



Published in final edited form as:

*J Am Chem Soc.* 2018 March 14; 140(10): 3510–3513. doi:10.1021/jacs.7b13546.

## O-GlcNAc transferase recognizes protein substrates using an asparagine ladder in the TPR superhelix

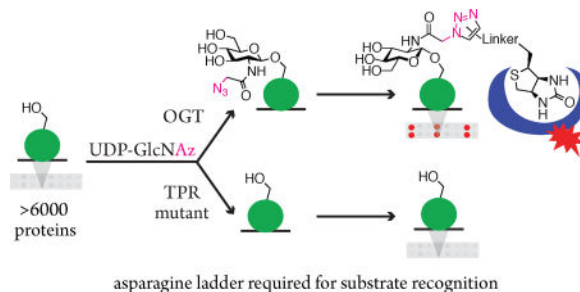
Zebulon G. Levine, Chenguang Fan, Michael S. Melicher, Marina Orman, Tania Benjamin, and Suzanne Walker\*

Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, United States

### Abstract

The essential mammalian enzyme O-GlcNAc Transferase (OGT) is uniquely responsible for transferring N-acetylglucosamine to over a thousand nuclear and cytoplasmic proteins, yet there is no known consensus sequence and it remains unclear how OGT recognizes its substrates. To address this question, we developed a protein microarray assay that chemoenzymatically labels *de novo* sites of glycosylation with biotin, allowing us to simultaneously assess OGT activity across >6000 human proteins. With this assay we examined the contribution to substrate selection of a conserved asparagine ladder within the lumen of OGT's superhelical tetratricopeptide repeat (TPR) domain. When five asparagines were mutated, OGT retained significant activity against short peptides, but showed limited glycosylation of protein substrates on the microarray. O-GlcNAcylation of protein substrates in cell extracts was also greatly attenuated. We conclude that OGT recognizes a majority of its substrates by binding them to the asparagine ladder in the TPR lumen proximal to the catalytic domain.

### Graphical abstract



$\beta$ -O-GlcNAc is a reversible post-translational modification that is ubiquitous in higher eukaryotes and has been implicated in the pathogenesis of cancer, diabetes, and

\*Corresponding Author. [suzanne\\_walker@hms.harvard.edu](mailto:suzanne_walker@hms.harvard.edu).

#### ASSOCIATED CONTENT

Supporting Information.

Experimental protocols, detailed descriptions of data analysis, supplemental figures, and tables containing processed microarray data are available free of charge on the ACS Publications website at DOI: 10.1021/XXXXXXXXXX.

No competing financial interests have been declared.

neurodegeneration.<sup>1</sup> O-GlcNAc Transferase (OGT), the sole enzyme responsible for adding this monosaccharide to serine and threonine residues on hundreds of nuclear and cytoplasmic proteins, is essential for cell survival.<sup>2</sup> Elucidating OGT's role in cellular physiology requires understanding how it recognizes protein substrates. Experimental and bioinformatic approaches have identified preferred amino acids flanking the glycosylation site, but have failed to reveal a useful consensus sequence for glycosylation.<sup>3</sup> Previous studies have implicated the N-terminal superhelical tetratricopeptide repeat (TPR) domain of OGT in substrate recognition.<sup>4</sup> This domain contains 13 tetratricopeptide repeats; truncating it to six repeats abrogates glycosylation of tested protein substrates,<sup>4d-f</sup> but does not prevent glycosylation of short peptides.<sup>4c-e</sup> Therefore, the TPR domain must contain structural features important for recognition of full-length proteins. Here, we show that a ladder of asparagines in the TPR lumen<sup>4a,g,h</sup> plays a major role in recognition of most OGT protein substrates. This finding demonstrates that residues distal to the site of catalysis are key players in substrate recognition; the results may lead to new approaches to predict OGT substrates and to inhibit O-GlcNAcylation in cells.

To investigate the role of the TPR asparagines in substrate selection, we needed a method to compare activity of wild-type and mutant OGT variants against a large number of protein substrates. We reasoned that protein microarrays, which display thousands of recombinant proteins on each array, would allow for such comparisons. We have previously used microarrays to discover new OGT substrates by using a monoclonal antibody to detect O-GlcNAcylation.<sup>5</sup> We found, however, that antibody detection resulted in background signal for some proteins due to pre-existing O-GlcNAc modifications, and we were also concerned about detection bias due to epitope preferences of the antibody.<sup>6</sup> Therefore, we sought a strategy that reported on OGT activity without using antibodies.

To detect only newly glycosylated proteins, we exploited OGT's ability to use UDP-GlcNAz as a sugar donor (Figure 1A, S1).<sup>7</sup> The azide on the *N*-acetyl moiety allows labeling of newly glycosylated proteins with biotin via strain-promoted azide-alkyne click chemistry.<sup>8</sup> We compared this method to antibody-based detection by treating commercial protein arrays with OGT and either UDP-GlcNAc or UDP-GlcNAz. Control arrays were treated with the sugar donors without OGT. UDP-GlcNAc-treated arrays were stained with an anti-O-GlcNAc antibody (CTD110.6) and a fluorescent secondary antibody as previously reported.<sup>5</sup> The GlcNAz-treated arrays were reacted with ADIBO-biotin<sup>9</sup> and stained with fluorescent streptavidin. As previously observed, background signal on some proteins was high for the antibody-treated control array; however, we found that a control array treated with ADIBO-biotin did not show high background signal (Figure 1B, S2).

We compared the two staining methods to assess how well each detected OGT biochemical activity. The ratio of signal intensity between OGT-treated and untreated arrays was determined for each protein (Figure S3A). The intensity ratios for each staining method followed a log-normal distribution; both staining methods had a tail of positive outliers we expected to be proteins glycosylated by OGT (Figure S3B-C). Proteins with a log<sub>2</sub> intensity ratio more than two standard deviations above the mean were identified as hits for each staining method (Figure S3B-C). Out of 6747 unique proteins analyzed, 415 scored as hits by at least one detection method (Figure 1C, S3B-C, Table S1), with 97 of these detected by

both methods. By data mining previous proteomics studies,<sup>10</sup> we identified 275 known O-GlcNAc proteins on the arrays. In cells, adaptor proteins are thought to recruit OGT to an unknown fraction of substrates,<sup>2,11</sup> but our assay system only interrogates direct recognition of substrates by OGT. Given that we do not know which proteins on the array are directly recognized by OGT, we were pleased to find that 60 of the known cellular O-GlcNAc proteins (22%) were detected as glycosylated on the array by at least one method, with 51 of these being detected by the biotinylation method (Figure 1D). The biotinylation method also detected nine of ten previously validated microarray hits discovered using the antibody method,<sup>5</sup> with the tenth not being glycosylated according to either staining method on this batch of microarrays (Table S2). We validated several additional microarray hits (Figure S4), of which four were detected by both methods, two were solely detected by the antibody, and two were solely detected by biotinylation (Table S3). Overall, the results show that both methods perform well, if not identically, at detecting known and new substrates on the microarray. We opted to use the biotinylation strategy for our microarray studies of the TPR asparagines due to its reduced background signal and increased hit profile.

To assess how broadly OGT uses the TPR domain asparagine ladder for substrate recognition, we wanted to compare microarrays treated with either wild-type OGT or a mutant in which the five asparagines closest to the active site were replaced with alanines (5N5A). These residues were previously shown to be important for binding two OGT substrates; structural studies revealed that the asparagine side chains anchor these proteins in the TPR lumen through bidentate hydrogen bonds to the peptide backbone (Figure 2A).<sup>4g,h</sup> Prior to treating microarrays with these OGT variants, we examined *in vitro* glycosylation of two peptides: a 17mer derived from casein kinase 2 (CK2) that does not contact the TPR domain,<sup>3e</sup> and a 35mer containing a C-terminal region that binds in the TPR lumen where it is anchored by the asparagine ladder (HCF-E10S, Figure 2A).<sup>12</sup> The 5N5A mutant showed similar activity to wild-type enzyme against the CK2 peptide, especially at low peptide concentrations (Figure 2B, S5A–B), but it did not glycosylate the HCF-E10S substrate (Figure 2C). Size exclusion chromatography and circular dichroism showed that there were no gross structural changes in the mutant (Figure S5C–D). The catalytic site of the 5N5A mutant appears intact based upon the CK2 results, yet the substrate recognized by the asparagine ladder is not glycosylated by this mutant. Therefore, we reasoned that we could use the dramatic decrease in activity with the 5N5A mutant to report on substrate binding in the lumen of the TPR domain.

To compare the activity of the 5N5A mutant to wild-type OGT across a large number of substrates, we treated microarrays with UDP-GlcNAz and either no enzyme or one of the two forms of OGT. Multiple arrays were treated per condition to allow us to distinguish effects due to the enzyme variant used on the array from noise between array replicates (Figure S6, S7A). Biotinylated control proteins spatially distributed across each array allowed us to normalize for variation in staining (Figure S7), and we corrected for variation in OGT activity using a known OGT substrate also present as a control throughout the arrays (Figure S8).<sup>14</sup> We noted that after these normalizations, almost all proteins had higher signal on the OGT-treated arrays than control arrays (Figure S9A–B, Table S4). This suggested that OGT has non-specific glycosyltransferase activity; however, the wild-type OGT-treated arrays also had a tail of positive outliers that we expected to be specifically recognized

substrates (Figure S9C). To focus only on these specific targets, we median normalized between the control, 5N5A, and wild-type OGT-treated array groups (Figure S9D).

A permutation test approach<sup>15</sup> determined which proteins were glycosylated by wild-type and 5N5A OGT variants. Those proteins with a greater than two-fold change in median signal between array treatments at a false-discovery rate of less than 5% were considered hits (Table S4). 739 proteins scored as hits when comparing wild-type OGT-treated arrays to controls (Figure 3A, S10), suggesting these proteins were specific OGT targets. Out of these 739 hits, 736 had a higher signal for wild-type OGT than 5N5A, with 555 of these meeting FDR cutoffs to be hits in a direct comparison between the 5N5A mutant and wild-type OGT. Notably, the loss of 5N5A activity relative to wild-type OGT against TAB1, a substrate against which the 5N5A mutant is known to have impaired glycosylation,<sup>4h</sup> was close to the average activity loss seen across all substrates of wild-type OGT (Figure 3A, green circle, S11). This observation implies that TAB1 is not a special case; instead, most OGT targets are poor glycosylation substrates for 5N5A. Based on these results, we concluded that the asparagine ladder is broadly used in substrate recognition.

We next wondered if the presence of cellular adaptor proteins might overcome the glycosylation defect seen on the microarrays by recruiting the 5N5A mutant to substrates independent of the mutated residues. To test this, we glycosylated HeLa cell extracts using exogenous wild-type or 5N5A mutant OGT (Figure 3B, S12). When UDP-GlcNAc but no OGT was added, minimal additional glycosylation was observed. Adding UDP-GlcNAc and wild-type enzyme led to robust glycosylation over time; in contrast, adding the 5N5A mutant resulted in a modest increase in glycosylation over time. This result demonstrated that the 5N5A mutant was defective for glycosylation in solution as well as on the microarrays. Furthermore, it showed that even in extracts that are presumed to contain adapter proteins, the 5N5A mutant is globally impaired for glycosylation. Given the agreement between the microarray and cell extract results, we have concluded that the asparagines in the TPR lumen proximal to the catalytic domain play a critical role in recognition of most substrates.

In summary, we have developed a protein microarray assay to determine how structural features of OGT affect activity against thousands of spatially separated proteins. This tool has allowed us to compare activity of wild-type OGT to a mutant in which five asparagines in the TPR lumen have been replaced with alanine. We showed that direct recognition of most OGT substrates depends upon having an intact asparagine ladder. Glycosylation in cell extracts is impaired, leading us to conclude that the majority of OGT substrates are recruited through binding interactions that involve contacts with the asparagines in the TPR lumen, regardless of adaptor protein involvement. The OGT active site itself accommodates a wide range of peptide sequences (Figure S13), and no clear consensus sequence for O-GlcNAcylation has been identified. It may be possible to predict glycosylation sites by including distal sequence elements that interact with the TPR domain. The microarray methods, which enable comparison of wild-type and mutant enzymes, can help identify the molecular interactions required to recruit OGT to particular substrates, and may lead to strategies to disrupt glycosylation of subsets of cellular proteins.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

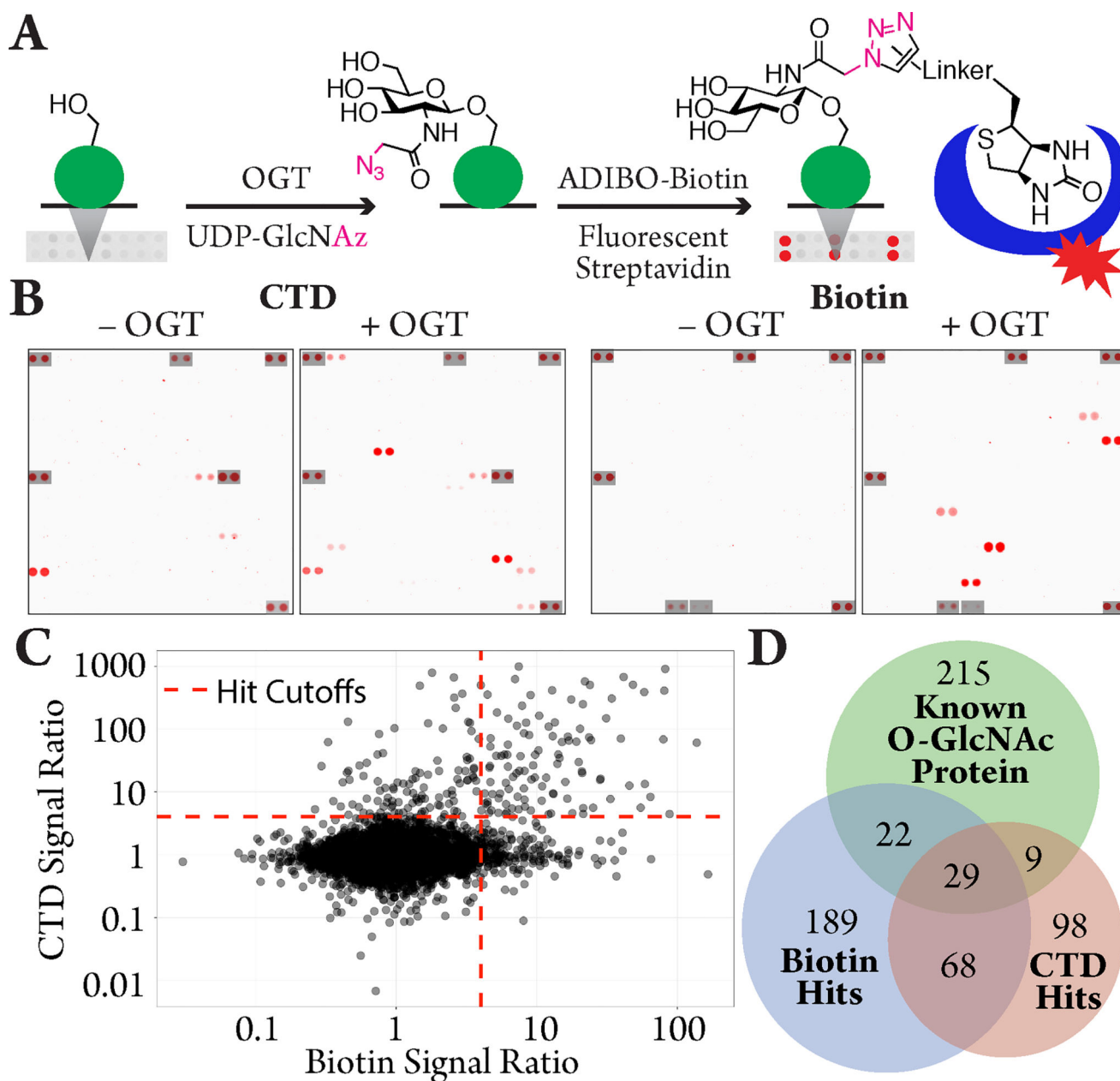
## Acknowledgments

We would like to thank Dr. Tao Wu for aid in producing HeLa S3 extracts, Dr. Cassandra M. Joiner and George Q. Fei for aid in manuscript editing, and the Center for Macromolecular Interactions at Harvard Medical School for aid in circular dichroism experiments. This research was supported by the National Institutes of Health (GM094263).

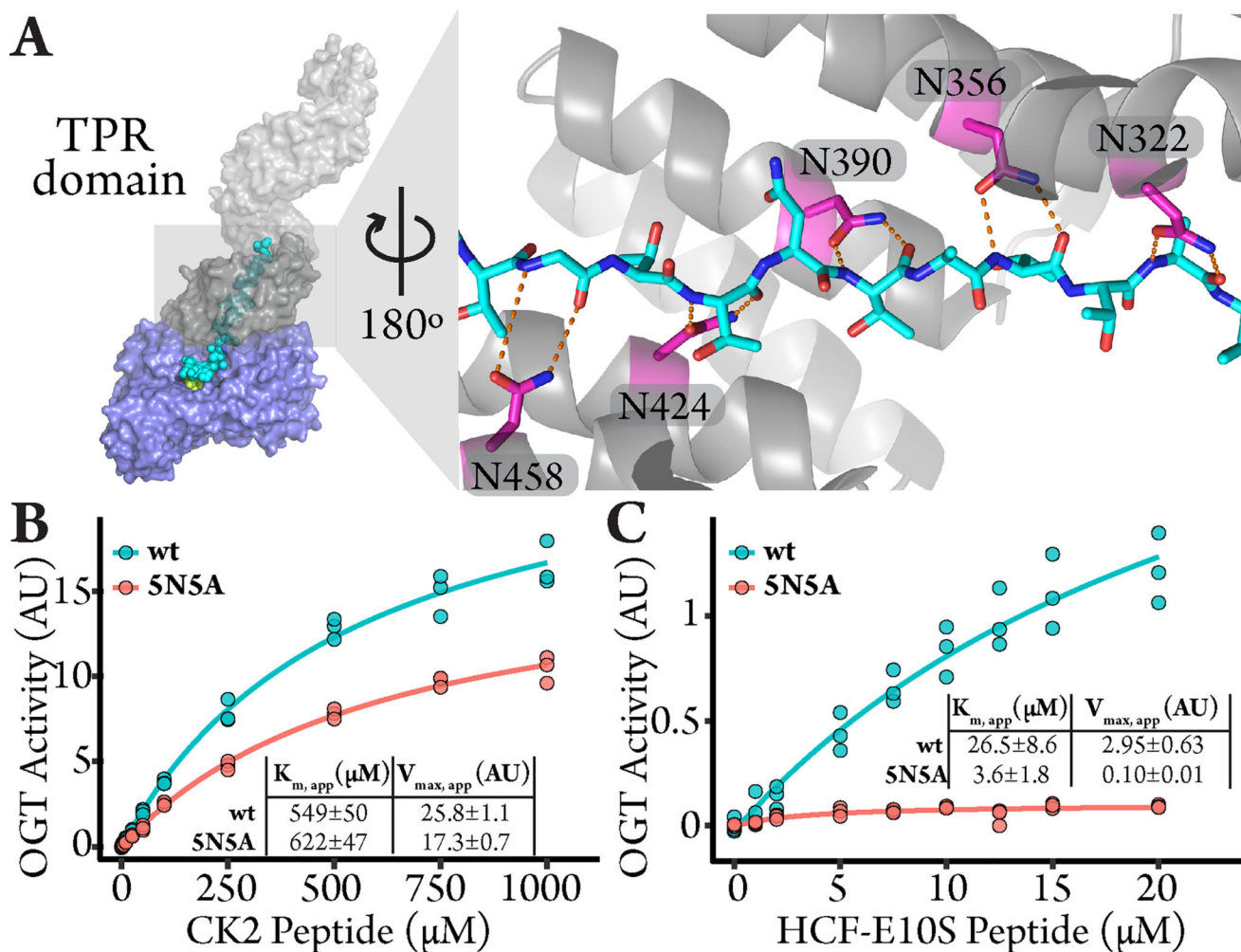
## References

- (a) Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. *Ann. Rev. Biochem.* 2011; 80:825–858. [PubMed: 21391816] (b) Ferrer CM, Reginato MJ. *Mol. Cell. Oncol.* 2015; 2:e961809. [PubMed: 27308381] (c) Vaidyanathan K, Wells L. *J. Biol. Chem.* 2014; 289:34466–34471. [PubMed: 25336652] (d) Yuzwa SA, Voadlo DJ. *Chem. Soc. Rev.* 2014; 43:6839–6858. [PubMed: 24759912]
- Levine ZG, Walker S. *Ann. Rev. Biochem.* 2016; 85:631–657. [PubMed: 27294441]
- (a) Leavy TM, Bertozzi CR. *Bioorg. Med. Chem. Lett.* 2007; 17:3851–3854. [PubMed: 17531489] (b) Pathak S, Alonso J, Schimpl M, Rafie K, Blair DE, Borodkin VS, Albarbarawi O, van Aalten DMF. *Nat. Struct. Mol. Biol.* 2015; 22:744–750. [PubMed: 26237509] (c) Liu X, Li L, Wang Y, Yan H, Ma X, Wang PG, Zhang L. *FASEB J.* 2014; 28:3362–3372. [PubMed: 24760753] (d) Wu H-Y, Lu C-T, Kao H-J, Chen Y-J, Lee T-Y. *BMC Bioinf.* 2014; 15:S1.(e) Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. *Nature.* 2011; 469:564–567. [PubMed: 21240259] (f) Wang J, Torii M, Liu H, Hart GW, Hu ZZ. *BMC Bioinf.* 2011; 12:91.(g) Wang J, Torii M, Liu H, Hart GW, Hu ZZ. *BMC Bioinf.* 2011; 12:91.(h) Jochman R, Holz P, Sticht H, Stürzl M. *Biochim. Biophys. Acta.* 2014; 1844:416–421. [PubMed: 24332980]
- (a) Britto-Borges T, Barton GJ. *PLoS One.* 2017; 12:e0184405. [PubMed: 28886091] (b) Jínek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E. *Nat. Struct. Mol. Biol.* 2004; 11:1001–1007. [PubMed: 15361863] (c) Kreppel LK, Hart GW. *J. Biol. Chem.* 1999; 274:32015–32022. [PubMed: 10542233] (d) Lubas WA, Hanover JA. *J. Biol. Chem.* 2000; 275:10983–10988. [PubMed: 10753899] (e) Comer FI, Hart GW. *Biochemistry.* 2001; 40:7845–7852. [PubMed: 11425311] (f) Iyer SPN, Hart GW. *J. Biol. Chem.* 2003; 278:24608–24616. [PubMed: 12724313] (g) Lazarus MB, Jiang J, Kapuria V, Bhuiyan T, Janetzko J, Zandberg WF, Voadlo DJ, Herr W, Walker S. *Science.* 2013; 342:1235–1239. [PubMed: 24311690] (h) Rafie K, Raimi O, Ferenbach AT, Borodkin VS, Kapuria V, van Aalten DMF. *Open Biol.* 2017; 7:170078. [PubMed: 28659383] (i) Hu C-W, Worth M, Fan D, Li B, Li H, Lu L, Zhong X, Lin Z, Wei L, Ge Y, Li L, Jiang J. *Nat. Chem. Biol.* 2017; 13:1267–1273. [PubMed: 29058723] (j) Clarke AJ, Hurtado-Guerrero R, Pathak S, Schüttelkopf AW, Borodkin V, Shepherd SM, Ibrahim AFM, van Aalten DMF. *EMBO J.* 2008; 27:2780–2788. [PubMed: 18818698]
- Ortiz-Meoz RF, Merbl Y, Kirschner MW, Walker S. *J. Am. Chem. Soc.* 2014; 136:4845–4848. [PubMed: 24580054]
- Teo CF, Ingale S, Wolfert MA, Elsayed GA, Nöt LG, Chatham JC, Wells L, Boons G-J. *Nat. Chem. Biol.* 2010; 6:338–343. [PubMed: 20305658]
- Voadlo DJ, Hang HC, Kim EJ, Hanover JA, Bertozzi CR. *Proc. Natl. Acad. Sci. U.S.A.* 2003; 100:9116–9121. [PubMed: 12874386]
- Agard NJ, Prescher JA, Bertozzi CR. *J. Am. Chem. Soc.* 2004; 126:15046–15047. [PubMed: 15547999]
- Kuzmin A, Poloukhina A, Wolfert MA, Popik VV. *Bioconjugate Chem.* 2010; 21:2076–2085.
- (a) Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M. *Nucleic Acids Res.* 2012; 40:D261–D270. [PubMed: 22135298] (b) Wang S, Yang F, Petyuk VA, Shukla AK, Monroe ME, Gritsenko MA, Rodland KD, Smith RD, Qian W-J, Gong C-X, Liu T. *J. Pathol.* 2017; 243:78–88. [PubMed: 28657654]
- (a) Cheung WD, Hart GW. *J. Biol. Chem.* 2008; 283:13009–13020. [PubMed: 18353774] (b) Cheung WD, Sakabe K, Housley MP, Dias WB, Hart GW. *J. Biol. Chem.* 2008; 283:33935–33941. [PubMed: 18840611] (c) Ruan HB, Han X, Li MD, Singh JP, Qian K, Azarhoush S, Zhao L,

- Bennett AM, Samuel VT, Wu J, Yates JR 3rd, Yang X. *Cell Metab.* 2012; 16:226–237. [PubMed: 22883232] (d) Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, Dawson MA, Volkmar M, Putmans P, Calonne E, Shih AH, Levine RL, Bernard O, Mercher T, Solary E, Urh M, Daniel DL, Fuks F. *EMBO J.* 2013; 32:645–655. [PubMed: 23353889] (e) Chen Q, Chen Y, Bian C, Fujiki R, Yu X. *Nature.* 2013; 493:561–564. [PubMed: 23222540]
12. Janetzko J, Walker S. *J. Am. Chem. Soc.* 2017; 139:3332–3335. [PubMed: 28207246]
13. Ortiz-Meoz RF, Jiang J, Lazarus MB, Orman M, Janetzko J, Fan C, Duveau DY, Tan Z-W, Thomas CJ, Walker S. *ACS Chem. Biol.* 2015; 10:1392–1397. [PubMed: 25751766]
14. (a) Jiang MS, Hart GW. *J. Biol. Chem.* 1997; 272:2421–2428. [PubMed: 8999954] (b) Cheng X, Hart GW. *J. Steroid Biochem. Mol. Biol.* 2000; 75:147–158. [PubMed: 11226831]
15. (a) Merbl Y, Kirschner MW. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:2543–2548. [PubMed: 19181856] (b) Storey JD, Tibshirani R. *Proc. Natl. Acad. Sci. U.S.A.* 2003; 100:9440–9445. [PubMed: 12883005] (c) Yang H, Churchill G. *Bioinformatics.* 2007; 23:38–43. [PubMed: 17077100]

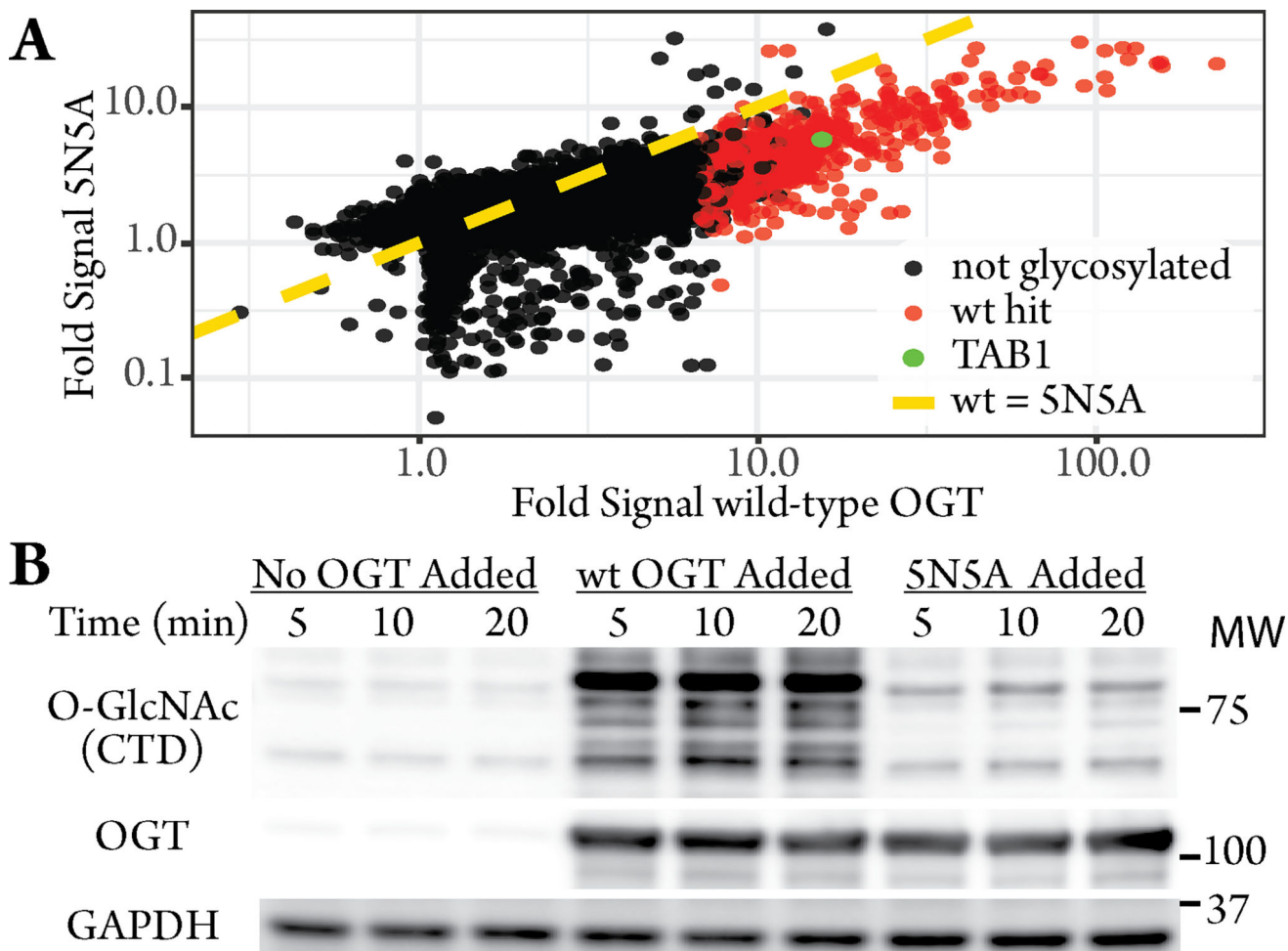
**Figure 1.**

Chemoenzymatic biotin labeling reduces background signal and improves detection *A*) Use of sugar analogue GlcNAz allows labeling with ADIBO-biotin and detection by fluorescent streptavidin (blue) *B*) Different sub-blocks from microarrays treated with or without OGT for antibody (CTD) or biotin detection methods. Grey boxes show control spots. *C*) Scatterplot comparing intensity ratios for the two detection methods for all proteins on the microarrays. Dashed red lines indicate hit cutoffs. *D*) Venn diagram of hits by each method compared to proteins on the microarray that are O-GlcNAcylated in cells.<sup>10</sup>



**Figure 2.** 5N5A is catalytically active but cannot glycosylate a substrate that binds the TPR asparagines *A*) Composite structure of OGT bound to an HCF-1 repeat (cyan), with an expanded view of the conserved asparagines (magenta) on the right. Dotted orange lines show bidentate hydrogen bonds to the HCF-1 peptide backbone. The composite was generated from PDB 4N3B,<sup>4g</sup> comprising the catalytic lobes (purple) and 4.5 TPRs (dark grey), and TPR structure PDB 1W3B.<sup>4b</sup> *B–C*) OGT wild-type and 5N5A mutant activity against *B*) non-TPR-binding CK2 peptide or *C*) TPR-binding HCF-E10S peptide (AU: arbitrary units, measured using UDP-Glo (Promega)<sup>13</sup>)  $n=3$  replicates per peptide concentration. Apparent  $K_m$  for peptide and apparent  $V_{max}$  at 100  $\mu\text{M}$  UDP-GlcNAc are listed  $\pm$  standard error.





**Figure 3.**

The asparagine ladder in the TPR lumen is critical for recognition of OGT substrates. *A*) Fold signal above control for every protein for wild-type OGT (x-axis) and 5N5A (y-axis) based upon median normalized data. The dashed line represents equivalent activity between wild-type and 5N5A enzymes. Red circles are hits for wild-type enzyme; black circles are proteins that do not score as glycosylated. TAB1, a known poor 5N5A substrate, is highlighted in green. *B*) Western blot of HeLa S3 cell extracts incubated with UDP-GlcNAc for indicated times with or without added OGT. OGT and GAPDH blots show enzyme and extract loading; CTD stains for O-GlcNAc. Molecular weight markers are indicated on the right.