Published in final edited form as:
Nat Struct Mol Biol. 2015 September ; 22(9): 744–750. doi:10.1038/nsmb.3063.

The active site of O-GlcNAc transferase imposes constraints on substrate sequence

Shalini Pathak[#a], **Jana Alonso**[#a], **Marianne Schimpl**[#a], **Karim Rafie**[a], **David E. Blair**[a], **Vladimir S. Borodkin**[a], **Osama Albarbarawi**[a], and **Daan M. F. van Aalten**[a,b,2]

[a]MRC Protein Phosphorylation and Ubiquitylation Unit and College of Life Sciences, University of Dundee, Dundee, UK

[b]Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee, UK

[#] These authors contributed equally to this work.

Abstract

O-GlcNAc transferase (OGT) glycosylates a diverse range of intracellular proteins with O-linked N-acetylglucosamine (O-GlcNAc), an essential and dynamic post-translational modification in metazoa. Although this enzyme modifies hundreds of proteins with O-GlcNAc, it is not understood how OGT achieves substrate specificity. In this study, we describe the application of a high-throughput OGT assay on a library of peptides. The sites of O-GlcNAc modification were mapped by ETD-mass spectrometry, and found to correlate with previously detected O-GlcNAc sites. Crystal structures of four acceptor peptides in complex with human OGT suggest that a combination of size and conformational restriction defines sequence specificity in the −3 to +2 subsites. This work reveals that while the N-terminal TPR repeats of hOGT may play a role in substrate recognition, the sequence restriction imposed by the peptide-binding site makes a significant contribution to O-GlcNAc site specificity.

Introduction

O-linked β-*N*-acetylglucosamine (O-GlcNAc) is an abundant and essential post-translational modification of intracellular proteins in metazoa and plants[1,2]. Hundreds of O-GlcNAc-modified nuclear and cytoplasmic proteins have been identified. O-GlcNAc proteins fall into a variety of functional classes including transcription factors, ribosomal proteins, translational factors, signalling proteins, cytoskeletal proteins and components of the nuclear pore complex as reviewed previously[3]. Like protein phosphorylation, O-GlcNAcylation is a rapidly reversible post-translational modification[4]. In recent years O-GlcNAc has been implicated in a multitude of cellular functions such as regulation of gene expression[5–7], circadian rhythm[8,9], vesicle trafficking and protein localization[10,11] and signalling[12–14]. Remarkably, unlike phosphorylation, which is orchestrated by over 600 Ser/Thr kinases and

[2]Correspondence to: dmfvanaalten@dundee.ac.uk.

Contributions

S.P., D.M.F.v.A conceived the study. S.P., D.E.B., K.R. performed the peptide assays; M.S., K.R. and D.M.F.v.A. performed structural biology; V.S.B. performed peptide synthesis; J.A.L., and O.A., performed MS studies; S.P., M.S., and D.M.F.v.A. interpreted the data and wrote the manuscript.

Europe PMC Funders Author Manuscripts

phosphatases in metazoa15, the O-GlcNAc modification is regulated by only two opposing enzymes. The enzyme catalyzing addition of O-GlcNAc onto proteins is O-GlcNAc transferase (OGT)16,17. The enzyme responsible for catalyzing cleavage of O-GlcNAc from modified proteins is a glycoside hydrolase known as O-GlcNAcase (OGA)18,19. In addition to the catalytic domain, OGT has an N-terminal domain comprising 13.5 tetratricopeptide repeats (TPRs) arranged in a superhelical spiral abutting the active site20–23 (Fig. 1a). This TPR domain is believed to mediate protein-protein interactions and is required for recognition of certain protein substrates24,25. However, recent reports of crystal structures of hOGT in complex with substrate peptides have shown these to bind the active site in an extended, but ordered fashion, suggesting that the enzyme my recognise features proximal to the O-GlcNAc site23,26,27.

A number of proteomics studies have been conducted to identify OGT substrates28–33. These studies exploit various new methods of enriching and labelling O-GlcNAc peptides derived from proteins in cell lysates/tissues for site mapping through mass spectrometry. On compiling the identified O-GlcNAc sites, certain sequence motifs have been observed flanking the O-GlcNAc modification32,34–36. It is not clear whether such motifs are restricted to the limited subset of proteins in the cell recognized by the OGT TPR repeats and/or whether they are imposed by restrictions on peptide binding in the OGT active site. Here, we aim to dissect the specificity of the hOGT active site using a substrate library of synthetic peptides. This approach defines the preferred hOGT peptide sequon as [TS][PT][VT]**S/T**[RLV][ASY] with several hits from this screen emerging as previously validated O-GlcNAc sites on proteins *in vivo/in vitro*. We report crystal structures of hOGT with four acceptor peptides, revealing conserved peptide conformations in the (−3 to +2) subsites, despite differences in amino acid sequence and a paucity of enzyme-substrate interactions of a sequence-specific nature.

Results

O-GlcNAc transferase possesses acceptor peptide specificity

To gain insight into possible sequence specificity, hOGT activity was measured against a library of 720 biotinylated 13-amino acid peptides derived from the human proteome37 (Table S1). We developed a high throughput-compatible OGT scintillation proximity assay, using UDP-[^{3}H]GlcNAc, and included an α-crystallin-derived peptide (NH_2-AIPVSREEK-(biotin)-COOH, CRYA1) as reference standard. The reaction product, radiolabelled O-GlcNAcylated peptide, was captured on Streptavidin-coated FlashPlates allowing for direct quantification on a scintillation counter, without the need for removing excess substrate. The averages of two independent screens of the peptide library are depicted as a heat map in Fig. 1b. The CRYA1 reference peptide is known to be a relatively poor OGT substrate38. Despite all peptides containing at least one potential Ser/Thr O-GlcNAc site (and on average 2.8), only 70 out of 720 peptides produced a radiometric signal greater than that obtained for the CRYA1 reference peptide. This implied that OGT has a significant level of substrate specificity dictated by the peptide binding site alone. However, almost all of the 70 top hits contained more than one Ser/Thr, preventing direct interrogation of sequence conservation around the O-GlcNAc acceptor site in absence of site mapping data.

O-GlcNAcylation of peptide library substrates is limited to specific positions

To allow for mass spectrometric mapping of the specific site(s) of O-GlcNAc modification on the best substrates from the peptide library screen, we resynthesized these peptides without the biotin tag (fragmentation products of biotin would otherwise impede the interpretation of mass spectrometry (MS) data). These peptides were enzymatically O-GlcNAcylated *in vitro* and analysed by Electron Transfer Dissociation (ETD) mass spectrometry. The ETD-MS/MS of the peptides generated fragmentation patterns covering the majority of *c*- and *z*-type ions, allowing the precise mapping of the O-GlcNAc sites. For instance, the ETD-MS/MS spectrum of the *in vitro* O-GlcNAcylated synthetic peptide "KENSPAVTPV**S**TA", the top hit from the screen (matching a peptide from the protein retinoblastoma-like protein 2, RBL2) is shown in Fig. 2a, with a strong c_{10}^{1+} peak corresponding to a serine + 203 Da, representing the sugar moiety. An expectation value below 0.1 was specified for each peptide fragmentation to ensure reliable designation of the O-GlcNAc sites (Fig. 2b, fragmentation spectra for all peptides are shown in Supplementary Fig. S1).

OGT peptide substrates are predictive for O-GlcNAc proteins

For those peptides shown to be good OGT substrates, we proceeded to investigate whether the corresponding proteins have previously been reported to be O-GlcNAc modified. The transcription factor FOXO1 has been reported to be O-GlcNAcylated at S31939, in agreement with our ETD-MS/MS data from the peptide alone (Fig. 2b). Similarly, a tryptic peptide from insulin receptor substrate-1 (IRS1) spanning residues 981-998 has been observed to bear an O-GlcNAc-modification on either Ser984 or Ser98540,41. Using ETD-MS/MS, we observed that both Ser984 and Ser985 were O-GlcNAc modified (Fig. 2b, Fig. S1). Interestingly, we also identified Ser400 as O-GlcNAc modified on a peptide derived from Tau; the same amino acid was recently published as an O-GlcNAc site42. Five of the O-GlcNAc peptides (RBL2, α-crystallin B chain, GSK3β Lamin A and Hsp27) identified from our screen are derived from proteins that have been reported to be O-GlcNAc proteins with either unknown or non-matching O-GlcNAc sites (Fig. 2b, Fig. S1)11,43–46.

Different substrate peptides bind OGT with a common conformation

Several studies have reported crystal structures of hOGT in complex with substrate peptides and the donor analog UDP-5S-GlcNAc47; notably a 14-aa peptide containing the O-GlcNAc site on Ser347 of casein kinase 2α (CK2)23,48, a 13-aa peptide based on Ser395 from the TAK1-binding protein TAB126 and recently a 26-aa peptide derived from the Host cell factor 1 (HCF-1)27 (although the fate of the latter is believed to be proteolytic cleavage by OGT rather than glycosylation). These studies have given the first insights into how hOGT substrates interact with the active site. It was noted that the peptides bind the active site of OGT in the same orientation and with similar, extended conformations. To explore possible links between peptide sequence and binding modes, we have determined the crystal structures of four peptide hits from the screen in complex with hOGT and UDP-5S-GlcNAc. Complexes with the peptides derived from Retinoblastoma-like protein 2 ($RBL2_{411—422}$ KENPAVTPV**S**TA), Proto-oncogene tyrosine-protein kinase receptor Ret ($Ret_{660—672}$ AQAFPVSY**S**SSGA), Keratin-7 ($KER7_{7—19}$ SPVFT**S**RSAAFSC) and Lamin B1

(LAMIN$_{179—191}$ KLSPSPSSRV**T**VS) were obtained by soaking hOGT crystals and refinement against synchrotron diffraction data (Fig. 3, Supplementary Fig. S2 and Table S2). Despite these peptides containing multiple serines and threonines, in all cases the position of target serine/threonine in the enzyme active site was in agreement with the site-mapping results obtained through mass spectrometry. As observed in the previously published hOGT-peptide complexes, the additional complexes reported here show peptides binding the active site in the same orientation and in an extended conformation. A comparison of all six complexes highlights several interesting common features (Fig. 3). Strikingly, all structures reveal a conserved backbone conformation of the peptides in the −3 to +2 subsites, with backbone torsion angles characteristic of the extended conformation observed in β-strands (−160° < phi < −50° and 100° < psi < 180°) (Fig. 3d). Beyond this −3 to +2 region, the peptide conformations diverge (Fig. 3a).

OGT is known to utilize an ordered bi-bi catalytic mechanism, where the donor substrate binds the enzyme first, and contributes to creating the binding site for the incoming acceptor peptide. UDP-GlcNAc interacts primarily with the C-terminal lobe of the catalytic domain (C-Cat), leaving part of the sugar nucleotide molecular surface exposed to the solvent (Fig. 3c). This area is subsequently covered by the peptide substrate binding in subsites −3 to 0. The side chain in the −3 subsite occupies a shallow pocket flanked by hydrophobic amino acids from the C-Cat domain and the uracil moiety of UDP-GlcNAc (Fig. 3b). Notably, the six OGT-peptide complexes contain isosteric amino acids, either a Val or a Thr, in this position, which fill the shallow pocket and form van der Waals interactions with the uracil and surrounding enzyme side chains. Significantly larger side chains would be difficult to accommodate in this position without some change in the peptide backbone conformation. The importance of these interactions and size restrictions was investigated by determining the effect of single amino acid substitutions on OGT activity (Figs. 4c, S4). Both the change to a very small (Ala) or a bulky (Phe) amino acid in −3 results in < 30% residual activity, thus corroborating the structural insights. The −2 subsite is occupied by a Pro in three out of the six structures. The conformational rigidity of Pro appears to be favourable in stabilising the extended conformation of the peptide. Nevertheless, larger side chains, such as the Phe and Arg are also accommodated, but project away from the enzyme and towards the solvent (Fig. 3a). Substrate peptides with very small side chains in the -2 subsite were disfavoured by OGT; Ala reduces the activity to 65%, but the effect was particularly pronounced for Gly (< 20%), which offers the least conformational rigidity (Fig. 4c). The −1 subsite appears to tolerate large side chains, such as the Tyr observed in Ret and TAB1, without undue distortion (the rotamer adopted by the Tyr side chain is observed in 13 % of tyrosines in high-resolution crystal structures of proteins), although OGT activity was higher on peptides with Val in the −1 position (100 %) than Tyr (42 %) or indeed Ala (21 %) (Fig. 4c).

The acceptor Ser or Thr in the 0 subsite is necessarily the position in the peptide with the least degree of positional and conformational flexibility. Both the backbone amide and the side chain hydroxyl participate in hydrogen bonds with the donor substrate (Fig. 3b), as has been discussed in detail in recent works investigating the catalytic mechanism of OGT26,48,49. The complex with the Lamin B1-derived peptide represents the first reported

OGT substrate complexes with a Thr acceptor. Interestingly, we observed a 50 % reduction of OGT activity upon substitution of Ser with Thr (Fig. 4c).

While substrate binding in the −3 to −1 subsites relies primarily on van der Waals interactions, the amino acids in the 0 to +2 subsites participate in hydrogen bonds through the peptide backbones. The carbonyl in the +1 position interacts with the Leu634 backbone amide, and the backbone amide in the +3 subsite interacts with the backbone of Tyr632, in a hydrogen-bonding pattern reminiscent of an antiparallel β-sheet (Fig. 3b). In the C-terminal part of the substrate peptides, in particular in subsites +1 and +3/4/5, the size and nature of side chains appears to be subject to fewer restrictions and greater flexibility (Fig. 3a). OGT activity is virtually unaffected by a Arg-to-Ala substitution in the +1 position (Fig. 4c). Sub-site +2, however, favours small side chains and is occupied by Ala or Ser in all five complexes and substitution by Phe reduces OGT activity to < 25% (Fig. 4c). Beyond the +3 subsite, most of the C-termini of substrate peptides are disordered in the crystal structures (Fig. 3a and Fig. S2).

Overall, the paucity of interactions involving side chains of the substrates is consistent with a degree of substrate tolerance of OGT. Analysis of the structures and OGT activity primarily suggests a preference for small amino acids in certain positions, most notably the −3 and also the +2 subsite. Where structural information is available beyond the +3 subsite, the C-terminal ends of the peptides appear to be in direct contact with the nearest TPR repeats (Fig. 3c). Indeed, in the recently reported structure of OGT in complex with the 26-amino acid HCF-1 peptide, the (proteolysis) substrate extends significantly further into the TPR domain, anchored by numerous interactions of a sequence-specific nature with the concave surface of the spiral formed by the TPRs27 (Fig. 1a). Whether such a recognition mode is generally employed by OGT for glycosylation of substrates remains to be established. The TPR domain partly occludes access to the active site (Fig. 3c) in all currently reported crystal structures of O-GlcNAc transferases, which is irreconcilable with the recognition and modification of amino acids in compact domains or secondary structure elements of substrate proteins. Opening of the active site groove through a hinge-like movement of the TPR repeats has been suggested based on molecular dynamics calculations23, but the available structural evidence points to OGT acting on unfolded regions of its substrates.

Sequence comparison of peptide library hits reveals a degenerate hexapeptide sequon

Guided by the structures of five different peptide substrates sharing the same binding mode in the −3 to +2 sub-sites, we analysed the best OGT substrates from the peptide library for sequence conservation in these positions. The corresponding hexapeptide sequences were extracted (Fig. S3) and a sequence logo was generated with the WebLogo server50 (Fig. 4b). In order to control for any compositional bias in the library, we also extracted all possible hexapeptide sequences with a Ser or Thr in the fourth position from the library, and subjected them to the same analysis (Fig. 4a), revealing a mostly random sampling of hexapeptide sequences in the library as a whole versus significant enrichment for particular amino acids in the O-GlcNAc sequence logo. Despite the limited sample size for the O-GlcNAc peptides (32 peptides) a number of interesting features emerge from inspection of

the logo (Fig. 4b). Most striking is the preference for a Pro residue in the −2 position; which is in agreement with similar observations arising from the comparison of O-GlcNAc sites determined on endogenous proteins28,34. This preference is also explained by the structural data (Fig. 3), since Pro would encourage the extended conformation observed in peptides bound in the OGT active site. Proline is the most sterically restricted of all the proteinogenic amino acids; the preferred backbone torsion angles of Pro, and the residue preceding proline (pre-Pro), coincide with the conformations observed for the amino acids in subsites −2 and −3, respectively (Fig. 3d). The stabilising action of Pro is particularly favorable in this position (−2/−3), where no hydrogen bonding interactions with the enzyme are possible. Conversely, in the +1 and +2 subsites, such backbone hydrogen bonds are in place (Fig. 3b), thus placing fewer restrictions on the sequence of substrate peptides, in particular for position +1 (Fig. 4b,c). Position +2 has previously been described as the smallest sub-site51. This is consistent with our structural analyses and OGT activity (Fig. 4c), and is also reflected in the sequence logo by a bias towards the smaller amino acids in this location (Fig. 4b). Indeed, the sequence logo suggests the degenerate OGT recognition sequence [TS][PT][VT]**S/T**[RLV][ASY] using an 0.5 probability cutoff (Fig. 4).

Discussion

One of the central questions about O-GlcNAc as a signalling modification is how a single OGT enzyme recognises and site-specifically modifies a plethora of substrates. OGT is a large enzyme, comprising over 1000 amino acids, and the catalytic domain accounts for less than half of its molecular weight. The main additional domain, consisting of helical TPR repeats that assemble into a super-spiral with two complete turns, has been proposed to play a role in substrate recruitment20. TPRs are frequently involved in mediating protein-protein interactions and the proposed role of the OGT TPRs in substrate recognition is based on studies showing differential activity of TPR truncations towards different protein substrates *in vitro*21,52. Considering that O-GlcNAc is believed to occur post-translationally and the acceptor substrates are frequently very large proteins, the OGT active site appears to be too small to accommodate such substrates in their entirety. OGT structures are available from several species in a range of crystal forms, all showing the TPR domain assuming the same conformation21–23. Moreover, this conformation is clearly catalytically competent, with two reports of glycosylation *in cristallo*23,26. This leaves three possible scenarios relying on a significant flexibility on the part of either the enzyme, or the substrates. 1) A hinge-like motion of the TPR domain has been proposed23 to expose the active site, and give access to larger substrates. Molecular dynamics calculations were used to support such an event, whereas there is currently no experimental data supporting such dramatic conformational changes of the OGT enzyme. 2) OGT acts on flexible regions in proteins, such as N- and C-terminal tails and interdomain regions; or binding to the enzyme induces a partial unfolding of the substrate that would allow the modification site to interact with the OGT active site in an extended conformation (Fig. 1a). In such a recognition mode, selective interactions with the OGT active site would be of primary importance. 3) Glycosylation by OGT may occur co-translationally for a subset of proteins; such O-GlcNAc sites could be found in less accessible regions of the protein, which could explain certain reported inaccessible O-GlcNAc sites such as on histone 2B53, α-crystallin (T170)54. Such sites are not, however,

likely to be dynamic. A recent study by Zhu *et al.* showed that co-translational O-GlcNAcylation occurs on Sp1 and Nup62, and appears to act as a quality control mechanism during translation55.

Perhaps none of these scenarios alone offers a satisfactory explanation for all reported O-GlcNAc sites, and it is possible that different subsets of substrates are targeted by one or more of these recognition modes. While there is an abundance of O-GlcNAc sites reported by proteomics studies28–34, very few of these have been thoroughly verified by mutagenesis and/or site-specific antibodies. Furthermore, OGT inhibitors were not used during the cell lysis step in these studies, and it is thus conceivable that OGT could modify non-physiological substrates during lysis and post-processing. The occasional reports of O-GlcNAc proteins from other subcellular compartments lend weight to such concerns. Such non-natural sites would seriously compromise the value of the collective information, and dilute efforts to extract generalised insights into physiological O-GlcNAc sites on proteins from MS data alone. Controlled *in vitro* experiments with peptide libraries thus offer a useful alternative approach to exploring OGT substrate specificity. Previous studies have generally used custom-made, small peptide panels. Haltiwanger *et al.* described the first peptide substrate for OGT, a synthetic peptide derived from three known sites16. They found that a proline was essential, although they could not site-map the modified sites and were thus unable to determine the proline position. Leavy and Bertozzi used a small library of peptides derived from α-crystallin, exploiting a novel assay using detection via ‘azido-ELISA’56. They show that the peptide can be improved in many places, suggesting an advantage of having a Pro in +2 (but not in −2). Finally, Liu *et al.* generated a peptide mutagenesis approach to explore motifs previously reported by MS studies suggesting preference for certain amino acids close to the O-GlcNAc and noting that sites predominantly occur in random coil regions57. Currently available O-GlcNAc site prediction servers – YinOYang58, dbOGAP59 and the more recently developed O-GlcNAcPRED60 are restricted to amino acid sequence and do not apply structural restraints around the modification site.

The work described here employs a large and diverse set of peptides derived from the human proteome to explore the specificity of OGT towards peptides independent of the role that the TPR domain may play in recognition of large protein substrates. Co-crystallisation of identified acceptor peptides with OGT showed that the peptide is tethered into a common binding mode by a combination of van der Waals interactions and hydrogen bonds that restrict torsional freedom in the −3 to +2 subsites only. These structural constraints together with ETD-MS/MS data then allowed identification of a degenerate sequon that is remarkably similar to what has been found for O-GlcNAc sites on proteins34. This is a significant advance in our understanding of how OGT selects its protein substrates as it suggests that constraints in the active site alone, independent of the TPRs, to a significant extent dictate OGT substrate preference. Future experiments could use a similar approach to address the effect of phosphorylation on OGT substrate recognition, to address the proposed interplay between the two regulatory post-translational modifications.

Methods

Peptide library screening

Peptide library was obtained from JPT Peptide Technologies GmbH37, comprising a library of 720 biotinylated peptide kinase substrates consisting of two 384 polypropylene plates each containing 360 peptides (0.25 nmol each in rows A1 to O24). Peptides were solubilised in 62.5 μl of 50 mM Tris-HCl (pH 7.5) to give a final concentration of 4 μM using a PlateMate Plus liquid handling robot (Thermo Scientific Matrix). A solution of either control (no protein) or $hOGT_{313-1046}$ protein (300 nM, from Glycobiochem UK) and UDP-GlcNAc substrate stock (1 μM), consisting of 910 nM cold UDP-GlcNAc (Sigma) and 90 nM UDP-[^{3}H]GlcNAc radioactive tracer (0.4 Ci/ mmol)(American Radiolabeled Chemicals, Inc.) was made up in buffer consisting of 50 mM Tris pH 7.5, 2 mM DTT and 0.1 mg/ml bovine serum albumin (BSA). The peptide screen was carried out in duplicate and consisted of 1 quality control plate containing: High controls, mutant α-A crystallin (CRYA1M)(NH_2 -AIPVSR**A**EK (biotin)-COOH) in columns 1–8, 17-24 and native α-A crystallin (CRYA1) (NH_2 -AIPVSREEK(biotin)-COOH in columns 9–16. The screening plates consisted of library peptides in rows A1–O24, with low controls in P1–P24. 10 μl of 4 μM peptide was added to all 384 polypropylene U bottom plates using a PlateMate Plus liquid handling robot as outlined above. The reaction was then initiated by the addition of 10 μl of control (no OGT / UDP-GlcNAc) to both QC plate (columns 9-16) and assay plates P13-P24 to provide low controls. All other positions received OGT / UDP-GlcNAc. Plates were incubated at room temperature on a micro titre plate shaker (Heidolph) for 4 h. The reaction was stopped by the addition of 40 μl of 100 μM phosphoric acid (pH 4) / 750 μM $MgCl_2$ using a Wellmate liquid handling robot (Thermo Scientific Matrix). The stopped reactions were incubated for a further 10 min at room temperature. 56 μl from the QC and screening plates was transferred to 384 well Streptavidin FlashPlates (PerkinElmer) using a PlateMate plus liquid handling robot. Plates were then sealed and incubated overnight at room temperature before determining scintillation counts using a Top Count plate reader (PerkinElmer).

The quality of the resultant screening data was assessed by coefficient of variance (CV), robust Z prime and SSMD calculations carried out on the QC plate and the high and low controls included on each of the screening replicates. Scintillation counts generated from the screen were converted into relative activity using the CRYA1 peptide as standard.

Peptide synthesis

Peptides were synthesized on microwave assisted CEM Liberty instrument using standard Fmoc protocol on Rink amide low load MBHA resin (0.24 mmol/g) (Novabiochem, Germany). Coupling steps were performed with 2.5 eq of amino acid in respect to the resin loading capacity, PyBOP (2.5 eq) and DIPEA (5 eq) at 70 °C for 5 min, except for Ser and His that were coupled at 50 °C for 10 min. Deprotection was carried out with 20 % piperidine/DMF at 70 °C for 0.5 min and then for 3 min with the fresh portion of the deprotection mixture. Peptides were manually cleaved from the resin with TFA:H_2O:TIPS:DODT 1:0.05:0.02:0.05 mixture (2 mL) for 2 h at RT. In the case of Met containing peptides the cleavage cocktail was supplemented with NH_4I (0.05 g) and Me_2S (0.1 mL). The peptides were precipitated with cold (0 °C) diethyl ether (40 mL) and

collected by centrifugation at 4000 *g*, at 4 °C for 15 min. The pellet was re-suspended in the cold ether and centrifuged again; this step was repeated twice. The peptides were dried with argon flow. The crude peptides were purified by HPLC on Gilson instrument using Waters Xbridge 19×100 mm Peptide separation technology column (flow rate 25 mL/min) with a linear gradient 5-95 % of MeCN in 0.1% TFA in water over 10 min. The appropriate fractions were pooled and concentrated *in vacuo*. Finally, the residue was re-dissolved in 20% AcOH (3 mL) and freeze-dried. The identity of the synthetic peptides was confirmed by LC-MS on Bruker microTOF instrument.

hOGT activity measurement

hOGT activity was determined in reactions containing 50 nM His_6-$hOGT_{312-1031}$ or ncOGT in 50 mM Tris-HCl pH 7.5, 0.1 mg/mL BSA, 10 μM sodium dithionate and 0.5 mM of peptide in a total volume of 100 μL. Reaction mixtures were pre-incubated for 15 min and initiated by addition of UDP-GlcNAc to a final concentration of 50 μM. The reaction was stopped after 30 min (short construct) or 4 h (ncOGT) at 21 °C by addition of 200 μL of 75 μM pyrocatechol violet in 25 mM HEPES pH 7.4, 10 mM NaCl, 50% (v/v) MeOH and 15 μM fluorophore, a UDP-sensitive xanthene-based Zn(II) complex described in 2–4. UDP formation was detected fluorimetrically on a Gemini EM fluorescence Microplate reader (Molecular Devices) at excitation and emission wavelengths of 485 nm and 530 nm, respectively. Turnover did not exceed 10 % for either substrate. Data are presented as averages of three measurements, with error bars showing the standard error of the mean.

In vitro O-GlcNAcylation of peptides for mass spectrometry

The re-synthesized peptides were O-GlcNAcylated in a final volume of 20 μl with the final concentrations of each component as peptide (2 μM), OGT enzyme (2 μM), UDP-GlcNAc (4 mM) and reaction buffer (25 mM ammonium bicarbonate (Sigma Aldrich LC-MS grade) and 1 mM DTT. The reaction mixture was incubated for 3 h at 30°C. For LC-MS analysis excess UDP-GlcNAc and OGT were removed by using NanoSep centrifugal columns (molecular weight cutoff 10 kDa, PALL Life Sciences). The resulting filtrate was dried in a SpeedVac (Thermo Fisher Scientific) and stored at −20 °C until further analysis.

O-GlcNAcylation Analysis by LC-MS/MS

Identification of the O-GlcNAc sites was performed by ESI-IT-ETD (ElectroSpray IonTrap Electron Transfer Dissociation) mass spectrometry coupled to a nano-LC system (Ultimate 3000 RSLC, Dionex, Netherlands). Dried peptides were resuspended in 100 μL of 0.5% HCOOH and 10 μL were injected for mass spectrometric analysis. *In vitro* O-GlcNAcylated peptides were concentrated on a trap column (2 cm x 100 μm, Dionex) at 10 μL/min and separated on a 15 cm x 75 μm Pepmap C_{18} reversed-phase column (Thermo Fischer Scientific). Peptides were eluted by a linear 60 min gradient of 95% A / 5% B to 90% B (A: H_2O, 0.1% HCOOH; B; 80% acetonitrile (ACN), 0.08% HCOOH) at 300 nl/min into a LTQ Velos ETD (Thermo Fisher Scientific). MS spectra were acquired in positive mode, firstly MS full scans were acquired followed by MS/MS in ETD mode. Up to five most intense precursors were selected for ETD fragmentation with an activation time of 300 ms and non-dynamic exclusion.

Proteome Discoverer v1.4 software (Thermo Scientific) was used to process raw LC-MS/MS data, applying the Mascot (version 2.3.02, Matrix Science, Boston, MA, USA) search engine algorithm against an in-house established peptides database with the following Mascot parameters: 2+, 3+, 4+ and 5+ ions; precursor mass tolerance 100 ppm; Da; fragment tolerance 0.6 Da and no missed cleavages. The variable modifications included were: Oxidation (M) (15.99 Da), Dioxidation (M) (31.98 Da) and HexNAc (ST) (+203.0794 Da). All MS/MS data and database results were manually inspected in detail to check the accurate assignment of fragment ions using the above software. Peptides with an expectation value (Exp Value) smaller than 0.1 are considered as a precise O-GlcNAc site assignment.

Crystallography

Human OGT (amino acids 312—1031) was recombinantly expressed as a cleavable GST-fusion in *E. coli* BL21(DE3)pLysS and purified and co-crystallised with the substrate analogue UDP-5S-GlcNAc47 as described previously26 for the complexes with peptides derived from Keratin-7 (SPVFTSRSAAFSC) and Ret (AQAFPVSYSSSGA). Crystals were soaked with excess of synthetic peptide for 20—60 min, and cryo-protected in a saturated lithium sulphate solution prior to flash-freezing in liquid nitrogen. For the complex with the RBL2 peptide (KENPAVTPVSTA), the reservoir solution was supplemented with 0.5 M $(NH_4)_2SO_4$, and hOGT was co-crystallised with 2.5 mM UDP-5S-GlcNAc and 2.5 mM peptide and crystal seeds. Seeds were generated from hOGT crystals grown in 1.3 M DL-Malic acid, 0.1 M Bis-Tris propane pH 6.4. Crystals were cryoprotected by a 2 s immersion in 2.5 M sodium malonate pH 7 before flash-freezing in liquid nitrogen. For the complex with the LaminB1 peptide (KLSPSPSSRVTVS), hOGT was crystallised in 1.3 M DL-Malic acid pH 6.4, 0.1 M Bis-Tris propane pH 6.4 and crystal seeds. Seeds were generated from hOGT crystals grown in 1.3 M DL-Malic acid, 0.1 M Bis-Tris propane pH 6.4. Crystals were soaked for 60 minutes with excess peptide and 2.5 mM UDP-5S-GlcNAc. Crystals were cryoprotected by a 2 s immersion in 2.5 M DL-Malic acid pH 7 before flash-freezing in liquid nitrogen. Data were collected at the European Synchrotron Radiation Facility (ESRF) beamlines ID23-2 and ID30A-3 and on Diamond Light Source beamline I04-1, and were processed using MOSFLM61, XDS62 and DENZO63. 2% (Ret, Keratin-7) or 5% (RBL2, Lamin) of all reflections were set aside as an R_{free} test set. Structures were solved with MOLREP64 using one molecule of PDB entry 3PE4 as a search model. Crystals belonging to *F*222 (RBL2, LaminB) have 1 molecule/asymmetric unit; whereas crystals belonging to space group *P*321 (Keratin-7, Ret) have 4 molecules/asymmetric unit, so non-crystallographic symmetry (NCS) restraints were imposed during refinement with REFMAC564, and four-fold NCS averaging was used to improve the maps for model-building in COOT65. Ligand topology for UDP-5S-GlcNAc was created using PRODRG66, and donor and acceptor substrates were manually placed. Supplementary Table S2 gives a summary of data collection and refinement statistics, and Fig. S2 shows the unbiased difference electron density prior to modelling of the ligands.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank ESRF and Diamond synchrotrons for beam time and assistance. This work was funded by a Wellcome Trust Senior Research Fellowship (WT087590MA) to DVA. X-ray diffraction data and refined structures have been deposited in the PDB (entries 4XI9, 4XIF, 4XIE, 5BNW).

References

1. Lubas WA, Frank DW, Krause M, Hanover JA. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. J Biol Chem. 1997; 272:9316–24. [PubMed: 9083068]
2. O'Donnell N, Zachara NE, Hart GW, Marth JD. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. Mol Cell Biol. 2004; 24:1680–90. [PubMed: 14749383]
3. Hart GW, Housley MP, Slawson C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. Nature. 2007; 446:1017–22. [PubMed: 17460662]
4. Hart GW, et al. O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization. Glycobiology. 1996; 6:711–6. [PubMed: 8953283]
5. Fujiki R, et al. GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. Nature. 2009; 459:455–9. [PubMed: 19377461]
6. Hanover JA, Krause MW, Love DC. Post-translational modifications: Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. Nat Rev Mol Cell Biol. 2012; 13:312–21. [PubMed: 22522719]
7. Sinclair DA, et al. Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). Proc Natl Acad Sci U S A. 2009; 106:13427–32. [PubMed: 19666537]
8. Durgan DJ, et al. O-GlcNAcylation, novel post-translational modification linking myocardial metabolism and cardiomyocyte circadian clock. J Biol Chem. 2011; 286:44606–19. [PubMed: 22069332]
9. Kim EY, et al. A role for O-GlcNAcylation in setting circadian clock speed. Genes Dev. 2012; 26:490–502. [PubMed: 22327476]
10. Geng F, Zhu W, Anderson RA, Leber B, Andrews DW. Multiple post-translational modifications regulate E-cadherin transport during apoptosis. J Cell Sci. 2012
11. Lefebvre T, et al. Evidence of a balance between phosphorylation and O-GlcNAc glycosylation of Tau proteins--a role in nuclear localization. Biochim Biophys Acta. 2003; 1619:167–76. [PubMed: 12527113]
12. Yang X, et al. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. Nature. 2008; 451:964–9. [PubMed: 18288188]
13. Pathak S, et al. O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release. EMBO J. 2012; 31:1394–404. [PubMed: 22307082]
14. Rotty JD, Hart GW, Coulombe PA. Stressing the role of O-GlcNAc: linking cell survival to keratin modification. Nat Cell Biol. 2010; 12:847–9. [PubMed: 20811358]
15. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002; 298:1912–34. [PubMed: 12471243]
16. Haltiwanger RS, Holt GD, Hart GW. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. J Biol Chem. 1990; 265:2563–8. [PubMed: 2137449]
17. Kreppel LK, Blomberg MA, Hart GW. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. J Biol Chem. 1997; 272:9308–15. [PubMed: 9083067]
18. Dong DL, Hart GW. Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. J Biol Chem. 1994; 269:19321–30. [PubMed: 8034696]

19. Gao Y, Wells L, Comer FI, Parker GJ, Hart GW. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. J Biol Chem. 2001; 276:9838–45. [PubMed: 11148210]
20. Jinek M, et al. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat Struct Mol Biol. 2004; 11:1001–7. [PubMed: 15361863]
21. Clarke AJ, et al. Structural insights into mechanism and specificity of O-GlcNAc transferase. EMBO J. 2008; 27:2780–8. [PubMed: 18818698]
22. Martinez-Fleites C, et al. Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. Nat Struct Mol Biol. 2008; 15:764–5. [PubMed: 18536723]
23. Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature. 2011; 469:564–7. [PubMed: 21240259]
24. Iyer SP, Hart GW. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. J Biol Chem. 2003; 278:24608–16. [PubMed: 12724313]
25. Kreppel LK, Hart GW. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. J Biol Chem. 1999; 274:32015–22. [PubMed: 10542233]
26. Schimpl M, et al. O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. Nat Chem Biol. 2012; 8:969–74. [PubMed: 23103942]
27. Lazarus MB, et al. HCF-1 is cleaved in the active site of O-GlcNAc transferase. Science. 2013; 342:1235–9. [PubMed: 24311690]
28. Vosseller K, et al. O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. Mol Cell Proteomics. 2006; 5:923–34. [PubMed: 16452088]
29. Cole RN, Hart GW. Cytosolic O-glycosylation is abundant in nerve terminals. J Neurochem. 2001; 79:1080–9. [PubMed: 11739622]
30. Wells L, et al. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. Mol Cell Proteomics. 2002; 1:791–804. [PubMed: 12438562]
31. Khidekel N, et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. Nat Chem Biol. 2007; 3:339–48. [PubMed: 17496889]
32. Chalkley RJ, Thalhammer A, Schoepfer R, Burlingame AL. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. Proc Natl Acad Sci U S A. 2009; 106:8894–9. [PubMed: 19458039]
33. Hahne H, et al. Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. J Proteome Res. 2013; 12:927–36. [PubMed: 23301498]
34. Trinidad JC, et al. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. Mol Cell Proteomics. 2012
35. Hahne H, Gholami AM, Kuster B. Discovery of O-GlcNAc-modified proteins in published large-scale proteome data. Mol Cell Proteomics. 2012
36. Alfaro JF, et al. Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. Proc Natl Acad Sci U S A. 2012
37. Schutkowski M, et al. High-content peptide microarrays for deciphering kinase specificity and biology. Angew Chem Int Ed Engl. 2004; 43:2671–4. [PubMed: 18629986]
38. Roquemore EP, et al. Vertebrate lens alpha-crystallins are modified by O-linked N-acetylglucosamine. J Biol Chem. 1992; 267:555–63. [PubMed: 1730617]
39. Housley MP, et al. A PGC-1alpha-O-GlcNAc transferase complex regulates FoxO transcription factor activity in response to glucose. J Biol Chem. 2009; 284:5148–57. [PubMed: 19103600]
40. Ball LE, Berkaw MN, Buse MG. Identification of the major site of O-linked beta-N-acetylglucosamine modification in the C terminus of insulin receptor substrate-1. Mol Cell Proteomics. 2006; 5:313–23. [PubMed: 16244361]
41. Klein AL, Berkaw MN, Buse MG, Ball LE. O-linked N-acetylglucosamine modification of insulin receptor substrate-1 occurs in close proximity to multiple SH2 domain binding motifs. Mol Cell Proteomics. 2009; 8:2733–45. [PubMed: 19671924]

42. Yuzwa SA, et al. Mapping O-GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. Amino Acids. 2011; 40:857–68. [PubMed: 20706749]
43. Wells L, Slawson C, Hart GW. The E2F-1 associated retinoblastoma-susceptibility gene product is modified by O-GlcNAc. Amino Acids. 2011; 40:877–83. [PubMed: 20680651]
44. Guo K, et al. Translocation of HSP27 into liver cancer cell nucleus may be associated with phosphorylation and O-GlcNAc glycosylation. Oncol Rep. 2012; 28:494–500. [PubMed: 22664592]
45. Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. Proc Natl Acad Sci U S A. 2004; 101:13132–7. [PubMed: 15340146]
46. Maury JJ, Ng D, Bi X, Bardor M, Choo AB. Multiple reaction monitoring mass spectrometry for the discovery and quantification of O-GlcNAc-modified proteins. Anal Chem. 2014; 86:395–402. [PubMed: 24144119]
47. Gloster TM, et al. Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. Nat Chem Biol. 2011; 7:174–81. [PubMed: 21258330]
48. Lazarus MB, et al. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. Nat Chem Biol. 2012; 8:966–8. [PubMed: 23103939]
49. Tvaroska I, Kozmon S, Wimmerova M, Koca J. Substrate-assisted catalytic mechanism of O-GlcNAc transferase discovered by quantum mechanics/molecular mechanics investigation. J Am Chem Soc. 2012; 134:15563–71. [PubMed: 22928765]
50. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–90. [PubMed: 15173120]
51. Gay LM, Zheng X, van Aalten DM. Molecular recognition: O-GlcNAc transfer: size matters. Nat Chem Biol. 2011; 7:134–5. [PubMed: 21321551]
52. Lubas WA, Hanover JA. Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity. J Biol Chem. 2000; 275:10983–8. [PubMed: 10753899]
53. Fujiki R, et al. GlcNAcylation of histone H2B facilitates its monoubiquitination. Nature. 2011; 480:557–60. [PubMed: 22121020]
54. Roquemore EP, Chevrier MR, Cotter RJ, Hart GW. Dynamic O-GlcNAcylation of the small heat shock protein alpha B-crystallin. Biochemistry. 1996; 35:3578–86. [PubMed: 8639509]
55. Zhu Y, et al. O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. Nat Chem Biol. 2015; 11:319–25. [PubMed: 25774941]
56. Leavy TM, Bertozzi CR. A high-throughput assay for O-GlcNAc transferase detects primary sequence preferences in peptide substrates. Bioorg Med Chem Lett. 2007; 17:3851–4. [PubMed: 17531489]
57. Liu X, et al. A peptide panel investigation reveals the acceptor specificity of O-GlcNAc transferase. FASEB J. 2014; 28:3362–72. [PubMed: 24760753]
58. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002:310–22. [PubMed: 11928486]
59. Wang J, Torii M, Liu H, Hart GW, Hu ZZ. dbOGAP - an integrated bioinformatics resource for protein O-GlcNAcylation. BMC Bioinformatics. 2011; 12:91. [PubMed: 21466708]
60. Jia CZ, Liu T, Wang ZP. O-GlcNAcPRED: a sensitive predictor to capture protein O-GlcNAcylation sites. Mol Biosyst. 2013; 9:2909–13. [PubMed: 24056994]
61. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 2011; 67:271–81. [PubMed: 21460445]
62. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–32. [PubMed: 20124692]
63. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. 1997; 276:307–326.
64. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53:240–55. [PubMed: 15299926]

65. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–32. [PubMed: 15572765]

66. Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr. 2004; 60:1355–63. [PubMed: 15272157]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

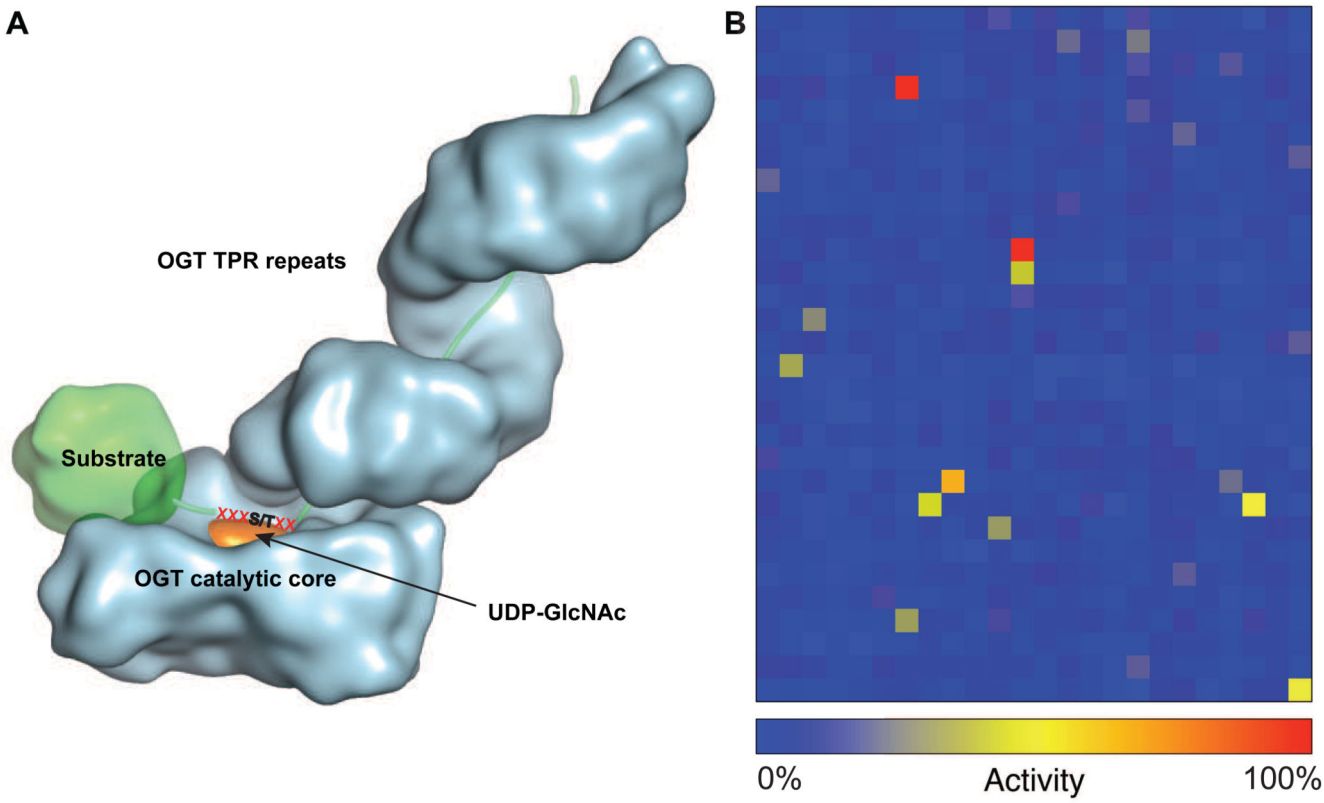


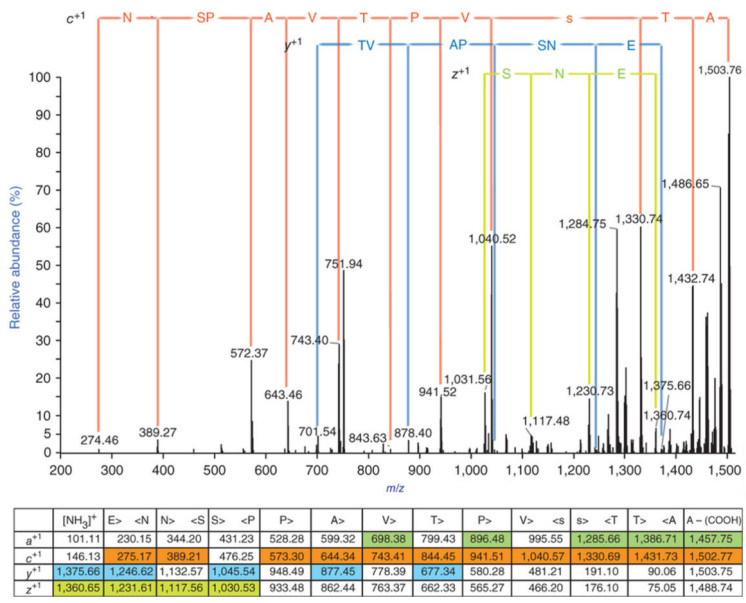
Figure 1. OGT shows substrate selectivity at the peptide level.

a. Schematic representation of OGT substrate recognition, highlighting three possible (and not mutually exclusive) binding modes. Acceptor substrates may be selected by 1) recognition of globular parts of the substrate protein by as yet unidentified regions on OGT, 2) recognition of unfolded regions by the TPR domain or 3) recognition of a target peptide sequence by the OGT active site, the binding mode that is the subject of investigation in this work.

b. A library of 720 synthetic peptides was *in vitro* O-GlcNAcylated with OGT and glycosylation of the peptides was quantified by a single point scintillation proximity assay. The average of two independent measurements is depicted in a heat map, scaled from no (0 %, blue) activity over background to the highest activity (100%, red) corresponding to the RBL2 peptide KENSPAVTPVSTA. The layout of the heatmap corresponds to peptide 1 from the Jerini peptide library (see Methods and Table S1) in the upper left corner, with peptide 720 in the lower right corner.

Europe PMC Funders Author Manuscripts

A



	[NH3]+	E> <N	N> <S	S> <P	P>	A>	V>	T>	P>	V> <s	s> <T	T> <A	A – (COOH)
a^{+1}	101.11	230.15	344.20	431.23	528.28	599.32	698.38	799.43	896.48	995.55	1,285.66	1,386.71	1,457.75
c^{+1}	146.13	275.17	389.21	476.25	573.30	644.34	743.41	844.45	941.51	1,040.57	1,330.69	1,431.73	1,502.77
y^{+1}	1,375.66	1,246.62	1,132.57	1,045.54	948.49	877.45	778.39	677.34	580.28	481.21	191.10	90.06	1,503.75
z^{+1}	1,360.65	1,231.61	1,117.56	1,030.53	933.48	862.44	763.37	662.33	565.27	466.20	176.10	75.05	1,488.74

B

PROTEIN	AMINO ACID NUMBER	PEPTIDE WITH O-GlcNAc SITES
RBL2	410-422	KENSPCVTPV**ST**A
M phase phospho protein -9	818-830	TKREIMLTPV**T**VA
ELK1	361-375	TLTPVLLTPS**S**LP
FOXO1	312-324	TFRPRT**S**SNASTI
ETS1	32-43	CADVPLLTPS**S**KE
α-crystallin B chain	38-50	FPT**S**TSLSPFYLR
Keratin 7	7-19	SPVFTSR**S**AAFSG
RET	680-692	AQAFPVSY**S**SSGA
MYB-related protein B	433-445	LDSCNSLTPK**S**TP
SCG10	55-67	ELILKPPSPI**S**EA
RBL2	655-667	GLGRSITSPT**T**LY
GSK3β	209-221	RGEPNV**S**YIA**S**RY
Kv2.1	489-501	KWTKR**T**LSET**S**SS
M phase phospho protein -9	11-23	TPV**T**VAYSPKRSP
Early E1A 32Kd	212-224	AILRRPTSPV**S**RE
Tau	389-401	GAEIVYKSPVV**S**G
Rhodopsin-OPSD	333-345	ASATV**S**KTET**S**QV
Insulin receptor substrate-1	982-994	VP**SS**RGDYMTMQM
Lamin B1	389-401	KLSPSPSSRV**T**VS
Lamin A	618-630	SSVTV**T**R**S**YR**S**VG
CGKI	52-64	THIGPR**T**TRAQGI
ELK1	376-389	IHFWS**T**LSPIAPR
BCKD kinase	45-57	ER**S**KTV**TS**FYNQS
Phospholipase C-γ-2	747-759	RDINSLYDV**S**RMY
HSP27	71-83	PAY**S**RALSRQLSS
Desmocollin-3	857-869	YNYEGRG**S**VAGSV
DOUBLIN	423-435	KDLYLPL**S**LDDSD

Figure 2. OGT modifies specific sites on peptide substrates.

Synthetic peptides were enzymatically O-GlcNAc modified with OGT, and the glycosylation sites were determined by ETD mass spectrometry.

a. ETD-MS/MS spectrum of RBL2 peptide [KENSPAVTPVSTA + GlcNAc + 2H]$^{2+}$ = 752.82 *m/z*, produced an intense ion at *m/z* 1503.76 corresponding to the monocharged precursor. The spectrum shows the c^{+1} fragment ions in orange, which unequivocally assign the O-GlcNAc site to S11, as well as the z^{+1} ions in green and the y^{+1} ions in blue. **b.** Peptides emerging as hits from the screen, ranked by OGT activity. The O-GlcNAc Ser/Thr

residue in each peptide is highlighted in red. Green shading marks peptides in which the site identified from the screen matches previous reports. Peptides derived from proteins that have been published in the literature as O-GlcNAcylated proteins (with unknown modification sites) are highlighted in blue.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

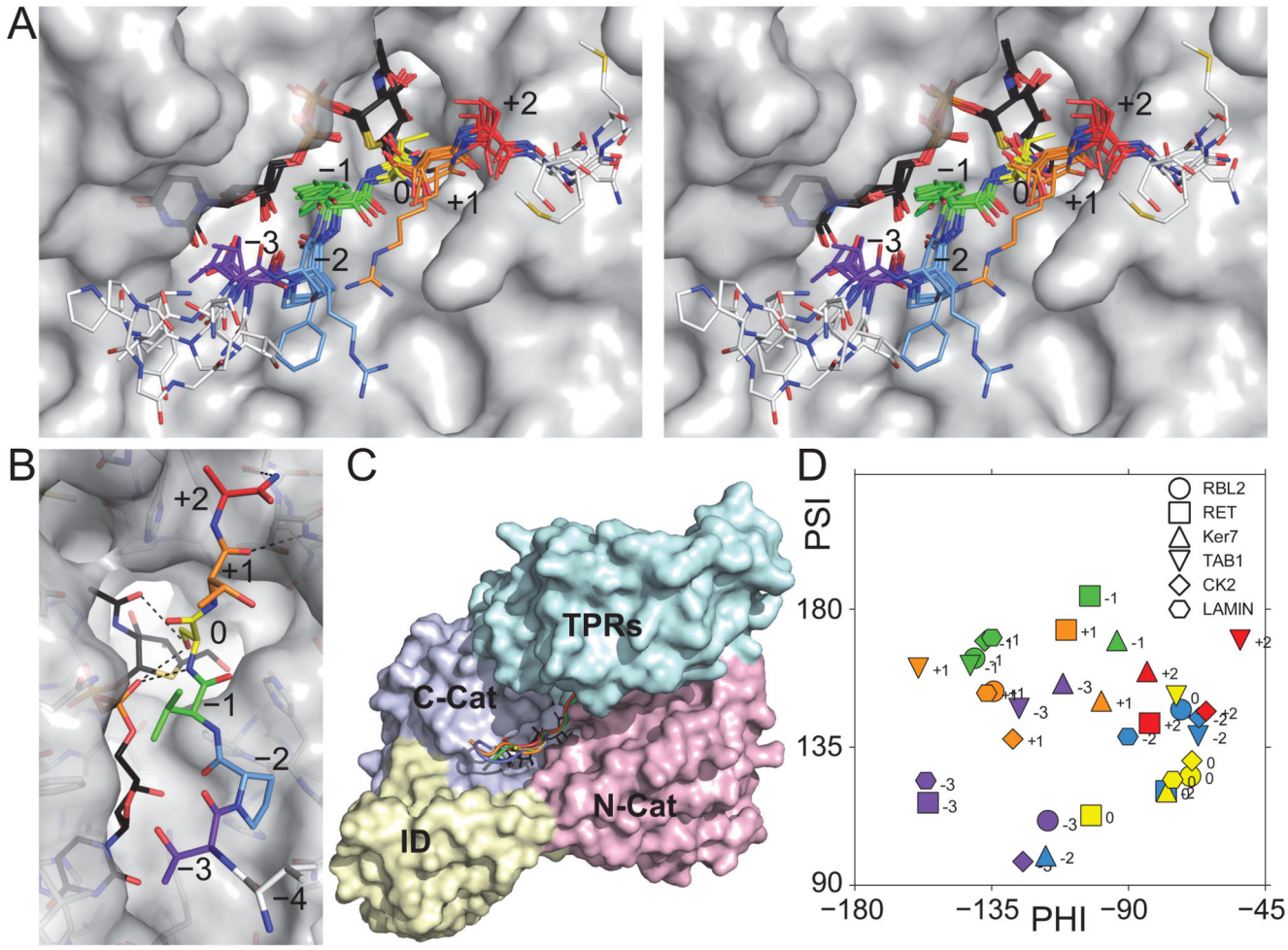


Figure 3. Substrate peptides bind the active site of OGT with similar conformations in the −3 to +2 subsites.

a. Stereo view of the active site region of OGT (grey surface), in complex with donor substrate analogue UDP-5S-GlcNAc (black sticks) and different substrate peptides derived from the following proteins: RBL2, Keratin-7, RET, Lamin B1, CK2 (PDBid 4GYY) and TAB1 (PDBid 4AY6). Peptides are colored by position in the active site to illustrate the conformational similarities within 3 amino acids before (−3 to −1 subsites), and 2 amino acids beyond (+1 to +2 subsites) the acceptor serine (the '0' subsite). Amino acids beyond the −3 to +2 subsites are shown as white sticks. For clarity, the TPR domain has been omitted. Individual structures including unbiased F_o-F_c electron density are shown in Supplementary Fig. S2.

b. Hydrogen bonding interactions tethering the RBL2 peptide in the active site of OGT are depicted as dashed lines. Colors as in panel **a**.

c. Domain structure of the crystallized fragment of OGT (amino acids 312—1031) in surface representation, with UDP-5S-GlcNAc shown as black sticks and substrate peptides in cartoon representation. The N-terminal TPR domain is formed by tetratricopeptide repeats, N-Cat and C-Cat are the amino-and carboxy-terminal lobes of the catalytic domain, ID is the 'intervening domain'.

d. Sub-section of a Ramachandran plot showing backbone torsion angles of substrate peptides. All phi and psi angles for the −3 to +2 sub-sites fall within the region characteristic of beta-strands. Shapes are used to differentiate the six different peptides, and the colors for subsites −3 to +2 are the same as in panel **a**.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

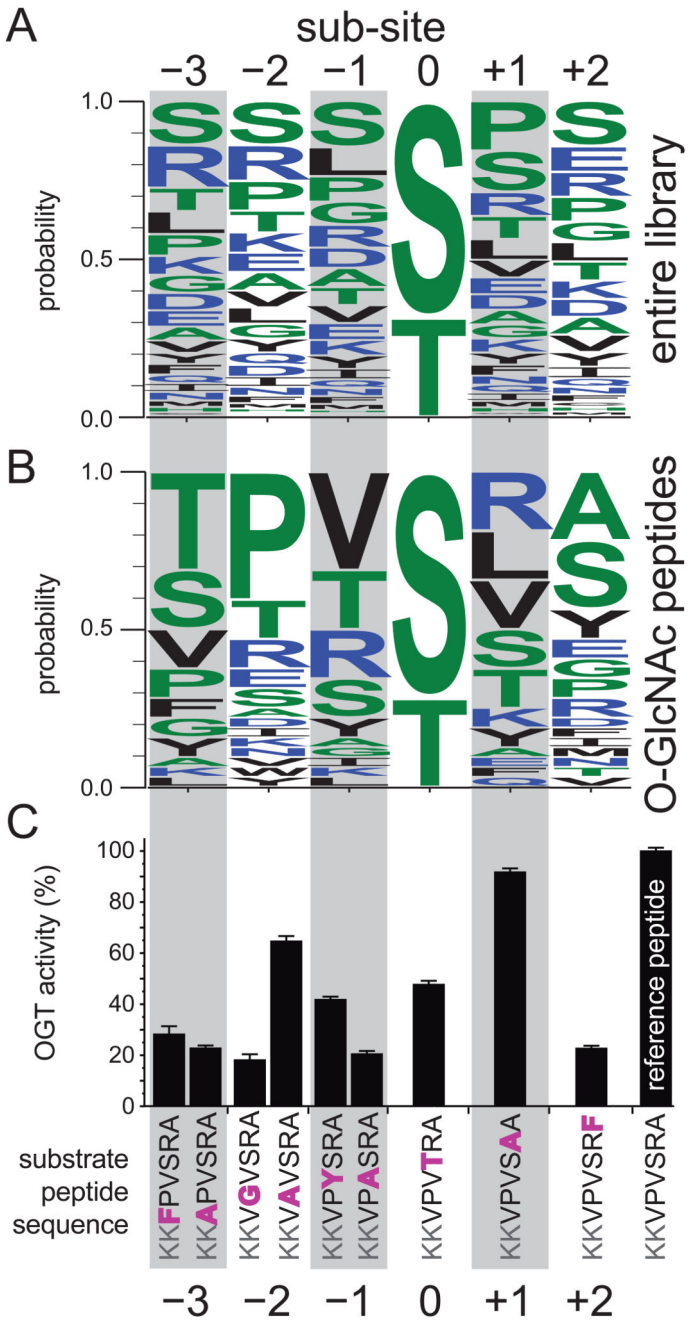


Figure 4. OGT hexapeptide sequon derived from the peptide library hits.

a. Sequence logo analyzing the overall composition of the library of 720 peptides. All possible hexapeptides with a Ser or Thr in the 4th position (corresponding to the 0 sub-site) were extracted, and a sequence logo was generated to illustrate the relative abundance of different amino acids (http://weblogo.threeplusone.com).

b. The O-GlcNAc sequon was derived from experimentally determined O-GlcNAc sites. Peptides were aligned by the modified Ser/Thr, and the sequence truncated to include only 3 residues to the N-terminus of the GlcNAc site ("−3") and 2 residues to the C-terminus of the

site (“+2”) as dictated by the ordered binding of peptides in only these subsites from the structural data. The peptide sequences used for generation of the logo are listed in Supplementary Fig. S3.

c. The tolerance of OGT for single amino acid substitutions was probed using the reference peptide KKVPVSRA, which represents the optimal OGT sequon except for position −3, where Val was used in order to avoid a potential second O-GlcNAc acceptor. Two Lys were added to the N-terminus to aid peptide solubility. Assay details are given in the Online Methods section. The mean of three measurements is shown, with error bars depicting the s.e.m. Supplementary Fig. S4 shows the activity of full-length human OGT (ncOGT) on the same panel of peptides.

Europe PMC Funders Author Manuscripts