**Streptococcus pneumoniae** endohexosaminidase D: Structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases.

D. Wade Abbott¹, Matthew S. Macauley², David J. Vocadlo²* and Alisdair B. Boraston¹*

¹Biochemistry & Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, BC, V8W 3P6, Canada.² Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada

¶ These authors contributed equally to this work.
*Correspondence should be addressed to: Alisdair B. Boraston (boraston@uvic.ca) or David J. Vocadlo (dvocadlo@sfu.ca)

**Running Title:** Structural and mechanistic analysis of an endohexosaminidase.
**Keywords:** NAG-thiazoline, anchimeric assistance, N-glycan, substrate-participation, oxazoline, enzyme mechanism.

**ABSTRACT**

Endo-beta-D-glucosaminidases from family 85 of glycoside hydrolases (GH85 endohexosaminidases) act to cleave the glycosidic linkage between the two N-acetylglucosamine units that make up the chitobiose core of N-glycans. Endohexosaminidase D, Endo-D, produced by *Streptococcus pneumoniae* is believed to contribute to the virulence of this organism by playing a role in the deglycosylation of IgG antibodies. Endohexosaminidases have received significant attention for this reason and, moreover, because they are powerful tools for chemoenzymatic synthesis of proteins having defined glycoforms. Here we describe mechanistic and structural studies of the catalytic domain (SpGH85) of Endo-D that provide compelling support for GH85 enzymes using a catalytic mechanism involving substrate-assisted catalysis. Furthermore, the structure of SpGH85 in complex with the mechanism based competitive inhibitor NAG-thiazoline ($K_d = 28 \, \mu M$) provides a coherent rationale for previous mutagenesis studies of Endo-D and other related GH85 enzymes. We also find GH85, GH56, and GH18 enzymes have a similar configuration of catalytic residues. Notably, GH85 enzymes have an asparagine in place of the aspartate residue found in these other families of glycosidases. We propose that this residue, as the imidic acid tautomer, acts analogously to the key catalytic aspartate of GH56 and GH18 enzymes. This topographically conserved arrangement of the asparagine residue and a conserved glutamic acid, coupled with previous kinetic studies, suggests these enzymes may use an unusual proton shuttle to coordinate effective general acid and base catalysis to aid cleavage of the glycosidic bond. These results collectively provide a blueprint that may be used to facilitate protein engineering of these enzymes to improve their function as biocatalysts for synthesizing glycoproteins having defined glycoforms and also may serve as a guide for generating inhibitors of GH85 enzymes.

**INTRODUCTION**

Both symbiotic and pathogenic bacteria possess diverse arrays of carbohydrate processing enzymes that are used to degrade a wide range of host glycoconjugates. *Streptococcus pneumoniae* (the pneumococcus), a gram positive, encapsulated pathogen that can cause acute...
respiratory disease (pneumonia), otitis media, meningitis, and several other serious diseases, is responsible for causing millions of deaths annually in the young and the elderly and relies on its carbohydrate processing enzymes for its full virulence. The ability of this bacterium to degrade hyaluronic acid and sialic acid-containing glycans in the host has long been correlated with the virulence of this pathogen (1). More recently, genome sequencing studies combined with large-scale efforts directed towards screening for new virulence factors have suggested that a large and diverse array of carbohydrate degrading enzymes is necessary for full virulence of the pneumococcus (2,3). Among this arsenal of carbohydrate processing enzymes are a group known as endohexosaminidases; enzymes that act to cleave the glycosidic linkage between the two N-acetylglucosamine units that make up the chitobiose core of N-glycans (Fig. 1a).

The functional role of pneumococcal carbohydrate-degrading enzymes is generally thought to be to harvest nutrients to promote bacterial growth (4,5). Endohexosaminidase D, Endo-D, produced by this organism, however, is also thought to contribute to virulence by playing a role in the deglycosylation of IgG molecules, decreasing the ability of the antibodies to interact with complement and thus compromising the host immune response (6).

There appears to be two main classes of endohexosaminidases. One group of endohexosaminidases is evolutionarily related to chitinases and these enzymes are found in the same family of glycoside hydrolases, GH18 (for a discussion of the CAZY classification system of the glycoside hydrolases see (7)). The second group, to which Endo-D belongs, is found in family GH85 of glycoside hydrolases and is of considerable interest owing to its potential use in biosynthesis of complex N-glycans. Elegant studies have detailed the ability of these enzymes to efficiently transfer en bloc the N-core of an appropriate carbohydrate donor to an N-acetylglucosaminylated asparagine residue of a polypeptide backbone and thereby chemoenzymatically generate glycoproteins having precisely defined glycoforms (8-10). More recent efforts have shown that these enzymes can act on a range of glycans having a sugar oxazoline at the reducing end as a glycosyl donor (11-14). These sugar oxazolines (Fig. 1b) are intermediates within the catalytic cycle of glycoside hydrolases from families 18, 20, 56, and 84. Enzymes from families 18, 20, 56, and 84 have previously been shown to use substrate assisted catalysis through both detailed kinetic studies (15-18) as well as X-ray structural studies (19-23). Therefore, the ability of GH85 endohexosaminidases to use sugar oxazoline donors strongly suggests that GH85 enzymes also use a substrate-assisted catalytic mechanism.

The substrate-assisted catalytic mechanism used by these four families of enzymes results in cleavage of the glycosidic bond with retention of configuration in two discreet steps. Two enzymic carboxylic acid residues play key catalytic roles (18,24-26). In the first step of the mechanism a carboxylic acid residue that hydrogen bonds to the glycosidic oxygen acts as a general acid, facilitating departure of the leaving group. Expulsion of the leaving group is also aided by concomitant attack of the carbonyl oxygen of the 2-acetamido group of the substrate. The acetamido group is oriented, and its nucleophilicity enhanced, through donation of a hydrogen bond to a second suitably positioned carboxylate residue. This residue has been postulated to act as a general base to aid formation of an oxazoline intermediate (26) or, alternatively, to stabilize an oxazolinium ion intermediate (23,27). The resulting intermediate is broken down through a process that is the near microscopic reverse of the first step; the general acid/base residue now facilitates the attack of water while the second residue facilitates departure of the 2-acetamido group to form the sugar hemiacetal product with overall retention of stereochemistry.
Like other glycoside hydrolases using substrate-assisted catalysis, the endohexosaminidases also process substrates having a 2-acetamido group. Detailed mutagenesis and chemical rescue studies have provided good support for a completely conserved carboxylic acid acting as the general acid/base residue (28). However, based on sequence alignments of known GH85 enzymes, there is no obvious carboxylate residue that might serve to enhance the nucleophilicity of the acetamido group. Very recently, Yamamoto and coworkers have proposed that a conserved asparagine fulfills the same function since they observed that deletion of the side chain by mutagenesis resulted in a dramatic loss of activity (29). Interestingly, endohexosaminidases in which either of these residues are mutated in a fairly conservative manner, are still able to catalyze the second step of the reaction; using oxazoline substrates as donor sugars and transferring the saccharide to a suitable acceptor (29,30). On the basis of these results, Yamamoto and coworkers have proposed that these enzymes use substrate-assisted catalysis (11,29). Given the non-canonical active site residues found in these enzymes, they also emphasized the need for more rigorous kinetic support and structural studies (29).

Here we describe mechanistic studies of *Streptococcus pneumoniae* Endo-D using small molecule probes (Fig. 1c,d) in conjunction with high-resolution X-ray crystallographic analysis of the uncomplexed enzyme and the enzyme in complex with an inhibitor that mimics either the oxazoline intermediate or a closely derived transition state. These data collectively provide compelling support for GH85 using a mechanism involving substrate-assisted catalysis. They also furnish detailed insight into the roles of active site residues and how these enzymes effect catalysis, suggesting that the GH85, GH18, GH56 enzymes, which have a similar configuration of catalytic residues, may use an unusual proton shuttle to coordinate effective general acid and base catalysis with concomitant cleavage of the glycosidic bond. Lastly, these findings provide a blueprint that may be used to facilitate protein engineering of these enzymes to improve their function as biocatalysts for synthesizing glycoproteins having defined glycoforms (13,30), a desirable feature of potential therapeutic biologicals (31).

**EXPERIMENTAL PROCEDURES**

**Cloning.** The gene fragment encoding the predicted catalytic module (nucleotides 475-2421; amino acids 159-807 of the complete protein) of Endo-D (here called SpGH85) was amplified by PCR from *S. pneumoniae* TIGR4 genomic DNA (American Type Culture Collection BAA-334D). The forward primer was GH85F (5’ - TATATACATATGGAAAAAGAAGTCA GCCAGATGTCCCTAAA -3’), which was engineered to have a 5’ *Nhe* I restriction endonuclease site, and the reverse primer was GH85R (5’- GCGCGCCTCGAGTTAATCTTTTACATACCCCAATCAAA -3’), which had a 3’ *Xho* I restriction site. The amplified DNA fragment was cloned into pET28a (Novagen) via the *Nhe* I and *Xho* I restriction sites using standard molecular biology procedures to generate pGH85. The DNA sequence of the resulting construct was verified by bidirectional DNA sequencing. The resulting gene fusion encoded a six-histidine tag fused to the SpGH85 catalytic domain by an intervening thrombin protease cleavage site.

**Protein production and purification.** SpGH85 was produced in *E. coli* BL21 Star (DE3) cells (Invitrogen) using LB media supplemented with 50 µg ml⁻¹ of kanamycin. Cultures were grown at 37°C until they reached an optical density at 600 nm of 0.5-0.7 whereupon protein production was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After further incubation at 37°C for 4 hours the cells were harvested by centrifugation and ruptured by chemical lysis. SpGH85 was purified by
Ni^{2+} immobilized metal affinity chromatography followed by size exclusion chromatography (SEC) using a Sephacryl S-200 column (GE Biosciences). SEC was performed using 20 mM Tris HCl, pH 8.0. Selenomethionine labelled SpGH85 (SeMet-SpGH85) was prepared using the procedures described previously (32) and purified as for the native protein. Protein concentration was determined by measuring the absorbance at 280 nm using the calculated molar extinction coefficient of 155160 cm^{-1} M^{-1} (33).

Chemical Synthesis.

Synthesis of 1,3,4,6-tetra-O-acetyl-2-deoxy-2-fluoroacetamido-\(\beta\)-D-glucopyranoside - Triethylamine (0.8 mL) and dry pyridine (20 mL) were added to a cooled (0 °C) solution of 2-amino-2-deoxy-1,3,4,6-tetra-O-acetyl-\(\beta\)-D-glucopyranose hydrochloride (34) (5 g) in a solution of DMF (100 mL). Sodium fluoroacetate (1.8 g) was added to a stirred mixture of dry DMF (90 mL) containing dried Dowex 50-H\(^{+}\) resin (12 g). After 1 h, dicyclohexylcarbodiimide (DCC) (3.2 g) and 30 ml of the fluoroacetic acid solution were added via cannula to the reaction vessel containing the hydrochloride salt. The resulting solution was allowed to stand for 16 h at 0 °C, after which time the reaction was judged complete by TLC analysis. The solvent was partially removed in vacuo and ethyl acetate (300 ml) and a solution of saturated sodium chloride (100 mL) were added. The organic layer was collected, and the aqueous layer was extracted twice with ethyl acetate. The combined organic extracts were washed successively with water, twice with saturated sodium bicarbonate, and finally with a solution of saturated sodium chloride. The organic extracts were dried over MgSO\(_4\) and filtered, and the solvent was removed in vacuo to yield colorless syrup. The desired product was then recrystallized from ethyl acetate and hexanes and the product was obtained with a 77 % yield. \(^1\)H NMR (500 MHz, CD\(_3\)OD) \(\delta\): 5.83 (1H, d, J\(_{H1,H2}\) = 8.8 Hz, H-1), 5.34 (1H, dd, J\(_{H3,H4}\) = 9.3 Hz, H-3), 5.00 (1H, dd, J\(_{H4,H5}\) = 10.0 Hz, H-4), 4.74 (2H, dd, J\(_{H7,F}\) = 47.0 Hz, H-7), 4.26 (1H, dd, J\(_{H6,H6'}\) = 12.5 Hz, H-6), 4.10 (1H, dd, J\(_{H12,H13}\) = 10.5 Hz, H-2), 4.07 (1H, dd, H-6'), 3.89 (1H, d, J\(_{H15,H16}\) = 4.6 Hz, J\(_{H15,H16}\) = 2.8 Hz, H-5), 2.04 (3H, s, OAc), 2.01 (3H, s, OAc), 1.98 (3H, s, OAc), 1.95 (3H, s, OAc) ppm.

Synthesis of 1,3,4,6-tetra-O-acetyl-2-deoxy-2-difluoroacetamido-\(\beta\)-D-glucopyranoside - Triethylamine (0.8 mL) and dry pyridine (20 mL) were added to a cooled (0 °C) solution of 2-amino-2-deoxy-1,3,4,6-tetra-O-acetyl-\(\beta\)-D-glucopyranose hydrochloride (5 g) in a solution of DMF (100 mL). Dicyclohexylcarbodiimide (DCC, 3 g) and difluoroacetic acid (1.2 mL) were added to the reaction mixture via syringe. The resulting solution was allowed to stand for 16 h at 0 °C, after which time another 0.5 mL of difluoroacetic acid were added. After a further 3.5 h at room temperature, the reaction was judged complete by TLC analysis. The solvent was partially removed in vacuo and ethyl acetate (300 ml) and a solution of saturated sodium chloride (100 mL) were added. The organic layer was collected, and the aqueous layer was extracted twice with ethyl acetate. The combined organic extracts were washed successively with water, twice with saturated sodium bicarbonate, and finally with a solution of saturated sodium chloride. The organic extracts were dried over MgSO\(_4\) and filtered, and the solvent was removed in vacuo to yield colorless syrup. The desired product was then recrystallized from ethyl acetate and hexanes and the product was obtained with an 83 % yield. \(^1\)H NMR (500 MHz, CD\(_3\)OD) \(\delta\): 4.96 (1H, t, J\(_{H7,F}\) = 53.6 Hz, H-7), 5.78
(1H, d, J$_{11,12}$ = 8.9 Hz, H-1), 5.28 (1H, dd, J$_{13,14}$ = 9.3 Hz, H-3), 4.98 (1H, dd, J$_{14,15}$ = 10.0 Hz, H-4), 4.23 (1H, dd, J$_{16,17}$ = 12.5 Hz, H-6), 4.06 (1H, dd, J$_{17,18}$ = 10.5 Hz, H-2), 4.05 (1H, dd, H-6'), 3.89 (1H, ddd, J$_{18,19}$ = 2.3 Hz, J$_{19,20}$ = 4.5 Hz, H-5), 2.01 (3H, s, OAc), 1.99 (3H, s, OAc), 1.95 (3H, s, OAc), 1.92 (3H, s, OAc) ppm.

Synthesis of 2-deoxy-1,3,4,6-tetra-O-acetyl-2-trifluoroacetamido-β-D-glucopyranoside - Triethylamine (0.8 mL) was added to a solution of 2-amino-2-deoxy-1,3,4,6-tetra-O-acetyl-β-D-glucopyranose chloride (1 g) dissolved in dry dichloromethane (20 ml) and cooled to 0 °C. Trifluoroacetic anhydride (0.6 mL) was added via syringe and the resulting solution was allowed to stand for 16 h at 0 °C, after which time the reaction was judged complete by TLC analysis. The solution was dilute in 50 ml of ethyl acetate and cooled (0 °C). Upon completion of the reaction the solvent was partially removed in vacuo to 1/3 of the volume. Dichloromethane (40 mL) was added and this solution was washed extensively with a saturated solution of sodium bicarbonate, water, and then saturated sodium chloride. The organic extracts were dried over MgSO$_4$ and filtered, and solvent was removed in vacuo. The compounds were not susceptible to crystallization, however, analysis of the products by NMR showed minimal contamination with yields ranging from 70-90%. Therefore, the desired compounds were used in the following steps without further purification. $^1$H NMR (500 MHz, CDCl$_3$) δ: 6.62 (1H, d, J$_{11,12}$ = 6.6 Hz, H-N), 6.21 (1H, d, J$_{13,14}$ = 3.7 Hz, H-1), 5.34 (1H, dd, J$_{14,15}$ = 9.3 Hz, H-3), 5.00 (1H, dd, J$_{15,16}$ = 10.0 Hz, H-4), 4.79 (2H, ABq, J$_{18,19}$ = 47.2 Hz, H-7), 4.26 (1H, dd, J$_{16,17}$ = 12.5 Hz, H-6), 4.10 (1H, dd, J$_{17,18}$ = 10.5 Hz, H-2), 4.07 (1H, dd, H-6'), 3.89 (1H, ddd, J$_{18,19}$ = 6.6 Hz, H-6'), 4.10 (1H, dd, J$_{19,20}$ = 2.3 Hz, H-5), 2.04 (3H, s, OAc), 2.11 (3H, s, OAc), 2.06 (3H, s, OAc), 2.05 (3H, s, OAc) ppm.

3,4,6-Tri-O-acetyl-2-deoxy-2-difluoroacetamido-β-D-glucopyranosyl chloride - $^1$H NMR (500 MHz, CDCl$_3$) δ: 6.73 (1H, d, H-N), 6.20 (1H, d, J$_{11,12}$ = 3.8 Hz, H-1), 5.89 (1H, t, J$_{11,12}$ = 53.9 Hz, H-7), 5.34 (1H, dd, J$_{13,14}$ = 9.3 Hz, H-3), 5.00 (1H, dd, J$_{14,15}$ = 10.0 Hz, H-4), 4.26 (1H, dd, J$_{16,17}$ = 12.5 Hz, H-6), 4.10 (1H, dd, J$_{17,18}$ = 10.5 Hz, H-2), 4.07 (1H, dd, H-6'), 3.89 (1H, ddd, J$_{18,19}$ = 6.6 Hz, H-6'), 4.10 (1H, dd, J$_{19,20}$ = 2.3 Hz, H-5), 2.04 (3H, s, OAc), 2.11 (3H, s, OAc), 2.06 (3H, s, OAc), 2.05 (3H, s, OAc) ppm.

General procedure for the synthesis of the 3,4,6-tri-O-acetyl-2-deoxy-2-acetamido-β-D-glucopyranosyl chlorides - In a similar manner that has been described previously for the non-fluorinated derivative(35), acetic anhydride (10 mL) was added to 1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-fluoroacetamido-β-D-glucopyranose (1 g) and cooled (0 °C). HCl$_3$was added until the solution was saturated and the reaction was allowed to proceed for 16 h at room temperature. As the reaction proceeded the starting material dissolved into the solvent. Upon completion of the reaction the solvent was partially removed in vacuo to 1/3 of the volume. Dichloromethane (40 mL) was added and this solution was washed extensively with a saturated solution of sodium bicarbonate, water, and then saturated sodium chloride. The organic extracts were dried over MgSO$_4$ and filtered, and solvent was removed in vacuo. The compounds were not susceptible to crystallization, however, analysis of the products by NMR showed minimal contamination with yields ranging from 70-90%. Therefore, the desired compounds were used in the following steps without further purification. $^1$H NMR (500 MHz, CDCl$_3$) δ: 5.81 (1H, d, J$_{11,12}$ = 8.9 Hz, H-1), 5.31 (1H, dd, J$_{13,14}$ = 9.3 Hz, H-3), 5.03 (1H, dd, J$_{14,15}$ = 10.0 Hz, H-4), 4.26 (1H, dd, J$_{16,17}$ = 12.5 Hz, H-6), 4.09 (1H, dd, J$_{17,18}$ = 10.5 Hz, H-2), 4.08 (1H, dd, H-6'), 3.91 (1H, ddd, J$_{18,19}$ = 4.5 Hz, J$_{15,16}$ = 2.3 Hz, H-5), 2.04 (3H, s, OAc), 2.02 (3H, s, OAc), 1.98 (3H, s, OAc), 1.94 (3H, s, OAc) ppm.

Synthesis of aryl 2-acetamido-2-deoxy-3,4,6-tri-O-acetyl-β-D-glucopyranosides - To a mixture of the respective fluorinated 2-N-acetyl-3,4,6-tri-O-acetyl-α-D-glucopyranosyl chloride (1 eq.), benzyltriethylammonium chloride (1 eq.), and the acceptor phenol (2 eq.) was added to
sufficient dichloromethane (1 volume) to yield a solution of 200 mM. An equal volume of 1 M NaOH was then added and the resulting mixture was stirred vigorously at room temperature. Ethyl acetate (5 volumes) was then added and the resulting organic phase was successively washed with 1 M NaOH (2 x 1 volume), water (1 volume), and saturated sodium chloride solution (1 volume). The organic layer was dried (MgSO₄), filtered, and concentrated. These materials were recrystallized using a mixture of ethyl acetate and hexanes to yield the desired glycosides in yields ranging from 40 to 60 %.

**3-fluoro-4-nitrophenyl 2-deoxy-2-fluoroacetamido-3,4,6-tri-O-acetyl-β-D-glucopyranoside** - ¹H NMR (500 MHz, CD₃OD) δ: 8.09 (1H, dd, J_H5Ar-H6Ar ≈ J_F3Ar-H2Ar = 8.9 Hz, H-5Ar), 6.91 (1H, dd, J_H2Ar-H6Ar = 2.5 Hz, J_H2Ar-F3Ar = 12.3 Hz, H-2Ar), 6.87 (1H, dd, J_H6Ar-HF = 0.9 Hz, H-6Ar), 6.62 (1H, dd, J_H7-F = 8.4 Hz, NH), 5.85 (1H, d, J_H1-H2 = 8.0 Hz, H-1), 5.55 (1H, dd, J_H3-H4 = 9.2 Hz, H-3), 5.13 (1H, dd, J_H4-H5 = 9.4 Hz, H-4), 4.76 (2H, dd, J_H7-F = 47.1 Hz, J_H7-H7' = 6.2 Hz, H-7), 4.26 (1H, dd, J_H6-H6' = 12.3 Hz, H-6'), 4.20 (1H, dd, H-6'), 4.12 (1H, dd, J_H2-H3 = 10.3 Hz, H-2), 4.00 (1H, ddd, J_H5-H6 = 6.1 Hz, J_H5-H6' = 2.6 Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), 2.07 (3H, s, OMe) ppm.

**3-fluoro-4-nitrophenyl 2-deoxy-2-difluoroacetamido-3,4,6-tri-O-acetyl-β-D-glucopyranoside** - ¹H NMR (500 MHz, CD₃OD) δ: 8.08 (1H, dd, J_H5Ar-H6Ar ≈ J_F3Ar-H2Ar = 8.8 Hz, H-5Ar), 6.92 (1H, dd, J_H2Ar-H6Ar = 2.5 Hz, J_H2Ar-F3Ar = 12.2 Hz, H-2Ar), 6.86 (1H, ddd, J_H6Ar-HF = 0.8 Hz, H-6Ar), 6.74 (1H, dd, J_NH-H2 = 8.6 Hz, NH), 5.86 (1H, t, J_H7-F = 54.0 Hz, H-7), 5.48 (1H, dd, J_H3-H4 = 9.2 Hz, H-3), 5.46 (1H, d, J_H1-H2 = 8.1 Hz, H-1), 5.14 (1H, dd, J_H4-H5 = 9.5 Hz, H-4), 4.26 (1H, dd, J_H6-H6' = 12.3 Hz, H-6'), 4.20 (1H, dd, J_H2-H3 = 10.4 Hz, H-2), 4.00 (1H, dd, J_H5-H6 = 6.0 Hz, J_H5-H6' = 2.7 Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), 3.08 (1H, s, OMe) ppm.

**Synthesis of 3-fluoro-4-nitrophenyl 2-deoxy-3,4,6-tri-O-acetyl-2-trifluoroacetamido-β-D-glucopyranoside** - The desired products were obtained from 3,4,6-tri-O-acetyl-2-deoxy-2-trifluoroacetamido-β-D-glucopyranosyl bromide (36) using AgCO₃ and the appropriate acceptor in a similar fashion as described previously (37).

**3-fluoro-4-nitrophenyl 2-deoxy-3,4,6-tri-O-acetyl-2-trifluoroacetamido-β-D-glucopyranoside** - ¹H NMR (500 MHz, CD₃OD) δ: 8.09 (1H, dd, J_H5Ar-H6Ar ≈ J_F3Ar = 8.9 Hz, H-5Ar), 6.92 (1H, dd, J_H2Ar-H6Ar = 2.6 Hz, J_H2Ar-F3Ar = 12.1 Hz, H-2Ar), 6.86 (1H, ddd, J_H6Ar-HF = 0.8 Hz, H-6Ar), 6.71 (1H, dd, J_NH-H2 = 8.7 Hz, NH), 5.41 (1H, d, J_H3-H4 = 9.2 Hz, H-3), 5.45 (1H, d, J_H1-H2 = 8.1 Hz, H-1), 5.16 (1H, dd, J_H4-H5 = 9.6 Hz, H-4), 4.28 (1H, dd, J_H2-H3 = 10.4 Hz, H-2), 4.26 (1H, dd, J_H6-H6' = 12.3 Hz, H-6'), 4.22 (1H, dd, H-6'), 3.99 (1H, d, J_H5-H6' = 5.9 Hz, J_H5-H6 = 2.7 Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), 3.08 (1H, s, OMe) ppm.

**General synthesis of aryl 2-N-acyl-2-deoxy-β-D-glucopyranosides** - To a stirred solution of the appropriate protected aryl glycosides in anhydrous methanol was added a spatula tip of sodium methoxide. The reaction mixture was then allowed to stir for approximately one hour at room temperature. After this time, the reaction was neutralized with Amberlite IR-120 resin (H⁺) and filtered. The filtrate was concentrated in vacuo and the desired products were recrystallized precipitate using a mixture of ethanol and ether. Yields of the desired product after one recrystallization ranged from 40 to 80 %. 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside (3F4NP-GlcNAc, 1a, Fig. 1b) is known previously and spectral characterization agrees with the literature (38).

**3-fluoro-4-nitrophenyl 2-deoxy-2-fluoroacetamido-β-D-glucopyranoside** (1b,
Fig. 1b) NMR (500 MHz, CD$_3$OD) δ: 8.08 (1H, dd, J$_{H5Ar-H6Ar} \approx$ J$_{H5Ar-F3Ar}$ = 9.1 Hz, H-5Ar), 7.02 (1H, dd, J$_{H2Ar-H6Ar}$ = 2.5 Hz, J$_{H2Ar-F3Ar}$ = 12.8 Hz, H-2Ar), 6.95 (1H, ddd, J$_{H6Ar-HF}$ = 1.1 Hz, H-6Ar), 5.25 (1H, d, J$_{H1-H2}$ = 8.4 Hz, H-1), 4.81 (2H, d, J$_{H7-F}$ = 47.1 Hz, H-7), 4.03 (1H, dd, J$_{H2-H3}$ = 10.3 Hz, H-2), 3.89 (1H, dd, J$_{H6-H6'}$ = 12.2 Hz, H-6'), 3.68 (1H, dd, J$_{H3-H4}$ = 2.5 Hz, H-3), 3.51 (1H, ddd, J$_{H5-H6}$ = 9.7 Hz, H-5), 3.40 (1H, dd, J$_{H4-H5}$ = 9.8 Hz, H-4 ppm.

3-fluoro-4-nitrophenyl 2-deoxy-2-difluoroacetamido-β-D-glucopyranoside (1c, Fig. 1b) - $^1$H NMR (500 MHz, CD$_3$OD) δ: 8.08 (1H, dd, J$_{H5Ar-H6Ar} \approx$ J$_{H5Ar-F3Ar}$ = 9.0 Hz, H-5Ar), 7.02 (1H, dd, J$_{H2Ar-H6Ar}$ = 2.5 Hz, J$_{H2Ar-F3Ar}$ = 12.8 Hz, H-2Ar), 6.94 (1H, ddd, J$_{H6Ar-HF}$ = 1.0 Hz, H-6Ar), 6.03 (1H, t, J$_{H7-F}$ = 53.9 Hz, H-7), 5.23 (1H, d, J$_{H1-H2}$ = 8.4 Hz, H-1), 3.98 (1H, dd, J$_{H2-H3}$ = 10.4 Hz, H-2), 3.90 (1H, dd, J$_{H6-H6'}$ = 12.1 Hz, H-6'), 3.68 (1H, dd, H-6'), 3.63 (1H, dd, J$_{H3-H4}$ = 8.5 Hz, H-3), 3.51 (1H, ddd, J$_{H5-H6}$ = 2.2 Hz, J$_{H5-H6'}$ = 6.0 Hz, H-5), 3.40 (1H, dd, J$_{H4-H5}$ = 9.6 Hz, H-4) ppm.

3-fluoro-4-nitrophenyl 2-deoxy-2-trifluoroacetamido-β-D-glucopyranoside (1d, Fig. 1b) - $^1$H NMR (500 MHz, CD$_3$OD) δ: 8.08 (1H, dd, J$_{H5Ar-H6Ar} \approx$ J$_{H5Ar-F3Ar}$ = 9.0 Hz, H-5Ar), 7.02 (1H, dd, J$_{H2Ar-H6Ar}$ = 2.5 Hz, J$_{H2Ar-F3Ar}$ = 12.8 Hz, H-2Ar), 6.94 (1H, ddd, J$_{H6Ar-HF}$ = 1.0 Hz, H-6Ar), 5.20 (1H, d, J$_{H1-H2}$ = 10.4 Hz, H-1), 3.97 (1H, dd, J$_{H2-H3}$ = 10.4 Hz, H-2), 3.90 (1H, dd, J$_{H6-H6'}$ = 12.1 Hz, H-6'), 3.69 (1H, dd, H-6'), 3.62 (1H, dd, J$_{H3-H4}$ = 8.6 Hz, H-3), 3.51 (1H, ddd, J$_{H5-H6}$ = 2.1 Hz, J$_{H5-H6'}$ = 5.9 Hz, H-5), 3.40 (1H, dd, J$_{H4-H5}$ = 9.7 Hz, H-4) ppm.

Enzyme assays. All kinetics was done using 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside (3F4NP-GlNAC, 1a, Fig. 1b) and were carried out in a manner similar to that described previously (38). Briefly, analysis of the pH-dependence was done in a buffer consisting of 50 mM sodium citrate, 50 mM sodium phosphate, 50 mM Ches, and 100 mM NaCl adjusted to the appropriate pH. Reactions were monitored continuously on a UV-Vis spectrophotometer at 37 °C. For kinetic analysis on the importance of the 2-acetamido group in catalysis, 3-fluoro-4-nitrophenyl 2-deoxy-2-fluoroacetamido-glucopyranoside, 3-fluoro-4-nitrophenyl 2-deoxy-2-difluoroacetamido-glucopyranoside, and 3-fluoro-4-nitrophenyl 2-deoxy-2-trifluoroacetamido-glucopyranoside, were assayed at a pH of 7.4. All assays contained between 1-20 μM of enzyme and rates were taken over the first 10 minutes where rates were linear.

NMR experiments. $^1$H NMR spectroscopy (600 MHz Bruker AMX spectrometer) was used to follow the progress and identify the products of the enzyme catalyzed reaction. The reaction was carried out in 0.5 ml containing 5 mM 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy-glucopyranoside in buffer containing 50 mM sodium phosphate and 100 mM sodium chloride which had been dissolved in D$_2$O, evaporated, and redissolved in 99.9% D$_2$O (Cambridge Isotopes) and adjusted to a pD of 7.81. Initiation of the reaction was done by the addition of the appropriate amount of lyophillized enzyme to a concentration of 400 μM. The reaction was placed at 37 °C immediately and placed in the NMR spectrometer at the appropriate time to acquire a spectrum. During acquisition, the temperature of the spectrometer was maintained at 20 °C to optimize the chemical shift of the water signal as well as to slow down mutarotation. All spectra represent 16 scans.

Isothermal titration calorimetry. ITC was performed with a VP-ITC (MicroCal, Northampton, MA) as described previously (39). SpGH85 was dialyzed extensively against 50 mM HEPES, pH 7.5, and its concentration determined by absorbance at 280 nm. 1,2-dideoxy-2'-methyl-α-D-glucopyranosyl[2,1-Δ2'-thiazoline (NAG-thiazoline) (16) was synthesized previously.
described and solutions were prepared by mass in buffer saved from the dialysis. All solutions were filtered and degassed prior to use. Twenty-five 10 µL injections of 2.1 mM NAG-thiazoline were titrated into 100 µM SpGH85. The heat of dilution was determined by an independent titration of NAG-thiazoline into buffer. The rate of complex formation was sufficiently slow to require spacings between inhibitor injections of 20-45 minutes. Due to the low C-value (40) of only ~3 that could be attained for the experiments the stoichiometry was fixed at 1 during fitting of the data (allowing the stoichiometry to be fit yielded a value of ~0.8). This approach was justified by the crystallographic data, which revealed a single NAG-thiazoline bound in the SpGH85 active site. The equilibrium association constant was determined by fitting the heat of dilution corrected data to a one-site binding model.

Structure determination. Crystals of SpGH85 and SeMet-SpGH85, both at 20 mg/ml, were obtained by the hanging-drop vapour diffusion method at 18 °C in 20% (v/v) polyethylene glycol (PEG) 3350, 0.2 M MgCl₂, 0.1 M Tri-Na citrate, pH 5.0. After a 10 second soak in crystallization solution supplemented with 15-20% (v/v) ethylene glycol as a cryoprotectant, crystals were flash cooled directly in a nitrogen stream at 113 K. Crystals of SpGH85 in complex with NAG-thiazoline were prepared by preincubating SpGH85 (20 mg/ml) with 60 mM NAG-thiazoline for 3 hours. This protein was used to set up crystallization experiments under the same conditions as for the uncomplexed protein. Because citrate molecules were observed in the active site of SpGH85 structures determined from different data sets, prior to data collection the SpGH85 crystals grown with NAG-thiazoline were soaked in a solution of 20% (v/v) PEG 3350, 0.2 M MgCl₂, 0.1 M sodium acetate, pH 5.0, containing 72 mM NAG-thiazoline. These crystals were cryoprotected in the soaking solution supplemented with 15% ethylene glycol.

Diffraction data were collected using a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic “blue” optics and an Oxford Cryostream 700 or on beamline X8C at the National Synchrotron Light Source (NSLS, Brookhaven National Laboratories, NY). Data was processed using Crystal Clear/d*trek (41). All data collection and processing statistics are shown in Table 1. The structure of SpGH85 was solved by single isomorphous replacement using a native data set collected to 1.8 Å and a SeMet-SpGH85 data set collected to 2.0 Å. Though a higher resolution native data set was obtained this proved to be non-isomorphous with the SeMet-SpGH85 crystals. The positions of the 8 selenium atoms expected for the single SpGH85 molecule in the asymmetric unit were determined with ShelxC/D (42). Initial phases were produced by refinement of the selenium substructure parameters with SHARP (43) followed by phase improvement with DM (44). Using the phases output from DM, ARP/wARP (45) was able to build a nearly complete model of SpGH85 with docked sidechains. This initial model was used with the high-resolution native data set to complete the model by successive rounds of building using COOT (46) and refinement with REFMAC (47). The structure of the NAG-thiazoline complex was solved using the native SpGH85 structure as a starting point and the model was completed by successive rounds of building using COOT and refinement with REFMAC. In all cases, the addition of water molecules was performed by the ARP/wARP function in REFMAC and manually checked. In both datasets, refinement procedures were monitored by flagging 5% of all observation as ‘free’ (48). Model validation was performed with SFCHECK (49) and PROCHECK (50). All model statistics are shown in Table 1. The coordinates and structure factors for SpGH85 and the SpGH85 NAG-thiazoline
complex have been deposited with the PDB codes 2w91 and 2w92, respectively.

RESULTS AND DISCUSSION

Enzyme kinetics studies. Endohexosaminidases have been the subject of considerable recent research activity. A number of elegant studies have now demonstrated that these enzymes can transfer chemically synthesized sugar oxazoline donor substrates to peptides and proteins bearing N-linked N-acetylglucosamine (11,13,14,51). Using this chemoselective transglycosylation strategy a wide range of homogenous complex glycoconjugates has been generated. The catalytic competence of these sugar oxazolines suggests they are intermediates of the endohexosaminidase-catalyzed hydrolysis reaction; supporting the notion that these enzymes use a substrate-assisted catalytic mechanism involving participation of the substrate acetamido group (11,29). To test this proposal in greater detail, and potentially provide compelling support for it, we set out to clone the catalytic domain of Endo-D(52) from S. pneumonia TIGR4. Endo-D is a 1646 amino acid protein that comprises a number of identifiable modules or domains including a family 85 glycoside hydrolase domain, which we more precisely defined by amino acid sequence comparison with other family members in conjunction with previous truncation-activity studies(53) of the enzyme. Using this information, we cloned the gene fragment encoding amino acids 159-807 and overproduced the protein product in E. coli for biophysical and biochemical analysis. For simplicity, we will refer to the truncated recombinant form of Endo-D as SpGH85 throughout.

With pure SpGH85 in hand we first aimed to establish a convenient enzyme assay. endo-Glycosidases, including endohexosaminidases, are generally most proficient on their natural substrates, which are typically large and structurally complex glycans. This preference for complex glycans complicates mechanistic studies, since these structures are generally not readily accessible on the large scale required to generate significant quantities of chromogenic or fluorogenic substrates required for structure-activity studies. Several endo-glycosidases, however, have been shown to be capable of hydrolyzing simplified unnatural substrates having fewer saccharide residues (24,54), sometimes even monosaccharide substrates if they bear activated leaving groups (55). We therefore assayed SpGH85 against several aryl β-D-glucosaminides having different leaving groups and found that the enzyme showed fair activity toward 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside (3F4NP-GlcNAc, 1a; Table 2). With this convenient substrate we next aimed to evaluate the pH-activity profile of SpGH85 to determine the optimal pH for activity assays as well as to gain insight into the number of titratable enzymic residues involved in catalysis. The pH-activity relationship of SpGH85 could be nicely fitted by a classical bell-shaped pH-activity profile (Fig. 2a), suggesting that the protonation states of two residues are critical in catalysis. Given that the pKₐ values of the two residues being titrated are well dispersed, we obtained good estimates of their kinetic pKₐ values. The basic limb reveals one residue (pKₐ ≈ 9.2 ± 0.1) that should be in its protonated form to enable catalysis; this limb is typically assigned to the general acid catalyst found in glycoside hydrolases using substrate-assisted catalysis (24,26). This kinetic pKₐ value appears high for a carboxylic acid, however, it has been established in studies of glycoside hydrolases that such perturbations in pKₐ values are common and are mediated by the electrostatic environment of the active site (56). The acidic limb reveals a second residue (pKₐ ≈ 5.7 ± 0.1) that is likely in its deprotonated state to enable catalysis; in glycosidases using substrate-assisted catalysis this residue is typically assigned to the second carboxylate that hydrogen bonds with the substrate amide (24,25). In the absence of further studies these titrations cannot be definitively assigned to specific
residues. It is notable, however, that GH85 enzymes have an asparagine in place of the carboxylate residue to which the acidic limb of the pH-activity profile is typically assigned.

Regardless, having a suitable substrate and knowing the pH-optimum of the SpGH85-catalyzed reaction (pH 7.0-8.0), we set out to establish the stereochemical outcome of the reaction catalyzed by this enzyme using proton nuclear magnetic resonance. Although the turnover of 3F4NP-GlcNAc is relatively slow, complicating the experiment and necessitating high concentrations of enzyme, we found on the basis of a time course of the reaction that SpGH85 cleaves this substrate to first form the β-hemiacetal of N-acetylglucosamine (see supplementary Figure S1). This stereochemical assignment is based on the chemical shift and 1,2 vicinal coupling of the anomeric proton of the first formed β-hemiacetal product and the later formed α-hemiacetal product, which both closely match values reported in the literature (39). Thus, SpGH85 carries out hydrolysis with retention of stereochemical configuration at the anomeric center, consistent with GH85 enzymes using the proposed catalytic mechanism involving substrate-assisted catalysis.

To directly assess the role of the substrate acetamido group we synthesized several analogues of 3F4NP-GlcNAc having various number of fluorine atoms substituted in place of hydrogen at the methyl group of the acetamido moiety. For all four substrates we observe Michaelian enzyme kinetics (Fig. 2b, Table 2) enabling us to determine the Michaelis-Menten parameters $K_M$ and $V_{max}$. By plotting log($V_{max}$/[E],$K_M$), which is proportional to the second-order rate constant, against the Taft electronic parameter ($\sigma^*$) of the N-acyl group we find a negative linear correlation with the number of fluorine groups (Fig. 2c). This negative linear correlation supports the proposed role of the carbonyl oxygen acting as the nucleophile to attack the anomeric center and also suggests that, as generally proposed for glycoside hydrolases (57), the mechanism involves electrophilic migration of the anomeric center. The slope of the correlation ($\rho = -0.31$) is a measure of the sensitivity of the reaction to different substituents.

The reaction parameter ($\rho$) we obtain for SpGH85 resembles those measured using related fluorine-containing substrates first for enzymes from unknown families (58,59), and more recently for GH families 18, 20, and 84, which all use substrate-assisted catalysis. The value we observe most closely resembles that seen for GH84 human O-GlcNAcase ($\rho$ ranging from -0.42 to -1.6) (15,19) but is significantly less that those observed for GH20 enzymes ($\rho^*$ ranging from -1.0 to -1.29)(15,37). The good correlation we observe here, coupled with the similarity of $\rho$ value to values observed for enzymes from GH84 provides strong kinetic evidence that these enzymes share in common a catalytic mechanism involving substrate assistance.

**Inhibition of SpGH85 by NAG-thiazoline.** NAG-thiazoline (Fig. 1d) is a potent inhibitor of glycoside hydrolases from GH20(15,16) and 84(15) by virtue of its mimicry of the enzyme catalyzed transition state(60) or the high-energy oxazoline intermediate (Fig. 1b). Recent studies have also shown that the related GH85 enzymes, Endo-A and Endo-M are strongly inhibited ($K_i = 220$ and 450 nM respectively) by a thiazoline derived from a decasaccharide N-glycan fragment (61). Given that SpGH85 is able to hydrolyze the monosaccharide substrate 3F4NP-GlcNAc, we predicted that NAG-thiazoline itself should bind to the active site. We therefore carried out both inhibition and isothermal titration calorimetry (ITC) studies of NAG-thiazoline binding to SpGH85. Our kinetic assays were hindered by the requirement for large concentrations of enzyme (1-10 μM) to obtain useful rates, which in turn complicated assaying inhibitor concentrations lower than 50 μM. On the basis of these preliminary studies, however,
we estimate a $K_i$ value of less than 50 $\mu$M (data not shown). To circumvent these issues and accurately measure the binding of NAG-thiazoline to SpGH85 we turned to using ITC. We found that NAG-thiazoline bound to SpGH85 with a $K_d$ of 28 ± 4 $\mu$M, a value that is consistent with the inhibition studies (see supplementary Figure S2). This relatively strong binding is somewhat surprising given that the structurally more complex (Man)$_9$-GlcNAc-thiazoline, which more faithfully resembles the natural substrate, inhibited the related enzymes Endo-A, Endo-M, and human ENGase with $IC_{50}$ values in the low micro to high nanomolar range (61). Although that study involves different enzymes and detail $IC_{50}$ values rather than true dissociation constants, the results tentatively suggest that saccharide residues binding at subsites beyond the core GlcNAc residue likely enhance binding by roughly 100-fold in GH85 enzymes.

The current lack of a three-dimensional structure of a family 85 glycoside hydrolase prevents detailed molecular-level interpretation of these mechanistic observations. Therefore, to gain more precise insight into the catalytic mechanism, coordinating these mechanistic findings and previous mutagenesis studies of Endo-D, Endo-A, and Endo-M, with detailed molecular features, we set out to obtain the structure of SpGH85. We also reasoned that, by using the knowledge obtained here regarding binding of NAG-thiazoline, we might be able to gain structural insight into the mechanism of SpGH85 while circumventing the need to obtain the structure of an enzyme-ligand complex or oligosaccharide substrate bound within the active site.

**Structure of SpGH85 and its catalytic module.** We were able to obtain high quality isomorphous crystals of SpGH85 and selenomethionine labeled SpGH85 (SeMet-SpGH85), allowing us to determine the structure of SpGH85 by single isomorphous replacement using data sets collected on our home source. Refinement and completion of the initial model using data to 1.4 Å provided us with a high-quality model of SpGH85 (Table 1). SpGH85 comprises 3 domains; an N-terminal $(\beta/\alpha)_8$-barrel with an incomplete complement of $\alpha$-helices followed in sequence by two $\beta$-sandwich domains, D1 and D2, containing two opposing 5-stranded and 4-stranded anti-parallel $\beta$-sheets, respectively (Fig. 3a). The $(\beta/\alpha)_8$-barrel of SpGH85 shows no primary structure identity with proteins outside of GH85 but, consistent with previous observations regarding their shared evolutionary ancestry, shows structural similarity with numerous members of GH family 18 as revealed by root-mean-square-deviations (rmsd) of 3.4-3.8 Å and Z-scores greater than 14 (as determined by DALI (62)). Structure similarity searches of the D1 domain revealed similarity to domains found in endoglucanases, mannanases, and chondroitin lyases; however, in general, the functions of these domains are unknown. Surprisingly, the highest structural identity of 2.4 Å over 134 matched C$\alpha$ (of 160 in D1) was obtained with a $\beta$-glucan binding family 44 carbohydrate-binding module (CBM44; PDB code 2c4x) from the Clostridium thermocellum endoglucanase CelJ (63). The conservation of some functional residues suggests a possible carbohydrate-binding function for the D1 domain (not shown). The last $\beta$-sandwich domain, D2, displays insignificant sequence identity with fibronectin domains but has low rmsds of 1.7-2.7 Å with proteins having a fibronectin fold (e.g. fibronectin, PDB code 1fnf; titin, PDB code 2nzi; and myosin binding protein C, PDB code 1x5y). Although FN3-like domains are commonly observed as components of carbohydrate processing enzymes, their functional role remains obscure.

Focusing on the $(\beta/\alpha)_8$-barrel domain, we find a pocket on one face within which the two completely conserved putative catalytic residues are positioned. The protein surface surrounding this center...
is highly featured, presenting a corrugated surface consistent with the known preference for Endo-D acting on branched high-mannose N-glycans (52) and suggesting these grooves may accommodate the glycan branches (Fig. 3b). Further structural studies, however, will be required to establish the mode of binding of the N-glycan to this surface and the structural determinants that govern substrate specificity of this enzyme.

Structure of SpGH85 in complex with NAG-thiazoline. Given that NAG-thiazoline binds surprisingly well to SpGH85, we set out to obtain crystals of the enzyme in complex with this inhibitor. Using a combined approach of co-crystallization with NAG-thiazoline followed by soaking in a solution containing a large excess of this inhibitor, we were able to obtain excellent diffraction data to 1.65 Å, which yielded clear electron density for a single molecule of NAG-thiazoline (Fig. 4a) bound within the pocket on the face of the (β/α)8-barrel that harbours the two conserved catalytic residues (Fig. 4b). This active site pocket cradling NAG-thiazoline must be the -1 saccharide binding subsite to which the N-acetylglucosamine residue of the chitiobiose core of the N-glycan of natural substrate is bound. This proposal is supported by the presence of the two catalytic residues and the obvious grooves that extend from this site across the surface of the (β/α)8-barrel to which the arms of the branched N-glycan most likely bind (Fig. 3b).

Despite the use of high concentrations of NAG-thiazoline and the unambiguous electron density for NAG-thiazoline, however, the presence of dual conformations of Tyr373 suggested to us that the active site was partially occupied by NAG-thiazoline. One conformation of Tyr373, which is the same as that in the uncomplexed protein, would result in steric clashes with the thiazoline group of the inhibitor and therefore prevent binding (Fig. 4c). The second conformation of Tyr373, clearly revealed by its F_{obs}-F_{calc} difference electron density when only the first conformation is modeled (Fig. 4b), is positioned to make optimal van der Waals interactions with the thiazoline group of NAG-thiazoline (Fig. 4d). On the basis of these observations, we modeled Tyr373 as having two conformations, each with an estimated 50% occupancy (Fig 4b). Accordingly, we also modeled NAG-thiazoline as having 50% occupancy of the active site and matched the occupied active site with the second conformation of Tyr373. This movement of Tyr373, the only obvious change in the binding site upon inhibitor binding, appears to be required to create a pocket at the bottom of the -1 subsite allowing this subsite to accommodate the methyl group of NAG-thiazoline (compare Figs. 4c and 4d).

The active site occupied by NAG-thiazoline is fairly open, consistent with the natural substrate being a complex glycan. Deeper within the pocket, however, the thiazoline ring of the substrate is encircled by several aromatic sidechains (Tyr373, Trp264, Phe471, Phe412) (Fig. 4b); a feature that has been observed within all families of GHs using substrate-assisted catalysis. The base of the pocket, which closely envelops the methyl group is also comprised of three hydrophobic residues (Leu229, Phe333, and Phe409). It is likely that these aromatic residues stabilize the transition state through favorable orbital overlap between the π system of the thiazoline ring, which on the basis of the Taft-like analysis must gain relative positive charge, and those of the aromatic residues. These residues also provide a well-defined surface that can properly recognize and orient the acetamido group of the substrate so that it can effectively participate in catalysis. For example, Tyr373 adopts a position placing its Oη 3.6 Å from the sulfur of the thiazoline ring, suggesting its normal function may be to orient and perhaps polarize the acetamido group of the substrate for efficient catalysis. Consistent with this view, mutation of this conserved tyrosine in Endo-A and Endo-M to phenylalanine
results in a dramatic loss of hydrolytic activity (29).

Hydrogen bonding between the enzyme and NAG-thiazoline is limited (Fig. 4b). O6 of NAG-thiazoline is appropriately positioned to hydrogen bond with Oε1 of Glu441. Asn335 of the enzyme appears to hydrogen bond with N2 of NAG-thiazoline, though this is slightly ambiguous. The orientation of the Asn335 amide group was modeled on the basis of B-factor analysis, which yielded one clear orientation that minimized the discrepancy between the Oδ1 and Nδ2 B-factors (Fig. 5). This orientation is also the preferred orientation in the uncomplexed SpGH85 structure (Fig. 5). The positioning of Asn335 placed Oδ1 at 3.2 Å from N2 of NAG-thiazoline, but with slightly less than ideal hydrogen bonding geometry. In this orientation, Nδ2 of Asn335 is 2.6 Å from N2 of NAG-thiazoline with ideal hydrogen bonding geometry, but, under normal circumstances, cannot accept a hydrogen bond (see below). Absent in this structure are interactions with O3 and O4, although it is possible that for large and complex substrates intramolecular hydrogen bonds may play an important role in recognition. The limited set of interactions presumably accounts for the micromolar affinity of NAG-thiazoline as compared to its nanomolar affinity for GH20 and GH84 (15,16,19). Notably, the sidechain of Glu337 is positioned to fulfill its putative function as the general acid/base in GH85 (Fig. 4b) and deliver a proton from the anti-protonation trajectory (64). Indeed, consistent with this assignment, this residue is structurally conserved among enzymes using substrate-assisted catalysis, where it functions as the general acid/base catalytic residue (7). Furthermore, when Glu337 is mutated to alanine in Endo-D the hydrolytic activity of the enzyme is abolished (53), an observation consistent with the deleterious effects of removing this carboxyl group in Endo-A and Endo-M (29,65).

Perhaps the most notable feature within the active site is the presence of the completely conserved asparagine residue, Asn335. This asparagine takes the place of an enzymic carboxylate that typically forms a hydrogen bond with the amide nitrogen of the acetamido group within enzymes from GH families 18, 20, 56, and 84 (19-23). Studies have variously proposed this carboxylate acts within these enzymes using substrate-assisted catalysis to stabilize an oxazolinium ion intermediate (23,27) or, alternatively, act as a general base to enhance the nucleophilicity of the acetamido group (15). Within GH85, however, previous studies into the functional consequences of deleting the side chain of this active site asparagine through site-directed mutagenesis have indicated that this conserved residue is absolutely critical; in the closely sequence-related Endo-M or Endo-A deletion results in apparently complete loss of activity toward a biantennary complex-type sialyl glycopeptide (29).

The role of Asn335. The structure of SpGH85 in complex with NAG-thiazoline described here, in conjunction with the pH-rate profile, offers a suggestion as to how this critically important asparagine residue may function in catalysis. The structure reveals that Asn335 forms a good hydrogen bond to the nitrogen of the thiazoline via one of its syn-lone pairs. Based on the B-factors obtained when modeling the two alternative orientations of the asparagine residue in both the inhibitor-enzyme complex and the free enzyme, we find that the most likely orientation of the amide moiety of Asn335 surprisingly positions the nitrogen, rather than the oxygen, in close proximity (2.6 Å) to N2 of the thiazoline ring. Since the amide nitrogen of asparagine cannot typically act as a hydrogen bond acceptor it is not immediately obvious as to how to assign a meaningful functional role for this residue in catalysis. The NH2 group of an amide side chain could not act to accept a hydrogen bond from the substrate acetamido group; therefore it might appear that this group would be unable to orient the acetamido group through hydrogen bonding, stabilize
an oxazolinium ion intermediate, or act as a general base.

The unexpected orientation of this amide group and its apparent inability to act in a catalytic manner in this tautomeric form suggested to us that, within the electrostatically well-defined enzyme active site, the imidic acid tautomer of Asn335 may be favoured. This tentative proposal readily accounts for several experimental observations.

An examination of the electron density of Asn335 in the high-resolution apo-structure reveals the highest density at the center we propose to be Oδ1, in keeping with this atom having a greater number of electrons than nitrogen, and coherent with the modeled conformation of this sidechain. Furthermore, the electron density at this atom is localized, which suggests that it forms only a single bond with Cγ. Consistent with this proposal is that the Oδ1-Cγ bond length is 1.39 Å when obtained using an unrestrained refinement. In contrast, delocalized electron density was observed between Cγ and Nδ2, which have an interatomic distance of 1.23 Å based on an unrestrained refinement, implying the presence of a double bond. By comparison, Asn441, which is a well-ordered asparagine very near but not in the active site and presumably in the amide form, displays the greatest electron density over Oδ1 but it is also delocalized over the 1.25 Å Oδ1-Cγ double bond (not shown). The Nδ2-Cγ single-bond length is 1.37 Å while the B-factor of Oδ1 is 1 Å less than for Nδ2. Thus, although not conclusive, the observed electron density and bond lengths of the Asn335 sidechain, in conjunction with its conformation established on the basis of B-factor analysis, are most consistent with the imidic acid tautomer of this sidechain.

Although residues titrated in the pH-activity study cannot be unambiguously assigned, there is a large body of evidence suggesting that in glycoside hydrolases the two limbs of bell-shaped pH-profiles typically arise from changes in the protonation state of the two key catalytic residues (66). The bell-shaped profile we observe for SpGH85 suggests that the activity outside of the optimum pH range approaches zero, consistent with titration of absolutely critical catalytic residues as seen in pH-profiles of other glycosidases (66). The basic limb having a kinetic pKₐ of 9.2 is most likely associated with titration of Glu441, the general acid/base catalytic residue. The assignment of the acidic limb is necessarily more speculative, however, we propose that it stems from protonation of the nitrogen of the imidic acid tautomer of Asn335.

This proposal is consistent with the need for this residue to have an available lone pair to facilitate catalysis through hydrogen bonding to amide nitrogen of the acetamido group. Furthermore, this proposal makes chemical sense since the pKₐ for N-deprotonation of an O-protonated imidic acid in solution is estimated at ≈ 7.5 whereas that of an amide is ≈ 18 (67). Accordingly, downward perturbation of the pKₐ of an amide moiety by over 12 units to generate a negatively charged amide anion seems most unlikely, particularly in the context of the hydrophobic and negatively charged environment of the active site. More reasonable is that the pKₐ of the protonated imidic acid tautomer would be perturbed to approximate the kinetic pKₐ of 5.7 seen in the pH profile. This tautomeric form of Asn335 is likely stabilized, and its pKₐ modulated, by its position immediately adjacent to, and hydrogen bonding with, the general acid catalytic residue (Fig. 4b).

Through acting as the imidic acid tautomer we posit that Asn335 fulfills an analogous role to the structurally conserved carboxylate found in other families of GHs using substrate-assisted catalysis. Instead of being negatively charged in the resting enzyme, however, this imidic acid would be neutral.

An alternative scenario in which the amide is oriented the other way, with the carbonyl oxygen accepting a hydrogen bond from the thiazoline ring and donating one to the general acid is also possible. Although,
based on our data, this appears less likely. Nevertheless, definitive assignment of the orientation and tautomeric state of Asn335 awaits further study.

A conserved proton shuttle. As the cyclization step of the reaction occurs, the \( pK_a \) of the amide proton of the substrate acetamido group goes from \( \sim 18 \) to \( \sim 5 \) for a putative oxazolinium ion intermediate (26,27). It therefore seems quite possible that in SpGH85, Asn335 in its imidic acid form, acts as a general base to facilitate catalysis since the \( pK_a \) of this group will be higher than that of the putative oxazolinium ion intermediate with which it forms a hydrogen bond. Furthermore, during the cyclization step of the reaction to form the oxazoline intermediate, the general acid Glu441 transfers its proton to the glycosidic oxygen and as this occurs the \( pK_a \) of the protonated imidic acid form of Asn335 must be perturbed upward. This proposal also makes good sense mechanistically since a neutral imidic acid cannot formally offer electrostatic stabilization to an oxazolinium ion intermediate, as proposed for the equivalent structurally conserved carboxylate in GH18 enzymes (23).

The structural conservation of an asparagine or aspartate in this position among GH18, 56, and 85 enzymes (Fig. 6a,b) suggests these residues may be tolerant to interchanging these residues. Notably, previous studies of Endo-D have shown that substitution of Asn335 with an aspartate only reduced the activity of Endo-D on (Man)_3 and (Man)_5 N-linked glycans by 2-5 fold (53). Deletion of the side-chain by mutagenesis to alanine, however, was deleterious. Furthermore, the opposite Asp to Asn mutation in GH18 ChiB from Serratia marcescens, revealed a less than a 10-fold decrease in the second-order rate constant (\( V_{max}/K_M \)), whereas substitution with alanine was deleterious (24).

Consistent with this proton shuttle proposal, we also observe that O\( \delta \)1 of Asn335 forms a 2.7 \( \AA \) hydrogen bond with O\( \varepsilon \)2 of the general acid residue Glu441. In SpGH85, the observation of this hydrogen bond in both the enzyme-inhibitor complex and the free enzyme lends further support that Asn335 operates as the imidic acid tautomer, since it would not be possible in both states if Asn335 was the amide tautomer oriented in the same way. Further, given that the \( pK_a \) of the general acid is predicted to be 9.2 while that of the \( N \)-protonated imidic acid of Asn335 is estimated at 5.7, it seems possible that even as it acts to deprotonate the forming oxazoline, the imidic acid of Asn335 could transfer the proton of O\( \delta \)1 to Glu441. If this shuttling of the proton from the amide group of the substrate through to the general acid occurs during catalysis, Asn335 is expected to exist as the amide tautomer in the enzyme-intermediate complex. More detailed studies at higher resolution, however, will be required to address this question.

The high level of structural conservation between these active site residues, despite GH85 enzymes having an asparagine in place of the aspartate found in GH18 and GH56, suggests the catalytic mechanism is conserved between these families of enzymes. This proposed mechanism (Fig. 6c) likely involves a cyclic series of bond making and bond breaking occuring in a concerted, but asynchronus, manner that is driven by cleavage of the glycosidic bond, formation of the oxazoline intermediate, and the consequent changes in \( pK_a \) of the groups involved. The intermediate operative in the catalytic mechanism of these families of enzymes is therefore likely an oxazoline and not an oxazolinium ion as is sometimes proposed (23). It seems quite possible that such a proton shuttle is conserved among enzymes from families 18, 56, and 85 since the topological arrangement of these two key catalytic residues, and their hydrogen bonding to each other, is highly conserved (Fig. 6a,b). Interestingly, mutation of Asp to Asn in GH84 and GH20 resulted in a greater than 1000-fold decrease in catalytic efficiency (19,25), highlighting the
differences between members of GH18, GH56, and GH85 versus those from GH20 and GH84. It is noteworthy that such a proton shuttle would effectively disperse charge across several residues, a favorable feature for realizing efficient catalysis.

Of course, based on the current data an alternative scenario, in which the amide is oriented the other way can be envisioned. This scenario would have the carbonyl oxygen interacting with the thiazoline ring. Although this possibility cannot be completely ruled out (Fig. 6d), on balance, the data currently suggest the other orientation. In this second formal scenario, Asn335 could act to orient the acetamido group but would not be able to stabilize the resulting oxazolinium ion intermediate through a formal charge-charge interaction. Furthermore, it is less likely Asn335 in this orientation could act as a general base since the $pK_a$ of the carbonyl of an amide is expected to be approximately 0 in solution (67). Regardless, further studies to address this putative proton shuttle and the role of the spatially conserved Asp and catalytically unusual Asn residue in these families of glycoside hydrolases await more definitive and detailed structural and mechanistic studies.

**Conclusion.** Our mechanistic and structural studies describing SpGH85, which contains the catalytic GH85 domain of Endo-D, offer unambiguous support for GH85 enzymes acting through a catalytic mechanism involving substrate-assisted catalysis. These studies also provide a framework for coherent assembly of a significant body of site-directed mutagenesis studies; offering a structural and mechanistic rationale for the importance of several key active site residues in binding and catalysis.

Interestingly, we observe the orientation of the two key catalytic residues of GH85 enzymes is structurally conserved in GH18 and 56 enzymes, even though these last two families have a carboxylate in place of the asparagine. The absence of this carboxylate in SpGH85, and GH85 endohexosaminidases more generally, argues against a role in stabilizing an oxazolinium ion intermediate, particularly in light of previous studies that have shown interchanging Asn and Asp residues at this position is not deleterious for either GH85 or GH18 enzymes. It appears more likely that this structurally conserved residue serves as a general base, the $pK_a$ of which is modulated by the general acid/base catalytic residue; a proposal that we favour more generally among the families of glycoside hydrolases using substrate-assisted catalysis.

We therefore propose a catalytic mechanism involving a concerted but asynchronous series of bond making and bond breaking that is driven by cleavage of the glycosidic bond and associated changes in the $pK_a$ values of key groups. Further support of this mechanistic proposal for GH85 awaits more detailed high-resolution structural studies.

The structure of the active site in complex with NAG-thiazoline, and its surprisingly potent binding, suggests a route to fairly small potent inhibitors of GH85 enzymes. The overall structure of SpGH85 also suggests sites at which more complex natural substrates may bind to Endo-D and homologous GH85 enzymes. These data therefore offer a valuable blueprint that could be used to select, or rationally design, mutant GH85 enzymes having optimized mechanistic features or altered substrate specificity that could be used in chemoenzymatic synthesis of homogenously N-glycosylated biomolecules; a topic of considerable importance to the pharmaceutical industry (31).
ACCESSION CODES
Protein Data Bank: Coordinates and structure factors have been deposited with accession codes of 2w91 (SpGH85) and 2w92 (SpGH85 NAG-thiazoline complex).

ACKNOWLEDGEMENTS
This work was supported by a grant from the Canadian Institutes of Health Research to ABB and DJV. ABB is a Canada Research Chair in Molecular Interactions and a Michael Smith Foundation for Health Research (MSFHR) Scholar. DJV is a Canada Research Chair in Chemical Glycobiology and a MSFHR Scholar. MSM is supported by a graduate scholarship from the Natural Sciences and Engineering Research Council of Canada and the MSFHR.

REFERENCES
1. Garcia-Suarez Mdel, M., Vazquez, F., and Mendez, F. J. (2006) Enferm Infecc Microbiol Clin 24, 512-517
2. 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., Feldblum, T. V., Anguoli, 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) Science 293, 498-506
3. Hava, D. L., and Camilli, A. (2002) Mol Microbiol 45, 1389-1406
4. Burnaugh, A. M., Frantz, L. J., and King, S. J. (2008) J Bacteriol 190, 221-230
5. King, S. J., Hippe, K. R., and Weiser, J. N. (2006) Mol Microbiol 59, 961-974
6. Koide, N., Nose, M., and Muramatsu, T. (1977) Biochem Biophys Res Commun 75, 838-844
7. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) Nucleic Acids Res
8. Yamamoto, K., Kadowaki, S., Watanabe, J., and Kumagai, H. (1994) Biochem Biophys Res Commun 203, 244-252
9. Takegawa, K., Tabuchi, M., Yamaguchi, S., Kondo, A., Kato, I., and Iwahara, S. (1995) J Biol Chem 270, 3094-3099
10. Mizuno, M., Haneda, K., Iguichi, R., Muramoto, I., Kawakami, T., Aimoto, S., Yamamoto, K., and Inazu, T. (1999) Journal of the American Chemical Society 121, 824-829
11. Fujita, M., Shoda, S., Haneda, K., Inazu, T., Takegawa, K., and Yamamoto, K. (2001) Biochim Biophys Acta 1528, 9-14
12. Rising, T. W., Claridge, T. D., Davies, N., Gamblin, D. P., Moir, J. W., and Fairbanks, A. J. (2006) Carbohydr Res 341, 1574-1596
13. Li, B., Song, H., Hauser, S., and Wang, L. X. (2006) Org Lett 8, 3081-3084
14. Li, B., Zeng, Y., Hauser, S., Song, H., and Wang, L. X. (2005) J Am Chem Soc 127, 9692-9693
15. Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D., and Vocadlo, D. J. (2005) J Biol Chem 280, 25313-25322
16. Knapp, S., Vocadlo, D., Gao, Z. N., Kirk, B., Lou, J. P., and Withers, S. G. (1996) Journal of the American Chemical Society 118, 6804-6805
17. Drouillard, S., Armand, S., Davies, G. J., Vorgias, C. E., and Henrissat, B. (1997) Biochemical Journal 328, 945-949
18. Arming, S., Strobl, B., Wechselberger, C., and Kreil, G. (1997) Eur J Biochem 247, 810-814
49. Vague, A. A., Richelle, J., and Wodak, S. J. (1999) Acta Crystallogr D Biol Crystallogr 55, 191-205
50. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Journal of Applied Crystallography 26, 283-291
51. Rising, T. W., Claridge, T. D., Moir, J. W., and Fairbanks, A. J. (2006) Chembiochem 7, 1177-1180
52. Muramatsu, H., Tachikui, H., Ushida, H., Song, X., Qiu, Y., Yamamoto, S., and Muramatsu, T. (2001) J Biochem 129, 923-928
53. Yamamoto, S., Muramatsu, H., and Muramatsu, T. (2005) Glycoconj J 22, 35-42
54. Saura-Valls, M., Faure, R., Brumer, H., Teeri, T. T., Cottaz, S., Driguez, H., and Planas, A. (2008) Journal of Biological Chemistry 283, 21853-21863
55. Tull, D., and Withers, S. G. (1994) Biochemistry 33, 6363-6370
56. Zeche, D. L., and Withers, S. G. (2001) Curr Opin Chem Biol 5, 643-649
57. Vocadlo, D. J., and Davies, G. J. (2008) Curr Opin Chem Biol
58. Jones, C. S., and Kosman, D. J. (1980) J. Biol. Chem. 255, 11861-11869
59. Yamamoto, K. (1974) J Biochem. (Tokyo) 75, 931-933
60. Whitworth, G. E., Macauley, M. S., Stubbs, K. A., Dennis, R. J., Taylor, E. J., Davies, G. J., Greig, I. R., and Vocadlo, D. J. (2007) J Am Chem Soc 129, 635-644
61. Li, B., Takegawa, K., Suzuki, T., Yamamoto, K., and Wang, L. X. (2008) Bioorganic & Medicinal Chemistry 16, 4670-4675
62. Holm, L., and Sander, C. (1993) J Mol.Biol. 233, 123-138
63. Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N., and Fontes, C. M. (2006) J Biol Chem 281, 8815-8828
64. Vasella, A., Davies, G. J., and Bohm, M. (2002) Current Opinion in Chemical Biology 6, 619-629
65. Fujita, K., and Takegawa, K. (2001) Biochem Biophys Res Commun 283, 680-686
66. Zeche, D. L., and Withers, S. G. (2000) Accounts of Chemical Research 33, 11-18
67. Eriksson, M. A. L., Hard, T., and Nilsson, L. (1995) Biophysical Journal 69, 329-339
68. Read, R. J. (1986) Acta Cryst.A 42, 140-149
69. Hansch, C., and Leo, A. (1979) Substituent constants for correlation analysis in chemistry and biology., Wiley, New York
FIGURE LEGENDS

Figure 1. The reaction catalyzed by GH85 endo-β-glucosaminidases including Endo-D and substrates and inhibitors used in this study. A) The reaction catalyzed by GH85 enzymes cleaves the chitobiose core of N-glycans to generate a free N-glycan bearing a single GlcNAc residue at the terminus, and the liberated protein in which a single GlcNAc residue is N-linked to the protein. The glycosidic bond cleaved is indicated with an arrow. B) The structures of the series of β-glucosaminide substrates used in this study that have varying degrees of fluorination in the acyl group. C) The structure of NAG-thiazoline.

Figure 2. Kinetic studies of SpGH85. A) pH dependence of log $V_{\text{max}}/K_M$ [E]$_0$. Full Michaelis-Menten parameters were determined at each pH using 3F4NP-GlcNAc as substrate. B) Initial velocity of the SpGH85-catalyzed hydrolysis of N-fluoroacetyl derivatives of (○) 3F4NP-GlcNAc-F$_0$ (1a); (●) 3F4NP-GlcNAc-F$_1$ (1b), (□) 3F4NP-GlcNAc-F$_2$ (1c), (■) 3F4NP-GlcNAc-F$_3$ (1d). C) Linear free energy analysis plotting the Taft parameter ($\sigma^*$) of the N-fluoroacetyl substituent of 3F4NP-GlcNAc substrate analogues (1a-1d) against the second order rate constant log $V_{\text{max}}/[E]_0 K_M$ values measured for each substrate with SpGH85 (●).

Figure 3. The structure of SpGH85. A) Divergent stereo cartoon representation of the 1.4 Å crystal structure of the catalytic region of SpGH85. The N-terminal catalytic domain (yellow) is followed by the D1 domain (purple) then the D2 domain (blue). Relevant active site residues are shown in green stick representation. B) Divergent stereo surface representation of SpGH85 shown from the same perspective as in panel A. The bound NAG-thiazoline molecule from the NAG-thiazoline complex is shown for reference. The surface area contributed by relevant active residues are shown in purple (acid/base, Glu337), green (catalytic asparagine, Asn335). The arrows approximate the parts of the active site that may be occupied by a branched, high-mannose substrate.

Figure 4. The structure of SpGH85 in complex with a mechanism-based inhibitor. A) The electron density of NGT (blue stick representation) bound in the active site of SpGH85 in divergent stereo. The maximum likelihood(47)/σ$_a$-weighted (68) $2F_{\text{obs}}-F_{\text{calc}}$ electron density map for the NAG-thiazoline is shown in magenta mesh and contoured at 1σ (0.37 e/Å$^3$). The $F_{\text{obs}}-F_{\text{calc}}$ omit map of NAG-thiazoline is shown in green mesh and contoured at 2.5σ (0.2 e/Å$^3$). B) Interactions in the SpGH85 active site. NAG-thiazoline is shown in blue stick representation and sidechains that interact with NAG-thiazoline are shown in grey stick representation and labeled; possible hydrogen bonds are shown as dashed lines. The maximum likelihood/σ$_a$-weighted $2F_{\text{obs}}-F_{\text{calc}}$ electron density omit maps for Tyr373 is shown in magenta mesh and contoured at 1σ (0.37 e/Å$^3$). The $F_{\text{obs}}-F_{\text{calc}}$ omit map of only the alternate conformation of Tyr373 is shown in green mesh and contoured at 2.5σ (0.2 e/Å$^3$). C) Cutaway view of the SpGH85 active site surface (transparent grey solvent accessible surface) when Tyr373 is in the conformation that does not permit NAG-thiazoline binding. This conformation is also found in the uncomplexed SpGH85 proteins. NAG-thiazoline is shown in yellow stick representation and Tyr373 in blue stick representation. D) As in panel C except with Tyr373 in its alternative conformation that allows NAG-thiazoline binding.

Figure 5. Orientation of Asn335. A) The modeled orientation of Asn335 where Nδ2 is in close proximity of the NAG-thiazoline N2. The B-factors for Oδ1 and Nδ2 obtained for an unrestrained refinement of the uncomplexed structure (indicated as apo) are shown in black text. The corresponding values for the NAG-thiazoline complex are shown, although due to this structure's lower resolution these values were obtained with a standard restrained refinement. Bond lengths for the sidechain in the uncomplexed structure are indicated in red. A 2F$_{\text{obs}}$-F$_{\text{calc}}$ electron density map contoured at 4σ (1.76 e/Å$^3$) obtained from unrestrained refinement of the uncomplexed structure with the side chain of Asn335 omitted is shown in green mesh. The $F_{\text{obs}}-F_{\text{calc}}$ electron density map contoured at 12σ (0.88 e/Å$^3$) from the same refinement is
shown in red mesh. B) The same as in panel A with Asn335 refined in the alternate conformation. Electron density maps are the same as those in panel A.

**Figure 6.** A structurally conserved proton shuttle and its role in the proposed catalytic mechanism of SpGH85. A) Superposition of catalytic residues of SpGH85 and a representative member of GH18 both in complex with ligands. The structure of SpGH85 in complex with NAG-thiazoline, in blue, is superimposed over the structure of the *S. marcescens* ChiB GH18 in complex with allosamidin, in orange (for simplicity only the allosamizoline residue is shown, PDB code 1E6R). The number of catalytic amino acids is shown in the appropriate colour for each complex. B) Superposition of catalytic residues of SpGH85 and a representative member of GH56 both in complex with ligands. The structure of SpGH85 in complex with NAG-thiazoline, in blue, is superimposed over the structure of the Bee venom hyaluronidase GH56 (PDB code 2FCV) in complex with hyaluronic acid tetrasaccharide, in orange (for simplicity only the terminal sugar-like allosamizoline residue is shown). C) Schematic showing the proposed catalytic mechanism of SpGH85 with the putative proton shuttle composed of the catalytic residues indicated. In this mechanism Asn335 is in the imidic acid tautomer with the nitrogen oriented to accept a hydrogen bond from the substrate amide. Asn335 acts as a general base to facilitate formation of an oxazoline intermediate and the proton shuttle acts to disperse charge within the active site. D) Schematic showing the proposed catalytic mechanism of SpGH85. In this mechanism Asn335 is the amide tautomer with the oxygen oriented to accept a hydrogen bond from the substrate amide. Catalysis is driven by cleavage of the glycosidic bond with Glu441 acting as the general acid and Asn335 acts to orient the acetamido group and offer some stabilization through hydrogen bonding interactions with the substrate amide.
Table 1: Data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

| Data collection       | SpGH85 Se-Met | SpGH85 Native 1 | SpGH85 Native 2 | SpGH85 NAG-thiazoline |
|-----------------------|---------------|-----------------|-----------------|-----------------------|
| **Beamline**          | MicroMax-002  | MicroMax-002    | NSLS X8C        | MicroMax-002          |
| **Wavelength** (Å)    | 1.5418        | 1.5418          | 1.1000          | 1.5418                |
| **Space group**       | C2            | C2              | C2              | C2                    |
| **Cell dimensions**   |               |                 |                 |                       |
| \( a, b, c (\text{Å}) \) | 134.5, 59.0, 95.5 | 134.3, 59.1, 95.4 | 135.1, 58.9, 96.7 | 134.9, 59.0, 96.5     |
| \( \beta \) (°)      | 110.1         | 110.0           | 110.4           | 110.3                 |
| **Resolution (Å)**    | 20.00-2.00 (2.07-2.00) | 20.00-1.80 (1.86-1.80) | 30.00-1.40 (1.48-1.40) | 20.00-1.65 (1.74-1.65) |
| \( R_{\text{sym}} \) or \( R_{\text{merge}} \) | 0.06 (0.224) | 0.046 (0.315) | 0.076 (0.410) | 0.050 (0.442) |
| \( I/\sigma I \)      | 15.0 (5.8)    | 15.4 (4.0)      | 18.5 (2.9)      | 16.6 (2.2)            |
| Completeness (%)      | 95.0 (92.2)   | 98.7 (96.9)     | 97.9 (86.2)     | 97.7 (95.4)           |
| Redundancy             | 6.4 (6.3)     | 5.2 (5.0)       | 6.7 (4.3)       | 3.5 (3.2)             |
| **Refinement**        |               |                 |                 |                       |
| **Resolution (Å)**    | 1.40          | 1.65            |                 |                       |
| No. reflections       | 129975        | 79465           |                 |                       |
| \( R_{\text{work}} \) / \( R_{\text{free}} \) | 0.129/0.162 | 0.149/0.187     |                 |                       |
| **B-factor model**    | anisotropic   | anisotropic     |                 |                       |
| **No. atoms**         |               |                 |                 |                       |
| Protein               | 5098          | 5106            |                 |                       |
| Ligand                | N/A           | 14              |                 |                       |
| Water                 | 978           | 935             |                 |                       |
| **B-factors**         |               |                 |                 |                       |
| Protein               | 12.4          | 18.2            |                 |                       |
| Ligand                | N/A           | 29.4            |                 |                       |
| Water                 | 24.9          | 31.4            |                 |                       |
| R.m.s. deviations     |               |                 |                 |                       |
| Bond lengths (Å)      | 0.017         | 0.019           |                 |                       |
| Bond angles (°)       | 1.590         | 1.550           |                 |                       |
| **Ramachandran**      |               |                 |                 |                       |
| Preferred (%)         | 99.8          | 99.6            |                 |                       |
| Allowed (%)           | 0.2           | 0.4             |                 |                       |
| Disallowed (%)        | 0.0           | 0.0             |                 |                       |
Table 2: Michaelis-Menten parameters for the SpGH85–catalyzed hydrolysis of a series of 3-fluoro-4-nitrophenyl 2-N-acetyl-2-deoxy-β-D-glucopyranosides.

| Substrate                  | $\sigma^a$ | $K_M$ (mM) | $V_{max} [E]_0$ (nmol min$^{-1}$ mg$^{-1}$) | $V_{max} [E]_0 / K_M$ (nmol mM$^{-1}$ min$^{-1}$ mg$^{-1}$) |
|----------------------------|------------|------------|------------------------------------------|------------------------------------------------------------|
| 3F4NP-GlcNAc ($1a$)       | 0.0        | 0.13 ± 0.01| 0.56 ± 0.01                              | 4.4 ± 0.5                                                  |
| 3F4NP-GlcNAc-F$_1$ ($1b$) | 0.8        | 0.14 ± 0.01| 0.33 ± 0.01                              | 2.3 ± 0.3                                                  |
| 3F4NP-GlcNAc-F$_2$ ($1c$) | 2.0        | 0.13 ± 0.01| 0.19 ± 0.01                              | 1.4 ± 0.2                                                  |
| 3F4NP-GlcNAc-F$_3$ ($1d$) | 2.8        | 0.13 ± 0.01| 0.066 ± 0.002                            | 0.52 ± 0.09                                               |

$^a$The Taft parameters ($\sigma^*$) used for each N-acyl substituent were obtained from Hansch and Leo (69).
Figure 1.

A

R = Protein,  R' = Glycan-Manβ(1-4)

B

C

D

Oxazoline intermediate

NAG-Thiazoline

1a, n = 3, F = 0, 3F4NP-GlcNAc
1b, n = 2, F = 1, 3F4NP-GlcNAc-F₁
1c, n = 1, F = 2, 3F4NP-GlcNAc-F₂
1d, n = 0, F = 3, 3F4NP-GlcNAc-F₃
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.
Streptococcus pneumoniae endohexosaminidase D: Structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases
D.Wade Abbott, Matthew S. Macauley, David J. Vocadlo and Alisdair B. Boraston

J. Biol. Chem.  published online January 30, 2009

Access the most updated version of this article at doi: 10.1074/jbc.M809663200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2009/01/30/M809663200.DC1