The Potential ‘O-GlcNAc-P’om’

I. Soo Moon\textsuperscript{1,2,}\textsuperscript{*}, HyunSook Lee\textsuperscript{2} and Hyung Jong Lee\textsuperscript{3}

\textsuperscript{1}Department of Anatomy, College of Medicine, Dongguk University, Gyeongju 780-714, Korea
\textsuperscript{2}Neuroscience Section, Medical Institute of Dongguk University, Gyeongju 780-714, Korea
\textsuperscript{3}Department of Obstetrics and Gynecology, Dongguk University College of Medicine, Gyeongju 780-714, Korea

Received February 19, 2012 /Revised February 27, 2012 /Accepted February 27, 2012

The addition and removal of N-acetylgalactosamine (GlcNAc) molecules on serine or threonine residues of a protein is called O-GlcNAcylation. This post-translational modification occurs on both cytoplasmic and nuclear protein, and is fast and reversible as comparable to phosphorylation. In contrast to the phospho-signaling cycles, this emerging moon-lightening signaling is cycled by only two enzymes, O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA). The simple machinery of a good evolutionary adaptation of a cell for quick accommodation to continuously fluctuating intra- and extracellular microenvironments. Rather than “switching” on or off a specific protein, this would be done by phosphorylation where numerous specific kinases and phosphatases are involved - O-GlcNAcylation would play a “rheostat” which would be much more delicately increase or decrease the efficacy of signal transmissions in response to cellular nutrient and stress conditions. Interestingly, recent evidence indicates that O-GlcNac is further modified by phosphorylation. The O-GlcNAc-P will upgrade the modulation efficiency of cellular processes to continuous ‘analogue’ level. So far, only one protein API80 was reported to have O-GlcNAc-P on Thr310. But, proteomic data from our laboratory indicate that there are multiple O-GlcNAc-P proteins, constituting “O-GlcNAc-P’om”. This will focus on the possibility of existence of ‘O-GlcNAc-P’om’.

Key words : NAGK, OGA, OGT, O-GlcNAc, O-GlcNAc-P

Introduction

The O-linked attachment of a single monosaccharide β-N-acetyl-D-glucosamine (GlcNAc) to serine and threonine residues (O-GlcNAc) was first reported in 1984 [26]. Since then, this type of post-translational modification of proteins (i.e., O-GlcNAcylation) was found in many different proteins. Most recently, a most advanced technology, the ‘Click’ chemistry, followed by LC-MS/MS afforded the identification of around 1500 O-GlcNAc proteins from a single cell line [9]. The O-GlcNAcylation is atypical, different from the classical glycosylation. Firstly, O-GlcNAcylated proteins are nucleocytoplasmic, whereas the classical glycosylation occurs on secreted extracellular proteins and on the extracellular segments of integral membrane proteins.

Secondly, O-GlcNAcylation is a single molecule (i.e., GlcNAc) addition, whereas the classical glyco-moteties are very complex. Thirdly, O-GlcNAcylation is a dynamic and reversible process of a single O-GlcNAc moiety [14]. In contrast, the classical glycosylation requires a series of irreversible addition reactions which occur in a co-translational or post-translational manner. Lastly, the sugar moiety O-GlcNAc was until recently thought to be a terminal modification, that is, the O-GlcNAc is not additionally modified in any way. This latter thought is now challenged and recent evidence indicates a possibility of phosphorylation on the O-GlcNAc to produce O-GlcNAc-P. Multiple O-GlcNAc-P proteins will constitute the dubbed ‘O-GlcNAc-P’om’. This review will present theoretical and experimental evidence that supports the existence of O-GlcNAc-P’om.

The O-GlcNAcylation

The GlcNAc molecule is energized by a high-energy covalent bond with uridine diphosphate (UDP). The transfer of the energized GlcNAc moiety in the UDP-GlcNAc onto a protein is catalyzed by the UDP-GlcNAc:polypeptide O-β/
-N-acetyl glucosaminyltransferase, or O-GlcNAc transferase (OGT). Conversely, the O-β-N-acetylglucosaminidase hexosaminidase C, O-GlcNAcase (OGA) removes this sugar moiety from proteins. O-GlcNAcylation can modulate protein functions in analogy to protein phosphorylation, and has been shown to be involved in almost all cellular processes including signal transduction and regulation of gene expression (see recent reviews [4, 22, 30]. An interesting finding is that O-GlcNAc and phosphate moieties can compete for the same or neighboring protein residues resulting in a complex interplay between the types of post-translational modification [3, 29].

The O-GlcNAc, a cellular 'rheostat'

The cycling of addition and removal of O-GlcNAc is comparable, in the time scale, to that of phosphorylation and dephosphorylation. However, there exists a fundamental difference between the two types of modification: while phosphorylation is regulated by hundreds of specific kinases and phosphatases, O-GlcNAcylation is irrespective of the target protein. The primary function of O-GlcNAcylation is thought to be the modulation of cellular processes in response to nutrients and to cellular stress. The modulation of protein functions by O-GlcNAcylation, however, is different in nature from phosphorylation. By analogy to an electrical circuit, the phosphorylation events represent "microswitches," which turn on or turn off protein activity. In contrast, O-GlcNAcylation represents a "rheostat", an adjustable resistor used in applications that require the adjustment of current or the varying of resistance in an electric circuit. Therefore, rather than controlling specific proteins, i.e. "microswitches", O-GlcNAcylation can tune the overall cellular processes to accommodate nutrient status and cellular stress [10].

The overall activity of OGT is regulated linearly by a wide range of intracellular concentrations of UDP-GlcNAc [15], which fluctuates in turn proportionally in response to the flow of nutrients (glucose, glutamine, energy) or stress [19]. Thus, the O-GlcNAcylation can play as a 'metabolic and stress sensor' by modulation of diverse proteins with just two enzymes (i.e., OGT and OGA) to adjust to the nutritional or stress status of the cell.

Is O-GlcNAc final modification?

Theoretical consideration I: O-GlcNAc is β-anomer, the right configuration for NAGK substrate

In the early step of UDP-GlcNAc biosynthesis, GlcNAc is phosphorylated to GlcNAc-6-phosphate (GlcNAc-6-P) by N-acetylglucosaminidase (GlcNAc kinase (GlcNAc kinase or NAGK; EC 2.7.159). By GlcNAc-6-P mutase, GlcNAc-6-P is converted to GlcNAc-1-P, which is coupled to UTP to produce UDP-GlcNAc. Phosphorylation of monosaccharide GlcNAc by NAGK is well known. On the other hand, it is an intriguing question to ask whether the O-GlcNAc, which is linked to the hydroxyl group of Ser or Thr of a protein, can serve as a substrate for NAGK. To answer this question it is most important to find the configuration of O-GlcNAc. Applying saturation transfer difference NMR experiments on human NAGK, Blume et al [2] found that NAGK exhibits a high preference for β-anomeric GlcNAc over other monosaccharides. Since O-GlcNAc transferase (OGT) also uses β-anomeric substrate, i.e. O-β-N-acetyl glucosamine, the O-GlcNAc is linked to a serine or threonine in β-configuration, which is the right anomeric configuration for a NAGK substrate (Fig. 1). Therefore, O-GlcNAc may serve as substrates for NAGK.

Theoretical consideration II: O-GlcNAc orientation and steric hindrance may not be a problem

Many different mammalian NAGK species were characterized [1, 5, 7]. Rat liver NAGK is a homodimer of 39 kDa subunits [13], and the human NAGK cDNA encodes a predicted molecular mass of 37.4 kDa [12]. Weihofer et al [28] elucidated crystal structures of homodimeric human NAGK in complexes with either GlcNAc or ADP/glucose. Unfortunately, they failed to crystallize a complex with both substrates, GlcNAc and ATP. Nonetheless, based on the two crystal structures, which they succeeded to determine, they established a 3D model for a NAGK complex with GlcNAc and ATP (Fig. 2). In this model, the N-terminal small and C-terminal large domains form a "V"-shaped structure, which acts as an active center for binding GlcNAc and ATP. In order for NAGK to bind its substrate GlcNAc, the O-GlcNAc must be presented in the right orientation, the C6-OH group of the glucopyranose pointing deep into the "V" valley. This means that the local structure of a substrate peptide with O-GlcNAc should be oriented such that the
Fig. 1. The O-GlcNAc modification. OGT couples GlcNAc moiety of UDP-GlcNAc to the hydroxyl group of a serine or threonine residue of a protein. The C1 hydroxyl group of GlcNAc moiety is β-anomeric, same as the monomer substrate GlcNAc for NAGK, which phosphorylates on its C6 hydroxyl group. The Ser/Thr-linked GlcNAc is called O-GlcNAc. The removal reaction is catalysed by OGA.

Fig. 2. Three-Dimensional representation of human NAGK in complex with ADP and glucose. NAGK forms a dimer in the cell (green and pink colors), through interaction of large domains. Between the small (broken circles) and large domain is “V” valley (broken curves) which is the substrate binding pocket. The 3D structure was from Weihofen et al. [28].

GlcNAc moiety thrust into the “V” pocket of NAGK. The -OH group of serine/threonine residues on a substrate protein is coupled to Cl –OH group of UDP-GlcNAc’s β-D-glucopyranose, so that the C6 -OH group is positioned in the free end side. When a substrate peptide thrusts into the “V” pocket, this is the right orientation as a natural monomer GlcNAc positions in the active site (Fig. 3). Another critical criterion for suitability of O-GlcNAc as a NAGK substrate is whether the “V” valley is spacious enough to avoid steric hindrance by the nearby structures surrounding the target O-GlcNAc peptide. Although no 3D structure of an O-GlcNAc bound NAGK complex is available at the present time, it can be simulated by putting the structure of substrate

Fig. 3. Crystal structures of the “V” valley of human NAGK in complex with ADP and Glucose. The orientation of substrates and the C1, C6 hydroxyl groups of glucose, and γ-phosphate of ATP are marked (A). Two examples of images viewed at different directions are shown (B). The inset in B is the 3D structure of TAB1 O-GlcNAc peptide in the similar size dimension as the NAGK. Note that “V” valley is spacious enough to accommodate the O-GlcNAc substrate peptide. Adapted from Weihofen et al. [28].
peptide into the “V” valley of NAGK’s substrate binding site. Taking advantage of the 3D viewing program CN 3D (version 4.3) we can view the crystal structure of human NAGK in complex with ADP and glucose [28] (NCBI MMDB ID: 41376, http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?Dopt=s&uid=41376). As shown in Fig. 3, the “V” pocket of NAGK active site was predicted spacious enough to allow the O-GlcNAc substrate peptide to thrust into. Therefore, it can be anticipated that a substrate protein can present its O-GlcNAc moiety into the “V” pocket in the right orientation deep enough to place it on the GlcNAc binding site without steric hindrance.

Theoretical consideration III: actin’s O-GlcNAc protrudes out of the peptide backbone

As is explained so far, it is essential requirement that a target O-GlcNAc peptide thrust out of the periphery. We examined through O-GlcNAc databases if this is the case. Contractile proteins actin and myosin are O-GlcNAcylated [11]. The crystal structure of the Drosophila 5C cytoplasmic actin with point mutations A204E/P243K (AP-actin) was determined in complex with cytochalasin D [20]. Actins are well conserved proteins and the Drosophila 5C actin is > 98% identical with human γ-cytoplasmic actin. The rat actin is O-GlcNAcylated on Ser198 [11], which corresponds to Ser199 of Drosophila 5C actin. The peptide loop containing Ser199 is positioned on the surface of the global protein (Fig. 4A). When viewed in various rotations, it is evident that the local loop as well as its regional domain protrudes out to periphery (Fig. 4B). Unfortunately, we could not find any other 3D structures of O-GlcNAc proteins. However, the actin’s case strongly indicates that at least some O-GlcNAc would serve as substrates for NAGK.

Theoretical consideration IV: possible substrate compatibility of O-GlcNAc peptides

The target loop containing O-GlcNAc protrusion must be such a dimension as to fit the substrate binding “V” pocket of NAGK. Since there is no such data available so far, we inferred the possible compatibility from several known O-GlcNAc peptides. Schimpl et al. [24] reported the molecular details of the interaction of a bacterial O-GlcNAcase (OGA) homolog with three different synthetic O-GlcNAc peptides derived from characterized O-GlcNAc sites in the human proteome: p53 (Ser149 [29]), TAK1-binding protein 1 (TAB1, Ser395 [22]), and hOGA itself (Ser405 [17]). The crystal structures of these peptides in complexes with the Clostridium perfringens OGA (CpOGA) was determined (Fig. 5A). The peptides bind a conserved OGA substrate binding groove with similar orientation and conformation (Fig. 5A; a, b, c). An important fact in our side of view is that, in addition to extensive contacts with the sugar, the OGA recognizes the peptide backbone. This fact indicates that the target O-GlcNAc peptides thrust deep into the substrate binding pocket of OGA, which, in turn, suggests that the local domain containing O-GlcNAc moiety thrust deep enough into the “V” valley of NAGK.

The deep thrust into the substrate binding pocket of OGA was well expected by the 3D structure of OGT, the first enzyme in O-GlcNAc cycling. The crystal structure of human OGT in complex with UDP-GlcNAc was determined by Lazarus et al. [16]. As shown in Fig. 5B, the binding site for UDP-GlcNAc is very deep in the protein. The deep position of the OGT’s substrate binding site requires the target O-GlcNAc peptides to thrust out far from the protein’s surface. Therefore, the O-GlcNAc cycling is, in nature, likely to occur on the peptides that protrude far out of the surface.

Experimental Evidence I: AP180 contains the O-GlcNAc-P mass, m/z 284

The first experimental evidence for the phosphorylation of O-GlcNAc was found by Graham et al. [8]. Assembly protein AP180, which is modified with both phosphate and O-GlcNAc, is involved in the assembly of clathrin coated vesicles in synaptic vesicle endocytosis. To determine the sites for phosphorylation and O-GlcNAcylation of the protein, they analysed rat brain AP180 phosphoepitides and
Fig. 5. Crystal structures of OGT and OGA showing their substrate binding pockets. The OGA structure in complex with its substrate GlcNAc peptide (A) and OGT structure with UDP-GlcNAc (B) are shown. The enlarged views of the OGA crystal structures with different substrate peptides in O-GlcNAcylated states were shown in the insets of A (a, b, c). Note the space large enough to allow the thrust of substrate peptides (insets of A) and deep positioning of UDP-GlcNAc in OGT substrate binding pocket (B). See details in the text. Adapted from Schimpl et al. [24] for Clostridium perfringens OGA (CpOGA) and Lazarus et al. [16] for human OGT (hOGT) in complexes with different known O-GlcNAc substrate peptides.

O-GlcNAc modified peptides by LC-MS/MS. To their surprise, they noticed that the more abundant signal at m/z 284 accounted for the difference between the nonmodified (M, 1545 Da) and modified API80 305-320 (M, 1828 Da) peptide by O-GlcNAc-P (Fig. 6A). A detailed analysis showed that the signal at m/z 284 is O-GlcNAc-P at Thr-310. The Thr-310 was either found in naked or O-GlcNAcylated or O-GlcNAc-P, but not phosphorylated alone (Fig. 6B). Interestingly, the flanking amino acids Ser306 and Ser313 were phosphorylated. A sequence with a high homology to API80 305-320 was repeated in the C-terminal portion API80 598-630. In this region, Ser621 and Ser627 were also phosphorylated (Fig. 6B). However, due to the low signal strength, they were unable to confirm the O-GlcNAcylation of Ser625. The interaction or cooperation of O-GlcNAc-P with nearby phosphorylation sites will offer diverse leverage for the control of such proteins. Multiple phospho-sites increase net negative charge and hydrophilicity. O-GlcNAc also increases hydrophilicity and therefore will affect the sol-

Fig. 6. API80 is modified by O-GlcNAc-P. A simplified MS profile shows two m/z (A). The m/z 284 and m/z 204 signals represents O-GlcNAc-P and O-GlcNAc, respectively at Thr-310. The local sequence including Thr310, phospho-sites flanking it, and a homologous sequence with potential O-GlcNAc site (T625) with flanking phospho-sites are marked. Adapted from Graham et al. [8].
utility and binding characteristics with other proteins. Subtle changes in the protein's character will be used to fine-tune its function to the fluctuation of the cellular physiology.

Experimental Evidence II: NAGK increases the phospho-signal of some select O-GlcNAc proteins

Our laboratory set out to identify potential O-GlcNAc-P proteins by proteomics. To minimize artificial inefficiency in the phosphorylation process we took advantage of cell cultures. The human embryonic kidney (HEK293T) cells were transfected with wild-type (pDsRed2-NAGKwt) or mutant (pDsRed2-NAGK<sup>D107A</sup>) NAGK-expression vectors. The D107A point-mutated NAGK retained almost no enzyme activities (manuscript in preparation). As reported in Moon et al. [18], we showed some of the phospho-signal increased proteins by NAGK overexpression were O-GlcNAc proteins. In brief, comparison of 2D-gel spots of pDsRed2-NAGK<sup>WT</sup>-transfected HEK293T cell extract with pDsRed2-NAGK<sup>D107A</sup>-transfected control culture one revealed 31 spots with altered phospho-signals. By eliminating spots that showed drastic increase or decrease in the protein amounts, we selected 7 spots for determination of protein identity. Among the 6 proteins, which we succeeded in ID determination, 4 of them were O-GlcNAcylated, and one was glycosylated. Furthermore, to prove the credibility of the data, the 4 O-GlcNAc protein spots were duplicates of the same two proteins, two HSP90<sup>β</sup> and two enolase 1 (ENO1) spots. The Ser<sup>436</sup> and Ser<sup>452</sup> of Hsp90<sup>β</sup> and the Ser<sup>461</sup> of Hsp90α are O-GlcNAcylated [23]. The Ser<sup>462</sup> is also a phosphorylation site [6]. Since ENO1 was reported to be phosphorylated by PKC [21], these proteins are probably another examples of complex interplay between phosphorylation and O-GlcNAcylation.

Existence of ‘O-GlcNAc-P’om? What does it mean?

The cycling of O-GlcNAcylation is peculiar in that it is regulated by just two enzymes, OGT and OGA. In contrast to phosphorylation cycling, which is regulated by numerous specific kinases and phosphatases, O-GlcNAcylation is irrespective of the target protein. Therefore, the primary function of O-GlcNAcylation is thought to adjust cellular processes in response to nutrients and to cellular stress. Therefore, rather than turning on or off of a single protein’s function (i.e., "microswitch"), O-GlcNAcylation is well adapted to control the overall flow rate of cellular signal transduction, because it requires only two enzymes to be controlled. By analogy to an electrical circuit, O-GlcNAcylation plays as a “rheostat” that is used to tune the rate of overall cellular processes to accommodate nutrient status and cellular stress [10]. Supporting this inference, the overall activity of OGT is regulated linearly by the concentrations of UDP-GlcNAc [15]. Since the concentration of UDP-GlcNAc sensitively fluctuates in response to the flow of nutrients (glucose, glutamine, energy) or stress [19], the O-GlcNAcylation can play as a “rheostat” to quickly adjust cellular processes to fluctuating energy and stress level.

The O-GlcNAc “rheostat” works in combination with phosphorylation. This combination will offer diverse leverage for the control over O-GlcNAc proteins (Fig. 7). This system is not binary with an ‘on’ or ‘off’ state for each signaling molecule but is ‘differential’. By adding another control tool ‘O-GlcNAc-P’, this system can set the flow rate of cellular processes at enormously diverse levels. Diverse changes in the protein’s character will adjust the cell to the subtle fluctuation of the cellular physiology. Such wide diversity will represent the continuous 'analogue' control of cellular processes instead of digital step-wise jump in the control.
Conclusion and prospect

O-GlcNAcylation is a moon-lightening signal controlling mechanism, which emerges only recently. The system is simple being controlled by only two enzymes. This, in turn, is a good evolutionary adaptation to quickly accommodate a cell to moment-moment fluctuating intra- and extracellular microenvironments. The O-GlcNAc "rheostat" would be much more delicately control the efficacy of signal transduction by adding another control lever, the phosphorylation of O-GlcNAc. It can be inferred that the key regulatory proteins in each signal transduction may be the target O-GlcNAc to be phosphorylated. These strategic proteins may constitute the O-GlcNAc-P group, dubbed 'O-GlcNAc-Pom'. Characterization of O-GlcNAc-Pom will shed light on our understanding normal and pathophysiology of cellular responses.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (grant number 2012006116).

References

초록: 'O-GlcNAc-P'om'의 존재 가능성

O-GlcNAc (O-GlcNAcyltransferase)는 단백질의 serine이나 threonine에 N-acetylglucosamine (GlcNAc)를 가리키는 것으로, 기존의 단백질의 단일 혹은 몇몇 양이 나와서 필요하다. 또한 수많은 단백질과의 상호 작용이 가능해서 인산화와 dephosphorylation이 이에 관계한다. 그러나 O-GlcNAc의 수치는 O-GlcNAc transferase (OGT)와 O-GlcNAcase (OGA)의 두 개의 효소에 의하여 이루어진다. 이러한 단순한 조절기작은 세포의 내질환에서의 작용을 통해 조절할 수 있도록 조절한 것으로 해석된다. 즉, O-GlcNAc 수치는 단백질의 이산화나 양화가 하나의 활성을 가리키고, 이는 단백질의 분자 내에서의 상호작용을 저하시킬 수 있는 결과를 일으킨다. O-GlcNAc 수치는 일반적으로 아미노산 혹은 고유의 아미노산 가입신호의 인산화와 dephosphorylation이 이에 관계한다. 최근 O-GlcNAc의 단백질의 O-GlcNAc 수치는 아미노산의 기여가 되고 있는 바, 본 연구에서는 이의 가능성을 이론적으로 설명하고, 실제 실험결과를 소개한다.