Chapter 5

Chemically Cleavable Tagging Method for the Enrichment and Detection of *O*-GlcNAc Glycosylated Proteins and Peptides

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.

5.1 Abstract

Post-translational modification by *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) plays many important roles in controlling protein function. However, 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 of *O*-GlcNAcylated proteins from cell lysates. The linker also allows subsequent quantitative 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 OGT. Notably, these novel sites on OGT lie in key functional domains of the protein, underscoring how this site identification method may reveal important biological insights into protein activity and regulation. We then describe the further optimization of this method to identify sites of *O*-GlcNAc modification on native proteins from complex lysate. Together, these experiments lay the groundwork for a straightforward method to label and enrich *O*-GlcNAcylated proteins and peptides for MS identification.

5.2 General Approach

Understanding the roles of *O*-GlcNAcylation in specific physiological contexts will require a more comprehensive characterization of the *O*-GlcNAc proteome and the modification sites on proteins. Notably, although thousands of proteins have been putatively shown to be *O*-GlcNAcylated,[1-7] relatively few glycosylation sites have been mapped. Mass spectrometric (MS) identification of *O*-GlcNAcylated peptides from complex mixtures has been challenging due to the substoichiometric nature of *O*-GlcNAcylation and is further exacerbated by suppression of *O*-GlcNAc peptide ionization in the presence of the unmodified peptide.[8] Thus, improved methods to enrich *O*-GlcNAcylated peptides or proteins are much needed, particularly approaches that can be directly used in conjunction with MS/MS sequencing to achieve a more comprehensive understanding of *O*-GlcNAc modification sites.

Robust enrichment of *O*-GlcNAcylated proteins can be accomplished using a two-step chemoenzymatic approach (Figure 5-1).[3, 9] First, the *O*-GlcNAc moiety is tagged with a non-natural azide group by treatment of cell lysates with UDP-GalNAz and a mutant galactosyltransferase (Y289L GalT)[10] that specifically recognizes terminal GlcNAc moieties. Next, a biotin group is attached via copper(I)-catalyzed azide–alkyne cycloaddition (CuAAC),[11] 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. In particular, harsh conditions are usually required to

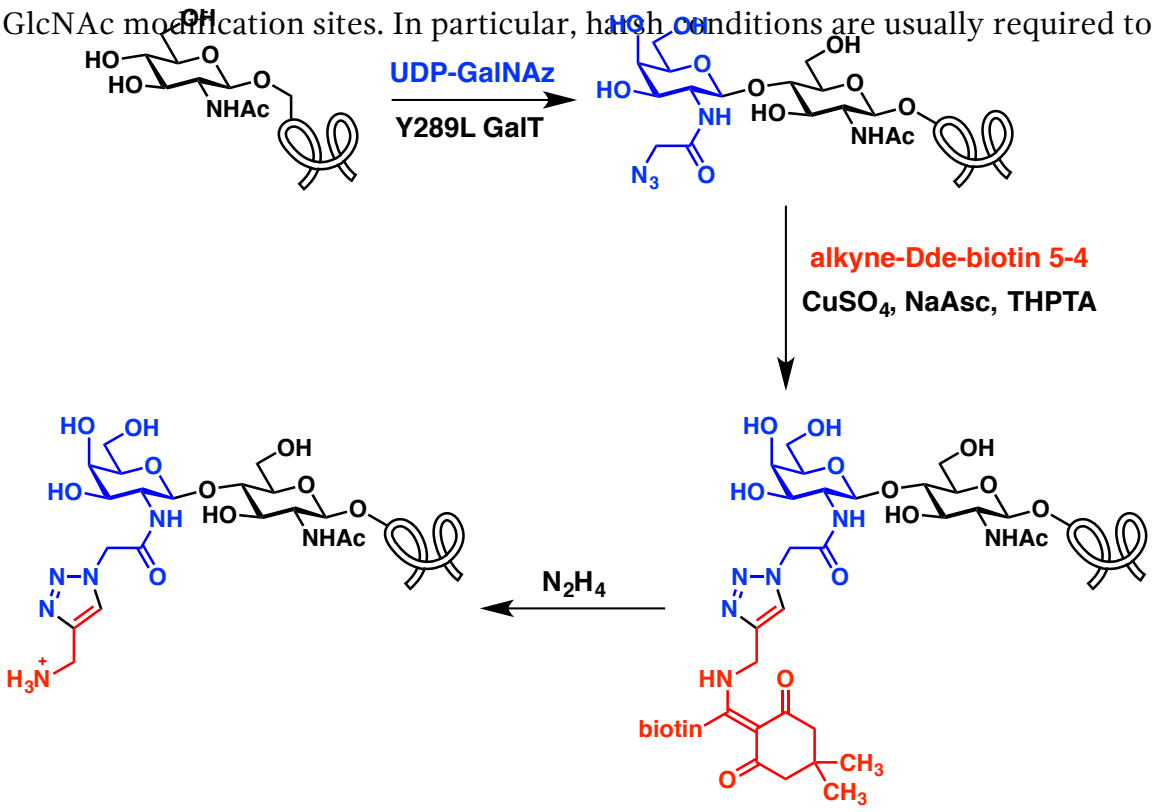


Figure 5-1. Two-step chemoenzymatic approach. *O*-GlcNAcylated substrates are modified by Y289L GalT with UDP-GalNAz to install an azide handle, which can be functionalized with a cleavable biotin linker using CuAAC. After enrichment, labeled *O*-GlcNAcylated substrates can be released by facile chemical cleavage with hydrazine.

disrupt the femtomolar biotin–streptavidin interaction,[12] which may hydrolyze the labile *O*-GlcNAc moiety. Additionally, many linkers contain a large spacer between the biotin group and the alkyne functionality,[12] which appends a relatively large mass to the glycopeptide and can preclude its identification. 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.[6-8, 13] However, each suffers from significant drawbacks for site identification. For example, a photocleavable linker (alkyne-PC-biotin **5-1**, Figure 5-2) was employed in conjunction with UDP-GalNAz and Y289L GalT to sequence modified peptides from mouse brain lysate.[7, 8] 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.[4, 14] Unfortunately, cleavage of the linker was found to be incomplete.[8] In a recent report, a dibromine-containing, acid-cleavable linker (**5-2**, Figure 5-2) was employed to identify various glycan modifications including *O*-GlcNAc.[13] However, cleavage of the linker revealed only a neutral hydroxyl group, and the halogenated glycopeptides demonstrated poor fragmentation efficiency using ETD. Recently, this linker was updated to use a mixture of ^{1}H- and ^{2}H to produce a similar isotopic signature (**5-3**, Figure 5-2); however, 36% of the modified peptides were only identified by the isotopic pattern and were unassignable to a specific sequence.[15] The analysis was further complicated by GALE-mediated epimerization of GalNAz into GlcNAz,[16] which allows simultaneous labeling of *O*-GalNAcylated and *O*-GlcNAcylated peptides that cannot be distinguished by MS and require manual annotation based on subcellular localization. Therefore, we aimed to develop a tagging approach that would be quantitatively appended and released, specifically modify only *O*-GlcNAcylation sites, and incorporate a positive charge upon cleavage to

facilitate ETD-MS detection.

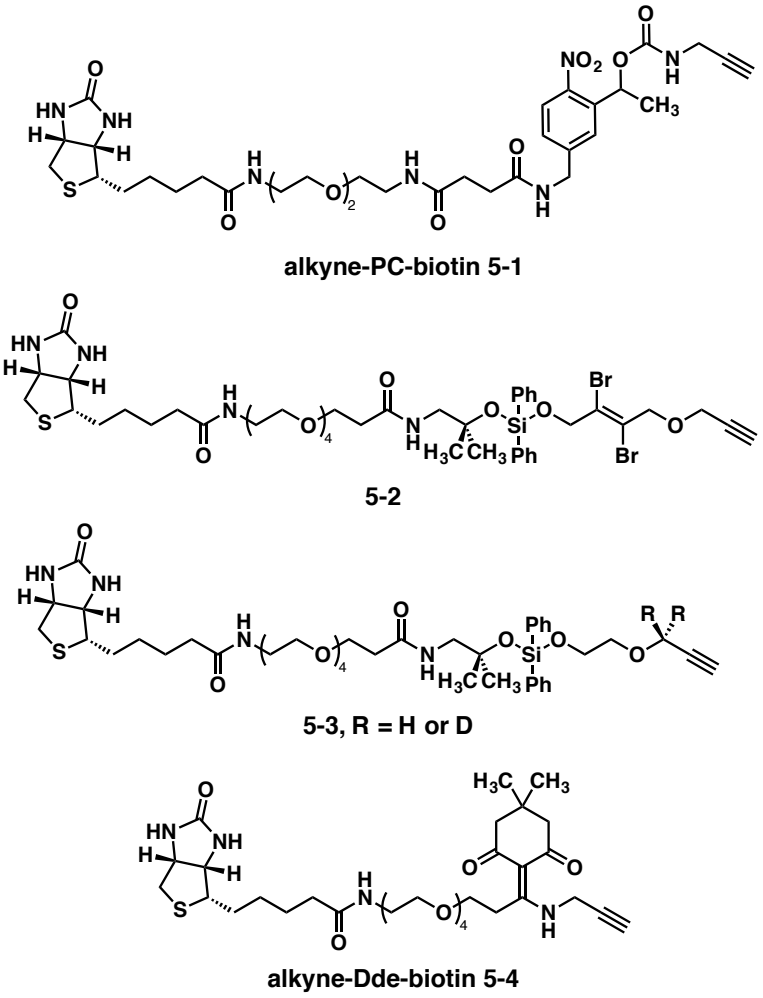


Figure 5-2. Cleavable biotin tags used for *O*-GlcNAc site identification.

5.3 Method Development

5.3.1 Validation of Chemically Cleavable Linker

To achieve these goals, we chose to examine the 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl (Dde) functional group. The Dde moiety has been used extensively as a protecting group for lysine in peptide synthesis,[17]

demonstrating its compatibility with biomolecules. The group is stable to both acid and base and can be quantitatively removed by hydrazine.[18] However, it

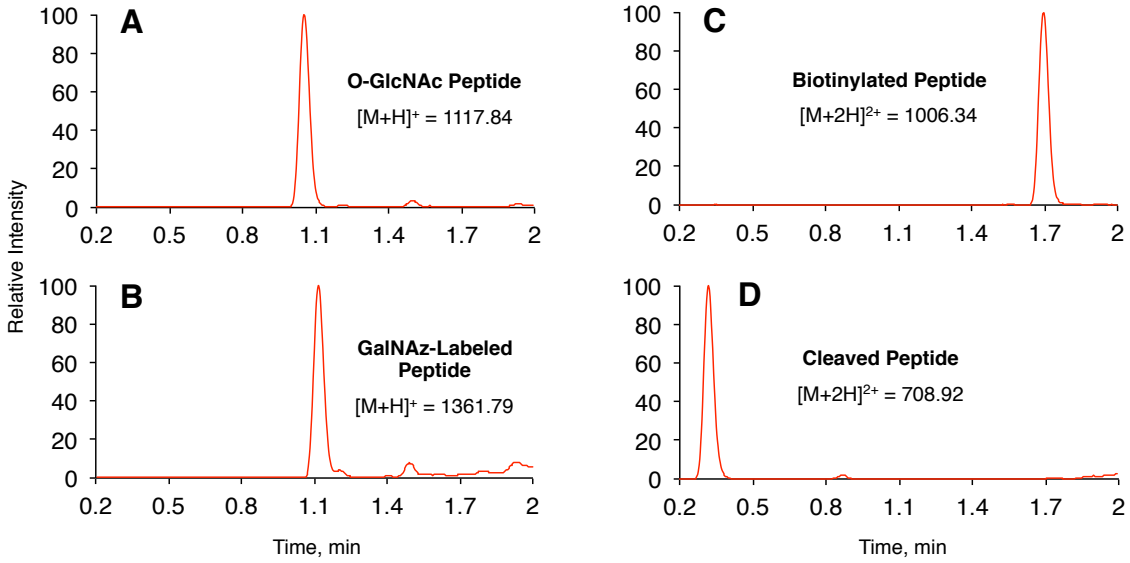
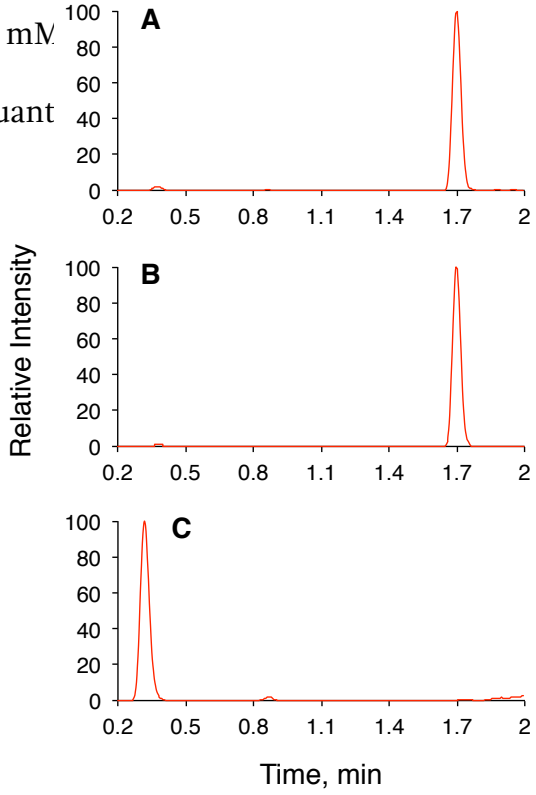


Figure 5-3. Validation of Dde linker 5-4. Reverse phase LC-MS analysis of O-GlcNAc peptide labelling reactions at (A) time 0, (B) 16 h after addition of 1 and Y289L GalT, (C) 1 h after CuAAC with 2, and (D) 1 h after cleavage with 2% aqueous hydrazine. (A) and (B) show base peak chromatograms. (C) and (D) show extracted ion chromatograms of the starting material and product within 1 m/z of calculated values.

was reported that the Dde group is incompatible with sodium dodecyl sulphate (SDS) and amine-containing buffers, common additives to protein labeling protocols.[19]

We first investigated the labeling of a model O-GlcNAcylated peptide with commercially available alkyne-Dde-biotin (**5-4**, Figure 5-2) followed by cleavage of the linker using liquid chromatography (LC)-MS (Figure 5-3). Commercially available peptide TAPT(gS)TIAPG (Figure 5-3a), where gS is the O-GlcNAcylated residue, was incubated with 100 ng mL Y289L GalT and 1 mM

UDP-GalNAz in 10 mM



ernight at 4 °C. LC-MS analysis revealed quant alNAz- labeled

Figure 5-4. Stability of Dde linker 5-4. Reverse-phase LC-MS analysis of alkyne-Dde-biotin-labelled 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*.

product (Figure 5-3b). Next, the azide-containing peptide was reacted with 100 mM of **5-4** in 10 mM sodium phosphate pH 7.6 containing 2 mM sodium ascorbate (NaAsc), 100 mM THPTA, and 1 mM $CuSO_4$. After 1 h, stoichiometric biotinylation of the peptide was observed (Figure 5-3c). 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 (Figure 5-3d).

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 5-4). In both cases, the linker remained intact, highlighting the compatibility of the linker with rigorous washing steps.

5.3.2 Comparing Dde and PC Linkers

We next tested the performance of our linker in comparison to the previously described, widely utilized photocleavable linker alkyne-PC-biotin **5-1**.[8] Briefly, HEK-293T cell lysate was subjected to chemoenzymatic labeling with UDP-GalNAz using Y289L GalT as described above. The azide-labeled protein was then split into two equal fractions and reacted with either **5-1** or **5-4** by CuAAC. An aliquot of each sample was reserved for analysis, and the remainder of each sample was subjected to cleavage using 2% hydrazine monohydrate or by UV irradiation at 365 nm. The samples were resolved by SDS-PAGE and probed for biotin using streptavidin conjugated to AlexaFluor 680 dye (Figure 5-5). Notably, a stronger biotin signal was observed for lysate labeled with **5-4** compared to **5-1**, suggesting higher labeling efficiency with the Dde linker. Furthermore, although both linkers cleaved well, the **5-1** showed slightly higher residual signal compared to **5-4**, suggesting that the Dde moiety was also released more efficiently than the photocleavable group. These results demonstrate that our new approach provides an improvement in both labeling efficiency and recovery of O-GlcNAcylated proteins compared to the most

widely used method.

5.3.3 Validation Using Known *O*-GlcNAcylated Proteins

We then evaluated the potential of the approach to pull down known O-

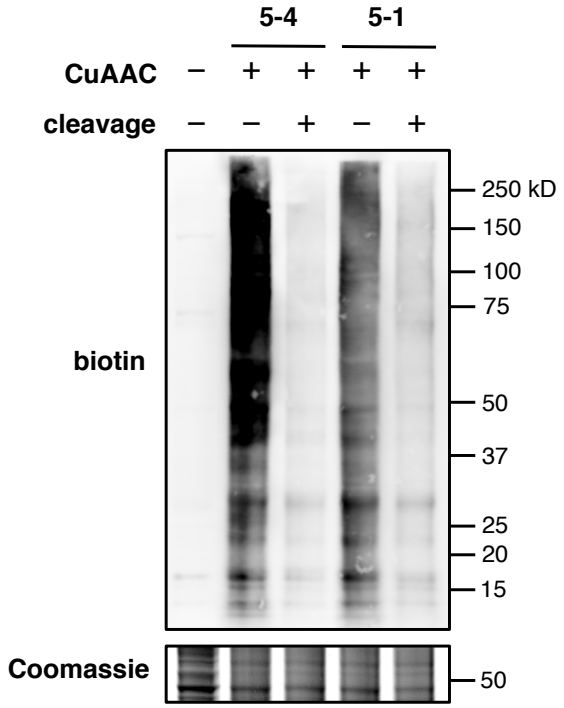


Figure 5-5. Comparison of Dde linker 5-4 and PC linker 5-1. Protein lysate was labeled using the chemoenzymatic approach using **5-4** or **5-1** and then resolved by SDS-PAGE. Protein lysates labelled with **5-4** show higher biotin signal after labelling (lane 2 vs. lane 4) and lower residual signal after cleavage (lane 3 vs. lane 5) compared to the PC linker.

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%).[20] Short-form OGT (sOGT)[21, 22] from Sf9 cells has multiple sites of O-GlcNAcylation[4] and was thus used to determine whether comprehensive site mapping could be achieved. To test the robustness of our method in a complex

mixture, each protein was added to 200 mg of adult mouse cortical lysate and 10 subjected to chemoenzymatic labeling and CuAAC using **5-2**. The labeled proteins were applied to high-capacity Neutravidin resin and washed with 0.5 mL of 1% SDS, 6 M urea, and phosphate buffered saline (PBS). The resin was then incubated for 1 h with 2% aqueous hydrazine to cleave the O-GlcNAcylated proteins from the resin. Eluted samples were 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.

Impressively, a large number of O-GlcNAcylation sites were identified on α-crystallin and sOGT (Table 5-1). The known O-GlcNAc site on α-crystallin A (Ser-162)[23] was readily recognized despite the low abundance of the O-GlcNAc modification at this site. Importantly, we observed both known and novel sites on sOGT.[4, 24] For example, we identified the previously reported Thr-662 site, 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) domains (Ser-10/Thr-12, Ser-20, Ser-52, and Ser-56) of sOGT, and we observed a doubly modified peptide at both Ser-10/Thr-12 and Ser-20, highlighting the sensitivity of the approach to identify novel glycosylation sites and multiply occupied states. As the TPR domains of OGT are thought to mediate protein-protein interactions,[25-27] such

modifications could play an integral role in OGT regulation and may provide a mechanism to selectively modulate its activity toward specific substrates.

Protein	Peptide sequence	Site(s)	Mascot ion score	Mascot delta ion score	MS
αCryA	AIPV*S*REEKPSSAPSS	S162	24.9	23.5	ETD
sOGT	I*SPT*FADAYSNMGNTLK	S10*/T12*	46.5	–	ETD
sOGT	ISPTFADAY*S*NMGNTLK	S20*	21.6	13.6	ETD
sOGT	I*SPT*FADAY*S*NMGNTLK	S10*/T12*, S20*	38.4	–	ETD
sOGT	EMQDVQGALQCY*T*R	T38	41.8	35.0	CID
sOGT	AIQINPAFADAH*S*NLASIHKDSGNIPEALASYR	S52*	53.5	7.9	ETD
sOGT	AIQINPAFADAHSNLA*S*IHKDSGNIPEALASYR	S56*	56.8	15.7	ETD
sOGT	LYLQMWEHYAAGNKPDHMIKPVEV*T*ESA	T662	33.1	8.0	ETD

Table 5-1. Validation of 5-4 to identify *O*-GlcNAc sites. *O*-GlcNAc sites identified following labeling with **5-4**, Neutravidin affinity purification, and hydrazine-mediated elution. Sites and regions of modification are denoted in red or maroon italics, respectively. Novel site identifications are marked by an asterisk.

5.4 Extension to Native *O*-GlcNAcylation Sites

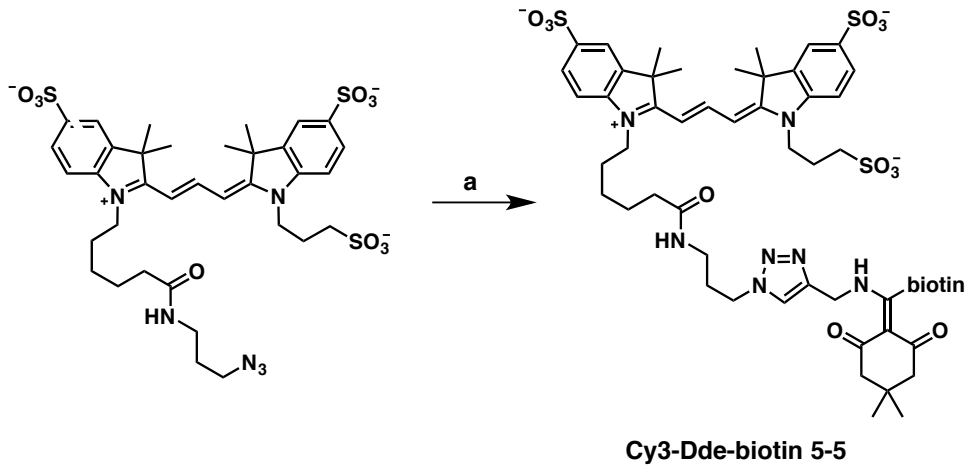
5.4.1 Choosing a Workflow

We next aimed to expand this enrichment and identification method to detect *O*-GlcNAcylation events on native proteins. Our original method called for the enrichment of intact proteins and tryptic digestion after elution.[28] Although this method was successful for exogenously added proteins and would be amenable to proteins purified by immunoprecipitation, the preponderance of unmodified peptides that would still be produced by the digestion of enriched

O-GlcNAcylated proteins may impede the detection of modified peptides by ion suppression. Therefore, two separate approaches were considered. First, we considered an approach for on-bead digestion. Here, labeled proteins would be immobilized onto Neutravidin resin, and the mixture would be subjected to tryptic digestion. Unmodified peptides would no longer be attached to the resin support and could be removed prior to chemical elution of the remaining, labeled peptides. This approach has been successfully used in a number of cases including the identification of *O*-GlcNAcylated sites in conjunction with isotopic labeling using the dibrominated or deuterated linker described earlier.[15, 29] Alternatively, we envisioned a digestion-first protocol, which would digest complex lysate directly and then proceed with labeling and enrichment steps using the tryptic peptides. This tactic had also been used successfully to enrich *O*-GlcNAcylated peptides using the photocleavable linker.[7, 8]

Selection of the proper workflow depended heavily on the extended stability of the linker to the conditions of on-bead tryptic digestion. Although the linker had shown stability to stringent wash conditions for 1 h, it was unclear if the linker could withstand elevated temperatures for 16 h. To test this, we took a practical approach by conjugating Cy3-azide dye to alkyne-Dde-biotin **5-4** using CuAAC conditions. The resulting compound, Cy3-Dde-biotin **5-5** (Figure 5-6), was used crude by incubating with Neutravidin beads, producing pink beads visible to the naked eye. Samples were then incubated for 24 h in a variety of buffer conditions (1x PBS, pH 7.4; 2 M urea, PBS; 1% SDS, PBS; 20 mM

HEPES pH 7.9, 1% SDS) at room temperature or at 37 °C with gentle rotation. The beads were then pelleted and visually inspected to see if the dye color was lost from the beads. Interestingly, all conditions incubated at 37 °C showed a



Cy3-Dde-biotin 5-5

Figure 5-6. Synthesis of Cy3-Dde-biotin 5-5. (a) **5-4** (1.2 eq), NaAsc (24 eq), THPTA (1.2 eq), $CuSO_4$ (12 eq), PBS, RT, 2 h.

decrease in dye color localized on the beads with a pink hue in the buffer (Figure 5-7), suggesting that incubation at 37 °C caused hydrolysis of the linker. To confirm that the hydrolysis was not catalyzed by some unknown additive found in the beads, solutions of **5-5** were incubated in 1x PBS, pH 7.4 overnight at room temperature or 37 °C with or without the addition of the Neutravidin beads. Again, both samples incubated at 37 °C showed less dye color on the beads regardless of whether the beads were present (Figure 5-8), suggesting that the elevated temperature alone is enough to facilitate hydrolysis.

Previous reports using Dde as a protecting group during peptide synthesis have described some instances of migration between Lys residues under

extended incubation times.[18, 30] These events only occurred when excess base was present. A previous report indicated that the Dde group would also degrade under similar b s pH 8.5.[19] Interestingly, the previously

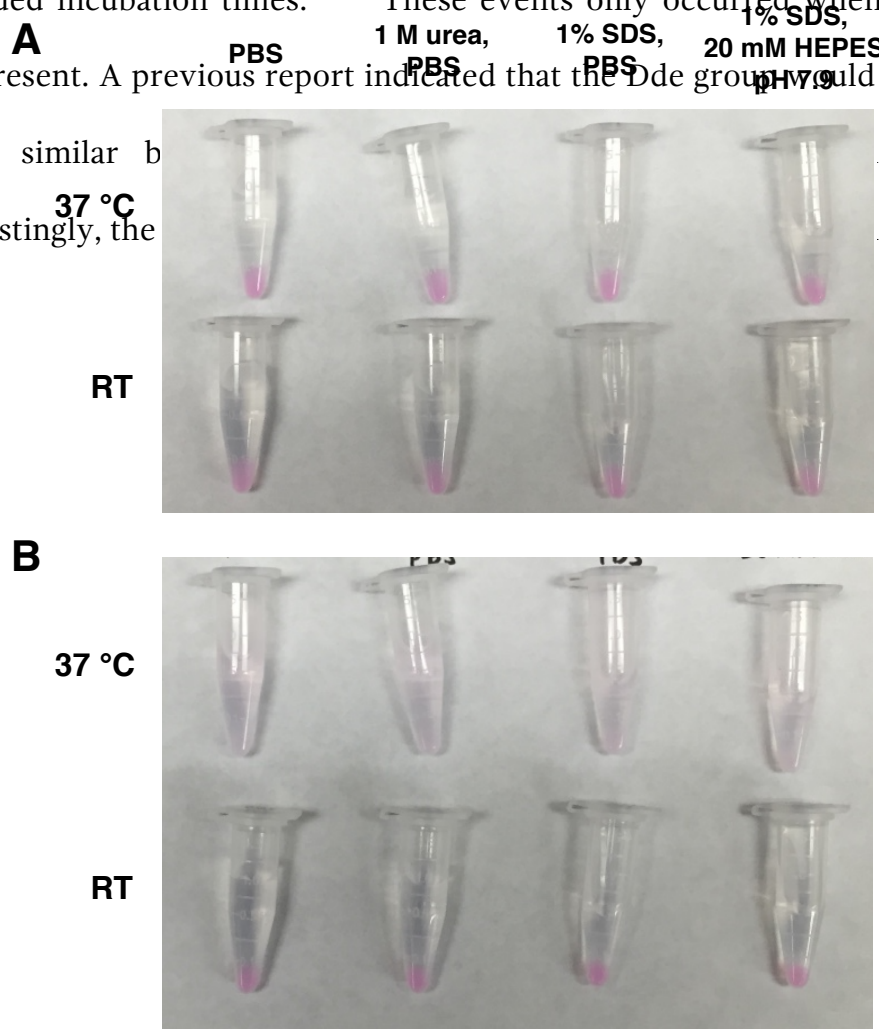


Figure 5-7. Long-term stability of Dde linker. Cy3-Dde-biotin **5-5** was (A) immobilized on Neutravidin beads and (B) incubated overnight in different buffers at RT and 37 °C. Beads incubated at 37 °C showed loss of dye to the solution, suggesting that the linker is sensitive to higher temperatures.

reported. To rectify the lability of the Dde group in peptide synthesis, a

sterically hindered alternative known as Ddiv or ivDde was developed by replacing the ethyl group of D sovaleryl group.[18] The biotin group provides a steric e e ivDde group; yet, the linker may be made more resista asing the steric bulk

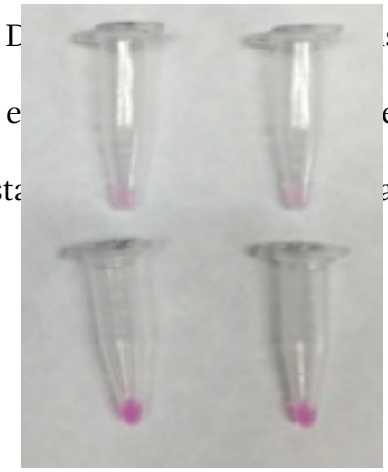


Figure 5-7. Long-term stability of Dde linker in solution. Cy3-Dde-biotin **5-5** was diluted in PBS and incubated overnight with or without beads at RT or 37 °C. Regardless of whether beads were present, less dye was immobilized onto beads when incubated at 37 °C.

adjacent to the cyclohexylidene group. Nevertheless, we decided to move forward with the alternative digestion-first approach.

5.4.2 Digestion-First Method

In this method, complex lysate would first be subjected to trypsin digestion. After numerous attempts using in solution digestion failed, we decided to turn our attention to filter-assisted sample preparation (FASP).[31] Here, filters are used to buffer exchange solubilized proteins after reduction and alkylation into a solution amenable to trypsin activity. There have also been recent reports of

improved FASP methods,[32, 33] and these approaches were attempted as well. Procedure details are given in the experimental methods section. Briefly, cell lysates from two 15-cm plates of HEK-293T cells at 75% confluency were lysed using SDS lysis buffer, quantified by the BCA assay, and adjusted to 2 mg/mL protein with lysis buffer. Samples (20 mg) were reduced by incubation for 5 min at 95 °C with 20 mM dithiothreitol (DTT) and then concentrated on Amicon centrifugal filter units (10 mg per filter unit). Samples were diluted with 8 M urea, 50 mM Tris pH 7.9 containing 20 mM IAA and allowed to react in the dark for 30 min at RT. The reduced and alkylated lysate was then concentrated and washed twice with urea buffer and twice with Tris buffer. Sequencing-grade trypsin was added to each sample at 1:50 w/w and incubated overnight at 37 °C. Peptides were separated from undigested material by centrifugation, and the filter was washed twice with Tris buffer. The filtrate was combined, acidified to pH 3 using TFA, and desalted with HLB columns. For the enhanced methods, the buffer in the final wash step was replaced with 0.2% w/v sodium deoxycholate, 50 mM Tris pH 7.9. Alternatively, the sample was first washed with 4% w/v sodium deoxycholate, 8 M urea, 50 mM Tris pH 7.9 before continuing with the urea and Tris washes. Acidification after filtration precipitated all remaining sodium deoxycholate, allowing for it to be easily separated prior to desalting. Trypsin is active in the presence of sodium deoxycholate, and the detergent is thought to remove excess, protein-bound SDS, which could inhibit trypsin activity. However, in all cases, we observed

yields of roughly 5-7 mg peptides from a starting amount of 20 mg protein. Therefore, we proceeded with the unmodified FASP method for all future sample preparation.

The remaining steps of the peptide labeling protocol were left largely unchanged from the original protein labeling protocol with a few exceptions. Generally, HLB desalting columns were used in place of protein precipitation as described in the original protocol. The clean-up step for the CuAAC step in other protocols used strong cation exchange (SCX) to separate the peptides from excess alkyne reagent.[7, 8, 34] However, repeated experiments using SCX columns failed to retain the peptides. Instead, we opted to remove excess Dde-alkyne-biotin by adding in azide agarose resin after the peptide reaction was complete. Remaining, unreacted reagent conjugated to the agarose, which could easily be filtered off. Finally, we decided to switch from 2% w/v hydrazine to a buffered solution containing 2% w/v hydroxylamine. The β-*O*-GlcNAc linkage is cleaved under basic conditions,[1] so we aimed to produce an elution buffer with a neutral pH. The pK_a of the hydrazinium ion is about 8, so a buffered solution of pH 7 will only contain about 10% of the reactive, deprotonated nucleophile. Conversely, hydroxylammonium has a pK_a of about 6, which would yield a solution containing roughly 90% of the deprotonated compound at pH 7.

5.4.3 Comparing Dde and PC Linkers with Endogenous *O*-GlcNAc Sites

With these adjustments, we conducted the protocol using both **5-1** and **5-4** as a direct comparison. Samples were eluted twice and analyzed separately to

ensure that complete elution was achieved and avoid incomplete cleavage that was observed previously with the **5-1**.[8] In total, 445 unique *O*-GlcNAcylated peptides on 256 different proteins were identified across all four samples using higher-energy collisional dissociation (HCD) MS/MS (Table 5-2). Here, peptides are counted as the number of unique peptide sequences; however, possible multiple modifications or different modification sites on the same peptide were not counted as additional unique peptides. Therefore, this count is likely an underestimation of the number of possible sites. Importantly, peptides containing either tag could easily be identified with their MS/MS spectra by the presence of up to three signature ions caused by the fragmentation of glycosidic bonds within the linker: 300.1 (GalNAz-GlcNAc), 318.1 (H_2O adduct), and 503.2 (GlcNAc-Ser/Thr) *m/z*, which are observed from both linkers due to the identical aminomethyltetrazolyl group left after cleavage.[8] **5-4** strongly outperformed **5-1** in peptide coverage. Of the 442 unique peptides identified in all samples, 414 were observed in the Dde samples, whereas only 257 were observed in the PC sample. In both cases, a single elution step was sufficient to release nearly all of the identified peptides, with only 7 unique peptides identified solely in the second Dde elution step and 16 in the second PC elution.

The modified peptides revealed by this analysis covered a variety of known *O*-GlcNAcylated proteins.[5, 7, 8, 13] As expected, sites on many multiply *O*-GlcNAcylated proteins such as host cell factor 1 (Hcfc1), Msx2-interacting protein (Mint), myosin phosphatase target subunit 1 (Mypt1), nucleoporins

(Nup98, Nup153, Nup214), glutamine and serine rich protein 1 (Qser1), and lysine deficient protein kinase 1 (Wnk1) were observed in the dataset. Interestingly, Qser1 and the related protein proline and serine rich protein 1 (Prsr1), which was also observed to be *O*-GlcNAcylated in this dataset, are ubiquitously expressed but have no known function. Thus, these results may help provide evidence for the role of these proteins in the cell and its regulation by *O*-GlcNAcylation.

5.5 Conclusions and Outlook

The results presented herein provide the framework for a new approach to identify sites of *O*-GlcNAcylation on proteins using a novel cleavable tag **5-4** based on the Dde functional group. This linker provides numerous benefits over existing linkers due to its complete cleavage by hydrazine and hydroxylamine as well as its ability to leave only the small aminomethyltetrazolyl group after cleavage. Moreover, the remaining linker provides an additional positive charge to the labeled peptide, aiding in ETD fragmentation. There still remain areas of improvement for the linker, including synthetic variants closer to the ivDde structure to avoid cleavage at elevated temperatures. Nevertheless, we envision that this linker can be combined with isotopic labeling methods such as SILAC or dimethyl labeling to allow for the quantitative measure of *O*-GlcNAcylation dynamics based on individual sites. These experiments will provide critical information regarding the regulation of protein *O*-GlcNAcylation to help delineate its physiological functions in diverse biological systems.

Table 5-2. *O*-GlcNAc sites identified from HEK-293T lysate. Samples were processed using a digestion-first protocol and labeled using the two-step chemoenzymatic approach with **5-4** or **5-1**. Peptides were identified by HCD-MS/MS and are presented as unique sequences.

Uniprot	Peptide sequence	Peptide counts			
		Dde-1	*Dde-2*	*PC-1*	*PC-2*
4ET_HUMAN	SVLHPPGSGSHAAAVSVQTTPQNVPSR	1			
AAK1_HUMAN	LTDPIPTTETSIAPR			1	
ABLM1_HUMAN	STSQGSINSPVYSR	1			
ADRM1_HUMAN	SQSAAVTPSSTTSSTR	1	1	1	2
AF10_HUMAN	NPGTTVSAASPFPQGSFSGTPGSVK	2			
AGFG1_HUMAN	APVGSVVSVPSQSSASSDK	6	3	5	3
	SSSADFGTFNTSQSHQTASAVSK	6	3	7	4
	VVASVHASISGSSASSTSSTPEVKPLK	1			
AGFG2_HUMAN	TLLGDPAPSLSVAASTSSQPVSQSHAR	4	1	2	
AHNK_HUMAN	GPQVSSALNLDTSK	4	2	3	2
AHNK2_HUMAN	ESEIPTSEIQTPSYGFSLLK	1			
AINX_HUMAN	SNVASSAACSSASSLGLGLAYR	1			
ALMS1_HUMAN	ISVASEPVDQTTGTPAVTSTSYSQYR	1		1	
	TETPSVSSSLYSYR	1			
	VSVAPGPVGQTTGAPTITSPSYSQHR	1		2	
AMRA1_HUMAN	TSASSVSLLSVLR	1			
ANKH1_HUMAN	NAFPLGAPTLVTSQATTLSTFQPANK	4	3	2	3
	SIHANFSSGVGTTAASSK	2			
	VSTSPVGLPSIDPSGSSPSSSSAPLASFSGIPGTR	2			
ANR17_HUMAN	IGSSAPTTTAANTSLMGIK	4		2	1
	LKVEDEPEVLTEPPSATTTTTIGISATWTTLAGSHGK	2			
	MTTVALSSTSQTATALTVPAISSASTHK	3		1	
	MTVPPLATSSAPVAVPSTAPVTYPMPQTPMGCPQPTPK	1			
	QHFSPLSLLTPCSSASNDSSAQSVSSGVR	2		1	
	TSNATTTTVTTTASNNNTAPTNATYPMPTAK	1			
APMAP_HUMAN	AGPNGTLFVADAYK			1	
ARI1A_HUMAN	GGTPGSGAAAAAGSKPPPSSSASASSSSSSFAQQR	2		1	1

	NPQMPQYSSPQPGSALSPR	1			
ARI1B_HUMAN	VMPTVPTSQVTGPPPQPPPIR	9	2	6	2
ARI3A_HUMAN	LPVSLAGHPVVAAQAAAVQAAAAQAAVAAQAAALEQLR	1			
ARIP4_HUMAN	VVTTTDIVIPGLNSSTDVQAR	5		4	1
ARNT_HUMAN	HSNPTQGATPTWTPTTR	2			3
ASPP2_HUMAN	ENLPVSSDGNLPQQAASAPSR			1	
ATF1_HUMAN	TTPSATSLPQTVVMTSPVTLTSQTTK	3	2	5	
ATF7_HUMAN	SAAEAVATSVLTQMASQR	1			
ATX1L_HUMAN	APSATSPSGQLPHHSSTQPLDLAPGR	2		2	
ATX2L_HUMAN	SAAPAPISASCPEPPIGSAVPTSSASIPVTSSVSDPGVGSISPASPK	3	1	2	1
BAG3_HUMAN	SQSPAASDCSSSSSSASLPSSGR			2	
BAZ2B_HUMAN	GGLSTGVASLSSTINPCGHLFR	2			
	LPSSAASSTTPTSSSTPSVASVVSK	2			
BCORL_HUMAN	TPPMPVLTPVHTSSK	1			
BPTF_HUMAN	FLFTPLATTATTASTTTTTVSTTAAGTGEQR	5		2	1
	GQPVSTAVSAPNTVSSTPGQK	4	2	2	2
	LEQQKPTVIATSTTSPTSSTTSTISPAQK			1	
	STVTTTTTTVTK	1			
	TVITEVTTMTSTVATESK	4	2	6	3
	VMVAPISGSVTTGTK	3	1	2	1
BRD8_HUMAN	LLEAGPTQFTTPLASFTTVASEPPVK			1	
CACL1_HUMAN	AAPAPTASSTININTSTSK			1	
CARF_HUMAN	SSSSTNTSLLTSK	1			1
CATL1_HUMAN	YSVANDTGFVDIPK	2		1	
CBL_HUMAN	VPVSAPSSSDPWTGR	3	1	2	2
CDK12_HUMAN	TSAVSSQANSQPPVQVSVK	5	2	4	2
CDK13_HUMAN	TENQHVPTTSSSLTDPHAGVK	2			
CDK8_HUMAN	VVPPTTTSGGLIMTSDYQR	3			
CE170_HUMAN	EINDVAGEIDSVTSSGTAPSTTVSTAATTPGSAIDTR	1			
CIC_HUMAN	GYGSAPSSSASSPASSSASAATSFSLGSGTFK	1		1	
CKAP5_HUMAN	ISTSTGISPQMEVTCVPTPTSTVSSIGNTNGEEVGPSVYLER	2		2	1
CLIP1_HUMAN	VQAEDEANGLQTTPASR	1			
CNOT1_HUMAN	APLAGQVSTMVTTSTTTTVAK	16	3	18	5
CNOT2_HUMAN	SLSQGTQLPSHVTPTTGVPTMSLHTPPSPSR	1			
CNOT4_HUMAN	SNPVIPISSSNHSAR	1			
CRTC2_HUMAN	SLQQPGLPSQSCSVQSSGGQPPGR	2	1	1	
CUL4B_HUMAN	MAEESSSSSSSSSPTAATSQQQQLK	1		2	
CUX1_HUMAN	QAPLSQSDITILTPK	1			
DAPLE_HUMAN	TCSTSATTTAPSNSTPIAR	1		1	
DIDO1_HUMAN	TYFPGPPGDGHPEPSPLEDLSPCPASCGSGVVTTVTVSGR		1	1	
	VLSSLKPAAPSPATAATTAAAASTAASSTASSASK	4		1	

DLG5_HUMAN	SLTPSTTVSSILR	1			
DSRAD_HUMAN	NAEFLTCNIPTSNASNNMVTTEK	1		4	1
E41L1_HUMAN	DVLTSTYGATAETLSTSTTTHVTK	2			
E41L2_HUMAN	TITYESPQIDGGAGGDSGTLLTAQTITSESVSTTTTTHITK	2		1	
ELF1_HUMAN	FILQAIPSSQPMTVLK	2			
ELF2_HUMAN	ALTPVSIAHGTPVMR	1			
	VAMQVPVVMTSLGQK	1			
EMD_HUMAN	LSPPSSSAASSYSFSDLNSTR	3			
EMSA1_HUMAN	VKEEQYLGHEGPGGAVSTSQPVELPPPSSLALLNSVVYGPER			1	
EMSY_HUMAN	IISSNIVSGTTTK	2			
	ITFTKPSTQTTNTTTQK	1			
	MSNIMQSIANSLPPHMSPVK	2			
	QTASQVEQPIITQGSSVTK	1			
	TTSGSIITVVPK	1		2	
	VIIVTTSPSSTFVPNILSK	3	1	3	2
EP300_HUMAN	SGSSPNLNMGVGGPGQVMASQAQQSSPGLGLINSMVK	1			
EP400_HUMAN	AAAAPFQTSQASASAPR	1			1
	AQPAITTGGSAAVLAGTIK	1			
	AVGSPATATPDLVSMATTQGVR	3	2	4	2
	AVTSVTASAVVTTNLTPVQTPAR	5	1	1	1
	SLVPQVSQATGVQLPGK	1	1	1	
	TAAPTTASAAPQGPLR	1			
	TQFLTTPISQAQK	3	1		1
	TSVTGTSMPTGAVSGNVIVNTIAGVPAATFQSINK	3			
EPC1_HUMAN	LTVPSSVATVNSIAPINAR	1			
EYA4_HUMAN	TEPLNSSETTATTGDGALDTFTGSVITSSGYSPR	1			
F193A_HUMAN	SPPSVSSASSGSGSSSPITIQQHPR	2			
	TATTTPGFVDTR			1	
	VVMATSSATSSVSCTATTVQSSNSQFR	2	1	1	
F208B_HUMAN	ETPLPVSLPSDK	2			
	SLSDTLVSTTAPSGIVNVSVK	1			
	VASYSGTVTQATFTR	1			
FLIP1_HUMAN	VTSTITITPVTTSSAR	2	1	1	2
FNBP4_HUMAN	ATEISTAVVQR		1	1	
FOXC1_HUMAN	GSPQSAAAELSSGLLASAAASSR	2			
FOXK1_HUMAN	EPAAAVAATATTTPATATTASASASSTGEPEVK	13	4	11	3
	HAVPTNSLAGNAYALTSPLQLLATQASSSAPVVVTR	1		2	
	VVTTSANSANGYILTSQGAAGGSHDAAGAAVLDLGSEAR	5	1	4	
FOXK2_HUMAN	FAQSAPGSPLSSQPVLITVQR	1			
GABPA_HUMAN	YVLASQEQQMNEIVTIDQPVQIIPASVQSATPTTIK	2			
GANP_HUMAN	SPTSVGAFPSTSAFGQEAGEIVNSGFGK	3	2	1	

GCR_HUMAN	VSASSPSLAVASQSDSK		1		
GEMI5_HUMAN	TVIESSPESPVTITEPYR	1			
GMEB2_HUMAN	VVSTLPSTVLGK	2			
GPKOW_HUMAN	EGVLPLTAASTAPISFGFTR	1			
GSCR1_HUMAN	QVPVSGYLASAAGPSEPVTLASAGVSPQGAGLVIQK	1			
GSE1_HUMAN	RVPMGPIIVPPGGHSVPSTPPVVTIAPTK	1			
HCFC1_HUMAN	APVTVTSLPAGVR	3	2	3	1
	AVTTVTQSTPVPGPSVPPPEELQVSPGPR	2			
	HSHAVSTAAMTR	1			
	IPPSSAPTVLSVPAGTTIVK	12	5	7	5
	ISVATGALEAAQGSK	7	3	5	2
	LVTPVTVSAVKPAVTTLVVK	6	3	3	1
	QEAAASLVTSTVGQQNGSVVR	7	2	5	1
	QTSATSTTMTVMATGAPCSAGPLLGPSMAR	7	3	8	1
	SGTVTVAQQAQVVTTVVGGVTK	16	4	13	3
	SPISVPGGSALISNLGK	2	1	1	1
	SPITIITTK	6	3	5	2
	SSVGAGEPR	1			
	TAAAQVGTSVSSATNTSTRPIITVHK	20		7	2
	TIPMSAIITQAGATGVTSSPGIK	25	13	18	9
	TMAVTPGTTTLPATVK	11	4	12	4
	VASSPVMVSNPATR	3		1	1
	VMSVVQTKPVQTSAVTGQASTGPVTQIIQTK	7	2	6	4
	VMTSGTGAPAK	1		1	2
	VTGPQATTGTPLVTMRPASQAGK	2		2	
	YDIPATAATATSPTPNPVPSVPANPPK	2		1	
HGS_HUMAN	AEPMPSASSAPPASSLYSSPVNSSAPLAEDIDPELAR	11	3	10	3
HSPB1_HUMAN	LATQSNEITIPVTFESR	3	1	1	1
HTF4_HUMAN	GSTSSSPYVAASHTPPINGSDSILGTR	1			
	LSYPPHSVSPTDINTSLPPMSSFHR	1			
I2BP2_HUMAN	AAASLAAVSGTAAASLGSAQPTDLGAHK	3		4	
	INGEAQPWLSTSTEGLK	2		2	
I2BPL_HUMAN	FEYPPPPVSLGSSSHTAR	8	3	5	2
	GPPTPAPPGAPGGPACLGGTPGVSATSSSASSSTSSSVAEVGVGAGGK	1		1	1
	SRFEYPPPPVSLGSSSHTAR	5		4	
IF4G3_HUMAN	EQEGQTSETTAIVSIAELPLPPSPTTVSSVAR	11		3	1
	KEQEGQTSETTAIVSIAELPLPPSPTTVSSVAR	2		1	
	STIAAPTSSALSSQPIFTTAIDDR	8	2	2	3
ITB1_HUMAN	VCECNPNYTGSACDCSLDTSTCEASNGQICNGR	2		5	2
ITSN2_HUMAN	AQSLIDLGSSSSTSSTASLSGNSPK	1		1	
JHD2C_HUMAN	HSVPQSLPQSNYFTTLSNSVVNEPPR	4		2	

	SPTHLTVSSTNTLR	1			
JUNB_HUMAN	LIVPNSNGVITTTPTPPGQYFYPR	1			
K2026_HUMAN	CLTSALQIPVTVALPTPATTSPK	1			
K2C8_HUMAN	TTSGYAGGLSSAYGGLTSPGLSYSLGSSFGSGAGSSSFSR	3		1	
KANL3_HUMAN	LPTPMQSLGAITTGTSTIVR	3			
	SSSSEGGVSASPVPSVVSSSTAPSALHTLQSR	2		1	
	VPTTITLTLR	1			
KCMF1_HUMAN	SNMHFTSSSTGGLSSSQSSYSPSNR			1	
KDM3B_HUMAN	NSILASSGFGAPLPSSSQPLTFGSGR	1	1	2	
	TLEQVGQGIVASAAVVTTASSTPNTVR	10	1	8	5
	TLVVQDEPVGGDTPASFTPYSTATGQTPLAPEVGGAENK	1			
	VEHSPFSSFASQASGSSSSATTVTSK	7	1	5	2
KMT2D_HUMAN	SLPSDPFSR	1			
KRT81_HUMAN	LCEGIGAVNVCVSSSR		1		
LAP2A_HUMAN	SSSSSSQPEHSAMLVSTAASPSLIK	1			
LAR4B_HUMAN	EPSVPASCAVSATYER	1			
	TLSADASVNTLPVVVSR	1			
LIMD1_HUMAN	TPSVSAPLALSCPR	1			
LIN54_HUMAN	LGAQTPVTISANQIILNK	1			
	TITISESGVIGSTLNSTTQTPNK	1		1	
LMNA_HUMAN	ASASGSGAQVGGPISSGSSASSVTVTR	9	5	8	3
	SVGGSGGGSFGDNLVTR	1	1		1
LMO7_HUMAN	TSTTGVATTQSPTPR			1	
LPP_HUMAN	STGEPLGHVPAR	2	1	1	2
LRIF1_HUMAN	ILATATTSTSGMVEASQMPTVIYVSPVNTVK	1			
MAFK_HUMAN	STELSSTSVPFSAAS	2	2	2	1
	VATTSVITIVK	2	1	2	1
MAP4_HUMAN	ASPSKPASAPASR	1			1
MATR3_HUMAN	DLSAAGIGLLAAATQSLSMPASLGR	3	1	2	1
MAVS_HUMAN	VPTTLMPVNTVALK	1			
MBD5_HUMAN	DIPNPLIAGISNVLNTPSSAAFPTASAGSSSVK	1			
MCAF1_HUMAN	NPTASAAPLGTTLAVQAVPTAHSIVQATR	1			
	NPVSLPSLPNPTKPNNVPSVPSPSIQR	1			
	TSLPTVGPSGLYSPSTNR	2			
MED15_HUMAN	FPPTTAVSAIPSSSIPLGR	9		7	
MGAP_HUMAN	IPGVSTPQTLAGTQK	1			
	TTGITTPVASVAFPK	2			
MINT_HUMAN	ADRPSLEKPEPIHLSVSTPVTQGGTVK	3		2	
	AQSTPSPALPPDTK	1		1	
	ASDVDTSSSTLR	1		1	1
	QPLFVPTTSGPSTPPGLVLPHTEFQPAPK	3		2	

	SLVSTPAGPVNVLK	1			
	VNTSEGVVLLSYSGQK	2	2	2	
MKL2_HUMAN	VSESPSPVTTNTPAQFASVSPTVPEFLK	1			
MLXIP_HUMAN	EGMLASTVSQSNVVIAPAAIAR	2			
MUC5A_HUMAN	GCPVTSTPVTAPSTPSGR			1	
MYPT1_HUMAN	DSVPTAVTIPVAPTVVNAAASTTTLTTTTAGTVSSTTEVR	3		1	3
	QDDLISSSVPSTTSTPTVTSAAGLQK	12	2	5	1
	RQDDLISSSVPSTTSTPTVTSAAGLQK	6		2	3
	TKPLASVTNANTSSTQAAPVAVTTPTVSSGQATPTSPIK	3		1	1
NASP_HUMAN	ATLVESSTSGFTPGGGGSSVSMIASR	2			
NCOA3_HUMAN	AVSLDSPVSVGSSPPVK	1			
NCOA6_HUMAN	SIVTTLVPSELISAVPTTK	3		1	1
NCOR1_HUMAN	HTSVVSSGPSVLR	1			
	IMPLPAGGPSISQGLPASR	2			
	NQVSSQTPQQPPTSTFQNSPSALVSTPVR	3	1	5	2
	SSHLEVSQASQLLQQQQQQQLR	1			
	YPPHSVQYTFPNTR	3			
NCOR2_HUMAN	AISSASIEGLMGR				1
	VVTLAQHISEVITQDYTR	1		1	
	YPPHSLSYPVQIAR	3			
NFIA_HUMAN	ASPHATPSTLHFPTSPIIQQPGPYFSHPAIR	3			
NFIC_HUMAN	LALPPATKPATTSEGGATSPTSPSYSPPDTSPANR	1			
NFRKB_HUMAN	IQTVPASHLQQGTASGSSK	1		2	
	LMPALGVSVADQK	1			
	QVPVSTTVVSTSQAGK	2		2	1
	TVAVASGAASTPISISTGAPTVR	6	1	2	2
NOTC1_HUMAN	CNCLLPYTGATCEVVLAPCAPSPCR			1	
NOTC2_HUMAN	YSCVCSPGFTGQR		1		1
NOTC3_HUMAN	CQCPAGYTGPLCENPAVPCAPSPCR	1			
NPM_HUMAN	MSVQPTVSLGGFEITPPVVLR	1			
NU153_HUMAN	ALTLTVVSESAETMTASSSSCTVTTGTLGFGDK	4		2	
	CQPVFSFGNSEQTK			1	
	CVSCMSEKPGSSVPASSSSTVPVSLPSGGSLGLEK	2		1	
	FGVSSSSSGPSQTLTSTGNFK	11	1	5	3
	GFDTSSSSSNSAASSSFK		2	1	2
	IGVSSDSGSINPMSEGFK	2	1	4	2
	ISLPITSSSLPTFNFSSPEITTSSPSPINSSQALTNK	8	4	6	2
	QQEPVTSTSLVFGK	6	2	4	2
	SSSAGFSFGTGVINSTPAPANTIVTSENK	6	4	3	3
	STEANVLPPSSIGFTFSVPVAK	31	6	12	10
	VQMTSPSSTGSPMFK	2	2		2

NU214_HUMAN	ASSTSLTSTQPTK	2	2		1
	EPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVPDAR	6	2	2	
	GGGFFSGLGGKPSQDAANKNPFSSASGGFGSTATSNTSNLFGNSGAK	1			
	KEPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVPDAR	8	1	4	
	LGELLFPSSLAGETLGSFSGLR	4	2	2	3
	NNPATPSTAMGSSVPYSTAK	4		4	1
	NPFSSASGGFGSTATSNTSNLFGNSGAK	10	4	5	3
	SSATVTGEPPSYSSGSDSSK	1			
	TFGGFASSSFGEQKPTGTFSSGGGSVASQGFGFSSPNK	5		2	
	TGGFGAAPVFGSPPTFGGSPGFGGVPAFGSAPAFTSPLGSTGGK	1		1	
NUFP2_HUMAN	TIQNSSVSPTSSSSSSSSTGETQTQSSSR	1		1	
NUP53_HUMAN	ASTSDYQVISDR	1			
NUP98_HUMAN	AASLMNIPSTSSWSVPPPLTSVFTMPSPAPEVPLK	1		3	
	FTSGAFLSPSVSVQECR	2		1	2
	GPQNQVGAGTTTGLFGSSPATSSATGLFSSSTTNSGFAYGQNK	1		1	1
	KGPQNQVGAGTTTGLFGSSPATSSATGLFSSSTTNSGFAYGQNK	3			1
P121A_HUMAN	APPTLQAETATKPQATSAPSPAPK	2			
P121B_HUMAN	QSFLFGTQNTSPSSPAAPAASSASPMFKPIFTAPPK	2			
P66A_HUMAN	GTTATSAQANSTPTSVASVVTSAESPASR	2		4	3
	TPLSTGGTLAFVSPSLAVHK	2			
P66B_HUMAN	LQQQAALSPTTAPAVSSVSK	2	1		
PACS1_HUMAN	LAQATSSSSSTSAAAASSSSSSTSTSTSMAVAVASGSAPPGGPGPGR	1			
PAPOA_HUMAN	EQLDTETSTTQSETIQTAASLLASQK			1	
PCF11_HUMAN	SPEEPSTPGTVVSSPSISTPPIVPDIQK			1	
PDLI5_HUMAN	ANNSQEPSPQLASSVASTR	2		2	1
	EVVKPVPITSPAVSK	4			
PF21A_HUMAN	FTPTTLPTSQNSIHPVR	3	1	3	1
PHAR4_HUMAN	FIISTSITTAPAATTAATSLAK	17	4	7	2
PHC1_HUMAN	QPGTAQAQALGLAQLAAAVPTSR	4		2	1
PHC3_HUMAN	STSQTQSLTICHNK	1		1	
PHF3_HUMAN	GSAVATSHFEVGNTCPSEFPSK	1			
PICAL_HUMAN	KPHTSLTTAASPVSTSAGGIMTAPAIDIFSTPSSSNSTSK	5		3	1
	SSGDVHLSISSDVSTFTTR			1	
PKCB1_HUMAN	SSAQTSAAGATATTSTSSTVTVTAPAPAATGSPVK	1			
	SSAQTSAAGATATTSTSSTVTVTAPAPAATGSPVKK	1			
	TPPSTTVGSHSPPETPVLTR	1		1	
PKHA5_HUMAN	IVNVSLADLR	4	1	2	1
PLEC_HUMAN	AGVAAPATQVAQVTLQSVQR	1			
PLIN3_HUMAN	TLTAAAVSGAQPILSK	3		2	1
	VASMPLISSTCDMVSAAYASTK	4	1	2	
PLRG1_HUMAN	MPSESAAQSLAVALPLQTK	1			

POGZ_HUMAN	LAPSFPSPPAVSIASFVTVK	6	3	2	1
	STPSTSTTPTATQPTSLGQLAVQSPGQSNQTTNPK	4		1	
PRC2B_HUMAN	DSDFSLPPGSASGPTGSPVVK	6			
	GGLPVSQSQEIFSSLQPFR	4	3	2	2
PRC2C_HUMAN	AVSEMSTEIGTMISVSSAEYGTNAK	3		3	
	ESVTDYTTPSSSLPNTVATNNTK	2	2	2	1
	ETIQQSSSLTSVPPTTFSLTFK	6	1	7	2
	RETIQQSSSLTSVPPTTFSLTFK	5		3	
PRSR1_HUMAN	GFLTSNDTNLINSSALSSAVTSGLASLSSLTLQNSDSSASAPNK	1			
	GPHPGTSDLHISSTPAATTLPVMIK	2			
PTN12_HUMAN	TNISTASATVSAATSTESISTR	2	1	2	
PUM1_HUMAN	SASSASSLFSPSSTLFSSSR	2	1	2	2
QSER1_HUMAN	KTEALQVATTSPTANTTGTATTSSTTVGAVK	2			
	QSSLSCSPIGDSTQVSNGGLQQK	1			
	SCSTEQPLTSTK			1	1
	TAQAAASGTTLLPQFR	2		1	
	TEALQVATTSPTANTTGTATTSSTTVGAVK	1	1	1	
	TSQGTVPTALAFER	3		1	2
RBM12_HUMAN	VNLPTTVSNFNNPSPSVVTATTSVHESNK	16		8	2
RBM14_HUMAN	AQPSVSLGAAYR	1			2
RBM27_HUMAN	AANIVIQTEPPVPVSINSNITR	5	4	7	1
	MMSKPQTSGAYVLNK	1			
	TQTQRPNLIGLTSGDMDVNPR	1			
RBM4_HUMAN	HLLPTSGAAATAAAAAAAAAAVTAASTSYYGR	5		1	
RBM7_HUMAN	TMDNMTSSAQIIQR	1		1	1
RBMS2_HUMAN	MLAQSALSPYLSSPVSSYQR	2	1		
RC3H2_HUMAN	VGVNNTVTTTAGNVISVIGSTETTGK	1			
RFX5_HUMAN	TAEVPVSEASGQAPPAK	1			1
RGPD3_HUMAN,RGPD4_HUMAN	EGFSIPVSADGFK	2			
	TDVIQGDDVADAASEVEVSSTSETTTK	1			
RIPK1_HUMAN	MQSLQLDCVAVPSSR		1		
RLA2_HUMAN	LASVPAGGAVAVSAAPGSAAPAAGSAPAAAEEK			1	
RPRD2_HUMAN	SAVSTSVPTKPTENISK	3		1	
	SFNYSPNSSTSEVSSTSASK	1		1	
	SLFSPQNTLAAPTGHPPTSGVEK	1	1	1	
	TPAPATTTSHNPLANILSK	2		1	
RPTOR_HUMAN	NYALPSPATTEGGSLTPVR	1		1	
RRP1B_HUMAN	TPTSSPASSPLVAK	1			
RSBNL_HUMAN	QLQPPAAPSPQSYGSPASWSFAPLSAAPSPSSSR	1			
S30BP_HUMAN	GTTTNATSTTTTTASTAVADAQK	7	2	8	2
	KGTTTNATSTTTTTASTAVADAQK	3		3	

	SKWDSAIPVTTIAQPTILTTTATLPAVVTVTTSASGSK	3			
	WDSAIPVTTIAQPTILTTTATLPAVVTVTTSASGSK	5	1	3	
SAP_HUMAN	DNATEEEILVYLEK	1			
	TNSTFVQALVEHVK	5		2	
SARNP_HUMAN	FGIVTSSAGTGTTEDTEAK			1	
SBNO1_HUMAN	FIQTTASTRPSVSAPTVR	1			
	TPPVTTNR			1	
SC16A_HUMAN	TLENPVNVYNPSHSDSLASQQSVASHPR	2			
SC24A_HUMAN	ASSQPTVSGNTSLTTNHQYVSSGYPSLQNSFIK	1	1		
SC24B_HUMAN	SSPVVSTVLSGSSGSSSTR	7	2	5	2
	TPPTANHPVEPVTSVTQPSELLQQK	9	2	4	2
SC24C_HUMAN	APPSSGAPPASTAQAPCGQAAYGQFGQGDVQNGPSSTVQMQR	5	1	2	
SCAF8_HUMAN	ETVQTTQSPTPVEK	1			1
	SSEPVKETVQTTQSPTPVEK	1			
SCML2_HUMAN	NPMYIHTSVSQDFSR	1	1		
SCYL2_HUMAN	SSASSTFTSVPSMGIGMMFSTPTDNTK	1			
SET1A_HUMAN	GSTPYSQDSAYSSSTTSTSFKPR	1			
	QDTPSSFGQFTPQSSQGTPYTSR	2			
SET1B_HUMAN	DFSFTPTFSEPSGPLLLPVCPLPTGR			1	
SFPQ_HUMAN	FPPLGGGGGIGYEANPGVPPATMSGSMMGSDMR	1			
SH3R1_HUMAN	IGVFPGNYVAPVTR	1			
	LQGNGVAGSPSVVPAAVVSAAHIQTSPQAK	1			
SI1L3_HUMAN	RPVSFPETPYTVSPAGADR	1			
SIX4_HUMAN	VLQSSANSATTTSYSPSVPVSFPGLIPSTEVK	2			
SKIL_HUMAN	TCTSVPETLHLNPSLK	1			
SLAI1_HUMAN	MPSTTAISSNISSPVTVR	2		1	
SMAP2_HUMAN	STAPVMDLLGLDAPVACSIANSK			1	
SMG7_HUMAN	KTPVSEAR	1			
	YPNNSMFNEVYGK	1			
SNRPA_HUMAN	AVQGGGATPVVGAVQGPVPGMPPMTQAPR	1			
SON_HUMAN	ILDSFAAAPVPTTTLVLK	4	2	2	2
	SMMSSYSAADR				1
SP2_HUMAN	TPSGEVQTVLVQDSPPATAAATSNTTCSSPASR	2		1	
SPAT2_HUMAN	CDSLLTCPPASKPSAFPSK	1			
SPTB2_HUMAN	DDEEMNTWIQAISSAISSDKHEVSASTQSTPASSR	2			
	HEVSASTQSTPASSR	3		2	2
SRC8_HUMAN	NASTFEDVTQVSSAYQK	5		2	1
SRCAP_HUMAN	AETQGANHTPVISAHQTR	1		2	
SRGP1_HUMAN	NSPTPATSTESLSPLHNVALR	2		1	
SRP14_HUMAN	AAAAAAAAAPAAAATAPTTAATTAATAAQ	2	2	4	
SRRM2_HUMAN	IPAASAAAMNLASAR	4	1	3	1

	MAPALSGANLTSPR	1			
	TPAALAALSLTGSGTPPTAANYPSSSR	4			
STAT3_HUMAN	FICVTPTTCSNTIDLPMSPR	2	1	3	
SUGP1_HUMAN	AQTSTDAPTSAPSAPPSTPTPSAGK	3		2	
SYNJ1_HUMAN	IDPFEDLSFNLLAVSK	1			
SYNPO_HUMAN	VSTPATTTSTFSR	1	1	1	1
TAB1_HUMAN	VYPVSVPYSSAQSTSK	2	1	2	2
TAB2_HUMAN	VVVTQPNTK	1		1	
TAF4_HUMAN	QVSQAQTTVQPSATLQR	3		4	1
TAF9_HUMAN	ASIPATSAVQNVLINPSLIGSK	2		2	
TAF9B_HUMAN	LSVGAVSSKPTTPTIATPQTVSVPNK	4			
TANC1_HUMAN	MSSSTSSLTSSSSFSDGFK	1			
TCF20_HUMAN	VGQFGQHYQSSASSSSSSSFPSPQR	3			
TIF1A_HUMAN	QWQISSGQGTPSTTNSTSSTPSSPTITSAAGYDGK	2			
TNR6A_HUMAN	ASNYNVPLSSTAQSTSAR	3		1	
	LTWSPGSVTNTSLAHELWK			1	
TNR6B_HUMAN	DNTTGSNSSLNTSLPSPGAWPYSASDNSFTNVHSTSAK	1			
TOB1_HUMAN	STQPLTFTTATFAATK	1			
TOX4_HUMAN	GLQLGQTSTATIQPSQQAQIVTR	5	2	3	2
	QMLPSSITMSQGGMVTVIPATVVTSR	6	1	3	
TPD54_HUMAN	SWHDVQVSSAYVK	1			
TPR_HUMAN	GIASTSDPPTANIKPTPVVSTPSK	2			
TRI33_HUMAN	QHSNPGHAGPFPVVSVHNTTINPTSPTTATMANANR	2			
TRIP6_HUMAN	TGSLKPNPASPLPASPYGGPTPASYTTASTPAGPAFPVQVK			1	
TROAP_HUMAN	TSVSQASGLLLETPVQPAFSLPK	1			
UBAP2_HUMAN	GVSVSSSTTGLPDMTGSVYNK	4	3	1	2
	IPYQSPVSSSESAPGTIMNGHGGGR	1			
	LLQLPSTTIENISVSVHQPQPK	6		3	
UBN1_HUMAN	TPASSSSALSHPAKPHSVSSAGSSYK	1			
UBP2L_HUMAN	QAFTPSSTMMEVFLQEK	1			
	SPAVATSTAAPPPPSSPLPSK	7	3	5	3
	TAQALAQLAAQHSQSGSTTTSSWDMGSTTQSPSLVQYDLK	1			
UBQL2_HUMAN	SQNRPQGQSTQPSNAAGTNTTSASTPR	1		2	4
UTRO_HUMAN	VVLVSSASDIPVQSHR	3		1	
VCIP1_HUMAN	TEPSVFTASSSNSELIR	3	3	2	3
VEZF1_HUMAN	KTPTTVVPLISTIAGDSSR	11		3	
	TPTTVVPLISTIAGDSSR	13	4	4	3
	TSLVSTIAGILSTVTTSSSGTNPSSSASTTAMPVTQSVK	5		2	
WAC_HUMAN	INEVLTAAVTQASLQSIIHK	1			
WNK1_HUMAN	EGPVLATSSGAGVFK	3	2	3	1
	FSAPGQLCISMTSNLGGSAPISAASATSLGHFTK	3		1	2

	QPIPASSMPQQIGIPTSSLTQVVHSAGR	7	1	5	1
	TLSPEMITVTSAVGPVSMAAPTAITEAGTQPQK	11		3	
	VFPSEITDTVAASTAQSPGMNLSHSASSLSLQQAFSELR	3			
WNK3_HUMAN	QIMAPVTNSSSYSTTSVR	1	1	2	
XRN1_HUMAN	EAQSSQATPVQTSQPDSSNIVK			2	
YETS2_HUMAN	QLTTGSVVQGTLGVSTSSAQGQQTLK	2			
YTHD3_HUMAN	IGGDLTAAVTK	2		2	
	TVGTALSSSGMTSIATNSVPPVSSAAPKPTSWAAIAR	3			
ZBT20_HUMAN	SVLQQPSVNTSIGQPLPSTQLYLR	1		1	
ZC3HE_HUMAN	TSQEELLAEVVQGQSR	1			
ZCH14_HUMAN	TAQQPALVVETSTAATGTPSTVLHAARPPIK	2		2	
ZEP1_HUMAN	LEQVYNIAVTSSVGLTSPSSR	2		1	
	QVFLLSVPSLDCLPITR	2		1	
	SNGPSAALVTTSTPSALPTGEK	1	1	1	
	SNSMPTTGYSAVPANIIPPPHPLR	1			
	VNIQEQSQQPVTSLSLFNIK	2			
ZEP2_HUMAN	ALYHNPPLSMGQYLQAEPIVLGPPNLR	1	2		
	APQTLPLGLESSIPLCLPSTSDSVATLGGSK	1			
	DGLQSGSSSFSSLSPSSSQDYPSVSPSSR	2			
	EQTYPCYSGASGLHPK	2		3	1
	SESAEQQISPPNTNAK	1			
	SFDYGNLSHAPVSGAAASTVSPSR	2		1	1
	SNSVPTSSATNLTIPPSLR	2			
	STETPSEQVLQEDFASANAGSLQSLPGTVVPVR	2		3	1
	TLVTNAAMQGIGFNIAQVLGQHAGLEK	2			1
ZFHX3_HUMAN	NFQHPLVSTANLIGPGHSFYGK	1		1	
ZFR_HUMAN	AGYSQGATQYTQAQQTR	5	2	3	5
	QAAAAAAAAAATAAWTGTTFTK	4		2	
	QQEAPPPPPPATTQNYQDSYSYVR	8	3	7	2
	QYYQQPTATAAAVAAAAQPQPSVAETYYQTAPK	3	1	3	
	STPVTSAVQIPEVK	2		1	
ZHX1_HUMAN	EEIVENPSSSASESNTSTSIVNR	2		1	
	ENEIKPDREEIVENPSSSASESNTSTSIVNR	1			
	IHPSTASTVVTPAAVLPGLAQVITAVSAQQNSNLIPK	1			
ZHX3_HUMAN	EGDHSFINGAVPVSQASASSAK	2		1	
	VPEVTCIPTTATLATHPSAK	2		1	
ZN281_HUMAN	TNESQISNNINMQSYSVEMPTVSSSGGIIGTGIDELQK	1			
ZN318_HUMAN	TVVAHTSPWMPVVTTSTQTK	1			
ZN384_HUMAN	GCGLAPPHYPTLLTVPASVSLPSGISMDTESK	3			
	SDQLTPHSQASVTQNITVVPVPSTGLMTAGVSCSQR	4	1	2	
ZN507_HUMAN	NETIPDIPVSVDNLQTHTVQTASVAEMGR	2			

ZN532_HUMAN	QVTIKPVATAFLPVSAVK	1			
ZNT6_HUMAN	GTDDLNPVTSTPAKPSSPPPEFSFNTPGK	1			
ZO1_HUMAN	TPSTEAAHIMLR	1			
ZYX_HUMAN	GPPASSPAPAPK	1			
	VSSGYVPPPVATPFSSK	11	4	4	2
	Total Peptides	*Dde-1*	*Dde-2*	*PC-1*	*PC-2*
	442	407	118	241	133

5.6 Experimental Methods

Reagents and materials. All proteins, chemicals, and reagents were of analytical grade, obtained from Sigma Aldrich (St. Louis, MO), and used without further purification unless specified. The *O*-GlcNAcylated peptide TAPT(gS)TIAPG, high-capacity Neutravidin agarose resin, spin columns, and C18 desalting tips were purchased from ThermoFisher Scientific (Waltham, MA). Thiamet G was received 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.[24] Cerebral cortices were obtained from adult C57BL/6 mice bred in house. All protein concentrations were measured using a

BCA assay kit (ThermoFisher Scientific). RapiGest and UDP-GalNAz were synthesized as referenced.[35, 36] Y289L GalT was expressed and purified as described previously. Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), alkyne-Dde-biotin **5-4**, the alkyne photocleavable biotin linker (alkyne-PC-biotin, **5-1**), Cy3-azide, and azide agarose beads were purchased from Click Chemistry Tools (Scottsdale, AZ).

Peptide labeling. The labeling protocol was adapted from a previously reported method. 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, 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 labeled peptide (10 μM final) was diluted into a 400 μL solution of 10 mM sodium phosphate pH 7.6, 100 μM alkyne-Dde-biotin **5-4**, 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.

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 (Thermo Scientific, Waltham, MA). Approximately 10 pmol peptide from each sample was injected onto a CORTECS UPLC C18+ column (2.1 x 50 mm, Waters Corp., Milford, MA). Flow rate was set at 0.4 mL/min. Solvent A (ddH2O, 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.

Linker comparison by protein labeling. Labeling with Y289L GalT and UDP-GalNAz was conducted as previously described.[3] Briefly, 500 μg of 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 and precipitated by adding three volumes of methanol, one volume of chloroform, and two volumes of ddH2O with vortex mixing after each addition. Protein was 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 MnCl2. Samples were briefly mixed by pipetting and placed on ice. Next, 25 μL of 0.5 mM UDP-GalNAz 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 into two equal portions (50 μL each) to be labeled with either alkyne-Dde-biotin **5-4** or alkyne-PC-biotin **5-1**. 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 **5-4** or alkyne-PC-biotin **5-1** (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 endover-end for 1 h at RT, and the reaction was halted by the addition of 25 μL EDTA pH 8.0. Samples were

acetone 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 Dde 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 PC 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.

Coomassie staining and streptavidin blotting. Aliquots corresponding to 20 μg of protein (5 μL) for each sample were 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 (ThermoFisher Scientific, 1.0 mm, 10-well). One duplicate was then stained with Imperial protein stain (ThermoFisher Scientific) according to the manufacturer's specifications and imaged by an Odyssey scanner (LI-COR Biosciences). The other set of samples was transferred onto an Immobilon-FL PVDF membrane

(EMD Millipore, 0.45 μM) and blocked for 1 h at RT with 5% bovine serum albumin (BSA) in 1x TBST (19 mM Tris pH 7.4, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween 20). The blot was then incubated with 1:20,000 AlexaFluor-680-conjugated streptavidin (ThermoFisher Scientific) in 5% BSA/TBST for 1 h at RT, washed thrice with TBST for 5 min, and imaged with an Odyssey scanner.

α-Crystallin and sOGT labeling. Labeling was conducted as described above using 200 μg of cortical lysate, 20 μg of α-crystallin, and 5 μg of sOGT.

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.

Processing proteins 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 IAA 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.

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 $C_{19}H_{30}N_6O_{10}$ to Ser or Thr (net addition of 502.202341 Da). For CID, a scoring neutral loss of $C_{19}H_{30}N_6O_{10}$ 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, filtered for 80% peptide confidence and modifications manually evaluated.

Testing linker in on-bead digestion conditions. Cy3-Dde-biotin **5-5** was synthesized using conditions similar to protein labeling. To 15 µL 20x PBS and 259 µL ddH_2O, the following reagents were added in order: 6 µL 5 mM alkyne-Dde-biotin **5-4**, 5 µL 5 mM Cy3-azide (0.83 eq), 6 µL 100 mM NaAsc, 3 µL 10 mM THPTA, and 6 µL 50 mM $CuSO_4$. The reaction was incubated for 1 h at RT, and the crude product was used directly without purification. Eight 40 µL aliquots of the reaction were diluted to 500 µL in PBS and were added to 20 µL settled volume high-capacity Neutravidin resin. The mixture was incubated for 1 h at RT, and then the beads were washed five times with PBS to remove excess dye. An image of the samples was taken. Samples were then incubated in 500 µL buffer (PBS; 2 M urea, PBS; 0.1% SDS, PBS; 0.1% SDS, 20 mM HEPES pH 7.9) overnight at RT or 37 °C with end-over-end rotation. Beads were pelleted the following day, and a second image of the samples was taken. Because of the obvious difference between samples by eye, Cy3 released into solution was not quantified. To test if the beads were playing a role in linker cleavage, four 40 µL aliquots were diluted to 500 µL in PBS, and only two were immobilized on Neutravidin beads as described above. The beads were then rotated end-over-end overnight at RT or 37 °C. Samples that had not been incubated with beads overnight were then immobilized on Neutravidin

beads as described above to capture all linker-bound Cy3. All beads were washed twice with PBS to remove all released Cy3, and an image of the samples was taken.

Cell culture. All reagents used for cell culture were obtained from ThermoFisher Scientific unless otherwise noted. The HEK-293T cell line and HUVECs were obtained from the American Type Culture Collection. HEK-293T cells were cultured in DMEM with 10% fetal bovine serum (FBS) and 1x penicillin/streptomycin (P/S) (referred to now on as complete DMEM).

Protein digestion. All work prior to trypsinization was performed in a laminar flow hood to minimize keratin contamination. Protein lysate from HEK-293T cells in 1% SDS, 50 mM Tris pH 7.6, 150 mM NaCl, 1x PIC- was quantified using the BCA assay (ThermoFisher Scientific), and 20 mg was diluted to 2 mg/mL in lysis buffer. Samples were reduced by incubation for 5 min at 95 °C with 20 mM DTT. Two 15-mL Amicon centrifugal filter units (10 kDa NMWL, UFC901024, EMD Millipore) were centrifuged at 4,000 x *g* with 5 mL ddH_2O to wet the filters, and both the retentate and filtrate were discarded. The reduced sample was split in two (10 mg per filter unit) and then concentrated to ~1 mL by centrifugation at 4,000 x *g* at RT. All remaining centrifugation steps occur at 4,000 x *g* at RT. No time is specified since the rate at which buffer exchange occurs is not always equivalent. All samples should be concentrated to ~1 mL before proceeding to the next step. Solutions of 8 M urea, 50 mM Tris pH 7.9 and 50 mM Tris pH 7.9 were sterile filtered to remove all particulate matter. Samples were diluted to 15

mL in the filter units with 8 M urea, 50 mM Tris pH 7.9 containing 20 mM IAA and allowed to alkylate for 30 min at RT in the dark. The samples were then concentrated and buffer exchanged twice with 15 mL 8 M urea, 50 mM Tris pH 7.9 and twice with 15 mL 50 mM Tris pH 7.9. Alternative 1: The final Tris exchange step was replaced with 15 mL 50 mM Tris pH 7.9, 0.2% w/v sodium deoxycholate that had been sterile filtered. Alternative 2: Prior to the first urea exchange step, an additional exchange step was added using 15 mL 8 M urea, 50 mM Tris pH 7.9, 4% w/v sodium deoxycholate that had been sterile filtered. Samples were then diluted up to 5 mL with 50 mM Tris pH 7.9, 1 mM $CaCl_2$ in the filter units and treated with 1:50 w/w sequencing-grade trypsin (90305, ThermoFisher Scientific). Samples were rotated end-over-end overnight at 37 °C. Peptides were separated from undigested protein by centrifugation for 5 min at 4,000 x *g*, and the filter was washed twice with 2 mL 50 mM Tris pH 7.9 and centrifuged. All filtrates were combined and acidified to pH 3 with TFA. Samples were centrifuged for 5 min at 15,000 x *g* without the filter cup to pellet any residual sodium deoxycholate. Solutions were combined and desalted using a 20 cc HLB cartridge. Peptide concentrations were measured using a NanoDrop 2000C (ThermoScientific) with absorbance at 205 nm ($\varepsilon_{205}{}^{1\ mg/mL} = 31$).

HLB desalting. Oasis HLB cartridges (3 cc, WAT094226; Plus short, 186000132; 20 cc, 186000117; Waters Corp.) were first conditioned with methanol (1 mL, 5 mL, 10 mL) and then equilibrated with 0.1% TFA (2 mL, 10 mL, 20 mL) on a vacuum manifold (WAT200677, Waters Corp.). Acidified samples were then

applied to the cartridge and allowed to flow through at a rate of ~1 drop per second. Columns were then washed with 0.1% TFA (2 mL, 10 mL, 20 mL), and samples were eluted in 70% acetonitrile, 0.1% TFA (1 mL, 5 mL, 10 mL). 0.1% FA was used for the wash and elution steps for the clean up step immediately prior to MS analysis. Peptides were aliquoted as 5-mg samples and were dried by either vacuum centrifugation or freeze-drying.

GalT labeling of peptides. The peptide sample (5 mg) was resuspended in 1 mL ddH_2O. Peptides may not dissolve readily and may require sonication and vortexing. To the solution was added 1 mL 100 mM HEPES pH 7.9, 275 μL 100 mM $MnCl_2$, 50 μL 5 M NaCl, and 2.15 mL ddH_2O. The pH was checked at this point to ensure that any residual TFA from HLB elution did not acidify the solution lower than pH 7.5. Then, to the solution was added 250 μL 0.5 mM UDP-GalNAz, 250 μL 2 mg/mL Y289L GalT, and 20 μL PNGase F (P0705L, New England Biolabs). The reaction was rotated at overnight at RT. Some Y289L GalT may precipitate out, but this does not affect labeling. The following day, the reaction was acidified to pH 3 with TFA. Any precipitated protein was pelleted by centrifugation, and the solution was desalted using an HLB Plus short cartridge.

CuAAC reaction of peptides. Dried peptides were resuspended in 1 mL ddH_2O with sonication and vortexing, if necessary. A 2x CuAAC reagent mixture was produced in the following order: 860 μL ddH_2O, 40 μL 5 mM alkyne-Dde-biotin **5-4** or 5 mM alkyne-PC-biotin **5-1**, 40 μL 100 mM NaAsc, 20 μL 10 mM THPTA, and 40 μL 50 mM $CuSO_4$. 1 mL of the 2x mixture was immediately added to the

peptide solution, and the reaction was rotated end-over-end for 2 h at RT. 100 μL settled volume azide agarose resin (1038-2, Click Chemistry Tools) was washed thrice with ddH_2O, resuspended in 1 mL ddH_2O, and then added to the reaction. The bead-reaction mixture was rotated end-over-end for 1 h at RT. The reaction was filtered, and beads were rinsed twice with 1 mL ddH_2O. Samples were acidified to pH 3 with TFA and desalted using an HLB Plus short cartridge.

Peptide enrichment and elution. Dried peptides were resuspended in 1 mL ddH_2O with sonication and vortexing, if necessary. The sample was diluted to 10 mL with PBS. 0.5 mL settled volume of high-capacity Neutravidin beads were washed thrice with PBS and then added to the sample. The suspension was rotated end-over-end for 2 h at RT. The suspension was transferred to a column, and beads were rinsed with 5 mL PBS, 5 mL PBS + 2 M NaCl, and 5 mL PBS. Beads were transferred to a clean tube for elution. For **5-4**, samples were rotated for 2 h at RT with 2 mL 100 mM sodium phosphate pH 6.6, 2% w/v hydroxylamine (final pH 7.2). For **5-1**, samples were irradiated (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. Elution buffer was removed, and beads were washed twice with 1 mL 100 mM sodium phosphate pH 7.2. Washes were combined with elution buffer as the first elution fraction. The elution procedure was repeated as described above and combined as the second elution fraction. Samples were acidified to pH 3 with TFA and desalted using and HLB 6 cc cartridge. Dried samples were stored at -20 °C prior to MS analysis as described above.

5.7 References

1. J. Ma, G. W. Hart. *O*-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics*. **2014**, *11*: 8.
2. N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson. Exploring the *O*-GlcNAc proteome: direct identification of *O*-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA*. **2004**, *101*: 13132-13137.
3. P. M. Clark *et al.* Direct in-gel fluorescence detection and cellular imaging of *O*-GlcNAc-modified proteins. *J Am Chem Soc*. **2008**, *130*: 11576-11577.
4. N. Khidekel *et al.* Probing the dynamics of *O*-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol*. **2007**, *3*: 339-348.
5. J. C. Trinidad *et al.* Global identification and characterization of both *O*-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics*. **2012**, *11*: 215-229.
6. B. W. Zaro, Y.-Y. Yang, H. C. Hang, M. R. Pratt. Chemical reporters for fluorescent detection and identification of *O*-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc Natl Acad Sci USA*. **2011**, *108*: 8146-8151.
7. J. F. Alfaro *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*: 7280-7285.
8. Z. Wang *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*: 153-160.

9. N. Khidekel *et al.* A chemoenzymatic approach toward the rapid and sensitive detection of *O*-GlcNAc posttranslational modifications. *J Am Chem Soc.* **2003**, *125*: 16162-16163.

10. B. Ramakrishnan, P. K. Qasba. Structure-based design of beta 1,4-galactosyltransferase I (β4Gal-T1) with equally efficient *N*-acetylgalactosaminyltransferase activity: point mutation broadens β4Gal-T1 donor specificity. *J Biol Chem.* **2002**, *277*: 20833-20839.

11. C. S. McKay, M. G. Finn. Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol.* **2014**, *21*: 1075-1101.

12. J. Szychowski *et al.* Cleavable biotin probes for labeling of biomolecules via azide–alkyne cycloaddition. *J Am Soc Chem.* **2010**, *132*: 18351-18360.

13. C. M. Woo, A. T. Iavarone, D. R. Spiciarich, K. K. Palaniappan, C. R. Bertozzi. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact *N*- and *O*-glycopeptide discovery and analysis. *Nat Methods.* **2015**, *12*: 561-567.

14. S. A. Myers, S. Daou, E. B. Affar, A. L. Burlingame. Electron transfer dissociation (ETD): the mass spectrometric breakthrough essential for *O*-GlcNAc protein site assignments—a study of the *O*-GlcNAcylated protein host cell factor 1. *Proteomics.* **2013**. *12*: 981-991.

15. C. M. Woo *et al.* Development of IsoTaG, a chemical glycoproteomics technique for profiling intact *N*- and *O*-glycopeptides from whole cell proteomes. *J Proteome Res.* **2017**, *16*: 1706-1718.

16. M. Boyce *et al.* Metabolic cross-talk allows labeling of *O*-linked β-*N*-acetylglucosamine-modified proteins via the *N*-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci USA*. **2011**, *108*: 3141-3146.

17. B. W. Bycroft, W. C. Chan, S. R. Chhabra, N. D. Hone. A novel lysine-protecting procedure for continuous-flow solid-phase synthesis of branched peptides. *J Chem Soc Chem Commun.* **1993**: 778-779.

18. S. R. Chhabra, H. Parekh, A. N. Khan, B. W. Bycroft, B. Kellam. An appraisal of new variants of Dde amine protecting group for solid phase peptide synthesis. *Tetrahedron Lett*. **1998**, *42*: 2189-2192.

19. Y. Yang, S. H. L. Verhelst, Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem Comm*. **2013**, *49*: 5366-5368.

20. R. J. Chalkley, A. L. Burlingame. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J Am Soc Mass Spectrom*. **2001**, *12*: 1106-1113.

21. J. A. Hanover *et al.* Mitochondrial and nucleocytoplasmic isoforms of *O*-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys*. **2003**, *409*: 287-297.

22. B. D. Lazarus, D. C. Love, J. A. Hanover. Recombinant *O*-GlcNAc transferase isoforms: identification of *O*-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. *Glycobiology*. **2006**, *16*: 415-421.

23. E. P. Roquemore *et al.* Vertebrate lens alpha-crystallins are modified by *O*-linked *N*-acetylglucosamine. *J Biol Chem*. **1992**, *267*: 555-563.

24. H.-C. Tai, N. Khidekel, S. B. Ficarro, E. C. Peters, L. C. Hsieh-Wilson. Parallel identification of *O*-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc*. **2004**, *126*: 10500-10501.

25. Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu. TET2 promotes histone *O*-GlcNAcylation during gene transcription. *Nature.* **2012**, *493*: 561-564.

26. N. Zeytuni, R. Zarivach. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure.* **2012**, *20*: 397-405.

27. S. P. Iyer, G. W. Hart. Roles of the tetratricopeptide repeat domain in *O*-GlcNAc transferase targeting and protein substrate specificity. *J Biol Chem.* **2003**, *278*: 24608-24616.

28. M. E. Griffin *et al.* Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag. *Mol Biosyst.* **2016**, *12*: 1756-1759.

29. C. M. Woo, A. Felix, L. Zhang, J. E. Elias, C. R. Bertozzi. Isotope-targeted glycoproteomics (IsoTaG) analysis of sialylated *N*- and *O*-glycopeptides on an Orbitrap Fusion Tribrid using azido and alkynyl sugars. *Anal Bioanal Chem.* **2017**, *409*: 579-588.

30. K. Augustyns, W. Kraas, G. Jung. Investigation on the stability of the Dde protecting group used in peptide synthesis: migration to an unprotected lysine. *J Pept Res.* **1998**, *51*: 127-133.

31. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann. Universal sample preparation method for proteome analysis. *Nat Methods.* **2009**, *6*: 359-362.

32. J. Erde, R. R. Loo, J. A. Loo. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res.* **2014**, *13*: 1885-1895.

33. D. Pellerin, H. Gagnon, J. Dubé, F. Corbin. Amicon-adapted enhanced FASP: an in-solution digestion-based alternative sample preparation method to FASP. *F1000Research.* **2015**, *4*: 140.

34. J. Yang *et al.* Global, *in situ*, site-specific analysis of protein S-sulfenylation. *Nat Protoc.* **2015**, *10*: 1022-1037.

35. P. J. J. Lee, B. J. Compton. U. S. Patent Office. **2007**, 7,229,539.

36. H. C. Hang, C. Yu, M. R. Pratt, C. R. Bertozzi. Probing glycosyltransferase activities with the Staudinger ligation. *J Am Chem Soc*. **2004**, *126*: 6-7.