosylational modifications, including sulfation, introduce additional complexity when identifying the precise sites of sulfation on glycans. Exo-glycosidase sequencing, which relies on the use of specific exo-glycosidases in conjunction with HPLC, capillary electrophoresis, and/or mass spectrometry (MS), can help elucidate the structures of specific glycoforms (10, 11). However, GHs with high specificity toward sulfated sugars are needed to help clarify the sites of sulfation on both *O*- and *N*-glycans. Recent work by Chuzel *et al.*(12), exploited functional metagenomics to discover a sugar-specific sulfatase capable of removing sulfate from 6-

* For correspondence: Stephen G. Withers, withers@chem.ubc.ca.

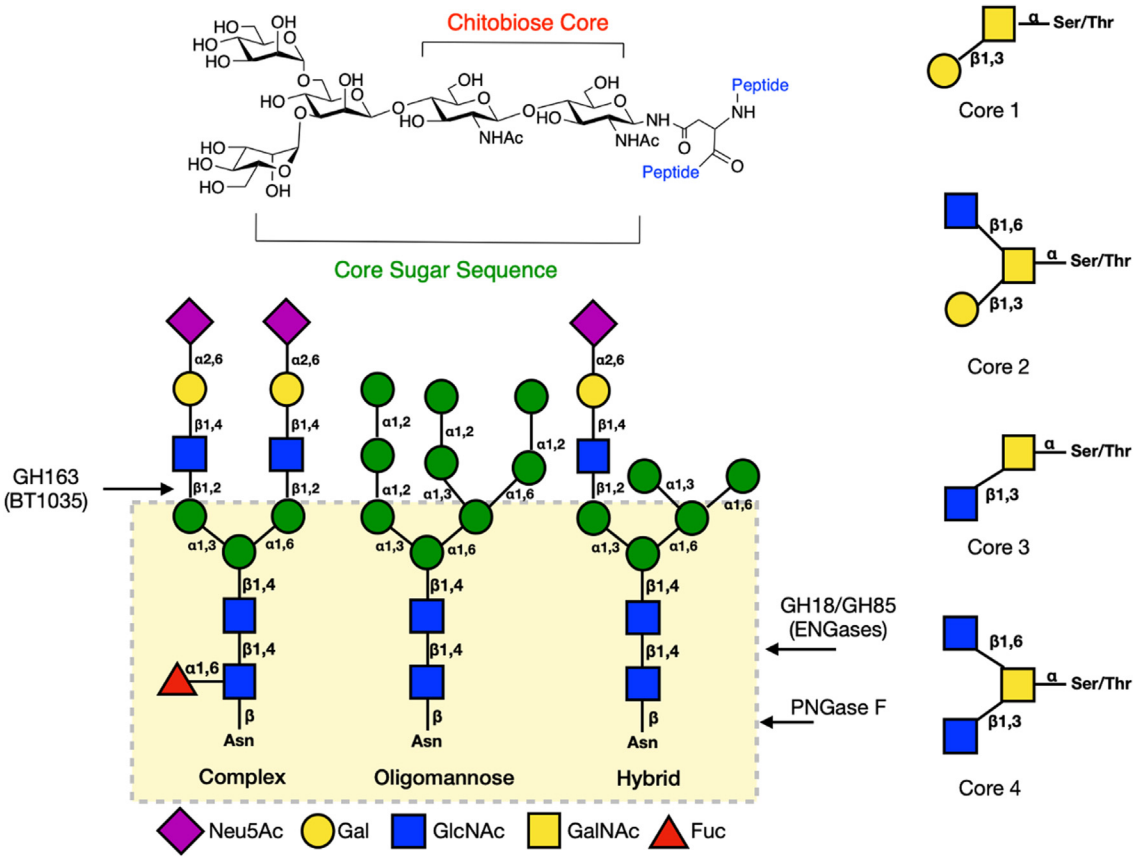

Figure 1. Thecommon structures observed for *N*- and *O*-glycans in eukaryotes. Sugars are depicted using the standard symbol nomenclature for glycans (SNFG). *Yellow box* highlights the core sugar sequence observed in all *N*-glycans.

sulfo-N-acetylglucosamine (6S-GlcNAc), as well as a 6-sulfoGlcNAcase that selectively cleaves 6S-GlcNAc from *N*-glycans. The discovery of additional enzymes with high specificity are needed to expand the tools available to glycobiologists, and the human gut microbiome is a promising source of enzymes of this type (13). Bacteria in the human gut not only degrade dietary glycoproteins as an energy source but also scavenge glycoproteins derived from their eukaryotic host. Access to novel enzyme activities from this large pool has been largely achieved through metagenomic approaches involving activity-based screening (6, 12, 14, 15) or elegant studies involving transcriptomics coupled with detailed biochemical characterization of enzymes involved in the degradation of target glycans (13, 16–20).

Recently it has become apparent that genes for the degradation of *N*-glycans and *O*-glycans are often organized in polysaccharide utilization loci in some microbes (21). Typically, the glycan, or a large fragment thereof, will be cleaved from the glycoprotein and imported into the bacterium *via* a TonB-dependent transporter, where it is further degraded *via* exo-GHs in the periplasm. Terminal anionic sugars such as sialic acids or anionic appendages such as sulfate moieties are often cleaved prior to import (22–25). Recent examples include the unraveling of the enzymology of high-mannose *N*-glycan and complex *N*-glycan degradation (23, 26, 27), as well as various enzymes involved in *O*-glycan degradation (2, 16, 22, 28). As part of our interest in the discovery of enzymes for glycan removal and modification (6, 13–15, 29, 30) we were intrigued by the identification of an endo-hexosaminidase from *Bacteroides thetaiotaomicron* (BT1035) that cleaves the GlcNAc-β1,2-Man linkage within complex type N-glycans (Fig. 1) to release LacNAc (Gal-β1,4-GlcNAc) and SialylLacNAc (Neu5Ac-α2,3-Gal-β1,4-GlcNAc or Neu5Ac-α2,6-Gal-β1,4-GlcNAc) fragments. This founding member of the GH163 family contains two distinct domains, the catalytic N-terminal DUF4838 (PF16126) domain with distant homology to the GH20 family, and a C-terminal F5/F8 domain proposed to be a carbohydrate binding module. An enzyme with such specificities could be highly valuable for *N*-glycan remodeling using approaches similar to those used with GH18 and GH85 family members. In these experiments, desired glycoforms have been introduced onto therapeutic proteins through the use of "glycosynthase" mutants to produce biotherapeutics with improved efficacy (31, 32).

In this article, we describe the characterization of Sp_0475 from *Streptococcus pneumoniae* TIGR4 which contains a DUF4838 domain (PF16126). Although Sp_0475 and BT1035 both contain the DUF4838 domain, the sequence similarity between them is sufficiently low that Sp_0475 has been assigned as the founding member of a new glycoside hydrolase family GH185. Through extensive screening of various substrates, we show that Sp_0475, from *S. pneumoniae* TIGR4, is an exo-acting 6-sulfo-β-GlcNAcase. Its domain of unknown function (DUF4838) shows distant homology with the GH20

family and, consistent with this, NMR analysis confirmed bond cleavage with net retention of anomeric stereochemistry. Inhibition studies with 6S-GlcNAc thiazoline suggest a substrate-assisted mechanism. Based upon kinetic analysis of active site mutants identified from an AlphaFold structural model, we identified critical active site residues and assigned their putative roles in catalysis. Lastly, by isolating the transglycosylation (TG) product produced by Sp_0475 we identify the probable natural substrate as the 6S-GlcNAc-β-1,6-GalNAc moiety. Taken together, this work suggests that Sp_0475 may be useful for the analysis and synthesis of sulfated oligosaccharides.

Results and discussion

Identification of subfamilies of DUF4838 and expression of a promising gene

To gain a better understanding of the diversity present within this domain we generated a sequence similarity network (SSN) of all genes containing the DUF4838 (PF16126) domain (33). The SSN was generated using 3826 sequences with an E-value cutoff of 10^{-74}, which produced clusters with pairwise sequence identity of over 30% for each cluster. Figure 2 highlights the significant sequence diversity that exists within the DUF4838 domain, suggesting that this domain could be involved in various biological activities. Cluster 3 drew our attention as essentially all 252 sequences lacked a signal peptide suggesting that these enzymes are localized within the cytoplasm and act on smaller oligosaccharides. This would be a sharp contrast to the surface-localized BT1035, which cleaves larger substrates such as *N*-glycans that have been released from glycoproteins and glycopeptides by a GH18 endo-N-acetylglucosaminidase (23). Furthermore, the members of cluster 3 share only low sequence homology to the GH163 family, suggesting that they may belong to a new glycoside hydrolase family with potentially distinct substrate specificities. To explore this, we cloned a *S. pneumoniae* TIGR4 gene (Sp_0475) from within this cluster and expressed and purified it as described in Methods.

Exploration of substrate specificity

Our initial screen of over 30 chromogenic and fluorogenic glycosides as potential substrates of Sp_0475 (Table S2) identified only two that were hydrolyzed—4-methylumbelliferyl N-acetyl-β-D-glucosaminide (MUGlcNAc) and 4-nitrophenyl N-acetyl-β-D-glucosaminide (pNPGlcNAc) and these at rather sluggish rates. Since GH163 enzymes cleave *N*-linked glycans, we screened Sp_0475 against the complex *N*-glycan purified from egg yolk powder, in both its native and sialidase-treated form (34, 35) but no hydrolysis was observed (Table S3). It thus seemed likely that some metal ion or other reaction condition was missing. So, using the basal N-acetyl-β-D-glucosaminidase activity observed, we embarked on a screen of divalent and monovalent metal ions using EDTA-treated enzyme, but could detect no rate enhancement in any case (Table S4).

Based on our experience with other glycosidases, we then proceeded to test a range of anions as potential activators of this basal activity. These included phosphate, sulfate, azide, acetate, formate, citrate, and nitrate (Table S4). Intriguingly, increases in the rate of hydrolysis of MUGlcNAc of approximately 4- and 10-fold were seen in the presence of 300 mM sodium azide, and sodium nitrate, respectively (Fig. S1), whereas the other anions had minimal effect. To try to

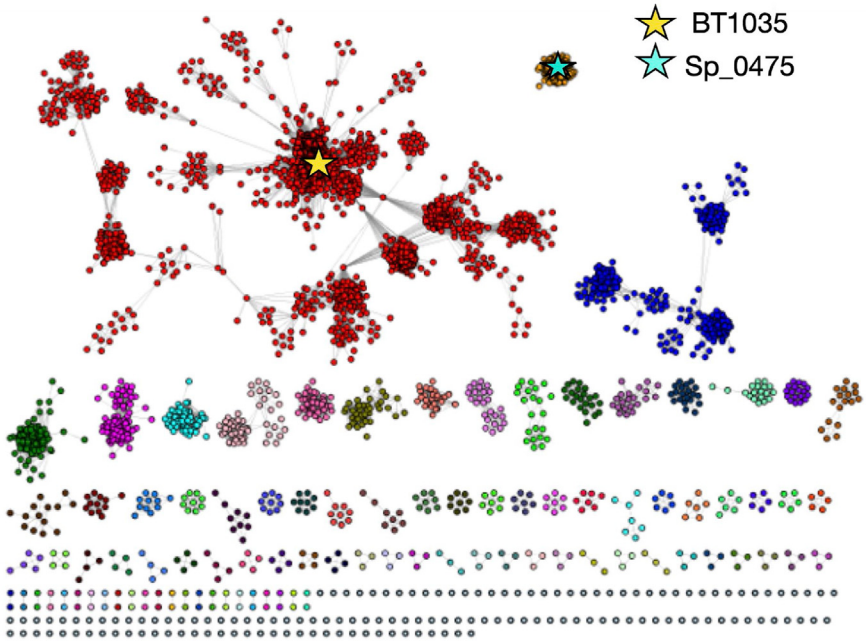

Figure 2. Sequence similarity network of PF16126 family (DUF4838 domain). The sequence similarity network was built from 3826 sequences of the PF16126 family (DUF4838 domain) using an alignment threshold of 10^{-74}. The founding member of the GH163 family, BT1035, is marked with an *yellow star*, while cluster 3 containing the sequence of Sp_0475 is marked with a *turquoise star*.

understand these effects we attempted to measure kinetic parameters for pNPGlcNAc cleavage in the absence and presence of 500 mM nitrate. K_M values were too high to allow saturation of rates but k_{cat}/K_M values could still be measured (Table 1). These revealed an approximate 10-fold increase in k_{cat}/K_M in the presence of 500 mM nitrate. Analysis of reaction mixtures by TLC did not reveal any products of TG, suggesting that the rate increases were due to enhanced interactions between the active site and the aryl glycoside substrate. This in turn suggested that perhaps the preferred substrate contained an anionic substituent, with the most probable candidates being those that are phosphorylated or sulfated at the 6-position.

Accordingly, we synthesized both the 6-phospho- and 6-sulfo analogs of MUGlcNAc and measured the kinetic parameters shown in Table 1. The phosphorylated analog remained a very poor substrate while, by contrast, the 6-sulfo analogue was rapidly hydrolyzed. Sp_0475 has a greater than 110,000 fold preference for 6S-GlcNAc substrates over the nonsulfated versions. Moreover, it exhibits Michaelis–Menten kinetics for pNP-6S-GlcNAc as highlighted in Figure 3*A*, with a K_M of 250 μM and k_{cat} of 65 s^{-1} (Table 1). This is well in the range of kinetic parameters observed for β-N-acetylhexosaminidases (36). Comparable parameters were measured for the methylumbelliferyl glycoside (MU-6S-GlcNAc) whereas the 6-sulfogalactosaminide (MU-6S-GalNAc) was found to be a considerably poorer substrate (∼ 11,000 fold). GH20 β-N-acetylhexosaminidases often cleave both GlcNAc and GalNAc substrates, typically with a preference for one epimer over the other (36). Interestingly, during our initial experiments we observed that Sp_0475 was inhibited by chloride ions when assayed in the absence of a reductant but not when a reductant was present. The mechanism underlying this unusual inhibition behavior remains unclear, but since Sp_0475 is likely located in the cytoplasm, no such inhibition should occur in its natural reducing environment. In line with this, we included tris(2-carboxyethyl)phosphine (TCEP) in all buffers used during purification or kinetic analysis.

The extremely strong preference for the sulfated substrate is impressive. A 110,000 fold difference in k_{cat}/K_M values as a consequence of sulfation corresponds to a decrease in activation free energy of some 28 kJ mol^{-1}. No rate increase is seen as a consequence of phosphorylation even though the bond lengths and geometries of the two species are similar. Earlier studies have assigned the differential recognition of the phosphates and sulfates as arising from their different hydrogen bonding architectures (37–39). The additional space around C6 allows the enzyme to also cleave pNP-6N_3-GlcNAc, indeed some 10-fold faster than it cleaves pNPGlcNAc. Overall, the evidence strongly suggests that Sp_0475 functions as an exo-acting 6-sulfo-N-acetylglucosaminidase. It is likely that Sp_0475 plays a crucial role in the scavenging of 6S-GlcNAc by *S. pneumoniae* where it typically resides in the upper respiratory tract.

Sp_0475 follows a substrate-assisted mechanism

The distant homology of Sp_0475 with GH20 enzymes suggested a substrate-assisted catalytic mechanism. In this mechanism, the substrate acetamide functions as the catalytic

Table 1
Kinetic parameters for hydrolysis of aryl glycosides by WT Sp_0475

Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} mM^{-1})	Approximate fold decrease relative to corresponding 6S-GlcNAc substrate	ΔΔG (kJ/mol) (Relative to corresponding 6S-GlcNAc substrate)
MU-6S-GlcNAc	55.0 ± 0.4	264 ± 7	210 ± 60	/	/
pNP-6S-GlcNAc	65.0 ± 0.3	250 ± 4	260 ± 80	/	/
MU-6S-GalNAc	/	/	(1.8 ± 0.03) × 10^{-2}	11,000	−23
MU-6P-GlcNAc	/	/	(4.1 ± 0.1) × 10^{-4}	525,000	/
pNP-GlcNAc	/	/	(2.2 ± 0.1) × 10^{-3}	118,000	−28
pNP-GlcNAc + 500 mM Sodium Nitrate	/	/	(2.3 ± 0.04) × 10^{-2}	11,000	−5.8*
pNP-6N_3-GlcNAc	/	/	(2.3 ± 0.1) × 10^{-2}	11,000	/

Fold change was determined using the respective 6S-GlcNAc aryl substrate.
ΔΔG (kJ/mol) was calculated using Equation 1, except for * where Equation 2 was used.

$$\Delta\Delta G^{\circ}_{binding} = -RTln\left(\frac{\left(\frac{k_{cat}}{K_M}\right)_{pNP\ or\ MU-6S-GlcNAc}}{\left(\frac{k_{cat}}{K_M}\right)_{modified\ pNP\ or\ MU\ substarte}}\right) \quad (1)$$

$$\Delta\Delta G^{\circ}_{binding} = -RTln\left(\frac{\left(\frac{k_{cat}}{K_M}\right)_{pNP-GlcNAc+500mM\ sodium\ nitrate}}{\left(\frac{k_{cat}}{K_M}\right)_{pNP-GlcNAc}}\right) \quad (2)$$

All measurements were performed at room temperature in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP.

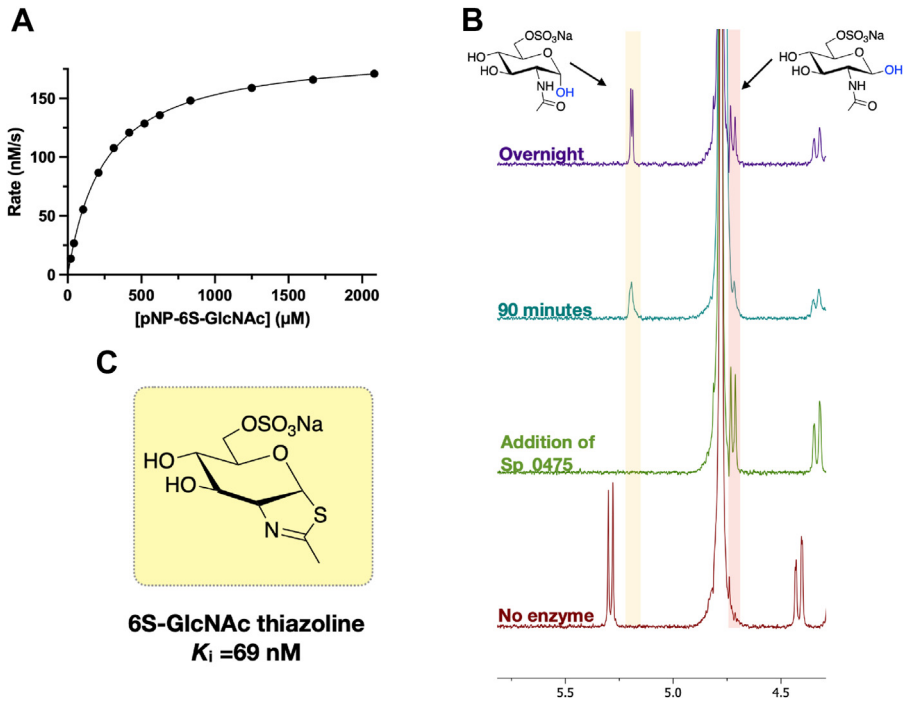

Figure 3. Biochemical characterization of Sp_0475. *A*, rates of cleavage of pNP-6S-GlcNAc by WT Sp_0475 as a function of substrate concentration. *B*, Sp_0475 cleaves with retention of anomeric configuration. Hydrolysis of MU-6SGlcNAc was monitored by NMR spectroscopy. After addition of enzyme a peak at 4.7 ppm corresponding to the β-anomer of 6S-GlcNAc forms (highlighted in *red*). Mutarotation is observed over time as shown by the slow formation of the 6S-GlcNAc α-anomer peak at 5.20 ppm (highlighted in *yellow*) *C*, chemical structure of 6S-GlcNAc-thiazoline shown which functions as a competitive inhibitor for Sp_0475 with a K_i 69 nM.

nucleophile to form an oxazoline/oxazolinium ion intermediate, with the reaction resulting in net retention of anomeric stereochemistry (Fig. 4). To assess whether Sp_0475 affects hydrolysis with the same stereochemical outcome we used NMR spectroscopy. Figure 3*B* shows the time course for hydrolysis of MU-6S-GlcNAc following the addition of enzyme (Fig. S3). Here, the peak at 4.7 ppm appearing immediately following the addition of a large amount of enzyme corresponds to the β-anomer of 6S-GlcNAc, confirming that Sp_0475 is indeed a retaining glycosidase. Subsequent traces follow the appearance of the α-anomer of 6S-GlcNAc (δ = 5.20 ppm), and the loss of the β-anomer (δ = 4.7 ppm) over

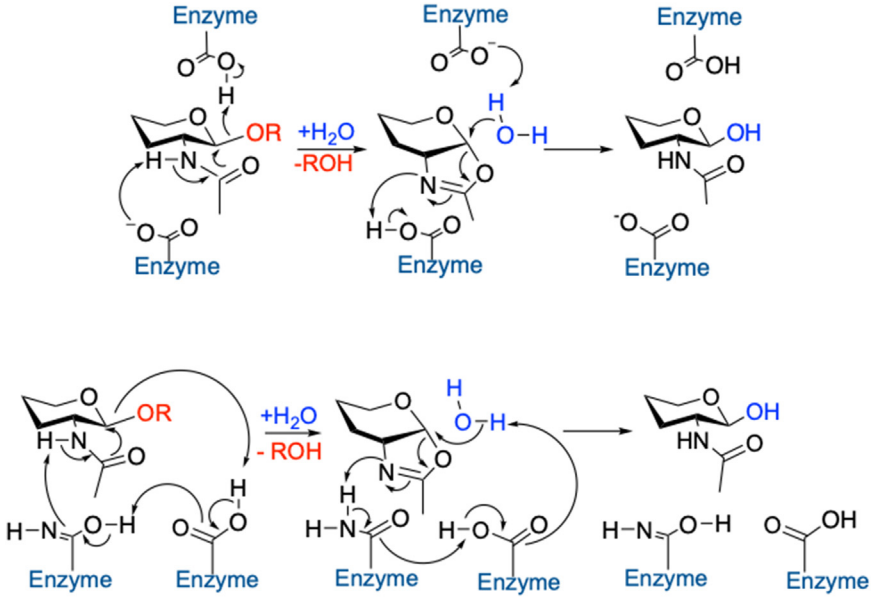

Figure 4. Substrate-assisted mechanismfollowed by members of the GH20 and GH85 family. (*Top*) the canonical mechanism followed by GH20 family members. Briefly, GH20s have an acid–base residue that aids in the cleavage of the glycosidic bond, as well as an additional carboxylate residue that stabilizes the oxazolinium ion intermediate formed during catalysis. (*Bottom*) an atypical substrate-assisted mechanism followed by members of the GH85 family, which also use a conserved carboxylic acid as the acid/base residue, but use a conserved asparagine, in its imidic tautomer to stabilize the oxazolinium ion intermediate through a conserved proton shuttle mechanism. Note: hydroxyl substituents are omitted for clarity.

time until equilibrium is reached through mutarotation. Evidence for an oxazoline intermediate was then obtained by synthesizing a stable mimic and measuring its inhibitory properties (40, 41). 6S-GlcNAc thiazoline (shown in Fig. 3*C*) was synthesized as previously described (42), subjected to kinetic analysis and found to function as a potent competitive inhibitor with a K_i of 69 nM (Fig. S16). The high affinity of this intermediate analogue, coupled with the stereochemical experiments, strongly support the proposal that Sp_0475 carries out catalysis *via* a substrate-assisted mechanism similar to that of the GH20 family (41). The nonsulfated GlcNAc thiazoline was also shown to be a competitive inhibitor, but a much weaker binder with a K_i value of 6.6 mM (Fig. S15). This almost 100,000-fold weaker binding nicely parallels the 110,000 fold difference in k_{cat}/K_M values between the sulfated and nonsulfated substrates, highlighting the fact that the inhibitor is recruiting essentially all the binding interactions derived between the enzyme and substrate sulfate at the reaction transition state. Furthermore, nitrate was observed to function as a weak competitive inhibitor with a K_i value of 68 mM (Fig. S14), thus contributing over 6 kJ/mol of binding energy (Table 2). This finding is further supported by the nearly equivalent binding energy of approximately 6 kJ/mol determined from k_{cat}/K_M values (Table 1).

Structure and active site of Sp_0475

Although our efforts to crystallize Sp_0475 in a useful form proved unfruitful, the AlphaFold structural model deposited in the AlphaFold Protein Structure Database (43) provided a convincing structural model of Sp_0475. Based on this, we made several mutants of residues believed to reside in the active site and evaluated their activity through kinetic analysis. As shown in Figure 5*C*, Sp_0475 adopts the canonical TIM barrel (triose-phophate isomerase) fold of the GH20 family, GH-K clan members, and several other clans within the CAZy database (40, 44). However, while the active sites of GH20 family members are located at the C-terminal end of the TIM barrel and contain a highly conserved D-E catalytic pair following a conserved sequence (H-X-G-G) (12) (Fig. 5*A*), Sp_0475 entirely lacks these conserved sequence signatures, making identification of active site residues challenging.

Moreover, sequence analysis of the members of cluster 3 of our SSN (Fig. 2) did not identify any clear candidates for catalytic residues near the C-terminal end of the TIM barrel (Fig. S4). We therefore performed a DALI search on the AlphaFold structure of Sp_0475 (45) and identified a GH115 (Protein Data Bank (PDB):4C90) from *Bacteroides ovatus* (46) as the structure with the highest Z-score (21.1). However, members of the GH115 family are α-glucuronidases that follow an inverting mechanism and are thus unlikely to be good models for a retaining glycosidase. The DALI search also revealed structural similarities between the retaining GH84 *O*-GlcNAcase from *Clostridium perfringens* (PDB:2V5C) and Sp_0475. GH84 family members follow a substrate-assisted mechanism using a pair of conserved carboxylates to carry out hydrolysis of the glycosidic linkage, similar to the GH20s (47). The superimposed structures of 2V5C and Sp_0475 reveal a single carboxylic acid (D309) overlaid with the acid/base residue of the GH84, as shown in Fig. S5, which is found in the same general area as the GH20 active site (Fig. 5*D*). Further sequence analysis shows D309 to be highly conserved as indicated by the sequence alignment shown in Figure 5*B* (and Fig. S4).

To tease out the function of this carboxylic acid residue we prepared the corresponding Ala mutant of D309 through site-directed mutagenesis. Kinetic parameters for the purified mutant are summarized in Table 3. Notably, K_M and k_{cat} values for the cleavage of MU-6S-GlcNAc by the D309A mutant are 170-fold, and 7800-fold lower, respectively, than those of WT, resulting in a more than 50-fold drop in k_{cat}/K_M . Attempts to use anion rescue studies on mutants to identify key catalytic residues (48) were confounded in this case by the anionic nature of the substrate and effects of anions on its binding. To gain further insight into potential roles we chose to measure the pH dependence of enzyme activity for the WT and mutant, as significant changes are expected when the essential acid/base residues are modified. The low K_M value measured with D309A made the measurement of the pH dependence of k_{cat}/K_M by depletion kinetics (14, 49) challenging. We therefore measured k_{cat} as a function of pH instead since those measurements could be done at substrate saturation. We first determined the pH range over which the enzyme retained at

Table 2
Inhibition constants, $\Delta G_{Binding}$ and $\Delta\Delta G_{Binding}$ for Sp_0475 with several competitive inhibitors

Inhibitor	K_i	$\Delta G_{Binding}\left(\frac{kJ}{mol}\right)$	$\Delta\Delta G_{binding}\left(\frac{kJ}{mol}\right)$
GlcNAc-thiazoline	6.6 ± 0.2 mM	−12.4	−28.4
Sodium nitrate	68 ± 13 mM	−6.6	/
6S-GlcNAc-thiazoline	69 ± 7 nM	−40.8	/

$\Delta G_{binding}$ (kJ/mol) and $\Delta\Delta G_{binding}$ (kJ/mol) were calculated using Equations 3, and 4, respectively.

$$\Delta G_{binding} = RTln(K_i) \tag{3}$$

$$\Delta\Delta G_{binding} = \Delta G_{6SGlcNAc-thiazoline} - \Delta G_{GlcNAc-thiazoline} = RTln\left(\frac{K_{i6SGlcNAc-thiazoline}}{K_{iGlcNAc-thiazoline}}\right) \tag{4}$$

All measurements were performed at room temperature in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP.

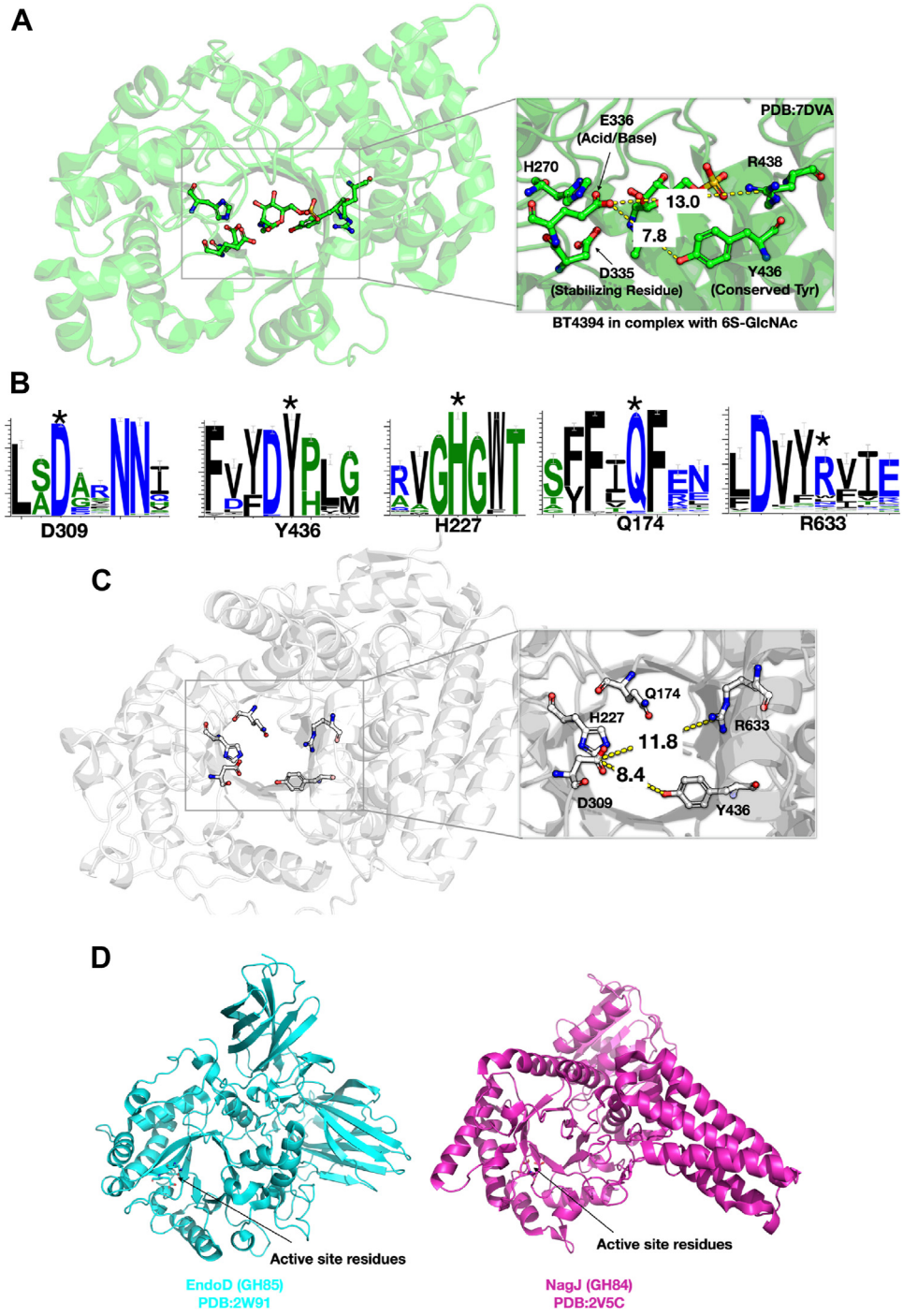

Figure 5. 3D structures of BT4394, Sp_0475, EndoD, and NagJ with active site residues highlighted for BT4394 and Sp_0475. *A*, PyMOL rendering of BT4394 with 6S-GlcNAc bound in the active site (PDB: 7DVA) (54). BT4394 is a GH20 family member with 6-sulfo-GlcNAcase activity. The three-dimensional structure of BT4394 is illustrated on the *left*, presenting a $(\beta/\alpha)_8$ barrel fold. The active sites of GH20 family members are situated at the C-terminal end of the TIM barrel fold. The inset (on *right*) emphasizes the essential active site residues that are critical to catalysis and binding of 6S-GlcNAc in the active site. *B*, sequence logos illustrate the conservation of several positions in Sp_0475. The height of each letter in the logo represents the degree of conservation at that specific position. Residues of interest are indicated below each logo with an *asterisk* and the corresponding residue position. These logos were generated using WebLogo (65) after performing a sequence alignment of all sequences found in cluster 3 (*C*) PyMOL rendering of the Sp_0475 AlphaFold structure obtained from the AlphaFold Database (43). The overall fold of Sp_0475 is shown on the *left*, which depicts the canonical $(\beta/\alpha)_8$ barrel fold. The inset to the right highlights the conserved residues of interest located at the C-terminal end of the $(\beta/\alpha)_8$ barrel fold. *D*, PyMOL rendering of EndoD (PDB:2W91) from the GH85 family and NagJ (PDB:2V5C) from the GH84 family highlighting the three-dimensional structure of each. The key catalytic residues are clearly marked, and are both found near the end of the $(\beta/\alpha)_8$ barrel fold. PDB, Protein Data Bank; TIM, triosephosphate isomerase.

least 80% activity during the time taken for an assay (Fig. S7). The pH dependence of k_{cat} within the stable pH range for both WT and D309A is shown in Fig. S7. While the k_{cat} of WT drops with increasing pH, the dependence is modest, amounting to only 3-fold over 2 pH units, thus not corresponding to an essential side chain ionisation in this range,

Table 3
Kinetic parameters for cleavage of aryl glycosides by Sp_0475 mutants

Mutant	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} mM^{-1})	Approximate fold change in k_{cat}/K_M	$\Delta\Delta G_{binding}$ $\left(\frac{kJ}{mol}\right)$
D309A	MU-6S-GlcNAc	$(7.0 \pm 0.1) \times 10^{-3}$	1.5 ± 0.1	4 ± 1	53	−9.8
	pNP-6S-GlcNAc	$(1.7 \pm 0.02) \times 10^{-1}$	4.0 ± 0.1	41 ± *13*	6	−4.6
R633Q	pNP-6S-GlcNAc	/	/	$(2.7 \pm 0.1) \times 10^{-2}$	9700	−22.8
	pNP-GlcNAc	/	/	$(1.4 \pm 0.02) \times 10^{-2}$	/	/
R633I	pNP-6S-GlcNAc	/	/	$(6.9 \pm 0.03) \times 10^{-2}$	3700	−20.4
	pNP-GlcNAc	/	/	$(2.9 \pm 0.2) \times 10^{-2}$	/	/
	pNP-GlcNAc + 500 mM Sodium Nitrate	/	/	$(2.8 \pm 0.1) \times 10^{-2}$	/	/
	pNP-6N_3-GlcNAc	/	/	$(1.0 \pm 0.03) \times 10^{-1}$	/	/
Y436F	pNP-6S-GlcNAc	$(1.5 \pm 0.02) \times 10^{-1}$	57 ± 4	3.0 ± 0.5	86	−11.1
H227N	pNP-6S-GlcNAc	17.5 ± 0.3	3000 ± 80	6.0 ± 4.0	43	−9.3
Q174A	pNP-6S-GlcNAc	/	/	1.3 ± 0.1	200	

Fold change of k_{cat}/K_M with respect to the corresponding aryl glycoside by WT Sp_0475. k_{cat}/K_M values of MU-6S-GlcNAc and pNP-6S-GlcNAc used to calculate fold change were 210 s^{-1} mM^{-1}, and 260 s^{-1} mM^{-1}, respectively.
All measurements were performed at room temperature in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP.
ΔΔG (kJ/mol) was calculated using Equation 5.

$$\Delta\Delta G_{binding} = -RTln\left(\frac{\left(\frac{k_{cat}}{K_M}\right)_{WT}}{\left(\frac{k_{cat}}{K_M}\right)_{Mutant}}\right) \quad (5)$$

while k_{cat} for the mutant does not show a similar decrease over the same pH range. Thus, while we cannot definitively assign D309 a role in acid/base catalysis, it is undoubtedly very important and the drop in K_M upon mutation is consistent with such a role (36, 48).

Superimposition of the structures of 2V5C and Sp_0475 did not identify any additional carboxylic acid residue candidates in the immediate vicinity of D309; thus, we expanded our search. While the acid/base and oxazoline-stabilizing residues of this class of hexosaminidases are typically carboxylic acids, this is not the case for enzymes of the GH85 family, which share similarities in fold and active site location (Fig. 5*D*). In those cases a conserved asparagine in its imidic tautomer has been proposed to stabilize the oxazoline through a conserved proton shuttle (50) as shown in Figure 4*B*. Therefore, we set out to identify other highly conserved residues within 10 Å of D309, including Gln, Asn, Glu, and Asp. By coupling the AlphaFold structural model to our sequence analysis, we identified Q174 as being within 8 Å of D309 (Fig. 5*C*). This residue is largely conserved within cluster 3 with the only substitutions being a small number of sequences containing Glu rather than Gln (Fig. 5*B*). Indeed, the mutation of Q174A resulted in a nearly 200-fold decrease in k_{cat}/K_M, suggesting that it may well play a role in stabilizing the oxazoline intermediate.

Another conserved residue identified nearby is H227, which lies within 3 Å of D309, and 5 Å of Q174 (Fig. 5*C*). A similarly placed His residue is found in some GH20s (Fig. 5*A*), and both computational and mutagenesis studies have pointed at important roles, possibly in setting the pKa of the acid/base residue (40, 51). We, therefore, prepared the H227N mutant and measured a k_{cat}/K_M value some 40-fold lower than that of WT pointing to similar importance. A second active site residue in GH20s that plays a critical role in stabilizing the oxazoline and orienting the incoming water is a conserved tyrosine (Fig. 5*A*), for which mutations to either Phe, His, or Gln significantly diminish activity (36, 52). In Sp 0475 Y436 is found within 8 Å (Fig. 5*C*) of D309 and is fully conserved in this cluster (Fig. 5*B*), suggesting it may well play a similar role. Consistent with this, the k_{cat} value of the Y436F mutant is 400-fold lower than that of the WT, with a more modest decrease in K_M (Table 3) as is typical for such mutants (36, 53). Based on the AlphaFold model and Michaelis–Menten parameters of the Y436F mutant, it probably shares a similar function to that of the conserved tyrosine in GH20s. It is worth noting that all predicted active site residues of Sp_0475 sit near the C-terminal end of TIM barrel, which is in the same general region the active site of GH20 is located. Though they are located largely at the end of the TIM barrel, they are arranged in a slightly different orientation as illustrated in overlaid structures of BT4394 and Sp_0475 in Fig. S6.

The residues involved in the recognition of sulfated sugars by 6-sulfoGlcNAcases belonging to the GH20 family have been extensively studied (54, 55). Figure 5*A* highlights the active site of BT4394, a GH20 family member that exhibits 6-sulfoGlcNAcase activity (54). Recent work has established the R438 residue of BT4394 as playing a critical role in stabilizing the sulfate (Fig. S13) (54). Using the AlphaFold structure, we identified R633 as a likely candidate to play such a role in Sp_0475 (Fig. 5*C*). Table 3 summarizes the kinetic parameters for cleavage of pNP-6S-GlcNAc by two mutants in which the positively charged Arg has been replaced by Ile or Gln. The two mutants cleave pNP-6S-GlcNAc with k_{cat}/K_M values almost 4000 to 10,000 fold lower than those of WT. By contrast, they cleave pNP-GlcNAc some 6-fold faster. Interactions between the sulfate and R633 contribute at least 22.8 kJ/mol to the

binding affinity, which is close to the nearly 28 kJ/mol stabilization provided by the sulfate group based upon relative k_{cat}/K_M values of the two substrates with WT. Moreover the 10-fold increase in the rate of hydrolysis of pNP-GlcNAc in the presence of nitrate is no longer observed for the R633I mutant, suggesting that R633 mediates the interaction between nitrate and pNP-GlcNAc. Interestingly, the sequence alignment shown in Fig. S4 shows that R633 is not highly conserved within this cluster (Fig. 5*B*), suggesting that other enzymes in cluster 3 recruit alternative residues to stabilize the sulfate group, similar to what has been reported recently (54, 55).

Investigation of GH1 found downstream of Sp_0475

As one approach toward the identification of a natural substrate of this GH we examined the genes surrounding Sp_0475 as shown in Fig. S8. In addition to both lactose type phosphoenolpyruvate-dependent sugar phosphotransferase system and ATP-binding cassette transporter components, a gene annotated as a 6-phospho-β-galactosidase (Sp_0477) from the GH1 family is found downstream. This organization suggests that Sp_0475 and the GH1 might work together to degrade a common natural substrate such as keratan sulfate. This glycosaminoglycan consists of repeating N-acetyllactosamine [3-Gal-β1,4-GlcNAc-β1] disaccharide units, with variable regions of sulfation at the 6 hydroxyls of Gal and/or GlcNAc (56). We therefore cloned Sp_0477 and tested its *in vitro* activity against keratan sulfate isolated from porcine lung, but TLC analysis indicated no hydrolysis products when incubated alongside Sp_0475. Testing against fluorogenic substrates was more successful, and while Sp_0477 could not hydrolyze aryl glycosides of 6-sulfo-galactose, it could hydrolyze the corresponding 6-sulfo- and 6-phospho-glucoside glycosides. However, it is not clear what oligosaccharide substrates this might suggest.

TG studies with WT Sp_0475

An alternative approach to identifying possible natural substrates, or at least to narrowing the probable substrate range, is to investigate whether the enzyme can carry out TG, and if so to screen for acceptor molecules that bind productively in the +1 site by identifying the product formed (Fig. 6). Using a high concentration of pNP-6S-GlcNAc as the activated donor substrate we monitored reaction mixtures by TLC for possible TG in the presence of various monosaccharides. Indeed, as shown in Figs. S17–S19 TG product was observed for all the acceptors surveyed, but the TG reaction was most dominant when Gal or GalNAc were the acceptors suggesting that these may naturally be the preferred sugar for the +1 site. Closer examination of the TLC reveals two clear TG products when Gal is the acceptor sugar. Mass spectrometric analysis indicates that they are likely linkage isomers, and further experiments show that 6S-GlcNAc does not serve as an acceptor for Sp_0475. Curiously, when GalNAc is used as the acceptor only one TG product is formed leading us to believe that GalNAc is the preferred monosaccharide in the +1 subsite (Fig. 6). To simplify purification and analysis of the product we used pNPGalNAc as the acceptor, though the poorer solubility of pNPGalNAc versus GalNAc adversely affected the yield (Fig. S20). 1-D and 2-D NMR analysis was used to characterize the isolated TG product (Table S5). The large coupling constant of 8.5 Hz associated with the anomeric proton of 6S-GlcNAc at 4.6 ppm indicates a β-glycosidic linkage of the isolated TG product. To assign the linkage we used (^{1}H, ^{13}C)-HMBC, which reveals three-bond correlations of [6S-GlcNAc

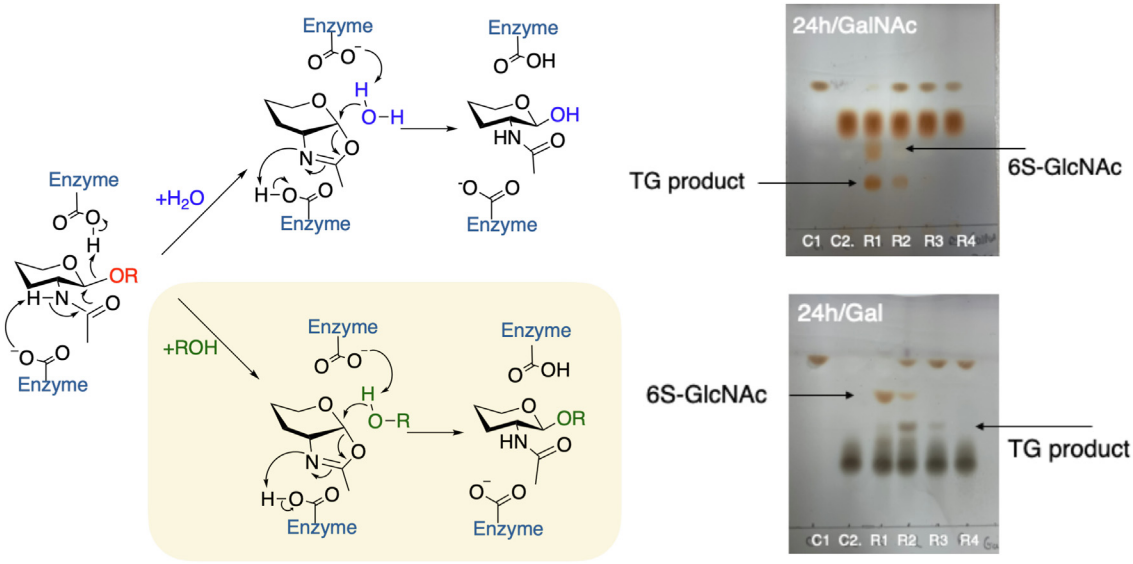

Figure 6. Transglycosylation activity of Sp_0475. (*Left*) Generalized scheme highlighting transglycosylation, and hydrolysis for GHs following a substrate assisted mechanism. (*Right*) TLC of transglycosylation reaction of GalNAc (*top*) and Gal (*bottom*). General reaction conditions consisted of 25 mM pNP-6S-GlcNAc and 50 mM Gal or GalNAc with vary amounts of WT Sp_0475 monitored after 24 h of incubation at room temperature. All reactions were incubated at room temperature in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP. Top TLC was run in 2:1:1 (butanol:water:acetic acid) mobile phase. Bottom TLC was run in 5:4:4:1 (butanol:methanol:ammonium hydroxide:water) mobile phase. C1: 25 mM pNP-6S-GlcNAc, C2: 50 mM acceptor (Gal or GalNAc) R1: 25 mM pNP-6S-GlcNAc, 50 mM acceptor (Gal or GalNAc), and 1 μM WT. GalNAc, N-acetylgalactosamine; TCEP, tris(2-carboxyethyl)phosphine.

C-1, pNP-GalNAc H-6] and [pNP-GalNAc C-6, 6S-GlcNAc H-1], suggesting a β1,6 linkage between 6S-GlcNAc and pNPGalNAc. Incubation of a sample of this isolated TG product with WT Sp_0475, resulted in complete hydrolysis after 1 h (Fig. S21) confirming that the 6S-GlcNAc-β1,6-GalNAc linkage can be cleaved and is likely the natural substrate. *S. pneumoniae* commonly colonizes the upper respiratory tracts of humans, where it uses host glycans as a major source of nutrients (24, 57). The mucus layer in the human body provides a rich source of carbohydrates, including heavily sulfated *O*- and *N*-glycans, that can be metabolized by this microbe (2). This GlcNAc-β1,6-GalNAc linkage is found within the structure of Core 2-*O*-glycans, as depicted in Figure 1. It is therefore quite plausible that sulfated Core 2 *O*-glycans, once released from the peptide backbone and transported into the cytoplasm, could serve as a natural substrate for Sp_0475. However, further *in vivo* studies would be needed to confirm this hypothesis.

Discussion and conclusion

In this work, we leveraged SSNs to aid in the discovery of a 6-sulfoGlcNAcase that appears to be the founding member of a new GH family. By constructing an SSN of the DUF4838 domain (PF16126) we illustrated the vast diversity present within this grouping, implying the existence of potentially unique substrate specificities within this large domain. Our discovery that Sp_0475 was indeed a 6-sulfoGlcNAcase was largely inspired by our serendipitous experimental observation that the presence of nitrate resulted in a substantial increase in the rate of hydrolysis of pNPGlcNAc. It will be interesting to see if similar rate activation by nitrate or other anions is seen in other sulfosugar-active enzymes when tested with non-sulfated substrates. If so, this may provide a useful screening strategy for the identification of other GHs suspected to be active on sulfated substrates.

Sp_0475 and presumably other members of cluster 3 represent a new GH family that has considerable structural and mechanistic parallels with the distantly related GH20 family. The success we observed in identifying the active site residues through our site-directed mutagenesis studies illustrate the power of new structure prediction methods. Furthermore, our biochemical studies of Sp_0475 may broaden or inspire our understanding of other DUF4838-containing GHs, such as the newly established GH163 family. Comparison of the active site of Sp_0475, with the hypothetical active site of BT1035 of the GH163 family reveals that many of the same catalytic residues investigated in this study are shared, suggesting similarities in mechanism, and catalytic residues (Fig. S22). Further investigation of the DUF4838 domain is likely to lead to the discovery of other novel glycosidases with distinct activities.

While the physiological role of Sp_0475 remains unclear, its specificity for sulfated oligosaccharides, its capacity for TG and its oxazoline mechanism suggest it may prove useful in the assembly of sulfated oligosaccharides. This could provide a useful alternative enzymatic approach to the use of highly specific sulfotransferases, which can be challenging enzymes to deploy. Further, the high cost of the activated donor, 3′-phosphoadenosine-5′-phosphosulfate makes scale-up a problem, even in the presence of substrate recycling schemes (58, 59). Selective chemical sulfation of pNPGlcNAc with trimethylamine sulfur trioxide complex ($Et_3N{:}SO_3$) would provide inexpensive access to the sulfated donor sugar, which could further be incorporated to target acceptors with a single engineered enzyme. The relatively promiscuous nature of Sp_0475 demonstrated by the formation of TG products when incubated with Glc, GlcNAc, Man, Gal, and GalNAc suggests that this one enzyme could be used to prepare a range of sulfated oligosaccharides.

Experimental procedures

Materials and reagents

All chemicals, buffers, and reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Biosynth/Carbosynth or Toronto Research Chemicals unless otherwise stated. Dry solvents were distilled under a nitrogen atmosphere over calcium hydride. All glassware used in synthesis was dried in an oven prior to use. TLCs were run on TLC Silica gel 60 F_{254} plates (Merck) and were visualized by either a p-anisaldehyde stain or 10% (v/v) sulfuric acid in methanol followed by heating. For TLCs stained with 10% (v/v) sulfuric acid in methanol a UV tray was used to visualize the TLC. Sep-Pak C8 35 cc Vac Cartridge, 10 g sorbent cartridges were purchased from Waters. The sodium form of DOWEX 50WX8 was routinely prepared from the DOWEX 50WX8 hydrogen form by washing the resin with 1 M sodium bicarbonate, followed by washing the resin with water until conductivity of flow-through stabilizes. All MS was performed at the UBC Mass Spectrometry Centre. Low resolution mass spectrometry was run on the Waters 2695 HPLC, with Waters ZQ equipped with an electrospray chemical ionization ion source. High resolution mass spectra were collected by the UBC Mass Spectrometry Centre. All PCR reactions used Phusion High-Fidelity DNA Polymerase from New England Biolabs. PCR clean ups and agarose gel extractions were done with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). For cloning and site-directed mutagenesis, PCR reactions were transformed into Subcloning Efficiency DH5α competent cells (Thermo Fisher Scientific). Plasmid DNA was extracted with the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Plasmid sequencing related to cloning and mutagenesis was done at Genewiz (Azenta Life Sciences).

Cloning and site directed mutagenesis of Sp_0475 and Sp_0477

All cloning for this work used the mega primer mediated molecular cloning strategy as previously described (60). Primers listed in Table S1 were used to clone Sp_0475 and Sp_0477 into pET-28a(+). Each construct contained an N-terminal hexahistidine tag and thrombin cleavage site. Briefly, primers in Table S1 were used to amplify desired genes directly from genomic DNA (*S. pneumoniae* TIGR4). Following the first PCR, an agarose gel was used to visualize the amplified

DNA, and the insert with the desired number of kilobases was extracted. Next, circular amplification was performed using the purified DNA, and pET-28-a(+). Following DPN1 digest, the PCR product was transformed into DH5α competent cells. Desired mutations were introduced using primers listed in Table S1.

Expression and purification of WT Sp_0475 and mutants

All constructs used were transformed into *E. coli* BL21 (DE3) cells for expression. For large scale growth, cultures were grown in ZY5052 auto induction media (61) for 72 h (18 °C,180 rpm). Cells were harvested through centrifugation (5000 rpm, 20 min, 4 °C), and stored at −20 °C till needed. Frozen cell pellets were resuspended in lysis buffer (25 mM tris(hydroxymethyl)aminomethane (tris) (pH 7.5), 300 mM sodium chloride (NaCl), 10 mM imidazole, 1% (w/v) glycerol, 1 mM TCEP, 1X protease inhibitor tablet EDTA free (Thermo Fisher Scientific Pierce), 2 U benzonase (EMD Millipore), 0.1 mg ml^{-1} lysozyme) and sonicated on ice for a total of 4 min (pulse time: 4 min pulse: 5 s, pause: 15 s, amplitude: 35%). Lysate was then clarified through centrifugation (13,500 rpm, 35 min, 4 °C) and supernatant was loaded directly onto either a 5 ml HisTrap HP column (GE HealthCare) using a peristaltic pump, or 5 ml of HisPur Ni-NTA resin pre-equilibrated with buffer A (25 mM tris (pH 7.5), 300 mM NaCl, 10 mM imidazole, 1% (w/v) glycerol and 1 mM TCEP). The 5 ml HisTrap HP Column (GE HealthCare) was attached to the ÄKTA GO protein purification system (Cytiva Life Sciences) and washed with buffer A till the absorbance at 280 nm stabilized. Then a linear gradient to 100% buffer B (25 mM tris (pH 7.5), 50 mM NaCl, 400 mM imidazole, 1% (w/v) glycerol + 1 mM TCEP) over 50 min was set, and 3 ml fractions were collected. Alternatively, after loading the crude lysate onto 5 ml of Ni-NTA resin HisPur, resin was washed with 30 column volumes (CVs) of buffer A, followed by 10 CVs of buffer C (25 mM tris (pH 7.5), 300 mM NaCl, 40 mM imidazole, 1% (w/v) glycerol, and 1 mM TCEP). The protein of interest was then eluted with approximately 8 CVs of buffer B. SDS-PAGE was used to confirm which fractions contained the desired protein. Fractions containing Sp_0475 were pooled and concentrated using an Amicon Ultra centrifugal filter unit with a molecular weight cutoff (MWCO) of 30 kDa and exchanged to buffer D (25 mM 2-(N-morpholino) ethanesulfonic acid (Mes) (pH 7.0), 100 mM NaCl and 1 mM TCEP). Sp_0475 was then loaded onto a Superdex 200 column equilibrated with buffer D. Fractions containing desired protein were confirmed through SDS PAGE gel (Fig. S2 for representative gel of purification), and then concentrated with Amicon Ultra centrifugal filter unit with a MWCO of 30 kDa. Molar extinction coefficients of WT Sp_0475 and mutants were calculated using an online protein calculator (https://protcalc.sourceforge.net/). Absorbance at 280 nm was measured with a Take^3 Micro-Volume Plate using a plate reader (BioTek Synergy) and protein concentrations were calculated accordingly. Typical yields from 1 L of autoinduction media varied anywhere from 10 to 30 mg depending on the protein purified. Concentrated protein was then flash frozen in liquid nitrogen and stored at −70 °C until needed.

Kinetic analysis of WT Sp_0475 and mutants

All kinetic assays were performed using a plate reader (Biotek Synergy) with Corning 96-Well plates (clear, flat bottom, half -area). All measurements were performed in triplicate and the total volume for each well was 90 µl. Substrates used were either purchased (Toronto Research Chemicals, Sigma-Aldrich, and Carbosynth) or prepared as described below. For para-nitrophenyl glycoside substrates, absorbance was monitored at 405 nm. For methylumbelliferyl glycosides absorbance (absorbance:360 nm) and fluorescence (excitation:360 nm, emission:450 nm) data were collected. All kinetic assays were performed at room temperature, in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP. The amount of enzyme used for assays varied depending on activity; see SI for the concentration of enzyme used for each mutant/assay (Figs. S9-S12). Initial rates were determined (either absorbance/s, or relative fluorescence units/s) and converted to M/s with appropriate standard curves under identical conditions. Data were either fit using nonlinear regression to obtain Michaelis–Menten kinetic parameters, or linear regression to obtain specificity constants using Prism (GraphPad) (https://www.graphpad.com/features).

Preparation of MU-6P-GlcNAc

ATP-dependent beta-glucoside kinase from *Klebsiella pneumoniae* was expressed and purified as previously described (62). For the synthesis of MU-6P-GlcNAc, the reaction included 0.2 mg ml^{-1} of beta-glucoside kinase, 5 mM MUGlcNAc, 8.7 mM adenosine 5′-triphosphate disodium salt, and 1 mM $MgSO_4$ in 50 mM tris (pH 7.5) with final concentration of 1.5% (v/v) of dimethylsulfoxide (DMSO). The reaction was incubated overnight at 37 °C, and TLC was used to check the progress of the reaction. When the reaction had gone to completion, the reaction mixture was lyophilized, redissolved in a minimal amount of water and loaded directly onto a Sep-Pak C8 column (10 g sorbent). The column was washed with water, and 2 ml fractions were collected. Fractions were analyzed through TLC, and those containing MU-6P-GlcNAc (2:1:1 butanol: water: acetic acid, Rf: 0.41) were pooled and immediately incubated with the sodium form of DOWEX 50WX8 resin for 2 h at 4 °C. Resin was removed through filtration, and solution was lyophilized to afford the desired product in the disodium form. NMR: ^{1}H NMR (600 MHz, D_2O) δ 7.26 (d, J = 8.8 Hz, 1H. H-5′), 6.80 (dd, J = 8.8, 2.4 Hz, 1H, H-6′), 6.58 (d, J = 2.5 Hz, 1H, H-8′), 5.80 (s, 1H, H-3′), 5.15 (d, J = 8.4 Hz, 1H, H-1), 4.32 to 4.08 (m, 2H, H-4&5), 4.03 (dd, J = 10.3, 8.4 Hz, 1H, H-2), 3.80 to 3.75 (m, 3H, H-6 & 3), 2.29 to 1.94 (m, 6H, NHAc, Me). ^{13}C NMR (151 MHz, D_2O) δ 174.2, 163.1, 158.7, 155, 152.6, 125.9, 114.0, 113.0, 110.4, 102.8, 98.1, 74.8, 74.7, 72.4, 68.5, 62.8, 54.9, 21.9, 17.4. HRMS: Calcd for $C_{18}H_{22}NO_{11}P$: ($[M+Na]^+$): 482.0792 found 482.0830.

Preparation of 6S-GlcNAc-thiazoline

GlcNAc-thiazoline (175 mg, 797 µmol) was suspended in dry pyridine (12 ml) and the reaction mixture was cooled to 0 °C. Sulfur trioxide triethylamine complex (139 mg, 998 µmol) was added to the mixture and stirred for 30 min. The reaction

mixture was then moved to a 4 °*C* incubator, and stirred for 18 h while stirring. The resulting white product was filtered off and washed thoroughly with pyridine. The residue was then dissolved in water (50 ml) and immediately incubated with the sodium form of DOWEX 50WX8 resin for 2 h at 4 °*C*. Resin was then removed through filtration, and solution was lyophilized to afford 110 mg (42%) of the sodium-6S-GlcNAc-thiazoline. NMR and ESI-MS was in agreement with literature values (42).

Competitive inhibition experiments

All competitive inhibition assays were performed using a plate reader (BioTek Synergy H1) with Corning 96-well plates (clear, flat bottom, half-area). The final concentration of WT Sp_0475 was 2.6 nM, and all inhibition measurements were performed at room temperature, in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP. Rates of hydrolysis of either MU-6S-GlcNAc or pNP-6S-GlcNAc were measured at four different substrate concentrations, in the presence of several different concentrations of 6S-GlcNAc-thiazoline (0–2.9 μM), sodium nitrate (0–467 mM), or GlcNAc-thiazoline (0–48.4 mM). Data were fit to a competitive inhibition model using nonlinear regression analysis on GraFit (version 7.0.3) (http://www.erithacus.com/grafit/), see SI (Figs. S14-S16).

Stereochemical outcome of Sp_0475 monitored through NMR spectroscopy

Sp_0475 was exchanged into deuterated buffer (5 mM sodium phosphate, pH 7) using an Amicon Ultra centrifugal filter unit with a MWCO of 30 kDa. To initiate reaction, 50 μl of WT Sp_0475 (16.9 mg ml^{-1}) was added to afford a final working concentration of 2.1 mg ml^{-1} of WT Sp_0475 and 5 mM MU-6S-GlcNAc in 5 mM deuterated sodium phosphate (pH 7), with a final d_6-DMSO concentration of 5% (v/v). 1H NMR spectra was acquired using a 400 MHz spectrometer (Bruker). Full time course hydrolysis of MU-6S-GlcNAc is shown in Fig. S3.

pH dependence of WT Sp_0475 and D309A

All pH stability and dependence data were collected using a plate reader (Biotek Synergy H1) with Corning 96-Well plates (clear, flat bottom, half -area). Since the K_M for D309A Sp_0475 is significantly lower than that of WT Sp_0475, substrate depletion experiments were not feasible, thus we monitored k_{cat} with respect to pH. Briefly, to determine the working pH range within which WT Sp_0475 and D309A did not lose more than 20% activity over the course of 5 min at a given pH, we performed enzyme stability experiments. For D309A Sp_0475 experiments, enzyme stocks were diluted to the appropriate concentration in 5 mM Hepes and 1 mM TCEP pH 7.2 and then reaction was initiated *via* a 7-fold dilution into an assay mixture (25 mM sodium citrate, 25 mM Mes, 25 mM Hepes, 25 mM glycine, and 1 mM TCEP) containing 131 μM pNP-6S-GlcNAc. For WT Sp_0475, enzyme stocks were prepared to the appropriate concentration in 5 mM Hepes and 1 mM TCEP pH 7.2, and the reaction was initiated with a 7-fold dilution into an assay mixture (25 mM citrate, 25 mM Mes, 25 mM Hepes, 25 mM glycine, and 1 mM TCEP) containing 1.3 mM pNP-6S-GlcNAc. For both WT and D309A Sp_0475 we also measured rates at the pH extremities using substrate at double the concentration to ensure we were at saturation conditions.

TG acceptor screening

All TG reactions were performed in 25 mM Mes (pH 7.0), 100 mM NaCl, and 1 mM TCEP, at a total volume of 20 μl. For monosaccharide screening, reactions contained 25 mM pNP-6S-GlcNAc, 50 mM monosaccharide acceptor (Gal, GalNAc, Man, Glc, GlcNAc, and 6S-GlcNAc), with varying amounts of WT Sp_0475 (1 μM-1 nM). All reactions stated above were incubated at room temperature and monitored through TLC. TLCs were visualized by p-anisaldehyde stain. In the cases where a product was generated, low resolution MS was used to confirm product formation.

General procedure for the preparation of pNP-6S-GlcNAc, and MU-6S-GalNAc

Aryl glycosides were dissolved in dry pyridine, to give a final concentration of approximately 37 mM, and mixture was subsequently cooled to 0 °C. Sulfur trioxide triethylamine complex (0.8 eq) was dissolved in dry pyridine, and then added to the aryl glycoside solution dropwise while mixing (final concentration of the aryl glycosides is approximately 30 mM). Following the addition of sulfur trioxide triethylamine complex, the reaction mixture was stirred at 4 °C, and incubated for an additional 56 h while stirring. The reaction is then quenched with methanol and dried under reduced pressure. The crude reaction mixture is then purified using a Sep-Pak C8 column (10 g sorbent) and washed with water to elute the product of interest. TLC was used to confirm which fractions contained the product of interest which were next pooled and incubated with the sodium form of DOWEX 50WX8 resin at 4 °*C* overnight. Resin was removed through filtration, and solution was lyophilized to afford the desired product as its sodium adduct. NMR and MS for pNP-6S-GlcNAc were in agreement with literature values (63). NMR (MU-6S-GalNAc) ^{1}H NMR (600 MHz, D_2O) δ 7.53 (d, *J* = 8.8 Hz, 1H, Ar), 6.97 (dd, *J* = 8.9, 2.4 Hz, 1H, Ar), 6.82 (d, *J* = 2.4 Hz, 1H, Ar), 6.08 (d, *J* = 1.5 Hz, 1H, H-3′), 5.16 (d, *J* = 8.4 Hz, 1H, H-1), 4.33 to 4.19 (m, 5H), 4.10 (d, *J* = 3.3 Hz, 1H, H-4), 3.96 (dd, *J* = 10.9, 3.4 Hz, 1H, H-3), 2.31 (s, 3H, Me), 2.08 (s, 3H, Ac). ^{13}C NMR (151 MHz, D_2O) δ 174.5, 163.7, 159.0, 155.4, 153.0, 126.1, 114.5, 113.1, 110.7, 103.2, 98.8, 72.7, 69.9, 66.9, 66.9, 51.6, 21.8, 17.4.

Isolation of TG product and NMR analysis

A reaction mixture with a total volume of 8160 μl was prepared in buffer A (25 mM Mes pH 7.0, 100 mM NaCl, and 1 mM TCEP) and contained 25 mM pNP-6S-GlcNAc and 15 mM pNPGalNAc with a final DMSO concentration of 20 v/v %. This reaction mixture was incubated with 2.5 μM WT Sp_0475 for 3 h at room temperature. After 3 h, the enzyme was removed with an Amicon Ultra centrifugal filter unit with a MWCO of 10 kDa to avoid substrate hydrolysis, and the flow through was lyophilized. The lyophilized reaction

mixture was then resuspended in 3 ml of 20% ethanol in water (v/v%), filtered, and loaded onto a P2 column equilibrated with 20% ethanol in water (v/v%). Fractions containing TG product were pooled based on TLC (2:1:1 butanol:water:acetic acid R_f:0.44). The pooled fractions were then lyophilized, resuspended in 500 µl water and loaded directly onto a Sep-Pak C8 column (2 g sorbent). TLC was used to determine which fractions contained the product of interest, which were then pooled and lyophilized to give 5 mg of target TG product. HRMS: Calcd for $C_{22}H_{31}N_3O_{16}S$: ([M + H]-): 624.1327 found 624.1349 ^{1}H NMR (400 MHz, D_2O) δ 8.27 (d, *J* = 9.3 Hz, 2H, Ar), 7.27 (d, *J* = 9.3 Hz, 2H, Ar), 5.27 (d, *J* = 8.5 Hz, 1H), 4.62 (d, *J* = 8.5 Hz, 1H), 4.41 (dd, *J* = 11.1, 2.2 Hz, 1H), 4.28 to 4.24 (m,1H), 4.25 (dd, J = 8.5,10.8 Hz, 1H),4.20 (dd, *J* = 11.0, 6.8 Hz, 1H), 4.01 to 3.97 (m, 3H), 3.87 (dd, *J* = 10.8, 3.3 Hz, 1H), 3.78 (dd, *J* = 10.3, 8.4 Hz, 1H), 3.76 to 3.69 (m, 1H), 3.60 to 3.55 (m, 2H), 2.03 (s, 3H,Ac), 1.90 (s, 3H,Ac). ^{13}C NMR (101 MHz, D_2O) δ 175.4, 174.7, 162.2, 142.8, 126.3, 116.8, 102.4, 99.2, 74.3, 74.2, 73.9, 71.0, 70.9, 70.2, 67.9, 67.6, 55.9, 52.3, 22.4, 22.3.

Data availability

The data supporting this study are accessible upon request from the corresponding author.

Supporting information—This article contains supporting information (47, 54, 64).

Author contributions—R. K. B. and S. G. W. conceptualization; R. K. B. and S. G. W. methodology; F. L. and S. A. N. investigation; J. F. W. and P. R. formal analysis; R. K. B. and S. G. W. writing–original draft; R. K. B., S. A. N., F. L., J. F. W., P. R., and S. G. W. writing–review and editing.

Funding and additional information—We thank the Natural Sciences and Engineering Research Council of Canada (Grant 05131) and the Canadian Glycomics Network, GlycoNet for support of this work. R. K. B. was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) doctoral postgraduate scholarship. S. A. N was supported by Vanier Canada graduate scholarship. J. F. W. was supported by Canadian Institutes for Health Research (CIHR) Canada graduate scholarship.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 6S-GlcNAc, 6-sulfo-N-acetylglucosamine; BT, *Bacteroides thetaiotaomicron*; CV, column volume; DMSO, dimethylsulfoxide; ENGases, endo-β-N-acetylglucosaminidase; GalNAc, N-acetylgalactosamine; MUGlcNAc, 4-methylumbelliferyl N-acetyl-β-D-glucosaminide; MS, mass spectrometry; MWCO, molecular weight cutoff; PDB, Protein Data Bank; pNPGlcNAc, 4-nitrophenyl N-acetyl-β-D-glucosaminide; SSN, sequence similarity network; TCEP, tris(2-carboxyethyl) phosphine; TG, transglycosylation.

References

1. Crouch, L. I. (2023) N-glycan breakdown by bacterial CAZymes. *Essays Biochem.* **67**, 373–385
2. Raba, G., and Luis, A. S. (2023) Mucin utilization by gut microbiota: recent advances on characterization of key enzymes. *Essays Biochem.* **67**, 345–353
3. Muthana, S. M., Campbell, C. T., and Gildersleeve, J. C. (2012) Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chem. Biol.* **7**, 31–43
4. Xiao, H., Suttapitugsakul, S., Sun, F., and Wu, R. (2018) Mass spectrometry-based chemical and enzymatic methods for global analysis of protein glycosylation. *Acc. Chem. Res.* **51**, 1796–1806
5. Tang, F., Zhou, M., Qin, K., Shi, W., Yashinov, A., Yang, Y., *et al.* (2020) Selective N-glycan editing on living cell surfaces to probe glycoconjugate function. *Nat. Chem. Biol.* **16**, 766–775
6. Wardman, J. F., Rahfeld, P., Liu, F., Morgan-Lang, C., Sim, L., Hallam, S. J., *et al.* (2021) Discovery and development of promiscuous O-glycan hydrolases for removal of intact sialyl T-antigen. *ACS Chem. Biol.* **16**, 2004–2015
7. Trastoy, B., Naegeli, A., Anso, I., Sjögren, J., and Guerin, M. E. (2020) Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from Akkermansia muciniphila. *Nat. Commun.* **11**, 1–14
8. Noach, I., and Boraston, A. B. (2021) Structural evidence for a proline-specific glycopeptide recognition domain in an O-glycopeptidase. *Glycobiology* **31**, 385–390
9. Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., *et al.* (2019) The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci.* **116**, 7278–7287
10. Kobata, A. (2013) Exo- and endoglycosidases revisited. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **89**, 97–117
11. Rudd, P. M., and Dwek, R. A. (1997) Rapid, sensitive sequencing of oligosaccharides from glycoproteins. *Curr. Opin. Biotechnol.* **8**, 488–497
12. Chuzel, L., Fossa, S. L., Boisvert, M. L., Cajic, S., Hennig, R., Ganatra, M. B., *et al.* (2021) Combining functional metagenomics and glycoanalytics to identify enzymes that facilitate structural characterization of sulfated N-glycans. *Microb. Cell Fact.* **20**, 1–17
13. Wardman, J. F., Bains, R. K., Rahfeld, P., and Withers, S. G. (2022) Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nat. Rev. Microbiol.* **20**, 542–556
14. Rahfeld, P., Wardman, J. F., Mehr, K., Huff, D., Morgan-Lang, C., Chen, H. M., *et al.* (2019) Prospecting for microbial α-N-acetylgalactosaminidases yields a new class of GH31 O-glycanase. *J. Biol. Chem.* **294**, 16400
15. Rahfeld, P., Sim, L., Moon, H., Constantinescu, I., Morgan-Lang, C., Hallam, S. J., *et al.* (2019) An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **4**, 1475–1485
16. Crouch, L. I., Liberato, M. V., Urbanowicz, P. A., Baslé, A., Lamb, C. A., Stewart, C. J., *et al.* (2020) Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nat. Commun.* **11**, 4017
17. Luis, A. S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., *et al.* (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* **3**, 210–219
18. Cartmell, A., Lowe, E. C., Baslé, A., Firbank, S. J., Ndeh, D. A., Murray, H., *et al.* (2017) How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7037–7042
19. Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., *et al.* (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169
20. Tailford, L. E., Crost, E. H., Kavanaugh, D., and Juge, N. (2015) Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 1–18
21. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol.* **199**, e00860–e00916
22. Luis, A. S., Jin, C., Pereira, G. V., Glowacki, R. W. P., Gugel, S. R., Singh, S., *et al.* (2021) A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332–337
23. Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., *et al.* (2019) Complex N-glycan breakdown by gut Bacteroides

involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat. Microbiol.* **4**, 1571–1581
24. Robb, M., Hobbs, J. K., Woodiga, S. A., Shapiro-Ward, S., Suits, M. D. L., McGregor, N., *et al.* (2017) Molecular characterization of N-glycan degradation and transport in Streptococcus pneumoniae and its contribution to virulence. *PLOS Pathog.* **13**, 1–33
25. Cao, Y., Rocha, E. R., and Smith, C. J. (2014) Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12901–12906
26. Cordeiro, R. L., Santos, C. R., Domingues, M. N., Lima, T. B., Pirolla, R. A. S., Morais, M. A. B., *et al.* (2022) Mechanism of high-mannose N-glycan breakdown and metabolism by Bifidobacterium longum. *Nat. Chem. Biol.* **19**, 218–229
27. Trastoy, B., Du, J. J., Klontz, E. H., Li, C., Cifuente, J. O., Wang, L. X., *et al.* (2020) Structural basis of mammalian high-mannose N-glycan processing by human gut Bacteroides. *Nat. Commun.* **11**, 1–11
28. Luis, A. S., Baslé, A., Byrne, D. P., Wright, G. S. A., London, J. A., Jin, C., *et al.* (2022) Sulfated glycan recognition by carbohydrate sulfatases of the human gut microbiota. *Nat. Chem. Biol.* **18**, 841–849
29. Armstrong, Z., Rahfeld, P., and Withers, S. G. (2017) Discovery of new glycosidases from metagenomic libraries. *Methods Enzymol.* **597**, 3–23
30. Higgins, M. A., Tegl, G., Macdonald, S. S., Arnal, G., Brumer, H., Withers, S. G., *et al.* (2021) N-glycan degradation pathways in gut- gnd soil-dwelling Actinobacteria share common core genes. *ACS Chem. Biol.* **16**, 701–711
31. Fairbanks, A. J. (2017) The ENGases: versatile biocatalysts for the production of homogeneous: N -linked glycopeptides and glycoproteins. *Chem. Soc. Rev.* **46**, 5128–5146
32. Fairbanks, A. J. (2019) Chemoenzymatic synthesis of glycoproteins. *Curr. Opin. Chem. Biol.* **53**, 9–15
33. Zallot, R., Oberg, N., and Gerlt, J. A. (2019) The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182
34. Alagesan, K., and Kolarich, D. (2019) Improved strategy for large scale isolation of sialylglycopeptide (SGP) from egg yolk powder. *MethodsX* **6**, 773–778
35. Liu, L., Prudden, A. R., Bosman, G. P., and Boons, G. J. (2017) Improved isolation and characterization procedure of sialylglycopeptide from egg yolk powder. *Carbohydr. Res.* **452**, 122–128
36. Tegl, G., Rahfeld, P., Ostmann, K., Hanson, J., and Withers, S. G. (2021) Discovery of β-N-acetylglucosaminidases from screening metagenomic libraries and their use as thioglycoligase mutants. *Org. Biomol. Chem.* **19**, 9068–9075
37. Kanyo, Z. F., and Christianson, D. W. (1991) Biological recognition of phosphate and sulfate. *J. Biol. Chem.* **266**, 4264–4268
38. Luecke, H., and Quiocho, F. A. (1990) High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature* **347**, 402–406
39. Thatcher, G. R. J., Cameron, D. R., Nagelkerke, R., and Schmitke, J. (1992) Selective hydrogen bonding as a mechanism for differentiation of sulfate and phosphate at biomolecular receptor sites. *J. Chem. Soc. Chem. Commun.* https://doi.org/10.1039/C39920000386
40. Mark, B. L., Vocadlo, D. J., Knapp, S., Triggs-Raine, B. L., Withers, S. G., and James, M. N. G. (2001) Crystallographic evidence for substrate-assisted catalysis in a bacterial β-hexosaminidase. *J. Biol. Chem.* **276**, 10330–10337
41. Knapp, S., Vocadlo, D., Gao, Z., Kirk, B., Lou, J., and Withers, S. G. (1996) NAG-thiazoline, an N-acetyl-β-hexosaminidase inhibitor that implicates acetamido participation. *J. Am. Chem. Soc.* **118**, 6804–6805
42. Liu, J., Shikhman, A. R., Lotz, M. K., and Wong, C. H. (2001) Hexosaminidase inhibitors as new drug candidates for the therapy of osteoarthritis. *Chem. Biol.* **8**, 701–711
43. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
44. Naumoff, D. G. (2011) Hierarchical classification of glycoside hydrolases. *Biochemistry* **76**, 622–635
45. Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215
46. Rogowski, A., Baslé, A., Farinas, C. S., Solovyova, A., Mortimer, J. C., Dupree, P., *et al.* (2014) Evidence that GH115 α-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. *J. Biol. Chem.* **289**, 53–64
47. Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J. H., Czjzek, M., Smith, S. P., *et al.* (2009) Portrait of an enzyme, a complete structural analysis of a multimodular β-N-acetylglucosaminidase from Clostridium perfringens. *J. Biol. Chem.* **284**, 9876–9884
48. MacLeod, A. M., Lindhorst, T., Withers, S. G., and Warren, R. A. J. (1994) The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* **33**, 6371–6376
49. Vocadlo, D. J., Mayer, C., He, S., and Withers, S. G. (2000) Mechanism of action and identification of Asp242 as the catalytic nucleophile of Vibrio furnisii N-acetyl-β-D-glucosaminidase using 2- acetamido-2-deoxy-5-fluoro-α-L-idopyranosyl fluoride. *Biochemistry* **39**, 117–126
50. Abbott, D. W., Macauley, M. S., Vocadlo, D. J., and Boraston, A. B. (2009) Streptococcus pneumoniae endohexosaminidase D,structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. *J. Biol. Chem.* **284**, 11676–11689
51. Cuxart, I., Coines, J., Esquivias, O., Faijes, M., Planas, A., Biarnés, X., *et al.* (2022) Enzymatic hydrolysis of human milk oligosaccharides. The molecular mechanism of Bifidobacterium bifidum lacto-N-biosidase. *ACS Catal.* **12**, 4737–4743
52. Slámová, K., Krejzová, J., Marhol, P., Kalachova, L., Kulik, N., Pelantová, H., *et al.* (2015) Synthesis of derivatized chitooligomers using transglycosidases engineered from the fungal GH20 β-N-acetylhexosaminidase. *Adv. Synth. Catal.* **357**, 1941–1950
53. Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., *et al.* (2013) Crystal structures of a glycoside hydrolase family 20 lacto-N-biosidase from Bifidobacterium bifidum. *J. Biol. Chem.* **288**, 11795–11806
54. Zhang, Z., Dong, M., Zallot, R., Blackburn, G. M., Wang, N., Wang, C., *et al.* (2023) Mechanistic and structural insights into the specificity and biological functions of bacterial sulfoglycosidases. *ACS Catal.* **13**, 824–836
55. Katoh, T., Yamada, C., Wallace, M. D., Yoshida, A., Gotoh, A., Arai, M., *et al.* (2023) A bacterial sulfoglycosidase highlights mucin O-glycan breakdown in the gut ecosystem. *Nat. Chem. Biol.* **19**, 778–789
56. Caterson, B., and Melrose, J. (2018) Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* **28**, 182–206
57. Hobbs, J. K., Pluvinage, B., and Boraston, A. B. (2018) Glycan-metabolizing enzymes in microbe–host interactions: the Streptococcus pneumoniae paradigm. *FEBS Lett.* **592**, 3865–3897
58. Huang, K., Li, C., Zong, G., Prabhu, S. K., Chapla, D. G., Moremen, K. W., *et al.* (2022) Site-selective sulfation of N-glycans by human GlcNAc-6-O-sulfotransferase 1 (CHST2) and chemoenzymatic synthesis of sulfated antibody glycoforms. *Bioorg. Chem.* **128**, 106070
59. Burkart, M. D., Izumi, M., Chapman, E., Lin, C. H., and Wong, C. H. (2000) Regeneration of PAPS for the enzymatic synthesis of sulfated oligosaccharides. *J. Org. Chem.* **65**, 5565–5574
60. Zhang, H., Liu, C. J., Jiang, H., Zhou, L., Li, W. Y., Zhu, L. Y., *et al.* (2017) Mega primer-mediated molecular cloning strategy for chimaeragenesis and long DNA fragment insertion. *Biosci. Rep.* **37**, 1–7
61. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234
62. Thompson, J., Lichtenthaler, F. W., Peters, S., and Pikis, A. (2002) β-Glucoside Kinase (BglK) from Klebsiella pneumoniae: purification, properties, and preparative synthesis of 6-phospho-β-D-glucosides. *J. Biol. Chem.* **277**, 34310–34321
63. Fukuda, T., Matsumoto, E., Onogi, S., and Miura, Y. (2010) Aggregation of alzheimer amyloid peptide (1-42) on the multivalent sulfonated sugar interface. *Bioconjug. Chem.* **21**, 1079–1086
64. Sievers, F., and Higgins, D. G. (2021) The clustal omega multiple alignment package. *Methods Mol. Biol.* **2231**, 3–16
65. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190