Biochem. J. (2010) **432**, 1–7 (Printed in Great Britain) doi:10.1042/BJ20101338

ACCELERATED PUBLICATION

Human OGA binds substrates in a conserved peptide recognition groove

Marianne SCHIMPL[1] , Alexander W. SCHÜTTELKOPF[1] , Vladimir S. BORODKIN and Daan M. F. VAN AALTEN[2]

Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, U.K.

Modification of cellular proteins with *O*-GlcNAc (O-linked *N*-acetylglucosamine) competes with protein phosphorylation and regulates a plethora of cellular processes. *O*-GlcNAcylation is orchestrated by two opposing enzymes, *O*-GlcNAc transferase and OGA (*O*-GlcNAcase or β-N-acetylglucosaminidase), which recognize their target proteins via as yet unidentified mechanisms. In the present study, we uncovered the first insights into the mechanism of substrate recognition by human OGA. The structure of a novel bacterial OGA orthologue reveals a putative substrate-binding groove, conserved in metazoan OGAs. Guided by this structure, conserved amino acids lining this groove in human OGA were mutated and the activity on three different substrate proteins [TAB1 (transforming growth factor-β-activated protein kinase 1-binding protein 1), FoxO1 (forkhead box O1) and CREB (cAMP-response-element-binding protein)] was tested in an *in vitro* deglycosylation assay. The results provide the first evidence that human OGA may possess a substrate-recognition mechanism that involves interactions with *O*-GlcNAcylated proteins beyond the GlcNAc-binding site, with possible implications for differential regulation of cycling of *O*-GlcNAc on different proteins.

Key words: β-N-acetylglucosaminidase, O-linked *N*-acetylglucosamine (*O*-GlcNAc), peptide recognition groove, protein glycosylation.

INTRODUCTION

The *O*-GlcNAc (O-linked *N*-acetylglucosamine) modification is a dynamic and reversible form of protein glycosylation occurring on specific serine and threonine residues of intracellular proteins [1,2]. Since the initial discovery of *O*-GlcNAc [3], technological advances have greatly facilitated its detection, and proteomics studies [4–6] have shown that a significant proportion of cellular proteins are *O*-GlcNAcylated. However, the functional importance of *O*-GlcNAc is only just emerging, with evidence to suggest that it may regulate protein activity in a manner analogous (and complementary) to phosphorylation [7]. *O*-GlcNAc levels are known to respond dynamically to nutrient availability [1] and stress [8], and to undergo changes during the cell cycle [9] and development [10]. *O*-GlcNAc has been shown to be associated with a range of human diseases [2]. Strikingly, only two enzymes orchestrate the *O*-GlcNAc modification. Both the OGT (*O*-GlcNAc transferase) and its antagonistic OGA (*O*-GlcNAcase or β-N-acetylglucosaminidase) are encoded by single genes in metazoa [11,12]. Little is known about the mode of substrate recognition of either enzyme, although studies have hinted at the existence of a defined *O*-GlcNAc peptide sequon [2,5]. In addition, it is thought that, for a single OGT to modify specific sites on a multitude of proteins, interactions with adapter proteins may be instrumental in achieving selectivity and regulation of activity [13].

OGA was first identified as a neutral-pH hexosaminidase, which was later associated with the gene *MGEA5* (meningioma-expressed antigen 5) [14]. The domain structure of hOGA (human OGA) is frequently oversimplified by designating an N-terminal hexosaminidase followed by a C-terminal domain with sequence homology with HATs (histone acetyltransferases) (referred to as the HAT domain in the present paper). A further putative domain may exist in the stretch of over 300 amino acids in between the hexosaminidase and HAT domains. hOGA (Ser^{405}) is itself a substrate for OGT [15]. Residues 404–548 have been shown to be required for interaction with OGT [16], and it is thought that the formation of such a complex negatively regulates *O*-GlcNAc removal.

Pharmacological inhibition of OGA has been an invaluable tool for investigating the cellular functions of *O*-GlcNAc; notably the non-selective inhibitor PUGNAc [*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate] has been used extensively to raise cellular *O*-GlcNAc levels [17]. Advances in understanding of the OGA reaction mechanism [18], in conjunction with structures of bacterial OGAs [19,20], have informed the development of a number of highly potent and selective inhibitors for hOGA [21].

Despite these recent advances in our understanding of OGA structure, mechanism and inhibition, we do not understand how OGA interacts with the protein part of its substrates. Such interactions could constitute the basis of substrate specificity in hOGA. In the present paper, we reveal that hOGA possesses a conserved substrate-binding groove that is required for the interaction with *O*-GlcNAc proteins. Different substrates require distinct surface residues for binding, affording a first glimpse into the substrate selectivity of human *O*-GlcNAcase.

Abbreviations used: *Cp*NagJ, *Clostridium perfringens* NagJ; CREB, cAMP-response-element-binding protein; Fmoc, fluoren-9-ylmethoxycarbonyl; FoxO1, forkhead box O1; GST, glutathione transferase; HAT, histone acetyltransferase; HEK, human embryonic kidney; LC, liquid chromatography; 4MU-GlcNAc, 4-methylumbelliferyl-β-*N*-acetylglucosamine; OGA, *O*-GlcNAcase or β-N-acetylglucosaminidase; hOGA, human OGA; *Og*OGA, *Oceanicola granulosus* OGA; OGT, *O*-GlcNAc transferase; *O*-GlcNAc, O-linked *N*-acetylglucosamine; *p*NP-GlcNAc, *p*-nitrophenyl-β-*N*-acetylglucosamine; PUGNAc, *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate; RMSD, root mean square deviation; TAB1, transforming growth factor-β-activated protein kinase 1-binding protein 1.

[1] These authors contributed equally to this work

[2] To whom correspondence should be addressed (email dmfvanaalten@dundee.ac.uk).

Structures have been deposited in the PDB under accession codes 2XSA and 2XSB.

MATERIALS AND METHODS

Cloning, expression and structure determination of *Og*OGA (*Oceanicola granulosus* OGA)

The *O. granulosus* gene for hyaluronoglucosaminidase (ZP_01156836.1) was cloned from genomic DNA into pGEX6-P1 for expression as a GST (glutathione transferase) fusion in *Escherichia coli*. The protein *Og*OGA was purified by affinity, anion-exchange and size-exclusion chromatography (see the Supplementary material at http://www.BiochemJ.org/bj/432/bj4320001add.htm). *Og*OGA was crystallized at 1.9 mg/ml from PEG [poly(ethylene glycol)]/$MgCl_2$ solutions (see the Supplementary Online Data). The structure of the apo-enzyme was solved using MAD (multiwavelength anomalous diffraction) phasing with a lead derivative and the structure of a PUGNAc complex was obtained by molecular replacement. Diffraction data and model statistics are shown in Supplementary Table S1 at http://www.BiochemJ.org/bj/432/bj4320001add.htm.

Glycopeptide synthesis and detection

A 3,4,6-triacetyl-*O*-GlcNAc-Fmoc (fluoren-9-ylmethoxycarbonyl)-Ser-OH derivative was synthesized as described previously [22]. Microwave-assisted solid-phase peptide synthesis was performed with a CEM Liberty instrument on low-load Rink amide MBHA (methylbenzhydrylamine) resin 100–200 mesh (Novabiochem) using standard Fmoc chemistry protocols on a 0.05 mmol scale. The structural homogeneity of the glycopeptides was verified by high-resolution mass spectra taken on Bruker micro-TOF (time-of-flight) instrument in the direct injection mode. The same instrument was used for the monitoring of the enzymatic deglycosylation in the LC (liquid chromatography)–MS mode using an Agilent 1200 binary analytical HPLC system equipped with a Waters X-Bridge C_{18} 3 μm particles, 3.2 mm×50 mm column and diode array UV–visible detector. Typically, the reaction was diluted to 0.3 ml with water, then passed through a 10 kDa molecular-mass cut-off concentrator to remove protein. Then, 10 μl of the flow-through was taken for the analysis using gradient elution (flow rate 0.3 ml/min) with a 5–95 % linear gradient of acetonitrile (0.1 % ammonia) in water (0.1 % ammonia).

hOGA cloning and expression

The coding sequence for full-length hOGA was cloned into the pEBG-6P vector [23] for transient expression in mammalian cells. Point mutations were introduced by site-directed mutagenesis (see Supplementary Table S2 at http://www.BiochemJ.org/bj/432/bj4320001add.htm for primers). The enzyme was expressed as an N-terminal GST fusion in HEK (human embryonic kidney)-293 cells cultivated in suspension. For transfection, 2×10^8 HEK-293 cells were serum-deprived for 4 h in 100 ml of Pro293sCDM medium before the addition of 10 ml of serum-free medium containing 100 μg of plasmid DNA and 600 μg of polyethylenimine. After another 4 h, the cells were diluted to 10^6 cells/ml, and the medium was supplemented with 1 % (v/v) fetal bovine serum. Cell pellets were harvested 72 h after transfection and lysed in 50 mM Tris/HCl (pH 7.5), 270 mM sucrose, 0.1 % 2-mercaptoethanol, 1 mM sodium orthovanadate, 1 mM EDTA, 1 mM EGTA, 5 mM pyrophosphate, 10 mM sodium 2-glycerophosphate, 50 mM NaF, 1 % Triton X-100, 1 mM benzamidine, 0.2 mM PMSF and 5 μM leupeptin. GST–hOGA was purified from cleared lysate by pull-down with glutathione–Sepharose (GE Healthcare), followed by elution with 25 mM reduced glutathione in 25 mM Tris/HCl buffer (pH adjusted to 7.5).

Colorimetric and fluorimetric OGA assays

Initial rates of hydrolysis of *p*NP-GlcNAc (*p*-nitrophenyl-β-*N*-acetylglucosamine) were measured in buffer containing 25 mM Tris/HCl (pH 7.5) and 0.1 mg/ml BSA, with substrate concentrations ranging from 0 to 10 mM with a DMSO concentration of 2 % throughout the assay. Reactions were performed with 5 nM enzyme in a 50 μl assay volume. After 30 min of incubation at 37°C, the reaction was stopped by addition of 100 μl of 3 M glycine/NaOH (pH 10.3), and product concentration was measured at A_{405}. 4MU-GlcNAc (4-methylumbelliferyl-β-*N*-acetylglucosamine) assays were performed in the same buffer, with substrate concentrations ranging from 0 to 8 mM, and the enzyme concentration was reduced to 0.2 nM and the assay time to 20 min. Product formation was determined fluorimetrically ($\lambda_{ex} = 360$ nm/$\lambda_{em} = 460$ nm). Inhibition constants for OGA inhibitors were obtained with 4MU-GlcNAc with substrate concentrations equivalent to the K_m. All experiments were performed in triplicate, and results were analysed and plotted using GraphPad Prism.

Kinetic parameters of 4MU-GlcNAc turnover by wild-type and mutant hOGA were determined as described above, with enzyme concentrations of 20 nM and a maximum substrate concentration of 2 mM (1 % DMSO in the assay). Competition assays with peptides were performed at 200 μM 4MU-GlcNAc. K_m determination for GlcNAc-modified peptides was performed following an approach using established equations for multisubstrate enzyme kinetics to derive kinetic constants of a substrate whose turnover cannot be detected directly [24]. The fluorigenic reporter substrate and the GlcNAcylated peptide are considered as competitive inhibitors of each other. From the initial hydrolysis rates of 4MU-GlcNAc in the presence and absence of the peptide, the Michaelis constant for the peptide (K_m') can be determined from eqn (1):

$$\frac{v_i}{v_0} = \frac{1 + \frac{K_m}{S}}{1 + \frac{K_m}{S}\left(1 + \frac{S'}{K_{m'}}\right)} \quad (1)$$

where v_i/v_0 is the relative activity in the presence of inhibitor, K_m and S are the Michaelis constant and substrate concentration of the reporter substrate respectively, and S' is the concentration of the peptide.

Determination of hOGA *in vitro* activity on *O*-GlcNAc proteins

The known *O*-GlcNAc targets FoxO1 (forkhead box O1) [25,26] and CREB (cAMP-response-element-binding protein) [27] were provided by the Division of Signal Transduction and Therapy, University of Dundee, Dundee, Scotland, U.K. Both are truncated constructs and N-terminal GST-fusion proteins, GST–CREB-(1–283), and GST–FoxO1-(329–655). A GST-fusion of the TAB1 (transforming growth factor-β-activated protein kinase 1-binding protein 1) N-terminal domain and full-length hOGT were purified as described elsewhere [27a]. A truncated construct lacking the first nine TPRs (tetratricopeptide repeats) (Δ9-hOGT) was expressed in *E. coli*.

For GlcNAcylation reactions, target proteins were incubated with 3.75 mM UDP-GlcNAc and recombinant OGT in 50 mM Tris/HCl (pH 7.5) containing 1 mM DTT (dithiothreitol).

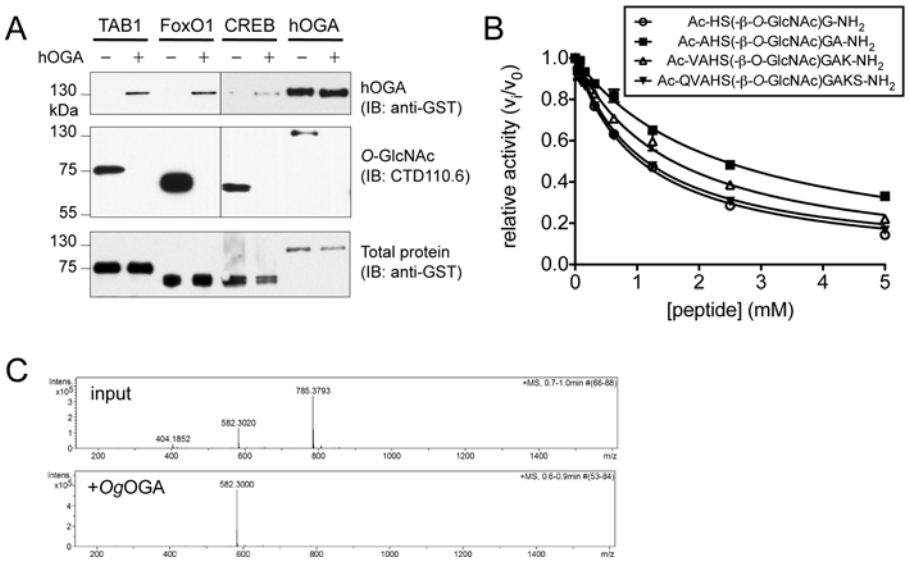

Figure 1 Activity of hOGA and *Og*OGA

(**A**) The *O*-GlcNAc substrate proteins (10 μM) TAB1, FoxO1 and were subjected to *in vitro* GlcNAcylation by recombinant OGT (see Supplementary Figure S1 at http://www.BiochemJ.org/bj/432/bj4320001add.htm). For the activity assay, these substrates, as well as GlcNAc-modified hOGA D175N, were incubated with hOGA (1 μM) at 37 °C, which led to complete deglycosylation within 6 h as detected by anti-*O*-GlcNAc antibody. The top panel shows the enzyme hOGA as detected through its GST tag, the middle panel shows *O*-GlcNAcylation using CTD110.6 anti-*O*-GlcNAc antibody, and the bottom panel shows all the substrate proteins through their GST tags (shorter exposure). (**B**) Competition of *O*-GlcNAc-modified peptides (tri-, penta-, hepta- and nona-peptides based on the *O*-GlcNAc site on hOGA Ser^{405}) with 4MU-GlcNAc turnover by hOGA affords an estimate of the affinity of hOGA for the peptide substrates (Table 1). (**C**) Mass spectrum showing the deglycosylation of Val-Ala-His-Ser-(β-*O*-GlcNAc)-Gly-Ala (molecular mass of 785 Da) to the deglycosylated form (molecular mass of 582 Da), also see Supplementary Figure S2 at http://www.BiochemJ.org/bj/432/bj4320001add.htm.

Enzyme/substrate protein ratios were 1:20, total protein concentration was 0.5 mg/ml. The reaction was allowed to proceed for 3 h at 37 °C (see Supplementary Figure S1 at http://www.BiochemJ.org/bj/432/bj4320001add.htm). Full-length hOGT was used to O-GlcNAcylate CREB and hOGA, whereas the truncated Δ9-hOGT was used for FoxO1 and TAB1. After the glycosylation reaction, target proteins were repurified by GST pull-down, eluted in 25 mM Tris/HCl buffer (pH 7.5) containing 25 mM reduced glutathione and spin-concentrated to approx. 5 mg/ml.

In vitro removal of *O*-GlcNAc was performed in a 5 μl volume in 25 mM Tris/HCl and 25 mM GST (pH 7.5), and contained 10 μM O-GlcNAcylated substrate protein and 1 μM wild-type or mutant hOGA. Reaction mixtures were incubated for 4–6 h at 37 °C, and a 1 μl aliquot was blotted for *O*-GlcNAc with CTD110.6 (Abcam). DM-17 rabbit anti-hOGT antibody was purchased from Sigma. GST and TAB1 antisera were obtained from the Division of Signal Transduction and Therapy [28].

Table 1 Michaelis constants of *O*-GlcNAc-modified peptides for hOGA

K_m values for peptides derived from the *O*-GlcNAc sites of two known substrate proteins were determined in a competitive assay with the fluorigenic substrate 4MU-GlcNAc. Ac, acetyl; amino acids in peptides are indicated using the single-letter code.

Substrate	K_m
hOGA-Ser^{405}-derived peptide series	
Ac-HS(-β-*O*-GlcNAc)G-NH$_2$	525 $\pm$ 13 μM
Ac-VAHS(-β-*O*-GlcNAc)GA-NH$_2$	1204 $\pm$ 43 μM
Ac-VAHS(-β-*O*-GlcNAc)GAK-NH$_2$	789 $\pm$ 36 μM
Ac-QVAHS(-β-*O*-GlcNAc)GAKAS-NH$_2$	528 $\pm$ 15 μM
TAB1-Ser^{395}-derived peptide series:	
Ac-YS(-β-*O*-GlcNAc)S-NH$_2$	2.2 $\pm$ 0.1 mM
Ac-PYS(-β-*O*-GlcNAc)SA-NH$_2$	6.3 $\pm$ 0.5 mM
Ac-VPYS(-β-*O*-GlcNAc)SAQ-NH$_2$	4.1 $\pm$ 0.3 mM
Ac-SVPYS(-β-*O*-GlcNAc)SAQS-NH$_2$	4.8 $\pm$ 0.3 mM

RESULTS AND DISCUSSION

Recombinant hOGA is active on *O*-GlcNAc-modified proteins and peptides *in vitro*

Although hOGA has been shown to be active on a GlcNAc-modified peptide substrate [29], most studies have been restricted to measuring the enzyme activity *in vitro* using pseudosubstrates, or monitoring OGA activity *in vivo* by Western blot analysis of cellular *O*-GlcNAc levels after overexpression, knock down or inhibition of OGA. We aimed to demonstrate *in vitro* activity of hOGA on recombinant glycoproteins. Substrates were prepared by *in vitro* glycosylation of the previously characterized *O*-GlcNAc proteins TAB1 and the transcription factors FoxO1 and CREB (see Supplementary Figure S1). OGA assays were then performed with these substrates as well as a glycosylated hOGA-inactive mutant, and reduction of *O*-GlcNAc was detected by anti-*O*-GlcNAc Western blot analysis, showing that hOGA is capable of removing *O*-GlcNAc from these glycoproteins (Figure 1A). We next sought to identify peptide substrates that would allow the determination of the Michaelis constant (K_m), and also allow for varying sequence and length to probe substrate specificity. Peptides covered the *O*-GlcNAc sites on TAB1 (Ser^{395}) (S. Pathak and D.M.F. van Aalten, unpublished work) and hOGA (Ser^{405}) [4] (Table 1). hOGA activity on these peptides was confirmed by LC–MS. The affinity of hOGA for these peptides was then quantified by their ability to compete with the reporter substrate 4MU-GlcNAc (Figure 1B and Table 1). The resulting K_m values for the hOGA-Ser^{405}- and TAB1-Ser^{395}-derived peptides are in the high-micromolar and low-millimolar range respectively, somewhat decreasing in affinity with increasing peptide length.

Discovery of a bacterial OGA with improved similarity to the human enzyme

hOGA is known to possess an N-terminal catalytic domain (residues 60–366) and the C-terminal HAT domain (residues 707–916) (Figure 2A). However, sequence analysis of putative metazoan OGAs shows that they also possess a low-complexity N-terminus (residues 1–59), and a 'middle' domain (residues 367–707), also containing a low-complexity region (residues 396–553). Only the catalytic domain is represented in the available (bacterial) OGA structures, yet any of the additional hOGA domains could contribute to substrate recognition/specificity.

In the absence of a hOGA structure to guide our substrate-specificity studies, we aimed to identify new bacterial OGAs with improved sequence identity and extending to beyond just the catalytic domain. A BLAST search identified a predicted protein from *O. granulosus* (*Og*OGA) that not only shows improved sequence identity (37 %) with the hOGA catalytic domain without any alignment gaps (Figure 2B), but also, intriguingly, possesses a 200-residue C-terminal domain that shares approx. 20 % identity with the hOGA middle domain (Figures 2A and 2B). An *Og*OGA structure would reveal the position of this domain with respect to the active site and serve as a guide to identifying residues or domains involved in glycoprotein substrate recognition.

To investigate the possible OGA activity of *Og*OGA, the gene was cloned from genomic DNA and expressed in *E. coli*. The enzyme shows OGA activity with *p*NP-GlcNAc ($K_m = 8.5 \pm 0.6$ mM, $k_{cat} = 24.0 \pm 1.0\ s^{-1}$) and 4MU-GlcNAc ($K_m = 2.9 \pm 0.2$ mM, $k_{cat} = 8.2 \pm 0.3\ s^{-1}$), with a pH optimum of 7.4–7.6 (see Supplementary Figure S2B at http://www.BiochemJ.org/bj/432/bj4320001add.htm). *Og*OGA is sensitive to the commonly used OGA inhibitors PUGNAc ($IC_{50} = 11\ \mu M$) and GlcNAcstatin G [29a] ($IC_{50} = 780$ nM). *Og*OGA was also able to quantitatively deglycosylate a glycopentapeptide, Val-Ala-His-Ser-(*β*-*O*-GlcNAc)-Gly-Ala, derived from the reported hOGA *O*-GlcNAc site [4], as detected by LC–MS (Figure 1C and Supplementary Figure S2).

*Og*OGA possesses a conserved peptide-binding groove and 'stalk'

We next explored the structural similarities of *Og*OGA with hOGA to allow exploitation of *Og*OGA as a structural guide for hOGA mutagenesis studies. *Og*OGA was crystallized and its structure solved, including a complex with PUGNAc (see Supplementary Table S1). The *Og*OGA structure reveals an N-terminal TIM barrel domain and a C-terminal *α*-helical domain (Figures 2A and 2C). The *Og*OGA catalytic domain is similar to that of *Cp*NagJ (*Clostridium perfringens* NagJ) [RMSD (root mean square deviation) = 1.5 Å (1 Å = 0.1 nm) for 254 C_αs; Figure 2C]. The *α*4–*β*5, *β*3–*α*3, *β*4–*α*4 and *β*5–*α*5 loops are of different lengths and/or adopt different conformations. Strikingly, sequence alignments show that these loops match hOGA, but not *Cp*NagJ, in length and conservation (Figure 2B). The C-terminal domain, hereafter referred to as the stalk domain, possesses good sequence similarity with hOGA (Figure 2B) and contains a core four-helical bundle (Figure 2C). Unexpectedly, this domain is structurally similar to the third domain of *Cp*NagJ (RMSD = 1.9 Å for 110 C_αs, yet <10 % sequence identity; Figure 2C), although the loops connecting the helices differ in length and conformation. In particular, the *α*10–*α*11 loop is nine residues longer in *Cp*NagJ compared with *Og*OGA (and, by alignment, hOGA; Figure 2B), reaching into the *Cp*NagJ active site, which is impossible in *Og*OGA (and thus unlikely in hOGA). Similarly, the *Og*OGA structure suggests structural homology between the stalk domain of *Bacteroides thetaiotaomicron* OGA [20] and hOGA (see Supplementary Figure S3 at http://www.BiochemJ.org/bj/432/bj4320001add.htm).

The *Og*OGA active site is formed by the TIM barrel *β*–*α* loops (Figure 2C), with the catalytic loop residues Asp^{115}/Asp^{116} on *β*4–*α*4 conserved with hOGA, compatible with the observed OGA activity (Figure 2C). This loop does not fully engage with the PUGNAc inhibitor, presumably related to the disordered phenylcarbamate group. Ligand-induced conformational changes in the *β*4–*α*4 loop have also been observed for *Cp*NagJ [30]. Upon PUGNAc binding to *Og*OGA, the *β*7–*η*1 loop shifts by up to 4.7 Å exposing the GlcNAc-binding site, associated with small displacements of the neighbouring *β*8–*α*8 and *β*1–*α*1 loops. Residues hydrogen-bonding with the sugar hydroxy groups (Gly^8, Tyr^{160}, Asp^{226} and Asn^{254}) or the *N*-acetyl moiety (Asn^{221}) are conserved in hOGA (Figures 2B and 2E). Similarly, the residues that show van der Waals/stacking interactions with the sugar and the *N*-acetyl group (Tyr^{10}, Tyr^{160} and Trp^{219}) are conserved in hOGA. Strikingly, the surface around the active site, together with the stalk domain, forms a groove lined by side chains that are highly conserved/identical between *Og*OGA and hOGA and also among metazoan OGAs (Figure 2D).

Residues lining the conserved binding groove are required for hOGA–substrate interactions

We next tested the hypothesis that the conserved groove represents the binding site for *O*-GlcNAc glycoprotein substrates by expressing GST-fusions of 20 individual hOGA point mutants of residues lining the groove (Figures 2B and 2D). To control for effects on protein folding and/or catalysis rather than protein substrate binding, we measured 4MU-GlcNAc hydrolysis by these mutants (Figure 3A). Mutations affecting 4MU-GlcNAc hydrolysis are mostly located close to the sugar/catalytic machinery (Tyr^{69}) or occupy positions shown to undergo a conformational change upon inhibitor binding (Tyr^{286} and Asp^{287} in the *β*7–*η*1 loop) (Figure 2). Strikingly, de-*O*-GlcNAcylation assays with the TAB1, FoxO1 and CREB glycoprotein substrates identified mutants of residues some distance from the active site (I176L, S190A, S190Q, T217A, E218S, F223S and K289A) with reduced activity against one or more of the protein substrates, yet unimpaired 4MU-GlcNAc hydrolysis (Figure 3).

Conclusions

Since its discovery more than three decades ago, it has become clear that *O*-GlcNAc is an important post-translational modification found on hundreds of intracellular proteins and plays a role in signalling pathways. However, it is not clear what determines specificity in the targeting and cycling of *O*-GlcNAc on these proteins. We have investigated whether OGA specifically recognizes the peptide component of its glycoprotein substrates. Our findings show that, by similarity to the structurally related bacterial *Og*OGA, hOGA possesses a putative substrate-binding groove that is highly conserved within metazoan OGAs, in addition to a stalk domain in close proximity to the active site. Indeed, mutations targeting this groove identified residues that are not required for hydrolysis of the 4MU-GlcNAc pseudosubstrate, but affected hydrolysis of glycoprotein substrates. Although a limited set of glycoprotein substrates (TAB1, FoxO1 and CREB) and peptides (derived from TAB1 and hOGA itself) was tested, these mutational studies hinted at differential substrate recognition. The differences between the residues required for the turnover of these substrates may reflect differences in the molecular environment of the *O*-GlcNAc site(s)

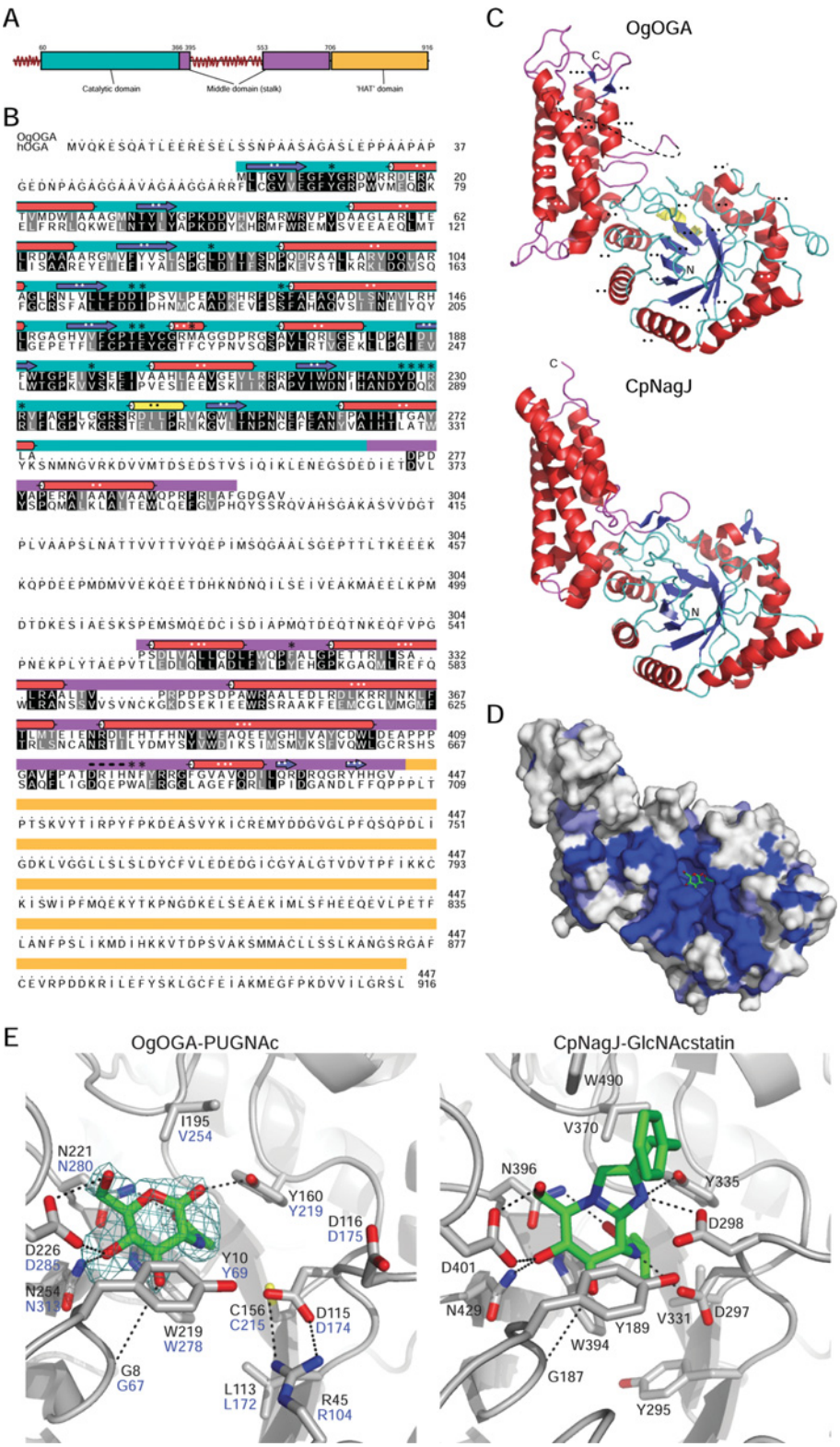

Figure 2 Structure of *Og*OGA and similarity to other OGAs

(**A**) Domain architecture of human OGA. Predicted ordered domains are shown as coloured boxes (cyan, catalytic domain; purple, stalk domain; orange, HAT domain), predicted disordered regions are indicated by red wavy lines. (**B**) Sequence alignment of *Og*OGA and hOGA. Secondary-structure elements from the *Og*OGA structure are shown (red, α-helices; blue, β-strands; yellow, 3_{10}-helices), broken lines denote residues disordered in apo-*Og*OGA. The proposed domains are indicated by boxes coloured as in (**A**). (**C**) Cartoon view of the *Og*OGA apo structure and domains 2 and 3 of *Cp*NagJ. (**D**) Surface view of *Og*OGA shaded by sequence conservation among metazoan OGAs (light blue) and metazoan OGAs+*Og*OGA (dark blue). The PUGNAc (green) complex is shown with additional ordered loops taken from the apo-structure. (**E**) Active sites of the *Og*OGA–PUGNAc and *Cp*NagJ–GlcNAcstatin complexes. The ligands are represented as sticks with green carbons, an unbiased $F_o - F_c;\varphi_c$ map (2.3 σ) is shown for PUGNAc. Hydrogen bonds are drawn as black broken lines. Hydrogen bonds to/from protein backbone amides are shown as originating from the Cα. Active-site residues in *Og*OGA are labelled in black, alongside the equivalent hOGA residues in blue.

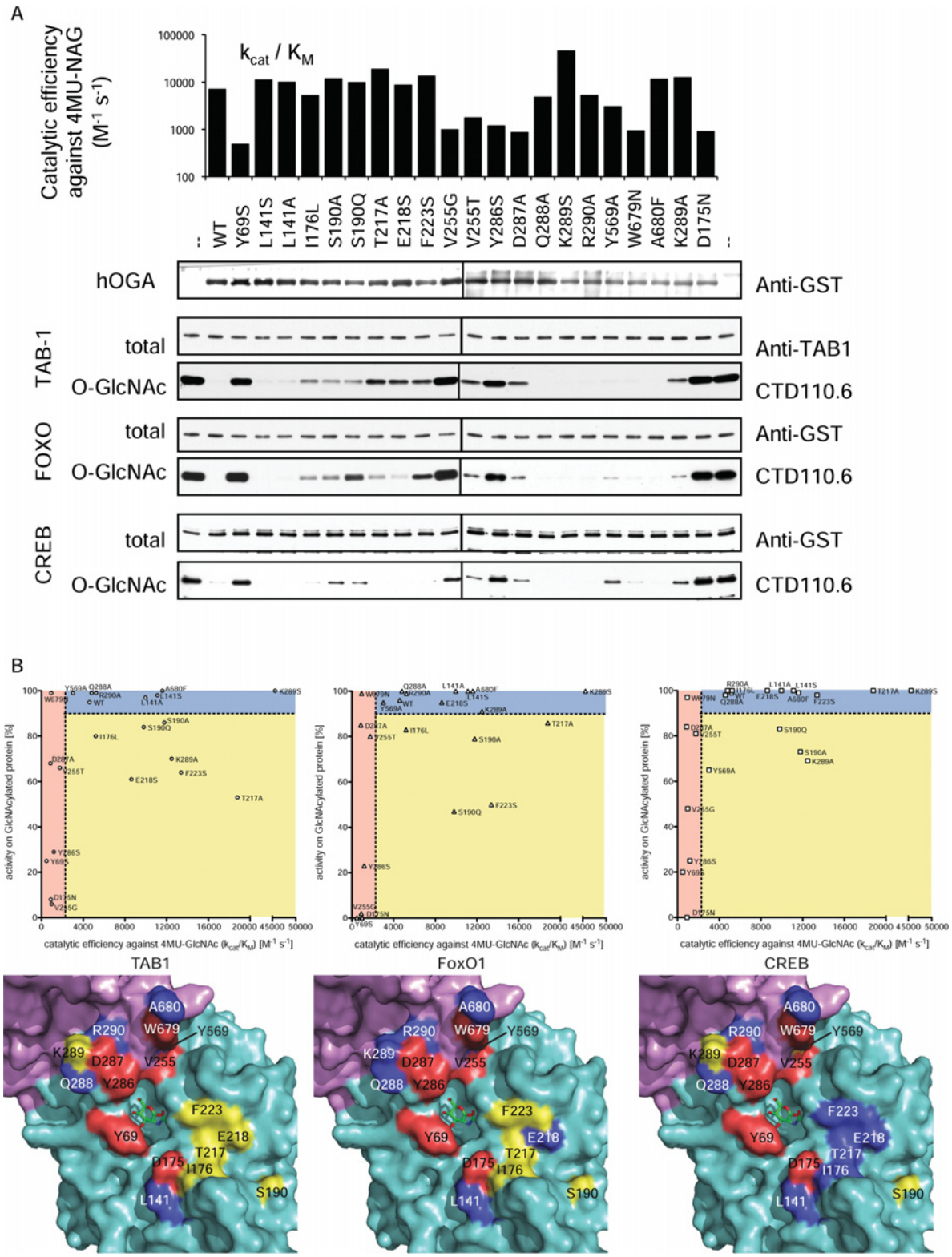

Figure 3 Activity of hOGA groove mutants

Twenty individual mutants of hOGA were generated based on the structural information obtained from *Og*OGA. (**A**) To evaluate the structural integrity of the hOGA mutants, the catalytic efficiency of 4MU-GlcNAc turnover was determined and compared with the wild-type ($k_{cat}/K_m = 7000\ M^{-1}{\cdot}s^{-1}$) and the catalytically impaired D175N mutant (negative control). hOGA mutants were also incubated with *O*-GlcNAc substrate proteins TAB1, FoxO1 and CREB, and deglycosylation was observed by Western blotting. (**B**) Visualization of the correlation between the mutants' ability to hydrolyse the pseudosubstrate 4MU-GlcNAc and glycoprotein substrates. Catalytically impaired mutants (k_{cat}/K_m for 4MU-GlcNAc $\leqslant$50 % of wild-type) locate in the areas shaded in red. Mutants unaffected in both 4MU-GlcNAc turnover and protein deglycosylation (cut-off 10 % reduction of wild-type activity) are shown in the blue areas. Mutations reducing OGA activity on protein substrates only are depicted in the yellow areas. Location of the mutations is shown on the surface around the active site of *Og*OGA (coloured by domain, see Figure 2A). The mutated amino acids are shaded according to correlation plots, thus amino acids affecting protein substrate recognition are highlighted in yellow.

on these proteins/peptides, and could tune *O*-GlcNAc cycling rates. Together, the results of the present study constitute the first evidence for specific interactions between hOGA and its substrate proteins beyond the sugar moiety, and may form the basis of an improved understanding of the enzyme's substrate specificity and the functional implications thereof.

AUTHOR CONTRIBUTION

The scientific hypothesis was devised by Marianne Schimpl, Alexander Schüttelkopf and Daan van Aalten. The experimental design was by Marianne Schimpl, Alexander Schüttelkopf, Vladimir Borodkin and Daan van Aalten. Experimentation was by Marianne Schimpl, Alexander Schüttelkopf and Vladimir Borodkin. Interpretation of results was by Marianne Schimpl, Alexander Schüttelkopf and Daan van Aalten who also wrote the paper.

ACKNOWLEDGEMENTS

We thank Sharon Shepherd for assistance with protein production, Lindsey Gray for assistance with peptide synthesis, Adel Ibrahim for help with molecular cloning, and the European Synchrotron Radiation Facility, Grenoble, for time on BM14, ID23-2 and ID14-1.

FUNDING

This work was supported by a Wellcome Trust Senior Research Fellowship to D.M.F.v.A. and a Wellcome Trust Ph.D. Studentship to M.S.

REFERENCES

1 Love, D. C. and Hanover, J. A. (2005) The hexosamine signaling pathway: deciphering the "O-GlcNAc code". Sci. STKE **2005**, re13
2 Hart, G. W., Housley, M. P. and Slawson, C. (2007) Cycling of O-linked β-*N*-acetylglucosamine on nucleocytoplasmic proteins. Nature **446**, 1017–1022
3 Torres, C. R. and Hart, G. W. (1984) 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
4 Khidekel, N., Ficarro, S. B., Clark, P. M., Bryan, M. C., Swaney, D. L., Rexach, J. E., Sun, Y. E., Coon, J. J., Peters, E. C. and Hsieh-Wilson, L. C. (2007) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. Nat. Chem. Biol. **3**, 339–348
5 Vosseller, K., Trinidad, J. C., Chalkley, R. J., Specht, C. G., Thalhammer, A., Lynn, A. J., Snedecor, J. O., Guan, S., Medzihradszky, K. F., Maltby, D. A. et al. (2006) O-linked *N*-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. Mol. Cell. Proteomics **5**, 923–934
6 Wang, Z., Udeshi, N. D., O'Malley, M., Shabanowitz, J., Hunt, D. F. and Hart, G. W. (2009) Enrichment and site mapping of O-linked *N*-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. Mol. Cell. Proteomics **9**, 153–160
7 Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., Shabanowitz, J., Hunt, D. F. and Hart, G. W. (2010) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. Sci. Signaling **3**, ra2
8 Zachara, N. E. and Hart, G. W. (2004) O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. Biochim. Biophys. Acta **1673**, 13–28
9 Slawson, C., Zachara, N. E., Vosseller, K., Cheung, W. D., Lane, M. D. and Hart, G. W. (2005) Perturbations in O-linked β-*N*-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. J. Biol. Chem. **280**, 32944–32956
10 Slawson, C., Shafii, S., Amburgey, J. and Potter, R. (2002) Characterization of the O-GlcNAc protein modification in *Xenopus laevis* oocyte during oogenesis and progesterone-stimulated maturation. Biochim. Biophys. Acta **1573**, 121–129
11 Comtesse, N., Maldener, E. and Meese, E. (2001) Identification of a nuclear variant of MGEA5, a cytoplasmic hyaluronidase and a β-*N*-acetylglucosaminidase. Biochem. Biophys. Res. Commun. **283**, 634–640
12 Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J. and Love, D. C. (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. Arch. Biochem. Biophys. **409**, 287–297
13 Cheung, W. D., Sakabe, K., Housley, M. P., Dias, W. B. and Hart, G. W. (2008) O-linked β-*N*-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. J. Biol. Chem. **283**, 33935–33941
14 Heckel, D., Comtesse, N., Brass, N., Blin, N., Zang, K. D. and Meese, E. (1998) Novel immunogenic antigen homologous to hyaluronidase in meningioma. Hum. Mol. Genet. **7**, 1859–1872
15 Copeland, R. J., Bullen, J. W. and Hart, G. W. (2008) Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity. Am. J. Physiol. Endocrinol. Metab. **295**, E17–E28
16 Whisenhunt, T. R., Yang, X., Bowe, D. B., Paterson, A. J., Van Tine, B. A. and Kudlow, J. E. (2006) Disrupting the enzyme complex regulating O-GlcNAcylation blocks signaling and development. Glycobiology **16**, 551–563
17 Haltiwanger, R. S., Grove, K. and Philipsberg, G. A. (1998) Modulation of O-linked *N*-acetylglucosamine levels on nuclear and cytoplasmic proteins *in vivo* using the peptide O-GlcNAc-β-*N*-acetylglucosaminidase inhibitor O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate. J. Biol. Chem. **273**, 3611–3617
18 Macauley, M. S., Stubbs, K. A. and Vocadlo, D. J. (2005) O-GlcNAcase catalyzes cleavage of thioglycosides without general acid catalysis. J. Am. Chem. Soc. **127**, 17202–17203
19 Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M. and van Aalten, D. M. (2006) Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. EMBO J. **25**, 1569–1578
20 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) Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. Nat. Struct. Mol. Biol. **13**, 365–371
21 Macauley, M. S. and Vocadlo, D. J. (2010) Increasing O-GlcNAc levels: an overview of small-molecule inhibitors of O-GlcNAcase. Biochim. Biophys. Acta **1800**, 107–121
22 Salvador, L. A., Elofsson, M. and Kihlberg, J. (1995) Preparation of building blocks for glycopeptide synthesis by glycosylation of Fmoc amino acids having unprotected carboxyl groups. Tetrahedron **51**, 5643–5656
23 Sanchez, I., Hughes, R. T., Mayer, B. J., Yee, K., Woodgett, J. R., Avruch, J., Kyriakis, J. M. and Zon, L. I. (1994) Role of SAPK/ERK kinase-1 in the stress-activated pathway regulating transcription factor c-Jun. Nature **372**, 794–798
24 Xie, D., Suvorov, L., Erickson, J. W. and Gulnik, A. S. (1999) Real-time measurements of dark substrate catalysis. Protein Sci. **8**, 2460–2464
25 Housley, M. P., Rodgers, J. T., Udeshi, N. D., Kelly, T. J., Shabanowitz, J., Hunt, D. F., Puigserver, P. and Hart, G. W. (2008) O-GlcNAc regulates FoxO activation in response to glucose. J. Biol. Chem. **283**, 16283–16292
26 Kuo, M., Zilberfarb, V., Gangneux, N., Christeff, N. and Issad, T. (2008) O-glycosylation of FoxO1 increases its transcriptional activity towards the glucose 6-phosphatase gene. FEBS Lett. **582**, 829–834
27 Lamarre-Vincent, N. and Hsieh-Wilson, L. C. (2003) Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. J. Am. Chem. Soc. **125**, 6612–6613
27a Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schüttelkopf, A. W., Borodkin, V., Shepherd, S. M., Ibrahim, A. F. and van Aalten, D. M. (2008) Structural insights into mechanism and specificity of O-GlcNAc transferase. EMBO J. **27**, 2780–2788
28 Cheung, P. C., Campbell, D. G., Nebreda, A. R. and Cohen, P. (2003) Feedback control of the protein kinase TAK1 by SAPK2a/p38α. EMBO J. **22**, 5793–5805
29 Gao, Y., Wells, L., Comer, F. I., Parker, G. J. and Hart, G. W. (2001) Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic β-*N*-acetylglucosaminidase from human brain. J. Biol. Chem. **276**, 9838–9845
29a Dorfmueller, H. C., Borodkin, V. S., Schimpl, M., Zheng, X., Kime, R., Read, K. D. and van Aalten, D. M. F. (2010) Cell-penetrant, nanomolar O-GlcNAcase inhibitors selective against lysosomal hexosaminidases. Chem. Biol., in the press
30 Pathak, S., Dorfmueller, H. C., Borodkin, V. S. and van Aalten, D. M. F. (2008) Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death. Chem. Biol. **15**, 799–807

Received 23 August 2010/9 September 2010; accepted 23 September 2010
Published as BJ Immediate Publication 23 September 2010, doi:10.1042/BJ20101338

Biochem. J. (2010) **432**, 1–7 (Printed in Great Britain) doi:10.1042/BJ20101338

online data

SUPPLEMENTARY ONLINE DATA
Human OGA binds substrates in a conserved peptide recognition groove

Marianne SCHIMPL[1], Alexander W. SCHÜTTELKOPF[1], Vladimir S. BORODKIN and Daan M. F. VAN AALTEN[2]
Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, U.K.

SUPPLEMENTARY MATERIALS AND METHODS

*Og*OGA cloning and expression

O. granulosus HTCC 2516 was obtained from the A.T.C.C. (Manassas, VA, U.S.A.) (BAA-861) and grown at 30 °C overnight in Difco marine agar/broth 2216. *O. granulosus* genomic DNA was prepared following the method of Wilson [1]. The full-length predicted *Og*OGA gene (protein accession number ZP_01156836.1) was then amplified by PCR using the forward primer 5′-TTATGGATCCATGCTGACCGGGGTGATCGAGG-3′ and the reverse primer 5′-TTTACTCGAGTCAGACGC-CATGGTGATACCGGC-3′ (the BamHI and XhoI restriction sites are underlined), cloned into pGEX 6P-1 (GE Healthcare) and verified by sequencing.

The resulting construct was transformed into *E. coli* BL21(DE3) pLysS cells, which were then grown overnight at 37 °C in 80 ml of Luria–Bertani medium supplemented with 100 μg/ml carbenicillin. This culture was then used to inoculate 8 litres of autoinduction medium [2] with carbenicillin, in which the cells were grown for 48 h at 25 °C before harvesting by centrifugation at 3500 ***g*** for 30 min. Cells were washed, then resuspended in 50 ml of 25 mM Tris/HCl (pH 7.5) and 250 mM NaCl and lysed by passage through a continuous-flow cell disrupter at 207 MPa (Constant Systems). After centrifugation and filtration, the soluble fraction of the lysate was incubated for 2 h at 4 °C with 4 ml of glutathione–Sepharose beads (GE Healthcare), then washed with lysis buffer. After addition of 400 μg of GST-tagged PreScission protease, the beads were incubated overnight at 4 °C. Beads were removed, and the filtrate was dialysed into 25 mM Tris/HCl (pH 7.5) before loading on to a 5 ml HiTrap Q FF column (GE Healthcare) equilibrated in the same buffer. After washing protein was eluted using a 0–500 mM NaCl gradient. Fractions containing *Og*OGA were pooled and concentrated to <5 ml for gel filtration using a XK26/60 Superdex 75 gel-filtration column equilibrated in 25 mM Tris/HCl (pH 7.5) and 150 mM NaCl. The resulting fractions containing pure *Og*OGA as verified by SDS/PAGE were pooled and concentrated to 1.9 mg/ml.

*Og*OGA structure determination

*Og*OGA was crystallized at 20 °C using sitting-drop vapour diffusion with drops of 1 μl of protein solution mixed with 1 μl of mother liquor over a 60 μl reservoir of 18 % poly(ethylene glycol) 4000, 100–250 mM $MgCl_2$ and 100 mM Tris/HCl (pH 8.5). Crystals appeared overnight as hexagonal prisms. Crystals were cryoprotected using a mixture of 80 % mother liquor and 20 % ethylene glycol. For the PUGNAc complex structure, solid PUGNAc was added to a drop containing crystals for 30 min before cryoprotection. For phasing, 0.17 μl of a solution of 200 mM lead acetate in mother liquor was added to a drop containing crystals for 1 h before cryoprotection. All data were collected at the European Synchrotron Radiation Facility (Grenoble, France) at a temperature of 100 K, and processed using HKL software [3] (Supplementary Table S1). *Og*OGA crystals are of space group C2 with one protein molecule per asymmetric unit. Data for the $PbAc_2$ derivative were collected at beamline BM14 with $\lambda = 0.946$ Å (peak) and $\lambda = 0.950$ Å (inflection point) and processed/scaled to 1.9 Å and 2.0 Å respectively. Phases were calculated using SHELXC/D/E [4,5] through the HKL2MAP interface [6], WarpNtrace [7] built 264 of a possible 447 residues. After manual completion using Coot [8], REFMAC5 [9] and other CCP4 software [10], the resulting model was used for molecular replacement into the native 2.0 Å dataset. Model building and refinement proceeded as for the derivative structure and resulted in a final model with an R_{work} of 0.186 and an R_{free} of 0.222 with good geometry throughout (Supplementary Table S1). The PUGNAc complex was solved by molecular replacement with the derivative structure and refined as above to an R_{work} of 0.242 and an R_{free} of 0.279 with good geometry. Both reported structures exhibit two poorly ordered regions, one comprising several residues directly preceding helix α4 (modelled in the apo structure), the other covering part of the α13–α14 loop (lacking in both models). Additionally, the 12 C-terminal residues are disordered in the PUGNAc complex. Interestingly, the residues lacking from the α13–α14 loop would not be able to span the distance (31.4 Å, in principle covered by five amino acids, i.e. 6.28 Å per amino acid) between the built amino acids on either side of the gap; it is possible that the actual connection is between two monomers and the protein thus forms a (sub)domain-swapped dimer (distance = 13.2 Å, giving an average of 2.64 Å span per residue). This has not been modelled as such, as *Og*OGA is monomeric in solution (by gel-filtration; results not shown). Alternatively, the distance between the disordered ends of the loop could be explained by the effects of an unknown proteolytic event. Finally, the termini of helices α13 and α14 are 29 Å apart, which can easily be covered by the α13–α14 loop (21 residues, ~1.3 Å per residue) after a degree of conformational change.

Ligand co-ordinates and topologies were generated with PRODRG [11]; during model building, it became apparent that the ligand electron density covered only the (pseudo)sugar ring and the rest of the molecule was either disordered or, more likely, had been hydrolysed. Consistent with the latter hypothesis, the ligand was modelled as a tetrahydropyran-2-one derivative. Figure 1(B) of the main text was created using Aline [12], Figures 2(C)–2(E) of the main text and the surface views in Figure 3(B) of the main text were rendered using PyMOL (DeLano Scientific http://www.pymol.org).

[1] These authors contributed equally to this work
[2] To whom correspondence should be addressed (email dmfvanaalten@dundee.ac.uk).
Structures have been deposited in the PDB under accession codes 2XSA and 2XSB.

Table S1 X-ray crystallographic data collection and structure refinement statistics

Values for the highest resolution shell are given in parentheses.

Statistic	Apo-*Og*OGA	*Og*OGA-PUGNAc
Beamline/wavelength (Å)	ID23-2/0.873	ID14-1/0.934
Resolution (Å)	25.00–2.00 (2.03–2.00)	20.00–2.10 (2.17–2.10)
Unit cell	$a = 84.4$ Å, $b = 95.3$ Å, $c = 60.5$ Å, $\beta = 104.0°$	$a = 84.6$ Å, $b = 96.3$ Å, $c = 60.0$ Å, $\beta = 104.0°$
Number of unique reflections	31362	26338
Redundancy	4.4 (4.1)	3.2 (2.9)
Completeness (%)	99.3 (94.3)	98.9 (95.7)
$I/\sigma(I)$	33.8 (5.6)	11.3 (2.3)
R_{merge}	0.072 (0.291)	0.095 (0.574)
Wilson B (Å^2)	36.9	41.0
R_{work}, R_{free}	0.186, 0.222	0.242, 0.279
Average B-factor		
Protein (Å^2)	41.4	37.0
Solvent (Å^2)	43.6	32.1
Ligand (Å^2)	n/a	39.3
RMSD from ideal geometry: bond lengths (Å)/bond angles (°)	0.014/1.30	0.012/1.30
Ramachandran statistics (most favoured/additionally allowed/generously allowed/disallowed)	91.8 %/7.9 %/0.3 %/0%	90.8 %/9.2 %/0 %/0%

Table S2 Oligonucleotides used for introducing point mutations in the coding sequence of hOGA

Oligonucleotide	Sequence (5′→3′)
Y69S_fwd	GTGGTGGAAGGATTTTCTGGAAGACCTTGGG
Y69S_rev	CCCAAGGTCTTCCAGAAAATCCTTCCACCAC
L141S_fwd	GATCTCACCTGGATCGGATATCACTTTTTC
L141S_rev	GAAAAAGTGATATCCGATCCAGGTGAGATC
L141A_fwd	GATCTCACCTGGAGCGGATATCACTTTTTC
L141A_rev	GAAAAAGTGATATCCGCTCCAGGTGAGATC
I176L_fwd	CTTTGCTTTTTGATGATCTAGACCATAATATGTG
I176L_rev	CACATATTATGGTCTAGATCATCAAAAAGCAAAG
S190A_fwd	CAAAGAGGTATTCAGTGCTTTTGCTCATGCCCAAG
S190A_rev	CTTGGGCATGAGCAAAAGCACTGAATACCTCTTTG
S190Q_fwd	CAAAGAGGTATTCAGTCAGTTTGCTCATGCCCAAG
S190Q_rev	CTTGGGCATGAGCAAACTGACTGAATACCTCTTTG
T217A_fwd	CTTTCCTCTTCTGTCCCGCAGAATACTGTGGCAC
T217A_rev	GTGCCACAGTATTCTGCGGGACAGAAGAGGAAAG
E218S_fwd	CTCTTCTGTCCCACATCATACTGTGGCACTTTC
E218S_rev	GAAAGTGCCACAGTATGATGTGGGACAGAAGAG
F223S_fwd	GAATACTGTGGCACTTCCTGTTATCCAAATGTG
F223S_rev	CACATTTGGATAACAGGAAGTGCCACAGTATTC
V255G_fwd	CAGGTCCCAAAGTTGGTTCTAAAGAAATTCC
V255G_rev	GGAATTTCTTTAGAACCAACTTTGGGACCTG
V255T_fwd	CAGGTCCCAAAGTTACTTCTAAAGAAATTCC
V255T_rev	GGAATTTCTTTAGAAGTAACTTTGGGACCTG
Y286S_fwd	CATTCATGCTAATGATTCTGATCAGAAGAGACTG
Y286S_rev	CAGTCTCTTCTGATCAGAATCATTAGCATGAATG
D287A_fwd	CATGCTAATGATTATGCTCAGAAGAGACTGTTTC
D287A_rev	GAAACAGTCTCTTCTGAGCATAATCATTAGCATG
Q288A_fwd	GCTAATGATTATGATGCGAAGAGACTGTTTCTGGG
Q288A_rev	CCCAGAAACAGTCTCTTCGCATCATAATCATTAGC
K289S_fwd	CTAATGATTATGATCAGTCGAGACTGTTTCTGGGCCCG
K289S_rev	CGGGCCCAGAAACAGTCTCGACTGATCATAATCATTAG
R290A_fwd	GATTATGATCAGAAGGCACTGTTTCTGGGCCCGTAC
R290A_rev	GTACGGGCCCAGAAACAGTGCCTTCTGATCATAATC
Y569A_fwd	CTATTCTACCTTCCTGCCGAGCATGGACCCAAAG
Y569A_rev	CTTTGGGTCCATGCTCGGCAGGAAGGTAGAATAG
W679N_fwd	GGAGACCAAGAACCCAATGCCTTTAGAGGTGGTC
W679N_rev	GACCACCTCTAAAGGCATTGGGTTCTTGGTCTCC
A680F_fwd	GACCAAGAACCCTGGTTCTTTAGAGGTGGTCTAG
A680F_rev	CTAGACCACCTCTAAAGAACCAGGGTTCTTGGTC
K289A_fwd	CTAATGATTATGATCAGGCGAGACTGTTTCTGGGCCCG
K289A_rev	CGGGCCCAGAAACAGTCTCGCCTGATCATAATCATTAG

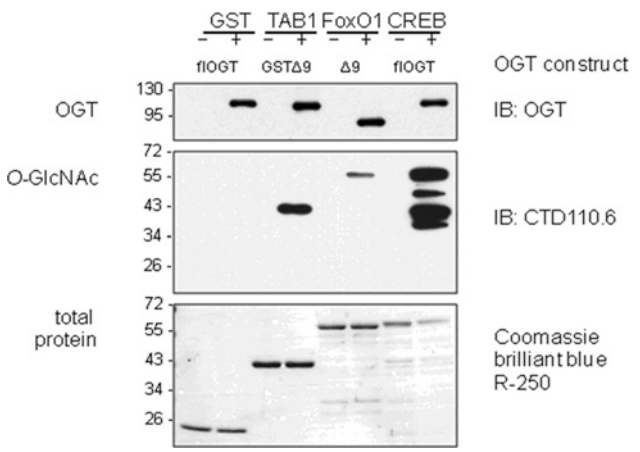

Figure S1 ***In vitro*** **O-GlcNAcylation of proteins with OGT generates OGA substrates**

GlcNAcylation reactions on GST (control), TAB1, GST–FoxO1 or GST–CREB. Full-length OGT was used for GST and CREB, a truncated (Δ9) construct for FoxO1 and a GST fusion thereof (GST–Δ9) for TAB1. Reactions contained 3.75 mM UDP-GlcNAc and recombinant OGT in 50 mM Tris/HCl (pH 7.5) plus 1 mM DTT (dithiothreitol). Enzyme/substrate protein ratios were 1:20, total protein concentration was 0.5 mg/ml. The reaction was allowed to proceed for 3 h at 37 °C. The top panel shows the different constructs of OGT, detected with DM-17 rabbit hOGT antiserum, the middle panel shows O-GlcNAcylation, as detected with the CTD110.6 antiserum, and the bottom panel shows total protein staining. This was necessary as the TAB1 used in this experiment did not carry a GST tag, but also serves to illustrate the presence of degradation products in the FoxO1 and CREB samples, some of which appear to bear the *O*-GlcNAc modification, as seen in the middle panel.

Glycopeptide synthesis

The 2,3,4,6-tetracetyl-*O*-GlcNAc-Fmoc-Ser derivative was synthesised according to the published procedure [13]. Solid-phase peptide synthesis was performed on automated microwave-assisted peptide synthesizer CEM Liberty on low-load Rink amide MBHA (methylbenzhydrylamine) resin 100–200 mesh (Novabiochem) using standard Fmoc chemistry protocols on 0.05 mmol scale. Fmoc deprotection was executed with 20 % piperidine-DMF (dimethylformamide) at 70 °C for 3 min.

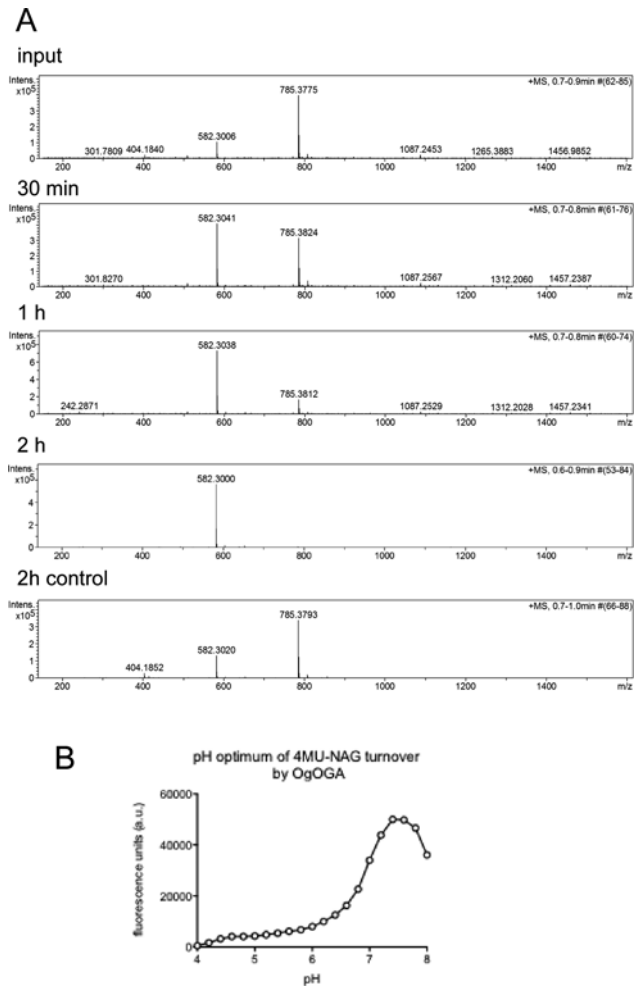

Figure S2 *Og*OGA pH optimum and time-dependent removal of *O*-GlcNAc from peptide by *Og*OGA

(**A**) Time-dependent removal of *O*-GlcNAc from peptide by *Og*OGA. A β-*O*-GlcNAc-modified peptide substrate was completely deglycosylated by *Og*OGA in the course of 2 h. Reactions were performed and analysed after incubation at 37 °C for the indicated lengths of time. The synthetic pentapeptide Ala-His-Ser-(β-*O*-GlcNAc)-Gly-Ala (molecular mass 785 Da) was progressively transformed into the deglycosylated form with a molecular mass of 582 Da. The presence of a peak at 582 Da in the control sample was most likely to be an artefact of the ionization procedure, as the glycopeptide sample appeared to be pure and no traces of contaminants were detected by NMR analysis. (**b**) pH optimum of *Og*OGA. Initial rates of hydrolysis of 4MU-GlcNAc were measured in buffer containing MacIlvaine buffers, ranging from pH 4 to pH 8, plus 0.1 mg/ml BSA, with substrate concentrations of 200 μM and 0.2 nM enzyme in a 50 μl assay volume. After 20 min of incubation at 37 °C, the reaction was stopped by addition of 100 μl of 3 M glycine/NaOH (pH 10.3). Product formation was determined fluorimetrically ($\lambda_{ex} = 360$ nm/$\lambda_{em} = 460$ nm) using a FL_x800 microplate fluorescence reader (Bio-Tek). The highest value measured was arbitrarily set to 100 %. *Og*OGA possesses a pH optimum of 7.4–7.6, comparable with hOGA (pH 6.5), and compatible with an intracellular location.

Coupling of Fmoc amino acids (four equivalents) was mediated by PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) (four equivalents) in the presence of DIPEA (*N*,*N*-di-isopropylethylamine) (four equivalents) in DMF/*N*-methyl-2-pyrrolidone 4:1 at 70 °C for 10 min except for serine and histidine that were coupled at 50 °C to prevent epimerization. Acetylation was achieved after removal of Fmoc protection from the N-terminus with 20 % acetic anhydride in pyridine at 70 °C for 5 min. After completion of the synthesis, the resin was collected in a 15 ml disposable PP (polypropylene) reservoir (Biotage) fitted with a PP fritted disk and PTFE [poly(tetrafluoroethylene)] stopcock. The resin was successively washed with DCM (dichloromethane) (3 × 5 ml) and methanol (3 × 5 ml). The acetate protections from the sugar residue were then removed by treatment with a solution of hydrazine hydrate in methanol 1:4 (2 ml) for 1 h at room temperature (23 °C). The reagent was filtered off and the treatment was repeated twice for 30 min with the same amount of fresh reagent. Finally, the resin was successively washed with methanol (3 × 5 ml) and DCM (3 × 5 ml). The resin was then treated with a cleavage cocktail (2 ml) made of TFA (92 %), water (2.5 %), phenol (2.5 %), DODT (3,6-dioxa-1,8-octanedithiol) (2.5 %) and TIS (tri-isopropylsilane) (1 %) for 2 h at room temperature. The cleavage cocktail was drained off and collected, and the resin was washed with neat TFA (trifluoroacetic acid) (2 × 1 ml). The combined filtrate was concentrated *in vacuo* to approx. 1 ml. The peptides were precipitated with ice-cold diethyl ether (35 ml) and recovered by centrifugation at 3000 ***g*** for 15 min. The pellet was washed twice with fresh diethyl ether (35 ml), then blow dried with argon, dissolved in water/acetic acid/acetonitrile (3:1:0.5, by vol.) (5 ml) and freeze-dried. The crude glycopeptides were purified by reverse-phase HPLC on a 21.1 × 100 Luna C_{18} (2) column (Phenomenex) using a linear gradient of 5–95 % acetonitrile in water (0.1 % ammonia). UV detection was at 214 nm. Appropriate fractions were collected and freeze-dried to afford the target glycopeptides in 50–70 % yield based on the resin-loading capacity.

The structural homogeneity of the glycopeptides was verified by high-resolution mass spectra taken using a Bruker micro-TOF (time-of-flight) instrument in the direct injection mode.

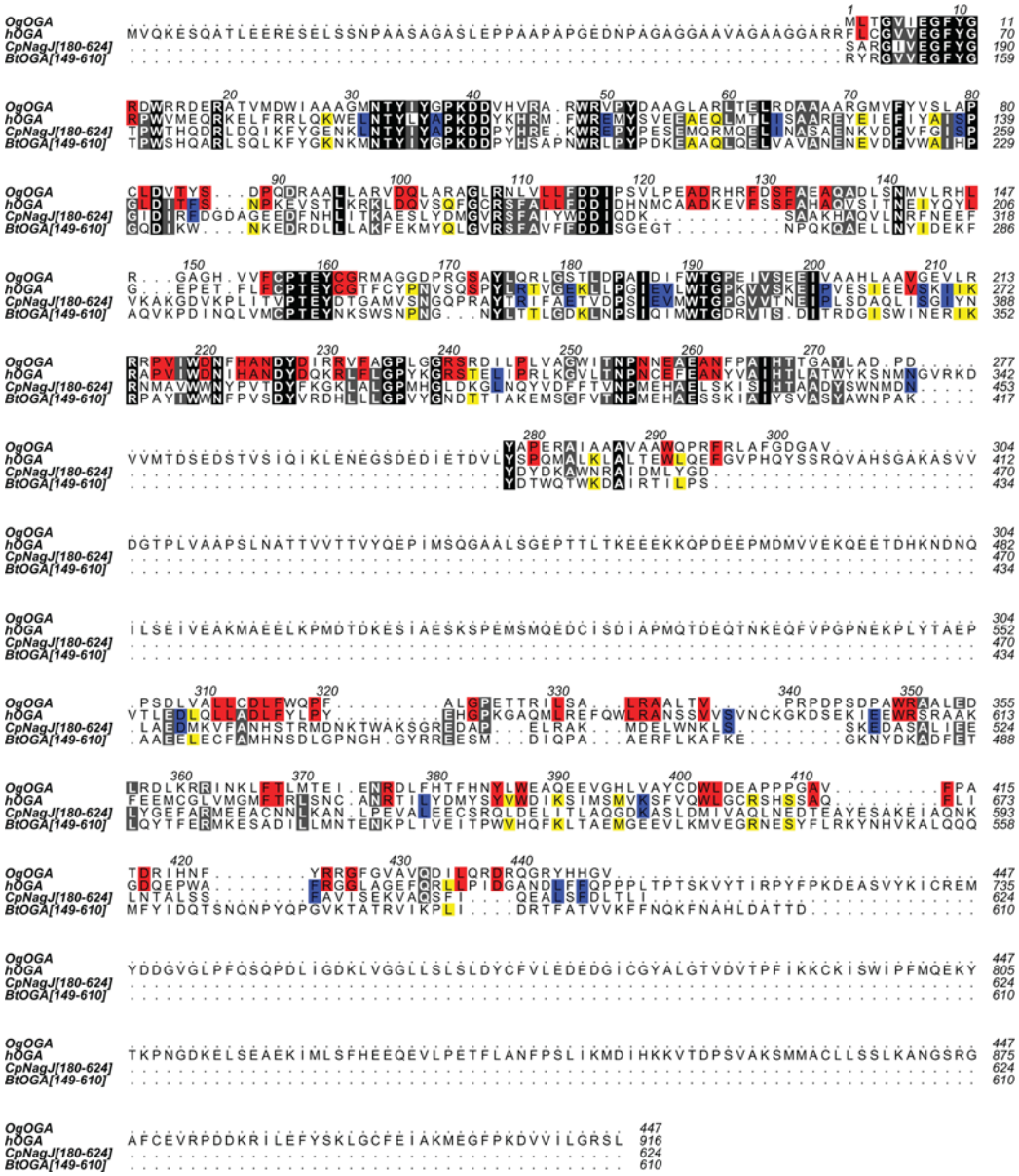

Figure S3 Sequence alignment of OGAs

Structure-based alignment of hOGA, *Og*OGA, *Cp*NagJ and *Bt*OGA, covering the catalytic and stalk domains. Black/grey boxes indicate sequence conservation, further coloured boxes indicate conservation between hOGA and *Og*OGA (red), hOGA and *Cp*NagJ (blue), hOGA and *Bt*OGA (yellow). *Bt*, *Bacillus thetaiotamicron*.

REFERENCES

1 Wilson, K. (2001) Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol., chapter 2, unit 2.4
2 Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. Protein Expression Purif. **41**, 207–234
3 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**, 307–326
4 Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. Z. Kristallogr. **217**, 644–650
5 Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. Acta Crystallogr. Sect. D Biol. Crystallogr. **58**, 1772–1779
6 Pape, T. and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. J. Appl. Crystallogr. **37**, 843–844
7 Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. **6**, 458–463
8 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 2126–2132
9 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D Biol. Crystallogr. **53**, 240–255
10 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. **50**, 760–763
11 Schüttelkopf, A. W. and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 1355–1363
12 Bond, C. S. and Schüttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. Sect. D Biol. Crystallogr. **65**, 510–512
13 Salvador, L. A., Elofsson, M. and Kihlberg, J. (1995) Preparation of bulding blocks for glycopeptide synthesis by glycosylation of Fmoc amino acids having unprotected carboxyl groups. Tetrahedron **5**, 5643–5656

Received 23 August 2010/9 September 2010; accepted 23 September 2010
Published as BJ Immediate Publication 23 September 2010, doi:10.1042/BJ20101338