Human O-GlcNAcase binds substrates in a conserved peptide recognition groove

Marianne Schimpl, Alexander Wolfgang Schuettelkopf, Vladimir S Borodkin, Daan Mf van Aalten

Biochemical Journal Immediate Publication. Published on 23 Sep 2010 as manuscript BJ20101338

Human *O*-GlcNAcase 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.

[1]*These authors contributed equally to this work*

[2]*To whom correspondence should be addressed, email: dmfvanaalten@dundee.ac.uk*

Classification: Biological Sciences (Biochemistry)

Abstract

Modification of cellular proteins with *N*-acetylglucosamine (*O*-GlcNAc) competes with protein phosphorylation and regulates a plethora of cellular processes. *O*-GlcNAcylation is orchestrated by two opposing enzymes, *O*-GlcNAc transferase and *O*-GlcNAcase, which recognize their target proteins via as yet unidentified mechanisms. Here, we uncover the first insights into the mechanism of substrate recognition by human *O*-GlcNAcase. The structure of a novel bacterial *O*-GlcNAcase orthologue reveals a putative substrate binding groove, conserved with metazoan *O*-GlcNAcases. Guided by this structure, conserved amino acids lining this groove in human *O*-GlcNAcase were mutated and the activity on three different substrate proteins (TAB1, FoxO1 and CREB) was tested in an *in vitro* de-glycosylation assay. The results provide the first evidence that human *O*-GlcNAcase 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.

Introduction

The *O*-GlcNAc modification (O-linked *N*-acetylglucosamine) 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 recent 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 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 *O*-GlcNAc transferase (OGT) and its antagonistic β-*N*-acetylglucosaminidase, *O*-GlcNAcase (OGA) are encoded by single genes in metazoa (11, 12). Little is known about the mode of substrate recognition of either enzyme, although recent 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 human OGA (hOGA) is frequently over-simplified by designating an N-terminal hexosaminidase followed by a C-terminal domain with sequence homology to histone acetyl transferases (referred to as the HAT domain here). A further putative domain may exist in the stretch of over 300 amino

acids in between the hexosaminidase and HAT domains. hOGA (Ser405) 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 has been used extensively to raise cellular *O*-GlcNAc levels (17). Recent 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 *O*-GlcNAcase interacts with the protein part of its substrates. Such interactions could constitute the basis of substrate specificity in human *O*-GlcNAcase. Here, 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.

Results & 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 (22), 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 TAK-1 binding protein-1 (TAB1) and the transcription factors FoxO1 and CREB (Fig. 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, showing that hOGA is capable of removing *O*-GlcNAc from these glycoproteins (Fig. 1A). We next sought to identify peptide substrates that would allow the determination of the Michaelis constant (K_M), and also allow for varying of sequence and length to probe substrate specificity. Peptides covered the *O*-GlcNAc sites on TAB1 (Ser395) (23) and hOGA (Ser405) (4) (Table I). hOGA activity on these peptides was confirmed by mass spectrometry. The affinity of hOGA for these peptides was then quantified by their ability to compete with the reporter substrate 4MU-GlcNAc (Fig. 1B, Table I). The resulting K_M values for the hOGA-Ser405 and TAB1-Ser395 derived peptides are in the high micromolar and low millimolar range, respectively, somewhat decreasing in affinity with increasing peptide length.

Discovery of a bacterial O-*GlcNAcase 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 (707-916, Fig. 2A). However, sequence analysis of metazoan OGAs shows that they also possess a low complexity N-terminus (residues 1-59), and a "middle" domain (367-707, also containing a low complexity region (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 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 *Oceanicola granulosus* (*Og*OGA) that not only shows improved sequence identity (37%) with the hOGA catalytic domain without any alignment gaps (Fig. 2B), but intriguingly also possesses a 200 residue C-terminal domain that shares about 20% identity with the hOGA middle domain (Figs. 2A,B). 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 possible *O*-GlcNAcase 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 ± 0.6 mM, k_{cat} = 24.0 ± 1.0 s^{-1}) and 4MU-GlcNAc (K_M = 2.9 ± 0.2 mM, k_{cat} = 8.2 ± 0.3 s^{-1}), with a pH optimum of 7.4-7.6 (Fig. S5). *Og*OGA is sensitive to the commonly used *O*-GlcNAcase inhibitors PUGNAc (IC_{50} = 11 μM) and GlcNAcstatin G (24) (IC_{50} = 780 nM). *Og*OGA was also able to quantitatively deglycosylate a glyco-pentapeptide, AHS(β-*O*-

GlcNAc)GA, derived from the reported hOGA *O*-GlcNAc site (4), as detected by LC-MS (Figs. 1C, 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 (Table S1). The *Og*OGA structure reveals an N-terminal TIM barrel domain and a C-terminal α-helical domain (Figs. 2A,C). The *Og*OGA catalytic domain is similar to that of *Cp*NagJ (RMSD = 1.5 Å for 254 C_αs, Fig. 2C). The α4-β5, β3-α3, β4-α4 and β5-α5 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 (Fig. 2B). The C-terminal domain, from hereon referred to as the stalk domain, possesses good sequence similarity with hOGA (Fig. 2B) and contains a core 4-helical bundle (Fig. 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, Fig. 2C), although the loops connecting the helices differ in length and conformation. In particular the α10-α11 loop is nine residues longer in *Cp*NagJ compared to *Og*OGA (and, by alignment, hOGA, Fig. 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 Fig. S3 in the Supplementary Material).

The *Og*OGA active site is formed by the TIM barrel β-α loops (Fig. 2C), with the catalytic loop residues Asp115/Asp116 on β4-α4 conserved with hOGA, compatible with the

observed *O*-GlcNAcase activity (Fig. 2C). This loop does not fully engage with the PUGNAc inhibitor, presumably related to the disordered phenylcarbamate group. Ligand-induced conformational changes in the β4-α4 loop have also been observed for *Cp*NagJ (25). Upon PUGNAc binding to *Og*OGA, the β7-η1 loop shifts by up to 4.7 Å exposing the GlcNAc binding site, associated with small displacements of the neighboring β8-α8 and β1-α1 loops. Residues hydrogen bonding with the sugar hydroxyls (Gly8, Tyr160, Asp226, Asn254) or the N-acetyl moiety (Asn221) are conserved in hOGA (Figs. 2B,E). Similarly, the residues that show van der Waals/stacking interactions with the sugar and the N-acetyl group (Tyr10, Tyr160, Trp219) 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 (Fig. 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 twenty individual hOGA point mutants of residues lining the groove (Figs. 2B,D). To control for effects on protein folding and/or catalysis rather than protein substrate binding, we measured 4MU-GlcNAc hydrolysis by these mutants (Fig. 3A). Mutations affecting 4MU-GlcNAc hydrolysis are mostly located close to the sugar/catalytic machinery (Tyr69) or occupy positions shown to undergo conformational change upon inhibitor binding (Tyr286 and Asp287 in the β7-η1 loop) (Fig. 2). Strikingly, de-*O*-GlcNAcylation assays with the TAB1, FoxO1 and CREB glycoprotein substrates identified mutants at 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 (Fig. 3).

Conclusions

Since its discovery more than three decades ago, it has become clear that *O*-GlcNAc is an important posttranslational modification found on hundreds of intracellular proteins and plays a role in signalling pathways. However, it is not clear what determines specificity in the targetting and cycling of *O*-GlcNAc on these proteins. We have investigated whether *O*-GlcNAcase specifically recognizes the peptide component of its glycoprotein substrates. Our data 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 proximity to the active site. Indeed, mutations targetting 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) on these proteins/peptides, and could tune *O*-GlcNAc cycling rates. Together, the data presented here 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.

Materials and Methods

Cloning, expression and structure determination of OgOGA

The *Oceanicola granulosus* gene for ZP_01156836.1 was cloned from genomic DNA into pGEX6-P1 for expression as a GST fusion in *E. coli*. The protein was purified by affinity, anion exchange and size exclusion chromatography (Supplemental Material). *Og*OGA was crystallized at 1.9 mg/ml from PEG/$MgCl_2$ solutions (Supplemental Material). The structure of the apo-enzyme was solved using MAD 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 Table S1.

Glycopeptide synthesis and detection

A 3,4,6-Triacetyl-*O*-GlcNAc-Fmoc-SerOH derivative was synthesized are described previously (26). Microwave assisted solid phase peptide synthesis was performed with a CEM Liberty on low load Rink amide MBHA resin 100-200 mesh (Novabiochem) using standard Fmoc chemistry protocols on 0.05 mmol scale. The structural homogeneity of the glycopeptides was verified by high resolution mass spectra taken on Bruker micro-TOF instrument in the direct injection mode. The same instrument was used for the monitoring of the enzymatic deglycosylation in the LC-MS mode using an Agilent 1200 binary analytical HPLC system equipped with Waters X-Bridge C-18 3 mm 3.2×50 mm column and diode array UV-VIS detector. Typically, the reaction was diluted to 0.3 ml with water, then passed trough a 10 kDa MWCO concentrator to remove protein. 10 µl of the flow-through was taken for the analysis using gradient elution (flow rate 0.3 ml/min) with a 5 to 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 (27) for transient expression in mammalian cells. Point mutations were introduced by site-directed mutagenesis (see Table S2 for primers). The enzyme was expressed as an N-terminal GST fusion in HEK293 cells cultivated in suspension. For transfection, 2×10^8 HEK293 cells were serum-deprived for 4 h in 100 ml Pro293sCDM medium prior to the addition of 10 ml serum-free medium containing 100 µg plasmid DNA and 600 µg polyethylenimine. After another 4 h, the cells were diluted to 1×10^6 /ml, and the medium supplemented with 1 % foetal calf serum. Cell pellets were harvested 72 h post transfection and lysed in 50 mM Tris/HCl pH 7.5, 270 mM sucrose, 0.1 % 2-mercaptoethanol, 1 mM Na-orthovanadate, 1 mM EDTA, 1 mM EGTA, 5 mM pyrophosphate, 10 mM Na-β-glycerophosphate, 50 mM NaF, 1 % Triton-X100, 1 mM benzamidine, 0.2 mM PMSF, 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 O-*GlcNAcase assays*

*Og*OGA: Initial rates of hydrolysis of *p*NP-β-GlcNAc were measured in buffer containing 25 mM Tris/HCl, 0.1 mg/ml BSA, with substrate concentrations ranging from 0−10 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 incubation at 37 °C, the reaction was stopped by addition of 100 µl 3 M glycine/NaOH (pH 10.3), and product concentration was measured (A = 405 nm). 4MU-β-GlcNAc assays were performed in the same buffer, substrate concentrations ranging from 0−8 mM, the enzyme concentration was reduced to 0.2 nM and

the assay time to 20 min. Product formation was determined fluorimetrically (λ_{ex} = 360 nM/λ_{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, data were analysed and plotted with 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 multi-substrate enzyme kinetics to derive kinetic constants of a substrate whose turnover cannot be detected directly (28). The fluorogenic 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 the equation below. (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, and S' the concentration of the peptide.)

$$\frac{v_i}{v_0} = \frac{1 + \frac{K_m}{S}}{1 + \frac{K_m}{S}\left(1 + \frac{S'}{K'_m}\right)}$$

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

The known *O*-GlcNAc targets FoxO1 (29, 30) and CREB (31) were provided by the Division of Signal Transduction and Therapy, University of Dundee (DSTT). 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 N-terminal domain and full-length human OGT were purified as described elsewhere (Clarke et al., 2008). A truncated construct lacking the first 9 TPR 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. 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 (Fig. S1). Full-length hOGT was used to *O*-GlcNAcylate CREB and hOGA, while the truncated Δ9-hOGT was used for FoxO1 and TAB1. After the glycosylation reaction, target proteins were re-purified by GST pull-down, eluted in 25 mM Tris/HCl buffer containing 25 mM reduced glutathione and spin-concentrated to approximately 5 mg/ml.

In vitro removal of *O*-GlcNAc was performed in 5 µl volume in 25 mM Tris/HCl, 25 mM glutathione pH 7.5, and contained 10 µM *O*-GlcNAcylated substrate protein and 1 µM wild-type or mutant hOGA. Reactions 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. Anti-GST and anti-TAB1 antisera were obtained from the DSTT (32).

Acknowledgements

This work was supported by a Wellcome Trust Senior Research Fellowship to DvA and a Wellcome Trust PhD Studentship to MS. We thank Sharon Shepherd for assistance with protein production, and the European Synchrotron Radiation Facility, Grenoble, for time on BM14, ID23-2 and ID14-1. Structures have been deposited in the Protein Data Bank (PDB id XXX, YYY - deposition in progress).

References

1. Love DC, Hanover JA (2005) The hexosamine signaling pathway: deciphering the "O-GlcNAc code" *Sci STKE* 2005: re13.

2. Hart GW, Housley MP, Slawson C (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins *Nature* 446: 1017-1022.

3. Torres CR, Hart GW (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, et al. (2007) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics *Nat Chem Biol* 3: 339-348.

5. Vosseller K, 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, et al. (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, et al. (2010) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis *Sci Signal* 3: ra2.

8. Zachara NE, Hart GW (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, et al. (2005) Perturbations in O-linked beta-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, 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, Meese E (2001) Identification of a nuclear variant of MGEA5, a cytoplasmic hyaluronidase and a beta-N-acetylglucosaminidase *Biochem Biophys Res Commun* 283: 634-640.

12. Hanover JA, et al. (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene *Arch Biochem Biophys* 409: 287-297.

13. Cheung WD, Sakabe K, Housley MP, Dias WB, Hart GW (2008) O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins *J Biol Chem* 283: 33935-33941.

14. Heckel D, et al. (1998) Novel immunogenic antigen homologous to hyaluronidase in meningioma *Hum Mol Genet* 7: 1859-1872.

15. Copeland RJ, Bullen JW, Hart GW (2008) Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity *Am J Physiol Endocrinol Metab* 295: E17-28.

16. Whisenhunt TR, et al. (2006) Disrupting the enzyme complex regulating O-GlcNAcylation blocks signaling and development *Glycobiology* 16: 551-563.

17. Haltiwanger RS, Grove K, Philipsberg GA (1998) Modulation of O-linked N-acetylglucosamine levels on nuclear and cytoplasmic proteins in vivo using the peptide O-GlcNAc-beta-N-acetylglucosaminidase inhibitor O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate *J Biol Chem* 273: 3611-3617.

18. Macauley MS, Stubbs KA, Vocadlo DJ (2005) O-GlcNAcase catalyzes cleavage of thioglycosides without general acid catalysis *J Am Chem Soc* 127: 17202-17203.

19. Rao FV, et al. (2006) Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis *Embo J* 25: 1569-1578.

20. Dennis RJ, et al. (2006) Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity *Nat Struct Mol Biol* 13: 365-371.

21. Macauley MS, Vocadlo DJ (2010) Increasing O-GlcNAc levels: An overview of small-molecule inhibitors of O-GlcNAcase *Biochim Biophys Acta* 1800: 107-121.

22. Gao Y, Wells L, Comer FI, Parker GJ, Hart GW (2001) 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.

23. Pathak S, van Aalten DMF (2010) O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release *manuscript in preparation*.

24. Dorfmueller HS, et al. (2010) Cell-permeable O-GlcNAcase inhibitors with improved potency and selectivity *manuscript in preparation*.

25. Pathak S, Dorfmueller HC, Borodkin VS, van Aalten DMF (2008) Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death *Chem Biol* 15: 799-807.

26. Salvador LA, Elofsson M, Kihlberg J (1995) Preparation of building blocks for glycopeptide synthesis by glycosylation of Fmoc amino acids having unprotected carboxyl groups *Tetrahedron* 51: 5643-5656.

27. Sanchez I, et al. (1994) Role of SAPK/ERK kinase-1 in the stress-activated pathway regulating transcription factor c-Jun *Nature* 372: 794-798.

28. Xie D, Suvorov L, Erickson JW, Gulnik AS (1999) Real-time measurements of dark substrate catalysis *Protein Sci* 8: 2460-2464.

29. Housley MP, et al. (2008) O-GlcNAc regulates FoxO activation in response to glucose *J Biol Chem* 283: 16283-16292.

30. Kuo M, Zilberfarb V, Gangneux N, Christeff N, Issad T (2008) O-glycosylation of FoxO1 increases its transcriptional activity towards the glucose 6-phosphatase gene *FEBS Lett* 582: 829-834.

31. Lamarre-Vincent N, Hsieh-Wilson LC (2003) Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation *J Am Chem Soc* 125: 6612-6613.

32. Cheung PC, Campbell DG, Nebreda AR, Cohen P (2003) Feedback control of the protein kinase TAK1 by SAPK2a/p38alpha *Embo J* 22: 5793-5805.

Figure legends

Figure 1: *O*-GlcNAcase activity of hOGA and *Og*OGA

A) The *O*-GlcNAc proteins (10 μM) TAB1, FoxO1 and were subjected to *in-vitro*-GlcNAcylation by recombinant OGT (Fig. S1). 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. 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 (longer exposure).

B) Competition of β-*O*-GlcNAc modified peptides (tri-, penta-, hepta- and nona-peptides based on the *O*-GlcNAc site on hOGA Ser405) with 4MU-GlcNAc turnover by hOGA affords an estimate of the affinity of hOGA for the peptide substrates (Table I).

C) Mass spectrum showing the deglycosylation of AHS(-β-*O*-GlcNAc)GA (MW 785) to the deglycosylated form (MW 582), also see Fig. S2.

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), dashed lines denote residues disordered in apo-*Og*OGA. The proposed domains are indicated by boxes coloured as in panel A.

C) Cartoon view of the *Og*OGA apostructure and domains 2-3 of *Cp*NagJ.

D) Surface view of *Og*OGA shaded by sequence conservation amongst 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;φ_c map (2.3 σ) is shown for PUGNAc. Hydrogen bonds are drawn as black dashed lines. Hydrogen bonds to/from protein backbone amides are shown as originating from the Cα. Active site residues in *Og*OGA are labeled 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 to the wild type (k_{cat}/K_M = 7000 M^{-1} s^{-1}) and the catalytically impaired D175N mutant (negative control). hOGA mutants were also incubated with substrate *O*-GlcNAc proteins TAB1, FoxO1 and CREB, and de-glycosylation was observed by Western blot.

B) Visualisation of the correlation between mutants' ability to hydrolyse the pseudosubstrate 4MU-GlcNAc and glycoprotein substrates. Catalytically impaired mutants (k_{cat}/K_M for 4MU-GlcNAc $\leq$ 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 *O*-GlcNAcase 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, Fig. 1A). The mutated amino acids are shaded according to correlation plots, thus amino acids affecting protein substrate recognition are highlighted in yellow.

Table I

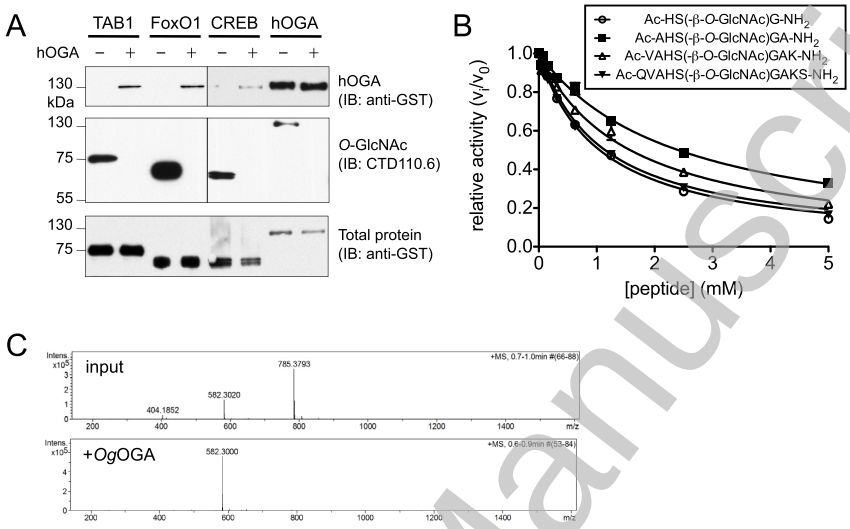
Michaelis constants of *O*-GlcNAc modified peptides for hOGA.

K_M values for peptides derived from the *O*-GlcNAc sites of two known substrate proteins were determined in a competitive assay with the fluorogenic substrate 4MU-GlcNAc.

Human *O*-GlcNAcase Ser405 derived peptide series:	K_M
Ac-HS(-β-*O*-GlcNAc)G-NH_2	(525 ± 13) μM
Ac-VAHS(-β-*O*-GlcNAc)GA-NH_2	(1204 ± 43) μM
Ac-VAHS(-β-*O*-GlcNAc)GAK-NH_2	(789 ± 36) μM
Ac-QVAHS(-β-*O*-GlcNAc)GAKAS-NH_2	(528 ± 15) μM
TAB1 S395 derived peptide series:	
Ac-YS(-β-*O*-GlcNAc)S-NH_2	(2.2 ± 0.1) mM
Ac-PYS(-β-*O*-GlcNAc)SA-NH_2	(6.3 ± 0.5) mM
Ac-VPYS(-β-*O*-GlcNAc)SAQ-NH_2	(4.1 ± 0.3) mM
Ac-SVPYS(-β-*O*-GlcNAc)SAQS-NH_2	(4.8 ± 0.3) mM

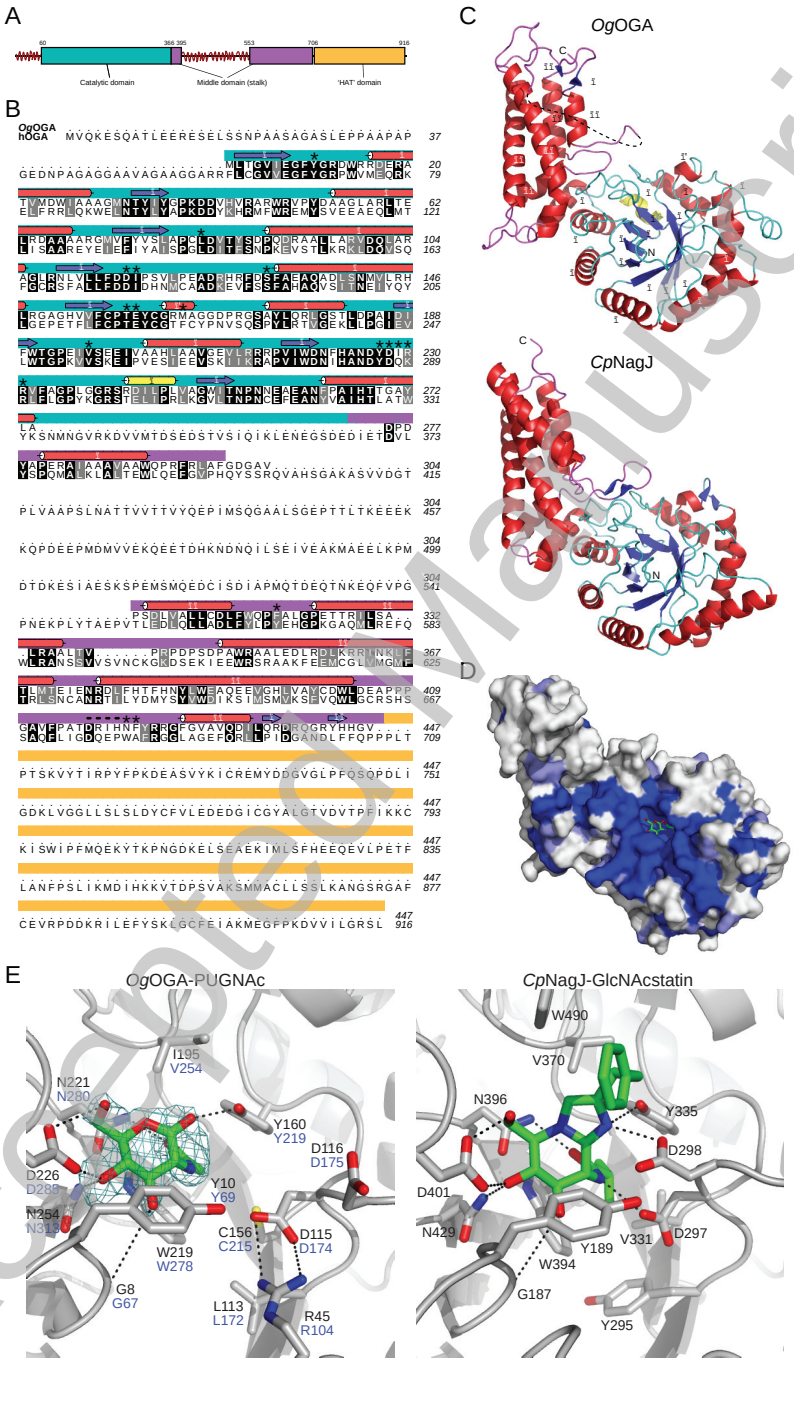
Biochemical Journal Immediate Publication. Published on 23 Sep 2010 as manuscript BJ20101338

Figure 1



Biochemical Journal Immediate Publication. Published on 23 Sep 2010 as manuscript BJ20101338

Figure 2



Biochemical Journal Immediate Publication. Published on 23 Sep 2010 as manuscript BJ20101338

Figure 3