LETTER

doi:10.1038/nature09638

Structure of human *O*-GlcNAc transferase and its complex with a peptide substrate

Michael B. Lazarus[1,4]*, Yunsun Nam[2,3]*, Jiaoyang Jiang[4], Piotr Sliz[2,3] & Suzanne Walker[4]

The essential mammalian enzyme O-linked β-*N*-acetylglucosamine transferase (*O*-GlcNAc transferase, here OGT) couples metabolic status to the regulation of a wide variety of cellular signalling pathways by acting as a nutrient sensor[1]. OGT catalyses the transfer of *N*-acetylglucosamine from UDP-*N*-acetylglucosamine (UDP-GlcNAc) to serines and threonines of cytoplasmic, nuclear and mitochondrial proteins[2,3], including numerous transcription factors[4], tumour suppressors, kinases[5], phosphatases[1] and histone-modifying proteins[6]. Aberrant glycosylation by OGT has been linked to insulin resistance[7], diabetic complications[8], cancer[9] and neurodegenerative diseases including Alzheimer's[10]. Despite the importance of OGT, the details of how it recognizes and glycosylates its protein substrates are largely unknown. We report here two crystal structures of human OGT, as a binary complex with UDP (2.8 Å resolution) and as a ternary complex with UDP and a peptide substrate (1.95 Å). The structures provide clues to the enzyme mechanism, show how OGT recognizes target peptide sequences, and reveal the fold of the unique domain between the two halves of the catalytic region. This information will accelerate the rational design of biological experiments to investigate OGT's functions; it will also help the design of inhibitors for use as cellular probes and help to assess its potential as a therapeutic target.

The ability to sense and respond to nutrient levels is critical for the growth of all living systems. In eukaryotes, a major mechanism for nutrient sensing involves the essential[11] protein glycosyltransferase OGT, which senses cellular glucose levels via UDP-GlcNAc concentrations, and responds by dynamically *O*-GlcNAcylating a wide range of nuclear and cytoplasmic proteins[1,12]. These include proteins involved in insulin-like signalling pathways[7] and transcriptional activators that regulate glucose levels by controlling gluconeogenesis[13]. As many known *O*-GlcNAcylation sites are also phosphorylation sites, OGT is proposed to play a major role in modulating cellular kinase signalling cascades[14]. OGT is also involved in widespread transcriptional regulation[15–17]. Prolonged hyperglycaemia, such as occurs in diabetes, or excessive glucose uptake, such as occurs in cancer cells, results in hyper-*O*-GlcNAcylation of cellular proteins by OGT, and this increased *O*-GlcNAcylation has been linked to harmful cellular effects[18]. Thus, strategies to modulate OGT activity may have therapeutic value for treating diabetic complications, cancer, and other diseases[13].

The lack of a crystal structure has been a major impediment to investigating OGT's molecular mechanisms, understanding substrate recognition, and developing inhibitors. OGT comprises two distinct regions: an N-terminal region consisting of a series of tetratricopeptide repeat (TPR) units[19,20] and a multidomain catalytic region. The TPR domain is proposed to scaffold interactions with other proteins, which may play a role in determining substrate selectivity[21]. A crystal structure comprising 11.5 TPR units of human OGT has been reported[21], but there have been no structures of the catalytic region. From sequence analysis and structures of bacterial glycosyltransferases[22–26],

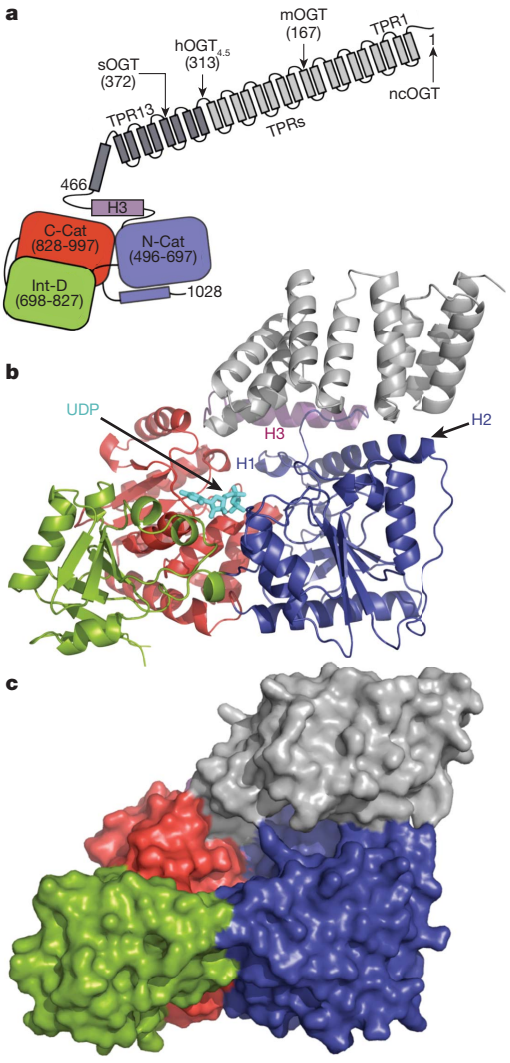

Figure 1 | Overall structure of human OGT complexed to UDP. **a**, Schematic of OGT domain architecture with the TPR units shown in grey, the transitional helix (H3) in purple, the N-Cat domain in blue, the Int-D domain in green, and the C-Cat domain in red. The native isoforms of OGT (sOGT, short OGT; mOGT, mitochondrial OGT; and ncOGT, nucleocytoplasmic OGT) and the crystallization construct differ only in the number of TPRs, as shown. **b**, Overall fold of OGT from the OGT–UDP complex in a ribbon representation. The colouring is the same as the schematic in **a**. The UDP is shown in cyan. The N-Cat domain helices unique to OGT are indicated as H1 and H2. **c**, Surface representation of the OGT–UDP complex. The colouring scheme is the same as in **a** and **b**.

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Laboratory of Molecular Medicine, Children's Hospital, Boston, Massachusetts 02115, USA. [4]Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

including a bacterial homologue of unknown function[25,26], OGT was predicted to be a member of the GT-B superfamily of glycosyltransferases (Gtfs)[27]. However, OGT is unusual because it is the only known member to glycosylate polypeptides and it contains a long uncharacterized intervening sequence (~120 amino acids) in the middle of the catalytic region. It is also proposed to contain a phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) binding domain involved in membrane recruitment in response to insulin signalling[7].

We report two crystal structures of a human OGT construct ($hOGT_{4.5}$) containing 4.5 TPR units and the catalytic domain. The catalytic properties of this construct are similar to those of the full-length enzyme (Supplementary Fig. 1)[28]. One structure (2.8 Å, referred to as OGT–UDP) is a complex with UDP; the other structure (1.95 Å, referred to as OGT–UDP–peptide) is a complex containing UDP and a well-characterized 14-residue CKII peptide substrate[28]. On the basis of currently available experimental data, we also present a model for the full-length enzyme (Supplementary Information). Details of structure determination are presented in Methods and Supplementary Tables 1 and 2.

The OGT–UDP complex is shown in Fig. 1. The catalytic region contains three domains: the amino (N)-terminal domain (N-Cat), the carboxy (C)-terminal domain (C-Cat), and the intervening domain (Int-D) (Fig. 1a, b). The N-Cat and C-Cat domains have Rossmann-like folds typical of GT-B superfamily members; however, the N-Cat domain is distinctive in containing two additional helices, H1 and H2, which form an essential part of the active site (Fig. 1b). The Int-D domain, which has a novel fold, packs exclusively against the C-Cat domain (Fig. 1c). The UDP moiety binds in a pocket in the C-Cat domain near the interface with the N-Cat domain[27]. This pocket is lined with conserved residues shown to be important for catalytic activity (Supplementary Table 3)[25,26]. A transitional helix (H3) links the catalytic region to the TPR repeats, which spiral along the upper surface of the catalytic region from the C-Cat domain to the N-Cat domain. The TPRs and the catalytic region are demarcated by a narrow horizontal cleft.

The OGT–UDP–peptide complex (Fig. 2), which crystallized in a different space group from the OGT–UDP complex, has a wider cleft between the TPR domain and the catalytic region than the OGT–UDP complex (Fig. 1c and Fig. 2a), and the CKII peptide binds in this cleft. This peptide, YPGGSTPVS*SANMM, contains three serines and a threonine, but only one serine (underlined; referred to as Ser*) is glycosylated by OGT[28]. The hydroxyl of Ser* points into the nucleotide-sugar binding site (Fig. 2b). The two residues N-terminal to Ser* lie over the UDP moiety; the residues C-terminal to Ser* traverse towards the back of the cleft along the H2 helix of the N-Cat lobe. Although OGT glycosylates a wide range of target peptides, it prefers sequences in which the residues flanking the glycosylated amino acid enforce an extended conformation (for example, prolines and β-branched amino acids; see Supplementary Fig. 2 and Supplementary Table 4). Consistent with these preferences, the peptide is anchored mainly by contacts from

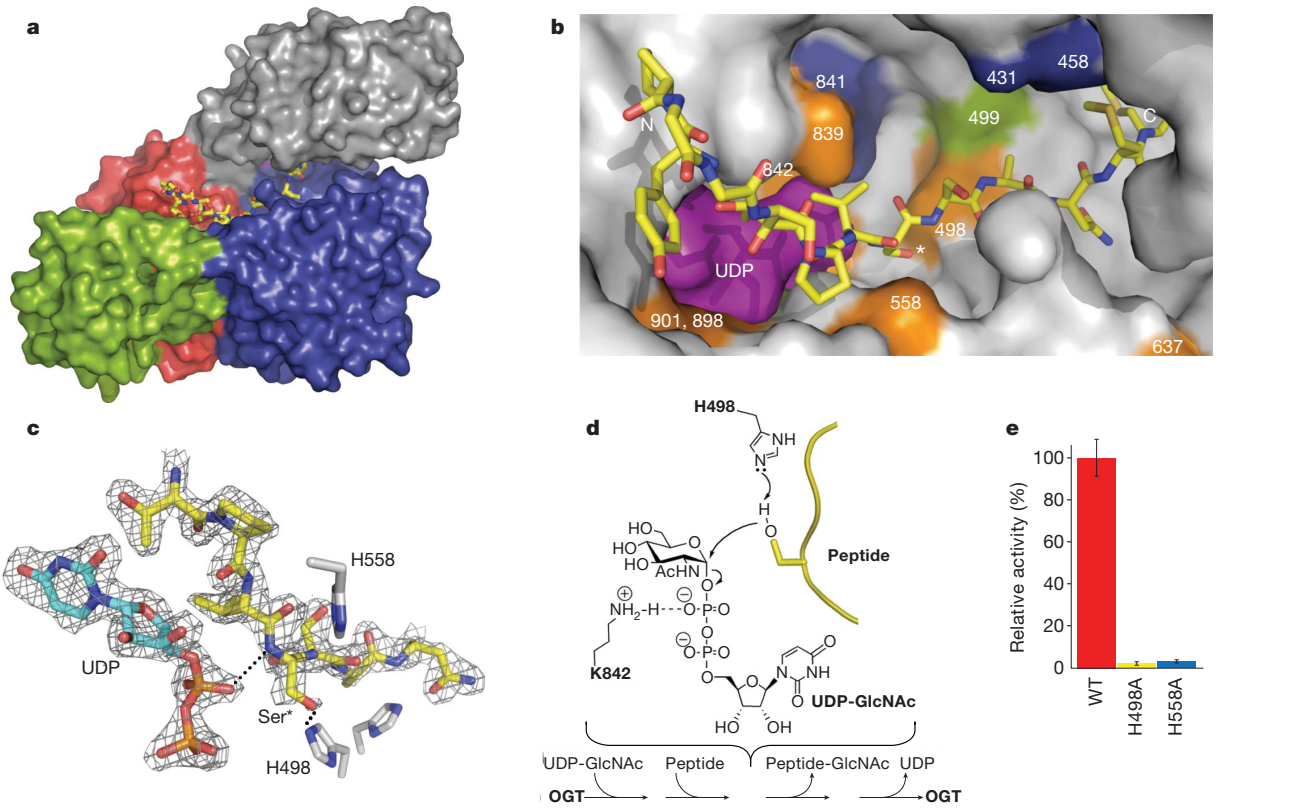

Figure 2 | Structure of the OGT–UDP–peptide complex. **a**, Surface rendering of the OGT complex with UDP and the CKII peptide substrate[28]. The view and the colouring is the same as in Fig. 1. The peptide, shown in yellow, lies over the UDP moiety, which is not visible in this orientation. **b**, Close-up surface rendering of the OGT active site (grey) containing the CKII peptide in a stick representation (carbon atoms shown in yellow) with the UDP (purple) in a space filling representation lying directly underneath it. The reactive serine is indicated by an asterisk. The peptide binds in the cleft between the TPR region and the catalytic region, and extends along the interface between the C-Cat and N-Cat domains. Protein residues implicated in catalytic activity are coloured orange, green, or blue in decreasing order of importance based on residual activity after mutation (Supplementary Table 3). Lysine 842 (orange) lies underneath UDP in this view. **c**, View of UDP (carbon atoms shown in cyan) and part of the CKII peptide (carbon atoms shown in yellow) with selected OGT side chains shown. Dashed lines indicate inferred hydrogen bonds based on distances in the OGT–UDP–peptide complex. The $2F_o - F_c$ omit map is contoured at the 1σ level. **d**, Proposed mechanism of OGT. The ordered sequential bi-bi kinetic mechanism shown is based on the structure of the ternary complex and supporting kinetic experiments (Supplementary Fig. 4). The peptide is depicted in yellow with only the reactive serine hydroxyl shown. H498 is the proposed catalytic base. Lys 842, also shown to be essential for activity[25,26], stabilizes the UDP moiety. **e**, Histogram showing the relative activities of the H498A and H558A mutants compared to the wild-type (WT) protein (average ± s.d., $n = 3$).

OGT side chains to the amide backbone, with an additional contact from the UDP moiety to the backbone amide of Ser*. The cleft is also filled with ordered water molecules, enabling it to serve as an adaptable interface to bind a range of polypeptides containing side chains of different sizes, polarity, and hydrogen bonding capabilities. As the peptide substrate is anchored by contacts to its backbone, it is reasonable to infer that protein substrates are glycosylated on flexible regions such as loops or termini that can bind in an extended conformation, exposing the amide backbone.

The closed conformation of the substrate-binding cleft in the OGT–UDP structure is stabilized by a 'latch' comprising contacts between TPRs 10/11 and the H2 helix of the catalytic domain (Fig. 2a and Supplementary Fig. 3). Opening of the cleft in the OGT–UDP–peptide complex occurs owing to a hinge-like motion around a pivot point between TPRs 12 and 13. The two structures suggest that glycosylation substrates enter the active site from the face of the enzyme shown in Fig. 2a, with the TPR domain restricting or allowing access, depending on its conformation and its interactions with the catalytic domain. Molecular dynamics simulations indicate that the 'hinge' between the catalytic domain and the TPR domain is capable of large motions that fully expose the active site, which would allow protein substrates to approach closely enough for surface loops to enter (Supplementary Movie 1). The molecular mechanisms that facilitate or stabilize opening of the cleft to allow access of protein substrates remain to be determined, but may involve interactions between protein substrates or adaptor proteins and the other regions of OGT.

The OGT–UDP–peptide complex, in addition to revealing how peptide substrates bind, provides unexpected insights into the kinetic mechanism. OGT was previously proposed to have a random sequential 'bi-bi' mechanism in which either substrate can bind first[28]. The structure, however, indicates that the peptide substrate binds over the nucleotide-sugar binding pocket, blocking access to it. Moreover, the α-phosphate of the UDP moiety contacts the backbone amide of Ser* (Fig. 2c), which helps orient the peptide. The peptide complex suggests an ordered mechanism in which UDP-GlcNAc binds before the polypeptide substrate. To assess the order of substrate binding, we analysed the product inhibition patterns for UDP. At saturating peptide concentrations, a competitive inhibition pattern was obtained for UDP with respect to UDP-GlcNAc, which is inconsistent with a random mechanism, but supports the ordered sequential bi-bi mechanism implied by the crystal structure (Supplementary Fig. 4).

Another insight from the crystal structure is the identity of the catalytic base. On the basis of analyses of other GT-B family members, including the bacterial OGT homologue, it was proposed that His 558 is the catalytic base. Although we have verified that this residue is critical for catalytic activity, the peptide complex shows that it is more than 5 Å away from the reactive serine hydroxyl and makes an apparent hydrogen bond with the backbone carbonyl of the preceding residue. In contrast, His 498, which is invariant in metazoan OGTs but absent in the homologous bacterial enzyme, protrudes from helix H1 into the active site within 3.5 Å of the Ser* hydroxyl. As His 498 is critical for activity and is located between the reactive serine hydroxyl and the GlcNAc binding pocket, it is the probable catalytic base in OGT.

We were unable to obtain a crystal of the OGT–UDP-GlcNAc complex owing to hydrolysis of the substrate, but according to the computational docking experiments we performed, the GlcNAc is oriented in a manner that exposes its β-face to the overlying peptide (Supplementary Fig. 5) and places the anomeric carbon near the reactive serine. This conformation is similar to the UDP-GlcNAc conformation observed in a complex of another GT-B family member[23], and its relevance is supported by evidence that the C2 N-acetyl moiety projects up from the OGT sugar binding pocket[29]. Furthermore, it is consistent with the enzymatic reaction, which

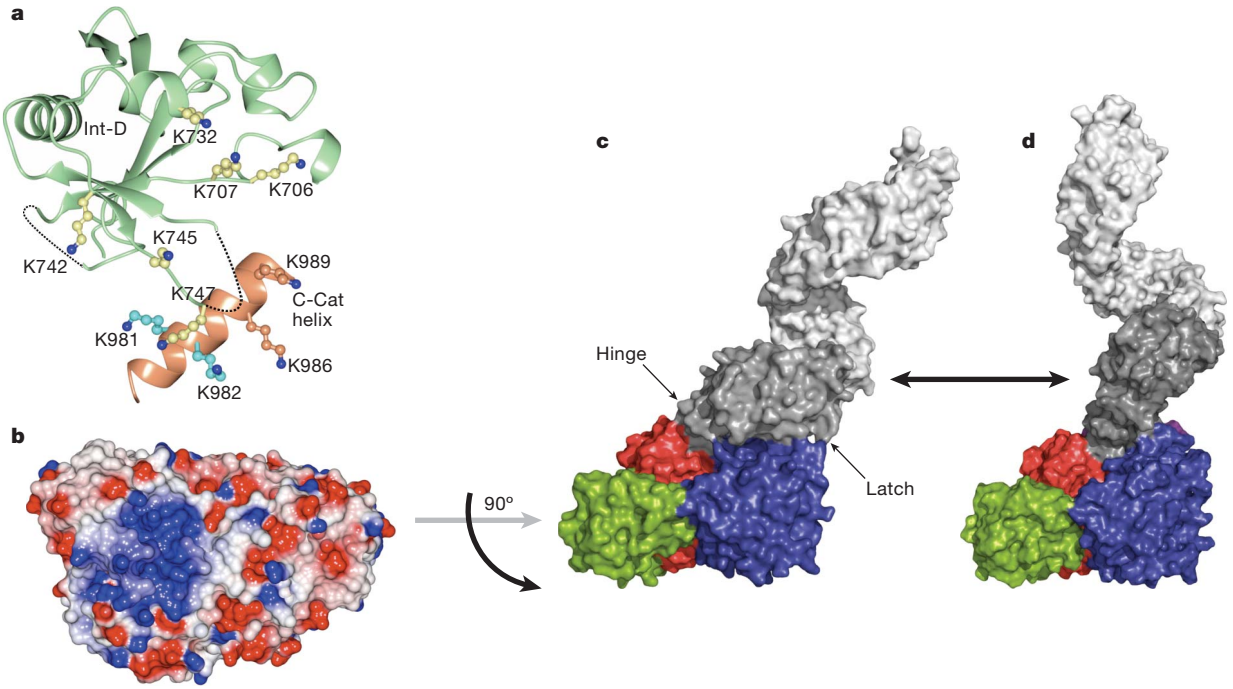

Figure 3 | Structure of the intervening domain and full-length models of human OGT. a, Ribbon representation of the intervening domain rendered in light green with missing loops represented by dotted lines. Lysine side chains that form an extensive positive surface (see **b**) are displayed in a 'ball-and-stick' representation. Shown in coral is a helix from the C-cat domain containing four basic residues that contribute to the positively charged surface[7]. **b**, Surface representation of OGT coloured according to electrostatic potential, with blue representing areas of positive charge and red representing areas of negative charge. The protein is rotated 90° around the *x*-axis from the representation shown in Figs 1, 2 and 3c, exposing the bottom surface of the catalytic region. **c**, Model of full-length human OGT, shown as a surface rendering and coloured as in Fig. 1a, based on the $hOGT_{4.5}$ structures and the previously reported TPR domain structure. The TPRs preceding the boundary of $hOGT_{4.5}$ are shown in light grey. The model is shown as a monomer, but OGT may exist in different oligomerization states in cells[21,28]. Hinge and latch regions are indicated by arrows. **d**, Model of full-length OGT opening to accommodate larger substrates. The 'open' conformation is based on molecular dynamics simulations (Supplementary Movie 1), as described in Methods. (Atomic coordinates for full-length models are available for download; see Supplementary Information.)

involves displacement of the α-UDP group to yield an inverted product. On the basis of the accumulated biochemical and structural data, we propose a general mechanism for the reaction (Fig. 2d).

The most unusual feature of OGT is the intervening domain between the catalytic lobes, which is only found in metazoans (Supplementary Figs 6 and 7). This polypeptide adopts a topologically novel fold with a seven-stranded β-sheet core stabilized by flanking α-helices (Fig. 3a). There are two long unstructured loops for which electron density is missing. An electrostatic surface rendering shows that the intervening domain and an adjacent helix of the C-Cat domain form a large basic surface comprising ten lysine residues (Fig. 3a and b). Among these are K981 and K982, which were previously reported to constitute part of a PIP_3 binding motif that recruits OGT to membranes[7]. We mutated eight of these ten lysines in various combinations (Supplementary Table 3). All mutants were catalytically active (Supplementary Fig. 8), but we were unable to identify a role for the Int-D domain in PIP_3 binding (Supplementary Table 5). We suggest that this domain is involved in other functions *in vivo*. These functions may include substrate selection, cellular localization, or interactions with regulatory factors or receptors. The reported structures and mutant data provide a crucial starting point for investigating the possible roles of the intervening domain.

The structures reported here show how OGT recognizes peptide sequences and provide new information on the enzymatic mechanism as well as a view of the intervening domain. Models of full-length human OGT in its open and closed states, constructed on the basis of crystal structures and molecular dynamics simulations, highlight the conformational changes that may regulate access of substrates to the active site (Fig. 3c and d). Our structures may assist in the development of inhibitors with possible therapeutic value for treating diseases associated with excessive *O*-GlcNAcylation.

METHODS SUMMARY

Human OGT residues 313–1031 (CPTH…KPVE) were expressed in *Escherichia coli* and purified by nickel affinity chromatography and gel filtration. Protein was then incubated with UDP or with UDP and a 17-residue substrate peptide (KKKYPGGSTPVSSANMM), which was cleaved to YPGGSTPVSSANMM in the crystallization drop (confirmed by mass spectrometry). The OGT–UDP structure was determined using the method of multiple isomorphous replacement with anomalous scattering (MIRAS) (Supplementary Table 2). The OGT–UDP–peptide complex structure was solved by molecular replacement using the refined OGT–UDP structure. The crystal packing for the two complexes is described in Supplementary Fig. 9. Kinetic analysis was performed using UDP-^{14}C-GlcNAc and a lysine tagged CKII peptide using our previously described filter binding assay[29]. The molecular dynamics simulation was performed by using the program Desmond[30] on an optimized 64-node Linux-based InfiniBand cluster.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 April; accepted 3 November 2010.
Published online 16 January 2011.

1. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017–1022 (2007).
2. 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).
3. Haltiwanger, R. S., Holt, G. D. & Hart, G. W. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *J. Biol. Chem.* **265,** 2563–2568 (1990).
4. Yang, X., Zhang, F. & Kudlow, J. E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110,** 69–80 (2002).
5. Dias, W. B., Cheung, W. D., Wang, Z. & Hart, G. W. Regulation of calcium/calmodulin-dependent kinase IV by O-GlcNAc modification. *J. Biol. Chem.* **284,** 21327–21337 (2009).
6. Fujiki, R. *et al.* GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. *Nature* **459,** 455–459 (2009).
7. Yang, X. *et al.* Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **451,** 964–969 (2008).
8. Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414,** 813–820 (2001).
9. Caldwell, S. A. *et al.* Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1. *Oncogene* **29,** 2831–2842 (2010).
10. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C. X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **101,** 10804–10809 (2004).
11. 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).
12. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci. STKE* **2005,** re13 (2005).
13. Dentin, R., Hedrick, S., Xie, J., Yates, J. III & Montminy, M. Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* **319,** 1402–1405 (2008).
14. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291,** 2376–2378 (2001).
15. Gambetta, M. C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* **325,** 93–96 (2009).
16. Sinclair, D. A. *et al. Drosophila* O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proc. Natl Acad. Sci. USA* **106,** 13427–13432 (2009).
17. Love, D. C. *et al.* Dynamic O-GlcNAc cycling at promoters of *Caenorhabditis elegans* genes regulating longevity, stress, and immunity. *Proc. Natl Acad. Sci. USA* **107,** 7413–7418 (2010).
18. Goldberg, H. J., Whiteside, C. I., Hart, G. W. & Fantus, I. G. Posttranslational, reversible O-glycosylation is stimulated by high glucose and mediates plasminogen activator inhibitor-1 gene expression and Sp1 transcriptional activity in glomerular mesangial cells. *Endocrinology* **147,** 222–231 (2006).
19. 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).
20. 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).
21. Jinek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nature Struct. Mol. Biol.* **11,** 1001–1007 (2004).
22. Ha, S., Walker, D., Shi, Y. & Walker, S. The 1.9 A crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* **9,** 1045–1052 (2000).
23. Hu, Y. *et al.* Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc. Natl Acad. Sci. USA* **100,** 845–849 (2003).
24. Wrabl, J. O. & Grishin, N. V. Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J. Mol. Biol.* **314,** 365–374 (2001).
25. Martinez-Fleites, C. *et al.* Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nature Struct. Mol. Biol.* **15,** 764–765 (2008).
26. Clarke, A. J. *et al.* Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J.* **27,** 2780–2788 (2008).
27. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
28. Kreppel, L. K. & Hart, G. W. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *J. Biol. Chem.* **274,** 32015–32022 (1999).
29. Gross, B. J., Kraybill, B. C. & Walker, S. Discovery of O-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **127,** 14588–14589 (2005).
30. Bowers, K. J. *et al.* Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. ACM/IEEE Conf. on Supercomputing (SC06)* (ACM Press, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Gross and C. Drennan for advice. We also thank the US National Institutes of Health, the US National Science Foundation, and the Harvard Biomedical Accelerator Fund for financial support. This work is based on research conducted at the Advanced Photon Source (Northeastern Collaborative Access Team beamlines) and Brookhaven National Laboratory (X25 and X29 beamlines).

Author Contributions S.W. conceived the project. M.B.L. obtained the crystallization construct and initial diffracting crystals. M.B.L., Y.N. and P.S. determined and refined the crystal structures. J.J. and M.B.L. performed the enzymatic assays. M.B.L., Y.N., J.J., P.S. and S.W. designed experiments, discussed results, and prepared the manuscript.

Author Information The structures of the OGT–UDP complex and the OGT–UDP–peptide complex have been submitted to the Protein Data Bank under accession numbers 3PE3 and 3PE4. Atomic coordinates for the full-length models of OGT as well as the docked UDP-GlcNAc structure are available for download from the Walker Laboratory website (see Supplementary Information). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.W. (enzymology; suzanne_walker@hms.harvard.edu) or P.S. (structural biology; piotr_sliz@hms.harvard.edu).

METHODS

Protein purification. Full length human OGT (ncOGT) was expressed as previously described. The $OGT_{4.5}$ construct (spanning residues 313–1031 based on the numbering of the full length human protein) was constructed from our previously reported *E. coli* codon-optimized construct using primers listed in Supplementary Table 6 after being cloned into a pET24b vector (Novagen)[29]. After plasmid transformation into BL21 (DE3), the protein was expressed as a fusion protein with an N terminus consisting of a T7 tag, followed by an 8-His tag, followed by an HRV3C protease cleavage site (LEVLFQGP). Cultures were grown at 37 °C after diluting an overnight culture 1 to 100 in fresh LB media. Cells were grown to an A_{600} of 1.1, at which point they were transferred to a temperature of 16 °C. After letting the cells grow at 16 °C for 30 min, they were induced with 0.2 mM IPTG and grown overnight at 16 °C for 16 h. Cells were pelleted, resuspended in TBS (20 mM Tris, pH 7.4, 250 mM NaCl) supplemented with 1 mM PMSF and 0.1 mg ml^{-1} lysozyme, lysed and the lysate was centrifuged at 5,000*g* for 20 min to remove unbroken cells. The supernatant was then centrifuged at 100,000*g* to further clean the lysate. Imidazole was then added to the supernatant to a final concentration of 40 mM before the lysate was incubated with Ni-NTA agarose superflow resin (Qiagen) which was prewashed with TBS + 40 mM imidazole for batch nickel affinity purification. After incubating the lysate and the resin with gentle rocking at 4 °C, the flowthrough was removed and the resin was washed with 10 column volumes of TBS + 50 mM imidazole. The protein was then eluted with 4 column volumes of TBS + 250 mM imidazole. The eluate was supplemented with 0.5 μM THP to prevent aggregation and then concentrated with centrifugal concentrators (Millipore). After protein concentration determination, the N-terminal tags were cleaved by adding HRV 3C protease (EMD) to the concentrated purified protein at a ratio 1 unit/150 μg of protein and incubating at 4 °C for 16 h. Following cleavage, the protein was further purified by gel filtration on a Superdex 200 column (GE Healthcare) in TBS (20 mM Tris, pH 8.0, 150 mM NaCl) + 0.5 μM THP (EMD). The fractions were collected and concentrated using centrifugal concentrators again. The $hOGT_{4.5}$ protein was monomeric in solution, as determined by gel filtration and sedimentation equilibrium analytical ultracentrifugation. The protein was then diluted 1:1 in water before setting up crystals.

Native crystals. All crystals were grown with the hanging drop method at room temperature. For the UDP structure, 7 mg ml^{-1} protein was incubated with 1 mM UDP for several hours at 4 °C. After screening, optimal crystals were obtained when 10 μl of protein was mixed with 5 μl of reservoir solution containing 1.45 M potassium phosphate dibasic, 8 mM EDTA, and 1% xylitol. After several days, hexagonal rod crystals grew, to a maximum size of about 400 × 100 × 100 μm. Crystals were flash frozen using a cryoprotectant consisting of 1.8 M potassium phosphate and 27% xylitol. For the peptide–UDP complex, OGT was incubated with 1 mM UDP and 2 mM CKII3K peptide[28,29] for several hours at 4 °C. Crystals were obtained by mixing 8 μl protein solution with 4 μl reservoir containing 1.6 M Li_2SO_4 and 0.1 M bis-tris propane-HCl pH 7.0 (1,3-bis(tris(hydroxymethyl) methylamino)propane). Trapezoidal crystals appeared after several days. Crystals were frozen in a cryprotectant consisting of 1.72 M Li_2SO_4, 0.05 M Bis Tris Propane, pH 7.0 and 28% xylitol.

Heavy metal soaks. Several heavy metal compounds were screened using the method of ref. 31. After identifying several promising heavy metal compounds, the following conditions gave useful derivatives: K_2PtCl_4, 10 mM, 1 h soak; sodium aurothiomalate, 10 mM, 15 min soak; K_2PtCl_4, 10 mM, 10 min soak; K_2PtBr_4, 1 mM, 1 h soak.

Data collection. All the data were collected at NSLS X29 or X25 at Brookhaven National Laboratory except for the gold derivative, which was collected at ID24C at APS at Argonne National Laboratory. The heavy metal derivatives were collected at the following peak wavelengths: gold at 1.0384 Å and platinum at 1.0715 Å. The UDP structure and all the derivatives belonged to the space group *P*321. The peptide complex crystals were *I*2. All data sets were processed with iMosflm[32] and scaled using SCALA[33].

Structure determination and refinement of the OGT–UDP structure. The structure of the native OGT–UDP complex was determined by using MIRAS with the program SHARP[34]. The native data set and all the heavy atom derivative data sets were processed with iMosflm and Scala. Heavy atom sites in the K_2PtCl_4 1 h soak data set were first determined by using HKL2MAP[35]. SAD phases were then obtained with the CCP4 program Phaser[36] (Experimental Phasing). These initial phases were then used to find the heavy atom sites in the other data sets using the CCP4[37] program FFT. After obtaining all the sites, multiple isomorphous replacement with anomalous scattering (MIRAS) phases to 4.4 Å were obtained using SHARP. The figures of merit at this resolution were 0.46329 (acentric) and 0.47049 (centric). After MIRAS phasing, the map was interpretable, and we confirmed that there were four monomers in the asymmetric unit. Density modification and phase extension to 2.78 Å with NCS averaging were performed using DM, yielding a map with clear side chains. A model was built using as a guide both the structure of the bacterial homologue (using a homology model generated with Swiss Model) and the heavy atom locations. There are two loops in the intervening domain for which there is no electron density, so these residues are omitted from the model. Twelve residues are missing from one loop and four from the other. The structure was refined with CNS[38]. Initial rigid body refinement optimized the placement of the monomers and then the components of each monomer. After several iterative rounds of simulated annealing, individual *B* factor refinement, and manual adjustments using COOT[39], the UDP and waters were added, and the structure refined to an R_{work} of 21% and an R_{free} of 24%. Refinement was completed in Phenix[40] using TLS refinement[41,42], minimization, and individual *B* factor refinement to give a final R_{work} of 18.5% and R_{free} of 21.8%. Figures were prepared using Pymol[43] and CCP4mg[44].

Structure determination and refinement of the OGT–UDP–peptide complex. Data were processed with iMosflm and Scala, and the structure was determined by molecular replacement. The refined OGT–UDP structure described above was used as a search model using the Phaser molecular replacement module[45] in CCP4. Initial molecular replacement efforts showed that whereas the catalytic domain was nearly identical in the UDP and UDP–peptide cocomplexes, the orientation of the TPRs relative to the catalytic domain was noticeably different. Therefore, the model was broken into three parts: the catalytic domain and two sections of the TPR domains. Using this approach, a good map and model were obtained, which confirmed the twofold NCS present in this structure. The peptide was built by hand, as the side chains were already clear enough at this point to place the residues properly. The peptide in the crystal structure was cleaved from KKKYPGGSTPVS SANMM to YPGGSTPVSSANMM, as confirmed by mass spectrometry. The model was then refined with Phenix. As before, repeated rounds of annealing and individual *B* factor refinement were interspersed with manual adjustments in COOT. Waters were then added and sulphate ions were added after refining the waters. The structure was completed with cycles of annealing, minimization, TLS and *B* factor refinement, leading to a final structure with R_{work} of 22.4% and R_{free} of 25.2%. The crystal packing for the two complexes is described in Supplementary Fig. 9.

Kinetics. Mutants were made from the full-length ncOGT using QuickChange mutagenesis and the primers shown in Supplementary Table 6. Kinetic measurements were performed using a previously described filter binding assay[29]. Briefly, reaction mixtures containing 500 μM CKII3K peptide (KKKYPGGSTPVSSANMM), 6 μM UDP-^{14}C-GlcNAc (300 mCi $mmol^{-1}$ specific activity, American Radiochemicals), 100 nM OGT (WT or mutant protein), and buffer (125 mM NaCl, 1 mM EDTA, 20 mM potassium phosphate, pH 7.4, and 500 μM tris(hydroxypropyl)phosphine) were incubated at room temperature for 30 min. Reactions were then quenched by spotting onto the Whatman P81 phosphocellulose disks, washed three times for five minutes in 0.5% phosphoric acid, and counted by liquid scintillation counting. Reactions proceeded to <10% conversion under these conditions. Positive and negative controls were conducted similarly without enzyme, and positive controls were detected by liquid scintillation counting without the phosphoric acid wash step. Data were analysed based on triplicate experiments. For product inhibition experiments, substrate concentrations were used as described in Supplementary Fig. 2. Reactions were allowed to proceed for either 30 min or 60 min and performed in triplicate and analysed with linear regression using GraphPad Prism5.

Model preparation. The $hOGT_{4.5}$ construct contains the residues 313–1031 (CPTH…KPVE) of the full-length ncOGT protein. Because the first two TPR units of $hOGT_{4.5}$ overlap with the last two TPR units of the previously crystallized human TPR domain (PDB code 1W3B)[21], we superimposed each of the $hOGT_{4.5}$ structures (PDB codes 3PE3 and 3PE4) with the TPR domain to create composite models of full length human OGT. Coordinates are provided in Supplementary Data Files 1 and 2.

Molecular dynamics. The coordinates of the OGT–UDP–peptide complex were optimized in the Protein Preparation Wizard (Schrodinger 2009) where hydrogens were added; water molecules, UDP and peptide were stripped; and the structure was minimized using the OPLS2001 forcefield. The 1-μm simulation used the CHARM27 forcefield[46], and the simple point charge model for water[47]. The CHARM27 forcefield was applied to the system using the VIPARR utility. The default Desmond relaxation was performed before simulation, and molecular dynamics were run at constant temperature (300 K) and pressure (1 bar). The simulation was performed by using the program Desmond, version 2.2.9.1.0[30] compiled by SBGrid on an optimized 64-node Linux-based InfiniBand cluster and took 75 days to complete. Molecular dynamics trajectories were processed and animated with VMD[48].

Lipid (PIP) binding assays. Recombinant OGT constructs (His- or GST-tagged full-length human OGT) were overexpressed in *E. coli* and purified by affinity chromatography, using agarose beads conjugated to nickel or glutathione, respectively. PIP binding assays were performed using PIP Strips (Echelon Biosciences). Each membrane was pre-incubated for 2 h at room temperature with a blocking solution

containing 0.1% ovalbumin (for GST fusion constructs) or 3% fatty acid free BSA (for His-tagged constructs) in buffer TBST (20 mM Tris pH 8.0, 50 mM NaCl, 0.1% Tween 20). Purified OGT proteins resuspended in TBST at various concentrations (0.2–2 μM) were applied to each membrane. Washing and developing steps were performed as outlined in the manufacturer's protocols, using the same TBST described above, and protein was detected using either anti-His or anti-GST antibodies and HRP-conjugated secondary antibodies.

31. Boggon, T. J. & Shapiro, L. Screening for phasing atoms in protein crystallography. *Structure* **8,** R143–R149 (2000).
32. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* **26,** 27–33 (1992).
33. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
34. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494 (1997).
35. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37,** 843–844 (2004).
36. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
37. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
38. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
40. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
41. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39,** 109–111 (2006).
42. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439–450 (2006).
43. DeLano, W. L. The Pymol Molecular Graphics System. (Delano Scientific, San Carlos, CA, 2002).
44. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60,** 2288–2294 (2004).
45. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41 (2007).
46. Mackerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).
47. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 331–342 (Reidel, 1981).
48. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 27–28, 33–38 (1996).