Sequence analysis of intrinsically disordered region of nucleus related human *O*-GlcNAcylated protein

Takuma Shirahase[1]
sj0002si@ed.ritsumei.ac.jp

Takayoshi Tomono[1]
sj004081@ed.ritsumei.ac.jp

Hisao Kojima[1]
hkojima@fc.ritsumei.ac.jp

Yukako Tohsato[1, 2]
yukako.tohsato@riken.jp

Masahiro Ito[1]
maito@sk.ritsumei.ac.jp

[1] Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan
[2] Current affiliation: Laboratory for Developmental Dynamics, RIKEN Quantitative Biology Center, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

Keywords: *O*-GlcNAcylation, Sequence analysis, Intrinsically Disordered Protein, Posttranslational modification

1 Introduction

Proteins acquire functions with altered structural and chemical properties due to posttranslational modifications such as glycosylation, phosphorylation etc. *O*-GlcNAcylation, one such modification, occurs on specific serine (Ser)/threonine (Thr) residues of proteins by attachment of the *N*-acetylglucosamine monosaccharide. Proteins with this modification are involved in transcriptional control and signal transduction [1]. Moreover, *O*-GlcNAcylated proteins are localized only in nucleus and cytoplasm, consensus sequence for *O*-GlcNAcylation remains unclear [2]. *O*-GlcNAcylation has been reported on Ser/Thr residues that phosphorylated themselves, or neighbor phosphorylation residue, suggesting that the functions and localization of *O*-GlcNAcylated protein involve the control of phosphorylation by *O*-GlcNAcylation [1].

Although proteins generally form certain three-dimensional structure by folding, some proteins do not fold on their own: there are referred to as intrinsically disordered proteins (IDPs). IDPs are of interest in the following sense: acquisition of functionality with changing structure due to interactions with other protein molecule, and structural stabilization by posttranslational modifications [3]. Most IDPs, localized in the nucleus, are involved in controlling signal transduction, and transcription factors have a high proportion of intrinsically disordered (ID) regions [3, 4].

This study was conducted to find relationship between *O*-GlcNAcylation and ID region for human *O*-GlcNAcylated proteins with nucleus-related localization, and to find characteristic of the modified region.

2 Methods and Results

Human *O*-GlcNAcylated proteins were acquired via protein databases such as UniProt, dbOGAP, and PhosphoSitePlus. Among these, 127 were nucleus-related proteins, with total of 291 *O*-GlcNAcylated residues. To determine the characteristics of the modified region, the amino acid composition around the residues was analyzed. Particular amino acids were found to be conserved, most of which were hydrophilic amino acids (Figure 1). We calculated the proportion of ID regions among human *O*-GlcNAcylated proteins and the entries set of recorded proteins in DICHOT, a database that classifies all human proteins into structured regions and ID regions, utilizing the acquired ID regions from DICHOT. We compared the proportion of ID region between whole human proteins, nucleus-related *O*-GlcNAcylated proteins, and cytoplasm-localized *O*-GlcNAcylated proteins. The resulting proportion showed that nucleus-related *O*-GlcNAcylated proteins had a high proportion (57.2 %) of ID regions, compared to whole human proteins (34.3 %) and cytoplasm-localized *O*-GlcNAcylated proteins (43.7 %) (Figure 2). Furthermore, the amount of *O*-GlcNAcylation in ID regions was compared between nucleus-related *O*-GlcNAcylated proteins and cytoplasm-localized *O*-GlcNAcylated proteins, and analyzed by the χ^2 test. The proportion of *O*-GlcNAcylation in nucleus-related proteins was found to be significantly high ($P < 0.01$) in ID regions (Figure 3). Next, we analyzed sequence properties around *O*-GlcNAcylated residues and phosphorylated residues, and analyzed the distances between these 2 types of residues. Half of the *O*-GlcNAc modifications were found on phosphorylation residue or their neighboring residues. Moreover, similar frequencies of *O*-GlcNAcylation were observed in ID regions and structured regions in the case of competitive modification between *O*-GlcNAcylation on the same residue, whereas the frequency was higher in ID regions in the case of

O-GlcNAcylated residues neighboring phosphorylated residues (Figure 4).

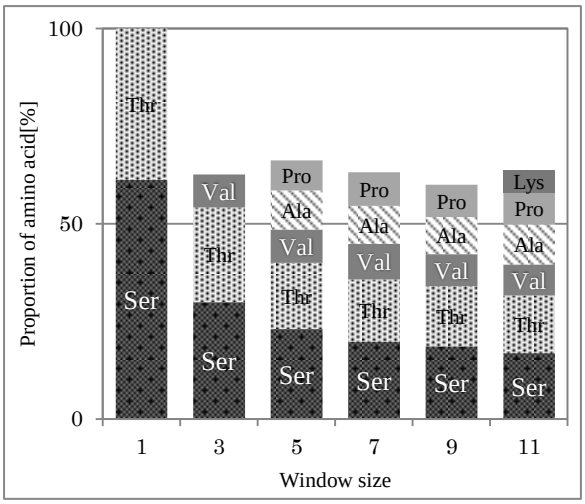

Figure 1: Amino acid component around *O*-GlcNAcylated residues

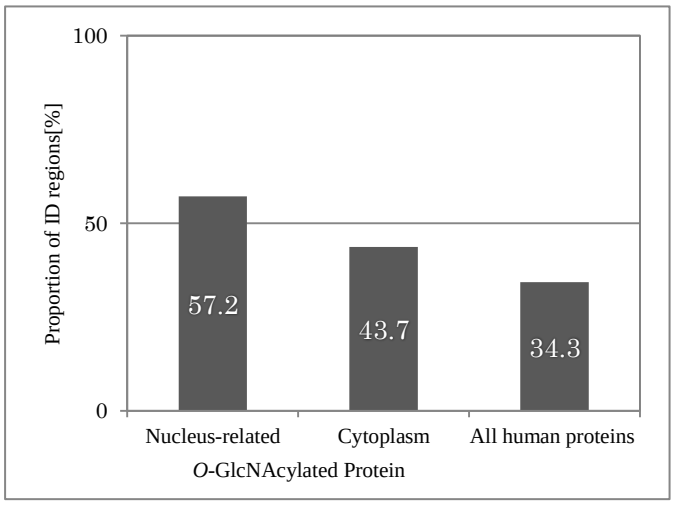

Figure 2: Proportion of ID region in all human proteins and *O*-GlcNAcylated proteins

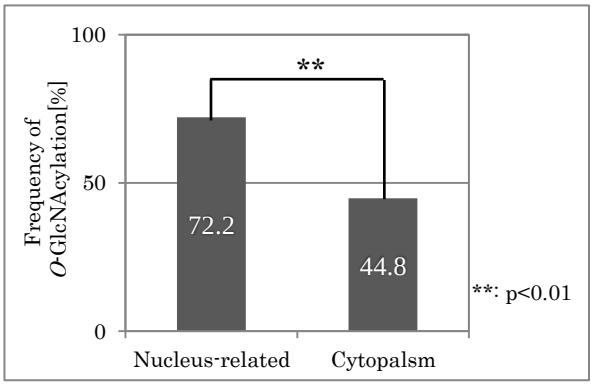

Figure 3: Frequency of *O*-GlcNAcylation in ID region

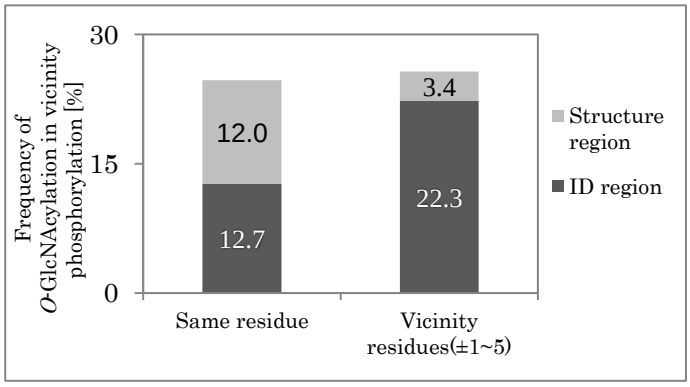

Figure 4: Frequency of *O*-GlcNAcylation residue that neighbor phosphorylation residues

3 Discussions

Because particular hydrophilic amino acids are conserved around *O*-GlcNAcylated residues, it is possible that *O*-GlcNAcylation is catalyzed by recognition of the hydrophilic amino acid composition around these regions. The proportion nucleus-related *O*-GlcNAcylated protein was high among those with ID regions, suggesting that *O*-GlcNAc modification interacts with ID regions. Moreover, the case in which *O*-GlcNAcylated residues was close to phosphorylated residues was high. This is consistent with a scenario in which transcriptional factor initiates transcription by transferring to nucleus as the result of phosphorylation being controlled by *O*-GlcNAc modification [1].

References

[1] G. Majumdar, A. Harrington, J. Hungerford, A. M. Hernandez, I. C. Gerling, R. Raghow, and S. Solomon, Insulin dynamically regulates calmodulin gene expression by sequential *O*-glycosylation and phosphorylation of sp1 and its subcellular compartmentalization in liver cells, *The Journal of Biological Chemistry*,**281**, 3642-3650, (2006)

[2] G. Walsh and R. Jefferis, Post-translational modifications in the context of therapeutic proteins, *Nature Biotechnology*, **24**, 1241-1252, (2006)

[3] V. N. Uversky, A. K. Dunker, Understanding protein non-folding, *Biochimica et Biophysica Acta*, **1804,** 1231-1264, (2010)

[4] S. Fukuchi, K. Hosoda, K. Homma, T. Gojobori and K. Nishikawa, Binary classification of protein molecules into intrinsically disordered and ordered segments, *BMC Structural Biology*, **11**, 29-38 (2011)