10334 *Biochemistry* **2009,** *48,* 10334–10341
DOI: 10.1021/bi9013825

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

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*[,‡]

[‡]*Glycobiology Program, Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, Canada K1A 0R6, and* [§]*Departments of Chemistry and Biochemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.* [||]*These authors contributed equally to this work.*

Received August 18, 2009; Revised Manuscript Received September 23, 2009

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 determined three-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. On the basis of 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. pK_a values of 5.3 and 7.2 were assigned to D764 and E796 on the basis of the pK_a 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 Alexafluor 488. These studies revealed binding to galacto-*N*-biose, consistent with a role for these domains in localizing the enzyme near its substrates.

Streptococcus 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 conducted to determine the factors that contribute to the virulence of this organism. Of major importance to this pathogen are the abilities 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 reference strain *S. pneumoniae* R6 and of a clinical isolate contain at least six cell surface-localized enzymes that are thought to play a role in virulence (*4*−*6*). These cell-associated enzymes include neuraminidases, hyaluronidases, *β*-galactosidases, and endo-*β*-*N*-acetylglucosaminidases, which have been studied in some detail (reviewed in ref *7*). A new member of this cell surface family of enzymes belonging to the GH101 family of glycoside hydrolases, endo-α-*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 nonessential 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 have been 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 determined 3D[1] structure of SpGH101 (*12*), in conjunction with sequence alignments, reveals that the GH101 enzymes are multimodular 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, and modules that resemble a degenerate legume lectin fold. The catalytic domain is a distorted $(\beta/\alpha)_8$ 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*).

[†]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 S.G.W. thanks the Canada Research Chairs Program for salary support.

*To whom correspondence should be addressed: National Research Council of Canada, Institute for Biological Sciences, Glycobiology Program, 100 Sussex Dr., Ottawa, ON, Canada K1A 0R6. E-mail: warren.wakarchuk@nrc-cnrc.gc.ca. Telephone: (613) 952-4299. Fax: (613) 941-1327.

[1]Abbreviations: TAg, Gal-*β*-1,3-GalNAc-α-R; DNP, 2,4-dinitrophenyl; pNP, *p*-nitrophenyl; FCHASE, 6-(fluorescein-5-carboxyamido)hexanoic acid succimidyl ester; 3D, three-dimensional; PDB, Protein Data Bank.

 Published on Web 09/29/2009

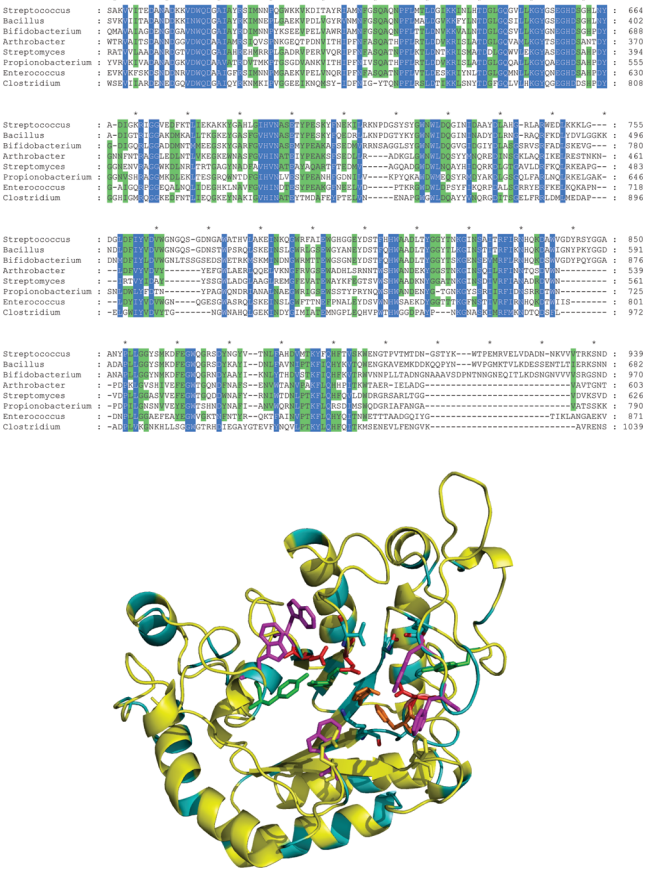

FIGURE 1: Sequence alignment of the catalytic domain of GH101 and structural representation of the catalytic domain of SpGH101. (A) Representative sequences from GH101 were used for a multiple-sequence alignment using Clustal 2.0, displayed with Genedoc, and the section defined as the catalytic domain was colored for those residues that are absolutely conserved in this region. Sequences: *Arthrobacter aurescens* TC1 (GenBank entry ABM10140.1), *Bacillus cereus* AH187 (GenBank entry ACJ78453.1), *B. longum* subsp. *longum* JCM 1217 (GenBank entry AAX44931.1), *Clostridium perfringens* strain Thirteen (genBank entry BAB80399.1), *Enterococcus faecalis* (GenBank entry BAG75105.1), *Propionibacterium acnes* KPA171202 (GenBank entry AAT83312.1), *S. pneumoniae* R6 (GenBank entry AAK99132.1), and *Streptomyces coelicolor* A3(2) (GenBank entry CAA20079.1). (B) Those residues which are absolutely conserved were mapped onto the α-carbon backbone (yellow) from PDB entry 3ECQ of the catalytic domain by being colored them blue. Those absolutely conserved residues that are in the active site pocket are represented by the stick structures on this backbone. F618, H657, D658, N692, H694, E699, Y701, W724/726, Y762, D764, V765, E796, W797, W810, Y816, and W867 are displayed. W residues are colored magenta, E/D residues red, H residues orange, Y residues green, F residues yellow, and other residues CPK colored. Images were created in PyMOL.

The degree of 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). On the basis of 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, respectively. 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 3D structure of BlGH101 has been determined, 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 about enzymes from this family is available. 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 in examining 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. *S. pneumoniae* R6 was obtained from the ATCC (catalog no. BAA-255) and was grown on Columbia Blood Agar at 37 °C with 5% CO_2. *Escherichia coli* AD202 (CGSC catalog no. 7297) 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 h and then induced with 0.5 μM isopropyl 1-thio-β-D-galactopyranoside and grown at 30 °C for 7 h. The plasmid used for cloning was a modified pCWori+ plasmid (*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, ON). PCR was performed using Phusion polymerase and the following program: 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 120 s, and finally 72 °C for 10 min. The gene was originally amplified from genomic DNA based on the Sp0328 sequence 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 listed in Table 1.

DNA was purified using either Qiaquick or Minelute kits from Qiagen Inc. Genes digested with *Nde*I and *Sal*I 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, QC). DNA sequencing was performed using an Applied Biosystems (Montreal, QC) 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 listed in Table 1.

Enzyme Preparation. Recombinant cultures grown in the presence of IPTG were lysed with an Avestin C5 Emulsiflex cell disrupter (Avestin, Ottawa, ON) and centrifuged first at 27000*g* to pellet unbroken cells and then at 100000*g* to pellet insoluble material. SpGH101 was partially purified using DEAE-Sepharose chromatography (GE Healthcare, Piscataway, NJ) with a gradient from 20 mM to 1 M NH_4OAc (pH 7.5). The enzyme

Table 1: Primers Used in the Cloning and Mutagenesis of SpGH101

primer name	sequence (5′−3′)	comments
SpGH101-F	GCTGGAGCTGGACATATGAGCGTGCAGTCTGGTTCCACGG-CGAACTTACC	forward primer to clone R6 gene Sp0328 without the 40-amino acid N-terminal signal sequence
SpGH101-R	GCTGGAGCTGGAGTCGACTTATTAGTCTTTCTTCGTTTTTAC-TACAAAGAGAGCAGATAGGGCTAG	reverse primer to clone R6 gene Sp0328
Δ200-C-F	GCTGGAGCTGGAGTCGACTTATTATGTCACAACAAGTTTCA-CATCACGCAAGTTACCATTTGAAC	reverse primer to make 200-amino acid C-terminal truncation
Δ200-C-R	GCTGGAGCTGGACATATGTTTAAAGATACCAAGAATAATGT-TTTTGTCGGTTACG	forward primer to make 200-amino acid N-terminal truncation
D658A-F	CCTTAAAGGATATGGTAGCGAAGGCCATGCCTCTGGTCACTTGAAC	forward primer to make D658A
D658A-R	GTTCAAGTGACCAGAGGCATGGCCTTCGCTACCATATCCTTTAAGG	reverse primer to make D658A
D764A-F	GGTCTCGACTTTATCTATGTGGCCGTTTGGGGTAATGGTCAATCAGG	forward primer to make D764A
D764A-R	CCTGATTGACCATTACCCCAAACGGCCACATAGATAAAGTCGAGACC	reverse primer to make D764A
D764F-F	GGTCTCGACTTTATCTATGTGTTTGTTTGGGGTAATGGTCAATCAGG	forward primer to make D764F
D764F-R	CCTGATTGACCATTACCCCAAACAAACACATAGATAAAGTCGAGACC	reverse primer to make D764F
E796A-F	GGCTGGCGCTTTGCGATCGCGTGGGGCCATGGTGGTGAG	forward primer to make E796A
E796A-R	CTCACCACCATGGCCCCACGCGATCGCAAAGCGCCAGCC	reverse primer to make E796A
E796Q-F	GGCTGGCGCTTTGCGATCCAGTGGGGCCATGGTGGTGAG	forward primer to make E796Q
E796Q-R	CTCACCACCATGGCCCCACTGGATCGCAAAGCGCCAGCC	reverse primer to make E796Q

was concentrated using Amicon Ultra-15 filtration devices (Millipore, Billerica, MA) and then applied to a Superdex-200 size exclusion column in 10 mM NH_4OAc (pH 7.5). The enzyme in the pooled fractions was finally applied to a Phenyl-Sepharose column (GE Healthcare) with a gradient from 1.5 M to 10 mM NH_4OAc (pH 7.5). Purity was assessed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (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-VGV[Gal-β-1,3-GalNAc-α-]TETP (FCHASE-[TAg]-IFNα2b). 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 cm × 80 cm) that had been equilibrated in 100 mM NH_4HCO_3. The peptide substrate FCHASE-[TAg]-IFNα2b was also synthesized as described previously (*18*). The purity of the products was determined by capillary electrophoresis on a PACE MDQ system as previously described (*19*).

Modeling of TAg in the Active Site. A PDB entry 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 version 1.1 (*22*).

Measurement of O-Glycanase Activity. Activity was measured using a substrate 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 and 0.5% sodium dodecyl sulfate and measured at 400 nm using an extinction coefficient of 14000 cm^{-1} M^{-1}. Assays with DNP-TAg and DNP-GalNAc were performed in continuous mode. For wild-type SpGH101 enzyme kinetics, 10−200 μM DNP-TAg was incubated followed by the addition of 0.2 nM enzyme. For SpGH101 mutants, 1−10 μM 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 instrument, and rates were calculated using an extinction coefficient of 10800 cm^{-1} M^{-1}. Kinetic data were analyzed using GraFit version 5.0. For the chemical rescue experiments, 10 mM to 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 SpGH101 E796Q mutant was prepared and left at room temperature overnight. Product analysis was performed by thin layer chromatography on 60 F_{254} silica gel aluminum plates (Merck) run in a 7:2:1 (v/v/v) ethyl acetate/methanol/water mixture and developed with 10% ammonium molybdate in 2 M H_2SO_4 followed by charring. We purified the final product by directly loading the crude reaction mixture onto a flash column for chromatography with Silicagel 60. The NMR spectra were recorded using a 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*). At low substrate concentrations ($[S] \ll K_m$), the reaction rates are given by the equation $v = k_{cat}[E]_o[S]/K_m$. The time course of the reaction can then be fit to a first-order equation and a pseudo-first-order rate constant, k_{obs}, extracted. Therefore, the k_{obs} values then correspond to $[E]_o k_{cat}/K_m$. Thus, k_{cat}/K_m values can be extracted by division of these obtained rate constants by the enzyme concentration. Assays were performed with 50 μM DNP-GalNAc or 5 μM DNP-TAg and the following buffers: 50 mM citrate-phosphate (pH 4.4−7.0), 50 mM sodium phosphate (pH 7.0−8.0), and 50 mM sodium borate (pH 8.0−9.0). Absorbance was measured using a Varian Cary AV-4000 instrument, and data were fitted to a first-order curve using GraFit version 5.0. 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

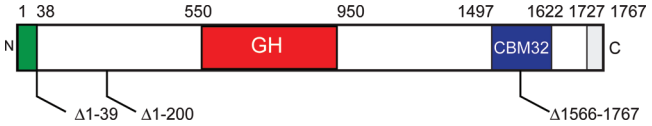

FIGURE 2: Schematic representation of the SpGH101 gene. Various deletion mutants were tested for expression and activity, and the truncation points are shown by the arrows under the schematic. The majority of the work was performed with the protein expressed without the leader sequence. The residues that define the catalytic domain are found between residues 550 and 950. There is a carbohydrate binding module of the CBM32 type at the C-terminal end of the protein.

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 glycan spot densities of both 100 and 10 μM for array version 3.1 with D764A, and at two 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 truncation of the protein 200 amino acids from either the N- or C-terminal end. 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 amino acids 40−1567), 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 SpGH101 amino acids 40−1767. The enzyme was purified to >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 wild-type (WT) enzyme. No significant differences in expression or purification were observed with these mutants.

Kinetic and Stereochemical Analysis of the Wild Type and Active Site Mutants. (*i*) *Wild Type*. Kinetic parameters for both DNP-GalNAc and DNP-TAg were measured and are listed in Table 2. The specificity of the enzyme for the TAg disaccharide was shown by the 26500-fold greater k_{cat}/K_m for the disaccharide substrate versus that for 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 improved leaving group ability of DNP makes this a better substrate than the *p*-nitrophenyl analogue. In addition, the low pK_a (4.0) of DNP ensures that the extinction coefficient of 2,4-DNP does not change significantly with at 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 SpGH101. The enzyme was also shown by NMR analysis of the reaction progress to act with retention of anomeric stereochemistry (data not shown) like *B. longum* enzyme BlGH101 (*11*).

Table 2: Kinetic Parameters for SpGH101 and Mutants

enzyme	substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
Δ1−39 (WT)	DNP-TAg	34 ± 2	541 ± 14	15.9
Δ1−39, Δ1568−1767	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
Δ1−39 (WT)	DNP-GalNAc	2400 ± 100	1.6 ± 0.1	0.0006
E796A A/B	DNP-GalNAc	2100 ± 220	5.0 ± 0.2	0.002
Δ1−39 (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

[a]These values could not be determined from the data obtained. [b]Using the substrate depletion method where [S] ≪ K_m, a direct estimate of k_{cat}/K_m can be obtained as explained in Experimental Procedures.

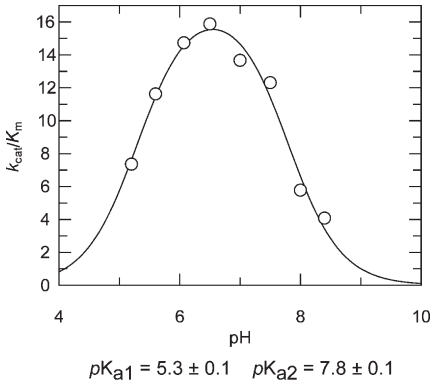

FIGURE 3: Dependence of k_{cat}/K_m upon pH for hydrolysis of DNP-TAg by SpGH101. The data were fit to a double-ionization pH curve by nonlinear regression using GraFit version 5.0.13 (Erithacus Software Limited, 2006).

(*ii*) *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 pK_a values of 5.3 ± 0.1 and 7.8 ± 0.1 (Figure 3).

Kinetic Analysis of Catalytic Site Mutants. (*i*) *Aryl Glycoside Substrates*. On the basis of the active site substrate complex model, mutants D658A, D764A, and E796A/Q were constructed and analyzed 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 WT enzyme. In fact, this is likely a maximum estimate of the activity since, on the basis of the similarity of its K_m value to that of the WT enzyme, it is quite probable that even this low activity is derived from a small amount of contaminating WT enzyme. The next most active is the D658A mutant, with a k_{cat}/K_m that is 16-fold lower than that of the WT enzyme. However, it did not exhibit 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 E796Q mutants, with k_{cat}/K_m values essentially the same as that of the WT enzyme. Importantly, however, the k_{cat} values are ~30-fold lower than that of the wild-type enzyme, with the K_m values also correspondingly lowered.

Table 3: Kinetic Parameters for the Cleavage of DNP-TAg by the E796A and E796Q Mutants of SpGH101 in the Presence of Anions

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

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*).

(*ii*) *"Natural" Glycopeptide Substrates*. To investigate the enzymatic activity of the active site mutants on a more natural substrate, a synthetic glycopeptide, FCHASE-[TAg]-IFNα2b, 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 addition of 34 nM WT SpGH101, 0.5 mM FCHASE-[TAg]-IFNα2b was rapidly cleaved, with hydrolysis being complete within 3 h at room temperature (data not shown). However, the addition of either 290 nM E796A or 220 nM E796Q to 0.5 mM FCHASE-[TAg]-IFNα2b resulted in no observable cleavage, even after incubation for 2 days.

(*iii*) *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), but these rescue data are unreliable given the low observed activity, and the possible presence of a low level of WT enzyme in the preparations. 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 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.

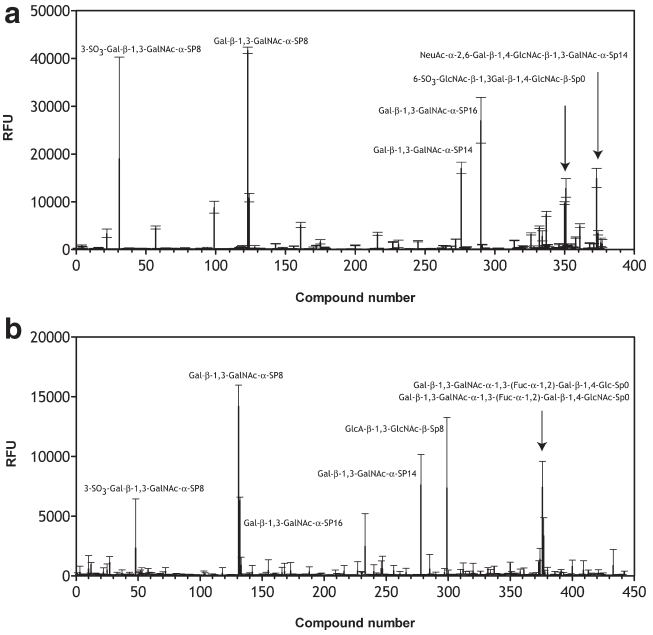

FIGURE 4: Analysis of the SpGH101 D764A/F mutant binding to a glycan array. SpGH101 mutants of the catalytic nucleophile were examined for binding to glycan on the glycan array from the Consortium for Functional Glycomics. The proteins were directly labeled with a fluorescent dye for measurement of binding.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

(*iv*) *Azide Rescue Product Identification*. TLC analysis of the reaction mixture of DNP-TAg, sodium azide, and the 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 ^{1}H NMR of the purified product. The relatively large chemical shift (δ 5.52) is also very typical for α-glycosyl azide since a much smaller chemical shift would be expected for a β-glycosyl azide: ^{1}H NMR (300 MHz, CD_3OD) δ 5.52 (1H, d, $J_{1,2}$ = 4.2 Hz, 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, $COCH_3$).

(*v*) *Glycan Array Experiments*. To probe the carbohydrate binding properties of the SpGH101 enzyme, we used the nucleophile mutant as a probe on a glycan microarray. The strategy was to use two versions of this mutation, D764A which the enzyme could potentially bind through the active site but which 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*-sulfated version of this disaccharide (Figure 4a). There were several other weak interactions that 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 (Figure 4b).

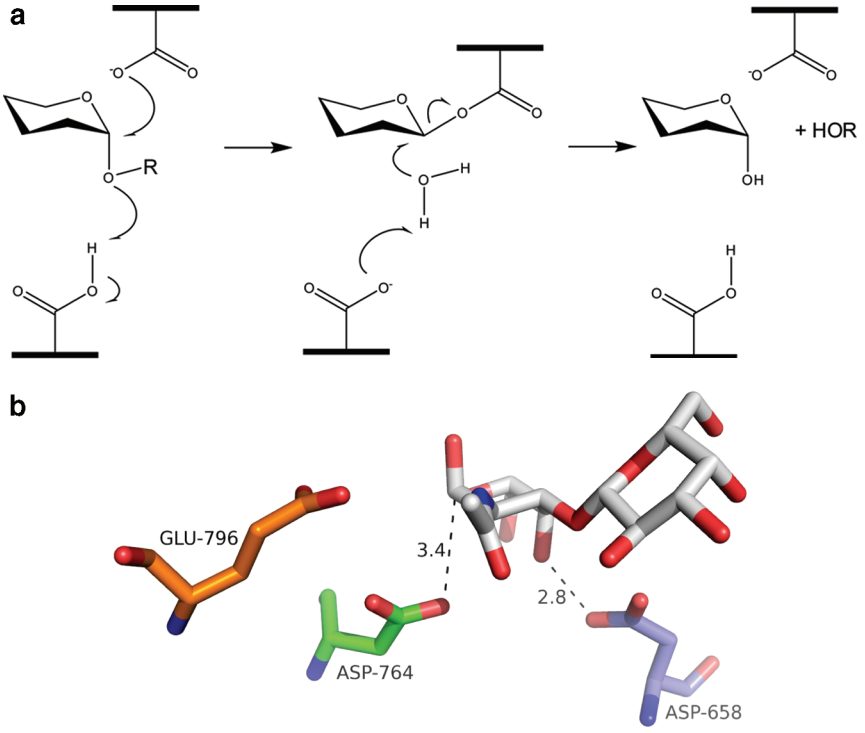

FIGURE 5: Mechanism and active site structure of SpGH101. (A) Schematic of the retaining glycoside hydrolysis model shown for reference. (B) Model of the SpGH101 active site with the TAg disaccharide generated using a docked TAg disaccharide and SpGH101 (PDB entry 3ECQ). The distances were measured using PyMOL.

DISCUSSION

The kinetic and mechanistic data that we 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 (Figure 5). These are the roles that were proposed on the basis of modeling of the substrate binding in 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. 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, nonactivated 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 incubation for 2 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 is derived 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 intermediate formed on the wild-type enzyme because of electrostatic screening from the deprotonated E796. Removal of that charge in the alanine mutant allows direct attack of azide or formate at the anomeric center, with associated increases in the steady state rate. Parallel studies with sodium chloride confirm that the rate increases are not due to salt effects.

The relatively small increases in the steady state rate observed in these anion rescue experiments are consistent with findings for 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" behavior seen in such experiments (Figure 6) is not a reflection of saturable reversible binding of the anion but rather is a consequence of a change in the 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 a substrate, consistent with the indications (from high K_m values) that the

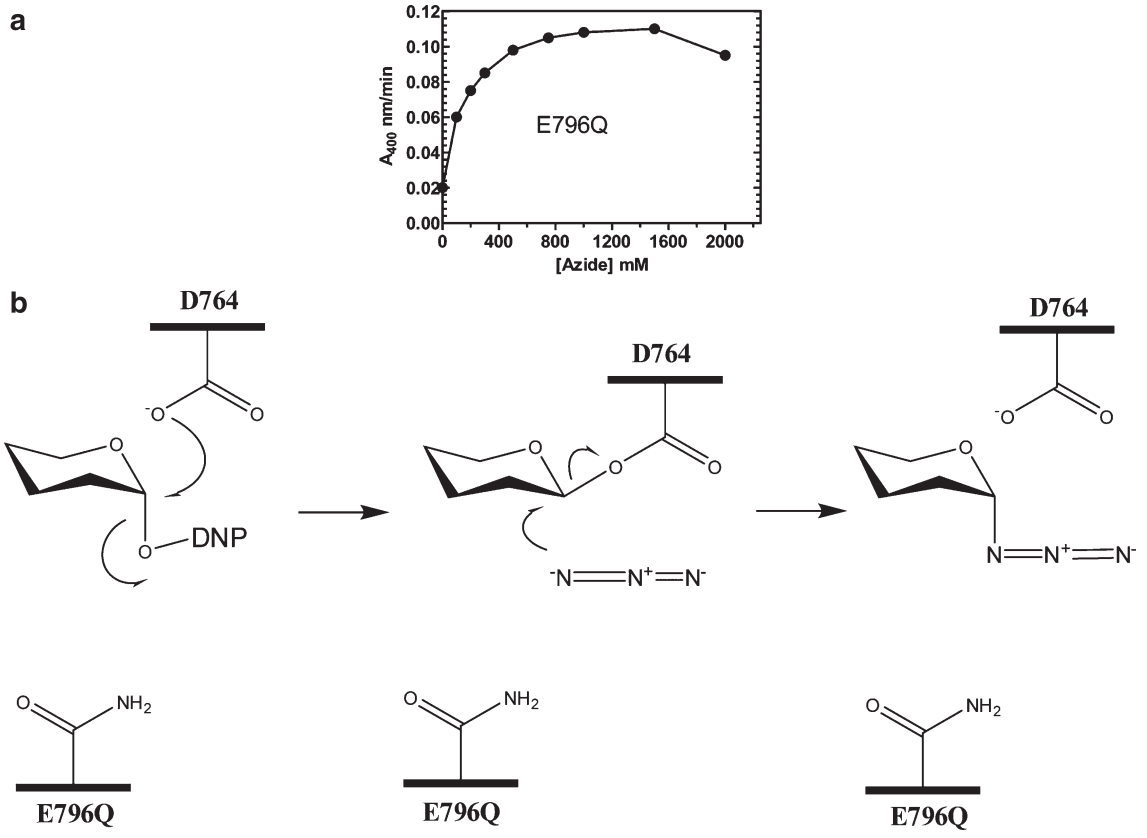

FIGURE 6: Chemical rescue of the SpGH101 E796Q mutant by increasing concentrations of sodium azide. (A) Plot of the activity of E796Q vs azide concentration. (B) Schematic of the mechanism of the azide rescue reaction.

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 and -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 exhibited significant binding. This could reflect an adaptation to the galacto-*N*-biose rich mucin-covered environment in which *Streptococcus* grows. The binding of SpGH101 D764A/F to 3-*O*-sulfated 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 mutant E764F to the 3-*O*-sulfated disaccharide was weaker; 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; however, the importance of these interactions is unknown, and at this point, it seems prudent not to overinterpret 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, a fact that provides an 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.

ACKNOWLEDGMENT

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.

REFERENCES

1. Scott, J. A. (2008) The global epidemiology of childhood pneumonia 20 years on. *Bull. W.H.O. 86*, 494C–496C.
2. Burnaugh, A. M., Frantz, L. J., and King, S. J. (2008) Growth of *Streptococcus pneumoniae* on Human Glycoconjugates Is Dependent upon the Sequential Activity of Bacterial Exoglycosidases. *J. Bacteriol. 190*, 221–230.
3. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An Expert Resource for Glycogenomics. *Nucleic Acids Res. 37*, D233–D238.
4. Hoskins, J., Alborn, W. E., Jr., Arnold, J., Blaszczak, L. C., Burgett, S., DeHoff, B. S., Estrem, S. T., Fritz, L., Fu, D. J., Fuller, W., Geringer, C., Gilmour, R., Glass, J. S., Khoja, H., Kraft, A. R., Lagace, R. E., LeBlanc, D. J., Lee, L. N., Lefkowitz, E. J., Lu, J., Matsushima, P., McAhren, S. M., McHenney, M., McLeaster, K., Mundy, C. W., Nicas, T. I., Norris, F. H., O'Gara, M., Peery, R. B., Robertson, G. T., Rockey, P., Sun, P. M., Winkler, M. E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C. A., Baltz, R. H., Jaskunas, S. R., Rosteck, P. R., Jr., Skatrud, P. L., and Glass, J. I. (2001) Genome of the Bacterium *Streptococcus pneumoniae* Strain R6. *J. Bacteriol. 183*, 5709–5717.
5. Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M. (2001) Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*. *Science 293*, 498–506.
6. Muramatsu, H., Tachikui, H., Ushida, H., Song, X. j., Qiu, Y., Yamamoto, S., and Muramatsu, T. (2001) Molecular Cloning and Expression of Endo-β-N-Acetylglucosaminidase D, Which Acts on the Core Structure of Complex Type Asparagine-Linked Oligosaccharides. *J. Biochem. 129*, 923–928.
7. Kadioglu, A., Weiser, J. N., Paton, J. C., and Andrew, P. W. (2008) The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol. 6*, 288–301.
8. Marion, C., Limoli, D. H., Bobulsky, G. S., Abraham, J. L., Burnaugh, A. M., and King, S. J. (2009) Identification of a Pneumococcal Glycosidase That Modifies O-Linked Glycans. *Infect. Immun. 77*, 1389–1396.
9. Umemoto, J., Bhavanandan, V. P., and Davidson, E. A. (1977) Purification and properties of an endo-α-N-acetyl-D-galactosaminidase from *Diplococcus pneumoniae*. *J. Biol. Chem. 252*, 8609–8614.
10. Bardales, R. M., and Bhavanandan, V. P. (1989) Transglycosylation and transfer reaction activities of endo-α-N-acetyl-D-galactosaminidase from *Diplococcus* (*Streptococcus*) pneumoniae. *J. Biol. Chem. 264*, 19893–19897.
11. Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H., and Yamamoto, K. (2005) 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. 280*, 37415–37422.
12. Caines, M. E. C., Zhu, H., Vuckovic, M., Willis, L. M., Withers, S. G., Wakarchuk, W. W., and Strynadka, N. C. J. (2008) The Structural Basis for T-antigen Hydrolysis by *Streptococcus pneumoniae*: A Target for Structure-Based Vaccine Design. *J. Biol. Chem. 283*, 31279–31283.
13. MacGregor, E. A., Janecek, S., and Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta 1546*, 1–20.
14. Suzuki, R., Katayama, T., Kitaoka, M., Kumagai, H., Wakagi, T., Shoun, H., Ashida, H., Yamamoto, K., and Fushinobu, S. (2009) Crystallographic and mutational analyses of substrate recognition of endo-α-N-acetylgalactosaminidase from *Bifidobacterium longum*. *J. Biochem. 146*, 389–398.
15. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. *Protein Sci. 3*, 467–475.
16. Willis, L. M., Gilbert, M., Karwaski, M. F., Blanchard, M. C., and Wakarchuk, W. W. (2008) Characterization of the α-2,8-polysialyltransferase from *Neisseria meningitidis* with synthetic acceptors, and the development of a self-priming polysialyltransferase fusion enzyme. *Glycobiology 18*, 177–186.
17. Chen, H., and Withers, S G. (2007) Facile Synthesis of 2,4-Dinitrophenyl α-D-Glycopyranosides as Chromogenic Substrates for α-Glycosidases. *ChemBioChem 8*, 719–722.
18. Bernatchez, S., Gilbert, M., Blanchard, M. C., Karwaski, M. F., Li, J., DeFrees, S., and Wakarchuk, W. W. (2007) Variants of the β-1,3-Galactosyltransferase CgtB from the Bacterium *Campylobacter jejuni* have Distinct Acceptor Specificities. *Glycobiology 17*, 1333–1343.
19. Wakarchuk, W. W., and Cunningham, A. M. (2003) Capillary electrophoresis as an assay method for monitoring glycosyltransferase activity. *Methods Mol. Biol. 213*, 263–274.
20. Bohne, A., Lang, E., and von der Lieth, C. W. (1999) SWEET: WWW-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics 15*, 767–768.
21. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res. 33*, W363–W367.
22. Delano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.
23. Vocadlo, D. J., Wicki, J., Rupitz, K., and Withers, S. G. (2002) 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 41*, 9736–9746.
24. Tull, D., and Withers, S. G. (1994) Mechanisms of Cellulases and Xylanases: A Detailed Kinetic Study of the Exo-β-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry 33*, 6363–6370.
25. Wicki, J., Schloegl, J., Tarling, C. A., and Withers, S. G. (2007) Recruitment of Both Uniform and Differential Binding Energy in Enzymatic Catalysis: Xylanases from Families 10 and 11. *Biochemistry 46*, 6996–7005.
26. Ly, H. D., and Withers, S. G. (1999) Mutagenesis of Glycosidases. *Annu. Rev. Biochem. 68*, 487–522.
27. Zechel, D. L., and Withers, S. G. (2000) Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Acc. Chem. Res. 33*, 11–18.
28. Zechel, D. L., and Withers, S. G. (2001) Dissection of nucleophilic and acid-base catalysis in glycosidases. *Curr. Opin. Chem. Biol. 5*, 643–649.
29. Rydberg, E. H., Li, C., Maurus, R., Overall, C. M., Brayer, G. D., and Withers, S. G. (2002) Mechanistic Analyses of Catalysis in Human Pancreatic α-Amylase: Detailed Kinetic and Structural Studies of Mutants of Three Conserved Carboxylic Acids. *Biochemistry 41*, 4492–4502.
30. Tarling, C. A., He, S., Sulzenbacher, G., Bignon, C., Bourne, Y., Henrissat, B., and Withers, S. G. (2003) 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. 278*, 47394–47399.
31. Wang, Q., Trimbur, D., Graham, R., Warren, R. A. J., and Withers, S. G. (1995) Identification of the Acid/Base Catalyst in *Agrobacterium faecalis* β-Glucosidase by Kinetic Analysis of Mutants. *Biochemistry 34*, 14554–14562.
32. Kuhns, W., Jain, R. K., Matta, K. L., Paulsen, H., Baker, M. A., Geyer, R., and Brockhausen, I. (1995) Characterization of a novel mucin sulphotransferase activity synthesizing sulphated O-glycan core 1,3-sulphate-Gal-β-1−3GalNAc-α-R. *Glycobiology 5*, 689–697.
33. Koutsioulis, D., Landry, D., and Guthrie, E. P. (2008) Novel endo-α-N-acetylgalactosaminidases with broader substrate specificity. *Glycobiology 18*, 799–805.