HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2018 April 23.

Published in final edited form as:

Nat Chem Biol. 2017 December ; 13(12): 1267–1273. doi:10.1038/nchembio.2494.

Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase

Chia-Wei Hu[1,4], **Matthew Worth**[2,4], **Dacheng Fan**[1,4], **Baobin Li**[1,4], **Hao Li**[1,4], **Lei Lu**[1], **Xiaofang Zhong**[1], **Ziqing Lin**[3], **Liming Wei**[3], **Ying Ge**[2,3], **Lingjun Li**[1,2], and **Jiaoyang Jiang**[1,*]

[1]Pharmaceutical Sciences Division, School of Pharmacy, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705

[2]Department of Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705

[3]Department of Cell and Regenerative Biology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705

Abstract

O-linked β-*N*-acetylglucosamine (O-GlcNAc) transferase (OGT) is an essential human glycosyltransferase that adds O-GlcNAc modifications on numerous proteins. However, little is known about how OGT recognizes various protein substrates. Here we report GlcNAc electrophilic probes (GEPs) to expedite the characterization of OGT-substrate recognition. Data from mass spectrometry, X-ray crystallization, and biochemical and radiolabeled kinetic assays

*Jiaoyang Jiang: jiaoyang.jiang@wisc.edu.

[4]These authors contributed equally to this work.

Accession codes

Primary accessions

Protein Data Bank

5VIF

5VIE

Referenced accessions

Protein Data Bank

4GZ5

4GYW

4GYY

3PE3

Author contributions

J.J. oversaw all aspects of the experiments and manuscript preparation. C-W.H. performed mass spectrometric, crosslinking, and cell culture experiments. M.W. and H.L. synthesized the compounds. D.F. performed mutagenesis, in-gel fluorescence scanning, and enzyme kinetic experiments. B.L. obtained protein crystals and determined structures. L.L. assisted with protein purification. X.Z., L.L., Z.L., L.W., and Y.G. provided access to the MS instruments. C-W.H. and J.J. wrote the manuscript with help from M.W., D.F., B.L., and H.L.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information and chemical compound information are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.J.

support the application of GEPs to rapidly report the impacts of OGT mutations on protein substrate or sugar binding and to discover OGT residues crucial for protein recognition. Interestingly, we found that the same residues on the inner surface of the N-terminal domain contribute to OGT interactions with different protein substrates. By tuning reaction conditions, a GEP enables crosslinking of OGT with acceptor substrates *in situ*, affording a unique method to discover genuine substrates that weakly or transiently interact with OGT. Hence, GEPs provide new strategies to dissect OGT-substrate binding and recognition.

INTRODUCTION

O-linked β-*N*-acetylglucosamine transferase (OGT) is an essential glycosyltransferase that catalyzes the transfer of monosaccharide *N*-acetylglucosamine (GlcNAc) from the sugar donor uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to serine and threonine residues of intracellular proteins.[1] This modification, termed O-GlcNAcylation, dynamically modulates the activity, stability, and interaction of proteins that participate in various biological processes such as signal transduction,[2] transcriptional regulation,[3] cell cycle,[4] and apoptosis.[5] Genetic knockout of OGT is embryonic lethal[6] and severe developmental defects have been observed in animal models with tissue-specific targeted deletion of OGT.[7] Dysfunction of OGT gives rise to aberrant protein O-GlcNAcylation, which has been detected in diseases such as diabetes,[8] cancer,[9] neurodegeneration,[10] and cardiovascular disorders.[11] Therefore, OGT is of fundamental importance for both physiological and pathological processes. To date, a large number of O-GlcNAcylated proteins have been reported.[12] However, the recognition mode of OGT towards diverse protein substrates remains largely unknown.

Human OGT is a multi-domain protein with an N-terminal tetratricopeptide repeat (TPR) domain folded into an extended α-helical tunnel, which has been proposed to be involved in protein-protein interactions.[13,14] The C-terminal catalytic domain of OGT is responsible for sugar transfer.[15,16] The crystal structure of a truncated OGT (called $OGT_{4.5}$) containing 4.5 of the total 13.5 TPRs of the full-length OGT, as well as its complexed structures with short peptide substrates, have been reported.[16] While these studies revealed that peptide substrates are mainly anchored through interactions with active-site bound UDP-GlcNAc and backbone interactions with neighboring TPR residues, no apparent sequence motif has been identified in OGT substrates. In addition, although $OGT_{4.5}$ is fully competent on glycosylating a number of peptide substrates, it has minimal activity towards proteins,[13,15] indicating that the extended TPR domain is crucial for protein binding. However, owing to the flexible nature of TPR, full-length OGT has not been amenable for crystallization. This poses a big hurdle to investigate the substrate recognition of OGT beyond the active site. Moreover, kinetic studies have shown that O-GlcNAcylation follows an ordered bi-bi mechanism,[13,15] in which OGT binds to UDP-GlcNAc prior to protein substrate. Hence, OGT mutation with impaired sugar binding also affects protein binding, which presents a great challenge on traditional protein-protein interaction methods (*e.g.*, isothermal titration calorimetry (ITC),[17] surface plasmon resonance (SPR),[18] and mass spectrometry (MS)[19]) to identify OGT structural features that contribute to specific protein binding. Therefore, new efficient

strategies are much needed to decipher the molecular basis for OGT substrate specificity and its precise role in physiology and disease.

In this study, we developed a class of GlcNAc Electrophilic Probes (GEPs) that can be applied to accelerate the characterization of OGT's structural features in protein substrate binding. We found that GEPs covalently modified OGT and/or O-GlcNAcylated proteins to generate distinct labeling patterns as direct readouts of OGT altered ability for sugar binding versus protein substrate binding. Applying these probes in click chemistry-based fluorescent assays, we rapidly screened a few OGT variants and identified asparagine residues on the inner surface of the 10th TPR contributing to OGT interactions with different protein substrates. Furthermore, we discovered that by adjusting the reaction conditions, a GEP can "lock" OGT with acceptor substrates *in situ* during the sugar transfer step, affording a unique strategy to discover genuine substrates that transiently or weakly interact with OGT. These features of GEPs will provide fundamental insights into the substrate specificity of full-length OGT and generate new perspectives for inhibitor development.

RESULTS

Synthesis and evaluation of GlcNAc Electrophilic Probes

To characterize OGT-substrate recognition, we rationally designed GlcNAc Electrophilic Probes (GEPs) based on the observation of a unique cysteine residue (C917) in close proximity (3.6 Å) to the *N*-acetyl group of UDP-GlcNAc in the OGT active site (PDB 4GZ5; Supplementary Results, Supplementary Fig. 1).[15,20] We hypothesized that a UDP-GlcNAc analogue containing a suitable electrophilic functionality might be able to react with the nucleophilic side chain of C917 and specifically label OGT in the absence of protein substrate (Figs. 1a,b). While in the presence of immediately available protein substrates, this probe can potentially conduct O-GlcNAcylation similarly as UDP-GlcNAc (Fig. 1b). According to the ordered bi-bi mechanism, this probe could potentially discriminate OGT's ability of sugar binding versus protein binding by generating varied levels of modification on OGT C917 and the protein substrate, both of which can be easily detected using click chemistry-based fluorescent assay (Fig. 1c).

To test this principle, we have synthesized a small panel of UDP-GlcNAc analogues containing different electrophiles extended from the *N*-acetyl group (*e.g.*, GEP1 and GEP2 in Fig. 1a; see Supplementary Note 1 for detailed compound synthesis and characterization). We first examined the ability of these compounds for covalent labeling of purified $OGT_{4.5}$ using intact protein MS analysis and discovered that GEP1, but not GEP2, irreversibly reacted with OGT (Supplementary Figs. 2a,b). Hence, our following experiments were focused on GEP1. To evaluate the reaction specificity, we employed a similar intact protein MS strategy to quantify the stoichiometry of GEP1 modification on OGT cysteines (+245.1 Da modification in Fig. 2a) in a range of probe concentrations. Remarkably, at 1:2 and 1:10 ratios of enzyme to GEP1, we only detected a single unit of modification, to an extent of 60% and 100%, respectively (Fig. 2a). In contrast, we did not detect any modification on the serine mutation of C917 of OGT (C917S) (Supplementary Fig. 2c). This suggests that GEP1 specifically targets C917 over a dozen other cysteines in OGT. To provide unambiguous evidence for the site-specific labeling, GEP1 (10-fold excess) was incubated with full-length

OGT followed by trypsin digestion and LC-MS/MS analysis. This experiment validated that GEP1 exclusively labeled C917 but not other cysteines of OGT (Fig. 2b and Supplementary Table 1). Furthermore, the covalent labeling of OGT proceeded rapidly in a few minutes and prolonged incubation generated UDP-hydrolyzed modification on C917 in a time-dependent manner (Supplementary Figs. 3a,b). This observation is in line with previous reports on unproductive hydrolysis of UDP from nucleotide sugars by OGT and other glycosyltransferases.[16,21] To obtain a more accurate reaction rate for GEP1 with C917, this compound was incubated with OGT at different ratios in a time course followed by trypsin digestion, dimethyl labeling, and quantitative LC-MS/MS analysis. The relative abundance of unmodified C917-containing peptides was applied to calculate the reaction rate as previously reported.[22] We found that the reaction rate of GEP1 with C917 was time-dependent (Supplementary Fig. 3c). The initial rate was determined to be around 0.01 s^{-1}, in a similar range as the k_{cat} of OGT in the glycosylation of protein substrates measured in our lab. Taken together, these results demonstrated that GEP1 is capable of efficiently and specifically labeling the C917 of OGT.

GEP1-derived O-GlcNAcylation

We next investigated whether GEP1 can be employed by OGT to conduct O-GlcNAcylation in the presence of acceptor substrates. Purified OGT was incubated with GEP1 and α-crystallin B chain peptide (FPTSTSLSPFYLR), a well-characterized substrate of OGT containing a single O-GlcNAcylation site.[23] Using LC-MS/MS, glycosylated peptide along with the GEP1 fragments as signature ions were detected (Fig. 3a). This result indicates that GEP1 is an efficient sugar donor for OGT, and the glycosylation occurs at a rate that exceeds the irreversible reaction with C917. To examine if there is any structural perturbation of OGT to accommodate GEP1 during sugar transfer, we solved the X-ray crystal structure (2.2 Å) of $OGT_{4.5}$ in the presence of GEP1 and a CKII peptide (YPGGSTPVSSANMM) that possesses a single O-GlcNAcylation site.[24] This structure clearly demonstrated that CKII peptide was glycosylated by GEP1 during crystallization (Figs. 3b,c; Supplementary Table 2), similarly as that reported for the crystallization of $OGT_{4.5}$ with UDP-GlcNAc and the same peptide (PDB 4GYW[15]). Importantly, these two complex structures are nearly identical, in terms of the overall structure of $OGT_{4.5}$ (RMSD value of 0.21 Å) and the binding conformations of UDP and the glycosylated peptides (Fig. 3d). Collectively, these data provided compelling evidence supporting that GEP1 does not alter OGT's conformation or substrate specificity, and thus is well suited for O-GlcNAcylation.

Fluorescent assay to characterize substrate recognition

The unique features of GEP1 promoted us to develop a facile assay for rapid elucidation of OGT's substrate binding residues. The principle of this assay is depicted in Figure 1b. We predicted that OGT mutations affecting sugar or protein binding would generate distinct patterns of labeled OGT and glycosylated protein substrate (Fig. 1c). Varied labeling patterns will provide a rapid readout of the impact on OGT's potentially altered sugar binding versus protein binding ability, which is indistinguishable using other O-GlcNAc chemical probes or otherwise requires extensive enzyme kinetic characterization. It is our expectation that this assay will aid in a better understanding of the substrate-specificity of OGT.

To facilitate detection of probe-modified proteins, an azide handle was introduced to the 6´-position of GEP1 (GEP1A in Fig. 1a, see Supplementary Note 1 for detailed compound synthesis and characterization). The new probe enables detection of modified proteins through click chemistry and in-gel fluorescence scanning (Fig. 1c). We anticipated that this new probe could be accommodated by OGT, since both GEP1 and reported UDP-6AzGlcNAc (Fig. 1a) are effective sugar donors of OGT.[25,26] Indeed, we confirmed the covalent labeling of GEP1A on OGT C917 by intact protein MS (Supplementary Fig. 4a) and LC-MS/MS analysis (Supplementary Table 1). More importantly, we also detected GEP1A-derived glycosylation on Lamin B1 peptide substrate[23] (Supplementary Fig. 4b). To examine the application of this probe, we incubated full-length OGT with GEP1A and recombinant nuclear pore protein 62 (NUP62), a well-known protein substrate that can be extensively glycosylated by OGT.[27] Reaction products were coupled to a fluorescent dye and separated on SDS-PAGE gel (representative gels in Fig. 4a and Supplementary Fig. 5). As expected, we detected a strong fluorescent band corresponding to the glycosylated NUP62, along with a marginally labeled OGT band (Fig. 4a). We further examined the reaction specificity by control experiments: i) OGT-C917S mutant displayed negligible labeling in the absence of protein substrate (Fig. 4a), which was consistent with the expected specificity of GEP1A for targeting C917; and ii) GEP1A-labeled NUP62 in the absence of OGT was minimal (Fig. 4a), indicating that labeling of NUP62 was attributed to GEP1A-derived glycosylation by OGT.

Next, we applied known OGT variants to examine if the labeling pattern can be exploited to distinguish OGT mutations that alter sugar binding versus protein binding. OGT-K842A contains a known mutation in the enzyme active site that abolishes sugar binding.[20] Consistent with our prediction, no significant fluorescent labeling on OGT-K842A or NUP62 was detected under the same conditions as mentioned for the wild-type enzyme (Fig. 4a). To assess protein binding, we tested $OGT_{4.5}$, which is deficient in protein binding but retains the ability for sugar binding.[14,16] Compared to the wild-type enzyme, the labeling of $OGT_{4.5}$ was dramatically augmented while the glycosylation of NUP62 was markedly reduced (Fig. 4a), which is in excellent agreement with the predicted pattern change for compromised protein binding (Fig. 1c). To achieve a more consistent comparison, background-fluorescence was subtracted and the signal was normalized to the protein amount based on Coomassie Blue staining of the same gel (Fig. 4a). In this way, the level of each modification can be relatively quantified (Fig. 4b). To determine the linear detection range, a serial dilution revealed that this fluorescent assay reproducibly detected at least a 16-fold linear dilution of the sample as long as the lowest signal-over-background was detected (Supplementary Fig. 6). We found that the ratio of fluorescent signal over Coomassie Blue was within 17% deviation range, suggesting that this assay is reliable for semi-quantification. To further validate that GEP1A can be efficiently utilized by OGT similarly as the native sugar donor, we conducted a competition assay, in which varied amounts of UDP-GlcNAc were added into the reaction containing a constant concentration of GEP1A, OGT and NUP62. In the presence of excess nucleotide sugars, GEP1A-glycosylated NUP62 was quantified using in-gel fluorescence scanning. This result tightly matched to the theoretical values calculated under the assumption that OGT had no preference for either sugar (Supplementary Fig. 7), indicating that GEP1A can be employed

similarly as UDP-GlcNAc. Taken together, these experiments validated the principle of the fluorescent assay and supported the application of GEP compounds to investigate the structural features of OGT for protein substrate recognition.

Defining key OGT residues in protein substrate binding

With the assay principle validated, we proceeded to demonstrate the application of GEP1A in characterizing key residues of OGT engaged in substrate binding. The labeling pattern for a protein-binding-compromised mutant is expected to be significantly different from a sugar-binding-impaired mutant (*e.g.*, $OGT_{4.5}$ compared to K842A in Fig. 4a), but it may not be distinguishable from a sugar-transfer-impaired mutant (Fig. 1c). To demonstrate the performance of GEP1A in these cases, we carried out assays with NUP62 as the acceptor substrate and investigated a set of OGT residues across the surface of the TPR, catalytic region, and intervening domain (Supplementary Fig. 8). While some mutations did not generate any noticeable change on the labeling pattern following the GEP1A fluorescent assay (Supplementary Fig. 9), we discovered that mutants of N321A/N322A and D554N dramatically reduced the glycosylation of NUP62 while enhancing the labeling of OGT compared to the wild-type enzyme (Figs. 4c,d), displaying a featured pattern of protein-binding-compromised or sugar-transfer-impaired mutants (Fig. 1c).

Since the catalytic base of OGT is still under debate,[13,15,28,29] we first examined the role of the D554 residue using the D554N mutant. A previous structural study proposed that during the GlcNAc transfer step, the negatively charged side chain of D554 shuttles the hydroxyl proton of the glycosylating serine/threonine via ordered water molecules away from the enzyme active site (Supplementary Fig. 10).[15] Thus, the D554N mutant could be defective for sugar transfer, which is in excellent agreement with what we detected in the GEP1A fluorescent assay (Figs. 4c,d). The radiolabeled kinetic experiments further supported the role of this residue and demonstrated that the O-GlcNAcylation turnover number (k_{cat}) of D554N dropped to 4%, while the impact on the Michaelis constant (K_m) for NUP62 was minor (Supplementary Fig. 11, Supplementary Note 2). On the other hand, radiolabeled kinetic experiments validated that the double mutation N321A/N322A impaired the binding ability of OGT towards NUP62 protein, as the K_m (NUP62) significantly increased with negligible impact on the k_{cat} (Supplementary Fig. 11, Supplementary Note 2). These kinetic results were consistent with our prediction based on the GEP1A fluorescent assay (Figs. 1c; 4c,d). Collectively, these experiments provided strong evidence supporting the essential role of D554 during sugar transfer and the important contribution of N321/N322 to OGT binding with NUP62 protein.

To further evaluate the robustness and sensitivity of GEP1A fluorescent assay, we then examined the performance of GEP1A on another acceptor substrate protein, O-GlcNAcase (OGA). Unlike NUP62 that can be extensively O-GlcNAcylated by OGT on multiple sites, OGA is known to be O-GlcNAcylated on a single site (S405) at a low stoichiometry (~10%) in cells.[30–32] *In vitro*, with recombinant OGT and a large excess of UDP-GlcNAc, we detected only ~20% O-GlcNAcylation on the purified OGA mutant D175N (OGA-D175N) that is defective in O-GlcNAc hydrolysis. With this OGA mutant as the acceptor substrate, we first performed control experiments to assess the labeling specificity of GEP1A and

examined the pattern changes with known OGT variants to discriminate sugar binding versus protein substrate binding. Due to the overall weak fluorescent signal (low glycosylation stoichiometry), the background labeling of OGA appeared to be stronger than that of NUP62 (compare Supplementary Fig. 12a with Fig. 4a). However, in the presence of WT OGT, we readily detected the glycosylation of OGA mutant over background (Supplementary Fig. 12a). More importantly, the pattern changes induced by OGT variants ($OGT_{4.5}$ and K842A) strongly matched the prediction (Supplementary Fig. 12a), supporting the robustness and sensitivity of GEP1A in characterizing OGT substrates with low glycosylation levels.

Next, we exploited the GEP1A fluorescent assay to investigate whether the mutants of OGT in the TPR domain (*e.g.*, N321A/N322A and E301A in Supplementary Fig. 8) affect the binding of different protein substrates equally. With OGA-D175N as the acceptor substrate, the E301A mutant exhibited a near identical labeling pattern as the wild-type OGT (Supplementary Fig. 12b), similar to when NUP62 was the acceptor substrate (Supplementary Fig. 9b). These data indicated that the E301 residue on the outside surface of TPR domain is dispensable for binding and glycosylation of these protein substrates. In striking contrast, the reaction with the N321A/N322A mutant displayed complete loss of O-GlcNAcylation on OGA-D175N and significantly increased OGT labeling (Supplementary Fig. 12b, Supplementary Note 3), a pattern in line with abolished binding of the acceptor substrate. In addition, our radiolabeled kinetic assay did not detect any glycosylation activity of this N321A/N322A mutant on OGA-D175N. Taken together, we demonstrated that GEP1A could be applied to characterize the binding of OGT with different protein substrates, accelerating the discovery of crucial residues for substrate recognition.

Crosslinking of OGT with acceptor substrates *in situ*

Another challenge with investigating the substrate recognition of OGT has been identifying genuine substrates that transiently or weakly interact with this enzyme. A previous study revealed that GlcNAc transfer involves limited movement of the sugar moiety from donor substrate to acceptor product (Supplementary Fig. 13).[15] We hypothesized that if GEP1 reacted with OGT C917 before glycosylation, it would crosslink OGT with acceptor substrate *in situ* during the sugar transfer step. To test this hypothesis, $OGT_{4.5}$ was briefly incubated with GEP1 before adding each of the biotinylated peptide substrates: RBL2, Lamin B1, and IRS1.[23] Indeed, we successfully detected crosslinked peptides at the molecular weight of $OGT_{4.5}$ using western blot with streptavidin-HRP (Supplementary Fig. 14a). Control experiments with the C917A mutant of $OGT_{4.5}$ or UDP-GlcNAc instead of GEP1 did not produce any detectable crosslinking, supporting the remarkable specificity of GEP1. To provide unambiguous evidence for GEP1-mediated crosslinking of OGT with substrates, we solved a crystal structure of crosslinked OGT:GEP1:CKII complex (2.6 Å) following a brief preincubation of $OGT_{4.5}$ with GEP1 (Fig. 5a and Supplementary Table 2). The asymmetric unit of the crystal contained two molecules of $OGT_{4.5}$. While GEP1-derived O-GlcNAcylation of CKII peptide was observed in one molecule, the other molecule depicted the GEP1-induced crosslinking of $OGT_{4.5}$ with CKII. This crosslinking was clearly defined by the continuous electron density observed from the OGT C917 side chain to GEP1 and further extended to the glycosylating serine of the CKII peptide (Fig. 5b). Notably, when

superimposed with the reported OGT:UDP:O-GlcNAcylated CKII complexed structure (PDB 4GYW), the crosslinked OGT:GEP1:CKII complex displayed identical binding conformations of the sugar and peptide substrates (Fig. 5c), supporting that GEP1 preincubation can capture OGT-substrate complexes *in situ*.

To examine the potential of GEP1 crosslinking OGT with protein substrates, we incubated full-length OGT with NUP62 following a similar preincubation strategy. The putative crosslinked complexes were detected as shifted bands on a SDS-PAGE gel (Supplementary Fig. 14b). In-gel tryptic digestion and LC-MS/MS analysis verified that the shifted bands were indeed OGT and NUP62 crosslinked complexes (Supplementary Table 3). Moreover, we demonstrated by western blot that GEP1 was capable of crosslinking full-length OGT with a number of proteins in MCF7 nuclear extracts (Supplementary Fig. 14c). Further quantitative LC-MS/MS analysis identified 102 potential OGT substrates from the extracts that were significantly enriched by wild-type but not C917S mutant of OGT following the preincubation with GEP1. Notably, 72 of these proteins (71%) have been reported as O-GlcNAcylated proteins (Supplementary Fig. 14d and Supplementary Table 4).[33–39] Hence, the crosslinking method opens a new avenue for identifying genuine substrates that transiently or weakly interact with OGT.

DISCUSSION

The lack of complete structural information of full-length OGT with bound protein substrate, and the paucity of efficient approaches for functional characterization of OGT-substrate recognition has hindered the investigation of OGT substrate specificity. We reported here the development of GEPs that can O-GlcNAcylate proteins or specifically label a unique active-site cysteine of OGT *in vitro*, depending on the relative reaction rates. OGT mutation-induced changes, as quantified by click chemistry-based fluorescent detection, has been exploited to report the impacts on sugar binding or protein substrate binding in an accelerated process to discover OGT residues crucial for protein recognition. Notably, we discovered that the N321/N322 residues localized on the inner surface of the 10^{th} TPR of OGT (Supplementary Fig. 8) made significant contributions to the interaction with NUP62 and OGA protein substrates. These asparagine residues are distant from the active site that was occupied by short peptide substrates,[15,16,23] implying an unusual recognition mode in which the flexible region of the protein substrate could penetrate part, if not all, of the super-helical TPR domain of OGT for efficient binding and glycosylation.[14,40] Since the TPR domain of OGT comprises over 500 amino acids, the GEP1A fluorescent assay provides a rapid screen of OGT mutants to identify important residues for more detailed characterization (*e.g.*, kinetic assays). Therefore, this new method is expected to substantially accelerate the discovery of key residues of OGT for protein substrate recognition.

As the N321A/N322A mutation impaired OGT binding with NUP62 and OGA substrates (Fig. 4c and Supplementary Fig. 12b), we propose that the asparagine residues on the inner surface of TPR domain are a generic binding region contributing to OGT interactions with different acceptor substrates. Since this OGT mutant retained partial glycosylation activity towards NUP62 but not OGA (Fig. 4c and Supplementary Fig. 12b), the significance of

N321/N322 residues appears to depend on the substrate, which may confer the substrate specificity of OGT. Moreover, these experiments highlighted a notable feature of GEP1A that is complementary to the conventional activity-based assays: even when no glycosylation is detected, GEP1A fluorescent assay can potentially still discriminate the impact of the mutated OGT residues on sugar binding versus protein substrate binding.

Furthermore, we discovered that preincubation of GEP1 induced crosslinking of OGT with a variety of acceptor substrates. When optimized, the crosslinking strategy is expected to afford more specific and straightforward detection of OGT substrates than many established methods in the field, such as antibody-,[41] lectin-,[42] or click chemistry-based methods.[43] In fact, it has been reported that sequence unrelated eOGT can O-GlcNAcylate a number of membrane proteins[44,45] and the commonly used metabolic labeling approach with Ac_4GlcNAz not only enriches O-GlcNAc but also S-GlcNAc[46] and O-GalNAc[47] modified proteins. Hence, the novel crosslinking method will be invaluable for identifying genuine substrates that transiently or weakly interact with OGT. We note that the allyl chloride group might be suboptimal for some protein substrates and that future work will be required to determine its generality and to evaluate additional analogues. In summary, GEPs provide a set of powerful innovative strategies to accelerate the characterization of OGT-substrate binding and recognition, which will promote drug discovery in this up-and-coming field.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ONLINE METHODS

Compound synthesis

Compound synthesis is described in the Supplementary Note.

NUP62 protein expression and purification

Human NUP62 expression plasmid in pET21a vector (a kind gift from Dr. Suzanne Walker's lab) was transformed into *E. coli* BL21(DE3) competent cells and expressed as a fusion protein containing a C-terminal His_6-tag. The transformant was grown at 37 °C in LB medium. After OD_{600} reached 0.4, the culture was then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h. Cells were pelleted and resuspended in lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5 M NaCl, and 1 mM PMSF). The cell suspension was lysed with ultra-high-pressure cell disrupter Emulsiflex-C5 (Avestin) followed by centrifugation at 20,500 g at 4 °C for 30 min. The collected pellets containing the inclusion body of NUP62 were then washed twice with 30 mL detergent solution (1.1 M urea, 2% Triton X-100) and resuspended in 30 mL denaturing buffer (8 M urea, 50 mM Tris pH 8.0, 1 mM EDTA, 2 mM DTT) and incubated at room temperature (RT) to dissolve the inclusion body. The lysate was then centrifuged at 16,100 g at 20 °C for 20 min to remove the unbroken pellets. The supernatant was centrifuged at 50,000 g at 20 °C for 15 min to further clear up the lysate. Proteins were dialyzed against 2 M urea in 20 mM Tris pH 7.5 for 4 h followed by 20 mM Tris pH 7.5 overnight at 4 °C. Following dialysis, samples were

centrifuged at 16,100 g at 4 °C for 20 min, the supernatant containing NUP62 was collected, concentrated, and stored at −80 °C until use. The purity of NUP62 protein was examined by SDS-PAGE gel.

$OGT_{4.5}$ protein purification for crystallization

The $OGT_{4.5}$ expression plasmid was a kind gift from Dr. Suzanne Walker's lab. Briefly, the $OGT_{4.5}$ construct (spanning residues 313–1031 based on the numbering of the full-length human protein) was cloned into a pET24b vector. The plasmid was transformed into *E. coli* BL21(DE3) competent cells, and the transformant was grown at 37 °C in LB medium. After OD_{600} reached 1.0, the culture was induced by 0.2 mM IPTG at 16 °C for 16 h. The cells were pelleted, re-suspended in TBS buffer (20 mM Tris pH 8.0, 150 mM NaCl) supplemented with 1 mM PMSF, and lysed with ultra-high-pressure cell disrupter. After centrifugation, the supernatant was subjected to Ni-NTA column for affinity purification. The recombinant protein was subsequently eluted with buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl, 250 mM imidazole and 0.5 mM Tris(3-hydroxypropyl)phosphine (THP). The N-terminal His_6-tag was cleaved by HRV3C protease. Further purification was performed by size-exclusion chromatography (Superdex 200 increase 10/300, GE Healthcare) in the buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl, and 0.5 mM THP. The purified protein was concentrated to 8 mg/mL for crystallization.

Crystallization

All crystals were generated by mixing 1 μL of protein with 1 μL of reservoir solution and were equilibrated against 160 μL of reservoir solution using the hanging-drop vapor-diffusion method at 20 °C. For the glycosylated OGT:GEP1:CKII complex, $OGT_{4.5}$ was incubated with 1 mM GEP1 and 2 mM CKII peptide for 1 h on ice. Crystals were obtained in the reservoir solution containing 0.08 M BIS-TRIS propane (pH 7.0), 0.02 M sodium cacodylate trihydrate (pH 6.5), 2.8 M sodium formate, 0.04 M ammonium sulfate, and 6% w/v polyethylene glycol 8,000. For the crosslinked OGT:GEP1:CKII complex, $OGT_{4.5}$ was preincubated with 1 mM GEP1 for 30 min at RT before adding 2 mM CKII peptide and incubated for another 2 h on ice. Crystallization condition contains 0.08 M Tris (pH 8.5), 0.02 M sodium cacodylate trihydrate (pH 6.5), 1.24 M ammonium sulfate, and 6% w/v polyethylene glycol 8,000. All crystals were transferred into cryoprotectant solution containing their respective mother liquor plus 10% (v/v) glycerol, before being flash-frozen in liquid nitrogen for storage.

Data collection and X-ray structure determination

All X-ray data were collected on the Life Sciences Collaborative Access Team (LS-CAT) beamline 21-ID-D at Advanced Photon Source, Argonne National Laboratory, IL. The wavelength for data collection was 0.9785 Å. Datasets were processed using HKL2000 package.[48] The crystal of glycosylated OGT:GEP1:CKII belongs to the space group of *F*222 and contains one molecule per asymmetric unit, with cell parameters: a = 139.1 Å, b = 152.7 Å, c = 199.4 Å, $\alpha = \beta = \gamma = 90°$. The crystal of crosslinked OGT:GEP1:CKII complex belongs to the space group of *C*2221 and has two molecules per asymmetric unit. The structures were solved by molecular replacement, using OGT as a search model (PDB 3PE3, ref.[16]). Iterative model building was performed in COOT[49] and refinement was completed in

PHENIX[50] and CCP4.[51] Final statistics were summarized in Supplementary Table 2. All structural figures were prepared using the program PyMOL (DeLano Scientific, http://www.pymol.org/). The coordinates and structure factors have been deposited with PDB codes 5VIF and 5VIE.

Mutagenesis and purification of OGA mutant

OGA-D175N and various OGT mutants were generated using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions. Human OGA and full-length OGT (plasmids were kind gifts from Dr. Suzanne Walker's lab) were used as the DNA templates along with the primers listed in Supplementary Table 5. The DNA sequences were verified by sequencing. Wild-type full-length OGT and mutant proteins were expressed and purified similarly as $OGT_{4.5}$ as mentioned above. For OGA-D175N, the mutant plasmid was transformed into *E. coli* Rosetta (DE3) competent cells, and the transformant was grown at 37 °C in LB medium. After OD_{600} reached 0.6, the culture was induced by 0.3 mM IPTG at 16 °C for 16 h. The cells were pelleted, lysed, and purified by Ni-NTA column followed by size-exclusion chromatography as mentioned above. The purified protein was concentrated to 5 mg/mL and stored at −80 °C before use.

Intact protein MS to detect the covalent modification of GEP1/GEP1A on OGT

A typical reaction was set up as below: purified OGT (20 μM wild-type $OGT_{4.5}$ or mutant) was incubated in a 25 μL reaction containing 200 μM of GEP1/GEP1A in the reaction buffer (10 mM Tris, pH 8.0, 75 mM NaCl, 0.5 mM THP, 200 U/mL CIP alkaline phosphatase, and 30 mM $MgCl_2$) at RT for 6 h followed by C8 StageTip[52] desalting. Samples were SpeedVac dried and re-dissolved in 20 μL 0.1% formic acid. Mass measurement of intact protein was performed on Q-TOF Maxis 4G (Bruker) or Impact II (Bruker) coupled with an ACQUITY UPLC (Waters). Samples were loaded onto a 2.1 × 100 mm BEH C4 column (1.7 μm, 300 Å, Waters). The injection volume was 9 μL with a flow rate of 300 μL/min. The mobile phases consisted of 0.1% formic acid (solvent A) and 0.1% formic acid in 95% acetonitrile (ACN) (solvent B). LC program: 5% B for 10 min, 5–60% B for 15 min, 60–90% B for 1 min, 90% B for 5 min, and 90–5% B for 1 min. MS analysis was operated in positive mode with electrospray voltage of 3.8 kV. The end plate offset and nebulizer pressure were −500 V and 2.1 bar, respectively. The interface heater temperature was set at 220 °C with the dry gas flow rate 10 L/min. Data was acquired using one full MS scan (*m/z* 700–3,000) with the scan rate at 1 Hz. Funnel 1 RF and multiple RF were set to 400 eV for ion transfer. The ion energy of quadrupole was 3 eV, the collision energy was 6 eV, the transfer time was 120 μs, and the pre pulse storage time was 25 μs. Time-course experiments were performed similarly as above by incubating the reactions at 37 °C in two replicates for each indicated time points. Dose-dependent experiments were carried out with 40 or 200 μM of GEP1 in a 10 μL reaction for 6 h at 37 °C. LC-MS data were processed and analyzed using Compass Data Analysis software (version 4.1, Bruker). The major LC peak corresponding to highest UV signal was selected for deconvolution performed by maximum entropy algorithm. The parameters of maximum entropy include mass range 50,000–100,000 Da, auto data point spacing, and the resolving power of 10,000. The increased mass resulting from labeling of GEP1/GEP1A was determined by subtracting the original protein mass from the deconvoluted protein mass.

Sample preparation for detection of GEP1/GEP1A-modified peptides

To detect GEP1/GEP1A-labeled OGT peptides, purified full-length OGT protein (70 μM) was incubated in a 10 μL reaction containing 350 or 700 μM of GEP1/GEP1A in the reaction buffer (10 mM Tris, pH 8.0, 75 mM NaCl, 0.5 mM THP, 200 U/mL CIP alkaline phosphatase, and 30 mM $MgCl_2$) at RT. The reaction was stopped at 6 h or other indicated time points by 10 mM DTT and 8 M urea. For trypsin digestion, the protein sample was diluted to 0.5 μg/μL in denature buffer (8 M urea and 10 mM Tris, pH 8.0). After reduction by 10 mM DTT for 30 min at RT and carbidomethylation with 55 mM iodoacetamide in the dark for 30 min at RT, alkylated proteins were seven-fold diluted using 25 mM NH_4HCO_3 and then digested by Trypsin/Lys-C mix (protein:protease = 40:1) (Promega) at RT for 18 h. Formic acid was added to the sample to reach final 0.5% for enzyme inactivation and sample acidification. Peptides were SpeedVac dried after SDB-XC StageTip desalting. To measure the irreversible reaction rate of GEP1 with C917, peptides (5 μg) from each reaction were re-dissolved in 50 μL of 0.1 M triethylammonium bicarbonate (TEAB) for dimethyl labeling.[53] The peptide samples prepared from the reactions of OGT with GEP1 at concentration ratios of 1:5 and 1:10 were mixed with 2 μL of 4% formaldehyde-H_2 (Sigma-Aldrich) and 4% formaldehyde-D_2 (Sigma-Aldrich), respectively. Each sample was then mixed with freshly prepared 0.6 M sodium cyanoborohydride (2 μL) and then incubated for 1 h at RT. The reaction was quenched by addition of 8 μL of 1% NH_4OH. Formic acid (10%, 10 μL) was added to further stop the reaction and acidify the samples. Finally, the H- and D-labeled samples were combined at 1:1 ratio followed by SDB-XC StageTips desalting and SpeedVac dry. To detect the GEP1/GEP1A-glycosylated peptides, α-crystallin B chain peptide (200 μM, peptide sequence: 38–50, FPTSTSLSPFYLR, synthesized by Biomatik) or Lamin B1 peptide (200 μM, peptide sequence: 389–401, KLSPSPSSRVTVSK-biotin, synthesized by Biomatik) was incubated with 20 μM purified OGT in a 5 μL reaction containing 1 mM GEP1/GEP1A in the reaction buffer (10 mM Tris, pH 8.0, 75 mM NaCl, 0.5 mM THP, 200 U/mL CIP alkaline phosphatase, and 30 mM $MgCl_2$) at RT for 6 h. Peptide samples were then purified by SDB-XC StageTips and dried by SpeedVac. All peptide samples were stored at −80 °C until LC-MS/MS analysis.

nanoLC-MS/MS analysis

Peptide samples were dissolved in 0.1% formic acid for LC-MS/MS analysis on an Orbitrap Q-Exactive (Thermo Scientific) equipped with a nanoAcaquity UPLC system (Waters). Peptides were loaded onto a 75 μm × 15 cm 1.7 μm BEH C18 column at a flow rate of 300 nL/min. Mobile phase A consisted of 0.1% formic acid, and solvent B was 0.1% formic acid in ACN. A linear gradient of 0–4% B for 0.1 min, 4–35% B for 30 min, 35–75% B for 0.1 min, 75% B for 9.5 min, 75–95% B for 0.1 min, 95% B for 9.5 min, 95–0% B for 0.5 min, and 0% B for 9.5 min was employed throughout this study. Mass spectra from full scans were acquired in a data-dependent mode (*m/z* 200–2,000). The resolution of survey scan was set to 17,500 at *m/z* 400 with an automated gain control (AGC) value of 10^6. The top 15 most-intense precursor ions were selected from the MS scan for subsequent higher energy collisional dissociation (HCD, normalized collision energy 30 eV) MS/MS scan. Peptide identification was performed by Mascot (v2.4, Matrix) and MaxQuant (v1.5.6.5 or v1.5.8.3)[54,55] against a composite target-decoy protein sequence database containing Uniprot database (release 2015_04, subset human, 20,265 protein entries).[56] The search

criteria used in this study include trypsin specificity allowing up to 2 missed cleavages, and variable modifications of GEP1 ($C_{10}H_{15}NO_6$) on Cys, GEP1A ($C_{10}H_{14}N_4O_5$) on Cys, GEP1 glycosylation ($C_{10}H_{14}ClNO_5$) on Ser/Thr, GEP1A glycosylation ($C_{10}H_{13}ClN_4O_4$) on Ser/Thr, carbamidomethyl on Cys, and oxidation on Met. The precursor mass tolerance and the fragment ion tolerance were set at ± 10 ppm and ± 0.6 Da, respectively. Peptide was considered identified based on the posterior error probability with a false discovery rate of 1%. The spectra of peptides were manually inspected.

To determine the reaction rate of GEP1 with OGT, raw MS spectra from two biological samples were processed using MaxQuant and precursor intensities was calculated. The search criteria and database were set up as described above. Peptide abundance was normalized to the total protein intensity of OGT detected in the same sample. Based on the previous report,[22] the reaction rate (k) of GEP1 with C917 of OGT in a second-order reaction can be calculated using eq. 1.

$$ln\left([P]\,[X_0]\,/\,[X]\,[P_0]\right) = -\,kt\,[X_0] + kt\,[P_0] \quad \text{[eq. 1]}$$

Where $[P_0]$ is the initial abundance of unmodified C917 peptide, $[X_0]$ is the initial concentration of GEP1, $[P]$ is the abundance of unmodified C917 peptide at time t, $[X]$ is the concentration of GEP1 at time t, which can be estimated using eq. 2.

$$X = X_0 \bullet \frac{(P_0 - P)}{P_0} \quad \text{[eq. 2]}$$

GEP1A fluorescent assay

To detect the modification of GEP1A on OGT and NUP62 protein substrate, purified OGT (1 μM of $OGT_{4.5}$, full-length OGT, or mutants) was incubated with NUP62 (5 μM) and GEP1A (25 μM) in a 12 μL reaction containing buffer (20 mM Tris pH 8.0, 150 mM NaCl, and 0.5 mM THP) and incubated at 37 °C for 30 min. For competition assay, different doses of UDP-GlcNAc ranging from 0–50 μM were added into the reaction mixture together with GEP1A. The reactions were quenched by boiling at 95 °C for 5 min. SDS (1%) was added and incubated at RT for 5 min to help dissolve precipitated proteins. The samples were further diluted to final 0.1% SDS with TBS buffer. Click chemistry reagents at the final concentration of 1 mM $CuSO_4$, 0.1 mM Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA), 50 μM fluor 488-alkyne and 1 mM sodium ascorbate were sequentially mixed and immediately added to each sample. Click chemistry reaction was performed at RT for 1 h in the dark. Proteins were precipitated in ice-cold methanol for 2 h at −80 °C followed by 16,000 g for 10 min to pellet the proteins. The protein pellet was washed by methanol and then re-dissolved in TBS buffer (20 mM Tris pH 8.0 and 150 mM NaCl) containing 4% SDS. The dissolved protein samples were separated on SDS-PAGE gel, followed by fluorescence detection and Coomassie Blue staining for relative quantification. Both in-gel fluorescence scanning and Coomassie Blue stained gels were detected using ChemiDoc-It 2 imager equipped with BioLite MultiSpectral light source (UVP) and quantified using ImageJ (v1.6.0_24).[57] The levels of GEP1A-modified proteins were normalized to the protein

amount and relatively quantified. For the detection of GEP1A modification on OGT and OGA-D175N protein substrate, purified OGT (0.35 μM of $OGT_{4.5}$, full-length OGT, or mutants) was incubated with OGA-D175N (16 μM) and GEP1A (25 μM) in a 12 μL reaction buffer (TBS with 0.5 mM THP) and incubated at 37 °C for 15 min. Subsequent click chemistry reactions were performed following the same protocol mentioned above. The reaction mixtures were then boiled with SDS loading buffer at 95 °C for 5 min without protein precipitation. In-gel fluorescence scanning and relative quantification were performed as mentioned above.

Radiolabeled kinetic analyses of OGT mutants

Purified OGT (45 nM for wild-type or N321A/N322A mutant, or 150 nM for D554N mutant) was incubated with 100 μM UDP-^{3}H-GlcNAc (92 mCi/mmol, PerkinElmer) at indicated concentrations of NUP62 protein in the reaction buffer (20 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM THP) at 37 °C for 0.5 or 1 h. Reactions were quenched by transferring the samples onto the nitrocellulose membrane, air-dried, and washed 5 min for four times in PBS buffer. The radioactivity on each nitrocellulose membrane was counted using Tri-Carb 2900TR liquid scintillation analyzer (PerkinElmer). A reaction in the absence of OGT was set up as negative control. Another reaction without the washing step was used to calculate the levels of enzymatic conversion. Three independent experiments were conducted for each condition. Data were analyzed using GraphPad Prism v5 (GraphPad Software).

MCF7 nuclear extracts

The MCF7 cell line was purchased from American Type Culture Collection (ATCC). The cell line has been authenticated following the guidelines from ATCC and was tested to be free of mycoplasma contamination. We periodically analyzed the growth curve and checked the morphology under microscopy to ensure the culture consistency and no adverse effect on cell behavior. MCF7 cells were maintained in Dulbecco's Modified Eagle Medium with 10% fetal bovine serum (FBS) at 37 °C in a 5% CO_2 incubator. Cells were plated at 70% confluence and starved in 2% FBS medium without glucose for 24 h before harvesting. Cell pellets were washed by PBS and lysed in buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, and 10 mM KCl). The sample was centrifuged at 10,000 g for 20 min at 4 °C to pellet nucleus. The nuclear proteins were extracted by resuspending the nuclear pellet in buffer (20 mM HEPES, pH 7.9, 25% (v/v) glycerol, 0.42 M NaCl, 1.5 mM $MgCl_2$, and 0.2 mM EDTA) and incubated at 4 °C for 1 h. Cell debris was then removed by centrifugation at 16,000 g for 10 min at 4 °C. The nuclear extracts were stored at −80 °C in aliquots before use.

GEP1 preincubation-induced crosslinking of OGT with substrates

To detect the crosslinking of OGT with substrates, purified OGT (1 μM) was preincubated with 100 μM GEP1 in the reaction buffer (10 mM Tris, pH 8.0, 75 mM NaCl, 1 mM THP, 100 U/mL CIP alkaline phosphatase, and 30 mM $MgCl_2$) for 30 min at 37 °C, followed by adding 5 μM of NUP62 protein or 100 μM of biotinylated peptides to react at 37 °C for another 3 h. The reactions were analyzed by SDS-PAGE gel or western blot. Biotinylated peptide RBL2 (410–422, KENSPAVTPVSTAK-biotin), Lamin B1 (389–401, KLSPSPSSRVTVSK-biotin), and IRS1 (982–994, VPSSRGDYMTMQMK-biotin) were custom synthesized at Biomatik. For OGT crosslinking with proteins from the MCF7

nuclear extracts, full-length OGT (0.2 μM) and 50 μM GEP1 were prepared similarly as described above and reacted with 40 μg of nuclear extracts at 37 °C for 5.5 h before western blot analysis. Nitrocellulose membrane (Life Technology) was blocked with 0.9% bovine serum albumin for 1 h and then probed with streptavidin-HRP (Sigma-Aldrich #GERPN1231) or anti-OGT antibody (Proteintech #11576-2-AP). All primary antibodies were used at 1:5,000 dilution. Secondary anti-lgG-HRP antibody (Proteintech #SA00001-2) was used at 1:50,000 dilution. Crosslinked peptides and proteins were detected using an enhanced chemiluminescence detection kit (Life Technology) on ChemiDoc XRS+ imager (Bio-Rad).

In-gel tryptic digestion of OGT-NUP62 crosslinked complexes

Each of the two major crosslinking bands (labeled as “1” and “2”) in the Coomassie Blue stained gel (Supplementary Fig. 14b) was excised into 1 mm^3 pieces for in-gel tryptic digestion. After de-staining and extensive wash steps using 50% ACN in 50 mM NH_4HCO_3, proteins were reduced by 10 mM DTT for 1 h at RT and alkylated by 55 mM iodoacetamide at RT for 1 h. Samples were then dehydrated by 100% ACN, vacuumed dried, rehydrated in 25 mM NH_4HCO_3, and digested with 0.05 μg/μL trypsin/Lys-C at RT for 18 h. The digestion was quenched by 0.5% formic acid. The tryptic peptides were recovered by 80% ACN/0.1% formic acid and subjected to SDB-XC StageTip desalting. Peptides were then vacuum dried and stored at −80 °C until use. nanoLC-MS/MS analysis and peptide identification were conducted as mentioned above.

Quantitative LC-MS/MS analysis of GEP1 preincubation-induced crosslinking of OGT with nuclear proteins

For crosslinking, purified His-tagged WT or C917S mutant of OGT (7.5 μM) was preincubated with 1 mM GEP1 in the reaction buffer for 10 min at RT, followed by reacting with 1 mg of MCF7 nuclear extracts (1 mg/mL) for another 12 h at RT. The samples were then subjected to affinity purification by Ni-NTA column as described above. The eluents containing OGT and its crosslinked complexes were precipitated by methanol and chloroform, and then re-dissolved in 8 M urea/50 mM TEAB. Proteins were subjected to in-solution digestion, desalting, and dried as described above. For dimethyl labeling, peptides were re-dissolved in 40 μL TEAB. The C917S and WT samples were first mixed with 1.6 μL formaldehyde-H_2 (light) and 4% formaldehyde-D_2 (heavy), respectively. Freshly prepared 0.6 M sodium cyanoborohydride (1.6 μL) (Sigma-Aldrich) was then added to each sample to mix well and incubated at RT for 1 h. The reactions were quenched by adding 6.4 μL of 1% NH_4OH followed by 8 μL of 10% formic acid as mentioned above. H- and D-labeled samples were mixed and subjected to SDB-XC StageTips for desalting. Peptides were vacuumed dried and stored at −80 °C until use. nanoLC-MS/MS and data analysis were performed as described above. Two biological samples (S1 and S2) and two technical repeats were applied for reliable quantitation. Precursor intensities were recalculated by MaxQuant using the “match between runs” option. All data were normalized to OGT intensity ($OGT_{[WT/C917S]} = 1$). Identified proteins with the relative ratios of [WT/C917S] > 1.5 in both biological samples were considered as potential substrates of OGT in the crosslinked complexes.

Statistical analysis

All of the data are shown as mean values with error bars representing ± s.d. Statistical significance was determined using Student's *t*-test. Significance is indicated as $*P < 0.01$.

Data availability

The data that support the findings of this study are available in the Supplementary Information, Supplementary Notes, and from the corresponding author upon reasonable request. X-ray crystallographic data that support the findings of this study have been deposited in the Protein Data Bank with the accession codes 5VIF and 5VIE for glycosylated OGT:GEP1:CKII and crosslinked OGT:GEP1:CKII complexes, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Suzanne Walker's lab for kindly sharing the expression plasmids of OGT, $OGT_{4.5}$, OGA, and NUP62. We would like to acknowledge the research funding support from University of Wisconsin-Madison (to J.J.), a Vilas Research Investigator Award (to J.J.), NIH R01 GM121718 (to J.J.), NIH R21 AG055377 (to L.L.), NIH R01 GM117058 (to Y.G.), and NIH R01 HL109810 (to Y.G.). We also thank NIH Shared Instrument Program Grant S10 RR029531 and high-end instrument grant S10 OD018475 for funding the MS instruments.

References

1. Vocadlo DJ. O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and enzyme regulation. Curr Opin Chem Biol. 2012; 16:488–497. [PubMed: 23146438]
2. Hardiville S, Hart GW. Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation. Cell Metab. 2014; 20:208–213. [PubMed: 25100062]
3. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. Annu Rev Biochem. 2011; 80:825–858. [PubMed: 21391816]
4. Capotosti F, et al. O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. Cell. 2011; 144:376–388. [PubMed: 21295698]
5. Ma Z, Vocadlo DJ, Vosseller K. Hyper-O-GlcNAcylation is anti-apoptotic and maintains constitutive NF-κB activity in pancreatic cancer cells. J Biol Chem. 2013; 288:15121–15130. [PubMed: 23592772]
6. 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. 2000; 97:5735–5739. [PubMed: 10801981]
7. Howerton CL, Bale TL. Targeted placental deletion of OGT recapitulates the prenatal stress phenotype including hypothalamic mitochondrial dysfunction. Proc Natl Acad Sci USA. 2014; 111:9639–9644. [PubMed: 24979775]
8. Ma J, Hart GW. Protein O-GlcNAcylation in diabetes and diabetic complications. Expert Rev Proteomics. 2013; 10:365–380. [PubMed: 23992419]
9. de Queiroz RM, Dias WB, Carvalho E. O-GlcNAcylation: the sweet side of the cancer. Front Oncol. 2014; 4:132. [PubMed: 24918087]
10. Yuzwa SA, Vocadlo DJ. O-GlcNAc and neurodegeneration: biochemical mechanisms and potential roles in Alzheimer's disease and beyond. Chem Soc Rev. 2014; 43:6839–6858. [PubMed: 24759912]
11. Laczy B, et al. Protein O-GlcNAcylation: a new signaling paradigm for the cardiovascular system. Am J Physiol Heart Circ Physiol. 2009; 296:H13–28. [PubMed: 19028792]

12. Ma J, Hart GW. O-GlcNAc profiling: from proteins to proteomes. Clin Proteomics. 2014; 11:8. [PubMed: 24593906]
13. Levine ZG, Walker S. The biochemistry of O-GlcNAc transferase: which functions make it essential in mammalian cells? Annu Rev Biochem. 2016; 85:631–657. [PubMed: 27294441]
14. Iyer SPN, Hart GW. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. J Biol Chem. 2003; 278:24608–24616. [PubMed: 12724313]
15. Lazarus MB, et al. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. Nat Chem Biol. 2012; 8:966–968. [PubMed: 23103939]
16. Lazarus MB, Nam YS, Jiang J, Sliz P, Walker S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature. 2011; 469:564–567. [PubMed: 21240259]
17. Pierce MM, Raman CS, Nall BT. Isothermal titration calorimetry of protein-protein interactions. Methods San Diego Calif. 1999; 19:213–221.
18. Hoa XD, Kirk AG, Tabrizian M. Towards integrated and sensitive surface plasmon resonance biosensors: a review of recent progress. Biosens Bioelectron. 2007; 23:151–160. [PubMed: 17716889]
19. Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends Biochem Sci. 2016; 41:20–32. [PubMed: 26654279]
20. Jiang J, Lazarus MB, Pasquina L, Sliz P, Walker S. A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. Nat Chem Biol. 2012; 8:72–77.
21. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. 2008; 77:521–555. [PubMed: 18518825]
22. Kaur P, Kiselar J, Yang S, Chance MR. Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (MS). Mol Cell Proteomics. 2015; 14:1159–1168. [PubMed: 25687570]
23. Pathak S, et al. The active site of O-GlcNAc transferase imposes constraints on substrate sequence. Nat Struct Mol Biol. 2015; 22:744–750. [PubMed: 26237509]
24. Kreppel LK, Hart GW. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. J Biol Chem. 1999; 274:32015–32022. [PubMed: 10542233]
25. Mayer A, Gloster TM, Chou WK, Vocadlo DJ, Tanner ME. 6”-Azido-6”-deoxy-UDP-N-acetylglucosamine as a glycosyltransferase substrate. Bioorg Med Chem Lett. 2011; 21:1199–1201. [PubMed: 21273069]
26. Chuh KN, Zaro BW, Piller F, Piller V, Pratt MR. Changes in metabolic chemical reporter structure yield a selective probe of O-GlcNAc modification. J Am Chem Soc. 2014; 136:12283–12295. [PubMed: 25153642]
27. Davis LI, Blobel G. Nuclear pore complex contains a family of glycoproteins that includes p62: glycosylation through a previously unidentified cellular pathway. Proc Natl Acad Sci USA. 1987; 84:7552–7556. [PubMed: 3313397]
28. Schimpl M, et al. O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. Nat Chem Biol. 2012; 8:969–974. [PubMed: 23103942]
29. Kumari M, et al. Exploring reaction pathways for O-GlcNAc transferase catalysis. A string method study. J Phys Chem B. 2015; 119:4371–4381. [PubMed: 25731954]
30. Lazarus BD, Love DC, Hanover JA. Recombinant O-GlcNAc transferase isoforms: identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. Glycobiology. 2006; 16:415–421. [PubMed: 16434389]
31. Rexach JE, et al. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. Nat Chem Biol. 2010; 6:645–651. [PubMed: 20657584]
32. Khidekel N, et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. Nat Chem Biol. 2007; 3:339–348. [PubMed: 17496889]
33. Zhao P, et al. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. J Proteome Res. 2011; 10:4088–4104. [PubMed: 21740066]

34. Lee A, et al. Combined antibody/lectin enrichment identifies extensive changes in the O-GlcNAc sub-proteome upon oxidative stress. J Proteome Res. 2016; 15:4318–4336. [PubMed: 27669760]
35. Hahne H, et al. Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. J Proteome Res. 2013; 12:927–936. [PubMed: 23301498]
36. Hahne H, Gholami AM, Kuster B. Discovery of O-GlcNAc-modified proteins in published large-scale proteome data. Mol Cell Proteomics. 2012; 11:843–850. [PubMed: 22661428]
37. Kamemura K, Hayes BK, Comer FI, Hart GW. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of Thr-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. J Biol Chem. 2002; 277:19229–19235. [PubMed: 11904304]
38. Malaker SA, et al. Identification of glycopeptides as posttranslationally modified neoantigens in leukemia. Cancer Immunol Res. 2017; 5:376–384. [PubMed: 28314751]
39. Phueaouan T, et al. Aberrant O-GlcNAc-modified proteins expressed in primary colorectal cancer. Oncol Rep. 2013; 30:2929–2936. [PubMed: 24126823]
40. Lazarus MB, et al. HCF-1 is cleaved in the active site of O-GlcNAc transferase. Science. 2013; 342:1235–1239. [PubMed: 24311690]
41. 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]
42. 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; 5:923–934. [PubMed: 16452088]
43. Vocadlo DJ, Hang HC, Kim EJ, Hanover JA, Bertozzi CR. A chemical approach for identifying O-GlcNAc-modified proteins in cells. Proc Natl Acad Sci USA. 2003; 100:9116–9121. [PubMed: 12874386]
44. Sakaidani Y, et al. O-linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell-matrix interactions. Nat Commun. 2011; 2:583. [PubMed: 22158438]
45. Müller R, Jenny A, Stanley P. The EGF repeat-specific O-GlcNAc-transferase eOGT interacts with notch signaling and pyrimidine metabolism pathways in Drosophila. PloS One. 2013; 8:e62835. [PubMed: 23671640]
46. Xiao H, Wu R. Global and site-specific analysis revealing unexpected and extensive protein S-GlcNAcylation in human cells. Anal Chem. 2017; 89:3656–3663. [PubMed: 28234450]
47. Zaro BW, Yang YY, Hang HC, Pratt MR. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. Proc Natl Acad Sci USA. 2011; 108:8146–8151. [PubMed: 21540332]

Online methods references

48. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
49. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
50. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948–1954. [PubMed: 12393927]
51. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol Crystallogr. 1994; 50:760–763. [PubMed: 15299374]
52. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007; 2:1896–1906. [PubMed: 17703201]
53. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009; 4:484–494. [PubMed: 19300442]
54. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008; 26:1367–1372. [PubMed: 19029910]

55. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016; 11:2301–2319. [PubMed: 27809316]

56. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014; 42:D191–198. [PubMed: 24253303]

57. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9:671–675. [PubMed: 22930834]

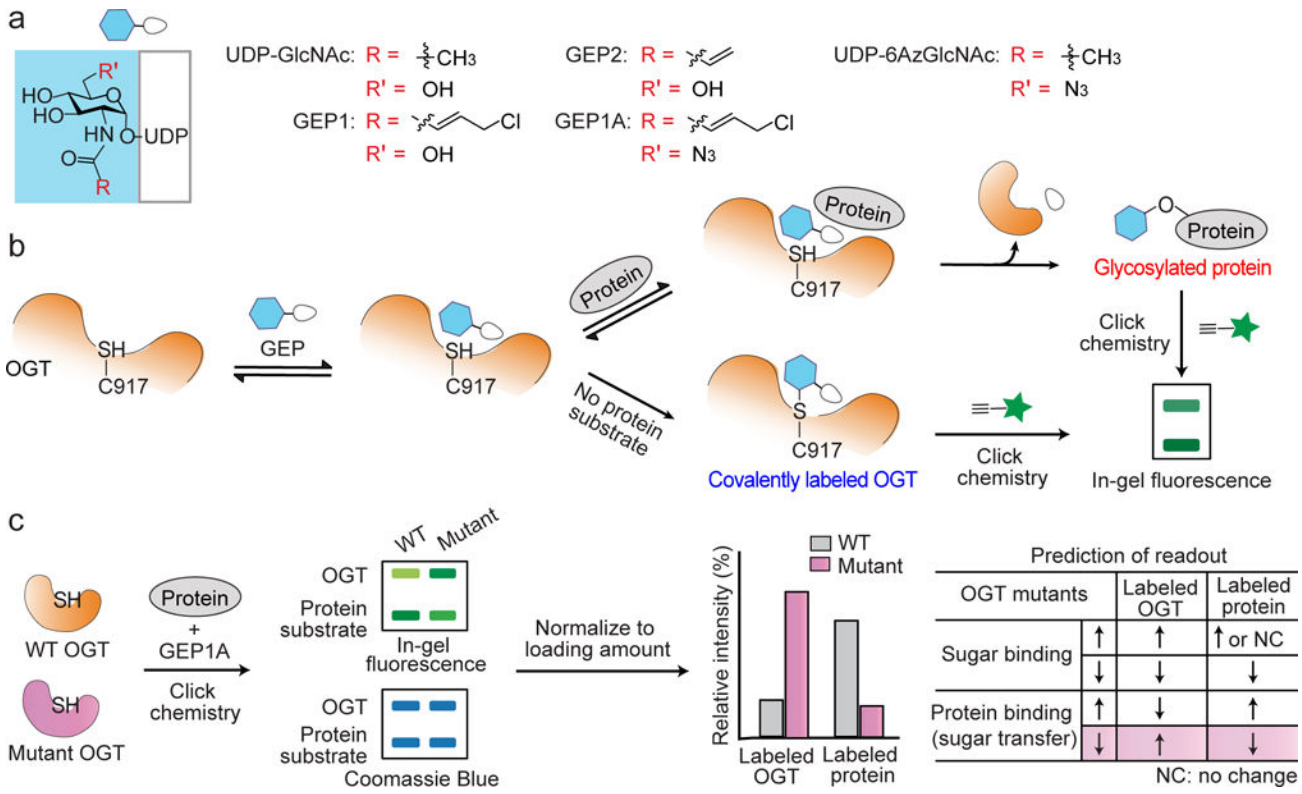

Figure 1.

The strategy of GlcNAc Electrophilic Probes (GEPs) for discerning altered OGT ability of sugar binding versus protein substrate binding. (**a**) The chemical structures of UDP-GlcNAc, UDP-6AzGlcNAc, and GEPs. (**b**) The reaction model of OGT and GEP in the presence or absence of protein substrate suggests generating differential modifications on OGT and the protein substrate. Modifications derived from GEP1A can be readily detected using click chemistry and in-gel fluorescence scanning. (**c**) OGT and its mutants are predicted to generate distinct levels of each modification following the reaction with GEP1A and a protein substrate. These pattern changes can be exploited to discriminate altered OGT ability of sugar binding versus protein substrate binding.

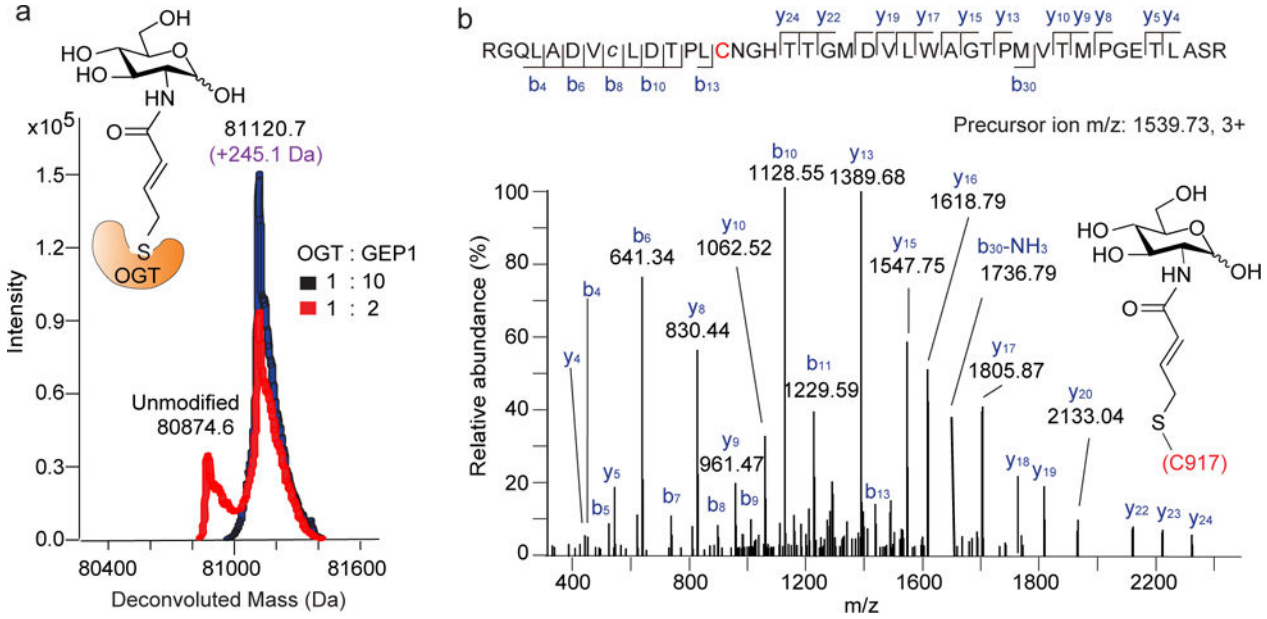

Figure 2.

MS data showed that GEP1 specifically labeled the C917 residue of OGT. (**a**) Overlay of the deconvoluted intact protein MS spectra from reactions of $OGT_{4.5}$ with GEP1 at concentration ratios of 1:2 and 1:10. (**b**) MS/MS spectrum mapped the modification site of GEP1 on C917 residue (highlighted in red) of full-length OGT. The carbamidomethylated cysteine residue was marked as *c*.

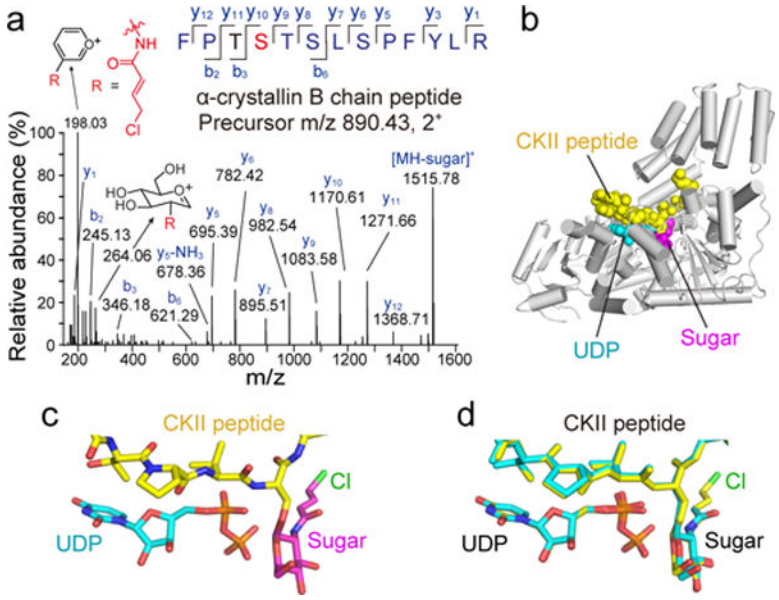

Figure 3.

LC-MS/MS and X-ray crystallography demonstrated GEP1-derived O-GlcNAcylation on peptide substrates of OGT. (**a**) MS/MS spectrum illustrated that GEP1 can be used by OGT to glycosylate α-crystallin B chain peptide. (**b**) X-ray crystal structure of the glycosylated OGT:GEP1:CKII complex with ligands highlighted in spheres. (**c**) Stick representation of GEP1-glycosylated CKII peptide and UDP in the active site of the glycosylated OGT:GEP1:CKII complex. The allyl chloride group of GEP1 was labeled as Cl. (**d**) Overlay of ligands from the glycosylated OGT:GEP1:CKII complex (yellow sticks) and OGT:UDP:O-GlcNAcylated CKII complex (cyan sticks) (PDB 4GYW) showed that GEP1-glycosylation displayed a similar conformation as regular O-GlcNAcylation and that the extended allyl chloride (labeled as Cl) can be tolerated by OGT.

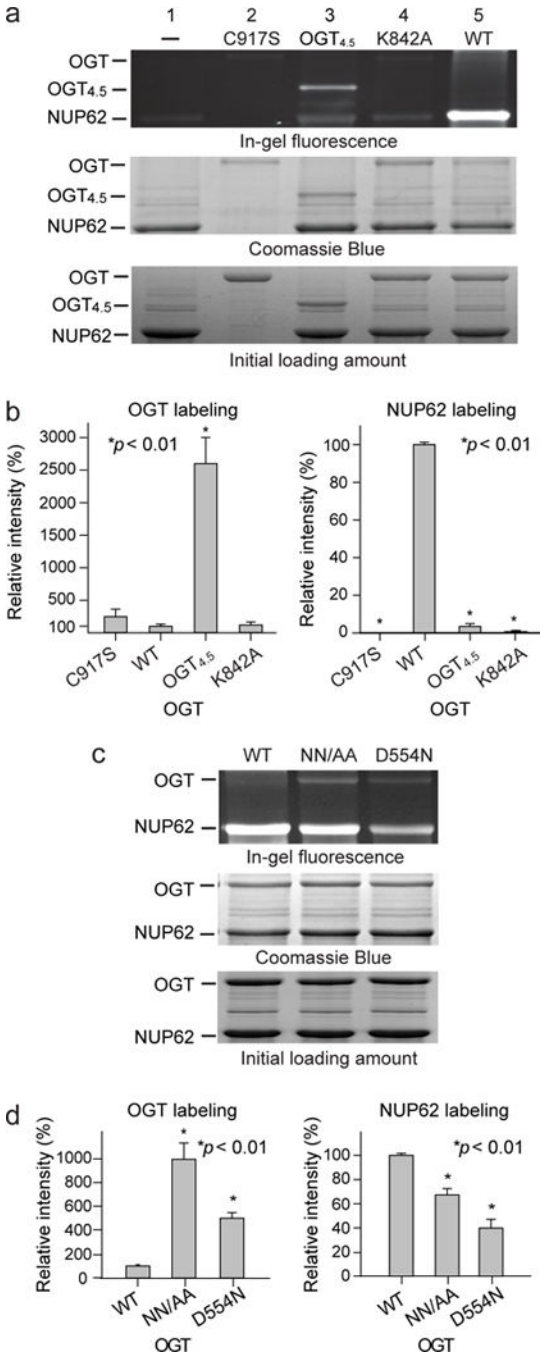

Figure 4.

Principle validation of GEP1A fluorescent assay and its application to characterize additional OGT mutants with altered ability on sugar binding versus protein substrate binding (or sugar transfer). (**a**, **c**) The reactions of OGT variants with GEP1A and NUP62 protein were coupled to an alkyne fluorescent dye and detected by in-gel fluorescence scanning (top panels). To normalize the protein amount, the fluorescent gels were further stained with Coomassie Blue (middle panels). The initial loading amounts were shown in the bottom panels (Coomassie Blue gels). Except $OGT_{4.5}$, all other OGT variants were full-

length proteins. WT, wild-type OGT; NN/AA, N321A/N322A double mutant of OGT. (**b**, **d**) Quantification of the relative intensities of GEP1A-modified OGT and NUP62 protein following normalization to the corresponding protein amounts as shown in the middle panels of (**a**) and (**c**). Similar reactions in the absence of OGT served as negative controls (background labeling). Full gels scans are available in Supplementary Figure 5. Statistical analysis was performed by Student's *t*-test (n = 3). Error bars represent ± s.d. $*P < 0.01$.

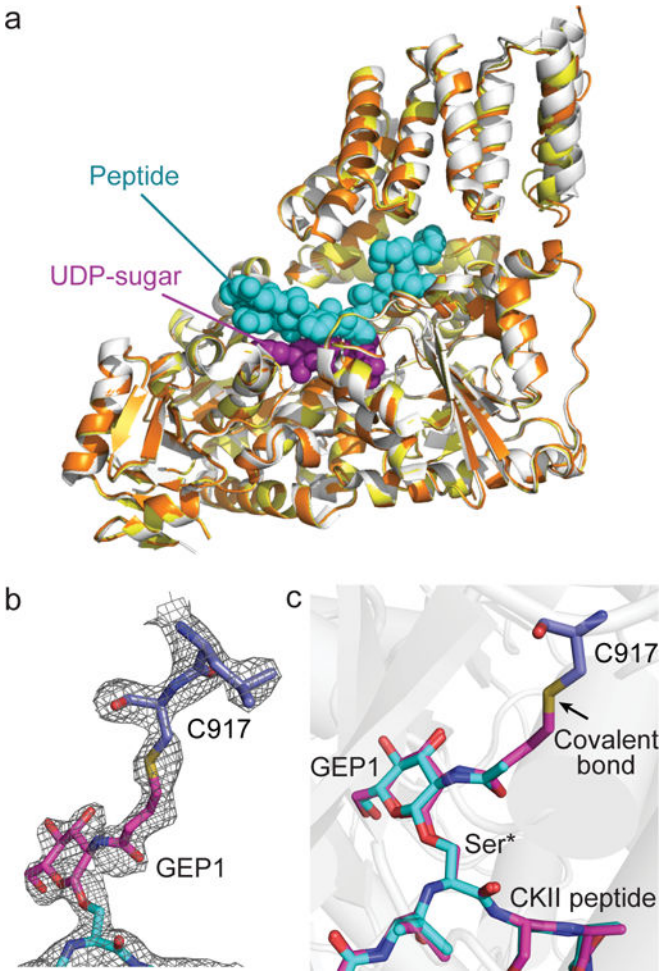

Figure 5.

Crystal structure of the crosslinked OGT:GEP1:CKII complex. (**a**) Superposition of crystal structures of OGT:UDP:O-GlcNAcylated CKII complex (PDB 4GYW, grey) with crosslinked complex of OGT:GEP1:CKII (orange). Ligands are shown in spheres. (**b**) Highlighted electron density of crosslinked complex of OGT:GEP1:CKII demonstrated that GEP1 bridged C917 of OGT and the glycosylating serine of CKII peptide (2*Fo-Fc* density map is shown at 1 σ). (**c**) Superposition of the ligands from the above two complexes: PDB 4GYW (cyan sticks) and crosslinked OGT:GEP1:CKII (magenta sticks) demonstrated that the ligands were crosslinked in a similar conformation as O-GlcNAcylation.