

# Streptococcus pneumoniae Endohexosaminidase D, Structural and Mechanistic Insight into Substrate-assisted Catalysis in Family 85 Glycoside Hydrolases<sup>\*[5]</sup>

Received for publication, December 23, 2008, and in revised form, January 22, 2009. Published, JBC Papers in Press, January 30, 2009, DOI 10.1074/jbc.M809663200

D. Wade Abbott<sup>†1</sup>, Matthew S. Macauley<sup>§1,2</sup>, David J. Vocadlo<sup>§3</sup>, and Alisdair B. Boraston<sup>†4</sup>

From the <sup>†</sup>Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6 and the <sup>§</sup>Department of Chemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

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 ( $K_d = 28 \mu\text{M}$ ) 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.

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 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 toward 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 is 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 harvest nutrients to promote bacterial growth (4, 5). Endohexosaminidase D (Endo-D)<sup>5</sup> 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 Ref. 7). The second group, to which Endo-D belongs, is found in family GH85 of glycoside hydrolases and is of considerable interest because of 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 gly-

\* This work was supported by a grant from the Canadian Institutes of Health Research (to A. B. B. and D. J. V.).

The atomic coordinates and structure factors (codes 2w91 and 2w92) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1 and S2.

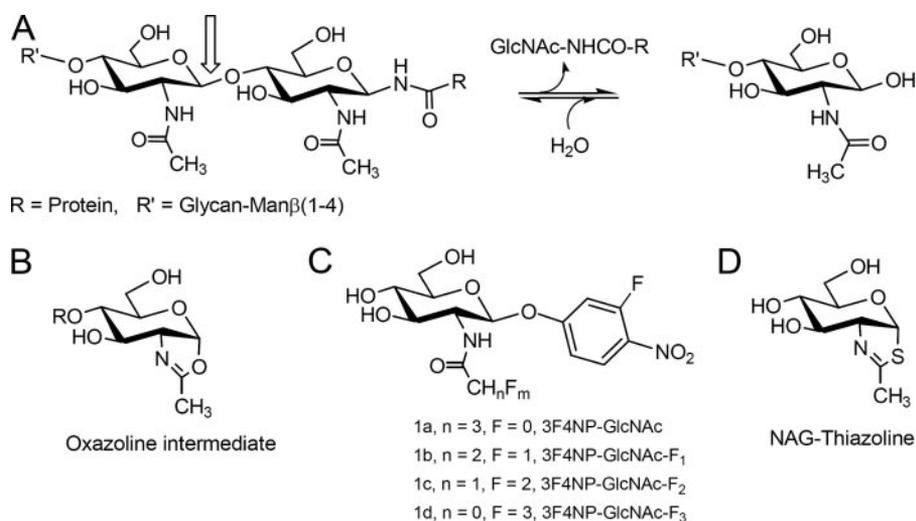
<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> Supported by a graduate scholarship from the Natural Sciences and Engineering Research Council of Canada and the Michael Smith Foundation for Health Research.

<sup>3</sup> Holds Canada Research Chair in Chemical Glycobiology and is a Michael Smith Foundation for Health Research Scholar. To whom correspondence may be addressed: E-mail: [dvocadlo@sfu.ca](mailto:dvocadlo@sfu.ca).

<sup>4</sup> Holds Canada Research Chair in Molecular Interactions and is a Michael Smith Foundation for Health Research Scholar. To whom correspondence may be addressed: E-mail: [boraston@uvic.ca](mailto:boraston@uvic.ca).

<sup>5</sup> The abbreviations used are: Endo-D, endohexosaminidase D; NAG-thiazoline, 1,2-dideoxy-2'-methyl- $\alpha$ -D-glucopyranosyl[2,1-d]- $\Delta$ 2'-thiazoline; SeMet, selenomethionine; PDB, Protein Data Bank; CHES, 2-(cyclohexylamino)ethanesulfonic acid.



**FIGURE 1. The reaction catalyzed by GH85 endo- $\beta$ -glucosaminidases including Endo-D and substrates and inhibitors used in this study.** *A*, 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*, Structure of the oxazoline intermediate proposed for glycoside hydrolases using a substrate assisted catalytic mechanism. *C*, structures of the series of  $\beta$ -glucosaminide substrates used in this study that have varying degrees of fluorination in the acyl group. *D*, structure of NAG-thiazoline.

cans having a sugar oxazoline at the reducing end as a glycosyl donor (11–14). These sugar oxazolines (Fig. 1*B*) 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) and 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 discrete 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, and 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 co-workers (29) have proposed that a conserved asparagine fulfills the same function because they observed that deletion of the side chain by mutagenesis resulted in a dramatic loss of activity. 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 co-workers (11, 29) have proposed that these enzymes use substrate-assisted catalysis. Given the noncanonical

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 *S. pneumoniae* Endo-D using small molecule probes (Fig. 1, *C* and *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, and 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. Finally, 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'-TATATACATATGGAAAAAGAAGTCCAGC-CAGATGTCCCTAAA-3'), which was engineered to have a 5' *Nhe*I restriction endonuclease site, and the reverse primer was GH85R (5'-GCGCGCCTCGAGTTAATCTTTTACAGT-CATACCCCAATCAAA-3'), which had a 3' *Xho*I restriction

## Structural and Mechanistic Analysis of an Endohexosaminidase

site. The amplified DNA fragment was cloned into pET28a (Novagen) via the *NheI* and *XhoI* 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 His<sub>6</sub> tag fused to the SpGH85 catalytic domain by an intervening thrombin protease cleavage site.

### Protein Production and Purification

SpGH85 was produced in *Escherichia coli* BL21 Star (DE3) cells (Invitrogen) using LB media supplemented with 50  $\mu\text{g mL}^{-1}$  of kanamycin. Cultures were grown at 37 °C until they reached an absorbance at 600 nm of 0.5–0.7, whereupon protein production was induced with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside. After further incubation at 37 °C for 4 h, the cells were harvested by centrifugation and ruptured by chemical lysis. SpGH85 was purified by Ni<sup>2+</sup>-immobilized metal affinity chromatography followed by size exclusion chromatography using a Sephacryl S-200 column (GE Healthcare). Size exclusion chromatography was performed using 20 mM Tris-HCl, pH 8.0. Selenomethionine-labeled 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 155,160  $\text{cm}^{-1} \text{M}^{-1}$  (33).

### Chemical Synthesis

*Synthesis of 1,3,4,6-Tetra-O-acetyl-2-deoxy-2-fluoroacetamido- $\beta$ -D-glucopyranose*—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 *N,N*-dimethylformamide (100 ml). Sodium fluoroacetate (1.8 g) was added to a stirred mixture of dry *N,N*-dimethylformamide (90 ml) containing dried Dowex 50-H<sup>+</sup> resin (12 g). After 1 h, dicyclohexylcarbodiimide (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<sub>4</sub> and filtered, and the solvent was removed *in vacuo* to yield colorless syrup. The desired product was purified using flash column silica chromatography using a gradient solvent system (2:1; hexanes/ethyl acetate) to yield the partially purified desired compound. The desired product was then recrystallized from ethyl acetate and hexanes, and the product was obtained with a 77% yield: <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 5.83 (1H, d,  $J_{\text{H1,H2}}$  = 8.8 Hz, H-1), 5.34 (1H, dd,  $J_{\text{H3,H4}}$  = 9.3 Hz, H-3), 5.00 (1H, dd,  $J_{\text{H4,H5}}$  = 10.0 Hz, H-4), 4.74 (2H, dd,  $J_{\text{H7-F}}$  = 47.0 Hz, H-7), 4.26 (1H, dd,  $J_{\text{H6,H6'}}$  = 12.5 Hz, H-6), 4.10 (1H, dd,  $J_{\text{H2,H3}}$  = 10.5 Hz, H-2), 4.07 (1H, dd, H-6'),

3.89 (1H, ddd,  $J_{\text{H5,H6}}$  = 4.6 Hz,  $J_{\text{H5,H6'}}$  = 2.3 Hz, H-5), 2.04 (3H, s, OAc), 2.01 (3H, s, OAc), 1.98 (3H, s, OAc), and 1.95 (3H, s, OAc) ppm.

*Synthesis of 1,3,4,6-Tetra-O-acetyl-2-deoxy-2-difluoroacetamido- $\beta$ -D-glucopyranose*—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 *N,N*-dimethylformamide (100 ml). Dicyclohexylcarbodiimide (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<sub>4</sub> and filtered, and the solvent was removed *in vacuo* to yield colorless syrup. The desired product was purified using flash column silica chromatography using a gradient solvent system (3:1; hexanes/ethyl acetate) to yield the partially purified desired compound. The desired product was then recrystallized from ethyl acetate and hexanes, and the product was obtained with an 83% yield: <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 4.96 (1H, t,  $J_{\text{H7-F}}$  = 53.6 Hz, H-7), 5.78 (1H, d,  $J_{\text{H1,H2}}$  = 8.9 Hz, H-1), 5.28 (1H, dd,  $J_{\text{H3,H4}}$  = 9.3 Hz, H-3), 4.98 (1H, dd,  $J_{\text{H4,H5}}$  = 10.0 Hz, H-4), 4.23 (1H, dd,  $J_{\text{H6,H6'}}$  = 12.5 Hz, H-6), 4.06 (1H, dd,  $J_{\text{H2,H3}}$  = 10.5 Hz, H-2), 4.05 (1H, dd, H-6'), 3.89 (1H, ddd,  $J_{\text{H5,H6}}$  = 2.3 Hz,  $J_{\text{H5,H6'}}$  = 4.5 Hz, H-5), 2.01 (3H, s, OAc), 1.99 (3H, s, OAc), 1.95 (3H, s, OAc), and 1.92 (3H, s, OAc) ppm.

*Synthesis of 1,3,4,6-Tetra-O-acetyl-2-deoxy-2-trifluoroacetamido- $\beta$ -D-glucopyranoside*—Triethylamine (0.8 ml) was added to a solution of 2-amino-2-deoxy-1,3,4,6-tetra-O-acetyl- $\beta$ -D-glucopyranose chloride (1 g) dissolved in dry dichloromethane (20 ml) and cooled (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 diluted in 50 ml of ethyl acetate and 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<sub>4</sub> and filtered, and the solvent was removed *in vacuo* to yield a white solid. The desired product was recrystallized from ethyl acetate and hexanes, and the product was obtained with a 91% yield: <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 5.81 (1H, d,  $J_{\text{H1,H2}}$  = 8.9 Hz, H-1), 5.31 (1H, dd,  $J_{\text{H3,H4}}$  = 9.3 Hz, H-3), 5.03 (1H, dd,  $J_{\text{H4,H5}}$  = 10.0 Hz, H-4), 4.26 (1H, dd,  $J_{\text{H6,H6'}}$  = 12.5 Hz, H-6), 4.09 (1H, dd,  $J_{\text{H2,H3}}$  = 10.5 Hz, H-2), 4.08 (1H, dd, H-6'), 3.91 (1H, ddd,  $J_{\text{H5,H6}}$  = 4.6 Hz,  $J_{\text{H5,H6'}}$  = 2.3 Hz, H-5), 2.04 (3H, s, OAc), 2.02 (3H, s, OAc), 1.98 (3H, s, OAc), and 1.94 (3H, s, OAc) ppm.

*General Procedure for the Synthesis of the 3,4,6-Tri-O-acetyl-2-acyl-2-deoxy- $\beta$ -D-glucopyranosyl Chlorides*—In a similar manner that has been described previously for the nonfluori-

nated derivative (35), acetic anhydride (10 ml) was added to 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-*N*-fluoroacetamido- $\beta$ -D-glucopyranose (1 g) and cooled (0 °C). HCl 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  $\frac{1}{3}$  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<sub>4</sub> 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 to 90%. Therefore, the desired compounds were used in the following steps without further purification: <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$ : 6.62 (1H, d,  $J_{\text{HN-H2}} = 6.6$  Hz, H-N), 6.21 (1H, d,  $J_{\text{H1,H2}} = 3.7$  Hz, H-1), 5.34 (1H, dd,  $J_{\text{H3-H4}} = 9.3$  Hz, H-3), 5.00 (1H, dd,  $J_{\text{H4-H5}} = 10.0$  Hz, H-4), 4.79 (2H, AB<sub>q</sub>,  $J_{\text{H7-F}} = 47.2$  Hz, H-7), 4.26 (1H, dd,  $J_{\text{H6-H6'}} = 12.5$  Hz, H-6), 4.10 (1H, dd,  $J_{\text{H2-H3}} = 10.5$  Hz, H-2), 4.07 (1H, dd, H-6'), 3.89 (1H, ddd,  $J_{\text{H5-H6}} = 4.6$  Hz,  $J_{\text{H5-H6'}} = 2.3$  Hz, H-5), 2.04 (3H, s, OAc), 2.11 (3H, s, OAc), 2.06 (3H, s, OAc), and 2.05 (3H, s, OAc) ppm.

**3,4,6-Tri-*O*-acetyl-2-deoxy-2-difluoroacetamido- $\beta$ -D-glucopyranosyl Chloride**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$ : 6.73 (1H, d, H-N), 6.20 (1H, d,  $J_{\text{H1,H2}} = 3.8$  Hz, H-1), 5.89 (1H, t,  $J_{\text{H7-F}} = 53.9$  Hz, H-7), 5.34 (1H, dd,  $J_{\text{H3-H4}} = 9.3$  Hz, H-3), 5.00 (1H, dd,  $J_{\text{H4-H5}} = 10.0$  Hz, H-4), 4.26 (1H, dd,  $J_{\text{H6-H6'}} = 12.5$  Hz, H-6), 4.10 (1H, dd,  $J_{\text{H2-H3}} = 10.5$  Hz, H-2), 4.07 (1H, dd, H-6'), 3.89 (1H, ddd,  $J_{\text{H5-H6}} = 4.6$  Hz,  $J_{\text{H5-H6'}} = 2.3$  Hz, H-5), 2.04 (3H, s, OAc), 2.11 (3H, s, OAc), 2.06 (3H, s, OAc), and 2.05 (3H, s, OAc) ppm.

**Synthesis of Aryl 2-Acetamido-2-deoxy-3,4,6-Tri-*O*-acetyl- $\beta$ -D-glucopyranosides**—To a mixture of the respective fluorinated 2-*N*-acyl-3,4,6-tri-*O*-acetyl- $\alpha$ -D-glucopyranosyl chloride (1 eq.), benzyltriethylammonium chloride (1 eq.) and the acceptor phenol (2 eq.) were 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  $\times$  1 volume), water (1 volume), and saturated sodium chloride solution (1 volume). The organic layer was dried (MgSO<sub>4</sub>), 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- $\beta$ -D-glucopyranoside**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 8.09 (1H, dd,  $J_{\text{H5Ar-H6Ar}} \approx J_{\text{H5Ar-F3Ar}} = 8.9$  Hz, H-5Ar), 6.91 (1H, dd,  $J_{\text{H2Ar-H6Ar}} = 2.6$  Hz,  $J_{\text{H2Ar-F3Ar}} = 12.3$  Hz, H-2Ar), 6.87 (1H, ddd,  $J_{\text{H6Ar-HF}} = 0.9$  Hz, H-6Ar), 6.62 (1H, d,  $J_{\text{NH-H2}} = 8.4$  Hz, NH), 5.58 (1H, d,  $J_{\text{H1-H2}} = 8.0$  Hz, H-1), 5.55 (1H, dd,  $J_{\text{H3-H4}} = 9.2$  Hz, H-3), 5.13 (1H, dd,  $J_{\text{H4-H5}} = 9.4$  Hz, H-4), 4.76 (2H, dd,  $J_{\text{H7-F}} = 47.1$  Hz,  $J_{\text{H7-H7'}} = 6.2$  Hz, H-7), 4.26 (1H, dd,  $J_{\text{H6-H6'}} = 12.3$  Hz, H-6), 4.20 (1H, dd, H-6'), 4.12 (1H, dd,  $J_{\text{H2-H3}} = 10.3$  Hz, H-2), 4.00 (1H, ddd,

$J_{\text{H5-H6}} = 6.1$  Hz,  $J_{\text{H5-H6'}} = 2.6$  Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), and 2.07 (3H, s, OMe) ppm.

**3-Fluoro-4-nitrophenyl 3,4,6-Tri-*O*-acetyl-2-deoxy-2-difluoroacetamido- $\beta$ -D-glucopyranoside**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 8.08 (1H, dd,  $J_{\text{H5Ar-H6Ar}} \approx J_{\text{H5Ar-F3Ar}} = 8.8$  Hz, H-5Ar), 6.92 (1H, dd,  $J_{\text{H2Ar-H6Ar}} = 2.5$  Hz,  $J_{\text{H2Ar-F3Ar}} = 12.2$  Hz, H-2Ar), 6.86 (1H, ddd,  $J_{\text{H6Ar-HF}} = 0.8$  Hz, H-6Ar), 6.74 (1H, dd,  $J_{\text{NH-H2}} = 8.6$  Hz, NH), 5.86 (1H, t,  $J_{\text{H7-F}} = 54.0$  Hz, H-7), 5.48 (1H, dd,  $J_{\text{H3-H4}} = 9.2$  Hz, H-3), 5.46 (1H, d,  $J_{\text{H1-H2}} = 8.1$  Hz, H-1), 5.14 (1H, dd,  $J_{\text{H4-H5}} = 9.5$  Hz, H-4), 4.26 (1H, dd,  $J_{\text{H6-H6'}} = 12.3$  Hz, H-6), 4.21 (1H, dd, H-6'), 4.20 (1H, dd,  $J_{\text{H2-H3}} = 10.4$  Hz, H-2), 4.00 (1H, ddd,  $J_{\text{H5-H6}} = 6.0$  Hz,  $J_{\text{H5-H6'}} = 2.7$  Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), and 3.08 (1H, s, OMe) ppm.

**Synthesis of 3-Fluoro-4-nitrophenyl 3,4,6-Tri-*O*-acetyl-2-deoxy-2-trifluoroacetamido- $\beta$ -D-glucopyranoside**—The desired products were obtained from 3,4,6-tri-*O*-acetyl-2-deoxy-2-trifluoroacetamido- $\beta$ -D-glucopyranosyl bromide (36) using AgCO<sub>3</sub> and the appropriate acceptor in a similar fashion as described previously (37).

**3-Fluoro-4-nitrophenyl 3,4,6-Tri-*O*-acetyl-2-deoxy-2-trifluoroacetamido- $\beta$ -D-glucopyranoside**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 8.09 (1H, dd,  $J_{\text{H5Ar-H6Ar}} \approx J_{\text{H5Ar-F3Ar}} = 8.9$  Hz, H-5Ar), 6.92 (1H, dd,  $J_{\text{H2Ar-H6Ar}} = 2.6$  Hz,  $J_{\text{H2Ar-F3Ar}} = 12.1$  Hz, H-2Ar), 6.86 (1H, ddd,  $J_{\text{H6Ar-HF}} = 0.8$  Hz, H-6Ar), 6.71 (1H, dd,  $J_{\text{NH-H2}} = 8.7$  Hz, NH), 5.41 (1H, dd,  $J_{\text{H3-H4}} = 9.2$  Hz, H-3), 5.45 (1H, d,  $J_{\text{H1-H2}} = 8.1$  Hz, H-1), 5.16 (1H, dd,  $J_{\text{H4-H5}} = 9.6$  Hz, H-4), 4.28 (1H, dd,  $J_{\text{H2-H3}} = 10.4$  Hz, H-2), 4.26 (1H, dd,  $J_{\text{H6-H6'}} = 12.3$  Hz, H-6), 4.22 (1H, dd, H-6'), 3.99 (1H, ddd,  $J_{\text{H5-H6}} = 5.9$  Hz,  $J_{\text{H5-H6'}} = 2.7$  Hz, H-5), 2.10 (3H, s, OMe), 2.08 (3H, s, OMe), and 3.08 (1H, s, OMe) ppm.

**General Synthesis of Aryl 2-*N*-Acyl-2-deoxy- $\beta$ -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 1 h at room temperature. After this time, the reaction was neutralized with Amberlite IR-120 resin (H<sup>+</sup>) 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- $\beta$ -D-glucopyranoside (3F4NP-GlcNAc, **1a**, Fig. 1C) was known previously, and spectral characterization agrees with the literature (38).

**3-Fluoro-4-nitrophenyl 2-Deoxy-2-fluoroacetamido- $\beta$ -D-glucopyranoside (1b, Fig. 1C)**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 8.08 (1H, dd,  $J_{\text{H5Ar-H6Ar}} \approx J_{\text{H5Ar-F3Ar}} = 9.1$  Hz, H-5Ar), 7.02 (1H, dd,  $J_{\text{H2Ar-H6Ar}} = 2.5$  Hz,  $J_{\text{H2Ar-F3Ar}} = 12.9$  Hz, H-2Ar), 6.95 (1H, ddd,  $J_{\text{H6Ar-HF}} = 1.1$  Hz, H-6Ar), 5.25 (1H, d,  $J_{\text{H1-H2}} = 8.4$  Hz, H-1), 4.81 (2H, d,  $J_{\text{H7-F}} = 47.1$  Hz, H-7), 4.03 (1H, dd,  $J_{\text{H2-H3}} = 10.3$  Hz, H-2), 3.89 (1H, dd,  $J_{\text{H6-H6'}} = 12.2$  Hz, H-6), 3.68 (1H, dd, H-6'), 3.64 (1H, dd,  $J_{\text{H3-H4}} = 8.8$  Hz, H-3), 3.50 (1H, ddd,  $J_{\text{H5-H6}} = 2.2$  Hz,  $J_{\text{H5-H6'}} = 6.0$  Hz, H-5), and 3.40 (1H, dd,  $J_{\text{H4-H5}} = 9.8$  Hz, H-4) ppm.

**3-Fluoro-4-nitrophenyl 2-Deoxy-2-difluoroacetamido- $\beta$ -D-glucopyranoside (1c, Fig. 1C)**—The following were obtained by <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ : 8.08 (1H, dd,  $J_{\text{H5Ar-H6Ar}} \approx J_{\text{H5Ar-F3Ar}} = 9.1$  Hz, H-5Ar), <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$ :

## Structural and Mechanistic Analysis of an Endohexosaminidase

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), and 3.40 (1H, dd,  $J_{H4-H5} = 9.6$  Hz, H-4) ppm.

**3-Fluoro-4-nitrophenyl 2-Deoxy-2-trifluoroacetamido- $\beta$ -D-glucopyranoside (1d, Fig. 1C)**—The following were obtained by  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$ : 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,  $J_{H1-H2} = 8.4$  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), and 3.40 (1H, dd,  $J_{H4-H5} = 9.7$  Hz, H-4) ppm.

### Enzyme Assays

All general kinetics were performed using 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy- $\beta$ -D-glucopyranoside (3F4NP-GlcNAc, **1a**, Fig. 1C) 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-visible 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 and 20  $\mu\text{M}$  of enzyme, and rates were taken over the first 10 min where rates were linear.

### NMR Experiments

$^1\text{H}$  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 3F4NP-GlcNAc in buffer containing 50 mM sodium phosphate and 100 mM sodium chloride, which had been dissolved in  $\text{D}_2\text{O}$ , evaporated, redissolved in 99.9%  $\text{D}_2\text{O}$  (Cambridge Isotopes), and adjusted to a pD of 7.81. Initiation of the reaction was done by the addition of the appropriate amount of lyophilized enzyme to a concentration of 400  $\mu\text{M}$ . 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

Isothermal titration calorimetry 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. NAG-thiazoline (**16**) was synthesized as described previously, and solutions were prepared by mass in buffer saved from the dialysis. All solutions were filtered and degassed prior to use. Twenty five 10- $\mu\text{l}$  injections of 2.1 mM NAG-thiazoline were titrated into 100  $\mu\text{M}$  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 min. Because of the low  $C$  value (40) of only  $\sim 3$  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  $\sim 0.8$ ). 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 vapor diffusion method at 18 °C in 20% (v/v) polyethylene glycol 3350, 0.2 M  $\text{MgCl}_2$ , 0.1 M Tri-sodium citrate, pH 5.0. After a 10-s 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 h. 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) polyethylene glycol 3350, 0.2 M  $\text{MgCl}_2$ , 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 (Brookhaven National Laboratories, New York). Data were 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. Although a higher resolution native data set was obtained, this proved to be nonisomorphous with the SeMet-SpGH85 crystals. The positions of the eight 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 side chains. 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).

TABLE 1

## Data collection and refinement statistics

Values in parentheses are for the highest resolution shell.

	SpGH85 Se-Met	SpGH85 native 1	SpGH85 native 2	SpGH85 NAG-thiazoline
<b>Data collection</b>				
Beamline	MicroMax-002	MicroMax-002	NLSL X8C	MicroMax-002
Wavelength	1.5418 Å	1.5418 Å	1.1000 Å	1.5418 Å
Space group	C2	C2	C2	C2
Cell dimensions				
<i>a, b, c</i>	134.5, 59.0, 95.5 Å	134.3, 59.1, 95.4 Å	135.1, 58.9, 96.7 Å	134.9, 59.0, 96.5 Å
$\beta$	110.1°	110.0°	110.4°	110.3°
Resolution	20.00–2.00 Å (2.07–2.00 Å)	20.00–1.80 Å (1.86–1.80 Å)	30.00–1.40 Å (1.48–1.40 Å)	20.00–1.65 Å (1.74–1.65 Å)
$R_{\text{sym}}$ 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)
<b>Refinement</b>				
Resolution			1.40 Å	1.65 Å
No. reflections			129,975	79,465
$R_{\text{work}}/R_{\text{free}}$			0.129/0.162	0.149/0.187
<i>B</i> -factor model			Anisotropic	Anisotropic
No. of atoms				
Protein			5098	5106
Ligand			NA <sup>a</sup>	14
Water			978	935
<i>B</i> -factors				
Protein			12.4	18.2
Ligand			NA	29.4
Water			24.9	31.4
Root mean square deviations				
Bond lengths			0.017 Å	0.019 Å
Bond angles			1.590°	1.550°
Ramachandran				
Preferred			99.8%	99.6%
Allowed			0.2%	0.4%
Disallowed			0.0%	0.0%

<sup>a</sup> NA means not applicable.

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 observations 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 homogeneous 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. pneumoniae* TIGR4. Endo-D is a 1646-amino acid protein that includes 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. Endoglycosidases, 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, because 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 endoglycosidases, 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  $\beta$ -D-glucosaminides having different leaving groups, and we found that the enzyme showed fair activity toward 3-fluoro-4-nitrophenyl 2-acetamido-2-deoxy- $\beta$ -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

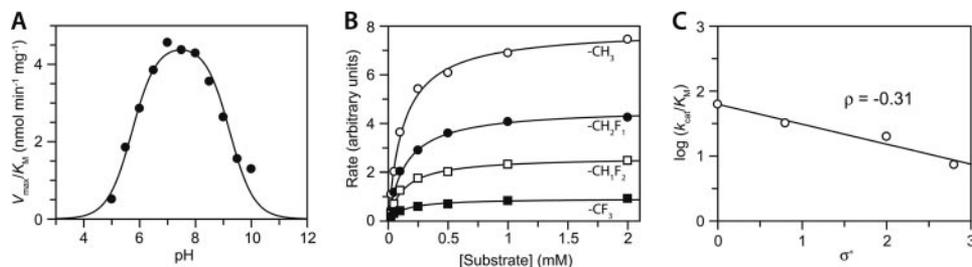
## Structural and Mechanistic Analysis of an Endohexosaminidase

**TABLE 2**

Michaelis-Menten parameters for the SpGH85-catalyzed hydrolysis of a series of 3-fluoro-4-nitrophenyl 2-*N*-acetyl-2-deoxy- $\beta$ -*D*-glucopyranosides

Substrate	$\sigma^{*a}$	$K_m$ mM	$V_{max} [E]_0$ nmol min <sup>-1</sup> mg <sup>-1</sup>	$V_{max} [E]_0 / K_m$ nmol mM <sup>-1</sup> min <sup>-1</sup> mg <sup>-1</sup>
3F4NP-GlcNAc (1a)	0.0	0.13 ± 0.01	0.56 ± 0.01	4.4 ± 0.5
3F4NP-GlcNAc-F <sub>1</sub> (1b)	0.8	0.14 ± 0.01	0.33 ± 0.01	2.3 ± 0.3
3F4NP-GlcNAc-F <sub>2</sub> (1c)	2.0	0.13 ± 0.01	0.19 ± 0.01	1.4 ± 0.2
3F4NP-GlcNAc-F <sub>3</sub> (1d)	2.8	0.13 ± 0.01	0.066 ± 0.002	0.52 ± 0.09

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



**FIGURE 2. Kinetic studies of SpGH85.** *A*, pH dependence of  $V_{max}/K_m$ . 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<sub>0</sub> (○) (1a); 3F4NP-GlcNAc-F<sub>1</sub> (●) (1b), 3F4NP-GlcNAc-F<sub>2</sub> (□) (1c), and 3F4NP-GlcNAc-F<sub>3</sub> (■) (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  $\log V_{max}/[E]_0 K_m$  values measured for each substrate with SpGH85.

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  $pK_a$  values of the two residues being titrated are well dispersed, we obtained good estimates of their kinetic  $pK_a$  values. The basic limb reveals one residue ( $pK_a \approx 9.2 \pm 0.1$ ) 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  $pK_a$  value appears high for a carboxylic acid; however, it has been established in studies of glycoside hydrolases that such perturbations in  $pK_a$  values are common and are mediated by the electrostatic environment of the active site (56). The acidic limb reveals a second residue ( $pK_a \approx 5.7 \pm 0.1$ ) 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  $\beta$ -hemiacetal of *N*-acetylglucosamine (see supplemental Fig. S1). This stereochemical assignment is based on the chemical shift and  $J_{1,2}$  vicinal coupling of the anomeric proton of the first formed  $\beta$ -hemiacetal product and the later formed  $\alpha$ -hemiac-

etal 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 substi-

tuted in place of hydrogen at the methyl group of the acetamido moiety. For all four substrates we observe Michaelian enzyme kinetics (Fig. 2B and Table 2) enabling us to determine the Michaelis-Menten parameters  $K_m$  and  $V_{max}$ . By plotting  $\log(V_{max}/[E]_0 K_m)$ , 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). The slope of the correlation ( $\rho = -0.31$ ) is a measure of the sensitivity of the reaction to different substituents. 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 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  $-1.6$  for GH84 enzymes) (15, 19) but is significantly less than 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 GH84 (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,

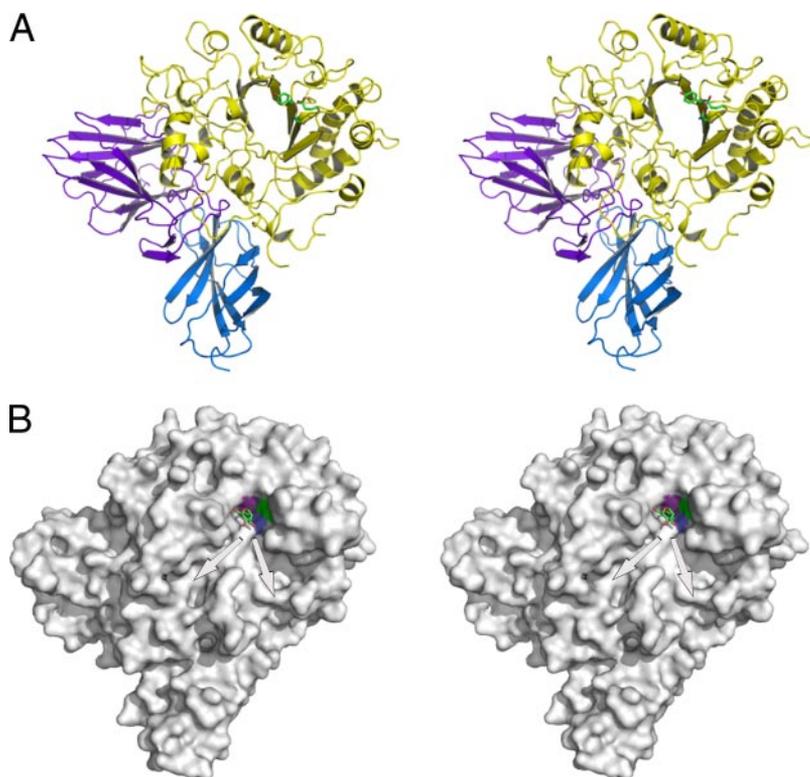


FIGURE 3. **The structure of SpGH85.** *A*, divergent stereo schematic representation of the 1.4 Å crystal structure of the catalytic region of SpGH85. The N-terminal catalytic domain (yellow) is followed by the D1 domain (purple) and then the D2 domain (blue). Relevant active site residues are shown in green stick representation. *B*, divergent stereo surface representation of SpGH85 shown from the same perspective as in *A*. The bound NAG-thiazoline molecule from the NAG-thiazoline complex is shown for reference. The surface areas contributed by relevant active residues are shown in purple (Tyr-373), green (acid/base, Glu-337), and blue (catalytic asparagine, Asn-335). The arrows approximate the parts of the active site that may be occupied by a branched, high mannose substrate.

are strongly inhibited ( $IC_{50} = 220$  and  $6000$  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 studies of NAG-thiazoline binding to SpGH85. Our kinetic assays were hindered by the requirement for large concentrations of enzyme ( $1-10$   $\mu$ M) to obtain useful rates, which in turn complicated assaying inhibitor concentrations lower than  $50$   $\mu$ M. 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  $28 \pm 4$   $\mu$ M, a value that is consistent with the inhibition studies (see supplemental Fig. S2). This relatively strong binding is somewhat surprising given that the structurally more complex (Man)<sub>9</sub>-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 micromolar to high nanomolar range (61). Although that study involves different enzymes and detailed  $IC_{50}$  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 of an oligosaccharide-based inhibitor 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 includes three domains as follows: 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 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 root mean square deviations of 1.7–2.7 Å with proteins having a fibronectin fold

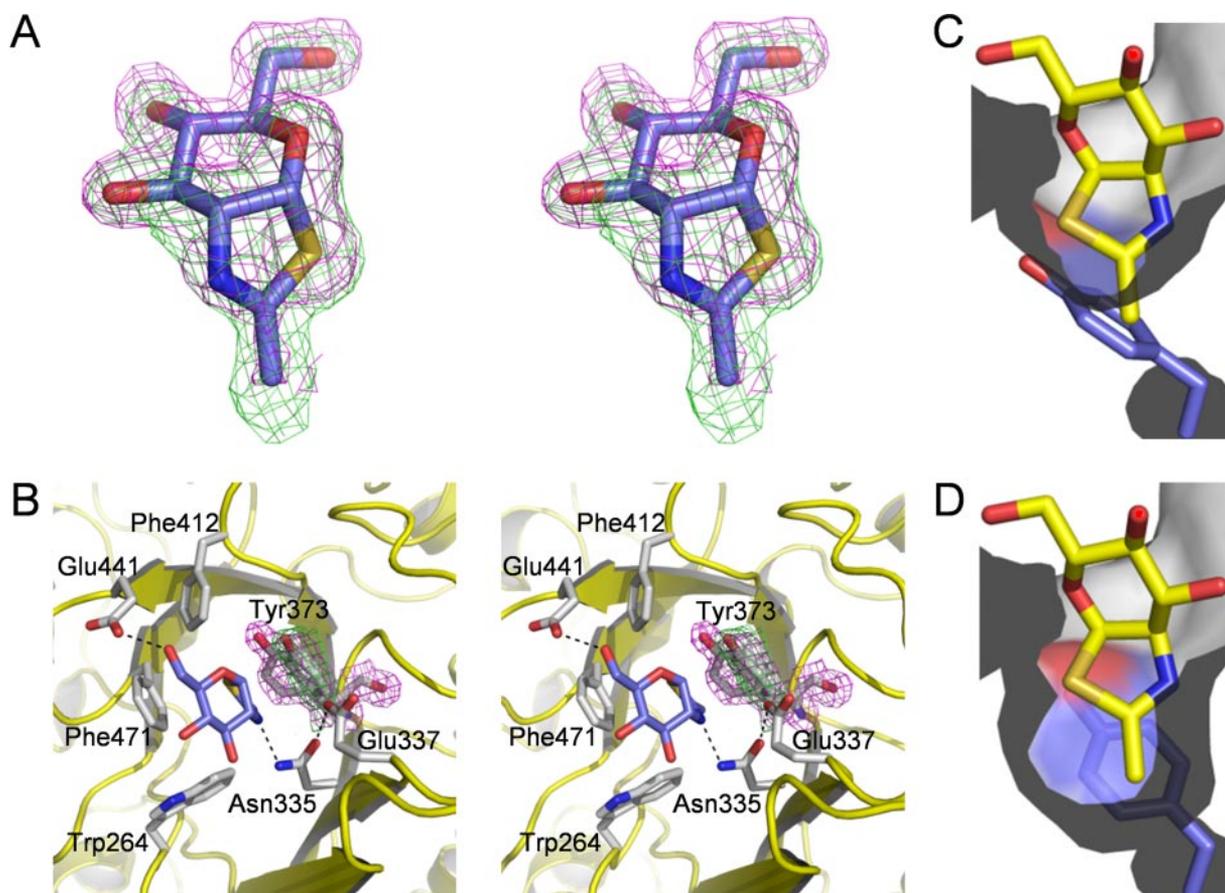


FIGURE 4. **The structure of SpGH85 in complex with a mechanism-based inhibitor.** *A*, electron density of NAG-thiazoline (blue stick representation) bound in the active site of SpGH85 in divergent stereo. The maximum likelihood (47)/ $\sigma_a$ -weighted (68)  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map for the NAG-thiazoline is shown in magenta mesh and contoured at  $1\sigma$  ( $0.37 \text{ e}^-/\text{\AA}^3$ ). The  $F_{\text{obs}} - F_{\text{calc}}$  omit map of NAG-thiazoline is shown in green mesh and contoured at  $2.5\sigma$  ( $0.2 \text{ e}^-/\text{\AA}^3$ ). *B*, interactions in the SpGH85 active site. NAG-thiazoline is shown in blue stick representation, and side chains that interact with NAG-thiazoline are shown in gray stick representation and labeled; possible hydrogen bonds are shown as dashed lines. The maximum likelihood/ $\sigma_a$ -weighted  $2F_{\text{obs}} - F_{\text{calc}}$  electron density omit maps  $F_{\text{obs}} - F_{\text{calc}}$  Tyr-373 are shown in magenta mesh and contoured at  $1\sigma$  ( $0.37 \text{ e}^-/\text{\AA}^3$ ). The  $F_{\text{obs}} - F_{\text{calc}}$  omit map of only the alternate conformation of Tyr-373 is shown in green mesh and contoured at  $2.5\sigma$  ( $0.2 \text{ e}^-/\text{\AA}^3$ ). *C*, cutaway view of the SpGH85 active site surface (transparent gray solvent-accessible surface) when Tyr-373 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 Tyr-373 in blue stick representation. *D*, as in *C* except with Tyr-373 in its alternative conformation that allows NAG-thiazoline binding.

(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. 3*B*). 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. 4*A*) bound within the pocket on the face of the  $(\beta/\alpha)_8$ -barrel that harbors the two conserved catalytic residues (Fig. 4*B*). This active site pocket cradling NAG-thiazoline must be the  $-1$  saccharide-binding subsite to which the *N*-acetylglucosamine residue of the chitobiose 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  $(\beta/\alpha)_8$ -barrel to which the arms of the branched *N*-glycan most likely bind (Fig. 3*B*).

Despite the use of high concentrations of NAG-thiazoline and the unambiguous electron density for NAG-thiazoline, however, the presence of dual conformations of Tyr-373 suggested to us that the active site was partially occupied by NAG-thiazoline. One conformation of Tyr-373, 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. 4*C*). The second conformation of Tyr-373, clearly

revealed by its  $F_{\text{obs}} - F_{\text{calc}}$  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 Tyr-373 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 Tyr-373. This movement of Tyr-373, 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 Fig. 4, C and D).

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 side chains (Tyr-373, Trp-264, Phe-471, and Phe-412) (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 (Leu-229, Phe-333, and Phe-409). It is likely that these aromatic residues stabilize the transition state through favorable orbital overlap between the  $\pi$  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, Tyr-373 adopts a position placing its  $O\eta$  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  $O\epsilon 1$  of Glu-441. Asn-335 of the enzyme appears to hydrogen bond with  $N2$  of NAG-thiazoline, although this is slightly ambiguous. The orientation of the Asn-335 amide group was modeled on the basis of  $B$ -factor analysis, which yielded one clear orientation that minimized the discrepancy between the  $O\delta 1$  and  $N\delta 2$   $B$ -factors (Fig. 5). This orientation is also the preferred orientation in the uncomplexed SpGH85 structure (Fig. 5). The positioning of Asn-335 placed  $O\delta 1$  at 3.2 Å from  $N2$  of NAG-thiazoline, but with slightly less than ideal hydrogen bonding geometry. In this orientation,  $N\delta 2$  of Asn-335 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  $O4$  of the inhibitor, 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 with its nanomolar affinity for GH20 and GH84 (15, 16, 19). Notably, the side chain of Glu-337 is positioned to fulfill its putative function as the general acid/base in GH85 (Fig. 4B)

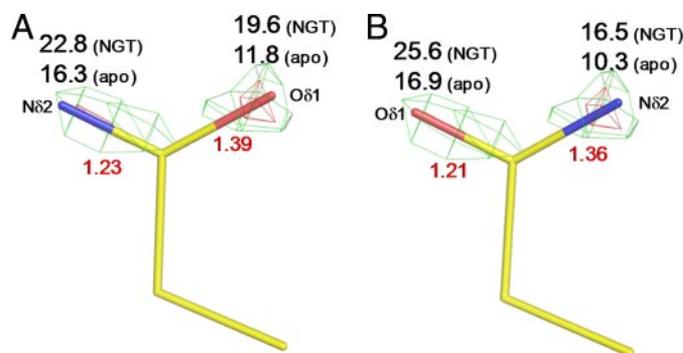


FIGURE 5. **Orientation of Asn-335.** A, modeled orientation of Asn-335 where  $N\delta 2$  is in close proximity of the NAG-thiazoline  $N2$ . The  $B$ -factors for  $O\delta 1$  and  $N\delta 2$  obtained for an unrestrained refinement of the uncomplexed structure (indicated as apo) are shown in black text. The corresponding values for the NAG-thiazoline complex are shown, although because of the lower resolution of this structure these values were obtained with a standard restrained refinement. Bond lengths for the side chain in the uncomplexed structure are indicated in red. A  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map contoured at  $4\sigma$  ( $1.76\text{ e}^-/\text{\AA}^3$ ) obtained from unrestrained refinement of the uncomplexed structure with the side chain of Asn-335 omitted is shown in green mesh. The  $F_{\text{obs}} - F_{\text{calc}}$  electron density map contoured at  $12\sigma$  ( $0.88\text{ e}^-/\text{\AA}^3$ ) from the same refinement is shown in red mesh. B, same as in A with Asn-335 refined in the alternate conformation. Electron density maps are the same as those in A.

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 Glu-337 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, Asn-335. 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).

**Role of Asn-335**—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 Asn-335 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 Asn-335 surprisingly positions the nitrogen, rather than the oxygen, in close proximity (2.6 Å) to  $N2$  of the thiazoline ring. Because the amide nitrogen of

## Structural and Mechanistic Analysis of an Endohexosaminidase

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  $\text{NH}_2$  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 Asn-335 may be favored. This tentative proposal readily accounts for several experimental observations.

An examination of the electron density of Asn-335 in the high resolution apo-structure reveals the highest density at the center we propose to be O $\delta$ 1, in keeping with this atom having a greater number of electrons than nitrogen, and coherent with the modeled conformation of this side chain. Furthermore, the electron density at this atom is localized, which suggests that it forms only a single bond with C $\gamma$ . Consistent with this proposal is that the O $\delta$ 1—C $\gamma$  bond length is 1.39 Å when obtained using an unrestrained refinement. In contrast, delocalized electron density was observed between C $\gamma$  and N $\delta$ 2, which have an interatomic distance of 1.23 Å based on an unrestrained refinement, implying the presence of a double bond. By comparison, Asn-411, 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 O $\delta$ 1, but it is also delocalized over the 1.25 Å O1—C $\gamma$  double bond (not shown). The N $\delta$ 2—C $\gamma$  single bond length is 1.37 Å, whereas the *B*-factor of O $\delta$ 1 is 1 Å less than for N $\delta$ 2. Thus, although not conclusive, the observed electron density and bond lengths of the Asn-335 side chain, in conjunction with its conformation established on the basis of *B*-factor analysis, are most consistent with the imidic acid tautomer of this side chain.

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  $pK_a$  of 9.2 is most likely associated with titration of Glu-337, 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 Asn-335.

This proposal is consistent with the need for this residue to have an available lone pair to facilitate catalysis through hydrogen bonding to the amide nitrogen of the acetamido group. Furthermore, this proposal makes chemical sense because the  $pK_a$  for *N*-deprotonation of an *O*-protonated imidic acid in solution is estimated at  $\approx 7.5$ , whereas that of an amide is  $\approx 18$  (67). Accordingly, downward perturbation of the  $pK_a$  of an amide moiety by over 12 units to generate a negatively charged amide anion seems

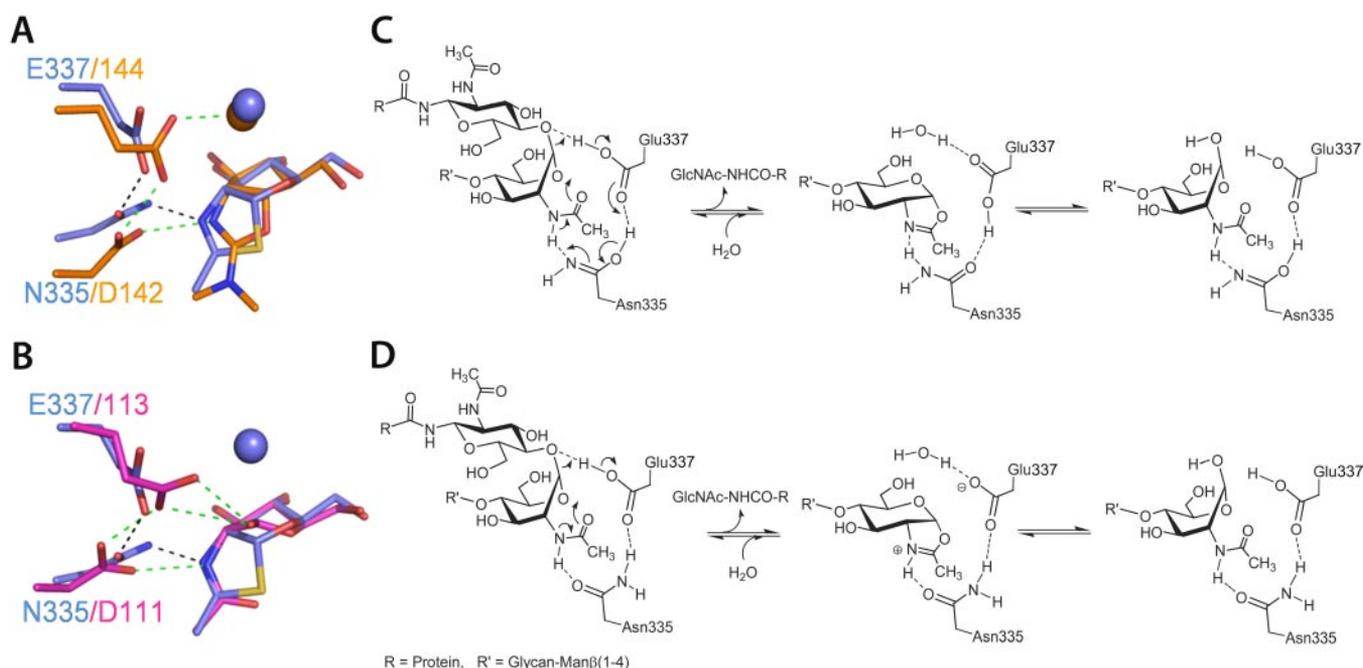
most unlikely, particularly in the context of the hydrophobic and negatively charged environment of the active site. More reasonable is that the  $pK_a$  of the protonated imidic acid tautomer would be perturbed to approximate the kinetic  $pK_a$  of 5.7 seen in the pH profile. This tautomeric form of Asn-335 is likely stabilized, and its  $pK_a$  is 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 Asn-335 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. Based on our data, however, this appears less likely. Nevertheless, definitive assignment of the orientation and tautomeric state of Asn-335 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, Asn-335 in its imidic acid form acts as a general base to facilitate catalysis because 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 Glu-337 transfers its proton to the glycosidic oxygen, and as this occurs the  $pK_a$  of the protonated imidic acid form of Asn-335 must be perturbed upward. This proposal also makes good sense mechanistically because 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, GH56, and GH85 enzymes (Fig. 6, A and B) suggests these residues may be tolerant to being substituted with one another. Notably, previous studies of Endo-D have shown that substitution of Asn-335 with an aspartate only reduced the activity of Endo-D on  $\text{Man}_3$  and  $\text{Man}_5$  *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 10-fold decrease in the second-order rate constant ( $V_{\text{max}}/K_m$ ), whereas substitution with alanine was deleterious (24).

Consistent with this proton shuttle proposal, we also observe that O $\delta$ 1 of Asn-335 forms a 2.7 Å hydrogen bond with O $\epsilon$ 2 of the general acid residue Glu-337. In SpGH85, the observation of this hydrogen bond in both the enzyme-inhibitor complex and the free enzyme lends further support that Asn-335 operates as the imidic acid tautomer, because it would not be possible in both states if Asn-335 was the amide tautomer oriented in the same way. Furthermore, given that the  $pK_a$  of the general acid is predicted to be 9.2 and that of the *N*-protonated imidic acid of Asn-335 is estimated at 5.7, it seems possible that even as it acts to deprotonate the forming oxazoline, the imidic acid of Asn-335 could transfer the proton of O $\delta$ 1 to Glu-337. If this



**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 allosamidin residue is shown, PDB code 1E6R). The number of catalytic amino acids is shown in the appropriate color 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 allosamidin 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 Asn-335 acts as the imidic acid tautomer with the nitrogen oriented to accept a hydrogen bond from the substrate amide. Asn-335 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 an alternate catalytic mechanism of SpGH85. In this mechanism Asn-335 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 Glu-337 acting as the general acid and Asn-335 acts to orient the acetamido group and offer some stabilization through hydrogen bonding interactions with the substrate amide.

shuttling of the proton from the amide group of the substrate through to the general acid occurs during catalysis, Asn-335 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 occurring in a concerted, but asynchronous, 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 because the topological arrangement of these two key catalytic residues, and their hydrogen bonding to each other, is highly conserved (Fig. 6, A and 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, Asn-335 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 Asn-335 in this orientation could act as a general base because the  $pK_a$  of the carbonyl of an amide is expected to be  $\sim 0$  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 GH56 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; this is a proposal that we favor 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 homogeneously *N*-glycosylated biomolecules; a topic that is of considerable importance to the pharmaceutical industry (31).

## REFERENCES

- Garcia-Suarez Mdel, M., Vazquez, F., and Mendez, F. J. (2006) *Enferm. Infecc. Microbiol. Clin.* **24**, 512–517
- 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) *Science* **293**, 498–506
- Hava, D. L., and Camilli, A. (2002) *Mol. Microbiol.* **45**, 1389–1406
- Burnaugh, A. M., Frantz, L. J., and King, S. J. (2008) *J. Bacteriol.* **190**, 221–230
- King, S. J., Hippe, K. R., and Weiser, J. N. (2006) *Mol. Microbiol.* **59**, 961–974
- Koide, N., Nose, M., and Muramatsu, T. (1977) *Biochem. Biophys. Res. Commun.* **75**, 838–844
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) *Nucleic Acids Res.* **37**, D233–D238
- Yamamoto, K., Kadowaki, S., Watanabe, J., and Kumagai, H. (1994) *Biochem. Biophys. Res. Commun.* **203**, 244–252
- Takegawa, K., Tabuchi, M., Yamaguchi, S., Kondo, A., Kato, I., and Iwahara, S. (1995) *J. Biol. Chem.* **270**, 3094–3099
- Mizuno, M., Haneda, K., Iguchi, R., Muramoto, I., Kawakami, T., Aimoto, S., Yamamoto, K., and Inazu, T. (1999) *J. Am. Chem. Soc.* **121**, 284–290
- Fujita, M., Shoda, S., Haneda, K., Inazu, T., Takegawa, K., and Yamamoto, K. (2001) *Biochim. Biophys. Acta* **1528**, 9–14
- Rising, T. W., Claridge, T. D., Davies, N., Gamblin, D. P., Moir, J. W., and Fairbanks, A. J. (2006) *Carbohydr. Res.* **341**, 1574–1596
- Li, B., Song, H., Hauser, S., and Wang, L. X. (2006) *Org. Lett.* **8**, 3081–3084
- Li, B., Zeng, Y., Hauser, S., Song, H., and Wang, L. X. (2005) *J. Am. Chem. Soc.* **127**, 9692–9693
- Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D., and Vocadlo, D. J. (2005) *J. Biol. Chem.* **280**, 25313–25322
- Knapp, S., Vocadlo, D., Gao, Z. N., Kirk, B., Lou, J. P., and Withers, S. G. (1996) *J. Am. Chem. Soc.* **118**, 6804–6805
- Drouillard, S., Armand, S., Davies, G. J., Vorgias, C. E., and Henrissat, B. (1997) *Biochem. J.* **328**, 945–949
- Arming, S., Strobl, B., Wechselberger, C., and Kreil, G. (1997) *Eur. J. Biochem.* **247**, 810–814
- Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocadlo, D. J., and Davies, G. J. (2006) *Nat. Struct. Mol. Biol.* **13**, 365–371
- Markovic-Housley, Z., Miglierini, G., Soldatova, L., Rizkallah, P. J., Muller, U., and Schirmer, T. (2000) *Structure (Lond.)* **8**, 1025–1035
- Tervisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B., and Dijkstra, B. W. (1995) *Biochemistry* **34**, 15619–15623
- Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., and Vorgias, C. E. (1996) *Nat. Struct. Biol.* **3**, 638–648
- van Aalten, D. M., Komander, D., Synstad, B., Gaseidnes, S., Peter, M. G., and Eijsink, V. G. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8979–8984
- Synstad, B., Gaseidnes, S., Van Aalten, D. M., Vriend, G., Nielsen, J. E., and Eijsink, V. G. (2004) *Eur. J. Biochem.* **271**, 253–262
- Williams, S. J., Mark, B. L., Vocadlo, D. J., James, M. N., and Withers, S. G. (2002) *J. Biol. Chem.* **277**, 40055–40065
- Cetinbas, N., Macauley, M. S., Stubbs, K. A., Drapala, R., and Vocadlo, D. J. (2006) *Biochemistry* **45**, 3835–3844
- Greig, I. R., Zahariev, F., and Withers, S. G. (2009) *J. Am. Chem. Soc.* **130**, 17620–17628
- Fujita, K., Sato, R., Toma, K., Kitahara, K., Suganuma, T., Yamamoto, K., and Takegawa, K. (2007) *J. Biochem. (Tokyo)* **142**, 301–306
- Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L. X., and Yamamoto, K. (2008) *J. Biol. Chem.* **283**, 4469–4479
- Heidecke, C. D., Ling, Z. L., Bruce, N. C., Moir, J. W. B., Parsons, T. B., and Fairbanks, A. J. (2008) *ChemBioChem* **9**, 2045–2051
- Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H., Jiang, Y. W., Rios, S., Bobrowicz, P., Stadheim, T. A., Li, H. J., Choi, B. K., Hopkins, D., Wischniewski, H., Roser, J., Mitchell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., and Gerngross, T. U. (2006) *Science* **313**, 1441–1443
- Ficko-Blean, E., and Boraston, A. B. (2005) *Acta Crystallogr. Sect. F Struct. Biol. Crystallogr. Commun.* **61**, 834–836
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) *Nucleic Acids Res.* **31**, 3784–3788
- Cunha, A. C., Pereira, L. O. R., de Souza, M. C. B. V., and Ferreira, V. F. (1999) *J. Chem. Educ.* **76**, 79–80
- Roy, R., and Tropper, F. D. (1991) *Can. J. Chem.-Rev. Can. de Chimie* **69**, 817–821
- Busca, P., and Martin, O. R. (1998) *Tetrahedron Lett.* **39**, 8101–8104
- Vocadlo, D. J., and Withers, S. G. (2005) *Biochemistry* **44**, 12809–12818
- Macauley, M. S., Stubbs, K. A., and Vocadlo, D. J. (2005) *J. Am. Chem. Soc.* **127**, 17202–17203
- Ficko-Blean, E., Stubbs, K. A., Nemirovsky, O., Vocadlo, D. J., and Boraston, A. B. (2008) *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6560–6565
- Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) *Anal. Biochem.* **179**, 131–137
- Pflugrath, J. W. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 1718–1725
- Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1772–1779
- Evans, G., and Bricogne, G. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 976–991
- Cowtan, K. D., and Zhang, K. Y. (1999) *Prog. Biophys. Mol. Biol.* **72**, 245–270
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**,

- 458–463
46. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132
  47. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
  48. Brunger, A. T. (1992) *Nature* **355**, 472–475
  49. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 191–205
  50. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **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. (Tokyo)* **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) *J. Biol. Chem.* **283**, 21853–21863
  55. Tull, D., and Withers, S. G. (1994) *Biochemistry* **33**, 6363–6370
  56. Zechel, 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.* **2**, 539–555
  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) *Bioorg. Med. Chem.* **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) *Curr. Opin. Chem. Biol.* **6**, 619–629
  65. Fujita, K., and Takegawa, K. (2001) *Biochem. Biophys. Res. Commun.* **283**, 680–686
  66. Zechel, D. L., and Withers, S. G. (2000) *Acc. Chem. Res.* **33**, 11–18
  67. Eriksson, M. A. L., Hard, T., and Nilsson, L. (1995) *Biophys. J.* **69**, 329–339
  68. Read, R. J. (1986) *Acta Crystallogr. Sect. A* **42**, 140–149
  69. Hansch, C., and Leo, A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, New York

***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.* 2009, 284:11676-11689.

doi: 10.1074/jbc.M809663200 originally published online January 30, 2009

---

Access the most updated version of this article at doi: [10.1074/jbc.M809663200](https://doi.org/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>

This article cites 68 references, 13 of which can be accessed free at

<http://www.jbc.org/content/284/17/11676.full.html#ref-list-1>