이학석사학위논문

A Multi-Method Approach to Analyzing *O*-GlcNAc Modified Peptides

융합 분석 방법론 기반 오글루넥 당화 수식화 연구

2017 년 2 월

서울대학교 대학원

분자의학 및 바이오제약학과

방 미 라

A Multi-Method Approach to Analyzing *O*-GlcNAc Modified Peptides

융합 분석 방법론 기반 오글루넥 당화 수식화 연구

지도교수 이 유 진

이 논문을 이학석사학위논문으로 제출함

2017년 1월

서울대학교 대학원

분자의학 및 바이오제약학 전공

방 미 라

방미라의 이학석사 학위논문을 인준함

2017년 1월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

A Multi-Method Approach to Analyzing *O*-GlcNAc Modified Peptides

By

Mi Ra Bang

(Directed by Eugene C. Yi, Ph.D)

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Molecular Medicine and Biopharmaceutical Sciences at the Seoul National University

January, 2017

Approved by thesis committee:

Professor _____ **Chairman**

Professor _____ **Vice chairman**

Professor _____

Abstract

A Multi-Method Approach to Analyzing *O*-GlcNAc Modified Peptides

Mi Ra Bang

Department of Molecular Medicine and Biopharmaceutical Sciences

Graduate School of Convergence Science and Technology

Seoul National University

The *O*-linked β-N-acetyl glucosamine (*O*-GlcNAc) modification, a dynamic post-translational modification (PTM) to serine or threonine residues of nuclear and cytoplasmic proteins from almost all functional classes, is involved in many different cellular processes including signal transduction, protein degradation, and regulation of gene expression. *O*-GlcNAc modification is known to be associated with several human disease states, such as diabetes, cancer, cardiovascular and neurodegenerative disorders. Despite the vital functional roles of protein *O*-GlcNAcylation in many cellular processes, the area of *O*-GlcNAc research field has been hampered, mainly due to the lack of techniques for the identification,

quantification and site mapping of *O*-GlcNAc modification in proteins. Proteomic analysis of *O*-GlcNAc modified proteins still presents significant challenges following reasons. First, site mapping of *O*-GlcNAcylation is very difficult due to its collision-induced dissociation (CID)-labile β-linkage between *O*-GlcNAc moiety and serine or threonine residues. Second, *O*-GlcNAc modified proteins are present in low stoichiometry at each site on proteins. Therefore, the research of *O*-GlcNAc modification has been limited by difficulties in mapping sites of *O*-GlcNAc modification. In this study, a new multi-method approach is introduced using enrichment and novel *N*-terminus tag based on the application of CID tandem mass spectrometry for *O*-GlcNAc proteome profiling. To compensate the substoichiometric occupancy and the lability of *O*-GlcNAc modified proteins, a new method for enrichment and detection is developed. The approach is combined with (i) lectin weak affinity chromatography (LWAC) and (ii) TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS). Collectively, with combined with LWAC and TEMPO-assisted FRIPS MS, this approach can be used in a variety of applications for *O*-GlcNAc research, all of which will provide insights into the many functions of *O*-GlcNAcylation in biological systems. The approach is not only improving enrichment for *O*-GlcNAc modified peptides but also allowing confirmation of a number of *O*-GlcNAcylation sites. We envisage that further application of the method to biological systems, which will elucidate the mechanism of specific role of protein *O*-GlcNAcylation in many biological processes. The new multi-method

provides an ideal platform to profile global *O*-GlcNAc peptides for quantitative liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis and will enable the study of functional roles of *O*-GlcNAc modified peptide in biological systems.

Keywords: *O*-linked β-N-acetyl glucosamine (*O*-GlcNAc), Lectin weak affinity chromatography (LWAC), TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS). Mass spectrometry (MS), Proteomics

Student Number: 2014-24866

Abbreviations

BEMAD ß-elimination followed by Michael addition with dithiothreitol

CID collision-induced dissociation

Con A concanavalin A

DMSO dimethyl sulfoxide

DTT dithiothreitol

ETD electron transfer dissociation

FA formic acid

FRIPS MS free radical-initiated peptide sequencing mass spectrometry

HCD higher energy collisional dissociation

IAA iodoacetamide

LC-MS/MS liquid chromatography-tandem mass spectrometry

LWAC lectin weak affinity chromatography

LTQ linear trap quadrupole

MeOH methanol

OGA *O*-GlcNAcase

***O*-GlcNAc** *O*-linked β-D-N-acetylglucosamine

OGT *O*-GlcNAc transferase

PTM post-translational modification

TEMPO 2-(2,2,6,6-tetramethyl piperidine-1-oxyl)

Table of Contents

List of Tables

List of Figures

Figure 10. Tandem Mass Spectrometry (MS/MS) spectra of *O*-GlcNAc peptides with high peptide spectrum matches (PSMs) scores and XCorr >2.5 identified from HEK 293 cells grown under high glucose conditions.

Figure 11. Tandem Mass Spectrometry (MS/MS) spectra of *O*-GlcNAc peptides with high peptide spectrum matches (PSMs) scores and XCorr >2.5 identified from HEK 293 cells grown under glucose deprivation conditions.

Introduction

While the human genome is estimated at about 25,000 genes, the human proteome size is expected to over 1 million proteins. Beyond the genomic information as a source of protein diversity, post-translational modification (PTM) of proteins further extend the range of protein functions by covalently attaching of chemical moieties to amino acid residues (Duan and Walther 2015). The diversity of proteins functions is generated by PTM that alter the structure and function and impact protein complex formation, biomolecule interactions, and enzyme catalysis (Cain, Solis et al. 2014). The biological roles of PTM are very diverse, spanning a wide spectrum, including cell-cell interactions, inter and intra trafficking, and host-pathogen recognition (Mechref 2012). Most biological functions are executed by the proteins rather than genes, study of protein PTM is crucial to gain further insights into complex biological systems. Furthermore, defects in PTM also have been linked to various developmental disorders and human diseases, highlighting the importance of PTM in maintaining normal cellular states (Wang, Peterson et al. 2014). Although many proteins undergo PTM, the changes in many function and character of the proteins as a result of this various modification are only starting to be understood working biological systems (Kudlow 2006). The comprehensive characterization of PTM still the analytical challenges due to inherent nature of low abundance. Therefore, the qualitative and quantitative evaluation of PTM is

fundamental for understanding the mechanisms of many biological events.

The modification of glycosylation proteins is the most common PTM in human and other species. Protein glycosylation associated with structural, stabilizing, and signaling roles in biological processes. The aberrant glycosylation is accompanied with oncogenesis and tumor progression, and it clinically used for cancer biomarkers (Huang, Wang et al. 2014). In mammalian cells, proteins are modified with a variety of glycans, which can be classified into two major groups, *N*- and *O*-glycans (Shirato, Nakajima et al. 2011). Unlike the complex glycans, the *O*-linked β-N-acetyl glucosamine (*O*-GlcNAc) modification is a seine or threonine linked monosaccharide (Figure 1). *O*-GlcNAcylation correlates with important enzyme regulation and disease relevant signaling pathways (Bond and Hanover 2015). *O*-GlcNAcylation occurs on nuclear and cytoplasmic proteins regulated by two enzymes: *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA), which add and eliminate the *O*-GlcNAc moiety rapidly. *O*-GlcNAc cycles on proteins with a timescale similar to protein phosphorylation at seine or threonine residues. It has an extensive cross talk with phosphorylation, where it serves as a nutrient and stress sensor to modulate signaling, transcription, and cytoskeletal functions (Hart, Slawson et al. 2011). *O*-GlcNAc modification is subject to an additional level of metabolic control through hexosamine signaling pathway (HSP). In the HSP, levels of uridine 5′-diphosphate (UDP)-GlcNAc respond to nutrient excess to activate *O*-GlcNAc modification with OGT (Love and Hanover 2005). *O*-GlcNAc, as a major nutrient sensor, underlies the progress of several metabolic diseases including

diabetes. Moreover, protein *O*-GlcNAcylation has been involved in cancer, Alzheimer's disease, and cardiovascular disorders. Recently, *O*-GlcNAc levels have been reported to be escalated in response to cellular injury in both *in vivo* and *in vitro* models of oxidative stress, heat stress, and ischemia reperfusion injury (Lima, Spitler et al. 2012). Therefore, sequencing of *O*-GlcNAc plays a significant role in identification of characteristic functions in many biological systems. As with any PTM, *O*-GlcNAc site mapping is a prerequisite for understanding the biological roles of the modifications.

Despite the vital functional roles of protein *O*-GlcNAcylation in many cellular processes, the area of *O*-GlcNAc research field has been hampered, mainly due to the lack of techniques for the identification, site mapping, and quantification of *O*-GlcNAc modification in proteins. There are some significant challenges of proteomic analysis of *O*-GlcNAc. First, collision-induced dissociation (CID) causes lability of β-linkage between *O*-GlcNAc moiety and seine or threonine residues. The extremely labile nature of *O*-GlcNAc modified protein makes difficult to identify modification sites by CID and higher energy collisional dissociation (HCD). Especially, deciphering the *O*-GlcNAc modification of proteins using proteomic approach remains to be an analytically challenging task because of the poor fragmentation of *O*-GlcNAc peptides. The concurrent *O*-GlcNAc peptide and site identification are difficult in CID conditions because peptides readily lose the GlcNAc moiety, which result in spectra containing neutral loss species along with the GlcNAc oxonium ions (m/z 204.08) and poor fragments

pattern (Chalkley and Burlingame 2001). Not only that, a series of fragments of the *O*-GlcNAc oxonium ions (i.e., m/z 186.07, m/z 168.06, m/z 144.06, m/z 138.05, and m/z 126.05) can also be generated at CID and higher-energy collisional dissociation (HCD) mode (Figure 2). CID methods are most widely employed in the analysis of peptides sequence. Therefore, stable CID spectra are necessary in order to accurately detect which site is *O*-GlcNAc modification (Ma and Hart 2014). Second, *O*-GlcNAc has substoichiometric ratio at each site on proteins and low abundance. Although some proteins may be abundantly *O*-GlcNAc modified the frequency in the whole proteome has been estimated to be far below that of phosphorylated proteins (Wang, Yang et al. 2014). To resolve the sub-stoichiometric occupancy and liability of *O*-GlcNAc modification, many methods have been developed for enrichment and detection, such as chemoenzymatic approach, tagging-via-substrate (TAS) method. But these methods are still not suitable for *O*-GlcNAc sites mapping for high throughput by direct CID mass spectrometry (Cao, Cao et al. 2013). Third, there are many forms of *O*-glycosylation, including *O*-GalNAc *O*-Mannose, *O*-Fucose. The other *O*-glycosylation species share much similarity with *O*-GlcNAc in that they can all undergo ß-elimination followed by Michael addition with dithiothreitol (BEMAD). Although BEMAD method has been well adopted for mass spectrometry-based *O*-GlcNAc site mapping, there are few weaknesses of the BEMAD method that it is indirect due to any moiety at a serine or threonine residues, which could result in a false assignment as an *O*-GlcNAc site and potential sample loss by multi-step

method. Therefore, a method development regarding *O*-GlcNAc profiling is needed to decipher their dynamic functions.

In this study, a new multi-method approach for proteomic *O*-GlcNAc profiling in CID is developed (Figure 3). After protein digestion with trypsin, *O*-GlcNAc peptides are enriched from the tryptic peptide mixture by lectin chromatography. Wheat germ agglutinin (WGA) is a lectin which specifically binds to GlcNAc and sialic acid and WGA achieves high affinity interaction with complex carbohydrates due to four carbohydrate binding sites. However, *O*-GlcNAc interaction with WGA is quite weak and immune-precipitation protocol including binding, washing, and elution have not been useful for isolating *O*-GlcNAc modified peptides. Therefore, we developed improved approach and implemented to enrich *O*-GlcNAcylated proteins using WGA based lectin weak affinity chromatography (LWAC). The strategy is coupled to high-performance liquid chromatography (HPLC) for efficiency of *O*-GlcNAc interaction with WGA. The LWAC method can enrich natively modified *O*-GlcNAc peptides without chemical modification. Therefore, the LWAC method essential for targeting *O*-GlcNAc peptides especially further steps for site mapping by spectrometry analysis. Next, *O*-GlcNAc peptides analyzed using TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS) (Figure 4). *O*-GlcNAc modification site mapping proved information as a result of neutral loss of the glycan under the conditions of CID. Direct site mapping, however, often results in the fact that the *O*-GlcNAc modification is labile during the standard mass

spectrometric readout of CID. Therefore, the modification site information is usually lost. To overcome the analytical challenges, electron transfer dissociation (ETD) mass spectrometry has been introduced but the technique also has shortcomings and notably poor overall sensitivity. Therefore, alternative methods for the site mapping have been developed in CID (Vocadlo, Hang et al. 2003). For example, several researchers have developed metabolic labeling of *O*-GlcNAc proteins by azide or alkyne-tagged Nacetylglucosamine and subsequently conjugated the modified proteins to an affinity probe using copper-catalyzed azide/alkyne Click chemistry (Sprung, Nandi et al. 2005), (Nandi, Sprung et al. 2006), (Zaro, Yang et al. 2011). However, these approaches did not enable the global profiling and site mapping for *O*-GlcNAc in CID due to its labile and large mass of tag attached to the *O*-GlcNAc. Another methods, nucleophilic replacement of the *O*-GlcNAc via ß-elimination followed by Michael addition with dithiothreitol (BEMAD) based enrichment method, has been shown to generate fragmentation at CID spectra contain more stable information than their native form (Gallop and Paz 1975). Previously, proteomic analysis approach has been able to identify the sites of *O*-GlcNAc modification of peptides after under β-elimination conditions of the sugar residue. But it is cleaved to leave an unmodified serine or threonine residues, leaving no residual indication that the residues are modified. To reduce the adverse BEMAD effects of the imitations associated with mass analysis of *O*-GlcNAc, the new method is attach of chemical tag to *O*-GlcNAc peptides for labile their native form. In TEMPO-assisted FRIPS MS, *p*-

TEMPO-Bz-Sc-NHS as radical precursor, is coupled with the *N*-terminus of the *O*-GlcNAc peptides and the *p*-TEMPO-Bz-Sc-NHS conjugated with peptides are results in homolytic cleavage. In this method, neutral loss of the TEMPO moiety produces a benzylic radical ion, which in turn induces peptide backbone fragmentations. A coupling of TEMPO reagents to the *N*-terminus of a *O*-GlcNAc peptide introduces a labile oxygen–carbon bond that can be selectively activated at CID mode to produce a radical ion as well as b and y ion (Marshall, Hansen et al. 2014). The *N*-terminus chemical tag approach provides radical generated fragmentation by soft ionization that retains the modification sites on peptides with a minimal sample loss by the single step process. Therefore, the *N*-terminus chemical tag was used for increase of confidence to identify *O*-GlcNAc modification site mapping. Quantification and sites mapping techniques coupling with mass spectrometry for *O*-GlcNAc profiling from a proteomics view, which should offer a systems perspective for the function of *O*-GlcNAc modification on multiple proteins in physiology and diseases. The multi-method platform implemented in this investigation is expected to be of general use and to facilitate more accurate identification of many *O*-GlcNAc modification protein involved in various biological processes.

The new method was developed and applied to biological samples using combined enrichment and *N*-terminus tag for mass spectrometry methods to perform global exploration of *O*-GlcNAc modification proteins specific in HEK 293 cells. Thus, for the first time, the new study definitely not only identified but also mapped that

some proteins were *O*-GlcNAc modification in global profiling. The new analytical platform could serve as a valuable multi-method to facilitate understanding of modification in *O*-GlcNAc research field. The new method will significantly advance in the overall analytical platform of *O*-GlcNAc analysis for enrichment and site mapping. Collectively, with combined with lectin affinity chromatography and TEMPO-assisted FRIPS MS, this method can be used in a variety of applications for *O*-GlcNAc research, all of which will accelerate the study *O*-GlcNAc modified proteins and provide insights into the many functions of *O*-GlcNAcylation.

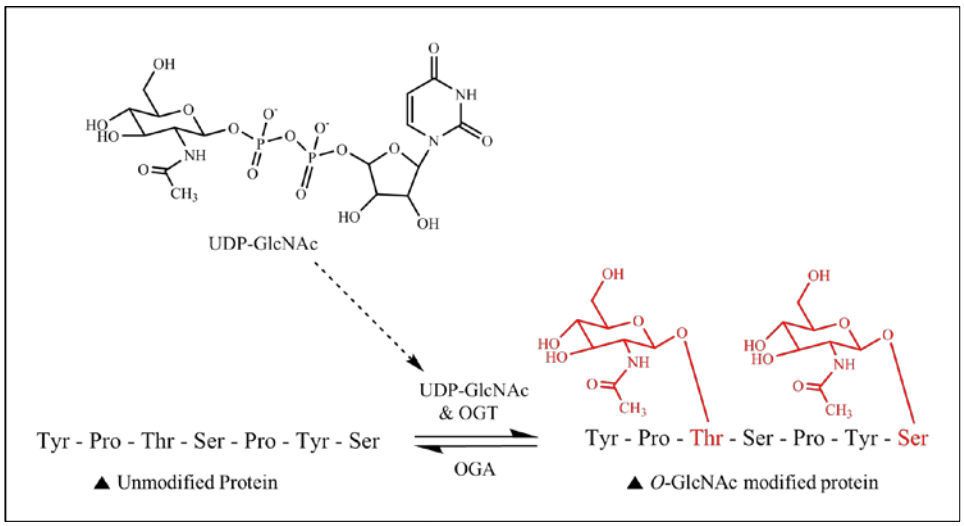

Figure 1. *O*-GlcNAcylation of serine or threonine residues in proteins.

O-GlcNAc occurs on nuclear and cytoplasmic proteins regulated by two enzymes: *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA), which add and eliminate the *O*-GlcNAc moiety rapidly. UDP-GlcNAc respond to nutrient excess to activate *O*-GlcNAc modification with OGT.

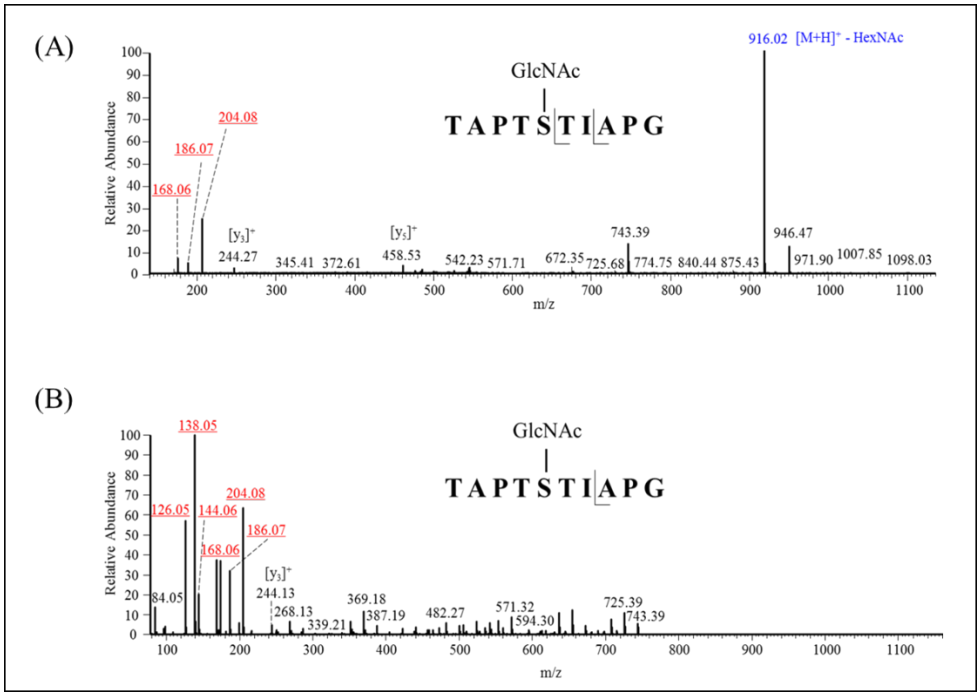

Figure 2. Lability of β-linkage between *O*-GlcNAc moiety and serine residue in *O*-GlcNAc standard peptide for modification site mapping in collision-induced dissociation (CID) (A) and higher energy collisional dissociation (HCD) (B).

Tandem mass spectrometry (MS/MS) spectra of CID (A) and HCD (B) of *O*-GlcNAc modified standard peptide (TAPTsTIAPG, where s is the *O*-GlcNAc modification site). The *O*-GlcNAc peptide liberated an intense neutral loss peak (m/z, 916.02) and loss of GlcNAc with a few b and y ions upon MS/MS fragmentation which appears to be not sufficient enough to identify the peptide and difficult to assign the modification site by the both CID and HCD. A series of intense *O*-GlcNAc oxonium ions (i.e., *m/z* 126.05, *m/z* 138.05, *m/z* 144.06, *m/z* 168.06, *m/z* 186.07, and *m/z* 204.08) were generated.

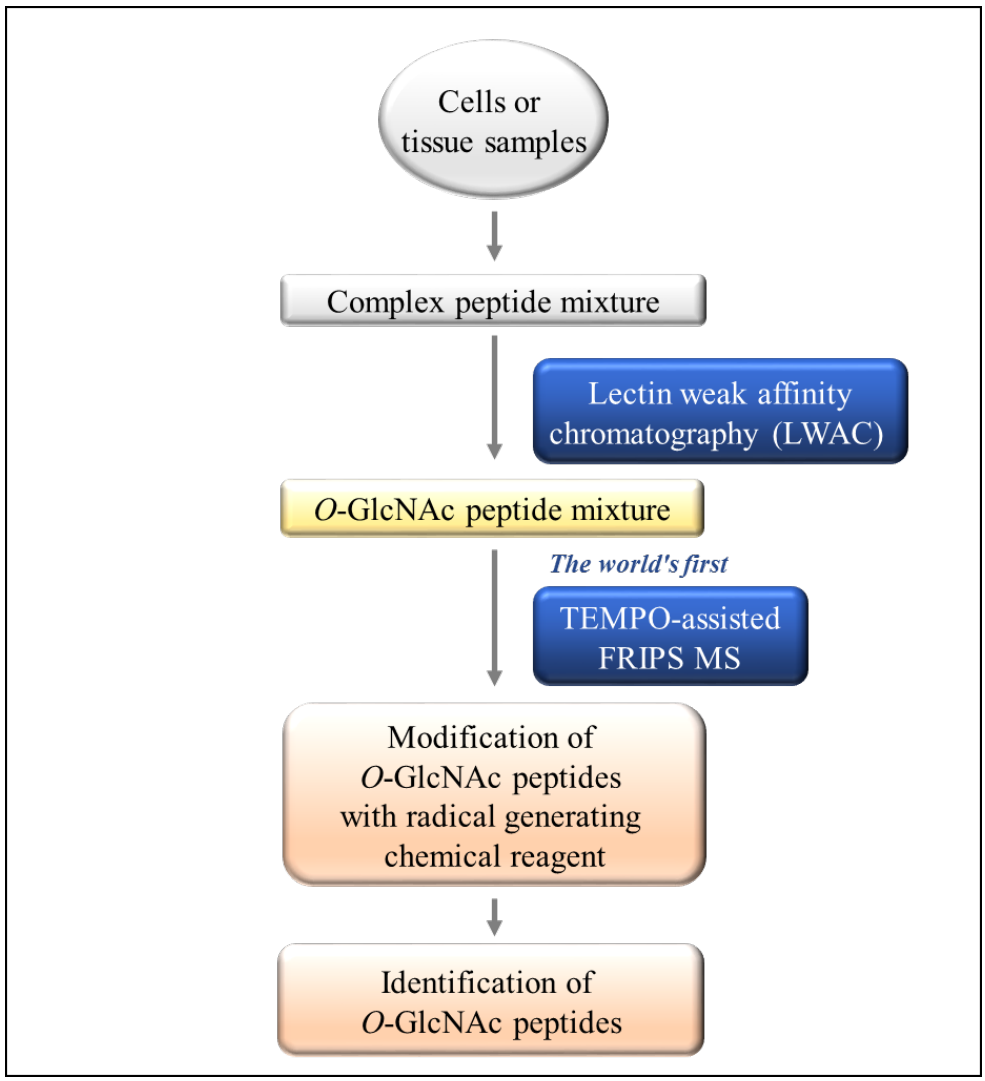

Figure 3. A multi-method approach for identification of *O*-GlcNAc peptides. *O*-GlcNAc peptide mixture is enriched from the complex peptide sample using lectin chromatography. Next, *O*-GlcNAc site mapping is carried out using TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS).

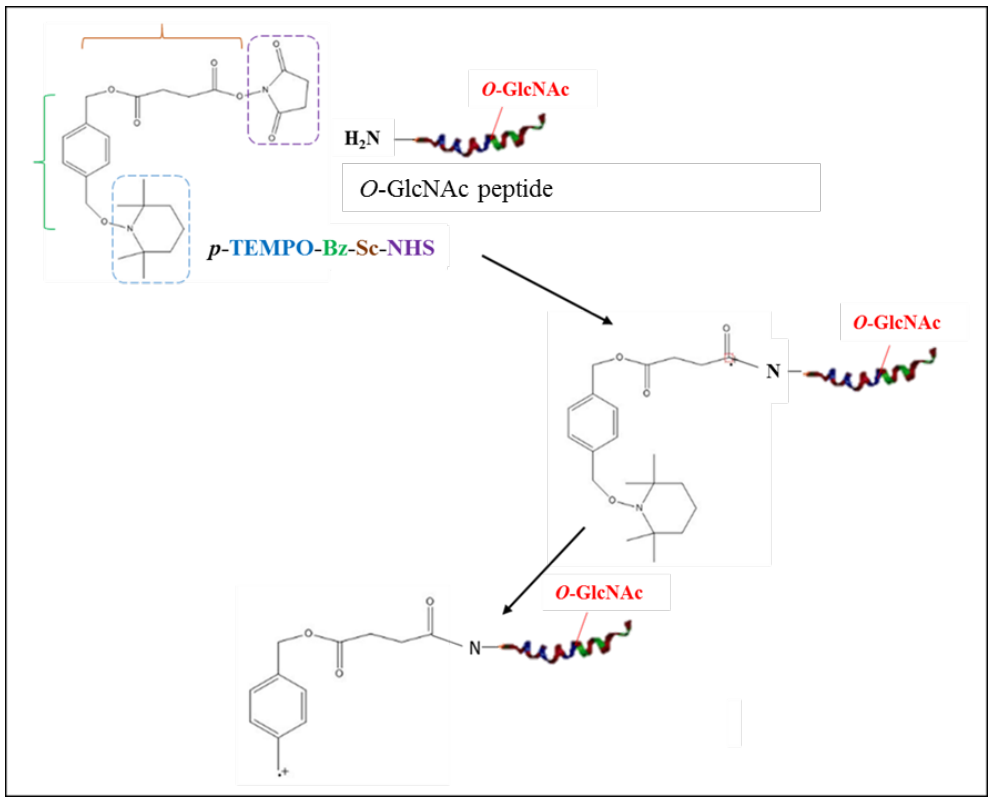

Figure 4. A mechanism of TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS) to analysis of *O*-GlcNAc peptides under collision-induced dissociation (CID).

The coupling TEMPO reagent to the *N*-terminus of O-GlcNAcylated peptides introduces a labile oxygen–carbon bond that can be selectively activated at a conventional CID mode to produce a benzylic radical ion which initiates a radical driven peptide fragmentation. The method offers a minimal loss by the single step chemical conjugation reaction with a primary amine-reactive NHS functional group of the reagent.

Materials and Methods

1. LWAC enrichment

For optimization of LWAC method, *O*-GlcNAc modified synthetic peptide (TAPTsTIAPG, where ‘s’ represents the *O*-GlcNAc modified serine) by commercially obtained (Thermo Fisher Scientific, Waltham, MA, USA) spiked in bovine serum albumin (BSA) peptides (Thermo Fisher Scientific). The ratio of *O*-GlcNAc modified synthetic peptide and BSA was 1:100. The mixture of *O*-GlcNAc modified synthetic peptide and BSA peptides dried and suspended in 40 ul of WGA buffer (50 mM sodium acetate (NaOAc) (Sigma-Aldrich, St. Louis, MO, USA), 0.2 M sodium chloride (NaCl) (Sigma-Aldrich) with 1 mM calcium chloride ($CaCl_2$,) (Sigma-Aldrich) 1 mM magnesium chloride ($MgCl_2$) (Sigma-Aldrich)) and loaded into the ProSwift™ (5 x 50 mm) con A-1S LC columns (Thermo Fisher Scientific) coupled with Ultimate 3000 standard LC system (Thermo Fisher Scientific). *O*-GlcNAc modified peptide was enriched by an isocratic 100 % WGA buffer at a flow rate of 0.5 ml/min, and 90 sec fractions were collected. During WGA chromatography, the pressure was never allowed to exceed 50 bar. For storage and reuse of WGA columns, WGA buffer containing 20 mM GlcNAc to stabilize the lectin (or recommend using 0.08 % sodium azide). And collected samples were acidified with addition formic acid (FA, CH_2O_2) (Sigma-Aldrich),

and collected peptides were cleaned up using C18 HARVARD desalting columns (Harvard Apparatus, Holliston, MA, USA), eluted in 0.5 % FA, 50 % acetonitrile (ACN) and 0.5 % FA, 80 % ACN, and dried using a Centrivap speed vacuum concentrator (Labconco, Kansas City, MO, USA) for subsequent analysis. For LC-MS/MS, the sample was dissolved in 0.1 % FA, and 1⁄5 of the sample was analyzed. All steps for enrichment were same to biological sample 1 mg of HEK 293 cells under high glucose and glucose deprivation conditions.

2. TEMPO-assisted FRIPS MS

For optimization of TEMPO-assisted FRIPS MS method, the conjugation of a *O*-GlcNAc standard peptide with TEMPO was performed with dimethyl sulfoxide (DMSO) solvent (Sigma-Aldrich) (Lee, Kang et al. 2009). 1 mg of *p*-TEMPO reagents dissolved in 100 ul of DMSO solvent. Because of the hydrophobic character of TEMPO reagents, TEMPO reagents was not soluble in water. A TEMPO solution was added to DMSO and 100 mM of tetraethylammonium bromide (TEAB) (Sigma-Aldrich). A dried *O*-GlcNAc peptide was mixed with final solution and then vortex for 1 min. The mixed solution was reacted at 37°C for overnight. The final *O*-GlcNAc peptide conjugated with TEMPO reagent was purified using C18 HARVARD columns (Harvard Apparatus) and eluted in 0.5 % FA, 50 % ACN and 0.5 % FA, 80 % ACN. The extracted peptides were dried using

Centrivap speed vacuum concentrator and safely stored at -20 ℃ until the LC-MS/MS analysis. All steps for enrichment were same to biological sample 1 mg of HEK 293 cells under high glucose and glucose deprivation conditions.

3. Application

Proteins were extracted from HEK 293 cells grown under high glucose and glucose deprivation. Cells were washed in Dulbecco's phosphate-buffered saline (PBS) (Thermo Fisher Scientific) and repeated 3 times, lysed in mammalian protein extraction reagent (M-PER) (Thermo Fisher Scientific). M-PER reagent used for extraction of cytoplasmic and nuclear protein. And then samples were centrifuged in 14,000 g at 37°C using micro centrifuge (Eppendorf, Hamburg, GER) and supernatants were isolated. Total protein concentration was measured using Pierce™ BCA protein assay kit (Thermo Fisher Scientific). The in-solution tryptic digestion was conducted following the referred protocol (Leon, Schwammle et al. 2013). Proteins were denatured with 6 M urea, (Sigma-Aldrich) in 50 mM Tris-HCI, pH 8.0 (Thermo Fisher Scientific) at 37 °C for an hour and reduced by 10 mM dithiothreitol (DTT) in 50 mM Tris-HCI, pH 8.0 at 37 °C for an hour. The samples were then alkylated with with freshly made 30 mM iodoacetamide (IAA) (Sigma-Aldrich) in 50 mM Tris-HCI, pH 8.0 in the dark for 45 minutes. The samples were then diluted to 1 M urea with 50 mM Tris-HCI, and endo proteinase

lys-C sequencing grade (Promega, Fitchburg, WI, USA) was added at a ratio of 1:100 (lys-C : protein), followed by incubation 2 hr at 37 °C. Next, Pierce trypsin protease MS grade (Thermo Fisher Scientific) was added at a ratio of 1:50 (trypsin: protein), followed by incubation overnight at 37 °C. Tryptic digested peptides were desalting with Oasis mixed-mode ion-exchange (MCX) cartridge (Waters, Milford, MA, USA), eluted in 5% ammonium hydroxide (NH_4OH) (Sigma-Aldrich) in 90% MeOH, and dried in a Centrivap speed vacuum concentrator (Labconco).

4. LC-MS/MS analysis

Peptides were resuspended in 0.1 % (v/v) FA and separated on an in-house packed with ReproSil-Pur 200 C18-AQ (Dr. Maisch GmbH, Ammerbuch, GER) to 12 cm of flexible fused silica capillary tubing, inner diameter 75 µm, outer diameter 375 µm (Polymicro Technologies, AZ, USA) by elution with a linear gradient of 2 – 35 % solvent B (0.1 % FA in 98 % ACN) in solvent A (0.1 % FA in HPLC grade water). In method optimization, sample analyzed for 50 min at a flow rate of 300 nl/min and HEK 293 cells were analyzed for 400 min at a flow rate of 300 nl/min. Samples were analyzed a linear trap quadrupole (LTQ)-velos (Thermo Finnigan, San Jose, CA, USA) in data-dependent mode coupled with a nano-ultra performance liquid chromatography system Easy nLC system (Proxeon Biosystems,

Odense, Denmark). The full scans were acquired in the mass analyzer at 400 − 1400 m/z by using a normalized collision energy of 35- % for collisional energy dissociation fragmentation. The spray voltage was set to 2.1 kV, and the temperature of the heated capillary was set to 325 °C. The LTQ-Velos was operated in data-dependent mode with one survey full MS scan followed by 5 MS/MS scan of the most abundant ions.

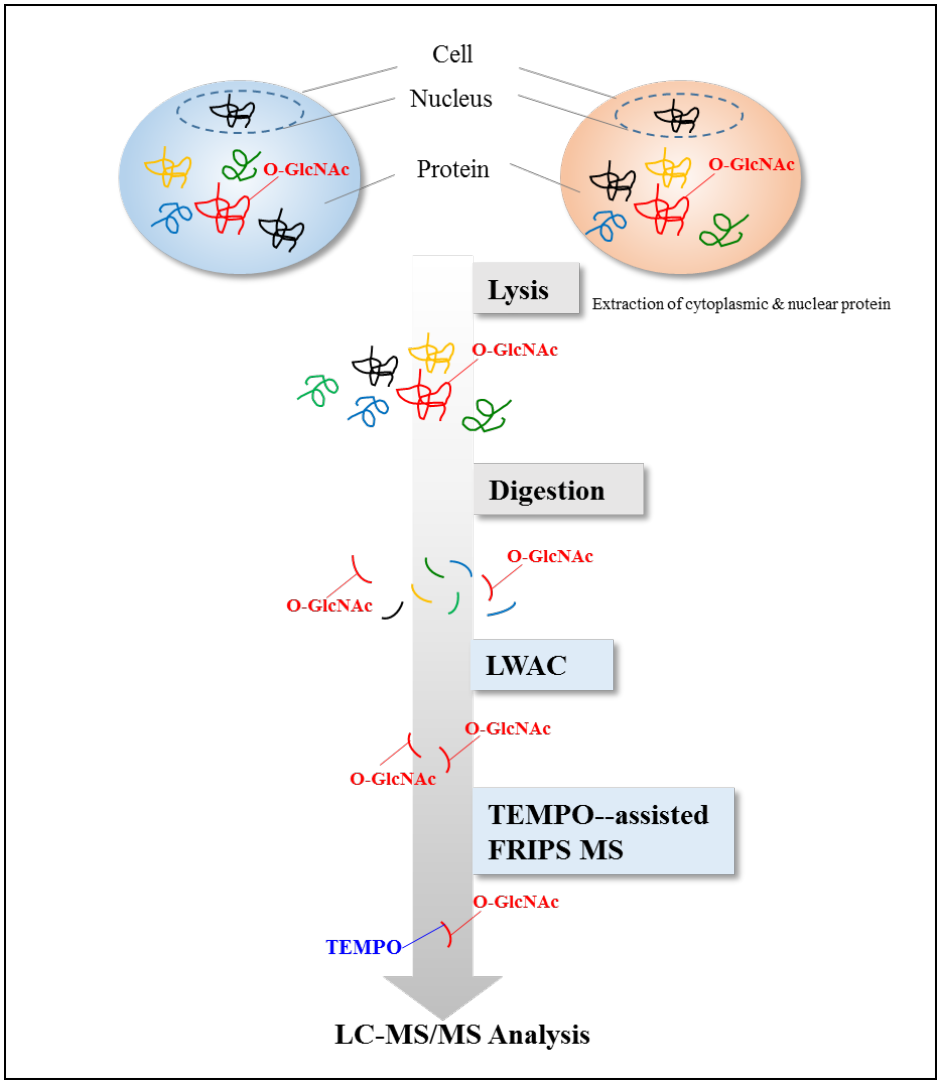

Figure 5. Experimental workflow for identification of *O*-GlcNAc peptides of HEK 293 cells under high glucose and glucose deprivation conditions.

Proteins are obtained using M-PER reagent for extraction of cytoplasmic and nuclear proteins and subjected to in-solution tryptic digestion followed by lectin chromatography. Enriched peptides are labeled by TEMPO-assisted FRIPS MS followed by LC-MS/MS analysis and database search.

5. Database search

Raw data files generated from HEK 293 cells grown under high glucose and glucose deprivation and were combined for protein database search and analyzed in Proteome Discoverer (v1.4) interfaced SEQUEST algorithm (Thermo Fisher Scientific). The database search was performed by database of uniprot *homo sapience* (Taxon identifier: 9606) and allowed for a mass TEMPO reagent tag with 360.466 Da (*N*-term), radical tag with 204.221 Da (*N*-term) and HexNAc (serine or threonine residues) with 204. Trypsin was selected for the enzyme and up to two missed cleavage sites were allowed during the database search. Peptides and proteins identification were filtered with charge state dependent cross correlation (Xcorr) ≥ 2.0 and high confidence with less than 1 % false discovery rate.

Results

1. Experimental strategy

O-GlcNAcylation is associated with many human diseases such as cancer, diabetes, and neurodegeneration. Therefore, understanding the variety roles in biological systems of *O*-GlcNAcylation will require more comprehensive characterization of the *O*-GlcNAc peptides. Despite the vital functional roles of *O*-GlcNAcylation in many cellular processes, *O*-GlcNAc research field has been turned over, due to the lack of techniques for the identification, quantification and site mapping of *O*-GlcNAc modification in proteins. Here, an improved multi-method approach has been developed to enrich and quantitatively identify using the *N*-term tag labeling *O*-GlcNAc for site identification (Figure 6). The approach is combined with (i) lectin-based *O*-GlcNAc enrichment and (ii) TEMPO-assisted FRIPS. The *N*-term chemical tag approach provides radical generated fragmentation by soft ionization that retains the modification sites with a minimal sample loss by single step method. TEMPO-assisted FRIPS MS is the new method for *O*-GlcNAc peptide identification that can be alternative to the ETD method.

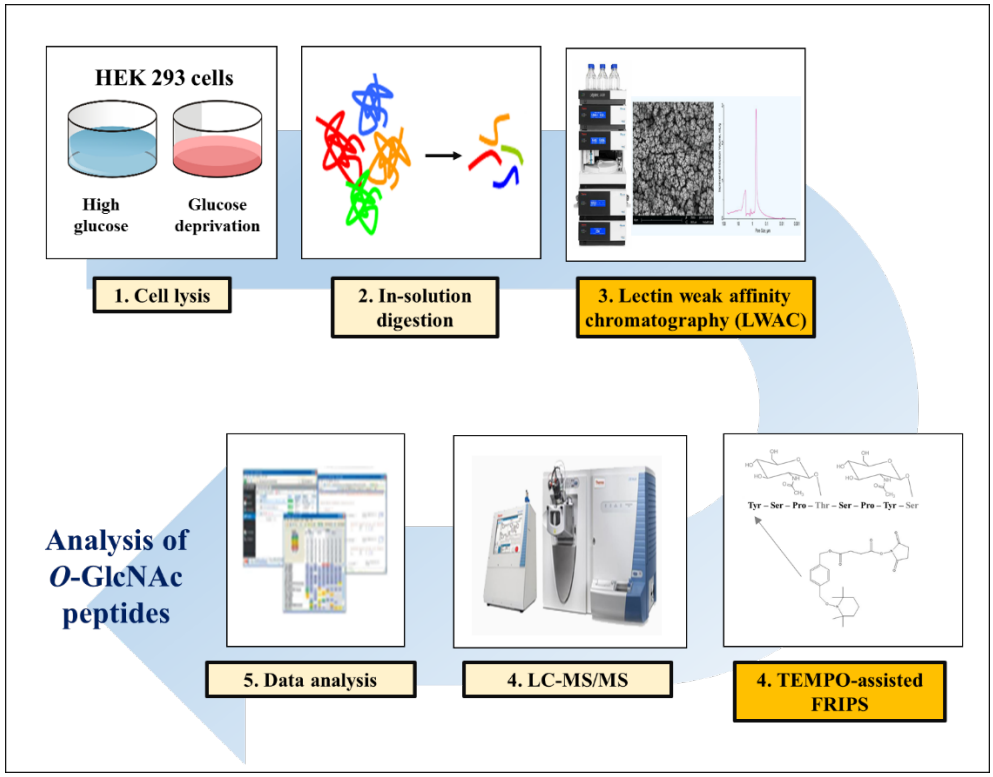

Figure 6. Experimental strategy.

Tryptic digested peptides mixtures were loaded into the lectin column for *O*-GlcNAcylated peptide enrichment followed by the TEMPO *N*-term labeling, mass spectrometry analysis, and database search.

2. O-GlcNAc modified peptides enrichment by LWAC

Enrichment of *O*-GlcNAc peptides was performed using a WGA column chromatography following the manufacture protocol with a slight modification. To optimize the *O*-GlcNAc peptide enrichment, a synthetic *O*-GlcNAc peptide spiked in a BSA peptide mixture was injected onto a ProSwift™ (5 x 50 mm) con A-1S LC columns with an isocratic 100 % of WGA buffer (50 mM sodium acetate (NaOAc) (Sigma-Aldrich), 0.2 M sodium chloride (NaCl) (Sigma-Aldrich) with 1 mM calcium chloride ($CaCl_2$,) (Sigma-Aldrich) 1 mM magnesium chloride ($MgCl_2$) (Sigma-Aldrich)) at a flow rate of 0.5 mL/min and monitored at 214 nm HPLC system. After eluting the major peak, fractions between 1.8 and 3.6 min were collated (Figure 7-A) and analyzed by LC-MS/MS (Figure 7-B). MS/MS spectrum of the standard *O*-GlcNAc peptide (m/z 559.78) extracted from the LC-MS/MS chromatographic trace showing a series of *O*-GlcNAc oxonium ion (i.e., m/z 126.05, m/z 138.05, m/z 144.06, m/z 186.07, m/z 168.06, and m/z 204.08). A prominent ion at m/z 204.08 corresponds to the GlcNAc oxonium ion generated by energetic dissociation of GlcNAc from the peptides (Figure 7-C).

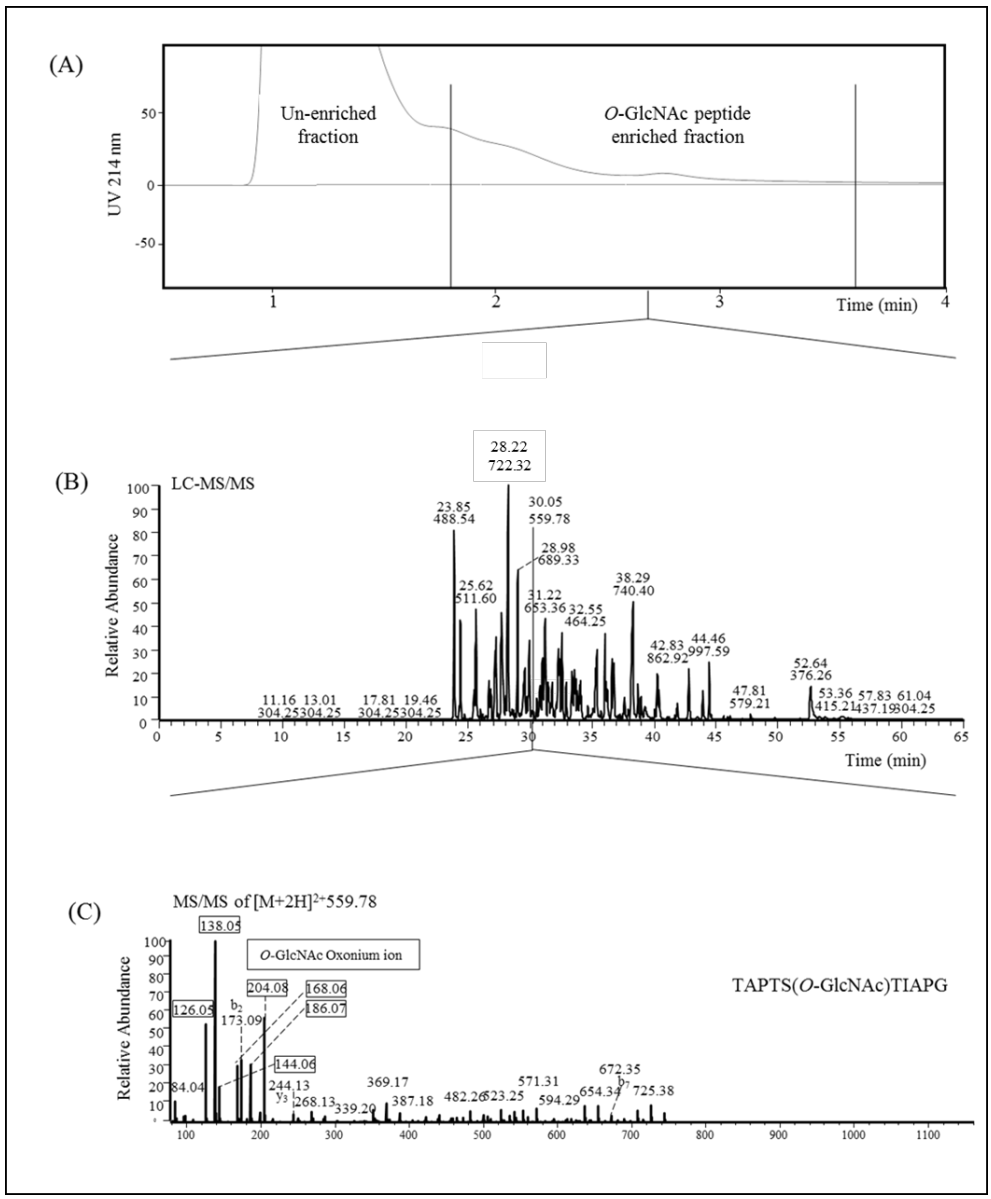

Figure 7. Lectin weak affinity chromatography (LWAC) enrichment of *O*-GlcNAc modified peptide by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

The UV peak marked 'LWGA-enriched' was analyzed by LC-MS/MS. Lectin enrichment of *O*-GlcNAc peptides (TAPTsTIAPG, where 's' represents the *O*-

GlcNAc modified serine) spiked with bovine serum albumin (BSA) peptides. A chromatographic trace of peptide elution with an isocratic 100 % buffer at a flow rate of 0.5 mL/min and monitored at 214 nm (A). LC-MS/MS trace of peptides collected from the fractions between 1.8 and 3.6 min (B). MS/MS spectrum of the standard *O*-GlcNAc peptide (m/z 559.78) extracted from the LC-MS/MS chromatographic trace showing a series of *O*-GlcNAc oxonium ion (i.e., m/z 126.05, m/z 138.05, m/z 144.06, m/z 186.07, m/z 168.06, and m/z 204.08) (C).

3. Identification of O-GlcNAc peptides by TEMPO-assisted FRIPS MS

TEMPO modified peptide generates singly and multiply charged ions by electrospray ionization. CID spectra showed the isolation and activation of $[M + nH]^{n+}$ ions (where n = 1, 2) of *O*-GlcNAc peptides-TEMPO generating 4 types of product ions (Figure 8); an intermediate form of *O*-GlcNAc standard peptide labeled by TEMPO reagent (Figure 8-A), and loss of HexNAc from intermediate form (Figure 8-B). An radical form of *O*-GlcNAc standard peptide (Figure 8-C) and and loss of HexNAc from radical form (Figure 8-D) were obserevd. Most fragmentation ions of TEMPO-assisted FRIPS mass spectrometry produced via radical-driven fragmentation pathway: dominant a/x- and c/z- type ions. *N*-term chemical tag approach using TEMPO provides radical-derived fragmentations by soft ionization that remained the *O*-GlcNAc moiety. TEMPO modified *O*-GlcNAc peptide provides a radical ion as well as b and y ion from the native peptide. To test our method, we identified that *O*-GlcNAc standard peptide conjugated with TEMPO reagent. A *p*-TEMPO-Bz-Sc conjugated peptide was subjected to collisional activation as reagent form in the platform of CID mass spectrometer (Figure 9-A, B). The data show the resulting collisional activation mass spectra of *p*-TEMPO-Bz-Sc-GlcNAc standard peptide (TAPTsTIAPG,‘s’ is modified *O*-GlcNAc) obtained at the collision energy. The major product ions were found to be

radical-adducted peptides as a result of the homolytic cleavage of the bond between the oxygen and the benzylic carbon of the TEMPO moiety (Figure 9-B, C).

In the previous TEMPO-assisted FRIPS MS studies, it was shown that two-step (MS^3) of collisional activation were necessary to obtain radical-driven peptide fragmentations. The two-step collisional activation processes needs to be addressed in the TEMPO-assisted FRIPS MS. However, the sensitivity of MS^3 is generally lower than that of MS^2 and the duty-cycle which is the percentage of time that a particular ion is getting through the quadrupole and reaching the detector is lower (Canas, Lopez-Ferrer et al. 2006). In this TEMPO-assisted FRIPS MS for site mapping of *O*-GlcNAc, CID which is mass spectrometers are known to be equipped only for a single-stage tandem MS, interestingly, one-step peptide backbone fragmentations are possible for analysis of *O*-GlcNAc modified peptide. This results indicate that the TEMPO-assisted FRIPS approach appears to become useful, particularly in CID for *O*-GlcNAc sites mapping.

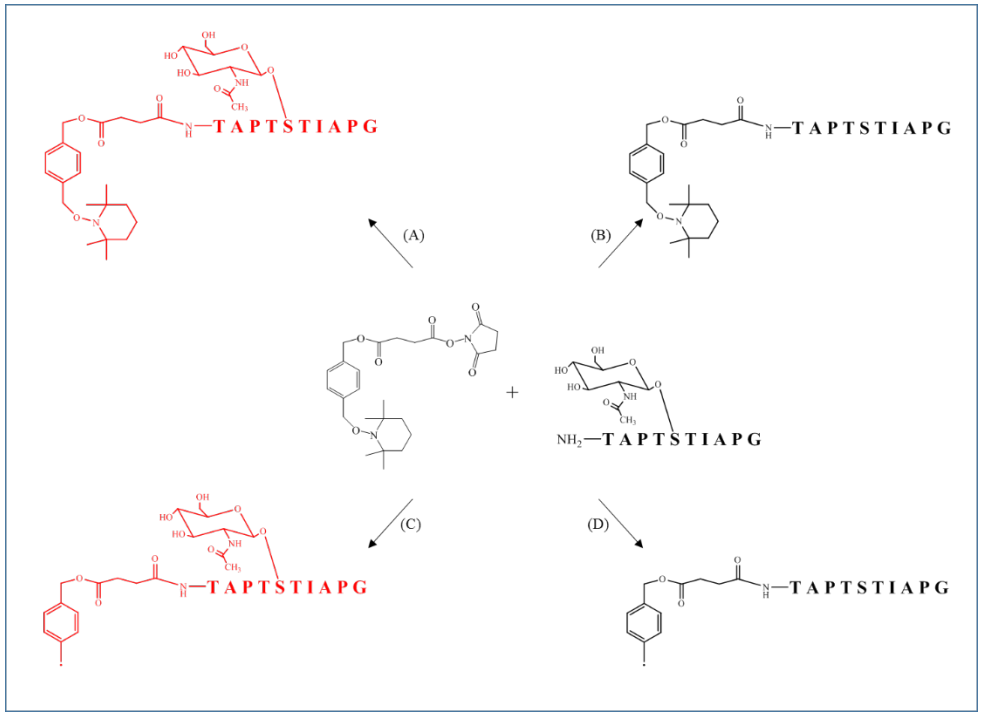

Figure 8. Suggested reaction forms of free radical precursor bound to *O*-GlcNAc standard peptide through the *N*-terminus.

Intermediate form of *O*-GlcNAc standard peptide (TAPTsTIAPG, where 's' represents the *O*-GlcNAc modified serine) labeled by TEMPO reagent (A), and loss of HexNAc from intermediate form (B). Radical form of *O*-GlcNAc standard peptide followed radical reaction labeled by TEMPO reagent (C), and loss of HexNAc from radical form (D). Most fragmentation ions of TEMPO-assisted FRIPS mass spectrometry produced via radical driven fragmentation pathway: dominant a/x- and c/z- type ions.

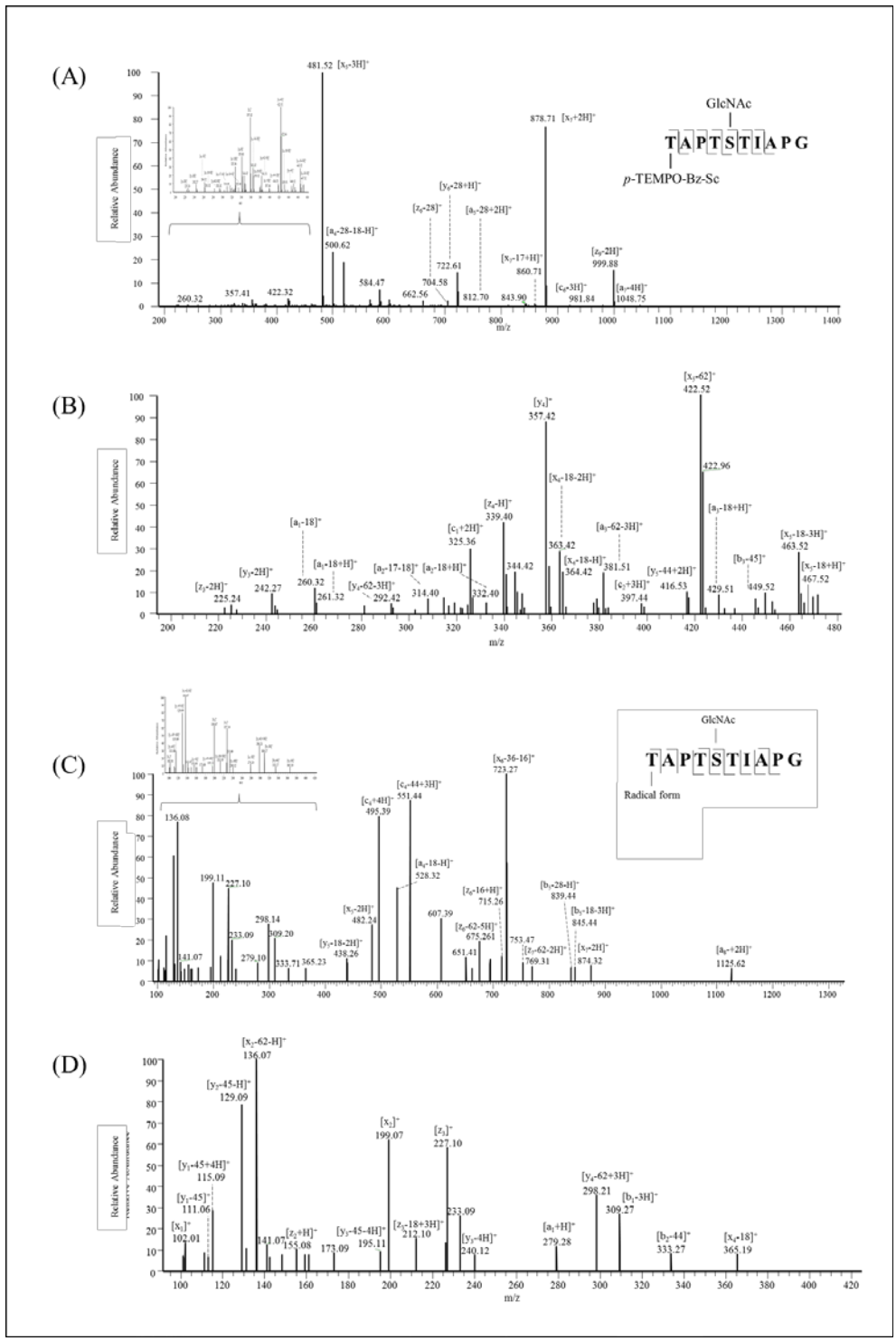
(A)
GlcNAc
TAPTSTIAPG
p-TEMPO-Bz-Sc
Relative Abundance
m/z
(B)
Relative Abundance
m/z
(C)
GlcNAc
TAPTSTIAPG
Radical form
Relative Abundance
m/z
(D)
Relative Abundance
m/z

Figure 9. Mass spectra following collision-induced dissociation (CID) of *O*-GlcNAc standard peptide using TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS) approach.

O-GlcNAc standard peptide (TAPTsTIAPG, 's' is modified *O*-GlcNAc) remains with *O*-GlcNAc moiety using TEMPO-assisted FRIPS MS. The data shows that CID on the *p*-TEMPO-Bz-Sc-NHS conjugated *O*-GlcNAc standard peptide produces radical ions during the peptide fragmentation process (Figure 9A and 9B). The *p*-TEMPO-Bz-Sc-NHS conjugated *O*-GlcNAc standard peptide produces peptide fragment ions and the *N*-terminus benzylic radical peptide fragment ions with TEMPO radical form (Figure 9C and 9D).

4. Analysis of O-GlcNAc proteins of HEK 293 cells under high glucose and glucose deprivation conditions

O-GlcNAc standard peptide (20 pmol) added to digested peptide mixtures isolated from HEK 293 cells for the experimental quality control. The sample was loaded onto a ProSwift™ (5 x 50 mm) Con A-1S LC columns with an isocratic 100 % of WGA buffer (50 mM sodium acetate (NaOAc) (Sigma-Aldrich), 0.2 M sodium chloride (NaCl) (Sigma-Aldrich) with 1 mM calcium chloride ($CaCl_2$,) (Sigma-Aldrich) 1 mM magnesium chloride ($MgCl_2$) (Sigma-Aldrich)) at a flow rate of 0.5 mL/min and monitored at 214 nm HPLC system. After enrichment, the sample was labeled by TEMPO reagent and further analyzed using LC-MS/MS for determination of possible modification sites on specific serine or threonine residues.

4.1. Analysis of O-GlcNAc proteins under high glucose and glucose deprivation conditions.

To identify proteins that are specifically modification with *O*-GlcNAc in HEK 293 cell grown under high glucose conditions, extraction of cytoplasmic and nuclear proteins was conducted using M-PER reagent. The samples was digested to peptides by in-solution tryptic digestion. The resulting peptide mixtures were subjected to lectin chromatography and TEMPO-assisted FRIPS MS. Enriched *O*-GlcNAc peptide mixtures were analyzed by LC-MS/MS analysis. Tandem mass spectra were analyzed by the SEQUEST based search platform. A total of 401 proteins (442 peptides) were identified with a false discovery rate (FDR) less than

0.1 % and further filtered with spectra scoring above certain XCorr cut-offs was included (XCorr >2.0 for all peptides) (Table 1). Fragment ions of selected MS/MS spectra were assigned, which provides detail information for the *O*-GlcNAc modification sites (Figure 10). Identical experimental condition was applied to the sample isolated from glucose deprivation conditions (Table 2 and Figure 11).

Table 1. The list of *O*-GlcNAc identified peptides and their quality under high glucose conditions.

All sequences of where first small letter means the radical conjugation and ‘s’ or ‘t’ is the *O*-GlcNAc modification site which is identified in high glucose conditions.

No.	Protein Group Accessions	Description	Sequence	# *PSMs	**XCorr
1	Q92878	DNA repair protein RAD50	vLAsLIIR	4	3.27
2	Q07666	KH domain-containing, RNA-binding, signal transduction-associated protein 1	gGGGGsRGGAR	1	3.10
3	Q14320	Protein FAM50A	gAAsEAGR	6	3.04
4	Q9Y3S1	Serine/threonine-protein kinase WNK2	dAsAPR	1	2.99
5	Q86X51	Uncharacterized protein CXorf67	aSsPSPPGR	2	2.92
6	C9JMR7	N-acetyl-D-glucosamine kinase	eVTGsGAVPR	1	2.90
7	A6NMB1	Sialic acid-binding Ig-like lectin 16	gsCSLVIR	28	2.86
8	O75683	Surfeit locus protein 6	gNLTPLtGR	1	2.86
9	C9JKY3	Epithelial cell adhesion molecule (Fragment)	gPsRGAASLR	1	2.84
10	Q8NDF8	Non-canonical poly(A) RNA polymerase PAPD5	asGGRAAGGGR	3	2.84
11	Q5VT97	Rho GTPase-activating protein SYDE2	dPAGsSVIR	21	2.83
12	E9PK46	Protein RIC-3 (Fragment)	gsGGGAGGGGSGR	3	2.82
13	Q9NZU7	Calcium-binding protein 1	aAAAAAsGGSR	11	2.82
14	A0A087WT25	Tubby-like protein	dPSGSPAsAR	1	2.81
15	C9JFF0	Kinesin-like protein KIF26A	aGPsVGAKAGR	1	2.81
16	Q6ZTR5	Cilia- and flagella-associated protein 47	gNVVtISPR	1	2.81
17	Q9BY12	S phase cyclin A-associated protein in the endoplasmic reticulum	aAELsSGR	2	2.81
18	Q96DN6	Methyl-CpG-binding domain protein 6	gPQtPR	1	2.80
19	Q6UXN2	Trem-like transcript 4 protein	aPACLGsGGPR	2	2.79
20	B5MCN9	Outer dense fiber protein 3B	aPAFtFGAR	8	2.78
21	A6NLJ0	C2 calcium-dependent domain-containing protein 4B	aPSsSPLSSR	8	2.77
22	Q9NZ56	Formin-2	gGGGGGGGGEsGK	1	2.77
23	P05997	Collagen alpha-2(V) chain	gDPGsHGR	1	2.77
24	Q86UU1	Pleckstrin homology-like domain family B member 1	gLAGAsGR	1	2.76

Table 1. Continued

25	Q8NFL0	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 7	eRQsAGGGR	3	2.75
26	Q9P232	Contactin-3	vAGNETsAR	4	2.75
27	Q9NPP4	NLR family CARD domain-containing protein 4	gVAAsDFIR	1	2.75
28	Q4VXP6	Potassium voltage-gated channel subfamily KQT member 2	gsPQAQTVR	1	2.75
29	Q9NYQ6	Cadherin EGF LAG seven-pass G-type receptor 1	aRLCGtGAR	1	2.72
30	Q2TAZ0	Autophagy-related protein 2 homolog A	aLDPKsTGR	1	2.71
31	Q8N7J2	APC membrane recruitment protein 2	gGAGAsVGVCR	1	2.70
32	P51957	Serine/threonine-protein kinase Nek4	vVtGSVSSSR	3	2.70
33	O76081	Regulator of G-protein signaling	lMHsPAGR	1	2.69
34	Q9P299	Coatomer subunit zeta-2	nVFNKTsR	21	2.68
35	Q5HY92	Fidgetin	fSsQSSR	2	2.67
36	Q9HAU4	E3 ubiquitin-protein ligase SMURF2	msNPGGR	1	2.67
37	Q9BXM7	Serine/threonine-protein kinase PINK1, mitochondrial	stGLLPGR	4	2.66
38	Q8N2Y8	Iporin	atGRGAR	1	2.65
39	Q5SNV9	Uncharacterized protein C1orf167	aQsKAHKR	4	2.65
40	Q5SYB0	FERM and PDZ domain-containing protein 1	vSsISAIR	2	2.65
41	A0A075B7C2	Uncharacterized protein C17orf50	gAAAGsGGGR	2	2.65
42	J3KPK3	RNA-binding motif protein, Y chromosome, family 1 member E	eYAPPsR	1	2.61
43	P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	vGSsLPGR	3	2.61
44	A0A0A0MR66	RNA binding motif protein 10, isoform CRA_d	gGRGDRtGR	1	2.61
45	Q9Y2I7	1-phosphatidylinositol 3-phosphate 5-kinase	gGsDYELAR	2	2.60
46	A0A087WZF2	Empty spiracles homolog 1 (Drosophila), isoform CRA_a	vQtAEAGGGR	3	2.60
47	B8ZZ34	Putative protein shisa-8	gsPHNSAGPR	1	2.60
48	Q5JPB2	Zinc finger protein 831	dHsQTAGR	1	2.59
49	Q17RG1	BTB/POZ domain-containing protein KCTD19	vVsLANR	6	2.59
50	Q9UN72	Protocadherin alpha-7	vGGtGGAVR	3	2.59
51	Q3MIS6	Zinc finger protein 528	cFLTsHQR	8	2.58
52	Q9HCE5	N6-adenosine-methyltransferase subunit METTL14	gGGRGGtSAGR	2	2.58
53	Q03113	Guanine nucleotide-binding protein subunit alpha-12	gsRVLVDAR	1	2.58
54	A0A087WTA8	Collagen alpha-2(I) chain	gPsGPQGIR	1	2.58
55	A0A087X0K8	Probable G-protein-coupled receptor 179	sALLsSGR	1	2.57

Table 1. Continued

56	Q02880	DNA topoisomerase 2-beta	dAAsPR	1	2.57
57	Q8N201	Integrator complex subunit 1	lGsSQVASR	2	2.57
58	Q5CZC0	Fibrous sheath-interacting protein 2	rASIsGR	3	2.57
59	Q9Y5I4	Protocadherin alpha-C2	sGTAQIsVR	2	2.56
60	P42684	Abelson tyrosine-protein kinase 2	gSSAARPsGR	1	2.56
61	Q5VV63	Attractin-like protein 1	tsGVR	1	2.56
62	Q8IVS8	Glycerate kinase	aLsLDPGGR	1	2.56
63	Q2LD37	Uncharacterized protein KIAA1109	eEISGSsDR	2	2.55
64	Q9UF83	Uncharacterized protein DKFZp434B061	asLTRTPSR	1	2.55
65	Q92522	Histone H1x	aAPGAAGsR	1	2.54
66	Q13939	Calicin	aAALsATSAGR	1	2.54
67	C9JY79	Non-erythrocytic beta-spectrin 4	eGGEGGGsR	1	2.54
68	Q5LJA9	Ubiquitin carboxyl-terminal hydrolase (Fragment)	gGsGRCVAR	1	2.54
69	Q6ZUU3	FOXL2 neighbor protein	aLQASsR	10	2.53
70	Q86T82	Ubiquitin carboxyl-terminal hydrolase 37	yLtSSR	4	2.53
71	Q5JR59	Microtubule-associated tumor suppressor candidate 2	gSSSGPsSPAR	5	2.53
72	Q8WY21	VPS10 domain-containing receptor SorCS1	vGAGGGsQAR	1	2.53
73	Q9Y6H8	Gap junction alpha-3 protein	aSRAsSGR	5	2.53
74	Q9UMD9	Collagen alpha-1(XVII) chain	tASLGGGsR	1	2.53
75	O75151	Lysine-specific demethylase PHF2	gsSLAAHGAR	5	2.52
76	Q6DN14	Multiple C2 and transmembrane domain-containing protein 1	gGtSDPYVK	4	2.52
77	E9PRU1	EGF-containing fibulin-like extracellular matrix protein 2	gEEVtR	3	2.52
78	Q5JR12	Protein phosphatase 1J	vMAtIGVTR	1	2.52
79	Q14593	Zinc finger protein 273	ssNLTR	3	2.51
80	A0A0A0MT07	Potassium voltage-gated channel subfamily KQT member 5	eQGEAsSR	5	2.51
81	Q02388	Collagen alpha-1(VII) chain	lVDtGPGAR	1	2.51
82	P20701	Integrin alpha-L	tsLLASGAPR	1	2.51
83	Q9HCE6	Rho guanine nucleotide exchange factor 10-like protein	lLtSGQR	2	2.50
84	B1AL05	39S ribosomal protein L43, mitochondrial	dGASsRGAR	6	2.50
85	Q9UF83	Uncharacterized protein DKFZp434B061	aSLTGtPPR	1	2.50
86	A0A087X208	Agrin	ttASVPR	3	2.50
87	Q5VT06	Centrosome-associated protein 350	sSsGSSR	2	2.50
88	Q9NPZ5	Galactosylgalactosylxylosylprotein3-beta-glucuronosyltransferase 2	mKsALFTR	3	2.50
89	C9JM79	Peptidyl-prolyl cis-trans isomerase G (Fragment)	tPsRSRSR	2	2.49

Table 1. Continued

90	Q8IYB3	Serine/arginine repetitive matrix protein 1	gASsSPQRR	3	2.49
91	P0C7T7	Putative uncharacterized protein FRMD6-AS1	gPPGsPVIR	1	2.49
92	Q7Z2K8	G protein-regulated inducer of neurite outgrowth 1	asPAPPR	1	2.48
93	Q9Y4B5	Microtubule cross-linking factor 1	vLHsPPAVR	1	2.48
94	Q76I76	Protein phosphatase Slingshot homolog 2	gNGSstPR	2	2.48
95	Q9Y566	SH3 and multiple ankyrin repeat domains protein 1	aPStSSSGR	3	2.48
96	Q96LD4	Tripartite motif-containing protein 47	gAsGAGGPGGAAR	1	2.48
97	Q96LX7	Coiled-coil domain-containing protein 17	gEELsR	1	2.48
98	Q9Y3S1	Serine/threonine-protein kinase WNK2	vGFVDStIK	3	2.47
99	E9PJN1	Ribosomal protein S6 kinase	aPVAsKGAPR	6	2.46
100	Q8IV04	Carabin	gIPsALR	1	2.46

*PSMs : The number of peptide spectrum matches. The number of PSM's is the total number of identified peptide spectra matched for the protein.

**XCorr: A measure of the goodness of fit of experimental peptide fragments to theoretical spectra.

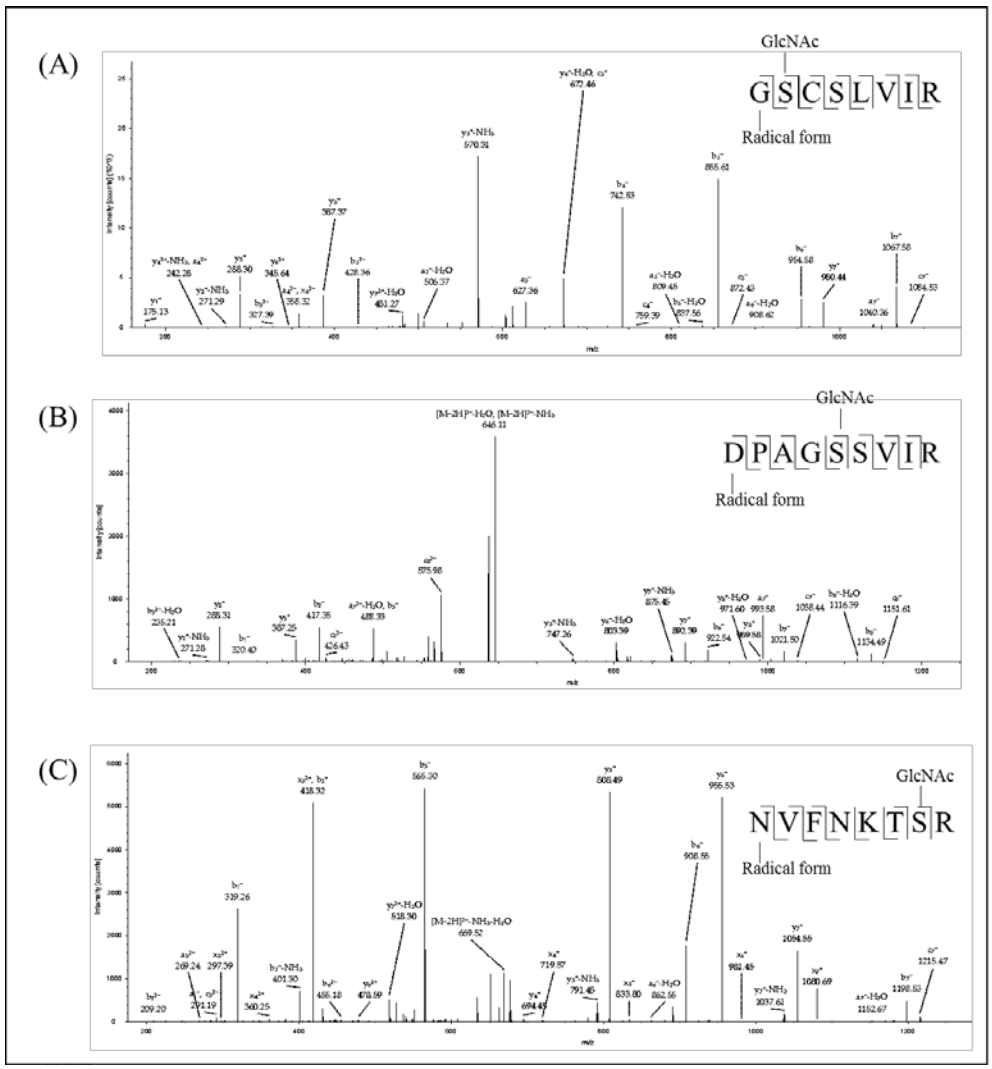

Figure 10. Tandem Mass Spectrometry (MS/MS) spectra of *O*-GlcNAc peptides with high peptide spectrum matches (PSMs) scores and XCorr >2.5 identified from HEK 293 cells grown under high glucose conditions.

Fragmentation of sialic acid-binding Ig-like lectin 16, Rho GTPase-activating protein SYDE2 and Coatomer subunit zeta-2.

Table 2. The list of *O*-GlcNAc identified peptides and their quality under glucose deprivation conditions..

All sequences of where first small letter means the radical conjugation and ‘s’ or ‘t’ is the *O*-GlcNAc modification site which is identified in high glucose conditions.

No.	Protein Group Accessions	Description	Sequence	# PSMs	XCorr
1	B8ZZ34	Putative protein shisa-8	gsPHNSAGPR	4	3.27
2	Q8WY21	VPS10 domain-containing receptor SorCS1	vGAGGGsQAR	1	3.21
3	Q9NZU7	Calcium-binding protein 1	gLsPALGLR	1	3.21
4	Q5VT97	Rho GTPase-activating protein SYDE2	dPAGsSVIR	15	3.08
5	O43281	Embryonal Fyn-associated substrate	aSGtQLAAPR	4	3.08
6	B7WPF5	Calpain-10	aGRGAtPAR	4	3.03
7	Q8WXI7	Mucin-16	tEITsSKR	17	3.02
8	A0A087WUK2	Heterogeneous nuclear ribonucleoprotein D-like	gAAAGGRGGtR	4	2.91
9	Q9UDW3	Zinc finger matrin-type protein 5	rLSsAPSSR	5	2.91
10	Q9ULH4	Leucine-rich repeat and fibronectin type-III domain-containing protein 2	gGGGsGGGEPPK	1	2.90
11	P51957	Serine/threonine-protein kinase Nek4	vVTGSVsSSR	6	2.88
12	Q9UIW0	Ventral anterior homeobox 2	aEsGGGGGR	1	2.86
13	Q96ME7	Zinc finger protein 512	eFVsESGVK	1	2.85
14	O75151	Lysine-specific demethylase PHF2	gsSLAAHGAR	3	2.85
15	Q6T4R5	Nance-Horan syndrome protein	vLStLDPK	1	2.85
16	O95996	Adenomatous polyposis coli protein 2	vAsALVPGR	3	2.83
17	Q96RL6	Sialic acid-binding Ig-like lectin 11	gsCSLVIR	12	2.82
18	Q9NS26	Sperm protein associated with the nucleus on the X chromosome A	qSsAGGVKR	7	2.82
19	Q8NDF8	Non-canonical poly(A) RNA polymerase PAPD5	asGGRAAGGGR	6	2.82
20	Q8WUM0	Nuclear pore complex protein Nup133	sVDKsSNR	1	2.81
21	Q9Y3S1	Serine/threonine-protein kinase WNK2	dAsAPR	2	2.81
22	O15230	Laminin subunit alpha-5	lAAsLDGAR	1	2.81
23	C9JKY3	Epithelial cell adhesion molecule (Fragment)	gPsRGAASLR	1	2.80
24	A0A0A0MSD9	Potassium voltage-gated channel subfamily KQT member 5	vLLNsAAAR	5	2.80
25	Q5JR59	Microtubule-associated tumor suppressor candidate 2	gSSSGPsSPAR	9	2.80

Table 2. Continued

26	J3QLS7	Coatomer subunit zeta-2 (Fragment)	nVFNKTsR	12	2.79
27	E9PDC3	Armadillo repeat protein deleted in velo-cardio-facial syndrome	gASSAGEAsEK	1	2.77
28	Q68DL7	Uncharacterized protein C18orf63	ntSVLGSPK	2	2.77
29	A0A0A0MT26	Sodium/potassium-transporting ATPase subunit alpha-3	gAAGsQAAGPR	1	2.76
30	Q9HAU4	E3 ubiquitin-protein ligase SMURF2	msNPGGR	4	2.76
31	Q9H1B5	Xylosyltransferase 2	dtDSSAGR	1	2.74
32	Q02846	Retinal guanylyl cyclase 1	wVsGAAVAR	1	2.74
33	Q92804	TATA-binding protein-associated factor 2N	gGGsGGGR	1	2.73
34	P09914	Interferon-induced protein with tetratricopeptide repeats 1	eAtKGQPR	3	2.73
35	Q5T9C9	Phosphatidylinositol 4-phosphate 5-kinase-like protein 1	tFStVSPAR	1	2.72
36	D6RGG3	Collagen alpha-1(XII) chain	gTGsSGPR	1	2.71
37	Q15643	Thyroid receptor-interacting protein 11	dtLLKER	1	2.69
38	Q9BRD0	BUD13 homolog	hGtPDPSPR	1	2.69
39	Q86W92	Liprin-beta-1	gGTRAtAGPR	3	2.69
40	Q9NZ56	Formin-2	gGGGGGGGGEsGK	2	2.69
41	Q9NZU7	Calcium-binding protein 1	aAAAAAsGGSR	6	2.69
42	Q9H091	Zinc finger MYND domain-containing protein 15	gAVGTsLEGR	4	2.69
43	Q96PE1	Adhesion G protein-coupled receptor A2	gLsGGVPGPAR	1	2.68
44	Q9Y3Q8	TSC22 domain family protein 4	gAsGGAGGR	1	2.67
45	Q96L93	Kinesin-like protein KIF16B	aDAtGATGVR	1	2.67
46	Q9UKP6	Urotensin-2 receptor	gPGsGGGR	2	2.65
47	Q9BXM7	Serine/threonine-protein kinase PINK1, mitochondrial	stGLLPGR	5	2.65
48	Q15772	Striated muscle preferentially expressed protein kinase	gSSAEsALPR	1	2.65
49	B4DFN8	Parathyroid hormone 2 receptor	atLADAR	3	2.64
50	Q9HCU4	Cadherin EGF LAG seven-pass G-type receptor 2	asCAAQR	6	2.64
51	Q92878	DNA repair protein RAD50	vLAsLIIR	1	2.64
52	O60725	Protein-S-isoprenylcysteine O-methyltransferase	vPtGLPFIK	3	2.63
53	A2A288	Probable ribonuclease ZC3H12D	aPGGsAGAR	2	2.63
54	Q9ULI4	Kinesin-like protein KIF26A	gRGLVAGGsR	2	2.63
55	B1AL05	39S ribosomal protein L43, mitochondrial	dGASsRGAR	2	2.62
56	A0A087WTA8	Collagen alpha-2(I) chain	gPsGPQGIR	1	2.62
57	Q5T1V6	Probable ATP-dependent RNA helicase DDX59	vIIAtPGR	2	2.62

Table 2. Continued

58	Q9UN72	Protocadherin alpha-7	vGGtGGAVR	1	2.62
59	Q13428	Treacle protein	kTNTtASAK	8	2.62
60	A0A075B7B7	GAS2-like protein 2	aIEAtTK	2	2.61
61	A0A087X0K4	CUB and sushi domain-containing protein 2	aDDACGGtLR	3	2.61
62	O75334	Liprin-alpha-2	lDNsTVR	2	2.61
63	Q14966	Zinc finger protein 638	rSVRSsDR	1	2.61
64	E7EU13	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1	gLGAGVsKVR	1	2.60
65	P20701	Integrin alpha-L	tsLLASGAPR	2	2.60
66	F5H7S1	Zinc finger protein 423	gtQTSPVPR	1	2.60
67	A0A087X1S4	Variable charge X-linked protein 1	ySPISEsSD	8	2.60
68	Q96NA8	t-SNARE domain-containing protein 1	msYGSIAR	1	2.60
69	P0C7T7	Putative uncharacterized protein FRMD6-AS1	gPPGsPVIR	1	2.60
70	Q9NUG4	Cerebral cavernous malformations 2 protein-like	aGGGGGGsLER	2	2.59
71	Q6ZN16	Mitogen-activated protein kinase kinase kinase 15	rNStGDR	1	2.59
72	E7EW31	Proline-rich basic protein 1	tsPLGGAR	1	2.59
73	B4DT28	Heterogeneous nuclear ribonucleoprotein R	gSRGsRGNR	1	2.58
74	Q9NZW5	MAGUK p55 subfamily member 6	aVVDAGItTK	14	2.58
75	Q7L8J4	SH3 domain-binding protein 5-like	gsDGGARGGR	1	2.58
76	Q9HCE5	N6-adenosine-methyltransferase subunit METTL14	gGGRGGtSAGR	2	2.58
77	Q6ZVN8	Hemojuvelin	gGGVGsGGLCR	3	2.58
78	Q5HY54	Filamin-A	gAGGQGKVAsK	3	2.57
79	F5H7K1	UPF0505 protein C16orf62	gAGsGGVR	1	2.57
80	Q12830	Nucleosome-remodeling factor subunit BPTF	vQSPsQTR	1	2.57
81	X6R3N0	Long-chain fatty acid transport protein 3	aAsSPGGSAPR	1	2.57
82	Q9BZZ5	Apoptosis inhibitor 5	gGtKEKR	3	2.56
83	Q15746	Myosin light chain kinase, smooth muscle	tSLsVDPSR	1	2.55
84	Q9H6K5	Proline-rich protein 36	rSALsAGAR	2	2.54
85	E9PGY3	Rap guanine nucleotide exchange factor 5	ltSAVQR	2	2.54
86	Q6IMN6	Caprin-2	gGtSGGPR	2	2.54
87	O75376	Nuclear receptor corepressor 1	gAGLsATIAR	1	2.53
88	Q14444	Caprin-1	gsGQSGPR	1	2.53
89	H7BZN7	Probable ATP-dependent RNA helicase DDX56 (Fragment)	lLGtADSPR	4	2.53
90	Q5VT06	Centrosome-associated protein 350	ssSGSSR	1	2.53

Table 2. Continued

91	Q9HCE5	N6-adenosine-methyltransferase subunit METTL14	gGtSAGR	2	2.52
92	Q14667	Protein KIAA0100	hSGtISQPR	1	2.52
93	E9PM69	26S protease regulatory subunit 6A	vIAAtNR	4	2.52
94	Q9NQ32	Uncharacterized protein C11orf16	aQtAVVGTTK	1	2.52
95	E9PJN1	Ribosomal protein S6 kinase	aPVAsKGAPR	2	2.52
96	Q5FWE3	Proline-rich transmembrane protein 3	gPQGGPGLsR	1	2.52
97	Q8IVF2	Protein AHNAK2	vMVTsAAR	1	2.51
98	H0YLZ0	Sodium/potassium/calcium exchanger 5	gGQtWAR	2	2.51
99	Q8NEH6	Meiosis-specific nuclear structural protein 1	nLsCSER	3	2.51
100	Q96PV4	Paraneoplastic antigen-like protein 5	vAMtPALR	1	2.51

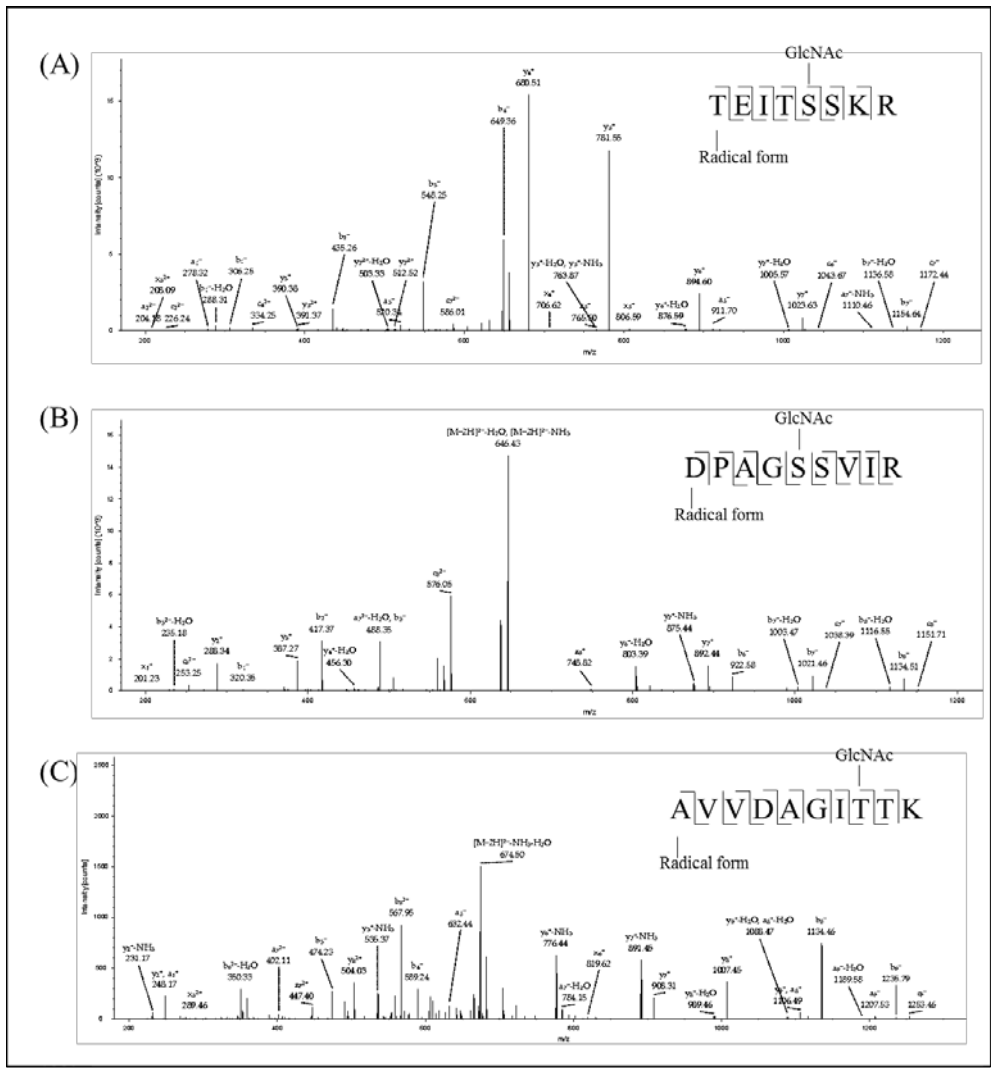

Figure 11. Tandem Mass Spectrometry (MS/MS) spectra of *O*-GlcNAc peptides with high peptide spectrum matches (PSMs) scores and XCorr >2.5 identified from HEK 293 cells grown under glucose deprivation conditions.

Site mapping of Mucin-16, Rho GTPase-activating protein SYDE2 and MAGUK p55 subfamily member 6 is provide information of *O*-GlcNAc modification.

4.2. Increased O-GlcNAcylation of HEK 293 cells in glucose deprivation conditions compared with high glucose conditions.

In this study, increased *O*-GlcNAcylation of HEK 293 cells in glucose deprivation conditions was identified (Table 3) and *O*-GlcNAc peptides were identified with accurate modification site mapping. First, E3 ubiquitin-protein ligase SMURF2 is associated with biological process including positive regulation of canonical Wnt signaling pathway, regulation of transforming growth factor beta receptor signaling pathway and protein ubiquitination involved in ubiquitin-dependent protein catabolic process (Bonni, Wang et al. 2001). Second, Growth arrest-specific protein is involved in cell cycle arrest and regulation of endoplasmic reticulum to golgi vesicle-mediated transport (Del Sal, Collavin et al. 1994). Third, Putative protein shisa-8 is associated with regulation of short-term neuronal synaptic plasticity (Nishiyama, Knopfel et al. 2002). HEK 293 cells were increased to *O*-GlcNAc in response by glucose deprivation.

Table 3. Increased *O*-GlcNAcylation of HEK 293 cells in glucose deprivation conditions compared with high glucose conditions

Glucose deprivation of HEK 293 cells stimulates an increase in total *O*-GlcNAc protein modification.

			#PSMs		
Protein Group Accessions	**Description**	**Sequence**	**HG**	**GD**	**GD vs. HG (con.)**
Q9HAU4	E3 ubiquitin-protein ligase SMURF2	msNPGGR	1	4	4.00
P54826	Growth arrest-specific protein 1	ssGGGGR	1	4	4.00
B8ZZ34	Putative protein shisa-8	gsPHNSAGPR	1	4	4.00
Q9Y4Z0	U6 snRNA-associated Sm-like protein LSm4	gRGGIPGtGR	1	4	4.00
Q9NS26	Sperm protein associated with the nucleus on the X chromosome A	qSsAGGVKR	2	7	3.50
O95714	E3 ubiquitin-protein ligase HERC2	gGLAGPDGtK	2	6	3.00
Q03001	Dystonin	tsLADDNLK	1	3	3.00
Q96LD4	Tripartite motif-containing protein 47	gAsGAGGPGGAAR	1	3	3.00
F8WAT8	DENN domain-containing protein 2A	lVNVKsR	1	3	3.00
Q8WXI7	Mucin-16	tEITsSKR	6	17	2.83
Q13428	Treacle protein	kTNTtASAK	3	8	2.67
O75052	Carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase protein	nSNsSGDPGR	2	5	2.50
Q9HCU4	Cadherin EGF LAG seven-pass G-type receptor 2	asCAAQR	3	6	2.00
Q8NDF8	Non-canonical poly(A) RNA polymerase PAPD5	asGGRAAGGGR	3	6	2.00
H7BZN7	Probable ATP-dependent RNA helicase DDX56 (Fragment)	lLGtADSPR	2	4	2.00
Q9NZ56	Formin-2	gGGGGGGGGEsGK	1	2	2.00
Q96JG9	Zinc finger protein 469	asGLRPR	1	2	2.00
P20701	Integrin alpha-L	tsLLASGAPR	1	2	2.00
Q9Y3S1	Serine/threonine-protein kinase WNK2	dAsAPR	1	2	2.00
P25054	Adenomatous polyposis coli protein	nASsIPR	1	2	2.00
C9J069	Uncharacterized protein C9orf172	aTERPsAR	1	2	2.00
Q68E01	Integrator complex subunit 3	gSSAVGsDSD	1	2	2.00

Table 3. Continued

Q9HCE5	N6-adenosine-methyltransferase subunit METTL14	gGtSAGR	1	2	2.00
A0A0A0MTS7	Titin	aGtSVKLR	1	2	2.00
Q02224	Centromere-associated protein E	aAQtGAAGVR	1	2	2.00
Q8TE68	Epidermal growth factor receptor kinase substrate 8-like protein 1	dNVtPR	1	2	2.00
Q5JR59	Microtubule-associated tumor suppressor candidate 2	gSSSGPsSPAR	5	9	1.80
Q9ULL0	Uncharacterized protein KIAA1210	mAVEGtSNK	5	9	1.80
Q5VVQ6	Ubiquitin thioesterase OTU1	ssPAFTKR	3	5	1.67
Q70EL4	Ubiquitin carboxyl-terminal hydrolase 43	yHTLsLGR	4	6	1.50
O14777	Kinetochore protein NDC80 homolog	sSVSsGGAGR	4	6	1.50
Q12986	Transcriptional repressor NF-X1	tsVISCR	2	3	1.50
O60245	Protocadherin-7	ssQAILR	2	3	1.50
A0A087WYC6	Dynein heavy chain 11, axonemal	lAAsQEIPR	2	3	1.50
Q6ZUU3	FOXL2 neighbor protein	aLQASsR	10	13	1.30
Q969Q1	E3 ubiquitin-protein ligase TRIM63	gSsVSMSGGR	12	15	1.25
Q8NDG6	Putative ATP-dependent RNA helicase TDRD9	iGAsSIAR	4	5	1.25
Q9BXM7	Serine/threonine-protein kinase PINK1, mitochondrial	stGLLPGR	4	5	1.25
Q9H2U1	ATP-dependent RNA helicase DHX36	gsGGGGGGGGGGR	18	20	1.11

Discussion

Although several reported methods for site-mapping *O*-GlcNAcylation have been successfully implemented to the biological systems, some methods have some limitations for detecting the modification sites due to its CID-labile β-linkage between *O*-GlcNAc moiety and serine or threonine residues and its sub-stoichiometric ratio at each site on proteins. To improve efficient enrichment and identification of *O*-GlcNAc peptides, a multi-method strategy has been developed using lectin chromatography and TEMPO-assisted FRIPS MS. To evaluate our strategy, we initially tested the method with the synthetic *O*-GlcNAc peptide. Lectin binds complex carbohydrates that contain *O*-GlcNAc and sialic acid residues. But lectin interaction with a single *O*-GlcNAc is not strong enough to enrich for modification peptides. Therefore, WGA column chromatographic separation for the enrichment of and *O*-GlcNAc peptides were implemented. We found that the majority of unmodified peptides eluted as a single peak followed by a chromatographic tail, which is enriched fractions of *O*-GlcNAc modified peptides. The BEMAD method has been adapted for the identification of modification sites based on selective BEMAD with DTT. However, the BEMAD method are not enough to provide correct information for *O*-GlcNAc peptide modification sites, because any moiety at a serine or threonine residues could result in a false

assignment as an *O*-GlcNAc site. Therefore, to overcome the limitations, an *N*-term chemical tag approach on peptides has been developed. The method provides radical generated peptide fragmentation by soft ionization in which retains the modification sites with a minimal sample loss by the single step method. Several chemical cleavable link reagents have been developed for the enrichment of *O*-GlcNAc proteins, but it suffers from significant disadvantages for site identification. For example, a photo-cleavable linker was conducted by conjunction with UDP-GalNAz 1 and Y289L GalT to sequence modified peptides and photo-cleavable linker resulted in incomplete cleavage (Wang, Udeshi et al. 2010). Importantly, the moiety of *O*-GlcNAc retained after cleavage provided an amine group, which increased the overall peptide charge and facilitated ionization by ETD. However, only a neutral hydroxyl group revealed a cleavage of the linker, and the halogenated glycosylated peptides demonstrated poor fragmentation efficiency using ETD. Therefore, we aimed to develop a tag that would be quantitatively attached, as well as cleavage to facilitate MS detection for *O*-GlcNAc peptide sequencing. TEMPO reagent was chosen to site mapping and evaluated the approach to identify sites of *O*-GlcNAc modification. Using the method, we identified 442 of *O*-GlcNAc modified peptides, 468 of *O*-GlcNAc modified peptides, and 125 of up regulated *O*-GlcNAc modified peptides under glucose deprivation conditions compared to the high glucose conditions. We identified some protein factors playing important roles in mediating the increased levels of *O*-GlcNAcylation that occur in cell signaling or pathway under glucose

deprivation conditions. Collectively, the *N*-term chemical tag has been quite successfully implemented for the study *O*-GlcNAc proteome in our biological working systems. We envisage that further application of the method to biological systems, which will elucidate the mechanism of specific role of protein *O*-GlcNAcylation in many biological processes.

Conclusion

We describe an improved multi-method for the comprehensive *O*-GlcNAc proteome analysis including accurate modification sites mapping. We experimented the refinement and validation of the multi-method based on lectin affinity chromatography and TEMPO-assisted FRIPS for the enrichment and identification of *O*-GlcNAc modified proteins. BEMAD is the widely used method for analysis of *O*-GlcNAc peptides. DTT attachment to monosaccharide in *O*-GlcNAc peptides provides a stable fragmentation pattern in MS analysis but there are some weakness including attach to any serine or threonine residues as false modification sites. In this study, the TEMPO reagent provides some benefits over other reported approaches including the BEMAD. First, the reagent induces peptide fragmentations providing accurate sequence information for mapping the modification sites. Second, the method provides a minimal sample loss by the single step reaction.

In conclusion, the new multi-method provides an ideal platform to profile global *O*-GlcNAc peptide for quantitative LC-MS/MS analysis and will enable the study of functional roles of *O*-GlcNAc modified proteins in biological systems.

References

Bond, M. R. and J. A. Hanover (2015). "A little sugar goes a long way: the cell biology of O-GlcNAc." J Cell Biol **208**(7): 869-880.

Bonni, S., H. R. Wang, C. G. Causing, P. Kavsak, S. L. Stroschein, K. Luo and J. L. Wrana (2001). "TGF-beta induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation." Nat Cell Biol **3**(6): 587-595.

Cain, J. A., N. Solis and S. J. Cordwell (2014). "Beyond gene expression: the impact of protein post-translational modifications in bacteria." J Proteomics **97**: 265-286.

Canas, B., D. Lopez-Ferrer, A. Ramos-Fernandez, E. Camafeita and E. Calvo (2006). "Mass spectrometry technologies for proteomics." Brief Funct Genomic Proteomic **4**(4): 295-320.

Cao, W., J. Cao, J. Huang, J. Yao, G. Yan, H. Xu and P. Yang (2013). "Discovery and confirmation of O-GlcNAcylated proteins in rat liver mitochondria by combination of mass spectrometry and immunological methods." PLoS One **8**(10): e76399.

Chalkley, R. J. and A. L. Burlingame (2001). "Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry." J Am Soc Mass Spectrom **12**(10): 1106-1113.

Del Sal, G., L. Collavin, M. E. Ruaro, P. Edomi, S. Saccone, G. D. Valle and C.

Schneider (1994). "Structure, function, and chromosome mapping of the growth-suppressing human homologue of the murine gas1 gene." Proc Natl Acad Sci U S A **91**(5): 1848-1852.

Duan, G. and D. Walther (2015). "The roles of post-translational modifications in the context of protein interaction networks." PLoS Comput Biol **11**(2): e1004049.

Gallop, P. M. and M. A. Paz (1975). "Posttranslational protein modifications, with special attention to collagen and elastin." Physiol Rev **55**(3): 418-487.

Hart, G. W., C. Slawson, G. Ramirez-Correa and O. Lagerlof (2011). "Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease." Annual Review of Biochemistry, Vol 80 **80**: 825-858.

Huang, J., F. Wang, M. Ye and H. Zou (2014). "Enrichment and separation techniques for large-scale proteomics analysis of the protein post-translational modifications." J Chromatogr A **1372C**: 1-17.

Kudlow, J. E. (2006). "Post-translational modification by O-GlcNAc: Another way to change protein function." Journal of Cellular Biochemistry **98**(5): 1062-1075.

Lee, M., M. Kang, B. Moon and H. B. Oh (2009). "Gas-phase peptide sequencing by TEMPO-mediated radical generation." Analyst **134**(8): 1706-1712.

Leon, I. R., V. Schwammle, O. N. Jensen and R. R. Sprenger (2013). "Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis." Mol Cell Proteomics **12**(10): 2992-3005.

Lima, V. V., K. Spitler, H. Choi, R. C. Webb and R. C. Tostes (2012). "O-GlcNAcylation and oxidation of proteins: is signalling in the cardiovascular system becoming sweeter?" Clin Sci (Lond) **123**(8): 473-486.

Love, D. C. and J. A. Hanover (2005). "The hexosamine signaling pathway:

deciphering the "O-GlcNAc code"." Sci STKE **2005**(312): re13.

Ma, J. and G. W. Hart (2014). "O-GlcNAc profiling: from proteins to proteomes." Clin Proteomics **11**(1): 8.

Marshall, D. L., C. S. Hansen, A. J. Trevitt, H. B. Oh and S. J. Blanksby (2014). "Photodissociation of TEMPO-modified peptides: new approaches to radical-directed dissociation of biomolecules." Phys Chem Chem Phys **16**(10): 4871-4879.

Mechref, Y. (2012). "Use of CID/ETD mass spectrometry to analyze glycopeptides." Curr Protoc Protein Sci **Chapter 12**: Unit 12 11 11-11.

Nandi, A., R. Sprung, D. K. Barma, Y. Zhao, S. C. Kim, J. R. Falck and Y. Zhao (2006). "Global identification of O-GlcNAc-modified proteins." Anal Chem **78**(2): 452-458.

Nishiyama, H., T. Knopfel, S. Endo and S. Itohara (2002). "Glial protein S100B modulates long-term neuronal synaptic plasticity." Proc Natl Acad Sci U S A **99**(6): 4037-4042.

Shirato, K., K. Nakajima, H. Korekane, S. Takamatsu, C. Gao, T. Angata, K. Ohtsubo and N. Taniguchi (2011). "Hypoxic regulation of glycosylation via the N-acetylglucosamine cycle." J Clin Biochem Nutr **48**(1): 20-25.

Sprung, R., A. Nandi, Y. Chen, S. C. Kim, D. Barma, J. R. Falck and Y. Zhao (2005). "Tagging-via-substrate strategy for probing O-GlcNAc modified proteins." J Proteome Res **4**(3): 950-957.

Vocadlo, D. J., H. C. Hang, E. J. Kim, J. A. Hanover and C. R. Bertozzi (2003). "A chemical approach for identifying O-GlcNAc-modified proteins in cells." Proc Natl Acad Sci U S A **100**(16): 9116-9121.

Wang, S., F. Yang, D. G. Camp, 2nd, K. Rodland, W. J. Qian, T. Liu and R. D.

Smith (2014). "Proteomic approaches for site-specific O-GlcNAcylation analysis." Bioanalysis **6**(19): 2571-2580.

Wang, Y. C., S. E. Peterson and J. F. Loring (2014). "Protein post-translational modifications and regulation of pluripotency in human stem cells." Cell Res **24**(2): 143-160.

Wang, Z., N. D. Udeshi, M. O'Malley, J. Shabanowitz, D. F. Hunt and G. W. Hart (2010). "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 **9**(1): 153-160.

Zaro, B. W., Y. Y. Yang, H. C. Hang and M. R. Pratt (2011). "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 **108**(20): 8146-8151.

국문초록

단백질 번역 후 변형과정 (post translational modifications, PTM) 중 하나인 오글루넥 당화 수식화 (*O*-linked β-N-acetylglucosamine, *O*-GlcNAc) 과정은 단백질과 아세틸글루코사민을 결합해 단백질의 기능을 조절하는 신호전달체계로써, 진핵 세포의 핵 및 세포질의 단백질에서 발견된다. 세린 또는 트레오닌 잔기의 side chain hydroxyl기가 *O*-linked β-N-acetylglucosamine으로 변형됨으로써 나타나며, *O*-GlcNAc transferase와 *O*-GlcNAcase 두 가지 효소에 의해 매우 빠르게 첨가되고 제거 되는 다이나믹한 특징을 가진다. 세포 내부 신호전달의 조절에 중요한 역할을 하는 것을 통해 세포 골격 단백질, 핵공 단백질, 암 억제 인자, 종양 형성 인자 및 다양한 전사 인자에 나타나는데 이러한 오글루넥 당화 수식화 과정은 포도당 영양상태와 밀접한 관계가 있으며 최근 암, 당뇨, 비만, 심혈관질환 및 알츠하이머 등 주요 질환 발병과 연관성이 높은 것으로 연구되고 있다. 오글루넥 당화 수식화 연구는 세포 내 신호전달 체계 및 유전자 발현 조절 이해 및 단백체의 기능 연구와 같은 기초 학문의 발전뿐 아니라, 당뇨와 신경 질환 등을 포함한 다양한 질병

메커니즘 이해 및 치료를 위해 그 중요성이 강조되고 있다. 그러나 오글루넥 당화 수식화 과정이 이러한 주요 질환 발병에 직접적인 연관성이 있음에도 불구하고 이에 대한 분석 연구는 활발히 진행되지 못한 한계를 지니고 있다. 그 이유는 오글루넥 당화 수식화 단백체가 생체 내에서 매우 적게 존재하여 enrichment에 어려움이 있으며, 변형 잔기의 결합이 매우 약해 분석 시 이온 단편화 과정에 사용되는 에너지에 의해 분해되어 정확한 변형 정보를 얻을 수 없기 때문이다. 따라서 본 연구에서는 오글루넥 당화 수식화 단백체 프로파일링을 위한 다각도의 접근 방법을 소개하고자 한다. 렉틴 크로마토그래피 방법을 통해 오글루넥 당화 수식화 펩타이드의 효율적인 enrichment 과정을 거친 후, TEMPO [2-(2,2,6,6-tetramethyl piperidine-1-oxyl)]-assisted free radical-initiated peptide sequencing mass spectrometry (FRIPS MS) 방법으로 변형 자리 분석 방법 최적화 및 HEK 293 세포에의 적용을 확인하였다. 특히 이러한 자유 라디칼 개시방법을 기반으로 한 TEMPO-assisted FRIPS MS은 기존에 널리 사용되는 ß-elimination followed by Michael addition with dithiothreitol (BEMAD) 방법 달리 단일 단계 반응을 통한 분석 시료가 유실되는 결과를 방지할 수 있으며, 오글루넥 당화 수식화 자리에 추가적인 화학적 변형이 이루어지지 않아 정확한 변형 자리 분석에 용이함을 밝혀 질량분석기 기반 충돌유발분해법을 응용한 변형 자리 분석 방법의 정확성을 검증하였다. 본 연구를 통해 오글루넥 당화

수식화의 확인 유무 및 정확한 변형 자리 분석이 가능함을 입증하였으며, 이를 바탕으로 추후 다양한 생물학적 시스템에서의 오글루넥 당화 수식화의 역할 및 질환 기전 연구에 적용 가능함을 예상한다. 또한 최적화된 방법을 통해 오글루넥 당화 수식화 단백체를 발굴하여 변형 자리를 정확하게 밝히는 것을 통해 생물학적 기능 및 세포신호전달 체계와의 관계성, 그리고 질환 치료 연구의 타겟을 연구하는 데에 기여할 것이다. 덧붙여 본 연구를 통해 정립된 연구법을 다양한 시료에 적용하여 질량분석기 기반 오글루넥 당화 수식화 분석 연구 플랫폼의 정확성을 증대시키는 데에 기반을 마련할 것이다.

주요어: 오글루넥 당화 수식화, 렉틴 크로마토그래피, 자유 라디칼 개시방법을 이용한 펩타이드 서열 분석법, 질량 분석기, 프로테오믹스

학번: 2014-24866