

Structure and mechanism of a bacterial β -glucosaminidase having O-GlcNAcase activity

Rebecca J Dennis¹, Edward J Taylor¹, Matthew S Macauley², Keith A Stubbs², Johan P Turkenburg¹, Samuel J Hart¹, Gary N Black³, David J Vocadlo² & Gideon J Davies¹

O-GlcNAc is an abundant post-translational modification of serine and threonine residues of nucleocytoplasmic proteins. This modification, found only within higher eukaryotes, is a dynamic modification that is often reciprocal to phosphorylation. In a manner analogous to phosphatases, a glycoside hydrolase termed O-GlcNAcase cleaves O-GlcNAc from modified proteins. Enzymes with high sequence similarity to human O-GlcNAcase are also found in human pathogens and symbionts. We report the three-dimensional structure of O-GlcNAcase from the human gut symbiont *Bacteroides thetaiotaomicron* both in its native form and in complex with a mimic of the reaction intermediate. Mutagenesis and kinetics studies show that the bacterial enzyme, very similarly to its human counterpart, operates via an unusual 'substrate-assisted' catalytic mechanism, which will inform the rational design of enzyme inhibitors.

The post-translational modification of nuclear and cytoplasmic proteins with the simple saccharide 2-acetamido-2-deoxy-D-glucopyranose (GlcNAc) is a common modification found within all cells of higher eukaryotes¹. This saccharide is attached to serine or threonine residues through a β -glycosidic linkage (O-GlcNAc) and is often found at sites known to be phosphorylated^{2,3} (Fig. 1). A reciprocal relationship is also observed between global levels of cellular O-GlcNAc and O-phosphorylation⁴. Consistent with its reciprocal relation to phosphorylation, O-GlcNAc is a dynamic post-translational modification with cycles of addition and removal occurring several times during the lifetime of a protein⁵. Many different classes of proteins have been shown to be modified with O-GlcNAc, including structural proteins such as nuclear pore proteins⁶ and cytoskeletal proteins⁷, transcription factors such as the glucose-responsive factor Sp1 (ref. 8) and important processing enzymes such as the proteasome⁹. In accord with its abundance on cellular targets, O-GlcNAc has been linked to several disease states, including type II diabetes.

To maintain an appropriate dynamic balance of cellular O-GlcNAc levels, two different carbohydrate-processing enzymes are vital. The enzyme catalyzing the installation of O-GlcNAc is a glycosyltransferase termed uridine diphosphate-N-acetyl-D-glucosamine: polypeptidyl transferase (OGTase)^{10,11}. The enzyme responsible for catalyzing the cleavage of O-GlcNAc from modified proteins and returning them to their unmodified state is the glycoside hydrolase known as O-GlcNAcase^{12,13}. Together, these enzymes are essential for regulating O-GlcNAc levels and maintaining health. OGTase is essential for survival at the level of single cells¹⁴, and recently a single-nucleotide polymorphism in the O-GlcNAcase gene has been linked to the

occurrence of type II diabetes in a Mexican population¹⁵ and elevation of O-GlcNAc levels in adipocytes has been implicated in the development of insulin resistance¹⁶.

Despite the importance of O-GlcNAcase and OGTase in mediating O-GlcNAc levels, little is known about the three-dimensional structure of either of these two enzymes beyond the structure of a noncatalytic region of OGTase¹⁷. O-GlcNAcase is crucial in cellular signaling and stress response¹⁸. The human O-GlcNAcase comprises two distinct domains: a C-terminal acyltransferase¹⁹ domain and an N-terminal glycoside hydrolase domain. Notably, the N-terminal domains of eukaryotic O-GlcNAcases have high sequence similarity to certain bacterial enzymes from human pathogens and symbionts. The function of these bacterial enzymes is less clear, although, on the basis of sequence similarities, they have been grouped with eukaryotic O-GlcNAcase into the glycoside hydrolase family GH84 (ref. 20).

To gain insight into the architecture and mechanism of these important signaling enzymes, we have investigated a close homolog of human O-GlcNAcase, termed BtGH84, identified within the genome²¹ of the human symbiont *B. thetaiotaomicron* VPI-5482. We show that BtGH84, like human O-GlcNAcase, is able to cleave O-GlcNAc from post-translationally modified proteins using a mechanism involving substrate-assisted catalysis (Fig. 1b). Furthermore, BtGH84 is strongly inhibited by both 1,2-dideoxy-2'-methyl- α -D-glucopyranoso-[2,1-d]- Δ 2'-thiazoline (NAG-thiazoline)²² (Fig. 1) and O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenyl-carbamate (PUGNAc)²³ as well as by structural homologs of these inhibitors. The X-ray crystal structure of BtGH84 in complex with the inhibitor NAG-thiazoline, further illuminated by kinetic data

¹York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, UK. ²Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. ³Biomolecular and Biomedical Research Centre, School of Applied Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK. Correspondence should be addressed to D.J.V. (dvocadlo@sfu.ca) or G.J.D. (davies@ysbl.york.ac.uk).

Received 19 December 2005; accepted 8 March 2006; published online 26 March 2006; doi:10.1038/nsmb1079

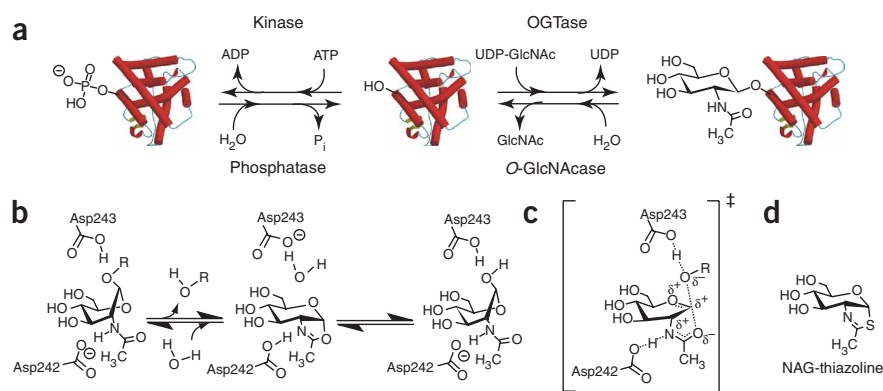


Figure 1 Schematic diagram of the *O*-GlcNAc modification (a) Dynamic interplay of phosphorylation and *O*-GlcNAc modification of intracellular proteins. (b) Mechanism of action of *O*-GlcNAc hydrolases using substrate-assisted catalysis (residue numbers are for the BtGH84 enzyme). (c) Transition state of the *O*-GlcNAc-catalyzed hydrolysis of *N*-acetylglucosaminides. (d) Structure of the mechanism-derived inhibitor NAG-thiazoline, which mimics the intermediate of the substrate-assisted mechanism.

on site-directed variants, shows that the active site of human *O*-GlcNAcase and the bacterial homolog are remarkably conserved.

RESULTS

Architecture of a bacterial *O*-GlcNAcase

Soluble *B. thetaiotaomicron* GH84 was expressed in the cytoplasm of *Escherichia coli* cells, without its N-terminal 21-residue signal peptide (see Methods). The structure of BtGH84 was solved initially using SAD data at 1.95 Å from a crystal grown in complex with the NAG-derived thiazoline, and data for the native enzyme were then obtained to 1.85 Å. The P1 asymmetric unit comprises two molecules of BtGH84 (which overlap with an r.m.s. deviation of 0.2 Å over all atoms), related by a two-fold noncrystallographic axis. The interface formed by the interaction of helical bundle domains buries approximately 950 Å² of solvent-accessible surface, and light-scattering data suggest that the protein is a dimer in solution as well (data not

shown). The structure can be traced from residue 4 of the mature protein to residue 589, with a short break from 44 to 53, in the first domain. The C-terminal domain is, however, extremely mobile in the crystal and can only be traced in parts.

Briefly, the *Bacteroides* enzyme comprises four distinct domains (Fig. 2a). The N-terminal domain, residues 4–127, is a six-stranded antiparallel β-sheet that also buries two α-helices against the surface of the next domain. This domain is similar in both topology (β1, β6, β4, β3, β2) and structure to the N-terminal domains of human²⁴ and bacterial²⁵ hexosaminidases from family GH20, with DALI²⁶ Z-scores > 10, corresponding to r.m.s. deviations of approximately 2.5 Å for 100 equivalent Cα positions, with approximately 21% sequence identity (Fig. 2b).

Domain 2, formed from residues 128–411, is a (β/α)₈-barrel domain that contains the catalytic machinery. As with the N-terminal β+α-domain, the (β/α)₈-barrel domain is highly similar to those of GH20 hexosaminidases; a notable exception is that, although the topological position of the catalytic center on strand β4 is invariant, the identities of these catalytic residues are different. The catalytic center, whose most important residues are Asp242 and Asp243, is on the C-terminal face of the barrel in a shallow surface depression. On the basis of sequence alone¹³, we expected the human *O*-GlcNAcase catalytic domain to have the same overall structure; the *Bacteroides* enzyme is over 40% identical to the human *O*-GlcNAcase in this region. The C-terminal region of the catalytic domain is terminated by two helices, residues 414–426 and 440–456, which form part of an all-helical bundle together with domain 3 of the bacterial enzyme.

After the catalytic domain, the *Bacteroides* enzyme contains two domains that are specific to bacterial as opposed to eukaryotic GH84 enzymes. The first of these is a helical bundle domain (residues 462–589) comprising four α-helices, α1 (462–484), α2 (492–516), α3 (522–545) and α4 (559–579), that are probably found only in GH84 enzymes from *Bacteroides* spp. and certain pathogenic *Clostridia* spp. The final domain of BtGH84, comprising approximately 125 residues, is highly mobile in the crystal and can only be built in part. It is an all-β-sheet domain, with

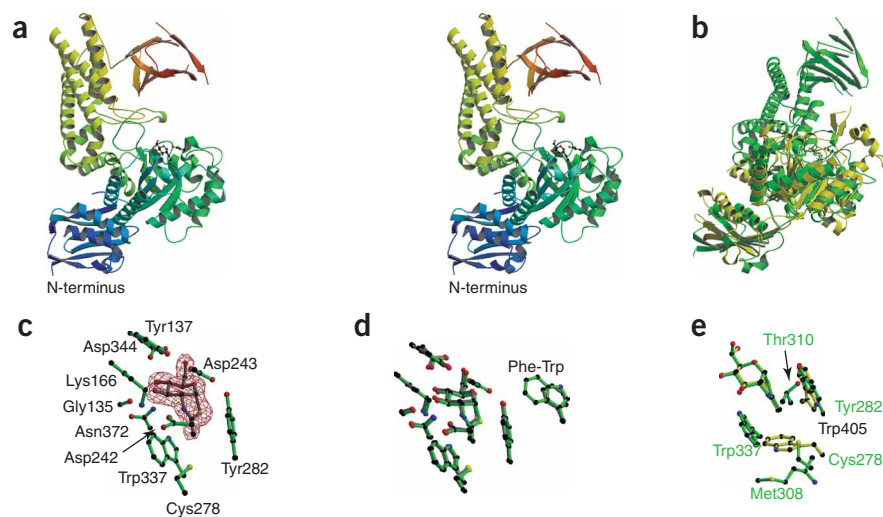


Figure 2 Three-dimensional structure of *B. thetaiotaomicron* GH84. (a) Divergent (wall-eyed) stereo cartoon, color-ramped from, N-terminus (blue) to C-terminus (red), with NAG-thiazoline and the acid/base Asp243 in ball-and-stick representation. (b) Overlay of BtGH84 (green) with the human hexosaminidase B from family GH20 (yellow). (c) Electron density ($F_{\text{obs}} - F_{\text{calc}}$, omit map at 3 σ) for the NAG-thiazoline inhibitor and its environment; kinetic parameters of mutants of these residues are given in Table 1. (d) Overlay of the experimentally determined BtGH84 (green) with a homology model of the human enzyme (gray) showing only a single Trp-Phe change near the active center. (e) Overlay of the acetamide pocket of BtGH84 (green) with human GH20 hexosaminidase B (yellow, black labels), revealing a smaller pocket in the latter by virtue of the steric blockage of Trp405. This difference in pocket volume has been exploited in the design of GH84-specific inhibitors, such as the *N*-butyl compounds used here. These figures were drawn with MolScript⁴⁵ and BobScript⁴⁶.

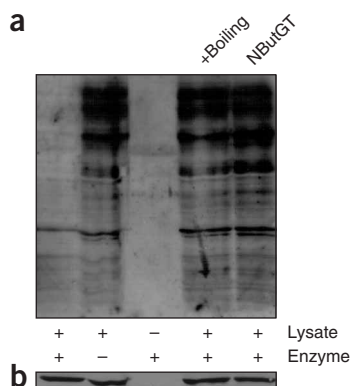


Figure 3 BtGH84 cleaves *O*-GlcNAc from post-translationally modified eukaryotic proteins. Cellular extracts from COS-7 cells were treated with boiled or untreated BtGH84 in the presence or absence of NButGT and analyzed by western blotting. **(a)** Western blot showing abundance of *O*-GlcNAc modifications on proteins from cell lysate, using a monoclonal antibody to *O*-GlcNAc, CTD 110.6, followed by an antibody recognizing mouse IgM that is conjugated to horseradish peroxidase. **(b)** Western blot from **a** stripped and reprobed using an antibody recognizing β -actin, followed by an antibody recognizing mouse IgG that is conjugated to horseradish peroxidase. Blot reveals equivalent β -actin levels in each sample. Human *O*-GlcNAcase has previously been shown by western blotting to process *O*-GlcNAc-modified proteins⁴⁷.

two sheets of 5 + 6 β -strands. Electron density peters out at one end of the sheets, where there is also no density for the interconnecting loops. We were able to model the sequences of 9 of these 11 strands, with the remaining two built as polyalanine. Sequence searches suggest this domain is found only in *Bacteroides* spp. GH84 enzymes and the NagI GH84 of *Clostridium perfringens*. Structural similarity searches using DALI revealed that this domain is topologically identical and structurally similar to a range of carbohydrate-binding modules from family 32 (ref. 27), most notably (DALI Z-scores >6) those from galactose oxidase and bacterial sialidases, which are believed to interact with cell-wall glycans²⁷. The carbohydrate-recognition site of these bacterial CBM domains (also called lectin domains) is in the loop region of the CBM that is disordered in the BtGH84 structure, probably because the correct ligand is absent. The role of this CBM-like domain may also be targeting of the catalytic domain through interaction with as-yet-undefined human glycans, for purposes of foraging²⁸.

The active site structure of GH84

To gain insight into which amino acid residues are crucial for substrate recognition and catalysis, the structure of BtGH84 was determined in complex with a strong competitive inhibitor of human *O*-GlcNAcase termed NAG-thiazoline²⁹, which mimics the intermediate (or a closely derived transition state) in the substrate-assisted catalytic mechanism (Fig. 1). The electron density maps clearly reveal the presence of the inhibitor (refining with a *B* value of 20 \AA^2 and unitary occupancy) within a depression on the C-terminal side of the $(\beta/\alpha)_8$ -barrel domain of BtGH84 (Fig. 2).

The pyranose ring of NAG-thiazoline adopts a slightly distorted 4C_1 (chair) conformation, forming numerous contacts with the protein (Fig. 2). A hydrophobic platform, provided by Trp337, stacks against the thiazoline ring of the inhibitor. In the native enzyme, an acetate molecule overlaps exactly with the thioacetimidate moiety of the inhibitor, and a glycerol derived from the cryo-protectant overlaps with C4, O4, C5, O5, C6 and O6 of the sugar portion of the inhibitor

(data not shown). The inhibitor makes several interactions with the enzyme (Supplementary Figure 1 online): both the O6 and O4 hydroxyls hydrogen-bond to Asp344, and the O4 hydroxyl also receives a hydrogen bond from the side chain amide of Asn372. The O3 hydroxyl donates a hydrogen bond to the main chain carbonyl of Gly135 and receives a hydrogen bond from Lys166. The nitrogen of the thiazoline group hydrogen-bonds to Asp242 of the catalytic Asp-Asp pair, with Asp243 acting as the catalytic acid/base (as discussed below).

Notably, this active site is composed almost entirely of residues that are conserved within human *O*-GlcNAcase (Fig. 2d). Such a stringent structural conservation suggests that there has been relatively little evolutionary pressure to adapt the enzyme for different functions in the bacterium. We therefore sought to evaluate whether BtGH84 could, like human *O*-GlcNAcase¹², process *O*-GlcNAc-modified proteins as well as thioglycosides³⁰. We found that BtGH84 does process *O*-GlcNAc-modified eukaryotic proteins (Fig. 3) and also cleaves p-nitrophenyl-S-GlcNAc (pNP-S-GlcNAc), albeit with a lower efficiency than that of the human enzyme (Table 1). Accordingly, we decided to evaluate further how appropriate a model BtGH84 is for human *O*-GlcNAcase, in terms of the details of the catalytic mechanism.

Similarities between human *O*-GlcNAcase and BtGH84

Almost the entire catalytic center, indeed all but one amino acid residue, is conserved within the active site architecture of BtGH84 when compared to human *O*-GlcNAcase (Fig. 2d). Combined with the generally high level of sequence conservation between the catalytic domains of these two enzymes, this structural similarity immediately suggests a common catalytic mechanism. Unlike the vast majority of 'retaining' glycoside hydrolases, which use a double-displacement mechanism involving the transient formation of a covalent glycosyl enzyme intermediate³¹, it has been proposed²⁹ that the human GH84 *O*-GlcNAcases use a catalytic mechanism involving substrate-assisted catalysis from the 2-acetamido group of the substrate (Fig. 1b). The enzyme-catalyzed reaction is thus thought to proceed in two discrete steps through the transient formation of an oxazoline or oxazolinium ion intermediate. By analogy to other glycosidases using such substrate-assisted mechanisms, it is also expected that two key catalytic carboxyl residues fulfill crucially important roles.

In the first, 'cyclization' step, departure of the aglycon leaving group is promoted by an appropriately positioned general-acid catalytic residue. Concomitant with expulsion of this leaving group, attack of the 2-acetamido group on the anomeric center is facilitated by a second carboxyl group that acts to polarize the acetamido group, thereby enhancing its nucleophilicity. This second carboxyl group acts to position the acetamido group and stabilize the resulting oxazoline intermediate, most probably as a catalytic general base. After the aglycon departs, a molecule of water takes its place and the second chemical step of the reaction, the 'ring-opening' step, occurs by a process that is almost exactly the reverse of the first step. In the ring-opening step, departure of the acetamido group from the anomeric center is facilitated by the second carboxyl group, which now most probably acts as a general acid. Concomitantly, the general acid of the first step acts as a general base, facilitating attack of the water molecule on the anomeric center. Both of these steps are thought to proceed through highly dissociative, oxocarbenium ion-like transition states (Fig. 1c), in an 'electrophilic migration' mechanism³¹.

Identities of the catalytic residues

The structure of BtGH84 in complex with NAG-thiazoline revealed here provides powerful insights into the identities of these key catalytic

Table 1 Michaelis-Menten parameters for aryl glycosides with BtGH84

Enzyme and substrate	σ^{*a}	K_m (mM)	$V_{max}/[E]_0$ ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	$V_{max}/K_m[E]_0$ ($\mu\text{mol mM}^{-1}\text{min}^{-1} \text{mg}^{-1}$)
WT (pH 7.4)				
pNP- <i>O</i> -GlcNAc	0	0.30 ± 0.04	3.2 ± 0.1	11 ± 2
pNP- <i>S</i> -GlcNAc	0	1.4 ± 0.1	0.054 ± 0.001	0.039 ± 0.003
Mu-GlcNAc	0	ND ^b	ND ^b	14 ± 1 ^c
Mu-GlcNAc-F ₁	0.8	ND ^b	ND ^b	0.25 ± 0.02 ^c
Mu-GlcNAc-F ₂	2.0	ND ^b	ND ^b	0.024 ± 0.01 ^c
Mu-GlcNAc-F ₃	2.8	ND ^b	ND ^b	0.00020 ± 0.00001 ^c
pNP- <i>O</i> -GlcNAc (pH 6.5)				
WT	0	0.28 ± 0.01	19.9 ± 0.2	71 ± 3
D242A	0	4.4 ± 0.3	0.110 ± 0.003	0.025 ± 0.002
D242N	0	1.47 ± 0.04	0.080 ± 0.001	0.054 ± 0.002
D243A	0	0.092 ± 0.04	0.059 ± 0.001	0.64 ± 0.04
D243N	0	0.016 ± 0.001	0.18 ± 0.002	11.3 ± 0.8
N372A	0	36 ± 18 ^d	40 ± 20 ^d	0.96 ± 0.02 ^c
Y137F	0	1.18 ± 0.03	42.5 ± 0.3	36 ± 1
Y282F	0	0.41 ± 0.02	0.92 ± 0.01	2.2 ± 0.2

^aTaft parameters (σ^*) used for *N*-acyl substituents were obtained from ref. 48. ^bValues were not determined, as saturation kinetics were not observed for these substrates owing to limitations in substrate solubilities. ^cValues were determined by linear regression of the second-order region of the Michaelis-Menten plot. ^dValues are estimates, as substrate concentrations used were less than K_m and saturation kinetics was not observed.

groups as well as their immediate environments and their hydrogen bond partners. The only carboxylic acid residue suitably positioned to fulfill the role of general-acid/base catalytic residue is Asp243, whose pK_a is probably elevated by its proximity to Asp242. This catalytic residue occupies a position above the plane of the pyranose ring and is flanked by two tyrosine residues, each one of which makes a hydrogen bond to a different O δ (Fig. 2c). These tyrosine residues position a *syn* lone pair of the carboxyl group in the appropriate orientation to deliver a proton to the glycosidic oxygen. Site-directed mutants in which either of the tyrosine groups has been changed to phenylalanine show substantially lower catalytic activity (Table 1). Immediately adjacent to Asp243 is another aspartate residue, Asp242, which is precisely positioned to form a good hydrogen bond with the nitrogen of the thiazoline ring of the inhibitor. Asp242 is thus an obvious candidate for the role of the oxazoline-stabilizing residue. Strong support for these proposals comes from the observation that Asp242 and Asp243 overlap with carboxylates in the structures of other enzymes that use the 2-acetamido group as a participating group, including GH18 chitinases³², GH20 hexosaminidases^{24,33} and GH56 hyaluronidases³⁴. Thus, the key catalytic residues of family 84 glycoside hydrolases are a tandem Asp-Asp pair; this contrasts with the other *exo*-acting enzymes using anchimeric assistance, from family 20, which use an adjacent Asp-Glu catalytic pair, and the *endo*-acting enzymes from families 18 (ref. 32) and 56 (ref. 34), which use an Asp-X-Glu motif (where X is any residue). Both in the family 20 human β -hexosaminidases^{24,33} and in the structure of BtGH84 described here in complex with NAG-thiazoline, it is the *anti* lone pair of the aspartate that is engaged with the nitrogen of the thiazoline ring (Fig. 2c). One of the *syn* lone pairs forms a hydrogen bond (Fig. 2c) with a conserved lysine residue (here 166) that may act to modulate the pK_a of this catalytic group.

Site-directed variants (Table 1 and Supplementary Figure 2 online) of these two catalytic groups (D242A and D243A) have markedly lower catalytic efficiencies (2,800- and 110-fold lower, respectively) than that of the wild-type enzyme. These deficits are similar to those observed for human *O*-GlcNAcase, where deletion of either of these carboxyl groups results in substantially lower catalytic efficiencies (3,000-fold for the residue corresponding to Asp242 and seven-fold

for the residue corresponding to Asp243)³⁵. Mutation of either of these residues to asparagine results in similar decreases in $V_{max}/K_m[E]_0$ values (where $[E]_0$ is the concentration of enzyme; Table 1). It is most likely that these two carboxyl residues must be in different protonation states for efficient catalysis³⁵. Consistent with such an expectation, we observed a bell-shaped pH-activity profile (Fig. 4a). For BtGH84, the pH optimum is approximately 6.0, as compared to 6.5 for human *O*-GlcNAcase. Unlike the human enzyme^{13,35}, which shows a normal, fairly broad bell-shaped pH-activity profile, the pH-activity profile for BtGH84 is extremely sharp, suggesting that in BtGH84 the pK_a values of the catalytic carboxyl groups are very similar.

Role of the 2-acetamido group in catalysis

The substrate-assisted catalytic mechanism previously proposed for enzymes cleaving the glycosidic linkage of GlcNAc residues such as chitinase³², chitobiase³³ and hyaluronidase³⁴ has recently been advanced for the human *O*-GlcNAcase^{29,30,35}. In this mechanism, the acetamido group of the substrate acts as a nucleophile in a double-displacement reaction (Fig. 1b). Accordingly, such enzymes use extensive noncovalent interactions to both enhance the nucleophilicity of the acetamido group as well as carefully position it to intercept the migrating electrophilic center. In addition to the

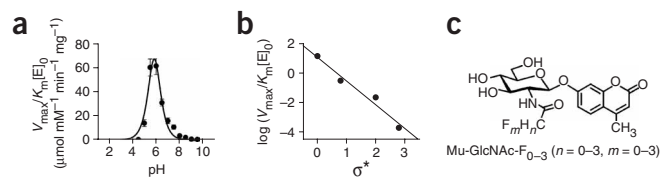


Figure 4 pH-activity profile of BtGH84 and Taft-like analysis. (a) pH dependence of $\log V_{max}/K_m[E]_0$. Full Michaelis-Menten parameters were determined at each pH using pNP-*O*-GlcNAc as substrate. (b) Linear free-energy analysis plotting the Taft parameter (σ^*) of the *N*-fluoroacetyl substituent of each 4-methylumbelliferyl-GlcNAc (Mu-GlcNAc) substrate analog against the $\log(V_{max}/K_m[E]_0)$ measured for that substrate with BtGH84. (c) General structure of Mu-GlcNAc-F₀₋₃ substrates used in the Taft-like analysis shown in b.

observation of potent inhibition by NAG-thiazoline, which resembles the reaction intermediate, a second unambiguous test of whether the acetamido group participates in catalysis is a kinetics study using substrates having differing levels of fluorine substitution on the *N*-acetyl group (Fig. 4). Increasing numbers of electron-withdrawing fluorine substituents successively decrease the basicity of the carbonyl oxygen, with the net effect for enzymes using substrate-assisted catalysis being a markedly lower second-order rate constant with each addition. For BtGH84, a plot of $\log(V_{\max}/K_m[E]_0)$ against the Taft electronic parameter (σ^*) of the *N*-acyl substituent (Fig. 4b) reveals a strong negative correlation (reaction constant $\rho = -1.6 \pm 0.2$), consistent with the carbonyl oxygen acting as a nucleophile to attack the anomeric center. The large magnitude of this slope resembles that measured for the family 20 human β -hexosaminidase ($\rho = -1.0 \pm 0.1$)²⁹. With human *O*-GlcNAcase, a negative correlation is also observed, but the slope is more gentle ($\rho = -0.42 \pm 0.08$)²⁹.

DISCUSSION

Within the family 20 β -hexosaminidases, recent X-ray crystal structures have shown that a tryptophan-lined pocket cradles the acetamido group^{24,25,33}, and kinetic studies of the human family GH20 β -hexosaminidase have revealed that this pocket does not tolerate increases in the bulk of the *N*-acyl side chain²⁹. In the structure of BtGH84, the acetamido group is also tightly nestled by several residues. It is noteworthy that the thiazoline ring of the inhibitor is sandwiched by the two aromatic residues Tyr282 and Trp337, and in the family 20 hexosaminidases two tryptophan residues adopt a similar configuration^{24,25,33}. This common strategic arrangement of aromatic residues in both families of enzymes must serve to accurately position the acetamido group. Furthermore, such a spatial arrangement probably also acts to stabilize positive charge development within the oxazoline ring at the transition state, through orbital interactions.

Greater differences between these two families of enzymes are seen, however, when the residues lining the bottom edge of the acetamido pocket are compared (Fig. 2e). The family 20 β -hexosaminidases have a pocket that is much more tightly packed, formed by a cage of only three tryptophan residues, a tyrosine residue and the general-base/acid catalyst^{24,25,33}. The structure of BtGH84 reveals a more spacious acetamido pocket composed of two aromatic residues and three other residues having greater conformational flexibility. The bottom of the pocket is sealed with Cys278 and Met308. Three further residues (Val342, Asn339 and Thr310) line the sides of this pocket, and the two aromatic groups (Tyr282 and Trp337) form the walls. Notably, the amino group of Asn339 makes a hydrogen bond with the sulfur of the thiazoline ring. Presumably, this group polarizes and directs the acetamido substituent in a manner resembling that seen for the family 20 enzymes, which use a tyrosine hydroxyl group^{24,25,33}. Toward the bottom of the pocket, Cys278 is 3.8 Å away from the methyl of the acetamido group, whereas Met308 is a distant 6.8 Å. Thus, some space is available within this pocket to accommodate substrates and inhibitors with somewhat bulkier *N*-acyl groups. Together, these observations suggest that the acetamido pocket of BtGH84 is considerably more accommodating than those of the family 20 β -hexosaminidases and that all the residues in this pocket are invariant between human *O*-GlcNAcase and BtGH84 (Fig. 2d).

The tight packing of the acetamido group within the family 20 β -hexosaminidases is underscored by observations showing that inhibitors with increasingly bulky 2-*N*-acyl groups rapidly become poor inhibitors as the size of the side chain increases^{29,36}. Within human *O*-GlcNAcase, however, the same series of substitutions are much better tolerated^{29,36}, a result supported by our crystallographic

observations of an enlarged acetamido pocket in the *Bacteroides* enzyme. To evaluate the plasticity and volume of this pocket and its similarity to human *O*-GlcNAcase, we determined the steric impacts of increasing the bulk of the pendent *N*-acyl group. Five inhibitors of *O*-GlcNAcase were used (Supplementary Figure 1). NAG-thiazoline is a strong inhibitor of human *O*-GlcNAcase ($K_i = 70$ nM), and an analog in which the acyl group is increased in length by two methylene units (NButGT) retains its potency ($K_i = 230$ nM)²⁹. A similar derivative of PUGNAc (PUGNBut), recently prepared as a series of inhibitors³⁶, binds considerably more poorly ($K_i = 2,400$ nM) than the parent compound ($K_i = 50$ nM)³⁶. Consistent with the great structural similarity of the catalytic domains of these enzymes, both NAG-thiazoline ($K_i = 200$ nM) and PUGNAc ($K_i = 10$ nM) are strong inhibitors of BtGH84 (Supplementary Table 1 online). Similar trends are observed as with the human enzyme, with NButGT binding eight-fold more poorly ($K_i = 2,000$ nM)²⁹ and PUGNBut binding 1,000-fold more poorly ($K_i = 10,000$ nM)³⁶. Streptozotocin (STZ) has previously been proposed to induce β -cell death by inactivating *O*-GlcNAcase through the covalent modification of a cysteine residue in the active center³⁷. We find, however, that STZ is a poor inhibitor of BtGH84 (Supplementary Table 1) and, despite the presence of a cysteine in the active site of these enzymes, it is not an irreversible inhibitor of BtGH84 (data not shown) or human *O*-GlcNAcase²⁹. The toxicity of STZ toward β -cells is most likely a consequence of its well-documented ability to alkylate DNA, which in turn results in DNA fragmentation³⁸. Elevation of *O*-GlcNAc levels is probably a downstream stress response¹⁸.

Although BtGH84 shows a narrower bell-shaped pH-activity profile than does the human *O*-GlcNAcase, and although the Taft-like analysis suggests greater nucleophilic participation in the transition states stabilized by BtGH84, the results described above collectively indicate that BtGH84 is an outstanding model for human *O*-GlcNAcase. The two enzymes share invariant structural elements within the active site, a conserved catalytic mechanism and similar affinities and plasticity toward inhibitors, and mutation of their key residues results in parallel defects in catalysis. In contradiction of frequent genomic functional annotations, we observed no hyaluronidase activity for BtGH84 (data not shown). This observation is consistent with the pocket shape of the active site, which precludes cleavage of oligosaccharides in the middle of the chain (Supplementary Figure 3 online). Existing inhibitors of *O*-GlcNAcase show reasonable potency^{7,29}, and more recently, selective inhibitors of human *O*-GlcNAcase over human β -hexosaminidase have been developed^{29,36} that discriminate on the basis of the size of the acetamido-binding pocket. These inhibitors may have limited bioavailability, and therefore their structure-guided modification, or the development of entirely new inhibitors, may result in compounds with improved properties. In addition to acting as potential therapeutics, such inhibitors may prove useful in modulating *O*-GlcNAc levels within various tissues for the purposes of gaining basic insights into the cellular role of *O*-GlcNAc.

Note added in proof: The structure of a related GH84 enzyme, NagJ from *Clostridium perfringens*, has also recently been determined³⁹. The structure of the NagJ PUGNAc complex, in tandem with site-directed variants, also provides strong evidence for a close similarity to the human *O*-GlcNAcase.

METHODS

Cloning and overexpression. The *B. thetaiotaomicron* GH84 ORF was truncated to produce an N-terminally tagged recombinant protein beginning at residue Gln22, to omit the predicted signal peptide. The ORF was amplified

from a *B. thetaiotaomicron* VPI 5482, ATCC 29148 (ref. 21) genomic DNA template using KOD Hot Start DNA polymerase (Invitrogen) and the pair of primers 5'-CAC CAC CAC CAC CAA TGG AAT GTT AGT CTG CAA CCT CCT CCT CA-3' and 5'-GAG GAG AAG GCG CGT TAT TAT TTC TTC TCT ATA GTC AGA ACA AAC TGG CG-3' according to the manufacturer's instructions. These primers contained ligation-independent cloning-compatible sequences; the resultant 2,184-base-pair product was annealed into ysbLIC pET28 to produce the expression construct pGH84.

Production and purification of BtGH84. N-terminally His₆-tagged BtGH84 was produced via overnight isopropyl IPTG induction of pGH84-containing *Escherichia coli* BL21 (DE3; Novagen) cultures grown at 37 °C to an absorbance of A₆₀₀ = 0.6. BtGH84 was purified from cell-free extracts via affinity chromatography using nickel-affinity chromatography followed by gel filtration. Purified protein was concentrated and exchanged into 20 mM HEPES (pH 7.5), using a 10-kDa-cutoff concentrator. Selenomethionine (SeMet) BtGH84 was produced using the methionine auxotroph B834 (DE3).

Kinetic analyses. All Michaelis-Menten kinetics experiments were carried out using substrate concentrations ranging from 1/5 to 5 times K_m at 37 °C. Reactions were carried out in thermally equilibrated 1.2-ml disposable acryl cuvettes (for the pH-activity studies) or 50- μ l quartz microcuvettes (for all other studies). Reactions were initiated by adding a small aliquot of enzyme using a syringe and then continuously monitored at the specified wavelength using a Cary 3E UV-VIS spectrophotometer equipped with a Peltier temperature controller. Studies on the pH dependence of enzyme-catalyzed hydrolysis of pNP-O-GlcNAc used continuous monitoring at 360 nm for assays at pH 4.5 and 5.0, and at 400 nm for assays at pH 6.0–9.5. Reaction velocities were determined by linear regression of the linear region of the reaction progress curve between the first and third minutes. A buffer of 50 mM citrate (for reactions at pH 4.5–5.5), phosphate (for reactions at pH 6.0–8.0) or Ches (for reactions at pH 8.0–9.5) with 100 mM NaCl was used. For inhibition studies, a minimum of eight inhibitor concentrations were tested, ranging from 1/5 to 5 times K_i. These reactions were carried out in the pH 6.5 phosphate buffer described above, using pNP-O-GlcNAc as a substrate, and were monitored at 400 nm. K_i values were determined by linear regression of data in Dixon plots. Thioglycosidase activity was monitored at 407 nm using pNP-S-GlcNAc as substrate in phosphate-buffered saline (pH 7.4). The enzyme concentration in this assay was 0.06 μ g μ l⁻¹.

Taft-like analyses. Enzymatic reactions were carried out in PBS buffer (pH 7.4) and monitored at 360 nm. Reaction velocities were determined by linear regression of the linear region of the reaction progress curve between the 1st and 20th minutes. The concentrations (μ g μ l⁻¹) of BtGH84 used in assays with fluorinated substrates were as follows: F₀, 0.00075; F₁, 0.0075; F₂, 0.075; F₃, 2.6.

Time-dependent inactivation assay using STZ. The possible time-dependent inactivation of BtGH84 was assayed by incubating BtGH84 (1 mg ml⁻¹) with either 10 mM STZ or no STZ in pH 6.5 buffer containing 50 mM NaH₂PO₄, 100 mM NaCl, 0.1% (v/v) BSA and 5 mM β -mercaptoethanol. An additional experiment was carried out where 10 mM STZ was incubated with the buffer but with no BSA or β -mercaptoethanol present. At several time intervals, the residual enzyme activity contained in the inactivation mixture was assayed. Enzyme assays were initiated by the addition of an aliquot (10 μ l) of the inactivation mixture to an assay mixture (1,500 μ l) containing 0.5 mM pNP-O-GlcNAc.

Activity toward O-GlcNAc-modified proteins. Four 10-cm plates of COS-7 cells were cultured to confluence in DMEM medium (Invitrogen) supplemented with 5%–10% (v/v) FBS (Invitrogen). Cells were harvested by scraping and pooled by centrifugation (200g, 10 min). After the supernatant was discarded, the cells were treated at 4 °C with cold lysis buffer (1 ml of PBS (pH 7.4), containing 1 mM EDTA, 1 mM PMSF, 1% (v/v) NP-40 and 0.5% (w/v) sodium deoxycholate). After 20 min at 4 °C the solution was centrifuged at 14,000 r.p.m. in an Eppendorf 5415C microcentrifuge and the supernatant was collected. Assays consisted of 250 μ l cell lysate, 0.15 mg ml⁻¹ recombinant BtGH84, 5 mM β -mercaptoethanol and 10 mM imidazole. In some cases, the enzyme was either preboiled for 10 min or 10 mM NAG-thiazoline was added. Assays (300 μ l) were then incubated overnight at

Table 2 Data collection, phasing and refinement statistics for the structure solution of *B. thetaiotaomicron* GH84 O-GlcNAcase

	Native	SeMet with thiazoline
Data collection		
Space group	P1	P1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	51.7, 94.5, 99.5	51.7, 94.5, 99.7
α , β , γ (°)	75.3, 93.9, 77.1	75.3, 93.8, 77.0
		<i>Peak</i>
Wavelength	0.9393	0.97950
Resolution (Å)	30–1.85 (1.95–1.85) ^a	50–1.95 (2.0–1.95) ^a
<i>R</i> _{sym} or <i>R</i> _{merge}	0.090 (0.49)	0.086 (0.45)
<i>I</i> / σ <i>I</i>	13 (3)	12 (3)
Completeness (%)	97 (96)	98 (97)
Redundancy	3.9 (3.8)	4 (4)
Refinement		
Resolution (Å)	1.85	1.95
No. reflections	135,175	117,960
<i>R</i> _{work} / <i>R</i> _{free}	0.18 / 0.22	0.18 / 0.23
No. atoms		
Protein	10,612	10,388
Ligand	32	28
Water	1,353	1,351
<i>B</i> -factors		
Protein	24	32
Ligand	22	21
Water	40	43
R.m.s deviations		
Bond lengths (Å)	0.015	0.017
Bond angles (°)	1.5	1.6

^aHighest-resolution shell is shown in parentheses.

37 °C and passed through a small nickel column to remove the His₆-tagged BtGH84. An aliquot of each reaction was then separated by SDS-PAGE, transferred onto nitrocellulose membrane (0.45 μ m; Bio-Rad) and probed as previously described²⁹.

Crystallization and structure solution. BtGH84 crystals were grown using hanging drop vapor diffusion from 0.5 M sodium acetate, 15% (v/v) PEG 3500, 0.1 M MES (pH 6.0) and 20% (v/v) glycerol. An additional 10% glycerol was included before flash-freezing in liquid nitrogen. Crystals of BtGH84 in complex with inhibitors were obtained by cocrystallization with 10 mM ligand. X-ray data were collected from single crystals at 100 K at the European Synchrotron Radiation Facility. Selenium data were collected on beamline ID 23-1 and native data were collected on ID 14-4. All X-ray data were processed with MOSFLM from the CCP4 suite⁴⁰ (Table 2).

The structure was solved using SeMet single-wavelength data collected at the selenium edge. The positions of the selenium sites were determined using SHELXD⁴¹ and initial maps calculated using SHELXE. Contrast and connectivity indicators as given by SHELXE clearly distinguished the correct hand. Five percent of the data were flagged for cross-validation, and REFMAC⁴² and ARP/wARP⁴³ were used to automatically build a partial (500 residues) initial model. COOT⁴⁴ was used to make manual corrections and augmentations to the model. Solvent molecules were added using COOT followed by ARP/wARP and checked manually.

Accession codes. Protein Data Bank: Coordinates have been deposited with accession codes 2CHN and 2CHO for the thiazoline complex and native enzyme, respectively.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

R.J.D. thanks the Biotechnology and Biological Sciences Research Council (BBSRC) for a PhD fellowship. This work was supported by grants from the BBSRC to G.J.D. and from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Protein Engineering Network of Centres of Excellence to D.J.V. D.J.V. is supported as a Tier II Canada Research Chair in Chemical Glycobiology. M.S.M. thanks NSERC and the Michael Smith Foundation for Health Research for fellowships. We also thank A.J. Bennet for access to equipment. This work was funded by the BBSRC, the Royal Society of the UK and the NSERC.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/nsmb/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Torres, C.R. & Hart, G.W. Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for *O*-linked GlcNAc. *J. Biol. Chem.* **259**, 3308–3317 (1984).
- Kamemura, K., Hayes, B.K., Comer, F.I. & Hart, G.W. Dynamic interplay between *O*-glycosylation and *O*-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J. Biol. Chem.* **277**, 19229–19235 (2002).
- Cheng, X., Cole, R.N., Zaia, J. & Hart, G.W. Alternative *O*-glycosylation/*O*-phosphorylation of the murine estrogen receptor beta. *Biochemistry* **39**, 11609–11620 (2000).
- Lefebvre, T. *et al.* Effect of okadaic acid on *O*-linked *N*-acetylglucosamine levels in a neuroblastoma cell line. *Biochim. Biophys. Acta* **1472**, 71–81 (1999).
- Chou, C.F., Smith, A.J. & Omary, M.B. Characterization and dynamics of *O*-linked glycosylation of human cytokeratin 8 and 18. *J. Biol. Chem.* **267**, 3901–3906 (1992).
- Davis, L.I. & Blobel, G. Nuclear pore complex contains a family of glycoproteins that includes p62: glycosylation through a previously unidentified cellular pathway. *Proc. Natl. Acad. Sci. USA* **84**, 7552–7556 (1987).
- Dong, D.L., Xu, Z.S., Hart, G.W. & Cleveland, D.W. Cytoplasmic *O*-GlcNAc modification of the head domain and the KSP repeat motif of the neurofilament protein neurofilament-H. *J. Biol. Chem.* **271**, 20845–20852 (1996).
- Jackson, S.P. & Tjian, R. *O*-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **55**, 125–133 (1988).
- Zhang, F. *et al.* *O*-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115**, 715–725 (2003).
- Lubas, W.A., Frank, D.W., Krause, M. & Hanover, J.A. *O*-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9316–9324 (1997).
- Kreppel, L.K., Blomberg, M.A. & Hart, G.W. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique *O*-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308–9315 (1997).
- Dong, D.L. & Hart, G.W. Purification and characterization of an *O*-GlcNAc selective *N*-acetyl-beta-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269**, 19321–19330 (1994).
- Gao, Y., Wells, L., Comer, F.I., Parker, G.J. & Hart, G.W. Dynamic *O*-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-*N*-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838–9845 (2001).
- Shafi, R. *et al.* The *O*-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl. Acad. Sci. USA* **97**, 5735–5739 (2000).
- Lehman, D.M. *et al.* A single nucleotide polymorphism in MGEA5 encoding *O*-GlcNAc-selective *N*-acetyl-beta-D-glucosaminidase is associated with type 2 diabetes in Mexican Americans. *Diabetes* **54**, 1214–1221 (2005).
- Vosseller, K., Wells, L., Lane, M.D. & Hart, G.W. Elevated nucleocytoplasmic glycosylation by *O*-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *Proc. Natl. Acad. Sci. USA* **99**, 5313–5318 (2002).
- Jinek, M. *et al.* The superhelical TPR-repeat domain of *O*-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004).
- Zachara, N.E. *et al.* Dynamic *O*-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J. Biol. Chem.* **279**, 30133–30142 (2004).
- Toleman, C., Paterson, A.J., Whisenhunt, T.R. & Kudlow, J.E. Characterization of the HAT domain of a bifunctional protein with activatable *O*-GlcNAcase and HAT activities. *J. Biol. Chem.* **279**, 53665–53673 (2004).
- Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316 (1991).
- Xu, J. *et al.* A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
- Knapp, S. *et al.* NAG-thiazoline, an *N*-acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation. *J. Am. Chem. Soc.* **118**, 6804–6805 (1996).
- Mohan, H. & Vasella, A. An improved synthesis of 2-acetamido-2-deoxy-D-glucohydroxymolactone (PUGNAc), a strong inhibitor of beta-*N*-acetylglucosaminidases. *Helv. Chim. Acta* **83**, 114–118 (2000).
- Mark, B.L. *et al.* Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J. Mol. Biol.* **327**, 1093–1109 (2003).
- Mark, B.L. *et al.* Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J. Biol. Chem.* **276**, 10330–10337 (2001).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *Trends Biochem. Sci.* **20**, 478–480 (1995).
- Newstead, S.L., Watson, J.N., Bennet, A.J. & Taylor, G. Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 1483–1491 (2005).
- Sonnenburg, J.L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
- Macauley, M.S., Whitworth, G.E., Debowski, A.W., Chin, D. & Vocadlo, D.J. *O*-GlcNAcase uses substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors. *J. Biol. Chem.* **280**, 25313–25322 (2005).
- Macauley, M.S., Stubbs, K.A. & Vocadlo, D.J. *O*-GlcNAcase catalyzes cleavage of thioglycosides without general acid catalysis. *J. Am. Chem. Soc.* **127**, 17202–17203 (2005).
- Vocadlo, D.J., Davies, G.J., Laine, R. & Withers, S.G. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835–838 (2001).
- Terwisscha van Scheltinga, A.C. *et al.* Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry* **34**, 15619–15623 (1995).
- Tews, I. *et al.* Bacterial chitinase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat. Struct. Biol.* **3**, 638–648 (1996).
- Markovic-Housley, Z. *et al.* Crystal structure of hyaluronidase, a major allergen of bee venom. *Struct. Fold. Des.* **8**, 1025–1035 (2000).
- Çetinbas, N., Macauley, M.S., Stubbs, K.A., Drapala, R. & Vocadlo, D.J. Identification of Asp174 and Asp175 as the key catalytic residues of human *O*-GlcNAcase by functional analysis of site-directed mutants. *Biochemistry* **45**, 3835–3844 (2006).
- Stubbs, K.A., Zhang, N. & Vocadlo, D.J. A divergent synthesis of 2-acyl derivatives of PUGNAc yields selective inhibitors of *O*-GlcNAcase. *Org. Biomol. Chem.* **4**, 839–845 (2006).
- Roos, M.D. *et al.* Streptozotocin, an analog of *N*-acetylglucosamine, blocks the removal of *O*-GlcNAc from intracellular proteins. *Proc. Assoc. Am. Physicians* **110**, 422–432 (1998).
- Boizan, A.D. & Bianchi, M.S. Genotoxicity of streptozotocin. *Mutat. Res.* **512**, 121–134 (2002).
- Rao, F.V. *et al.* Structural insights into the mechanism and inhibition of eukaryotic *O*-GlcNAc hydrolysis. *EMBO J.* (in the press).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
- Schneider, T.R. & Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
- Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
- Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Esnouf, R.M. An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J. Mol. Graph. Model.* **15**, 132–134 (1997).
- Wells, L. *et al.* Dynamic *O*-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic beta-*N*-acetylglucosaminidase, *O*-GlcNAcase. *J. Biol. Chem.* **277**, 1755–1761 (2002).
- Hansch, C. & Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley, New York, 1979).