Published in final edited form as:

Biochemistry. 2009 November 3; 48(43): 10334–10341. doi:10.1021/bi9013825.

Mechanistic investigation of the endo- α -N-acetylgalactosaminidase from *Streptococcus pneumoniae* R6

Lisa M. Willis^{‡, *}, Ran Zhang^{¶, *}, Anne Reid[‡], Stephen G. Withers[¶], and Warren W. Wakarchuk[‡]

[‡] National Research Council of Canada, Institute for Biological Sciences, Glycobiology Program, 100 Sussex Drive, Ottawa, ON., Canada K1A 0R6

[¶] Departments of Chemistry and Biochemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

Abstract

The large (1767 amino acid) endo- α -N-acetylgalactosaminidase from *Streptococcus pneumoniae* (SpGH101) specifically removes an O-linked disaccharide Gal- β -1,3-GalNAc- α from glycoproteins. While the enzyme from natural sources has been used as a reagent for many years, very few mechanistic studies have been performed. Using the recently solved 3-dimensional structure of the recombinant protein as a background, we report here a mechanistic investigation of the SpGH101 retaining α -glycoside hydrolase using a combination of synthetic and natural substrates. Based on a model of the substrate complex of SpGH101 we propose D764 and E796 as the nucleophile and general acid/base residues respectively. These roles were confirmed by kinetic and mechanistic analysis of mutants at those positions using synthetic substrates and anion rescue experiments. pKa values of 5.3 and 7.2 were assigned to D764 and E796 on the basis of the pKa values derived from the bell-shaped dependence of k_{cat}/K_m upon pH. The enzyme contains several putative carbohydrate binding modules whose glycan binding specificities were probed using the printed glycan array of the Consortium for Functional Glycomics using the inactive D764A and D764F mutants that had been labeled with Alexfluor 488. These studies revealed binding to galacto-N-biose, consistent with a role for these domains in localizing the enzyme near its substrates.

Keywords

Glycosidase; retaining mechanism; chemical rescue; *S. pneumoniae*; galacto-N-biose

S. pneumoniae is a gram-positive organism that colonizes the respiratory tract of humans as a commensal organism but can become a pathogen that causes pneumonia, meningitis, bacteraemia and otitis media with significant morbidity and mortality, especially in children (1). Because *S. pneumoniae* is considered a leading cause of death worldwide, a great deal of research has been directed to determining the factors that contribute to the virulence of this organism. Of major importance to this pathogen is the ability to degrade host glycoproteins, and to metabolize the resultant carbohydrates (2). Consistent with this, *S. pneumoniae* has a large number of glycoside hydrolases, with 21 CAZY glycoside hydrolase families represented by at least one gene in the sequenced reference strains (3). These hydrolases are both secreted and cell-associated. Indeed the genome sequences of the reference strain *S. pneumoniae* R6

Address correspondence to: Dr. Warren Wakarchuk, National Research Council of Canada, Institute for Biological Sciences, Glycobiology Program, 100 Sussex Drive, Ottawa, ON., Canada K1A 0R6. warren.wakarchuk@nrc-cnrc.gc.ca Telephone: 613-952-4299, FAX: 613-941-1327.

*These authors contributed equally for this work

and of a clinical isolate contain at least 6 cell surface-localized enzymes that are thought to play a role in virulence. (4) (5) (6). These cell-associated enzymes include neuraminidases, hyaluronidases, β -galactosidases, and endo- β -*N*-acetylglucosaminidases, which have been studied in some detail (reviewed in (7)). A new member of this cell surface family of enzymes belonging to the GH101 family of glycoside hydrolases, endo- β -*N*-acetylgalactosaminidase (SpGH101), specifically cleaves Gal- β -1,3-GalNAc- α -Ser/Thr (T-antigen, galacto-*N*-biose), which is the core 1 type *O*-linked glycan common to mucin glycoproteins. This CAZY GH101 family (3) currently contains proteins from 12 species of bacteria, most of which are commensal human bacteria, though some may also be human pathogens. The *S. pneumoniae* enzyme has been shown recently to be non-essential, but is involved in aiding adherence of *Streptococcus* to human airway epithelial cells (8). The substrate specificity of this enzyme, SpGH101, has been studied in some detail with material purified from *Streptococcus* culture supernatants (9)(10). A related enzyme from *Bifidobacterium longum* was expressed as a recombinant protein, shown to be a retaining glycoside hydrolase, and candidate catalytic residues suggested through kinetic analyses of mutants of conserved carboxylic amino acids (Glu and Asp) (11). However the catalytic machinery was not positively identified. The recently solved 3D structure of SpGH101 (12), in conjunction with sequence alignments, reveals that the GH101 enzymes are multi-modular proteins with a catalytic domain as well as domains that resemble carbohydrate binding modules (CBMs). These include a CBM32 module, along with degenerate CBM4 and CBM16-like folds, plus modules that resemble a degenerate legume lectin fold. The catalytic domain is a distorted (β/α)₈ barrel flanked by a domain of all β -sheet structure: very analogous to the arrangement seen in GH13 α -amylases. The flanking domain likely serves to shield the active site from water during catalysis (13).

Sequence similarity between these GH101 enzymes is highest in the catalytic domain, with many of the absolutely conserved residues being in the active site pocket of this GH domain (Figure 1). Based on the active site structure of the α -amylase from *Thermoactinomyces vulgaris* R-47 complexed with substrate, we modeled a substrate complex for SpGH101 (12). Earlier kinetic analyses of mutants made at the sites of conserved Glu and Asp residues of the BlGH101 enzyme had suggested two conserved residues to be important for enzyme activity, BlGH101 D682 and BlGH101 D789, which correspond to SpGH101 D658 and SpGH101 D764. However inspection of the active site geometry of these acidic residues in the SpGH101 structure with the substrate modeled into place, based on the α -amylase structure, suggested that the likely catalytic residues are instead D764 (nucleophile) and E796 (acid/base). More recently the three-dimensional structure of BlGH101 has been solved and mutation of E822 (E796 in SpGH101) confirms this as an important catalytic residue, though no detailed studies were performed, simply assays (14).

There is considerable interest in enzymes of this type, both as reagents for removal of *O*-linked glycans and potentially also for de-novo glycosylation of proteins when used in transglycosylation mode, possibly in engineered versions. However, only limited functional information on enzymes from this family is available. In order to correct this deficiency we have performed a detailed mechanistic characterization of this enzyme to identify the catalytic residues. We have also probed the binding specificities of the putative carbohydrate binding modules by use of a catalytically impaired mutant to examine binding to a carbohydrate microarray containing > 400 carbohydrate structures, many of which would be found in a human host.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Plasmids, and Growth Conditions

Streptococcus pneumoniae R6 was obtained from the ATCC (ATCC # BAA-255) and was grown on Columbia Blood Agar at 37 °C and 5 % CO₂. *E. coli* AD202 (CGSC # 7297 or Ref

3) was used for expression of cloned enzymes. Recombinant *E. coli* strains were grown in 2YT broth with 150 µg/mL ampicillin at 37 °C for 1.5 hours then induced with 0.5 µM isopropyl 1-thio-β-D-galactopyranoside and grown at 30 °C for 7 hours. The plasmid used for cloning was a modified pCWori+ (15).

Basic Recombinant DNA Method

All DNA isolations, restriction enzyme digestions, ligations, and transformations were performed as recommended by the supplier. Enzymes were obtained from New England Biolabs (Mississauga, ON). Genomic DNA was isolated using a DNeasy Tissue kit (Qiagen Inc., Mississauga, Ontario). PCR was performed using Phusion polymerase and the program: 94 °C for 5 min, 30 cycles of 94 °C for 30 sec, 55 °C for 30 sec and 72 °C for 120 sec, and finally 72 °C for 10 min. The gene was originally amplified from genomic DNA based on the sequence Sp0328 from *S. pneumoniae* R6 with primers to remove a 40 amino acid signal sequence at the N-terminal end of the gene. The primers for this amplification and other sequence truncations are shown in Table 1.

DNA was purified using either Qiaquick or Minelute kits from Qiagen Inc (Mississauga, Ontario). Genes digested with *NdeI* and *SalI* were ligated into pCW and then used to transform *E. coli* AD202 by electroporation. Plasmids were isolated using a High Pure Plasmid Isolation kit (Roche Diagnostics, Laval, Quebec). DNA sequencing was performed using an Applied Biosystems (Montreal, Quebec) model 3100 automated DNA sequencer and the manufacturer's cycle sequencing kit.

Site-directed mutagenesis

Mutations were introduced using the pull-through PCR method described previously (16). The primers for construction of D658A, D764A/F, and E796A/Q are given in Table 1.

Enzyme preparation

Recombinant cultures grown in the presence of IPTG were lysed with an Avestin C5 Emulsiflex cell disrupter (Avestin Ottawa, Ontario) and centrifuged first at 27,000 x g to pellet unbroken cells, and then at 100,000 x g to pellet insoluble material. SpGH101 was partially purified using DEAE-Sepharose chromatography (GE Healthcare, Piscataway, NJ) with a gradient of 20 mM – 1 M NH₄OAc pH 7.5. The enzyme was concentrated using Amicon Ultra-15 filtration devices (Millipore, Billerica, MA) and then applied to a Superdex-200 size exclusion column in 10 mM NH₄OAc pH 7.5. The enzyme in the pooled fractions was finally applied to a Phenyl-Sepharose column (GE Healthcare, Piscataway, NJ) with a gradient of 1.5 M – 10 mM NH₄OAc, pH 7.5. Purity was assessed by SDS-PAGE analysis and protein concentrations were determined using a BCA assay (Pierce Biotechnology Inc., Rockford, IL).

Synthesis of Gal-β-1,3-GalNAc-α-2,4-dinitrophenyl (DNP-Tag) and FCHASE-[Gal-β-1,3-GalNAc-α-] [VGVETEP]

DNP-GalNAc (the α anomer) was synthesized as previously described (17). Gal-β-1,3-GalNAc-α-DNP (DNP-Tag) was synthesized enzymatically from GalNAc-α-DNP using the *Campylobacter jejuni* CgtB galactosyltransferase as described previously (18). The disaccharide product was purified by chromatography on a Superdex Peptide 30 column (GE Healthcare, 1.5 × 80 cm) which had been equilibrated in 100 mM NH₄HCO₃. The peptide substrate FCHASE-[T-Ag]-IFNα2b was also synthesized as described previously (18). Purity of the products was determined by capillary electrophoresis on a PACE MDQ system as previously described (19).

Modeling of T-Ag in the active site

A PDB file for the T-antigen disaccharide ligand was made using the SWEET modeling program (20). The ligand was fitted to the SpGH101 PDB entry 3ECQ structure using the PatchDock server (21) and figures were generated using PyMOL v1.1 (22).

Measurement of O-glycanase activity

Activity was measured using the substrates Gal- β -1,3-GalNAc- α -pNP (pNP-TAg), obtained from Toronto Research Chemicals), Gal- β -1,3-GalNAc- α -DNP (DNP-TAg), or DNP-GalNAc. Assays were performed at 37 °C in 50 mM citrate-phosphate buffer pH 6.5 with 0.1 mg/mL acetylated bovine serum albumin. Assays with pNP-TAg were performed in stopped mode with an equal volume of 0.2 M sodium carbonate, 0.5% sodium dodecylsulphate and measured at 400 nm using an extinction coefficient of 14,000 cm⁻¹M⁻¹. Assays with the DNP-TAg and DNP-GalNAc were performed in continuous mode. For wild type SpGH101 enzyme kinetics, 10 μ M – 200 μ M of DNP-TAg was incubated followed by the addition of 0.2 nM enzyme. For SpGH101 mutants, 1 μ M – 10 μ M of DNP-TAg was incubated followed by the addition of 2 nM enzyme. Release of the dinitrophenolate ion was followed at 400 nm using a Varian Cary AV-4000 and rates calculated using an extinction coefficient 10,800 cm⁻¹M⁻¹. Kinetic data were analysed using Grafit 5.0 software. For the chemical rescue experiments, 10 mM – 1 M sodium acetate, sodium formate, sodium azide, sodium acetate or potassium fluoride was added and the pH was checked to ensure it remained at 6.5.

Azide rescue product identification

A reaction mixture containing 3.5 mM DNP-TAg, 1 M sodium azide and 5 nM of SpGH101 E796Q mutant was prepared and left at room temperature overnight. Product analysis was performed by thin layer chromatography on 60 F₂₅₄ silica gel aluminum plates (Merck) run in 7:2:1 (v/v/v) ethyl acetate/methanol/water and developed with 10% ammonium molybdate in 2 M H₂SO₄ followed by charring. Purification of the final product was performed by directly loading the crude reaction mixture onto a flash column chromatography with Silicagel 60. The NMR spectra were recorded using the Bruker AV-300 MHz spectrometer. High resolution mass spectral data were collected by the mass spectrometry laboratory at the University of British Columbia.

pH profile

The pH dependence of k_{cat}/K_M for SpGH101 was measured using the substrate depletion method at low substrate concentrations ($[S] \ll K_M$, (23)). Assays were performed with 50 μ M DNP-GalNAc or 5 μ M DNP-TAg and the following buffers: 50 mM citrate-phosphate pH 4.4 – 7.0, 50 mM sodium phosphate pH 7.0 – 8.0, 50 mM sodium borate pH 8.0 – 9.0. Absorbance was measured using a Varian Cary AV-4000 and data fitted to a first order curve using Grafit 5.0 software. The k_{cat}/K_M values were obtained from these fits by dividing by the enzyme concentration.

Glycan array analysis

SpGH101 D764A, and D764F (1 mg) were labeled with Alexafluor 488 according to the instructions supplied with the dye (Invitrogen). The labeled protein was purified from the unreacted dye by chromatography on Sephadex G-25 equilibrated in PBS buffer pH 7.4. The labeled protein was analyzed for binding to the printed glycan array from the Consortium for Functional Glycomics version 3.1(D764A) or 4.0 (D764F), by the CORE H facility of the CFG. The protocols are available online at <http://www.functionalglycomics.org/glycomics>. Protocols cfgPTC_243, and cfgPTC_248 were used to analyze SpGH101 on the array. Binding was examined on arrays with both 100 μ M and 10 μ M glycan spot density for array version 3.1 with D764A, and at 2 protein concentrations of D764F for the version 4 array.

RESULTS

Cloning, expression and purification

The gene from *S. pneumoniae* R6 was amplified from genomic DNA using primers designed to eliminate the signal sequence by starting at amino acid 40 (Figure 2). Attempts were made to prepare shorter versions by truncating the protein 200 amino acids from either the N- or C-terminal ends. The 200 residue N-terminal truncation had only 6.6% of the specific activity of the full length protein (data not shown), and therefore was not examined further. The 200 residue C-terminal truncation, SpGH101 aa40–1567, however, maintained the specific activity of the full length enzyme, but as we were interested in the C-terminal CBM-like domain we chose to work with the SpGH101 aa 40–1767. The enzyme was purified to greater than 90% purity in a three step purification procedure with an 80-fold enrichment in specific activity. Catalytic site mutants were grown and purified as described for the WT enzyme. No significant differences in expression or purification were observed with these mutants.

Kinetic and stereochemical analysis of wild-type and active site mutants

Wildtype—Kinetic parameters for both DNP-GalNAc and DNP-TAg were measured and are shown in Table 2. The specificity of the enzyme for the T-Ag disaccharide was shown by the 26,500-fold greater k_{cat}/K_m for the disaccharide substrate than the monosaccharide. This large difference in specificity is due not only to an increase in affinity, as seen by the 70-fold decrease in K_m , but also to a 400-fold increase in k_{cat} . DNP-TAg is therefore a vastly superior substrate for use in mechanistic studies. The 2,4-dinitrophenyl glycoside was chosen for the majority of the kinetic analyses since the greater leaving group ability of DNP makes this a better substrate than the *p*-nitrophenyl analog. In addition the low pKa (4.0) of DNP ensures that the extinction coefficient of 2,4-DNP does not change significantly with pH above pH 5 and that assay is much more sensitive than with the *p*-nitrophenyl substrate at pH 6.5, which is the optimal pH for the -SpGH101. The enzyme was also shown by NMR analysis of the reaction progress to act with retention of anomeric stereochemistry (data not shown) similar to the *B. longum* enzyme BIGH101(11).

pH dependence—The bell shaped curve obtained from the plot of k_{cat}/K_M versus pH indicated that two ionizable groups are involved in catalysis with pKa values of 5.3 ± 0.1 and 7.8 ± 0.1 . (Figure 3).

Kinetic analysis of catalytic site mutants

Aryl glycoside substrates—Based on the active site substrate complex model the mutants D658A, D764A, and E796A/Q were constructed and analysed using the disaccharide DNP-TAg substrate. The least active of these mutants is D764A, with a k_{cat}/K_m value some 700 times lower than that of the wild type enzyme. In fact this is likely a maximum estimate of the activity since, based upon the similarity of its K_m value to that of the wild type enzyme, it is quite probable that even this low activity derives from a small amount of contaminating wild type enzyme. The next most active is the D658A mutant, with a k_{cat}/K_m that is 16-fold lower than that of the wild type enzyme. However it did not show saturation kinetics with the DNP-TAg substrate, thus individual kinetic constants k_{cat} and K_m , could not be determined for this mutant. The most active were the E796A and Q mutants, with k_{cat}/K_m values essentially the same as the wild type enzyme. Importantly however the k_{cat} values are about 30 fold lower than that of the wild type enzyme, with the K_m values also correspondingly lowered.

The monosaccharide substrate DNP-GalNAc was cleaved only by the WT enzyme and the E796A mutant relatively efficiently, but at an extremely low rate for the E796Q mutant. Notably the k_{cat} value for this substrate with E796A was 3 fold higher than for the WT enzyme, though the K_m values were similar. However the k_{cat} value was some 300–400 - fold lower

than that of the disaccharide substrate, and the K_m value some 80-fold higher. These high K_m values are suggestive of glycosylation step being rate-limiting. Interestingly a parallel situation was seen with the xylanase Cex from *Cellulomonas fimi*, where the “addition” of a glucose residue to a monosaccharide substrate was seen to change the rate-limiting step from glycosylation to deglycosylation, and a rationale for this was developed. (24) (25)

‘Natural’ glycopeptide substrates—In order to investigate the enzymatic activity of the active site mutants on a more natural substrate, a synthetic glycopeptide, FCHASE-[TAG]-IFN α 2b was prepared as previously described (18). This substrate contains a T-Antigen disaccharide attached to the threonine side chain of a seven-amino-acid peptide via an α -glycosidic linkage, with a fluorescent label on the N-terminus. Upon the addition of 34 nM of WT SpGH101, 0.5 mM of FCHASE-[TAG]-IFN α 2b was rapidly cleaved, with hydrolysis being complete within three hours at room temperature, (data not shown). However, the addition of either 290 nM of E796A or 220 nM of E796Q to 0.5 mM of FCHASE-[TAG]-IFN α 2b, resulted in no observable cleavage, even after incubation for two days.

Chemical rescue of catalytic residue mutants—To investigate the roles of D764 and E796 in catalysis, assays were performed in the presence of the exogenous anionic nucleophiles (Table 3). In the case of the proposed acid/base (E796) mutants, both azide and formate increased the steady state activity in a dose-dependent manner, with the greatest increase being seen with azide and the E796Q mutant. Azide rescue was also observed with the E796A mutant, but to a lesser extent. The only anion that restored activity to the nucleophile mutant (D764A), was azide, which has the highest nucleophilicity of those tested (azide, formate, acetate and fluoride). In order to rule out the possibility of these changes in k_{cat} and K_m simply resulting from salt effects, the kinetic parameters of the E796 mutants in the presence of either 100 mM or 500 mM NaCl were measured. However, in none of the cases were the K_m values of the mutants affected by the addition of Cl^- . Anion rescue experiments with the E796A/Q mutants were also attempted using the monosaccharide substrate DNP-GalNAc, however, no anion rescue was observed, consistent with the conclusion that the glycosylation step is rate-limiting for the monosaccharide substrates.

Azide rescue product identification—TLC analysis of the reaction mixture of DNP-TAG, sodium azide and SpGH101 E796Q mutant revealed a new non-UV active spot ($R_f = 0.32$), which is distinct from DNP-TAG ($R_f = 0.49$) and TAG ($R_f = 0.14$). After purification, high resolution mass spectra confirmed that the new product has the molecular formula for the TAG-azide. (Calcd. for $C_{14}H_{24}N_4O_{10}+Na^+$: 431.1390, Found: 431.1393). The α configuration of the anomeric azide was clearly demonstrated by the small $J_{1,2}$ coupling constant of 4.2 Hz from the 1H -NMR of the purified product. The relatively big chemical shift ($\delta = 5.52$ ppm) is also very typical for α -glycosyl azide since a much lower chemical shift would be expected for a β glycosyl azide. 1H -NMR (300 MHz, CD_3OD) δ : 5.52 (1H, d, $J_{1,2} = 4.2$, H-1), 4.45 (1H, dd, $J_{2,3} = 11.2$ Hz, $J_{1,2} = 4.2$ Hz, H-2), 4.41 (1H, d, $J_{1',2'} = 7.5$ Hz, H-1'), 4.20 – 3.43 (m, 11H), 1.89 (3H, s, $COCH_3$).

Glycan array experiments—In order to probe the carbohydrate binding properties of the SpGH101 enzyme we used the nucleophile mutant as probe on a glycan microarray. The strategy was to use two versions of this mutation, D764A in which the enzyme could potentially bind through the active site, but would not hydrolyze glycans on the array, and D764F which would block binding through the active site by the presence of the bulky Phe residue. In the latter mutant only CBM type interactions should mediate binding to the glycan array. The binding profile revealed strong binding to the disaccharide that is the natural substrate, as well as some binding to the 3-*O*-sulphated version of this disaccharide (Figure 4A). There were several other weak interactions which disappear when the glycan array contained only low

density carbohydrates (data not shown). The binding profile with the D764F mutant was essentially the same although the total binding strength appeared to be lower.

DISCUSSION

The kinetic and mechanistic data accumulated provide very strong evidence that D764 acts as the catalytic nucleophile ($pK_a = 5.3$) and that E796 functions as the acid/base catalyst ($pK_a = 7.2$) in the double-displacement mechanism followed by SpGH101 (Fig. 5). These are the roles that were proposed on the basis of modeling of substrate binding into the active site using the structure of an α -amylase/substrate complex as a guide. Strong support for the assignment of D764 as the catalytic nucleophile comes from the observation that mutation to alanine results in essentially complete abrogation of enzyme activity. Even with the most active 2,4-dinitrophenyl T-antigen substrate the activity was some 700-fold lower than that of the wild type enzyme. Indeed, even this activity was most probably due to contaminating wild type enzyme. Such essentially complete loss of activity has been seen in essentially all retaining glycosidases when their catalytic nucleophiles have been mutated (26)

Evidence in support of a role for E796 as the acid/base catalyst is particularly strong. The mutants E796A and Q efficiently cleave substrates containing the highly reactive dinitrophenyl leaving group, which does not need acid catalytic assistance, with k_{cat}/K_m values essentially identical to that of the wild type enzyme. By contrast, cleavage of natural, non-activated glycopeptide substrates, which do need acid catalytic assistance, was severely compromised. Indeed no cleavage whatsoever was observed using sensitive fluorescent substrates, even after incubations with almost 10 times as much enzyme and incubating for two days. This contrasting behavior with the two different substrates is what has been seen for other retaining glycosidases in which the acid catalyst is mutated, and is consistent entirely with the role. (27)(28)

A more detailed inspection of the kinetic data for the dinitrophenyl glycoside also reveals very low K_m values for the acid mutants. This is not a consequence of improved affinity *per se*, but rather indicates the accumulation of the glycosyl enzyme intermediate. This arises because the glycosylation step remains fast (as shown by k_{cat}/K_m values) as a consequence of the excellent leaving group ability of the dinitrophenyl group, but the deglycosylation step is slowed (as seen in k_{cat}) due to the removal of general base catalysis.

Further, very strong evidence for E796 as the acid/base catalyst derives from the rescue of steady state activity seen in the presence of the anionic nucleophiles azide and formate. Increases in k_{cat} of up to 5 fold were observed, with no effect upon k_{cat}/K_m . This arises because azide is a much better nucleophile than water, but cannot attack the glycosyl-enzyme formed on the wild-type enzyme due to electrostatic screening from the deprotonated E796. Removal of that charge in the alanine mutant allows direct attack of azide or formate at the anomeric centre, with associated increases in steady state rate. Parallel studies with sodium chloride confirm that the rate increases are not due to salt effects.

The relatively small increases in steady state rate observed in these anion rescue experiments are consistent with findings on other α -glycosidases (29) (30) where increases of 5–10 fold have been seen. This contrasts with the much larger increases typically seen for catalytic acid mutants of β -glycosidases. (27)(31). The ‘saturation’ behaviour seen in such experiments (Fig. 6) is not a reflection of saturable reversible binding of the anion, but rather is a consequence of a change in rate-limiting step at higher anion concentrations. The smaller increases seen for the α -glycosidases compared to those seen for the β -glycosidases are, then, a reflection of the inherent relative rates of glycosylation and deglycosylation steps in α - and β -glycosidases. This most likely arises primarily from the somewhat greater reactivity of β -glycosidases than their α -anomers. Interestingly, no anion rescue was observed when DNP GalNAc was used as

substrate, consistent with the indications (from high K_m values) that the glycosylation step, rather than the deglycosylation step, is rate-limiting in that case. Further confirmation of the mechanism shown in Figure 6 is provided by the isolation of the α -configured azide adduct of the T-antigen substrate from such reaction mixtures.

The mechanistic role of D658 is less clear. Structural alignment with α -amylase places this residue on top of Y82, a residue that has been suggested to form hydrophobic interactions with the sugar. Clearly such a role is not possible for D658, both since the residue is not hydrophobic and because the hydroxyl is axial-configured, thereby decoupling the hydrophobic face. Indeed, modeling would suggest the formation of hydrogen bonds between D658 and the axial hydroxyl of the substrate. The large increase in K_m value (no saturation was seen) is consistent with a role in substrate binding, as is the 15-fold decrease in k_{cat}/K_m seen upon mutation. Indeed, in the related B1GH101 enzyme, a similar substrate model also suggests that the residue is involved in hydrogen bonding with the substrate (14).

The sugar binding specificity of the putative binding domain(s) was probed using two separate catalytically compromised mutants D764A/F. The mutations were made both to avoid cleavage of the glycans on the array and, in the case of the D764F mutant, to avoid active site binding. Only the disaccharide that functions as the substrate showed significant binding. This could reflect an adaptation to the galacto-*N*-biose rich mucin-covered environment that *Streptococcus* grows in. The binding of SpGH101 D764A/F to 3-*O*-sulphated disaccharide is interesting as this is a possible modification found in human mucin (32), although it is not known if this would be a substrate for SpGH101. The binding of the mutant E764F was weaker to the 3-*O*-sulphated disaccharide however both mutants showed large variations (100% standard deviation) in binding to this glycan, so it may not reflect a true ligand. Other weak binding to more complex ligands is seen in both experiments, but the importance of these interactions is unknown, and at this point it seems prudent not to over interpret such qualitative binding experiments. Low level binding to several glycans is common in these experiments, and more detailed binding studies using quantitative methods such as SPR or ITC will be required to evaluate the binding of individual glycans to the protein.

The strength of binding of the enzyme to the array suggests the binding is stronger than what would be seen only with an interaction through the active site, since the enzyme must stay bound during washing steps; this binding is more typical of a carbohydrate binding domain. Further analysis is required to determine where in the protein structure this binding is taking place, however an obvious target is the C-terminal CBM32 module. The CBM32 family has been shown to bind to various galactosides (see <http://www.cazy.org/fam/CBM32.html>) and this type of CBM is present in other cell surface enzymes from *Streptococcus*, although the specificity of the CBM32 modules is not yet known in these enzymes.

These GH101 proteins appear to be cell surface anchored which provides the opportunity to use carbohydrate binding as a tether for binding to host surfaces. It is clear that these enzymes have different substrate specificities (33) and more work on the various family members will be required to determine if the CBM like domains have the same specificity. Three dimensional structures of other related protein will be required to decipher how binding specificity is achieved by the various GH101 enzymes.

Acknowledgments

We acknowledge Dr. Hongming Chen for synthesis of substrates, Cynthia Bainbridge for technical assistance with the purification of proteins, and Dr. Chris Whitfield for his generosity during the preparation of the enzymes for the glycan array work. We thank Dr. David F. Smith at Core H of the Consortium for Functional Glycomics (National Institutes of Health Grant GM62116) for glycan array screening.

Funding sources

This work was funded by the Natural Sciences and Engineering Research Council of Canada and The Canadian Institutes for Health Research. R.Z thanks the BC Innovation Council for receipt of a scholarship and SGW thanks the Canada Research Chairs Program for salary support.

Abbreviations

T-Ag	Gal- β -1,3-GalNAc- α -R
DNP	2,4-dinitrophenyl
pNP	para-nitrophenyl
FCHASE	6-(fluorescein-5-carboxyamido)-hexanoic acid succinidyl ester

References

1. Scott JA. The global epidemiology of childhood pneumonia 20 years. *Bulletin of the World Health Organization* 2008;86:494–496C. [PubMed: 18568283]
2. Burnaugh AM, Frantz LJ, King SJ. Growth of *Streptococcus pneumoniae* on Human Glycoconjugates Is Dependent upon the Sequential Activity of Bacterial Exoglycosidases. *J Bacteriol* 2008;190:221–230. [PubMed: 17981977]
3. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucl Acids Res* 2009;37:D233–D238. [PubMed: 18838391]
4. Hoskins J, Alborn WE Jr, Arnold J, Blaszcak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu DJ, Fuller W, Geringer C, Gilmour R, Glass JS, Khoja H, Kraft AR, Lagace RE, LeBlanc DJ, Lee LN, Lefkowitz EJ, Lu J, Matsushima P, McAhren SM, McHenney M, McLeaster K, Mundy CW, Nicas TI, Norris FH, O’Gara M, Peery RB, Robertson GT, Rockey P, Sun PM, Winkler ME, Yang Y, Young-Bellido M, Zhao G, Zook CA, Baltz RH, Jaskunas SR, Rosteck PR Jr, Skatrud PL, Glass JI. Genome of the Bacterium *Streptococcus pneumoniae* Strain R6. *J Bacteriol* 2001;183:5709–5717. [PubMed: 11544234]
5. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald LA, Feldblyum TV, Angiuoli S, Dickinson T, Hickey EK, Holt IE, Loftus BJ, Yang F, Smith HO, Venter JC, Dougherty BA, Morrison DA, Hollingshead SK, Fraser CM. Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*. *Science* 2001;293:498–506. [PubMed: 11463916]
6. Muramatsu H, Tachikui H, Ushida H, Song Xj, Qiu Y, Yamamoto S, Muramatsu T. Molecular Cloning and Expression of Endo- β -N-Acetylglucosaminidase D, Which Acts on the Core Structure of Complex Type Asparagine-Linked Oligosaccharides. *J Biochem* 2001;129:923–928. [PubMed: 11388907]
7. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Micro* 2008;6:288–301.
8. Marion C, Limoli DH, Bobulsky GS, Abraham JL, Burnaugh AM, King SJ. Identification of a Pneumococcal Glycosidase That Modifies O-Linked Glycans. *Infect Immun* 2009;77:1389–1396. [PubMed: 19139197]
9. Umemoto J, Bhavanandan VP, Davidson EA. Purification and properties of an endo- α -N-acetyl-D-galactosaminidase from *Diplococcus pneumoniae*. *J Biol Chem* 1977;252:8609–8614. [PubMed: 21877]
10. Bardales RM, Bhavanandan VP. Transglycosylation and transfer reaction activities of endo- α -N-acetyl-D-galactosaminidase from *Diplococcus (Streptococcus) pneumoniae*. *J Biol Chem* 1989;264:9893–19897.
11. Fujita K, Oura F, Nagamine N, Katayama T, Hiratake J, Sakata K, Kumagai H, Yamamoto K. Identification and Molecular Cloning of a Novel Glycoside Hydrolase Family of Core 1 Type O-

- Glycan-specific Endo- α -N-acetylgalactosaminidase from *Bifidobacterium longum*. J Biol Chem 2005;280:37415–37422. [PubMed: 16141207]
12. Caines MEC, Zhu H, Vuckovic M, Willis LM, Withers SG, Wakarchuk WW, Strynadka NCJ. The Structural Basis for T-antigen Hydrolysis by *Streptococcus pneumoniae*: A TARGET FOR STRUCTURE-BASED VACCINE DESIGN. J Biol Chem 2008;283:31279–31283. [PubMed: 18784084]
 13. MacGregor EA, Janecek S, Svensson B. Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 2001;1546:1–20.
 14. Suzuki R, Katayama T, Kitaoka M, Kumagai H, Wakagi T, Shoun H, Ashida H, Yamamoto K, Fushinobu S. Crystallographic and mutational analyses of substrate recognition of endo- α -N-acetylgalactosaminidase from *Bifidobacterium longum*. J Biochem 2009:mvp086.
 15. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M. Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. Protein Sci 1994;3:467–475. [PubMed: 8019418]
 16. Willis LM, Gilbert M, Karwaski MF, Blanchard MC, Wakarchuk WW. Characterization of the α -2,8-polysialyltransferase from *Neisseria meningitidis* with synthetic acceptors, and the development of a self-priming polysialyltransferase fusion enzyme. Glycobiology 2008;18:177–186. [PubMed: 18000029]
 17. Chen H, Withers SG. Facile Synthesis of 2,4-Dinitrophenyl α -D-Glycopyranosides as Chromogenic Substrates for alpha-Glycosidases. ChemBioChem 2007;8:719–722. [PubMed: 17373018]
 18. Bernatchez S, Gilbert M, Blanchard MC, Karwaski MF, Li J, DeFrees S, Wakarchuk WW. Variants of the β -1,3-Galactosyltransferase CgtB from the Bacterium *Campylobacter jejuni* have Distinct Acceptor Specificities. Glycobiology 2007;17:1333–1343. [PubMed: 17766267]
 19. Wakarchuk WW, Cunningham AM. Capillary electrophoresis as an assay method for monitoring glycosyltransferase activity. Methods Mol Biol 2003;213:263–274. [PubMed: 12619996]
 20. Bohne A, Lang E, von der Lieth CW. SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides. Bioinformatics 1999;15:767–768. [PubMed: 10498779]
 21. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl Acids Res 2005;33:W363–W367. [PubMed: 15980490]
 22. Delano, WL. The PyMOL Molecular Graphics System. 2002. <http://www.pymol.org>
 23. Vocadlo DJ, Wicki J, Rupitz K, Withers SG. A Case for Reverse Protonation: Identification of Glu160 as an Acid/Base Catalyst in *Thermoanaerobacterium saccharolyticum* β -Xylosidase and Detailed Kinetic Analysis of a Site-Directed Mutant. Biochemistry 2002;41:9736–9746. [PubMed: 12146939]
 24. Tull D, Withers SG. Mechanisms of Cellulases and Xylanases: A Detailed Kinetic Study of the Exo- β -1,4-glycanase from *Cellulomonas Fimi*. Biochemistry 1994;33:6363–6370. [PubMed: 8193153]
 25. Wicki J, Schloegl J, Tarling CA, Withers SG. Recruitment of Both Uniform and Differential Binding Energy in Enzymatic Catalysis: Xylanases from Families 10 and 11. Biochemistry 2007;46:6996–7005. [PubMed: 17503782]
 26. Ly HD, Withers SG. MUTAGENESIS OF GLYCOSIDASES. Annu Rev Biochem 1999;68:487–522. [PubMed: 10872458]
 27. Zechel DL, Withers SG. Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. Accounts of Chemical Research 2000;33:11–18. [PubMed: 10639071]
 28. Zechel DL, Withers SG. Dissection of nucleophilic and acid-base catalysis in glycosidases. Current Opinion in Chemical Biology 2001;5:643–649. [PubMed: 11738173]
 29. Rydberg EH, Li C, Maurus R, Overall CM, Brayer GD, Withers SG. Mechanistic Analyses of Catalysis in Human Pancreatic α -Amylase: Detailed Kinetic and Structural Studies of Mutants of Three Conserved Carboxylic Acids. Biochemistry 2002;41:4492–4502. [PubMed: 11914097]
 30. Tarling CA, He S, Sulzenbacher G, Bignon C, Bourne Y, Henrissat B, Withers SG. Identification of the Catalytic Nucleophile of the Family 29 α -L-Fucosidase from *Thermotoga maritima* through Trapping of a Covalent Glycosyl-Enzyme Intermediate and Mutagenesis. J Biol Chem 2003;278:47394–47399. [PubMed: 12975375]

31. Wang Q, Trimbur D, Graham R, Warren RAJ, Withers SG. Identification of the Acid/Base Catalyst in *Agrobacterium faecalis* β -Glucosidase by Kinetic Analysis of Mutants. *Biochemistry* 1995;34:14554–14562. [PubMed: 7578061]
32. Kuhns W, Jain RK, Matta KL, Paulsen H, Baker MA, Geyer R, Brockhausen I. Characterization of a novel mucin sulphotransferase activity synthesizing sulphated O-glycan core 1, 3-sulphate-Gal- β -1-3GalNAc- α -R. *Glycobiology* 1995;5:689–697. [PubMed: 8608271]
33. Koutsioulis D, Landry D, Guthrie EP. Novel endo- α -N-acetylgalactosaminidases with broader substrate specificity. *Glycobiology* 2008;18:799–805. [PubMed: 18635885]

Figure 1. Sequence alignment of the catalytic domain of GH101 and structural representation of the catalytic domain of SpGH101

A. Representative sequences from GH101 were used for a multiple sequence alignment using clustal 2.0, displayed with Genedoc and the section defined as the catalytic domain was coloured for those residues that are absolutely conserved in this region. Sequences are: *Arthrobacter aurescens* TC1 genebank ABM10140.1; *Bacillus cereus* AH187 genebank ACJ78453.1; *Bifidobacterium longum* subsp. *longum* JCM 1217 genebank AAX44931.1; *Clostridium perfringens* str. 13 BAB80399.1; *Enterococcus faecalis* genebank BAG75105.1; *Propionibacterium acnes* KPA171202 genebank AAT83312.1; *Streptococcus pneumoniae* R6 genebank AAK99132.1; *Streptomyces coelicolor* A3(2) genebank CAA20079.1

B. Those residues which are absolutely conserved were mapped onto the alpha-carbon backbone (yellow) from PDB 3ECQ of the catalytic domain by colouring them blue. Those absolutely conserved residues which are in the active site pocket are represented by the stick structures on this backbone. Residues displayed are F618; H657; D658; N692; H694; E699; Y701; W724/726; Y762; D764; V765; E796; W797; W810; Y816; W867. W residues are

magenta; E/D residues are red; H residues are orange; Y residues are green; F residues are yellow; other residues are CPK coloured. Images were made in Pymol.

Figure 2. Schematic representation of the SpGH101 gene

Various deletion mutants were tested for expression and activity, and the truncation points are shown by the arrows under the schematic. The majority of the work was performed with the protein expressed without the leader sequence. The residues which define the catalytic domain are found between residues 550 and 950. There is a carbohydrate binding module of the CBM32 type in the C-terminal end of the protein.

$$pK_{a1} = 5.3 \pm 0.1 \quad pK_{a2} = 7.8 \pm 0.1$$

Figure 3. Dependence of k_{cat}/K_m upon pH for hydrolysis of DNP-TAg by SpGH101
The data were fit to a double ionization pH curve by nonlinear regression using the program GraFit 5.0.13 (Erithacus Software Limited, 2006).

B

Figure 4. Analysis of SpGH101 D764A/F mutant binding to a glycan array

SpGH101 mutants of the catalytic nucleophile were examined for binding to glycan on the glycan array from the Consortium for Functional Glycomics. The proteins were directly labeled with a fluorescent dye for measurement of binding. D764A was analyzed on the array version 3.1, D764F was analyzed using array version 4.0. The error bars (blue) show the standard deviation of the RFU value from the average of 4 spots on the array.

A

B

Figure 5. Mechanism and active site structure of SpGH101

(A) A schematic of the retaining glycoside hydrolysis model is shown for reference.

(B) A model of the SpGH101 active site with the T-Ag disaccharide was generated using a docked T-Ag disaccharide and SpGH101 (PDB 3ECQ). The distances were measured using Pymol.

Figure 6. Chemical rescue of the SpGH101 E796Q mutant by increasing concentrations of sodium azide

(A) Plot of the activity of E796Q versus azide concentration.

(B) Schematic of mechanism of azide rescue reaction

Table 1

Primers used in the cloning and mutagenesis of SpGH101

Primer name	Sequence 5' – 3'	Comments
SpGH101-F	GCTGGAGCTGGACATATGAGCGTGAGTCTGGTTCCACGGCGAACTTACC	Forward primer to clone R6 gene Sp0328 without the 40aa N-terminal signal sequence
SpGH101-R	GCTGGAGCTGGAGTCGACTTATTAGTCTTCTTCGTTTTACTACAAAGAGAGCAGATAGGGCTAG	Reverse primer to clone R6 gene Sp0328
Δ200-C-F	GCTGGAGCTGGAGTCGACTTATTATGTCACAACAAGTTTCACATCACGCAAGTTACCATTGAAC	Reverse primer to make 200aa C-terminal truncation
Δ200-C-R	GCTGGAGCTGGACATATGTTTAAAGATACCAAGAATAATGTTTTGTGCGTTACG	Forward primer to make 200aa N-terminal truncation
D658A-F	CCTTAAAGGATATGGTAGCGAAGGCCATGCCTCTGGTCACTTGAAC	Forward primer to make D658A
D658A-R	GTC AAGTGACCAGAGGCATGGCCTTCGCTACCATATCCTTTAAGG	Reverse primer to make D658A
D764A-F	GGTCTCGACTTTATCTATGTGGCCGTTTGGGGTAATGGTCAATCAGG	Forward primer to make D764A
D764A-R	CCTGATTGACCATTACCCCAAACGGCCACATAGATAAAGTCGAGACC	Reverse primer to make D764A
D764F-F	GGTCTCGACTTTATCTATGTGTTTGTGGGGTAATGGTCAATCAGG	Forward primer to make D764A
D764F-R	CCTGATTGACCATTACCCCAAACAAACACATAGATAAAGTCGAGACC	Reverse primer to make D764A
E796A-F	GGCTGGCGCTTTGCGATCGCGTGGGGCCATGGTGGTGAG	Forward primer to make E796A
E796A-R	CTCACCACCATGGCCCCACGCGATCGCAAAGCGCCAGCC	Reverse primer to make E796A

Primer name	Sequence 5' – 3'	Comments
E796Q-F	GGCTGGCGCTTTGCGATCCAGTGGGGCCATGGTGGTGAG	Forward primer to make E796Q
E796Q-R	CTCACCACCATGGCCCCACTGGATCGCAAAGCGCCAGCC	Reverse primer to make E796Q

Table 2

Table of kinetic parameters for SpGH101 and mutants

Enzyme	Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \text{mM}^{-1}$)
$\Delta 1-39$ (WT)	DNP-TAg	34 ± 2	541 ± 14	15.9
$\Delta 1-39, \Delta 1568-1767$	DNP-TAg	48 ± 3	748 ± 19	15.6
D658A	DNP-TAg	ND ^a	ND ^a	0.987^b
D764A NUC	DNP-TAg	33 ± 2	0.76 ± 0.002	0.023
E796A A/B	DNP-TAg	1.3 ± 0.1	18 ± 1	13.8
E796Q A/B	DNP-TAg	1.2 ± 0.1	14 ± 1	11.7
$\Delta 1-39$ (WT)	DNP-GalNAc	2400 ± 100	1.6 ± 0.1	0.0006
E796A A/B	DNP-GalNAc	2100 ± 220	5.0 ± 0.2	0.002
$\Delta 1-39$ (WT)	pNP-TAg	62 ± 3	579 ± 11	9.3
E796A A/B	pNP-TAg	21 ± 1	1.0 ± 0.1	0.048
E796Q A/B	pNP-TAg	212 ± 17	2.1 ± 0.1	0.0099

^aThese values could not be determined from the data obtained.^bUsing the substrate depletion method (ref) where $[S]$ is $\ll K_m$, a direct estimate of k_{cat}/K_m can be obtained.

Table 3

Kinetic parameters for the cleavage of DNP-T-Ag by the E796A and E796Q mutants of SpGH101 in the presence of anions.

Mutant/anion	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \text{mM}^{-1}$)
E796A + 100 mM NaN_3	5.6 ± 0.6	33 ± 1	5.9
E796A + 100 mM NaCl	1.5 ± 0.1	16.6 ± 1	11
E796A + 500 mM sodium formate	11 ± 1	48 ± 1	4.4
E796A + 500 mM NaCl	1.6 ± 0.1	14 ± 1	8.75
E796Q + 500 mM NaN_3	15 ± 1	74 ± 2	4.9
E796Q + 500 mM sodium formate	9.7 ± 0.8	32 ± 1	3.3
E796Q + 500 mM NaCl	1.7 ± 0.2	9.1 ± 0.3	5.4