C h a p t e r 3

Development of biological and chemical tools for discovery of the OGT interactome and *O*-GlcNAcome

Portions of this chapter are published as:

Griffin ME, **Jensen EH**, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. "Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag." *Mol. Biosys*. 2016, 12: 1756-1759. doi: 10.1016/j.chembiol.2015.12.007. Research article.

The work of this chapter was truly a collaborative effort throughout involving Matt Griffin, Yao Xiao, Yelena Koldobskaya, and Priya Choudhry, precluding the ability to sufficiently credit throughout the text. Please assume that Yelena, Yao, Matt, Priya, and I conducted the OGT interactome project and Matt and I performed the O-GlcNAcome project experiments. Shirley Pease generated the OGT-FLAG-HA mouse at Caltech (Genetically Engineering Mouse (GEM) Services). The peptide LCMS for the *O*-GlcNAcome experiments were performed by Mona Shahgholi at the Caltech Mass Spectrometry Laboratory. The protein *O*-GlcNAcome MS was performed by Daniel Mason at the Genomics Institute of the Novartis Research Foundation (GNF). The OGT interactome proteomic MS was performed by Annie Moradian, Sonja Hess, and Michael Sweredoski at the Caltech Proteome Exploration Laboratory (PEL).

3.1 Abstract

The intracellular post-translational modification of serine or threonine residues of proteins with a single N-acetylglucosamine monosaccharide (O-GlcNAcylation) is essential for neuronal homeostasis and a variety of cellular processes. This modification is cycled by only two enzymes in mammalian cells: O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA). OGT contains several protein-binding TPR domains, which allow it to bind to specific interactors. A prevalent hypothesis in the field suggests that OGT is targeted OGT to specific substrates by certain key interactors. In order to identify these interactors, we used CRISPR/Cas9 to develop a novel mouse with a minimally tagged OGT in order to identify the OGT brain interactome using tandem affinity purification and proteomic methods. The preliminary OGT brain interactome showed agreement with previous OGT interactome studies and O-GlcNAcome studies although these studies were performed in different species and cell types. The identified OGT interactors are enriched for ribosomal and cytoskeletal proteins in addition to axonal, dendritic, and neuronal cell body proteins.

In addition to the OGT interactome, we sought to uncover OGT substrates in order to fully test the OGT interactor/substrate hypothesis. OGT modifies over 1000 different proteins, but the lack of a well-defined consensus sequence has hampered efforts to predict sites a priori. Furthermore, relatively few O-GlcNAc modification sites have been mapped due to the difficulty of enriching and detecting O-GlcNAcylated peptides from complex samples. Here, we describe an improved approach to quantitatively label and enrich O-GlcNAcylated proteins for site identification. Chemoenzymatic labeling followed by copper(I)-catalyzed azide-alkyne cycloaddition

(CuAAC) installs a new mass spectrometry (MS)-compatible linker designed for facile purification and release of O-GlcNAcylated proteins for downstream MS analysis. We validate the approach by unambiguously identifying several established O-GlcNAc sites on the proteins α-crystallin and O-GlcNAc transferase (OGT) as well as discovering new, previously unreported sites on both proteins. Notably, these novel sites on OGT lie in key functional domains of the protein, underscoring how this site identification method can reveal important biological insights into protein activity and regulation.

3.2 Overview of OGT interactome and *O*-GlcNAcome approach

The enzymes OGT and OGA cycle *O*-GlcNAc modification on over 1000 proteins in mammalian cells.[1] Although the recent crystal structures of OGT have augmented our understanding of the structural constraints imposed by the active site of OGT, OGT appears to lack a well-defined consensus sequence, which has hampered efforts to predict substrates and their sites.[2] Without a well-defined consensus sequence or site prediction, the choice of OGT substrate remains poorly understood. Several studies have demonstrated that key OGT interactors can effectively recruit OGT to specific substrates.[3-5] This has lead to the OGT interactor/substrate hypothesis whereby the protein-binding TPR (tetratricopeptide repeat) domain of OGT interacts with certain interactors, which in turn mediate the *O*-GlcNAcylation of specific target proteins.[6] While this hypothesis has been shown for specific substrate-interactor pairs, this hypothesis has not been tested in a global proteome-wide fashion. Toward that end, we have developed chemical and biological tools in order to identify the global OGT interactome and *O*-GlcNAcome.

In order to identify the OGT interactome, we sought to develop a biological system that would enable facile pull down of OGT to identify physiologically-relevant interactors (Figure 3.1). Through the utilization of Crispr/Cas9 genome editing methods, we installed two small FLAG and HA tags upon the C-terminus of OGT in mice for tandem affinity purification. There are multiple advantages of our approach: (1) tandem affinity purification methods have been shown to successfully enrich interactors with minimal off-target effects compared to single enrichment methods; (2) small tags are likely to minimally perturb interactors and OGT function; and (3) by labeling endogenous OGT instead of previously employed overexpression/knockdown strategies or yeast two-hybrid assays, we endeavor to maximize endogenous interactors and avoid false positive interactors.[7] It is important to note that tandem affinity purification will identify strong and stable interactions rather than transient interactions.

In addition to the OGT interactome, we also needed to identify OGT's substrates (the *O*-GlcNAcome). We employed chemoenzymatic-labeling methods in order to quantitatively label and then enrich *O*-GlcNAcylated proteins and peptides for site identification. Using this approach, a GalNAz group was appended to *O*-GlcNAcylated proteins using a mutant enzyme. Then, a novel biotin-Dde-alkyne linker was covalently appended to the GalNAz handle (Dde = 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl). Now biotinylated *O*-GlcNAcylated proteins were enriched using streptavidin beads and eluted using mild hydrazine cleavage of the Dde group. The remaining amine moiety conferred a positive charge to peptides facilitating mass spectrometric identification. Through this method, we were able to identify known and novel sites on the *O*-GlcNAcylated proteins, α-crystallin and OGT itself. Finally, current

efforts are underway to expand the *O*-GlcNAcome approach to identify novel *O*-GlcNAc sites across the proteome in the brain (Figure 3.1). Once the *O*-GlcNAcome and OGT interactome have been identified, we can determine the important hub interactors as well as key substrates and interactors in different tissue-specific contexts. Importantly, with the OGT-FH mouse for OGT interactome enrichment, these tools and methods have broad applicability to different tissues, developmental time points, disease states (when OGT-FH mice are bred with other disease mouse models), or physiological relevant stimuli.

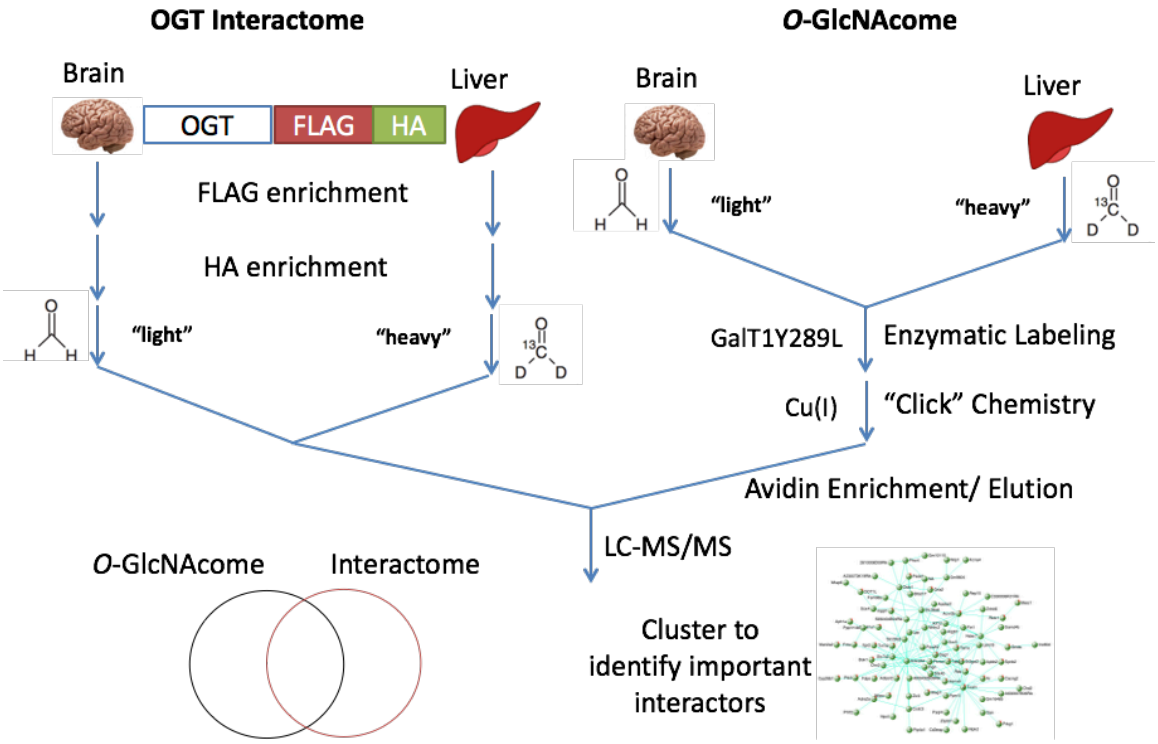


Figure 3.1 Overview of tissue-specific OGT interactome and *O*-GlcNAcome dual approach. Shown is the overall workflow that we have been developing in order to identify the OGT interactome and *O*-GlcNAcome (OGT substrates). Briefly, we tag endogenous OGT with the small FLAG and HA tags and perform tandem affinity purification followed by mass spectrometry (left side). In parallel, the *O*-GlcNAcome is obtained through differential sample labeling, chemoenzymatic labeling, avidin enrichment, and a mild elution prior to mass spectrometry (right side). After these dual analyses, the *O*-GlcNAcome and OGT interactome can be integrated to identify critical hub interactors and determine the differential regulation of *O*-GlcNAc across different tissues.

3.3 Development of biological tools for identifications of the OGT interactome

3.3.1 Validation of tandem affinity purification and C-terminal tagged OGT using OGT activity assay

Prior to generating a mouse with the FLAG-HA tag inserted into OGT, we first wanted to validate the tandem affinity purification and that the tags would not interfere with OGT activity. In order to check the first criterion, we used HEK293T cells that had been treated with lentivirus that inserted OGT-FLAG-HA (C-terminal tagged OGT) downstream of a doxycycline-inducible promoter. The workflow of the tandem affinity purification is shown in Figure 3.2. Briefly, we lysed cells or tissue expressing either OGT (wildtype or -dox) or OGT-FLAG-HA (OGT-FH or +dox) and enriched the OGT-FH using anti-FLAG beads. After washing and elution with a 3xFLAG tag peptide, we then enriched the OGT-FH with anti-HA beads, washed, and eluted using 3M NaSCN (Figure 3.3C). Prior to cell lysis, we also explored the usage of crosslinking techniques in order to obtain the transient OGT interactome (Figure 3.3A-B). While the crosslinking method appeared to pull down Nup62, a known interactor and substrate of OGT, we decided to forego the crosslinking in order to simplify the workflow and also focus on the strong, stable interactors.

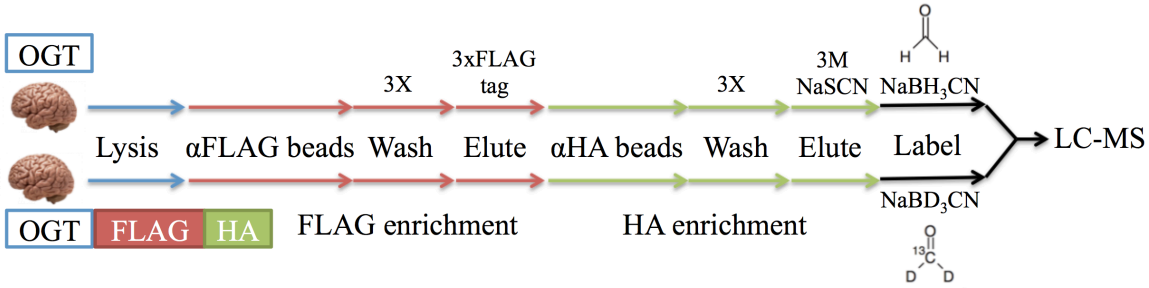


Figure 3.2 Workflow of tandem affinity purification for OGT interactome identification. In order to verify selective OGT-FH enrichment, we used HEK293T cells that would express OGT-FH in response to doxycycline treatment.

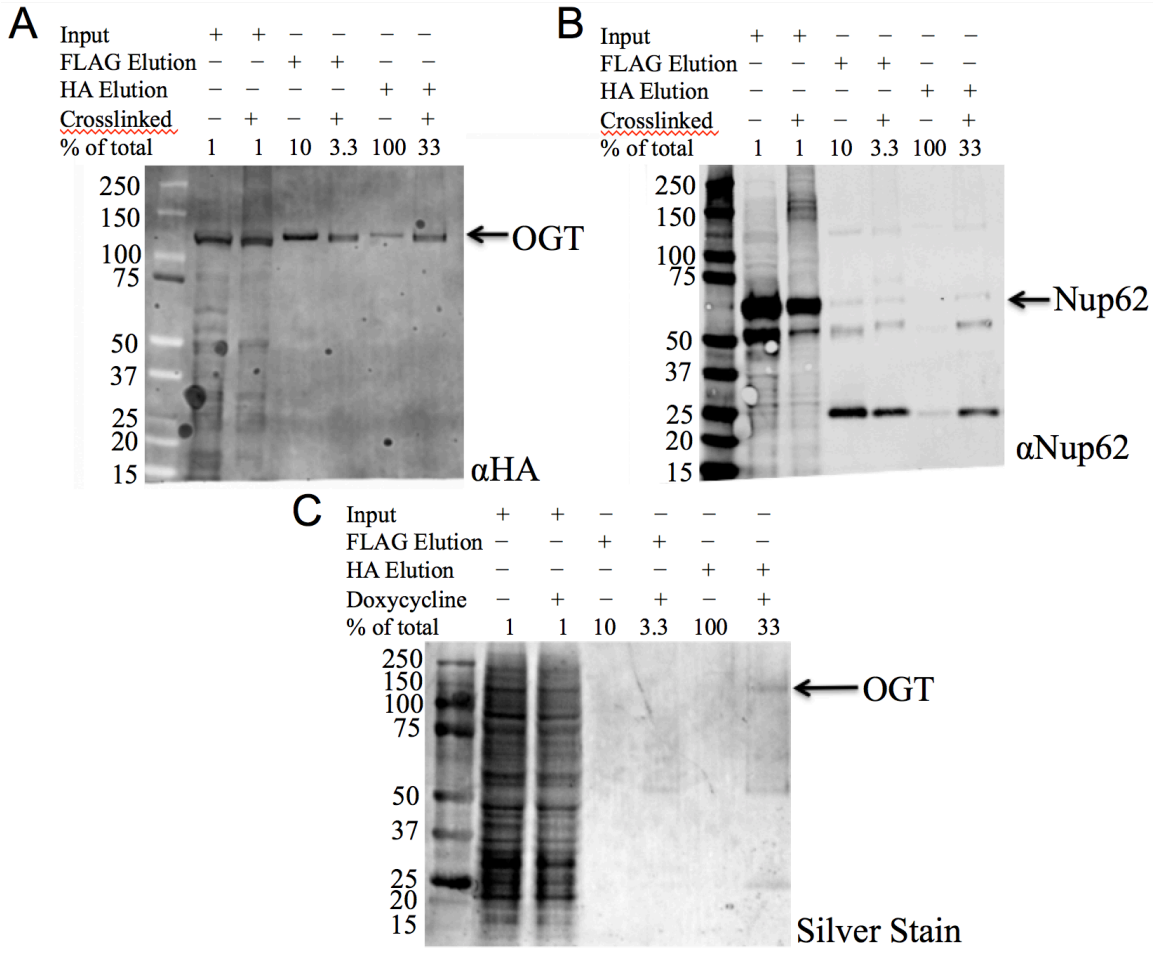


Figure 3.3 Validation of crosslinking conditions and doxycycline induction of OGT-FH expression. (A) We showed pull down of OGT itself with and without crosslinking using a HA antibody. (B) Nup62, a known substrate and interactor with OGT, is enriched in crosslinking conditions. (C) Using a silver stain, we show that we are able to pull down OGT using our tandem affinity purification methods only in the doxycycline-induced expression conditions but not without doxycycline.

In addition to validating the efficiency of tandem affinity purification, we sought to compare the activity of N-terminal tagged OGT to the C-terminal tagged OGT. The active site of OGT is closer to the C-terminus while the N-terminus of OGT contains the protein-protein interacting TPR domains. Toward that end, we transfected HEK293T cells with C-terminal and N-terminal tagged OGT, purified the OGT with a FLAG pull down, and then performed an activity assay. We observed no difference in the activity of the N-terminal and C-terminal tagged OGT as measured by an *in vitro* OGT activity assay (Figure 3.4). Importantly, the N-terminus of OGT undergoes differential splicing to

form mitochondrial (mOGT, 11.5 TPR domains) and short OGT (sOGT, 4.5 TPR domains). This would mean that an N-terminal tag would only tag the full-length isoform of OGT (13.5 TPR domains) while leaving the sOGT and mOGT untagged. We decided to move forward with tagging OGT on the C-terminus as this strategy would (1) tag all isoforms of OGT thereby ensuring the complete OGT interactome would be enriched and (2) not interfere with OGT activity.

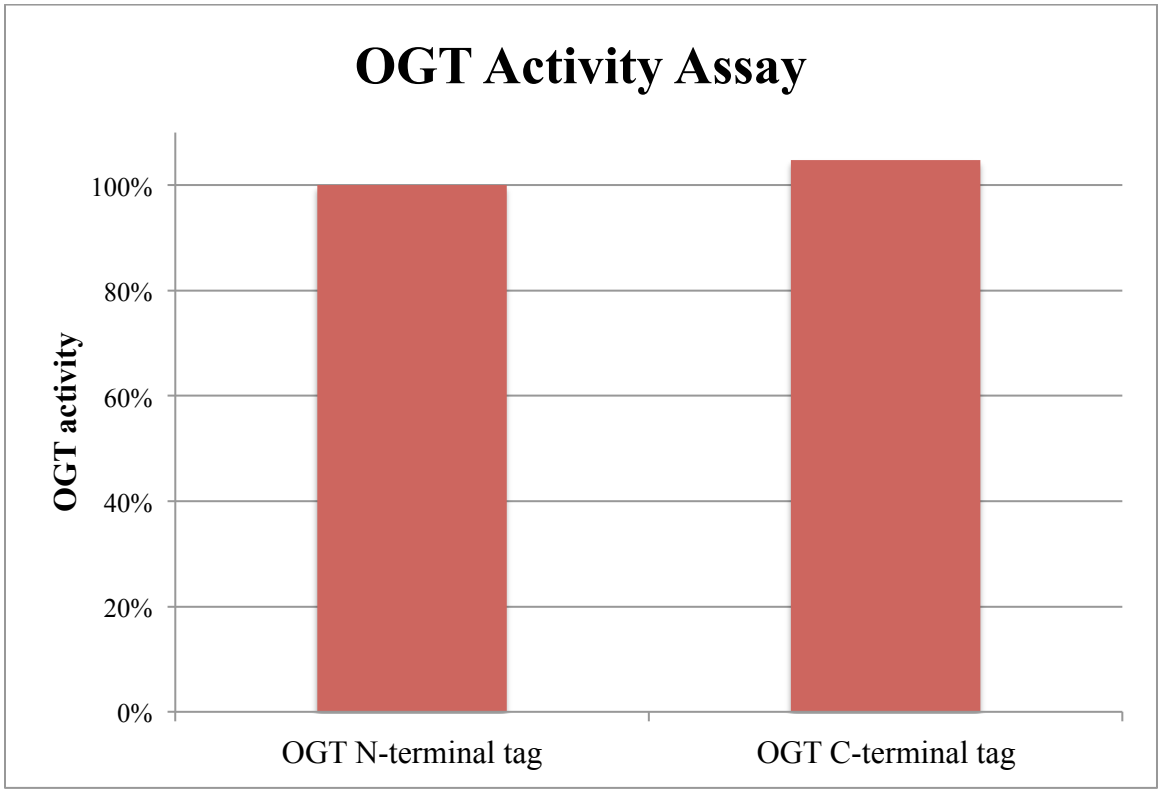


Figure 3.4 FLAG-HA C-terminal and N-terminal tagging of OGT does not affect OGT activity. Using the UDP-Glo™ Glycosyltransferase Assay from Promega, we showed that there was no discernible difference in the OGT activity. The average fluorescence was normalized to the OGT levels as measured by Western.

3.3.2 Validation of OGT targeting sgRNA for CRISPR/Cas9

Using the guide RNA program from the Feng Zhang, we obtained five sgRNA candidates that we then tested for cutting and homologous recombination efficiency.[8] Briefly, the guide RNA needs to have a requisite 3'-protospacer-adjacent motif (PAM) sequence (N(G or A)G) that allows the sgRNA and Crispr/Cas9 machinery to bind

efficiently to the cut site and ideally, good sequence homology for the target site while having low off-target cut site homology across the entire genome. We cloned the C-terminus of mouse OGT into the pCAG-EGxxFP construct that would produce EGFP upon dsDNA cleavage and successful homologous recombination (Figure 3.5). All five of the sgRNA sequences were efficient based on this assay, but we decided to use the highest rated (first) sgRNA as predicted by Feng Zhang lab's program that maximizes on-target cleavage while minimizing off-targets cleavage [S_{guide} = 87% = 100/[100 – Σ(i = 1 → n_{mm}) $S_{hit}(h_i)$]] (Figure 3.5).[8]

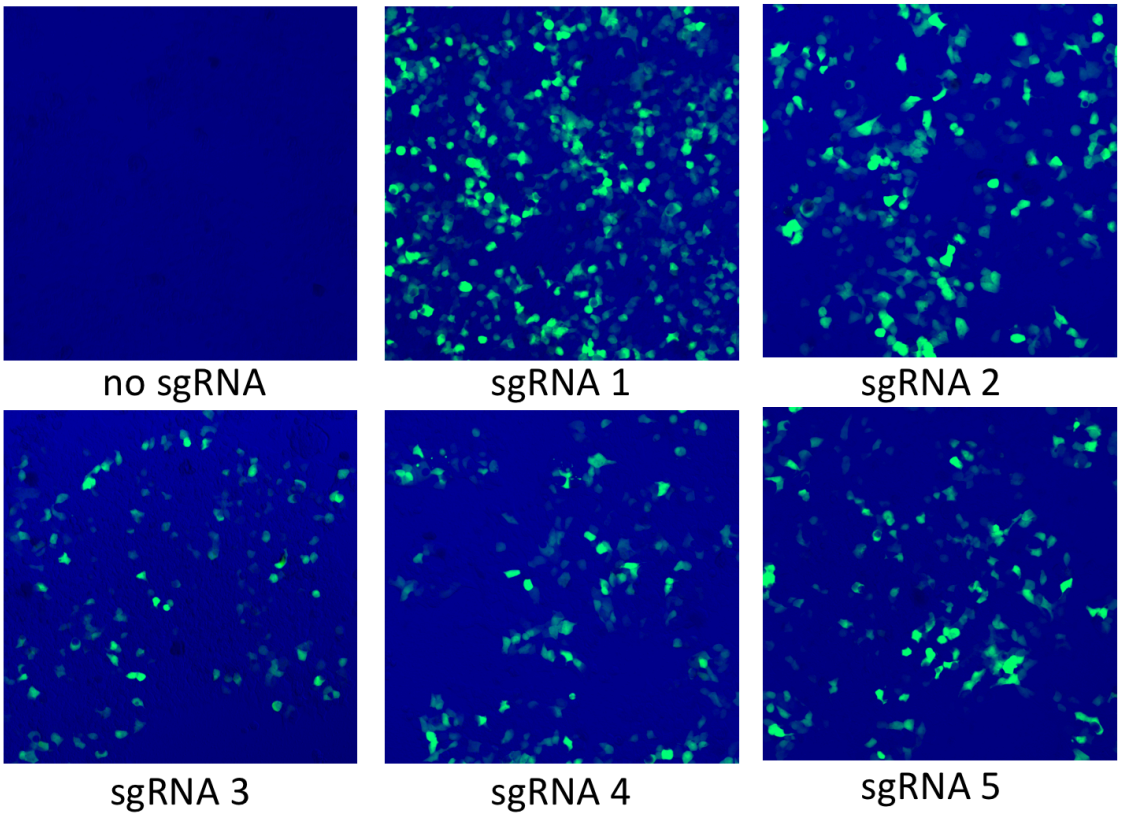


Figure 3.5 Screening of sgRNAs for efficiency. The top five sgRNA candidates were cloned into the pX330 plasmid while the C-terminus OGT target was cloned into the pCAG-EGxxFP construct. Then, the pX330-sgRNA and pCAG-EG_ogt_FP constructs were co-transfected into HEK293T cells. Upon CRISPR/Cas9/sgRNA mediated dsDNA cleavage and HDR, the cells fluoresced, indicating that all five sgRNA candidates were competent. Figure from Matt Griffin.

3.3.3 CRISPR/Cas9 to make novel OGT-FLAG-HA mice

After verifying the proper targeting of the sgRNA, we sought to design the ssODN used for homologous recombination to insert the short tags into the C-terminus of OGT. Based on previous studies exploring the efficacy of homologous recombination for small inserts following Crispr/Cas9 cleavage and our own needs, we sought to design the ssODN with the following criteria: (1) at least 60 bp of OGT homology arms on either side of the cut site, (2) inclusion of two novel cut sites for facilitation of PCR genotyping after incorporation, (3) codon optimization in order to ensure maximal translational efficiency, (4) removal of PAM sequence in the final construct (to prevent recutting by Crispr/Cas9), and (5) reintroduction of a stop codon to terminate the OGT sequence.[9] The final ssODN sequence along with its relevant features is shown in Figure 3.6.

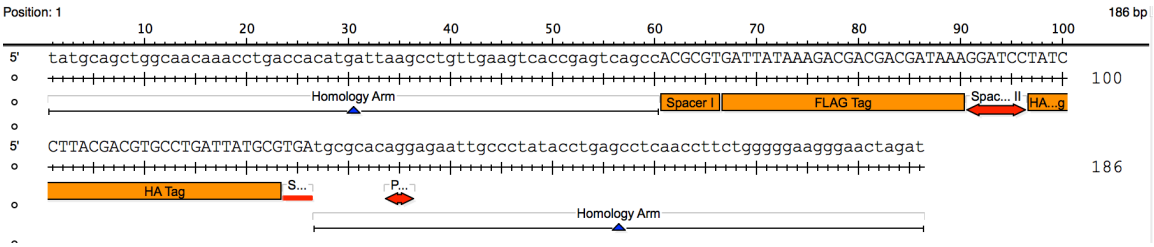
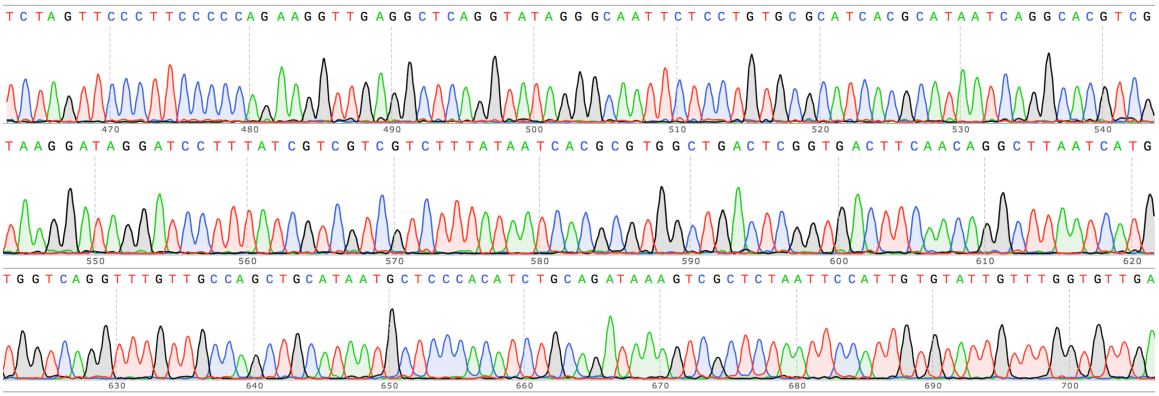


Figure 3.6 Sequence for ssODN for homologous recombination. The single-stranded DNA oligonucleotide was designed with the following components (from 5' to 3'): (1) 60 bases homologous to the C-terminus of OGT prior to the sgRNA cut site, (2) Spacer I (MluI cut site), (3) FLAG tag, (4) Spacer II (BamHI cut site), (5) HA tag, (6) stop codon, (7) 60 bases homologous to the region 3' to the original stop codon of OGT with the PAM sequence mutated out to prevent recutting by the CRISPR/Cas9/sgRNA system after proper HDR insertion.

In order to perform Crispr/Cas9 genome editing in mice, we followed the procedure previously published by Jaenisch and colleagues, where ssODN, Cas9 mRNA, and the sgRNA were directly injected into fertilized mouse C57B/L6 wildtype embryos.[9] Using pronuclear and cytoplasmic injections, only a single female mouse was heterozygous for the correct on-target sequence out of 10 total mice that were born although 4 died shortly after birth. This deviated significantly from the experimental efficiencies reported by Jaenisch and colleagues (30-50% efficiency).[9] Upon closer

examination, the Crispr/Cas9 targeting produced insertions and deletions in 5/10 mice (~5/12 chromosomes we were able to test) although only one had the appropriate insertion. This overall efficiency suggests that the Crispr/Cas9 targeting and cleavage had a high efficiency (~42%) while the homologous recombination efficiency was inefficient (<10%). Thus, the ssODN could have been optimized potentially with longer homology arms for example in order to maximize successful incorporation and reach the Jaenisch and colleagues reported efficiency (30-50%) for small insertion incorporation. After breeding this heterozygous female, we obtained the first group of homozygous mice (all males initially due to the OGT's location on the X chromosome). These homozygous mice exhibited no abnormalities in growth, behavior, or breeding and were indistinguishable from wildtype and heterozygous mice, consistent with our ultimate goal of integrating a minimally disruptive small tag. We verified the genotypes of the mice using on-target sequencing and on-target PCR (Figures 3.7 & 3.8). Importantly, the program provides the most likely off-target sites, which, in the case of sgRNA 1, was within a gene called *Rhox11*. We verified the proper sequence of *Rhox11* around the predicted sgRNA off-target site using sequencing (Figure 3.9). Finally, we verified expression of OGT-FLAG-HA protein in brain tissue of the adult homozygous and heterozygous mice using Western blot (Figure 3.10). In conclusion, we were able to generate a novel OGT-FLAG-HA (OGT-FH) mouse with the desired on-target incorporation and without off-target CRISPR/Cas9 mutation.



OGTFH7
GTCTGCAACACAGAACTACACAGATACACACAAAGCATGTTATCTAGTTCCCTTCCCCCA 480
ssODN ----------atctagttcccttccccca 19
****

OGTFH7
GAAGGTTGAGGCTCAGGTATAGGGCAATTCTCCTGTGCGCATCACGCATAATCAGGCACG 540
ssODN gaaggttgaggctcaggtatagggcaattctcctgtgcgcaTCACGCATAATCAGGCACG 79
****

OGTFH7
TCGTAAGGATAGGATCCTTTATCGTCGTCGTCTTTATAATCACGCGTGGCTGACTCGGTG 600
ssODN TCGTAAGGATAGGATCCTTTATCGTCGTCGTCTTTATAATCACGCGTggctgactcggtg 139
****

OGTFH7
ACTTCAACAGGCTTAATCATGTGGTCAGGTTTGTTGCCAGCTGCATAATGCTCCCACATC 660
ssODN acttcaacaggcttaatcatgtggtcaggtttgttgccagctgcata---------- 186
****

Legend
3' C-terminal OGT homology arm Stop codon HA tag BamHI cut site FLAG tag MluI cut site 5' C-terminal OGT homology arm

Figure 3.7 Sequencing validation of tag insert. After one round of breeding, we obtained homozygous mice and were able to verify by sequencing the proper insertion of the tag. Shown here is the autoscaled sequencing results for the first homozygous mouse produced in SnapGene® Viewer 4.0.5. The homology alignment was performed using Clustal Omega version 1.2.4.

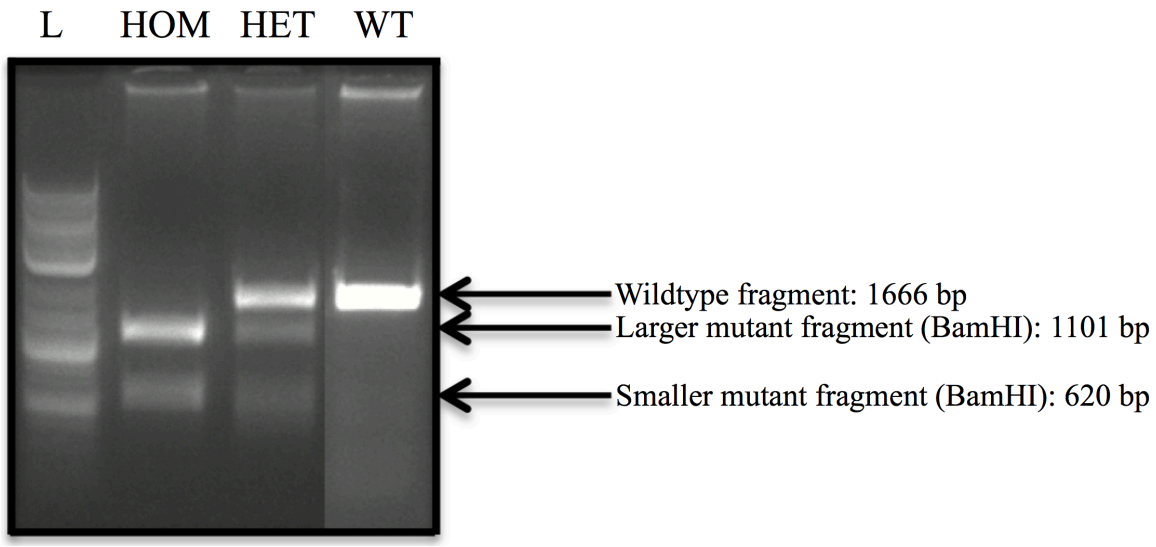
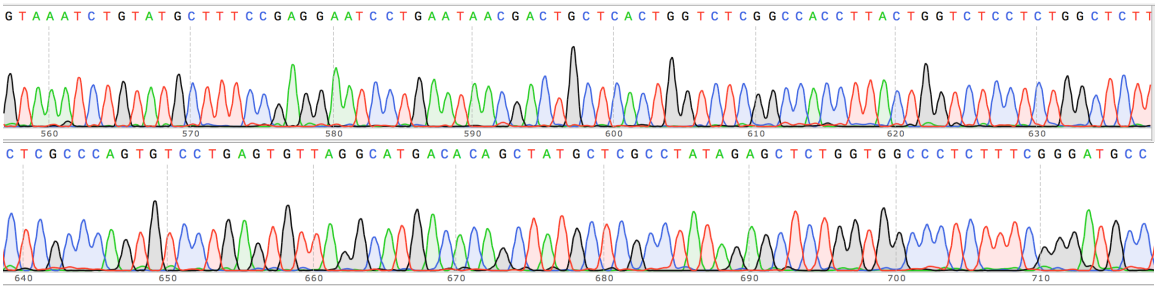


Figure 3.8 PCR and gel genotyping. We designed primers that amplified the regions surrounding the inserted tag. Then, the amplified DNA was cleaved using the BamHI restriction enzyme and then run on a 1% DNA agarose gel. Sample genotyping results are shown here where the fragment sizes are (1) 1666 bp for the wildtype, (2) 1101 bp for the larger mutant fragment, and (3) 620 bp for the smaller mutant fragment. As indicated, homozygous (HOM) mice have 2 distinct bands (2 and 3), heterozygous (HET) mice have three distinct bands (1, 2, and 3), and the wildtype (WT) mice have a single band (1).



```
off_F           actgattttctacgaacactgcttgcagttcccacagctgtccaggtgtaaatctgtatg          8100          2_F
ACTGATTTTCTACGAACACTGCTTGCAGTTCCCACAGCTGTCCAGGTGTAAATCTGTATG 571
    ************************************************************                  off_F
ctttccgaggaatcctgaataacgactgctcactggtctcggccaccttactggtctcct                8220          2_F
CTTTCCGAGGAATCCTGAATAACGACTGCTCACTGGTCTCGGCCACCTTACTGGTCTCCT 631
    ************************************************************
off_F ctggctcttctcgcccagtgtcctgagtgttaggcatgacacagctatgctcgcctatag 8280
2_F    CTGGCTCTTCTCGCCCAGTGTCCTGAGTGTTAGGCATGACACAGCTATGCTCGCCTATAG
751
          ************************************************************
```

Figure 3.9 Sequencing validation of most likely off-target site. We validated that the most likely off-target site had no discernible insertions or deletions using the original heterozygous OGT-FH we obtained. The most likely target site is highlighted in yellow. Shown here is the autoscaled sequencing results for the first homozygous mouse produced in SnapGene® Viewer 4.0.5. The homology alignment was performed using Clustal Omega version 1.2.4.

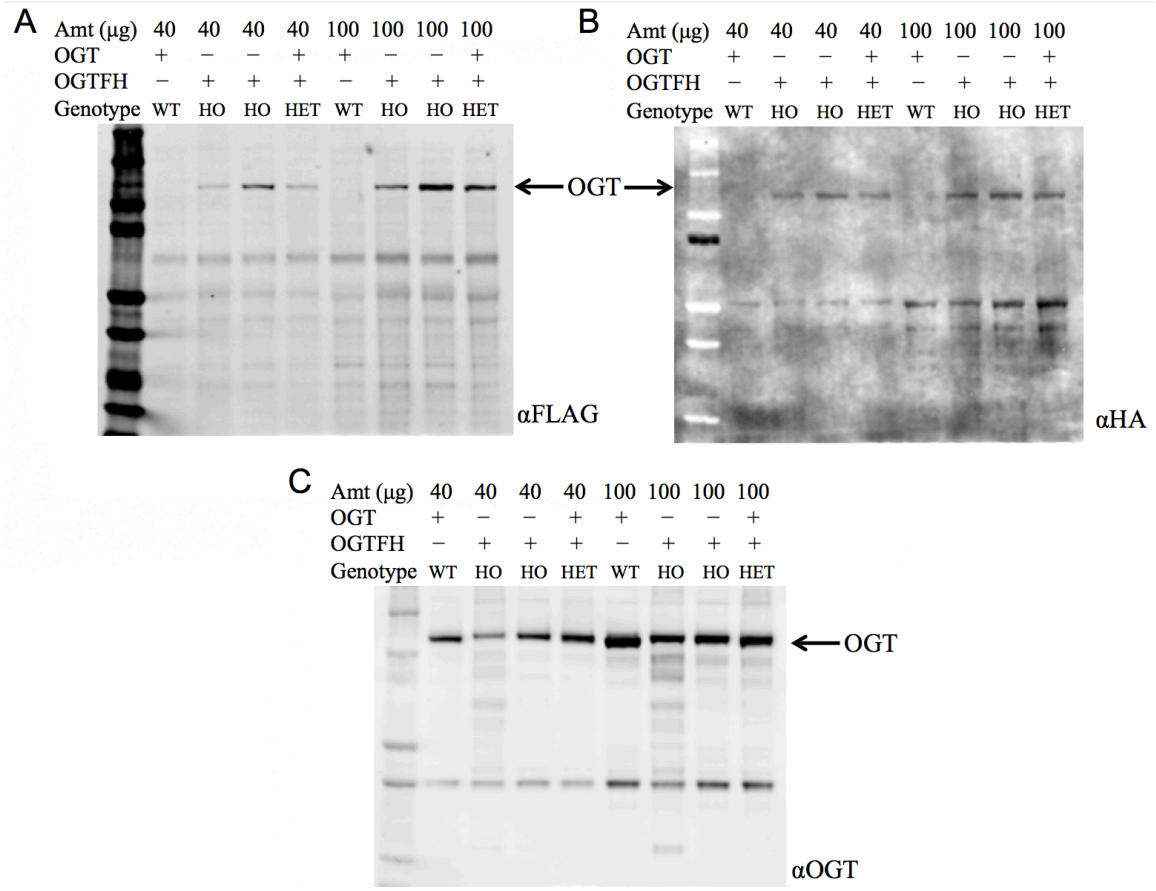


Figure 3.10 Western blotting verification of OGT-FLAG-HA protein expression in mice. We took 40 µg and 100 µg of adult (2 month old) mouse brain lysate of homozygous (HO), heterozygous (HET), and wildtype (WT) littermates. The homozygous and heterozygous, but not wildtype mice show expression of FLAG-tagged OGT (A) and HA-tagged OGT (B). (C) Homozygous, heterozygous, and wildtype mice showed expression of OGT.

3.3.4 OGT interactome preliminary results

After verification and validation of our OGT-FLAG-HA mouse, we began to optimize lysis conditions in order to ensure lysis of nuclear proteins while maintaining protein-protein interactions. We found that 0.5-1% (vol/vol) NP-40 and Triton X-100 were both capable of lysing brain tissue with douncing (Figure 3.11). After optimizing the tandem affinity pull down followed by differential labeling of the OGT and OGT-FH peptides, we were able to get a preliminary list of 76 proteins in the brain OGT interactome using a cutoff SILAC ratio of at least 5.0 (data from Yao Xiao). Shown below is the gene ontology enrichment for this initial list of 76 proteins (Table 3.1). The

proteins were enriched for phosphorylated (68), acetylated (32), methylated (21), and ubiquitylated (14) proteins consistent with the literature showing considerable crosstalk between *O*-GlcNAc glycosylation and other post-translational modifications.[1,10,11] Furthermore, the interactors were enriched for cytoskeletal proteins including those previously reported to be *O*-GlcNAc glycosylated such as bassoon (BSN); piccolo (PCLO); microtubule associated proteins 1B, 2, and 6 (MAP1B, MAP2, MAP6); α-internexin (INA); tubulin α1c, β2A, and β3 chains (TUBA1C, TUBB2A, TUBB3); and actin β (ACTB).[12] In a similar vein, we found several proteins involved in cytoskeletal dynamics that are known to be *O*-GlcNAc glycosylated: α-actinin 1 (ACTN1); spectrin α and β chains (SPTAN1, SPTBN1); myosin-10 (MYH10); unconventional myosin-Va (MYO5A); trafficking kinesin-binding proteins 1 and 2 (TRAK1, TRAK2); ankyrin-3 (ANK3); dynamin-1 (DNM1); and erythrocyte membrane protein band 4.1-like proteins 1 and 3 (EPB41L1, EPB41L3).[13-16] As discussed in Chapter 1, OGT has been shown to interact with and *O*-GlcNAcylate TRAK1/Milton, leading to changes in mitochondrial motility and OGT substrate *O*-GlcNAcylation.[17,18]

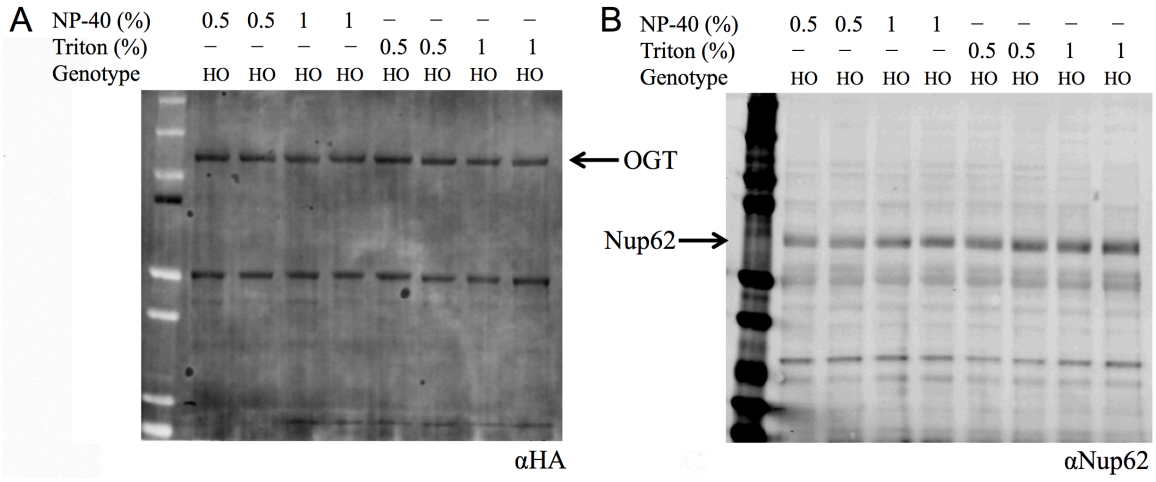


Figure 3.11 Lysis condition screen for brain tissue. We looked at the ability of different mild lysis conditions to lyse OGT and also the nuclear protein and known OGT interactor and substrate, Nup62.

Table 3.1 Preliminary OGT interactome from OGT-FH brain.

GO term	Proteins	FDR
Phosphoprotein	Atp1b1, Srcin1, Tubb2a, Bap1, Hsph1, Atp2b4, Ash2l, Ank3, Wdr77, Stk39, Agap2, Nsf, Tubb3, Jakmip1, Ina, Caskin1, Rbbp5, Arhgef2, Slc25a4, Pfkp, Wnk1, Bsn, Pfkm, Pclo, Mfn2, Mapk1, Ksr2, Atp5c1, Bin1, Carm1, Hcfc1, Klc2, Brsk1, Cxxc1, Camkv, Tpi1, Klc1, Ppp1r12a, Ppp3cb, Camk2b, Ppp3ca, Hspa5, Clasp2, Gapdh, Iqsec1, Tubb4a, Rtcb, Tubb4b, Plp1, Ncdn, Immt, Map1b, Csnk2b, Atp1a3, Atp1a2, Dpysl2, Ywhah, Map2, Hivep3, Pspc1, Ywhaq, Hivep2, Ahcyl1, Hivep1, Atp5a1, Map6, Hnrnph1, Dnm1	$1.8x10^{-21}$
Myelin sheath	Ina, Plp1, Atp1b1, Slc25a4, Slc25a5, Immt, Atp1a3, Dpysl2, Atp1a2, Slc25a12, Atp5c1, Atp5a1, Hspa5, Gapdh, Dnm1, Nsf, Tubb4a, Tubb4b	$5.5x10^{-16}$
Cytoplasm	Srcin1, Tubb2a, Hcfc1, Bap1, Brsk1, Klc2, Hsph1, Klc1, Ank3, Wdr77, Ppp1r12a, Camk2b, Stk39, Clasp2, Hspa5, Agap2, Gapdh, Iqsec1, Nsf, Tubb4a, Tubb3, Rtcb, Jakmip1, Tubb4b, Caskin1, Arhgef2, Ncdn, Map1b, Wnk1, Pfkp, Bsn, Dpysl2, Pfkm, Mapk1, Tet3, Ksr2, Ywhah, Map2, Hivep3, Pspc1, Ywhaq, Ahcyl1, Map6, Carm1, Bin1, Dnm1	$6.9x10^{-12}$
Protein binding	Atp1b1, Srcin1, Bap1, Wbp2, Hsph1, Ash2l, Ank3, Wdr77, Stk39, Agap2, Tubb3, Nsf, Ina, Rbbp5, Arhgef2, Slc25a4, Wnk1, Bsn, Pfkm, Pclo, Mfn2, Mapk1, Ksr2, Bin1, Carm1, Hcfc1, Klc2, Brsk1, Klc1, Ppp1r12a, Camk2b, Ppp3ca, Hspa5, Clasp2, Plp1, Ncdn, Immt, Map1b, Csnk2b, Dpysl2, Atp1a2, Tet3, Ywhah, Map2, Ywhaq, Pspc1, Ahcyl1, Atp5a1, Dnm1	$4.8x10^{-11}$
Microtubule	Arhgef2, Tubb2a, Map1b, Klc2, Dpysl2, Hsph1, Klc1, Map2, Map6, Clasp2, Bin1, Tubb3, Tubb4a, Dnm1, Tubb4b, Jakmip1	$1.4x10^{-9}$
Methylation	Caskin1, Srcin1, Ncdn, Slc25a5, Tubb2a, Map1b, Bsn, Hcfc1, Dpysl2, Brsk1, Tpi1, Tet3, Atp2b4, Ash2l, Pspc1, Atp5a1, Hspa5, Hnrnph1, Carm1, Gapdh, Dnm1	$2.2x10^{-8}$
Cytoskeleton	Arhgef2, Srcin1, Tubb2a, Map1b, Bsn, Klc2, Dpysl2, Brsk1, Mapk1, Klc1, Ank3, Map2, Camk2b, Clasp2, Map6, Gapdh, Dnm1, Tubb3, Tubb4a, Tubb4b, Jakmip1	$1.6x10^{-7}$
Acetylation	Tubb2a, Hcfc1, Cxxc1, Hsph1, Tpi1, Ppp3cb, Stk39, Hspa5, Ppp3ca, Gapdh, Nsf, Tubb4b, Ina, Arhgef2, Ncdn, Slc25a4, Immt, Slc25a5, Map1b, Pfkp, Csnk2b, Pfkm, Slc25a12, Mapk1, Ywhah, Ywhaq, Pspc1, Atp5c1, Ahcyl1, Atp5a1, Hnrnph1, Bin1	$3.3x10^{-6}$
Nucleotide-binding	Tubb2a, Pfkp, Wnk1, Atp1a3, Brsk1, Pfkm, Atp1a2, Mfn2, Mapk1, Hsph1, Atp2b4, Ksr2, Rhot1, Stk39, Camk2b, Atp5a1, Hspa5, Agap2, Dnm1, Nsf, Tubb3, Tubb4a, Rtcb, Tubb4b	$4.9x10^{-6}$
Coiled coil	Srcin1, Hcfc1, Bap1, Klc2, Cxxc1, Hsph1, Atp2b4, Trak2, Klc1, Ank3, Ppp1r12a, Clasp2, Hspa5, Iqsec1, Tubb3, Jakmip1, Tubb4b, Ina, Arhgef2, Immt, Map1b, Wnk1, Bsn, Atp1a2, Pclo, Mfn2, Tet3, Pspc1, Hivep3, Bin1	$4.4x10^{-5}$
Axon	Mapk1, Srcin1, Klc1, Ank3, Map1b, Atp1a3, Bsn, Hcfc1, Dpysl2, Bin1, Pclo, Tubb3	$1.6x10^{-4}$
cGMP-PKG signaling pathway	Mapk1, Atp1b1, Atp2b4, Slc25a4, Slc25a5, Atp1a3, Ppp3cb, Ppp1r12a, Ppp3ca, Atp1a2	$1.9x10^{-4}$
Microtubule binding	Arhgef2, Map2, Map1b, Map6, Clasp2, Dpysl2, Gapdh, Dnm1, Jakmip1	$2.6x10^{-3}$
Beta tubulin	Tubb2a, Tubb4a, Tubb3, Tubb4b	$3.1x10^{-3}$
Neuron projection	Atp2b4, Srcin1, Ncdn, Klc1, Ank3, Map2, Bsn, Klc2, Atp1a2, Dpysl2, Pclo	$4.7x10^{-3}$
Neuronal cell body	Arhgef2, Srcin1, Ncdn, Klc1, Map2, Map1b, Bsn, Hcfc1, Dpysl2, Pclo, Tubb4a, Tubb3	$5.9x10^{-3}$
Calmodulin-	Camkv, Atp2b4, Map2, Ppp3cb, Camk2b, Ppp3ca, Map6	$7.4x10^{-3}$

binding		
Set1c/ COMPASS complex	Rbbp5, Ash2l, Hcfc1, Cxxc1	$1.0x10^{-2}$
Protein kinase binding	Mapk1, Srcin1, Ppp1r12a, Wnk1, Stk39, Camk2b, Brsk1, Dpysl2, Agap2, Dnm1, Nsf	$1.8x10^{-2}$
Dendrite	Srcin1, Ncdn, Ank3, Map2, Map1b, Bsn, Hcfc1, Camk2b, Dpysl2, Pclo, Tubb3	$1.8x10^{-2}$
Substantia nigra development	Ina, Plp1, Ywhah, Ywhaq, Hspa5	$2.5x10^{-2}$
Microtubule-based process	Tubb2a, Map1b, Tubb4a, Tubb3, Tubb4b	$3.4x10^{-2}$
ATP-binding	Pfkp, Wnk1, Atp1a3, Atp1a2, Pfkm, Brsk1, Mapk1, Hsph1, Atp2b4, Ksr2, Camk2b, Stk39, Atp5a1, Hspa5, Nsf, Rtcb	$3.4x10^{-2}$
GTPase activity	Mfn2, Tubb2a, Rhot1, Agap2, Tubb4a, Dnm1, Tubb3, Tubb4b	$3.7x10^{-2}$
Postsynaptic density	Arhgef2, Srcin1, Map2, Map1b, Bsn, Camk2b, Pclo, Nsf	$4.0x10^{-2}$
Extracellular vesicle	Atp1b1, Tubb2a, Atp1a3, Atp1a2, Tubb4b	$4.3x10^{-2}$

Table 3.1 Shown here are the top GO categories for the significantly enriched OGT interactome proteins. FDR = false discovery rate. OGT-FH mouse brain MS data from Yao Xiao.

If we lowered the cutoff to an average SILAC ratio of 5.0, the list becomes 188 putative OGT interactors, which are heavily enriched for ribosomal proteins (*FDR* < $2.7x10^{-23}$, 29 proteins, 17.4-fold enrichment). This is consistent with previous studies showing that the majority of ribosomal proteins appear to glycosylated in addition to translational regulators such as the elongation factors Tu translation elongation factor, mitochondrial (TUFM) and elongation factor 1-α1 (EEF1A1).[19,20] We also compared the brain OGT interactome to the OGT interactomes found in a tandem affinity purification experiment performed in HEK293T cells and a human protein microarray experiment (Figure 3.12). The top two gene ontology general categories for the overlapping proteins in the two tandem affinity purification experiments were ribosomal and cytoskeletal proteins suggesting that OGT interacts with these highly abundant proteins across cell types and species (Benjamini-corrected $P < 1.8x10^{-19}$ and $P < 1.3x10^{-6}$ respectively). In addition, we saw 75 of the 188 proteins in our OGT interactome overlapping with known *O*-GlcNAcylated proteins (Figure 3.13). The top annotation cluster 1 for those

overlapping 75 proteins includes dendritic, axonal, and neuronal cell body proteins (FDR < $2.9x10^{-5}$, $1.2x10^{-3}$, $4.6x10^{-3}$ respectively) (Table 3.2). The next two annotation clusters were enriched for calmodulin binding and cytoskeletal proteins (FDR < $1.7x10^{-5}$ and $2.3x10^{-4}$) (Table 3.2). In conclusion, our preliminary OGT brain interactome results reveal similar proteins as identified in other OGT interactome and *O*-GlcNAcome studies as well as some novel protein interactors. Experiments are ongoing within the lab to verify the interaction of OGT with these putative interactors and test to see if the *O*-GlcNAcome is influenced by the interaction of OGT with these interactors.

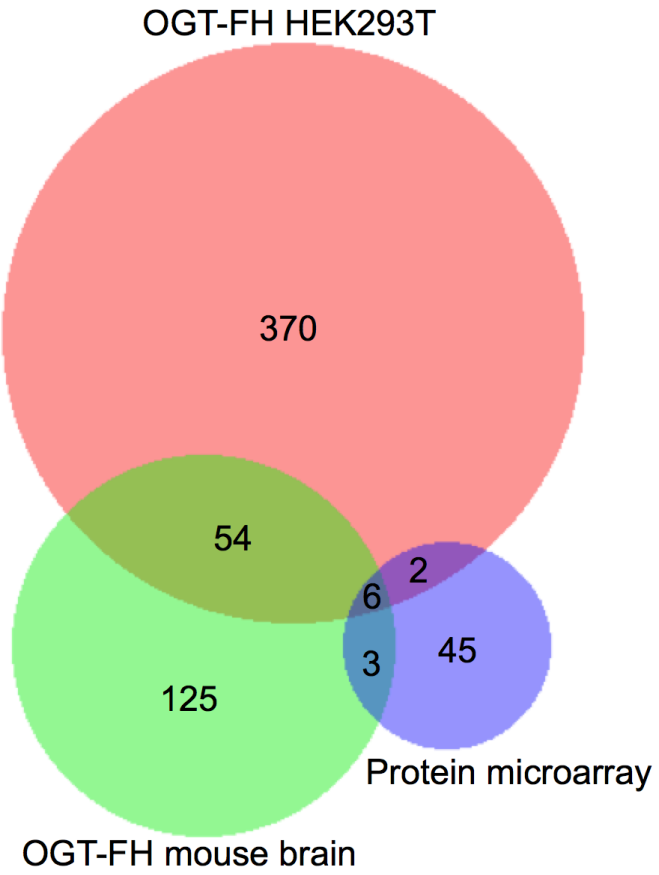


Figure 3.12 Comparison between OGT-FH mouse brain OGT interactome and other OGT interactome experiments. Shown here is a Venn Diagram displaying the number of common proteins in our OGT interactome study and other OGT interactome studies.[21,22] OGT-FH mouse brain MS data from Yao Xiao.

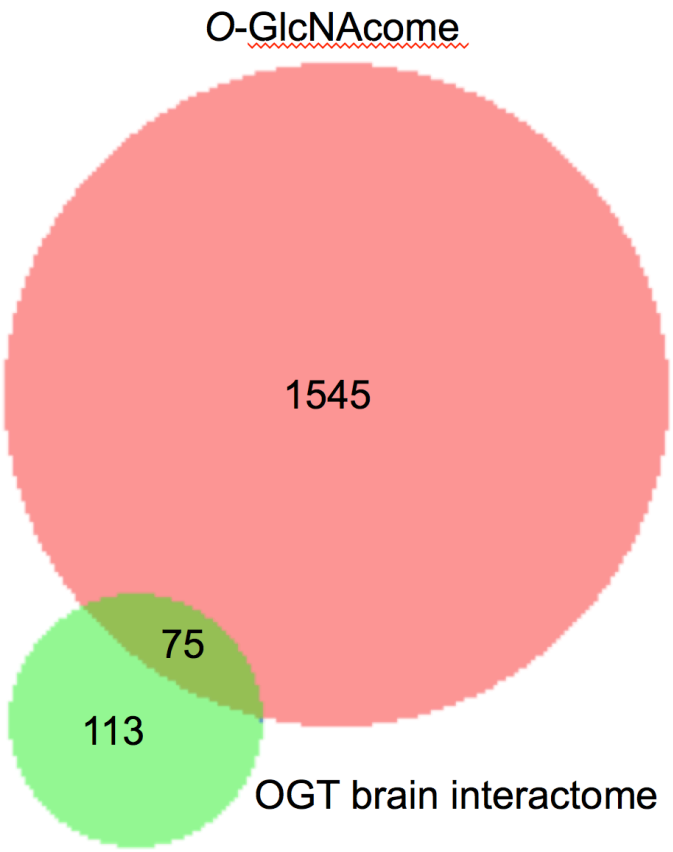


Figure 3.13. Comparison between *O*-GlcNAcome and the OGT brain interactome. Here is a Venn Diagram overview of the *O*-GlcNAcome and the OGT brain interactome. The *O*-GlcNAcome was obtained from several different species, cell types, and methods as previously described.[12,14-16,23,24] OGT-FH mouse brain MS data from Yao Xiao.

Table 3.2 Top 5 GO clusters for OGT brain interactome and *O*-GlcNAcome common proteins.

GO Term	#	FDR	Proteins	FE
Dendrite	14	2.89×10^{-5}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	7.7
Axon	11	1.20×10^{-3}	Myo5a, Actb, Ank3, Map1b, Bsn, Hcfc1, Pclo, Camk2a, Ywhae, Tubb3, Myh10	8.0
Neuronal cell body	12	4.61×10^{-3}	Myo5a, Ncdn, Map2, Map1b, Bsn, Hcfc1, Ogt, Pclo, Camk2a, Tubb3, Mbp, Myh10	6.1
Calmodulin-binding	9	1.67×10^{-5}	Myo5a, Camkv, Map2, Sptbn1, Ppp3ca, Map6, Camk2a, Myh10, Sptan1	20
Cytoskeleton	17	2.27×10^{-4}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	4.9
Microtubule	8	2.46×10^{-2}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb5	9.4
Nucleotide-binding	23	1.51×10^{-5}	Myo5a, Ncdn, Map1b, Bsn, Hcfc1, Rps6, Pclo, Rps3, Gnb1, Ank3, Map2, Pabpc1, Camk2a, Tubb3	4.1
Postsynaptic density	11	2.06×10^{-5}	Actb, Epb41l3, Bcas1, Map2, Map1b, Bsn, Sptbn1, Pclo, Camk2a, Nsf, Dclk1	12

Table 3.2 shows the top gene ontology (GO) clusters for the overlapping proteins between the OGT brain interactome and the *O*-GlcNAcome. The clusters are indicated by color: red (cluster 1: enrichment 6.4), orange (cluster 2: enrichment 6.1), green (cluster 3: enrichment 5.0), blue (cluster 4: enrichment 4.4), purple (cluster 5: enrichment 3.3). Categories with FDR > 0.05 or that were redundant were removed from

the table. FE = Fold Enrichment; FDR = False Discovery Rate. OGT-FH mouse brain MS data from Yao Xiao.

3.4 Development of chemoenzymatic tools for the *O*-GlcNAcome

3.4.1 Overview of chemoenzymatic approach

Robust enrichment of O-GlcNAcylated proteins can be accomplished using a two-step chemoenzymatic approach.[25,26] First, the O-GlcNAc moiety is tagged with a non-natural azide group by treatment of cell lysates with UDP-GalNAz **1** and a mutant glycosyltransferase (Y289L GalT) that specifically recognizes terminal GlcNAc moieties.[27] Next, a biotin group is attached via copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), which allows for affinity purification. Although a limited set of alkyne-biotin linkers are commercially available, many existing linkers are not ideal for mapping O-GlcNAc modification sites.[28] In particular, harsh conditions are usually required to disrupt the femtomolar biotin-streptavidin interaction, which may hydrolyze the labile O-GlcNAc moiety.[29] Additionally, many linkers contain a large spacer between the biotin group and the alkyne functionality, which appends a relatively large mass to the glycopeptide and can complicate its identification.[29] Therefore, a facile method to release the labeled peptides and proteins with minimal added mass would greatly facilitate downstream analysis.

Several cleavable linkers have been previously developed for enrichment of O-GlcNAcylated proteins.[30-33] However, each suffers from significant drawbacks for site identification. For example, a photocleavable linker was employed in conjunction with UDP-GalNAz **1** and Y289L GalT to sequence modified peptides from mouse brain lysate.[30,31] Importantly, the moiety retained after cleavage provided a positively-charged amine group, which increased the overall peptide charge and facilitated ionization by

electron-transfer dissociation (ETD), the most successful MS/MS method for O-GlcNAc peptide sequencing.[34,35] Unfortunately, cleavage of the linker was found to be incomplete.[30] In a recent report, a dibromine-containing, acid-cleavable linker was employed to identify various glycan modifications including O-GlcNAc.[32] However, cleavage of the linker revealed only a neutral hydroxyl group, and the halogenated glycopeptides demonstrated poor fragmentation efficiency using ETD. Therefore, we aimed to develop a tag that would be both quantitatively released and incorporate a positive charge upon cleavage to facilitate MS detection by ETD.

3.4.2 Validation of Dde cleavable linker

We aimed to develop a tag that would be both quantitatively released and incorporate a positive charge upon cleavage to facilitate MS detection by ETD. To achieve these dual goals, we chose to examine the 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl (Dde) functional group (Figure 3.14A). The Dde moiety has been used extensively as a protecting group for lysine in peptide synthesis, demonstrating its compatibility with biomolecules.[36] The group is stable to both acid and base and can be quantitatively removed by hydrazine.[37] However, it was reported that the Dde group is incompatible with sodium dodecyl sulfate (SDS) and amine-containing buffers, common additives to protein labeling protocols.[38]

We first investigated the labeling of a model O-GlcNAcylated peptide with commercially available alkyne-Dde-biotin **2** followed by cleavage of the linker using liquid chromatography (LC)-MS (Figure 3.14B and Figure 3.15). Commercially available peptide TAPT(gS)TIAPG (Figure 3.15A), where gS is the O-GlcNAcylated residue, was incubated with 100 ng/μL Y289L GalT and 1 mM UDP-GalNAz of **1** in 10 mM HEPES

pH 7.9, 5.5 mM MnCl2 overnight at 4 °C. LC-MS analysis revealed quantitative conversion to the desired GalNAz-labeled product (Figure 3.15B). Next, the azide-containing peptide was reacted with 100 µM of **2** in 10 mM sodium phosphate pH 7.6 containing 2 mM sodium ascorbate (NaAsc), 100 µM THPTA, and 1 mM $CuSO_4$. After 1 h, stoichiometric biotinylation of the peptide was observed (Figure 3.15C). Treatment with 2% aqueous hydrazine for 1 h at RT resulted in quantitative cleavage of the linker to afford a minimal, positively-charged aminomethyltriazolyl group (Figures 3.14B and 3.15D). To test whether the linker would be stable under stringent wash conditions, we incubated the labeled peptide with 1% RapiGest, a MS-compatible analogue of SDS, or 6 M urea for 1 h at RT (Figure 3.16). In both cases, the linker remained intact, highlighting the compatibility of the linker with rigorous washing steps.

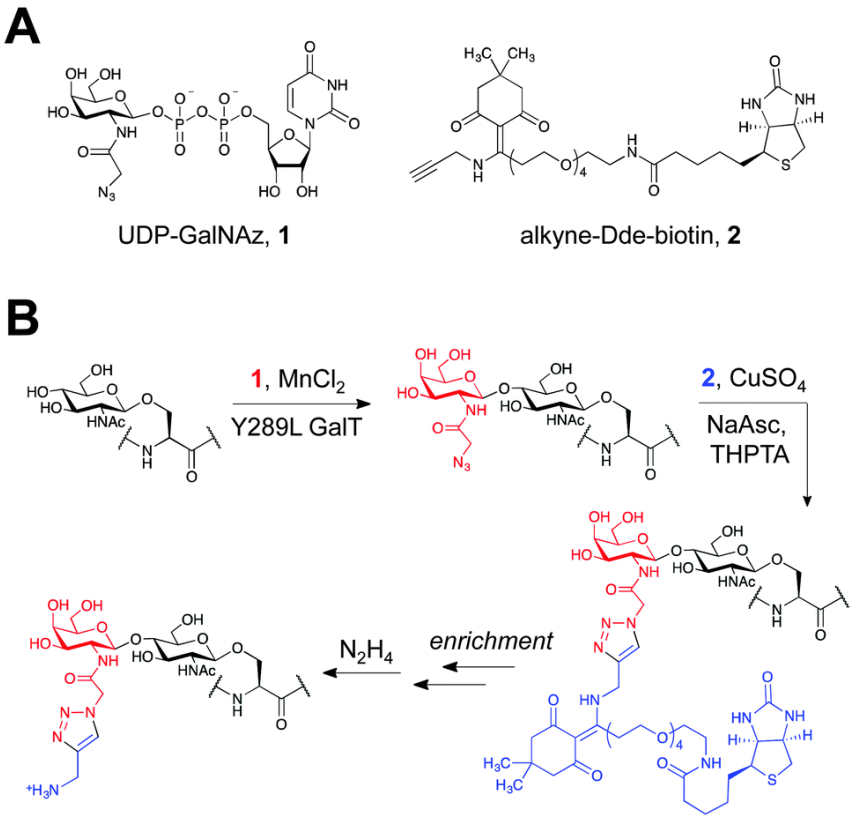


Figure 3.14 Overview of chemicals and workflow of chemoenzymatic linker labeling. (A) Chemicals used in the labeling protocol. (B) Schematic of *O*-GlcNAc protein enrichment and elution using two-step chemoenzymatic/CuAAC labeling protocol.

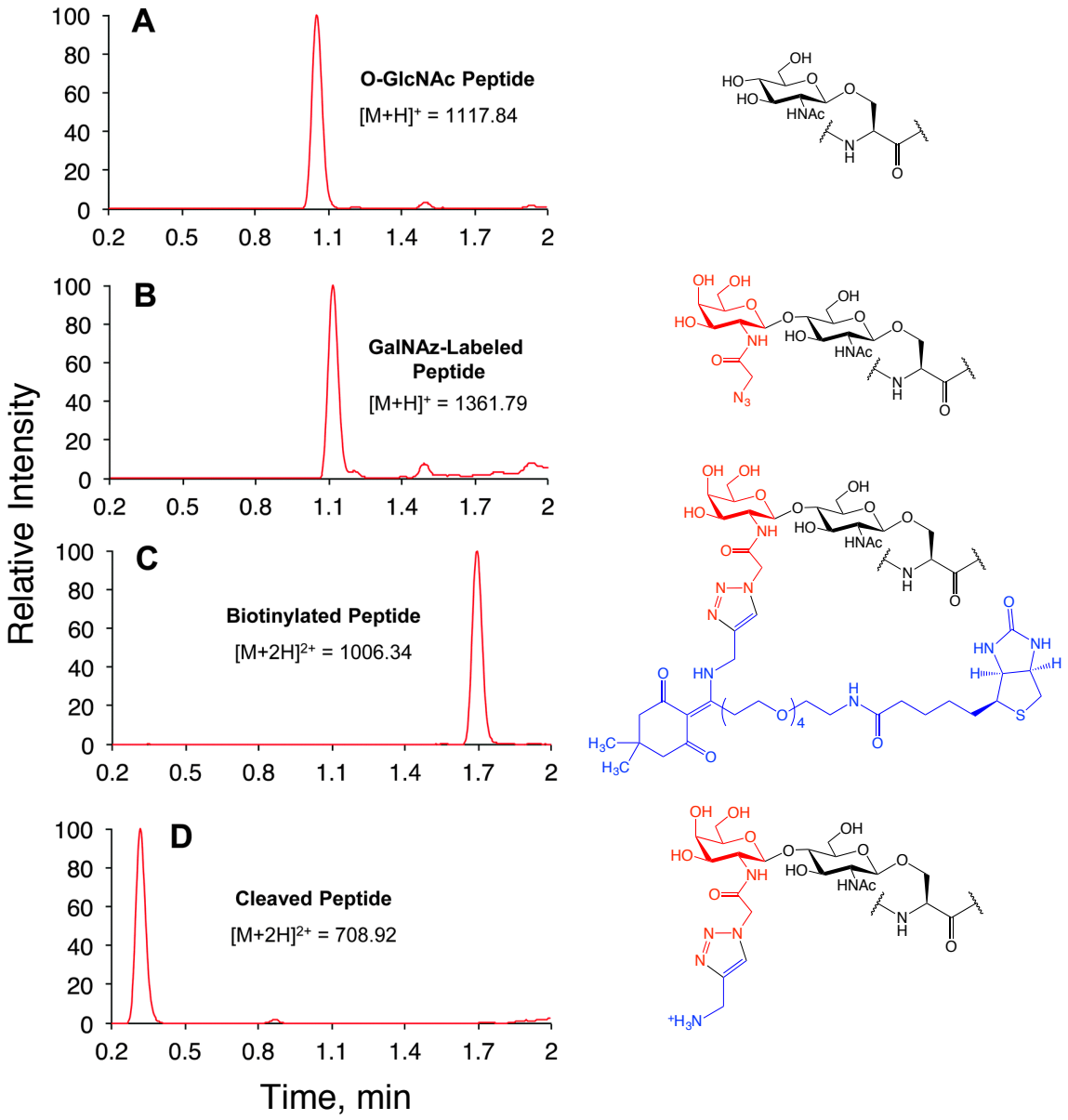


Figure 3.15 Labeling and cleavage reactions proceed quantitatively. Reverse phase LC-MS analysis of O-GlcNAc peptide labeling reactions at (A) time 0, (B) 16 h after addition of 1 and Y289L GalT, (C) 1 h after CuAAC with 2, and (D) 1 h after cleavage with 2% aqueous hydrazine. See methods for experimental details. (A) and (B) show base peak chromatograms. (C) and (D) show extracted ion chromatograms of the starting material and product within ±1 m/z of calculated values.

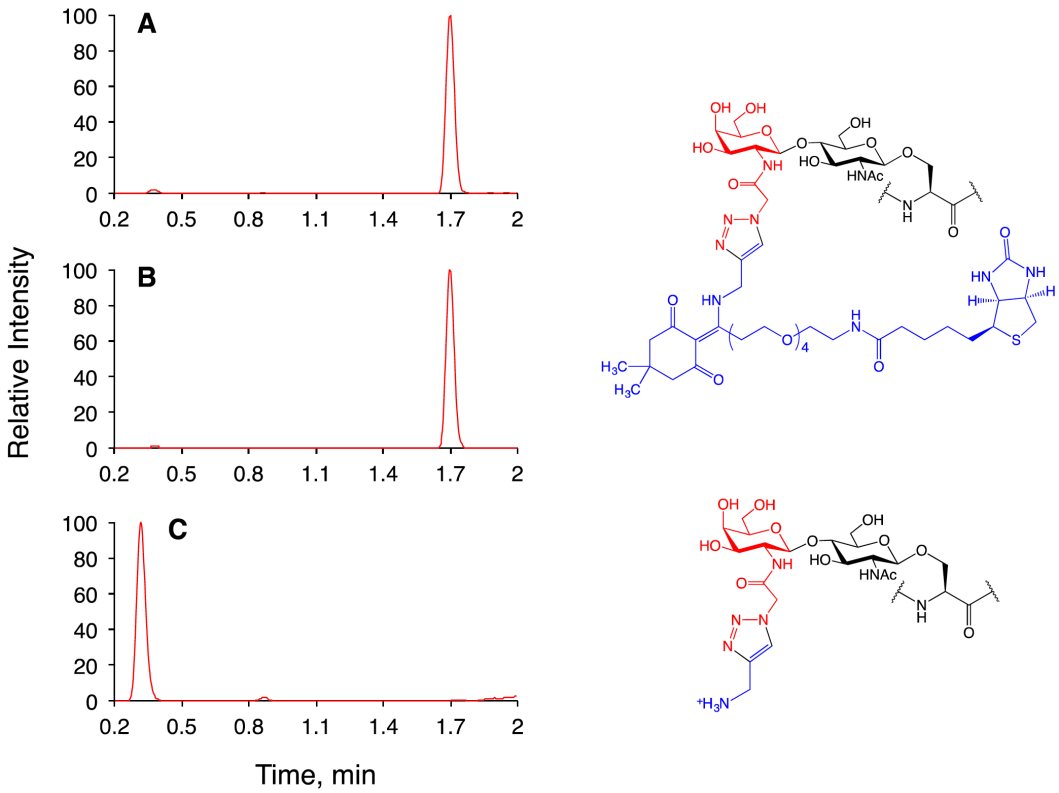


Figure 3.16 Labeled peptide is stable to wash conditions. Reverse-phase LC-MS analysis of alkyne-Dde-biotin-labeled peptide after 1 h incubation with (A) 6 M urea, (B) 1% RapiGest, or (C) 2% hydrazine. All graphs show extracted ion chromatograms of the starting material and possible product within ± 1 *m/z*.

3.4.3 Comparison with a photocleavable linker

We then tested the performance of our linker in comparison to the previously described photocleavable linker (alkyne-PC-biotin).[30] Briefly, 500 µg HEK-293T cell lysate was subjected to chemoenzymatic labeling with **1** using Y289L GalT as described above. The azide-labeled protein was split into two fractions of 200 µg and reacted with **2** or alkyne-PC-biotin by CuAAC. A sample of each biotinylated mixture was reserved, and the remaining material was subjected to cleavage by 2% hydrazine monohydrate or UV irradiation by 365 nm. Samples were resolved by SDS-PAGE and probed for biotin using fluorescently tagged streptavidin (Figure 3.17). Notably, lysate labeled by **2** showed a stronger biotin signal than the photocleavable linker, suggesting higher labeling efficiency. Furthermore, although both linkers cleaved well, slightly higher residual

signal was observed for the photocleavable linker compared to **2**, implying that the Dde moiety was also released more efficiently than the photocleavable group.

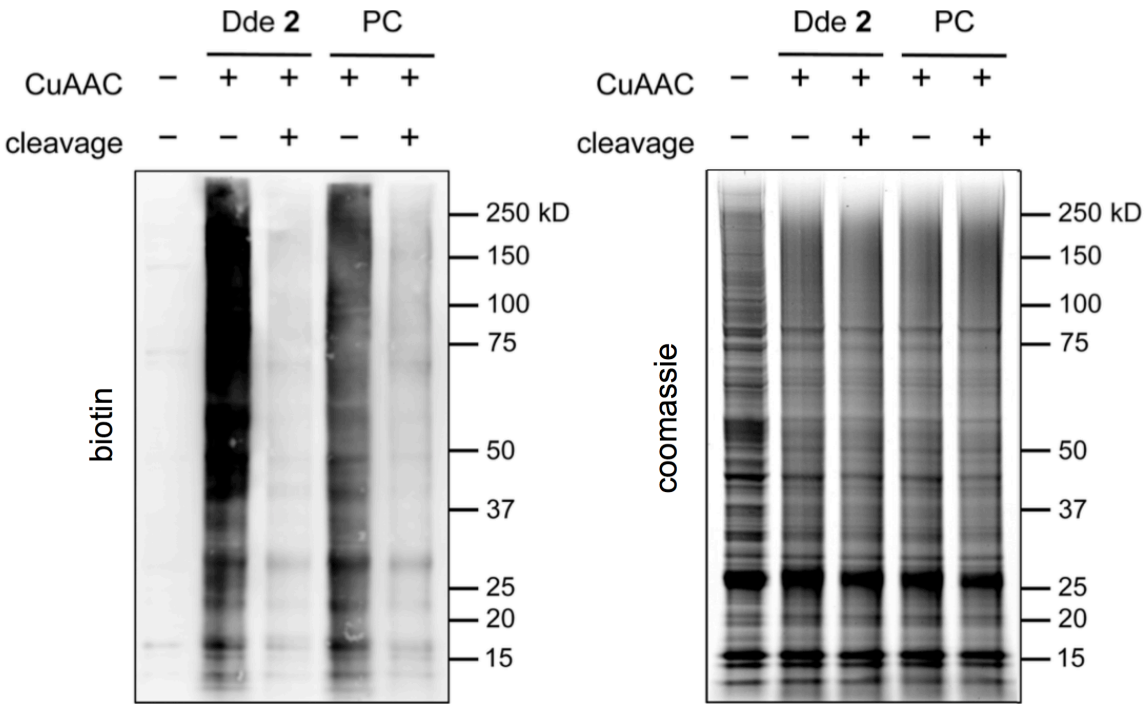


Figure 3.17 Dde linker 2 outperforms photocleavable linker. Protein (20 μg) labeled with alkyne-Dde-linker **2** shows both higher biotin signal after labeling (lane 2 vs. lane 4) and lower residual signal after cleavage (lane 3 vs. lane 5) compared to alkyne-PC-linker (PC). Coomassie staining on the right confirms equal protein loading in each lane.

3.4.4 Validation using known *O*-GlcNAcylated proteins- α-crystallin and *O*-GlcNAc transferase

We next evaluated the potential of the approach to pull down known *O*-GlcNAcylated proteins and identify sites of modification. The well-characterized *O*-GlcNAcylated protein α-crystallin was selected to assess the sensitivity of the method because it has a relatively low glycosylation stoichiometry (< 10%).[39] Short-form OGT (sOGT) has multiple sites of *O*-GlcNAcylation and was thus used to determine whether comprehensive site mapping could be achieved.[35,40,41] To test the robustness of our method in a complex mixture, each protein was added to 200 μg of adult mouse cortical lysate and subjected to chemoenzymatic labeling and CuAAC using **2**. The labeled

proteins were applied to high-capacity Neutravidin resin and washed five times each with 0.5 mL of 1% SDS, 6 M urea, and phosphate buffered saline (PBS). The resin was then rotated end-over-end for 1 h with 2% aqueous hydrazine to cleave the O-GlcNAcylated proteins from the resin. Eluted samples were acetone precipitated, re-dissolved in denaturing buffer and subjected to reduction, alkylation, and proteolytic digestion. Digested peptides were separated by nanoLC-MS and analyzed on an LTQ-Velos by a combination of collision-induced dissociation (CID) and ETD-MS.

A large number of *O*-GlcNAcylation sites were identified on α-crystallin and sOGT (Table 3.3). First, the known *O*-GlcNAc site on α-crystallin A (Ser-162) was readily recognized despite the low abundance of the *O*-GlcNAc modification at this site.[42] Importantly, we observed both known and novel sites on sOGT.[35,43] For example, we identified the previously reported Thr-662 site, which is found in the catalytic domain of sOGT.[16] The new linker design also as a neighboring modification site at Ser-664, which to our knowledge has not been reported.[31] Furthermore, we mapped a novel site of glycosylation at Ser-580, which is found in the catalytic domain of sOGT. The new linker design also revealed a number of new *O*-GlcNAcylation sites within the *N*-terminal tetratricopeptide repeat-containing (TPR) domain (Ser-10, Thr-12, Ser-20, and Thr-23) of sOGT, and we observed a doubly modified peptide at both Thr-12 and Ser-20. As the TPR domains of OGT are thought to mediate protein-protein interactions, such modifications could play an integral role in OGT regulation and may provide a mechanism to selectively modulate its activity toward specific substrates.[5,44,45]

Table 3.3 *O*-GlcNAc sites identified on alpha-crystallin and OGT

Protein	Peptide Sequence	Site(s)	Mascot ion score	Mascot delta ion score	Method
α-crystallin A	AIPVSREEKPSSAPSS	Ser-162	24.9	23.5	ETD

sOGT	ISPTFADAYSNMGNTLK	Ser-10*/ Thr-12*	46.5	-	ETD
sOGT	ISPTFADAYSNMGNTLK	Ser-20*	21.6	13.6	ETD
sOGT	ISPTFADAYSNMGNTLK	Ser-10*/ Thr-12*/ Ser-20*	38.4	-	ETD
sOGT	EMQDVQGALQCYTR	Thr-38	41.8	35.0	ETD
sOGT	AIQINPAFADAHSNLASIHKDS GNIPEAIASYR	Ser-52*	53.5	41.0	CID
sOGT	AIQINPAFADAHSNLASIHK	Ser-56*	56.8	15.7	CID
sOGT	LYLQMWEHYAAGNKPDHMIK PVEVTESA	Thr-662	33.1	8.0	CID

Table 3.3 *O*-GlcNAc sites identified following labeling with **2**, Neutravidin affinity purification, and hydrazine mediated elution. Sites within the peptide sequence are denoted in red. Novel site identifications are marked by an asterisk.

3.5 Discussion

We describe an improved method to facilitate the comprehensive mapping of *O*-GlcNAc modification sites. Chemoenzymatic attachment of an azide-containing monosaccharide onto *O*-GlcNAc sugars provides a stoichiometric, bioorthogonal handle, which is further functionalized with an alkyne-Dde-biotin linker to isolate and enrich *O*-GlcNAcylated proteins. The cleavable Dde linker provides numerous benefits over other reported structures. First, the linker is commercially available and inexpensive. Second, it is stable to rigorous, denaturing wash conditions and can be quantitatively cleaved under mild chemical conditions. Finally, the cleaved moiety that remains on the modified peptide minimally changes the peptide mass and generates an additional positive charge, which facilitates peptide sequencing by ETD. Together, in combination with the commercially available chemoenzymatic labeling kit, the method provides an accessible and practical system for the broader community.[26] Using this approach, we identified both established and new modification sites on α-crystallin and sOGT, including potentially novel sites for OGT regulation. Our results showcase the method's capacity to comprehensively profile protein *O*-GlcNAcylation. The stoichiometric nature of the chemoenzymatic labeling, CuAAC reaction, and elution steps

provides an ideal platform for future quantitative MS analyzes to profile global *O*-GlcNAcylation and will enable the discovery of novel functional roles for *O*-GlcNAcylation in diverse biological contexts. Our lab is currently developing next generation linkers that have improved stability and differential labelling capabilities in order to obviate the dimethyl-labelling step. These next-generation linkers are currently being synthesized in the lab and will be used to identify novel *O*-GlcNAc sites as well as perform quantitative site-specific *O*-GlcNAcylation changes that accompany neuronal activation.

Furthermore, we describe the generation of a novel OGT-FLAG-HA mouse that facilitates tandem affinity purification of OGT in order to identify the OGT interactome. Importantly, the integration of the minimal tag into the mouse allows for pull down of endogenous OGT while avoiding the false positives and other potential pitfalls that could befall overexpression-based tandem affinity purification. Using our novel OGT-FH mouse, we were able to obtain the brain-specific OGT interactome, which is consistent with previous OGT interactome studies. In addition, many of the proteins within the brain-specific OGT interactome are also known to be OGT substrates. The next steps are to validate the novel OGT interactors and then to test the OGT interactor/substrate hypothesis using interactor knockdown or ideally OGT-interactor specific interaction disruption followed by our newly optimized *O*-GlcNAcome identification methods. In this way, we can directly test the OGT interactor/substrate hypothesis and identify the key “hub” interactors that mediate substrate targeting in the brain. In summary, we have developed robust OGT interactome and *O*-GlcNAcome tools that facilitate OGT

interactor and substrate identification toward the goal of understanding the key role of *O*-GlcNAc in neuronal function.

3.6 Methods

3.6.1 Reagents and materials for OGT interactome

Fetal bovine serum (FBS); ViraPower™ Lentiviral Packaging Mix, Lipofectamine™ 3000; Na pyruvate (100X, Gibco); Penicillin/Streptavidin (P/S, 100X, Gibco); Nonessential amino acids (100X, Gibco); HEPES (100X, Gibco); DMEM (Dulbecco's Modified Eagle Media) High Glucose and GlutaMAX (Gibco); FastDigest MluI; NuPAGE™ Novex™ 4-12% Bis-Tris protein gels (1.0 mm, 10-well); Triton®-X-100, Electrophoresis Grade; ElectroMax™ DH5α-E™ Competent Cells; Pierce™ BCA Protein Assay Kit; Bovine Serum Albumin, Fraction V; MEGAshortscript™ T7 Transcription Kit; MEGAclear™ Transcription Clean-up Kit; UltraPure™ DNase/RNase-Free Distilled Water; FastDigest SalI; FastDigest MluI; Pierce™ Trypsin Protease, MS Grade; DTSSP (3,3'-dithiobis(sulfosuccinimidyl propionate)); and BCA assay reagents were obtained from ThermoFisher Scientific (Waltham, MA). We purchased BD Bacto™ Yeast Extract, Tryptone, and Agar from BD Biosciences (San Jose, CA). Amicon Ultra-15 Centrifugal Filter Units with Ultracel-10 membrane (10 kDa); Amicon Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (30 kDa); phosphate buffered saline powder, pH 7.4; DL-Dithiothreitol (DTT); 3x FLAG peptide; ANTI-FLAG M2® affinity gel; α-Tubulin mouse mAb; Anti-*O*-GlcNAc transferase (DM-17) rabbit antibody; α-Tubulin mouse mAb (T9026); iodoacetamide, BioUltra; Ammonia Solution, 25%; Monoclonal ANTI-FLAG® M2 mouse antibody; Triethylammonium Bicarbonate (TEAB) Buffer, Volatile Buffer for HPLC; Sodium

cyanoborohydride ($NaBH_3CN$) were purchased from MilliporeSigma (Burlington, MA). HA-Tag (C29F4) Rabbit mAb was obtained from Cell Signaling (Danvers, MA). Q5® Site-Directed Mutagenesis Kit, Q5® Hot Start High-Fidelity 2X Master Mix, BamHI-HF®, Quick Ligation® Kit were obtained from New England BioLabs, Inc. (Ipswich, MA). GenePulser/MicroPulser Cuvettes (0.1 cm gap sterile electroporation cuvettes) and Silver Stain Plus were purchased from Bio-Rad (Hercules, CA). The pX330-U6-Chimeric_BB-CBh-hSpCas9 construct was obtained from Addgene (Cambridge, MA). DNeasy Blood & Tissue Kits were purchased from Qiagen (Hilden, Germany). The CMV6-Entry construct was obtained from Origene (Rockville, MD). Polyethylenimine, linear transfection grade (PEI, MW 25,000) was purchased from Polysciences, Inc. (Warrington, PA). Casein kinase II (CKII) peptide was synthesized and purchased from GenScript (Piscataway, NJ). cOmplete™ protease inhibitor cocktail without EDTA (PIC-EDTA) was purchased from Roche Diagnostics Corp. (Indianapolis, IN).

3.6.2 Tandem affinity purification protocol for OGT pull down and lentiviral production

The conditions for tandem affinity purification were optimized using HEK293T cells transfected with lentivirus expressing rat OGT-FLAG-HA and FLAG-HA-OGT. In order to generate lentivirus, 10 cm plates were coated with 0.1% wt/vol. gelatin A for 2 hours followed by rinsing twice with sterile water and air drying. Then, HEK293FT cells were plated in COMPLETE media (10% FBS, Na pyruvate, P/S, nonessential amino acids, and HEPES in DMEM High Glucose). At 60% confluence, the media was removed, and then the cells were plated with serum-free DMEM High Glucose with GlutaMAX transfected with ViraPower™ Lentiviral Packaging Mix (pLP1, pLP2, and

pLP/VSVG), the Tet-OGT-FH or Tet-FH-OGT and Lipofectamine™ 3000 according to the manufacturer's instructions. After 6 hours, the serum-free media was replaced with COMPLETE media and then the media was collected and then replaced every 12 hours thereafter for 72 hours total. The collected media was then filtered using a 0.45 μm sterile filter or spun 2000 x *g* for 5 min in order to remove or pellet cellular debris. Then, the media was concentrated using 50-ml pre-washed Amicon-15 30 kDa centrifugal concentrators. HEK293T cells were transduced with the lentivirus in order to create HEK293T cells that would express OGT-FH with the addition of doxycycline (Tet-OGT-FH HEK293T cells).

Tet-OGT-FH HEK293T cells cultured with or without 1 μg/ml doxycycline or OGT-FH mouse brains or WT mouse brains were then lysed in 1% Triton X-100 in TBS with PIC-EDTA followed by enrichment using ANTI-FLAG M2® M2 affinity gel agarose beads with end-over-end rotation overnight at 4°C. After this, the FLAG beads were washed three times with TBS (Tris Buffered Saline with Triton X-100). Following washing, the OGT-FH was eluted using 150 ng/μL 3xFLAG peptide in TBS (19 mM Tris, pH 7.4, 137 mM NaCl, 2.7 mM KCl) with 1% Triton X-100. Then, the eluted proteins were added to HA beads and enriched with end-over-end rotation overnight at 4°C. The beads were washed with TBS three times followed by elution with 3M NaSCN. Proteins in the final tandem affinity purification elution were separated on NuPAGE™ Novex™ 4-12% Bis-Tris protein gels in MOPS buffer (50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7). Following protein gel electrophoresis, the proteins were either then prepared for mass spectrometry (see below), silver stain, or transferred onto an Immobilon-FL PVDF membrane in transfer buffer (25 mM Tris, 190 mM

glycine, 20% methanol) for western blot. For silver stain, the gel was stained using the Silver Stain Plus kit according to the manufacturer's instructions. For western blot, the membrane was blocked for 1 h at RT with 5% bovine serum albumin (BSA) in 1x TBST (TBS with 0.1% Tween 20). The blot was then incubated for 1 h at RT or overnight at 4°C with the indicated primary antibodies (1:1,000 dilution): HA-Tag (C29F4) Rabbit mAb, ANTI-FLAG M2® mouse mAb, α-Tubulin mouse mAb, and OGT rabbit Ab. After primary incubation, the membrane was rinsed three times with 1x TBST and then incubated with the appropriate Alexa Fluor® 680 or 800 conjugated secondary antibody (1:5,000) in 5% BSA in TBST for 1 h at RT, washed three times with TBST for 5 min, and imaged with an Odyssey scanner.

For crosslinking, the cells were placed on ice, washed once with PBS/Ca^{2+}/Mg^{2+} buffer (0.1 mM $CaCl_2$, 1 mM $MgCl_2$ in PBS), then incubated for 2 hours in either 1 mM DTSSP (crosslinking solution) in PBS/Ca^{2+}/Mg^{2+} or PBS/Ca^{2+}/Mg^{2+} (non-crosslinking solution), and quenched in 20 mM Tris, pH 7.4 in PBS/Ca^{2+}/Mg^{2+} buffer for 15 min. Following crosslinking, the cells were lysed in 0.5% Triton X-100 in Buffer A (10 mM HEPES, 150 mM NaCl, 1 mM EGTA, 0.1 mM $MgCl_2$, pH 7.4) with PIC-EDTA for 30 min with rocking. Following lysis, the cells were scraped from the plate and centrifuged at 15,000 x *g* for 15 min. All crosslinking and lysis conditions were performed on ice or at 4°C with ice-cold buffers. Then, the supernatant was analyzed for protein concentration using the BCA assay according to the manufacturer's instructions. Afterwards, the tandem affinity purification proceeded with 10 mg of each sample in parallel as described above.

3.6.3 Mass spectrometry for OGT interactome

Following tandem affinity purification, the final HA pull down eluent was added to two prewetted 10K MWCO Amicon filter units and washed 3 times in TEAB buffer (100 mM triethylammonium bicarbonate). Then the concentrate was spun down to 50 μL of liquid and then 10 mM DTT in 100 mM TEAB was incubated with gentle rotation for 60 min at 37°C. After reduction, the solution was washed twice with 100 mM TEAB, concentrated to 50 μL, and then alkylated with 50 mM 2-iodoacetamide (IAA) in 100 mM TEAB in the dark with rotation for 40 min at RT. The solution was washed twice with 100 mM TEAB, concentrated to 20 μL, and then digested overnight at 37°C with trypsin in 1 mM $CaCl_2$, 100 mM TEAB with a final protease to protein ratio of 1:20 (wt/wt). The peptides were washed twice with 100 mM TEAB and then combined for a final volume of 100 μL. Then, 4 μL of either 4% (vol/vol) CH_2O or CD_2O (for light or intermediate labelling respectively) was added to the samples, mixed, and then spun down as previously described.[46] Afterwards, 4 μL of 0.6M $NaBH_3CN$ was added to the samples and then rotated for 1 hour at RT. The solution was quenched by adding 16 μL of 1% (vol/vol) ammonia solution and then acidified with 8 μL formic acid. The sample was desalted using HPLC and then resuspended in 10 μL of 0.1% formic acid. At this point, the light- and intermediate-labelled samples were mixed and analyzed by LC-MS/MS at the Caltech Proteomics Exploration Laboratory. About 250 ng of digested peptides were analyzed with an EASY-nLC 1000 and Thermo Orbitrap Fusion™ Tribrid™ Mass Spectrometer (ThermoFisher Scientific) as previously described.[47]

3.6.4 Activity assay to check activity of OGT tags

N-terminal tagged (FLAG-HA-OGT) and C-terminal tagged (OGT-FLAG-HA) were cloned into the CMV6-Entry construct using Q5® Site-Directed Mutagenesis Kit

according to the manufacturer's instructions (New England BioLabs). HEK293T cells were cultured on a 10 cm plate in DMEM, high glucose, GlutaMAX, with 10% FBS and were transfected with these two constructs with a 3:1 PEI:DNA ratio (wt:wt). Twenty-four hours following transfection, the media was removed from cells prior to lysis and scraping in ice-cold lysis buffer (50 mM Tris/HCl pH 7.4, 250 mM mannitol, 50 mM NaF, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM Na_2VO_4, 1 % Triton X-100, and 1X protease inhibitors). The cell lysate was dounced 10 times with a 2 ml douncer on ice followed by centrifugation at 20,000 x *g* for 10 min at 4°C. Next, 500 µL of soluble lysate was added to 40 µL of ANTI-FLAG M2® M2 affinity gel agarose beads, washed 1X with 0.2% triton in TBS, washed 3X in TBS, and then resuspended in 80 µL of 1X OGT buffer (250 mM Tris, 125 mM $MgCl_2$, 0.6 mg/ml BSA, 10 mM DTT) and then divided evenly into four wells. Then, we added a mixture of 1 mM UDP-GlcNAc with or without 100 µM CKII peptide (KKKYPGGSTPVSSANMM) for 1 hour at RT with gentle mixing. After 1 hour, we added the UDP-Glo™ Solution in order to convert all free UDP to ATP followed by a luciferase/luciferin reaction for visualization. All conditions were performed in duplicate or triplicate per plate and the C-terminal and N-terminal tagged OGT luminescence with peptide was normalized to the luminescence lacking peptide. Following completion of the UDP-Glo™ Glycosyltransferase assay, the beads were collected from the wells and then boiled in 1X loading buffer. The boiled elution was then resolved using protein gel electrophoresis and then western blotting as described above. The membrane was blotted using anti-HA and anti-FLAG in order to quantify the levels of OGT that were used in the assay. After quantification, the normalized

luminescence levels were normalized to the average blot intensities of HA and FLAG tagged OGT as determined by Western blotting.

3.6.5 Design and screen of CRISPR/Cas9 sgRNA

We used the Dr. Feng Zhang's lab's sgRNA design website to generate potential guide RNAs (http://crispr.mit.edu/guides). The top five sgRNA sequences (with scores) were (1) 5'-CCTGAATAAAGACTGCGCAC-3' (87%); (2) 5'-GCTGACTCGGTGACTTCAAC-3' (84%); (3) 5'-CCTGTGCGCAGTCTTTATTC-3' (77%); (4) 5'-CTTCAACAGGCTTAATCAT-3' (75%); and (5) 5'-ACAGGCTTAATCATGTGGTC-3' (74%). These sgRNA candidates were cloned into the pX330 plasmid (U6-sgRNA-Chimeric_BB-CBh-hSpCas9), which was a gift from Feng Zhang (Addgene plasmid #42230).

We cloned the genomic region neighboring the C-terminal end of mouse OGT into the pCAG-EGxxP plasmid in order to generate the pCAG-EG_ogt_FP plasmid (Figure 3.18). Briefly, we extracted gDNA from mouse tails using a DNeasy kit and amplified the C-terminal region of OGT with Q5® Hot Start High-Fidelity DNA polymerase using the following primers: OGT_F: 5'-CAACCACTGAGGATCCTGCCCACAGACAAGTTTGA-3' and OGT_R: 5'-TATCGAATTCGTCGACTCAATCCACGGCATCACAA-3' [98°C for 30 sec; [98°C for 10 sec; 62°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] to produce the following amplicon: chrX:101681950+101682750 (https://tinyurl.com/yajf2aqt). Then, we double digested the amplicon and the pCAG-EGxxP using SalI and BamHI restriction enzymes for 1 hour at 37°C followed by PCR clean-up. The Quick Ligation® Kit was used to ligate the cleaned up amplicon and cut

pCAG-EGxxP plasmid as per the manufacturer's instructions. The newly ligated plasmid was transformed into DH5α electrocompetent cells using the Bio-Rad MicroPulser™ Electroporation Apparatus at the 18 kV/cm (Ec1) setting using 0.1 cm gap electroporation cuvettes. The transformed E. coli was grown overnight at 37°C in 1 mg/mL ampicillin in sterile lysogeny broth (LB) (12 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with shaking and then streaked on 100 µL/mL ampicillin sterile agar (10 g/L agar in LB) plates. The final plasmid, pCAG-EG_ogt_FP is shown in Figure 3.18B. Using HEK293T cells grown in DMEM, high glucose, GlutaMAX, with 10% FBS, we screened the top 5 guide RNA candidates for cleavage and homologous repair efficiency by co-transfecting the pCAG-EG_ogt_FP construct the pX330-sgRNA with Lipofectamine™ 3000 according to the manufacturer's instructions. Fluorescent images were taken using the Zeiss 700 LSM confocal microscope.

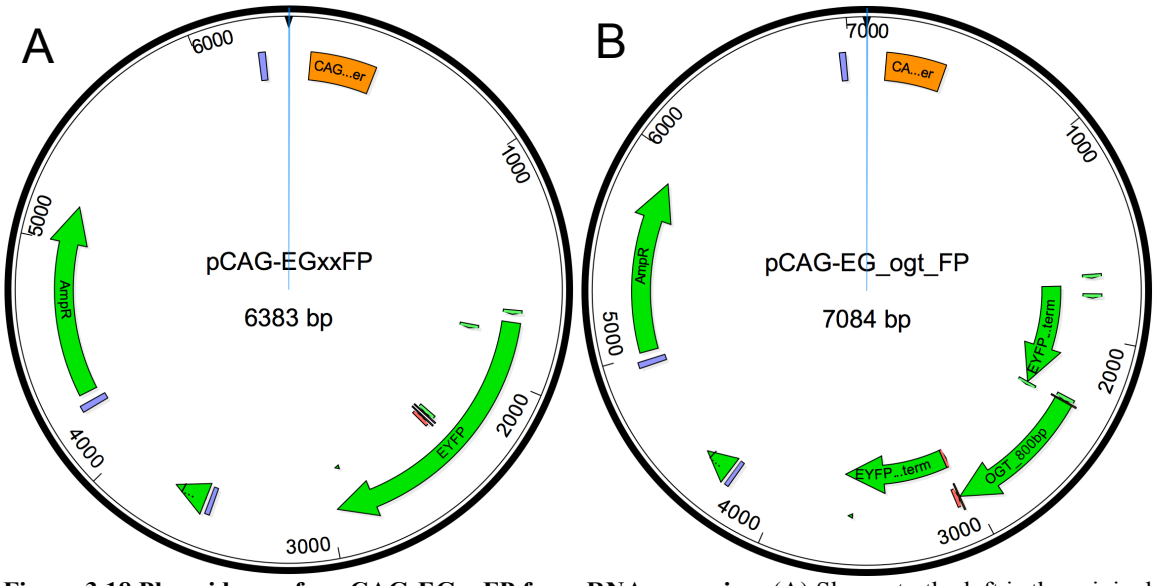


Figure 3.18 Plasmid map for pCAG-EGxxFP for sgRNA screening. (A) Shown to the left is the original pCAG-EGxxFP plasmid. (B) Shown to the right is the pCAG-EG_ogt_FP plasmid with the C-terminus of OGT inserted into the plasmid.

3.6.6 Generation of OGT-FLAG-HA mice using CRISPR/Cas9 and genotyping

We synthesized the sgRNA by performing PCR with Q5® Hot Start High-

Fidelity DNA polymerase on the pX330-sgRNA1 construct (10 ng) using the following primers T7-sgRNA_F: 5'-ttaatacgactcactataggCCTGAATAAAGACTGCGCAC-3'; and T7-sgRNA-R: 5'-AAAAGCACCGACTCGGTGCC-3' [98°C for 30 sec; [98°C for 10 sec; 60°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min]. We then gel purified the T7-sgRNA PCR product using the QIAQuick PCR purification kit and *in vitro* transcribed the T7-sgRNA PCR product with the MEGAshortscript™ T7 Transcription Kit followed by MEGAclear™ Transcription Clean-up Kit according to the manufacturer's instructions and Yang *et al.* protocol.[9] Some of the sgRNA was diluted to 500 ng/μL in RNase-free water and tested for quality using a 2% (wt/vol.) agarose gel in TAE buffer. OGT-FLAG-HA mouse was generated using C57/BL6 wildtype zygotic pronuclear and cytoplasmic injections as previously described.[9,48,49] For pronuclear injections, 2.5 ng/μL sgRNA, 10 ng/μL ssODN, and 5 ng/μL Cas9 mRNA in TE buffer (10 mM Tris-HCl, 0.1 mM Na_2EDTA, pH 8.0) [final concentrations: 0.36 ng/μL sgRNA, 1.43 ng/μL ssODN, and 0.71 ng/μL Cas9 mRNA] were injected into the nucleus of mouse zygotes. For the cytoplasmic injections, 50 ng/μL sgRNA, 100 ng/μL ssODN, and 100 ng/μL Cas9 mRNA in TE buffer [final concentrations per embryo: 7.14 ng/μL sgRNA, 14.3 ng/μL ssODN, and 14.3 ng/μL Cas9 mRNA] were injected into the cytoplasm of mouse zygotes. Following injection, the zygotes were grown to the two-cell stage and implanted into pseudopregnant foster mothers (up to 30 two-cell zygotes).[9] Approximately 19.5 days after implantation, the pups were delivered, and 3 weeks after birth the pups were tailed and separated by gender.

Genotyping of the OGT-FLAG-HA mice originally involved direct sequencing and DNA agarose gel genotyping. The DNA samples were isolated from tail tips using

the DNeasy Blood & Tissue Kit. Then, PCR amplification was performed on 50 ng of template genomic DNA using the Q5® Hot Start High-Fidelity DNA polymerase and OGT-FH_geno_F1/OGT-FH_geno_R1 primer pair [98°C for 30 sec; [98°C for 10 sec; 65°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] producing the following amplicon: chrX:101681261+101682926 (https://tinyurl.com/ya2zt875). After amplification, the amplicon was either submitted for sequencing at Laragen, Inc. using the sequencing primer OGT-FH_Seq_R or incubated with BamHI-HF® enzyme for two hours at 37°C. After amplicon cleavage, the DNA was separated using agarose electrophoresis (1.5% agarose gel in TAE buffer) to distinguish wildtype (WT, 1666 bp band), heterozygotes (HET, 1068 bp + 658 bp bands), and homozygotes (HOM, 1666 bp + 1068 bp + 658 bp bands). The primers used were as follows: OGT-FH_geno_F1: 5'-CCATCTCACCAGCCCAATAC-3', OGT-FH_geno_R1: 5'-ACTGACAGTGCCAAGCATTA-3'; and OGT-FH_Seq_R: 5'-ACTGATATAGGCTCATGTGGTTT-3'.

For off-target sequence validation, we performed PCR on 50 ng of the DNeasy-derived gDNA using the Q5® Hot Start High-Fidelity DNA polymerase (New England BioLabs, Inc.) and the off_F/off_R primer pair [98°C for 30 sec; [98°C for 10 sec; 67°C for 20 sec; 72°C for 30 sec (34 cycles)], and then 72°C for 2 min] producing the following amplicon near the *Rhox11* gene: chrX:38076598+38085139 (https://tinyurl.com/ybpvwll2). Then, the amplicon and the off_F primer were submitted for sequencing at Laragen, Inc. The primers used were as follows: off_F: 5'-CTTCGTGGGTTTAAGGCCGA-3' and off_R: 5'-CTCACACAGGTTTGTGAGTTGAAG-3'.

3.6.7 Reagents and materials for *O*-GlcNAcome

All chemicals and reagents were of analytical grade, obtained from Millipore Sigma, and used without further purification unless specified. The NuPAGE™ Novex™ 4-12% Bis-Tris protein gels (1.0 mm, 10-well), Imperial™ protein stain, *O*-GlcNAcylated peptide TAPT(gS)TIAPG, high capacity Neutravidin agarose resin, Alexa Fluor® 680 conjugated streptavidin, Alexa Fluor® 680 Goat Anti-Rabbit IgG (H+L), Alexa Fluor® 790 Goat Anti-Rabbit IgG (H+L), Alexa Fluor® 680 Goat Anti-Mouse IgG (H+L), Alexa Fluor® 790 Goat Anti-Mouse IgG (H+L), Accela LC, PAL autosampler, highly cross-adsorbed, spin columns, and C18 desalting tips were purchased from ThermoFisher Scientific (Waltham, MA). All protein concentrations were measured using the BCA assay (ThermoFisher Scientific). Immobilon-FL PVDF membrane (EMD Millipore, 0.45 μM), gelatin from porcine skin (Type A, lyophilized powder, γ-irradiated, BioXtra, suitable for cell culture), α-crystallin from bovine eye lens were obtained from Millipore Sigma (St. Louis, MO). The CORTECS UPLC C18+ column (2.1 x 50 mm) was received from Waters Corp. (Milford, MA). Thiamet G was obtained from Tocris Biosciences (Avonmouth, Bristol, UK). cOmplete™ protease inhibitor cocktail without EDTA (PIC-EDTA) was purchased from Roche Diagnostics Corp. (Indianapolis, IN). Baculovirus preparation and protein expression of short-form OGT (sOGT) in *Spodoptera frugiperda* (Sf9) cells was performed as previously described.[50] Cerebral cortices were obtained from adult C57BL/6 mice bred in the Caltech Animal Facility according to NIH guidelines. RapiGest and UDP-GalNAz **1** were synthesized as referenced.[51,52] Y289L GalT was expressed and purified as described previously.[53] Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), alkyne-Dde-biotin **2**, and the

photocleavable alkyne-biotin (alkyne-PC-biotin) were purchased from Click Chemistry Tools (Scottsdale, AZ). The Odyssey scanner was purchased from LI-COR Biosciences.

3.6.8 *O*-GlcNAcylated peptide labeling

The labeling protocol was adapted from a previously reported method.[54] The peptide TAPT(gS)TIAPG (20 μM final) was dissolved in a 200 μL solution of 10 mM HEPES pH 7.9, 5.5 mM $MnCl_2$, 1 mM UDP-GalNAz **1**, and 100 ng/μL Y289L GalT and rotated end-over-end overnight at 4°C. Prior to enzyme addition, an aliquot was removed as an initial time point for LC-MS analysis. The reaction was acidified to 0.1% TFA, desalted using a C18 tip, and an aliquot was saved for analysis. The labelled peptide (10 μM final) was diluted into a 400 μL solution of 10 mM sodium phosphate pH 7.6, 100 μM alkyne-Dde-biotin **2**, 2 mM sodium ascorbate, and 100 μM THPTA. $CuSO_4$ was added (1 mM final), and the reaction was incubated while rotating end-over-end at RT for 1 h. After removing a sample, the reaction was acidified and desalted again. The peptide (10 μM) was then split into fractions of 50 μL containing 25 mM sodium phosphate pH 7.6 and either 1% RapiGest, 6 M urea, or 2% hydrazine monohydrate and incubated for 1 h at RT. Samples were acidified, desalted, and subjected to LC-MS analysis.

3.6.9 LC-MS analysis of *O*-GlcNAc peptide labeling

Liquid chromatography and mass spectrometry (LC-MS) were performed using an LTQ linear ion trap mass spectrometer combined with an Accela LC and PAL autosampler. Approximately 10 pmol peptide from each sample was injected onto a CORTECS UPLC C18+ column (2.1 x 50 mm). Flow rate was set at 0.4 mL/min. Solvent A (ddH_2O, 1% formic acid) and Solvent B (acetonitrile, 1% formic acid) were used to create a gradient. The gradient consisted of 0-0.2 min, 5% B; 0.2-3.5 min 5-65% B, 3.5-

4.0 min 65% B with injection into the MS starting at 0.2 min to avoid salt contamination. All peptide products were found to elute during the linear gradient between 0.2 and 2.0 min. For the biotinylated and cleaved peptide, alkyne reagent **2** was not sufficiently removed by the C18 tips. Therefore, the reaction was monitored using an extracted ion chromatogram by extracting all ions within ± 1 *m/z* of the calculated masses.

3.6.10 Chemoenzymatic labeling using Dde and photocleavable linkers

Labeling with Y289L GalT and UDP-GalNAz **1** was conducted as previously described.[26] Briefly, 500 μg HEK-293T cell lysate in 1% SDS, 1x PBS pH 7.4 (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl), 10 μM Thiamet-G, and 1x PIC-EDTA was diluted to a protein concentration of 1 mg/mL using 1% SDS, 1x PBS pH 7.4. For the α-crystallin and sOGT labeling, 200 μg cortical lysate was spiked with 20 μg of α-crystallin and 5 μg of sOGT. Cortical lysate was obtained from adult 2 month old mice and dounced 10 times in 2% SDS, 1x PBS, 10 μM Thiamet-G, and 1x PIC-EDTA and then diluted to a protein concentration of 1 mg/mL using 1% SDS, 1x PBS pH 7.4, 10 μM Thiamet-G, and 1x PIC-EDTA. Protein was precipitated by adding three volumes of methanol, one volume of chloroform, and two volumes of ddH_2O with vortex mixing after each addition and then pelleted at the aqueous-organic interface by centrifuging at 21,000 x *g* for 5 min. The top, aqueous layer was removed, and one volume of methanol was added with mixing. The protein was pelleted again, and all liquid was removed. After the pellets were air-dried, samples were redissolved at 5 mg/mL (100 μL) in 1% SDS, 20 mM HEPES pH 7.9 by sonication. To each sample, the following were added in the given order: 10 μL of 50x PIC-EDTA, 112.5 μL of ddH_2O, 200 μL of 2.5x labeling buffer (50 mM HEPES pH 7.9, 125 mM NaCl, 5% NP-40), and 27.5 μL of 100 mM $MnCl_2$.

Samples were briefly mixed by pipetting and placed on ice. Next, 25 µL of 0.5 mM UDP-GalNAz **1** was added followed by pipetting to mix. Finally, 25 µL of 2 mg/mL Y289L GalT was added, and samples were rotated end-over-end for 16 h at 4 °C. Proteins were then precipitated again, and pellets were air dried. Pellets were next dissolved at 4 mg/mL (125 µL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of the GalT-labeled sample was removed (25 µL), and the remaining sample was split in two (50 µL each) to be labeled with either alkyne-Dde-biotin **2** or alkyne-PC-biotin. All manipulations with the photocleavable linker were performed in the dark.

To each sample, the following were added in the given order with mixing: 4 µL of 50x PIC-EDTA, 78 µL of ddH_2O, and 10 µL of 20x PBS pH 7.4. Next, the CuAAC reagents were added with vortex mixing after each addition: 4 µL of 5 mM alkyne-dde-biotin **2** or alkyne-PC-biotin (stock in DMSO), 4 µL of 100 mM sodium ascorbate (freshly prepared), 10 µL of 2 mM THPTA (stock in 4:1 *t*BuOH/DMSO), and 4 µL of 50 mM $CuSO_4$ (freshly prepared). Samples were rotated end-over-end for 1 h at RT, and the reaction was halted by the addition of 25 µL EDTA pH 8.0. Samples were precipitated, and the pellet was washed once with 1 mL MeOH to remove residual, unreacted linker. The pellet was then air-dried and redissolved at 4 mg/mL (50 µL) in 1% SDS, 20 mM HEPES pH 7.9. An aliquot of each sample was reserved (25 µL), and the remaining sample was cleaved. For the alkyne-Dde-biotin sample, the mixture was diluted to 1 mg/mL with 2% hydrazine monohydrate, and the sample was rotated end-over-end for 1 h at RT. For the alkyne-PC-biotin sample, the protein was diluted to 1 mg/mL with ddH_2O, and the liquid was irradiated from the open top of the tube (2 cm distance) with 365 nm UV light (UVGL-25 handheld UV lamp, 1.5 mW/cm^2) for 1 h at RT with mixing

every 10 min. Both samples were then precipitated by addition of four volumes of -20 °C acetone and storage at -20 °C for 1 h. Samples were air-dried and then redissolved at 4 mg/mL (25 µL) in 1% SDS, 20 mM HEPES, pH 7.9.

3.6.11 Coomassie staining and western blotting

Aliquots corresponding to 20 µg protein (5 µL) for each sample was resolved by SDS-PAGE as follows. Samples were added to 10 µL ddH_2O and 5 µL 4x SDS-PAGE loading buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol) and were then used directly without boiling to avoid cleaving the linkers. The mixtures were loaded in duplicate on NuPAGE™ Novex™ 4-12% Bis-Tris protein gels and separated using protein gel electrophoresis in MOPS buffer. One duplicate was then stained with Imperial™ protein stain according to the manufacturer's specifications and imaged by an Odyssey scanner. The other set of samples was transferred onto an Immobilon-FL PVDF membrane in transfer buffer and blocked for 1 h at RT with 5% bovine serum albumin (BSA) in 1x TBST. The blot was then incubated with 1:20,000 Alexa Fluor® 680 conjugated streptavidin in 5% BSA/TBST for 1 h at RT, washed three times with TBST for 5 min, and imaged with an Odyssey scanner.

3.6.12 Enrichment and elution of labeled proteins

Labeled samples were diluted to 1 mL using 1x PBS pH 7.4, 1x PIC-EDTA. For each sample, 25 µL (settled volume) of high-capacity Neutravidin agarose was washed twice with 500 µL of 1x PBS in spin columns, and samples were added to the washed beads. Mixtures were rotated end-over-end for 1 h at RT. Lysate was removed by centrifugation at 2,000 x *g* for 30 s. The beads were washed with 1% SDS (5 x 0.5 mL), 6 M urea (5 x 0.5 mL), and 1x PBS pH 7.4 (5 x 0.5 mL). Beads were then resuspended in

50 µL of 2% hydrazine monohydrate in ddH_2O and rotated end-over-end for 1 h at RT. The elution volume was removed, and beads were washed with 50 µL of 1x PBS pH 7.4. The wash volume was combined with the elution volume, and samples were flash-frozen and stored at -80 °C prior to analysis.

3.6.13 *O*-GlcNAcome sample processing for MS analysis

Samples were thawed and precipitated by addition of four volumes of -20 °C acetone. Samples were stored at -20 °C for 1 h and centrifuged at 21,000 x *g* for 5 min. Pellets were redissolved in 20 µL of 8 M urea, 100 mM Tris pH 8.0, 10 mM DTT and incubated at 60 °C with shaking for 20 min. Cysteine residues were blocked by addition of 25 mM iodoacetamide for 45 min. Samples were diluted four-fold with 100 mM Tris pH 8.0. Samples were split in two and digested with 0.01 mg/mL trypsin or chymotrypsin for 4-16 h at 37 °C. A portion of the trypsin digests was further digested with 7 µg/mL AspN for 6 h at 37 °C. Digests were acidified to a final concentration of 0.5% formic acid, 0.05% TFA.

3.6.14 LC separation and MS analysis

The digests were analyzed by nanoLC/MS on the LTQ-Velos with a 0 to 30% B in 120 min gradient with top 5 MS/MS (A: ddH_2O, 0.1% formic acid; B: acetonitrile, 0.1% formic acid). Samples were desalted on a 360 x 100 µm Kasil fritted pre-column (2 cm Monitor C18) prior to separation on a 360 x 75 µm (10 cm BEH130 C18, 1.7 µm) analytical column/tip. Full scan MS was acquired at 60,000 resolution followed by top 5 tandem MS in the linear ion trap alternating between ETD and CID modes of the same precursor. The ETD reaction time was 100 ms with supplemental activation. RAW files were converted to MGF files for Mascot searching using Proteome Discoverer with CID

and ETD spectra extracted to separate MFG files. Data was searched against a custom database with fixed modifications of carbamidomethyl (C) and variable mods of oxidation (M) and a custom modification for the tagged *O*-GlcNAc. The custom modification was defined as addition of C(19) H(30) N(6) O(10) to Ser or Thr (net addition of 502.202341 Da). For CID, a scoring neutral loss of C(19) H(30) N(6) O(10) was included, but this was omitted for ETD. Enzyme specificity was trypsin (KR), chymotrypsin (FLYW), or trypsin-AspN_ND (cleave C-term KR and N-term ND). Mass tolerances were 25 ppm and 0.8 Da for precursor and fragments ions, respectively. The instrument type was chosen as either ESI-TRAP or ETD-TRAP. Search results were combined in Scaffold 4.4 Proteome Software, filtered for 80% peptide confidence and modifications manually evaluated.

3.7 References

1 Hart, G.W., 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**, 1:825-858.

2 Pathak, S. *et al.* The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat Struct Mol Biol* (2015) **22**, 9:744-750.

3 Yang, X., Zhang, F. & Kudlow, J.E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* (2002) **110**, 1:69-80.

4 Cheung, W.D., Sakabe, K., Housley, M.P., Dias, W.B. & Hart, G.W. O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J Biol Chem* (2008) **283**, 49:33935-33941.

5 Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* (2013) **493**, 7433:561-564.

6 Rafie, K. *et al.* Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats. *Open Biol* (2017) **7**, 6.

7 Dalvai, M. *et al.* A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* (2015) **13**, 3:621-633.

8 Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* (2013) **31**, 9:827-832.

9 Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* (2014) **9**, 8:1956-1968.

10 Leney, A.C., El Atmioui, D., Wu, W., Ovaa, H. & Heck, A.J.R. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proc Natl Acad Sci U S A* (2017) **114**, 35:E7255-E7261.

11 Ruan, H.B., Nie, Y. & Yang, X. Regulation of protein degradation by O-GlcNAcylation: crosstalk with ubiquitination. *Mol Cell Proteomics* (2013) **12**, 12:3489-3497.

12 Wang, J., Torii, M., Liu, H., Hart, G.W. & Hu, Z.Z. dbOGAP - an integrated bioinformatics resource for protein O-GlcNAcylation. *BMC Bioinformatics* (2011) **12**, 1:91.
13 Nandi, A. *et al.* Global identification of O-GlcNAc-modified proteins. *Anal Chem* (2006) **78**, 2:452-458.
14 Clark, P.M. *et al.* Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J Am Chem Soc* (2008) **130**, 35:11576-11577.
15 Skorobogatko, Y.V. *et al.* Human Alzheimer's disease synaptic O-GlcNAc site mapping and iTRAQ expression proteomics with ion trap mass spectrometry. *Amino Acids* (2011) **40**, 3:765-779.
16 Alfaro, J.F. *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) **109**, 19:7280-7285.
17 Iyer, S.P. & Hart, G.W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J Biol Chem* (2003) **278**, 27:24608-24616.
18 Pekkurnaz, G., Trinidad, J.C., Wang, X., Kong, D. & Schwarz, T.L. Glucose regulates mitochondrial motility via Milton modification by O-GlcNAc transferase. *Cell* (2014) **158**, 1:54-68.
19 Ohn, T., Kedersha, N., Hickman, T., Tisdale, S. & Anderson, P. A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. *Nat Cell Biol* (2008) **10**, 10:1224-1231.
20 Zeidan, Q., Wang, Z., De Maio, A. & Hart, G.W. O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins. *Mol Biol Cell* (2010) **21**, 12:1922-1936.
21 Deng, R.P. *et al.* Global identification of O-GlcNAc transferase (OGT) interactors by a human proteome microarray and the construction of an OGT interactome. *Proteomics* (2014) **14**, 9:1020-1030.
22 Ruan, H.B. *et al.* O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. *Cell Metab* (2012) **16**, 2:226-237.
23 Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* (2015) **12**, 6:561-567.
24 Ma, J. *et al.* O-GlcNAcomic Profiling Identifies Widespread O-Linked beta-N-Acetylglucosamine Modification (O-GlcNAcylation) in Oxidative Phosphorylation System Regulating Cardiac Mitochondrial Function. *J Biol Chem* (2015) **290**, 49:29141-29153.
25 Khidekel, N. *et al.* A Chemoenzymatic Approach toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. *J. Am. Chem. Soc.* (2003) **125**, 52:16162-16163.
26 Clark, P.M. *et al.* Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J. Am. Chem. Soc.* (2008) **130**, 35:11576-11577.
27 Ramakrishnan, B. & Qasba, P.K. Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. *J Biol Chem* (2002) **277**, 23:20833-20839.
28 McKay, C.S. & Finn, M.G. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol* (2014) **21**, 9:1075-1101.
29 Szychowski, J. *et al.* Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc* (2010) **132**, 51:18351-18360.
30 Wang, Z. *et al.* Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell Proteomics* (2010) **9**, 1:153-160.
31 Alfaro, J.F. *et al.* Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. USA* (2012) **109**, 19:7280-7285.
32 Woo, C.M., Iavarone, A.T., Spiciarich, D.R., Palaniappan, K.K. & Bertozzi, C.R. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* (2015) **12**, 6:561-569.
33 Zaro, B.W., Yang, Y.-Y., Hang, H.C. & Pratt, M.R. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc. Natl. Acad. Sci. U.S.A.* (2011) **108**, 20:8146-8151.
34 Myers, S.A., Daou, S., Affar, E.B. & Burlingame, A.L. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments—a study of the O-GlcNAcylated protein …. *Proteomics* (2013).

35 Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature Chemical Biology* (2007) **3**, 6:339-348.
36 Bycroft, B.W., Chan, W.C., Chhabra, S.R. & Hone, N.D. A Novel Lysine-Protecting Procedure for Continuous-Flow Solid-Phase Synthesis of Branched Peptides. *J. Chem. Soc. Chem. Comm.* (1993) 9:778-779.
37 Chhabra, S.R., Parekh, H., Khan, A.N., Bycroft, B.W. & Kellam, B. A Dde-based carboxy linker for solid-phase synthesis. *Tet. Lett.* (2001) **42**, 2189-2192.
38 Yang, Y. & Verhelst, S.H.L. Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem. Comm.* (2013) **49**, 47:5366-5368.
39 Chalkley, R.J. & Burlingame, A.L. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *Journal of the American Society for Mass Spectrometry* (2001) **12**, 10:1106-1113.
40 Hanover, J.A. *et al.* Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* (2003) **409**, 2:287-297.
41 Lazarus, B.D., Love, D.C. & Hanover, J.A. Recombinant O-GlcNAc transferase isoforms: identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology* (2006) **16**, 5:415-421.
42 Roquemore, E.P. *et al.* Vertebrate lens alpha-crystallins are modified by O-linked N-acetylglucosamine. *J Biol Chem* (1992) **267**, 1:555-563.
43 Tai, H.C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel identification of O-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* (2004) **126**, 34:10500-10501.
44 Zeytuni, N. & Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* (2012) **20**, 3:397-405.
45 Iyer, S.P. & Hart, G.W. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. *J. Biol. Chem.* (2003) **278**, 27:24608-24616.
46 Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J.R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protocols* (2009) **4**, 4:484-494.
47 Sung, M.-K., Reitsma, J.M., Sweredoski, M.J., Hess, S. & Deshaies, R.J. Ribosomal proteins produced in excess are degraded by the ubiquitin–proteasome system. *Mol Biol Cell* (2016) **27**, 17:2642-2652.
48 Behrenger, R., Gertsenstein, M., Nagy, K.V. & Nagy, A. *Manipulating the Mouse Embryo: A Laboratory Manual*. Fourth edn, 5, 6, and 7 (Cold Spring Harbor Laboratory Press, 2014).
49 Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* (2013) **154**, 6:1370-1379.
50 Tai, H.-C., Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. Parallel Identification of *O*-GlcNAc-Modified Proteins from Cell Lysates. *J. Am. Chem. Soc.* (2004) **126**, 10500-10501.
51 Lee, P.J.J. & Compton, B.J. Destructible surfactants and uses thereof. (2007) US patent 7,229,539.
52 Hang, H.C., Yu, C., Pratt, M.R. & Bertozzi, C.R. Probing glycosyltransferase activities with the Staudinger ligation. *J. Am. Chem. Soc.* (2004) **126**, 6-7.
53 Clark, P.M., Rexach, J.E. & Hsieh-Wilson, L.C. in *Curr Prot Chem Biol* (John Wiley & Sons, Inc., 2009).
54 Khidekel, N. *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J. Am. Chem. Soc.* (2003) **125**, 52:16162-16163.