Ubiquitous and temporal glycosylation of nuclear and cytoplasmic proteins

Gerald W. Hart, Kenneth D. Greis, L.-Y. Dennis Dong, Melissa A. Blomberg, Teh-Ying Chou, Man-Show Jiang, Elizabeth P. Roquemore, Doris M. Snow, Lisa K. Kreppel, Robert C. Cole and Bradley K. Hayes. Dept. Biochem. & Molecular Genetics, University of Alabama at Birmingham, UAB Station, 1918 University Avenue, Birmingham, AL 35294-0005; Phone: (205) 934-4753; FAX: (205) 975-6685; Internet: GWHART@BMG.BHS.UAB.EDU.

We have described a novel form of nuclear and cytoplasmic protein glycosylation (O-GlcNAc), which is as abundant and as transient on intracellular proteins as protein phosphorylation. O-GlcNAc consists of single, non-modified N-acetylglucosamine residues O-glycosidically attached to Ser(Thr) hydroxyl moieties at sites similar to those used by growth-factor kinases. O-GlcNAc occurs on 'hundreds' of intracellular proteins in all eukaryotes. Proteins bearing O-GlcNAc include cytoskeletal-, viral-, nuclear pore-, heat shock-, and transcriptional regulatory proteins. Available data suggest that O-GlcNAc is a regulatory modification that mediates subunit-subunit interactions, and in many cases blocks phosphorylation.

Ten-years ago while probing the surfaces of murine lymphocyte sub-populations for terminal N-acetylglucosamine (GlcNAc) moieties using bovine milk galactosyltransferase and UDP-[3H]galactose, we found that many lymphocyte proteins contained single O-glycosidically linked GlcNAc residues (1). Detergent latency experiments also demonstrated that the vast majority of these O-GlcNAc glycoproteins were nucleoplasmic or cytosolic. Studies of rat liver subcellular fractions showed O-GlcNAc to be greatly enriched in nuclear fractions (2). Later studies of highly-purified, 100% viable lymphocytes showed O-GlcNAc to be localized exclusively in nucleoplasm or cytoplasm (3). Fig. 1, which shows a 2D-gel analysis of O-GlcNAc proteins from liver nuclei that are radiolabeled by galactosyltransferase, illustrates the large numbers of these glycoproteins in the nucleus.

NEPHGE GEL

<table>
<thead>
<tr>
<th>OH^-</th>
<th>H^+</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% SDS-PAGE</td>
<td></td>
</tr>
</tbody>
</table>

Exposure Time

7 Days 30 Days

Fig. 1 2D Gel Electrophoresis of Rat Liver Nuclei O-GlcNAc Glycoproteins Radiolabeled with UDP-[3H]galactose and bovine milk galactosyltransferase.
O-GlcNAc proteins have been found in all eukaryotes, but not in bacteria (4-6). To date only a few of the O-GlcNAc-bearing proteins have been identified (Table 1). While these O-GlcNAc proteins have a diverse range of functions, they share common features: 1) All are also phosphoproteins; 2) All reversibly form multimeric associations depending upon intracellular signals or stages of the cell cycle; and 3) All are localized in the cytosolic or nucleoplasmic compartments of cells.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Reference</th>
<th>Proteins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear pore proteins</td>
<td>(7-12)</td>
<td>Human Erythrocyte Band 4.1 &amp; 65 kDa protein</td>
<td>(13)</td>
</tr>
<tr>
<td>RNA Polymerase II catalytic subunit</td>
<td>(14)</td>
<td>Cytokeratins 13, 8, 18</td>
<td>(15-18)</td>
</tr>
<tr>
<td>Many RNA Pol II</td>
<td>(19-22)</td>
<td>Neurofilaments H, M &amp; L</td>
<td>(23)</td>
</tr>
<tr>
<td>Transcription Factors</td>
<td></td>
<td>Rotavirus NS26</td>
<td>(25)</td>
</tr>
<tr>
<td>65 kDa Nuclear tyrosine phosphatase</td>
<td>(24)</td>
<td>67 kDa RBC kinase</td>
<td>(28)</td>
</tr>
<tr>
<td><em>Aplasia</em> neuron proteins</td>
<td>(26,27)</td>
<td>Synapsin I</td>
<td>(30)</td>
</tr>
<tr>
<td>v-Erb-a Oncoprotein</td>
<td>(29)</td>
<td>92 kDa smooth ER protein</td>
<td>(32)</td>
</tr>
<tr>
<td>c-myc Oncoprotein</td>
<td>(31)</td>
<td>Estrogen Receptor</td>
<td>(Jiang &amp; Hart, unpub.)</td>
</tr>
<tr>
<td>(murine, bovine, human)</td>
<td></td>
<td>Bovine Lens α-crystallins</td>
<td>(33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(small heat shock proteins)</td>
<td></td>
</tr>
<tr>
<td>p43/hnRNP G canine autoantigen</td>
<td>(34,35)</td>
<td>Many Schistosome proteins</td>
<td>(36)</td>
</tr>
<tr>
<td>Adenovirus Fiber protein</td>
<td>(37)</td>
<td>Talin</td>
<td>(38)</td>
</tr>
<tr>
<td>HCMV UL32 (basic phosphoprotein)</td>
<td>(39,40)</td>
<td>Baculovirus gp41 tegument protein</td>
<td>(41)</td>
</tr>
<tr>
<td>tegument protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Many Chromatin Proteins of <em>Drosophila</em></td>
<td>(42)</td>
<td>Many Trypanosome proteins</td>
<td>[Kelly&amp;Hart, unpublished]</td>
</tr>
</tbody>
</table>

Addition and removal of O-GlcNAc to proteins appears to be highly-dynamic, analogous to phosphorylation (17,43,44). Mitogen activation of murine T-lymphocytes results in rapid and transient changes in the levels of O-GlcNAc on both nuclear and cytoplasmic proteins (43). Pulse-chase studies on human cytoskeletons show that O-GlcNAc turnover many-fold more rapidly than the peptide backbone (44). Addition and removal of O-GlcNAc is also regulated during the cell cycle on many proteins, including cytoskeletons (17) and nuclear pore proteins (Kelly, Roquemore and Hart, unpublished). Several in vivo sites of attachment of O-GlcNAc have been mapped (Table 2) While no consensus sequence is obvious at the primary sequence level, many of the sites are very similar to those same sites used by the proline-directed kinases that regulate much of the cell's metabolism (45,46). This similarity in site utilization has led to the hypothesis that O-GlcNAc is a regulatory modification that often competes for the same hydroxyl moieties used by kinases. In fact, in some examples studied, such as the CTD of RNA Pol II (14) and the human cytoskeletons (17,44) reciprocal relationships between glycosylation and phosphorylation have been clearly documented. O-GlcNAc attachment sites also exhibit high 'PEST' scores (47-49) which may mean that O-GlcNAc is involved in regulating the targeting of proteins for degradation.

© 1995 IUPAC, Pure and Applied Chemistry 67, 1637–1645
TABLE 2: Identified In Vivo Sites of O-GlcNAc

<table>
<thead>
<tr>
<th>Protein</th>
<th>Glycosylated Peptide (With a PV(S/T) motif underlined)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Erythrocyte 65 kDa</td>
<td>DSPVSQPSLVGSK..</td>
<td>(50)</td>
</tr>
<tr>
<td>Bovine Lens a-A-crystallin</td>
<td>158DIPVSREKH..</td>
<td>(33)</td>
</tr>
<tr>
<td>Human Serum Response factor</td>
<td>302YLAPYSVSPSAYSA318.</td>
<td>(22)</td>
</tr>
<tr>
<td>Rat Neurofilament (NF-L)</td>
<td>43YSAPVSSSLSVR..</td>
<td>(23)</td>
</tr>
<tr>
<td>HCMV (UL32) BPP</td>
<td>914PPSVPVSGSAPGR927.</td>
<td>(40)</td>
</tr>
<tr>
<td>Rhesus monkey a-B-crystallin</td>
<td>164EEKPAVTTAAPK174.</td>
<td>(33)</td>
</tr>
<tr>
<td>Neurofilament (NF-M) from Rat spinal cord</td>
<td>44GSPSTVSSSYK..</td>
<td>(23)</td>
</tr>
<tr>
<td></td>
<td>427QPSVTISSK..</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>Glycosylated Peptide (No PV(S/T) motif)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talin from chicken gizzard</td>
<td>1475MAXQLNVDPAXTQ1488</td>
<td>(38)</td>
</tr>
<tr>
<td></td>
<td>1886NQLTDYGOLAQ1889</td>
<td></td>
</tr>
<tr>
<td>RNA Pol II from calf thymus</td>
<td>(S/T)PS(S/T)SP..</td>
<td>(14)</td>
</tr>
<tr>
<td>Human Serum Response Factor</td>
<td>269VTLNPGLTSTTAPSTT289 280TQTSSGTVTLPATIM305</td>
<td>(22)</td>
</tr>
<tr>
<td>Rat Nuclear Pore p62</td>
<td>MAGGAPADTDPL</td>
<td>(7)</td>
</tr>
<tr>
<td>Human Erythrocyte Band 4.1</td>
<td>AQTITSETPSSTTT..</td>
<td>(13)</td>
</tr>
<tr>
<td>Rat Neurofilament (NF-L)</td>
<td>18YVETPRVHLSVR30.</td>
<td>(23)</td>
</tr>
<tr>
<td>HCMV (UL32) BPP</td>
<td>935STTPYAVTVPPSSTAK..</td>
<td>(40)</td>
</tr>
</tbody>
</table>

Enzymes of O-GlcNAc Addition/Removal

Using synthetic peptides based upon the sites mapped in vivo (Table 2), we have identified (51) and purified to homogeneity (52) a UDP-GlcNAc:polypeptide N-acetylglucosaminyltransferase (O-GlcNAc transferase) from rat liver cytosol. Some of the properties of this enzyme include: (1) The active site of the enzyme has been localized to the cytosol based on latency studies. (2) UDP-GlcNAc has been demonstrated to be the nucleotide sugar donor with a Km of about 545 nM, while UDP, UTP and UDP-GlcNAc, and to a 100-fold lesser extent, UMP and UDP-GalNAc have been shown to inhibit the transfer from UDP-[3H]GlcNAc. (3) The enzyme appears to contain 2 subunits of Mr of 110 (α-subunit) and 78 kDa (β-subunit) after purification of greater than 30,000 fold. (3) The holoenzyme is very large, based on gel filtration and sedimentation data, with an apparent molecular weight of 340 kDa, thus suggesting a heterotrimer of α2β configuration. (4) Photoaffinity labeling studies with 4-(β-32P)thio-UDP indicated that the α-subunit was likely to contain the active site. Specificity of binding by 4-thio-UDP to the active site was confirmed by competition with cold UDP. Recent studies have also shown that both subunits of the enzyme are tyrosine phosphorylated (Blomberg, Kreppel and Hart, unpublished). Differential substrate activity in homogenates from various tissues suggests that multiple forms of the O-GlcNAc Transferase with distinct peptide specificity’s exist.

A neutral O-GlcNAc-specific, β-D-N-acetylglucosaminidase (O-GlcNAcase) activity from rat spleen cytosol has also been identified (53) and recently purified and characterized (54). O-GlcNAcase is distinguished from lysosomal hexosaminidase by several features: 1) pH optimum of 6.4; 2) inactivity towards GalNAc; and 3) a higher relative activity to O-GlcNAc bearing peptides.
The O-GlcNAcase is effectively inhibited by various analogs of GlcNAc including 1-amino-GlcNAc, 1-azido-GlcNAc, LOGNAC ($k_i = 1.7$ mM) and PUGNAC ($k_i = 52$ nM). It is likely that the excellent inhibitory activities of these analogs will be important tools when studying the dynamics of O-GlcNAc on various cellular proteins.

**Nuclear Pore Proteins Bearing O-GlcNAc**

Nuclear envelope is particularly enriched in O-GlcNAc-bearing proteins (55). Most of the monoclonal antibodies originally prepared against nuclear pore proteins (nucleoporins) (56) have O-GlcNAc as a large part of their epitopes. Nucleoporins are involved in the transport of macromolecules into and out of the nucleus (57,58). Microinjection of a lectin, WGA, which binds GlcNAc, (59) or a monoclonal antibody to O-GlcNAc (56) prevents both import of proteins and export of RNA from the nucleus (60). Steric problems associate with such a large lectin (or antibody) were ruled-out since WGA appeared to only block proteins translocation with little effect on binding of proteins to the pore complex (61). Reconstitution experiments have shown that the nucleoporins are essential for nuclear transport but not the gross structure of the nuclear pores themselves (62).

**Chromatin Proteins Bearing O-GlcNAc**

Early work showed that chromatin contained a large number of O-GlcNAc bearing proteins (55). In fact, WGA binding and galactosyltransferase labeling studies of *Drosophila* embryo polytene chromosomes have shown that O-GlcNAc is highly abundant along the entire length of the chromosomes (42). Furthermore, the O-GlcNAc moieties appear to be particularly concentrated in condensed regions of the chromatin resulting in a banded pattern, while much less O-GlcNAc is present in ‘puff’ regions which have been associated with areas of active transcription.

Recently, RNA polymerase II (RNA Pol II) (14) and all of its transcription factors, thus far investigated (19-22), have been shown to be modified by O-GlcNAc. Glycosylated transcription factor Sp1 has been reported to be more transcriptionally active than the non-glycosylated form (19). Glycosylated Sp1 appears to play a role in insulin-dependent stimulation of growth factor TGFα (63), and Sp1 appears to be differentially glycosylated in a cell-type specific manner (64).

The catalytic subunit of RNA Pol II contains a highly conserved domain at the carboxyl-terminus consisting of up to 52 repeat units with consensus sequence -(Tyr-Ser-Pro-Thr-Ser-Pro-Ser)-[for review see (65). Genetic analysis has demonstrated that this C-terminal domain (CTD) is required for cell viability. Furthermore, various forms of the RNA Pol II subunits have been demonstrated in vivo based the phosphorylation state of the CTD and on its mobility by SDS-PAGE. These include $\Pi_0$ (Mr 240), which is the highly phosphorylated over the entire repeat region of the CTD and is associated with the transcription complex during the transition from initiation to elongation (66), and $\Pi_A$ [Mr 215 kDa] the non-phosphorylated form found in the preinitiation complex (67). We have shown that the CTD of $\Pi_A$, but not $\Pi_0$, is extensively modified by O-GlcNAc over the entire conserved repeat region (14). Thus, the phosphorylation and glycosylation of the CTD appear to be mutually exclusive events suggesting a high degree of regulation. We hypothesize that the O-GlcNAc modified form ($\Pi_A$) is involved in the formation of the initiation complex, possibly by oligomerization with O-GlcNAc modified transcription factors via lectin-like protein(s) or by carbohydrate-carbohydrate interactions. After formation of the initiation complex, it appears that the RNA polymerase IIa is rapidly deglycosylated and then rapidly phosphorylated to form the IIo form involved in transcriptional elongation. Fig. 2 illustrates this model of the putative role of O-GlcNAc in transcriptional initiation. The putative transcriptional regulatory role for O-GlcNAc is presently under active investigation.

In addition to ‘normal’ transcription factors, nuclear oncogene proteins, such as c-myc (31) and v-Erb-a (29) are also modified by O-GlcNAc. Interestingly, c-myc appears to be predominantly modified in its transactivation domain, suggesting a role for the saccharide in its action as a transcription factor activating neoplastic growth.

Several other nuclear proteins have been demonstrated to contain O-GlcNAc modifications (Table 1) including bovine and mouse estrogen receptor (Jiang & Hart, unpublished), *Aplasia* neuron 83 kDa protein (26,27) and canine autoantigen p43 (34,35). For the estrogen receptor, considerable evidence is now available that phosphorylation may be crucial to the regulation of estrogen-responsive
Glycosylation of nuclear and cytoplasmic proteins

promoters (68). Here again, like other O-GlcNAc bearing proteins, the glycosylation may act as an antagonist to phosphorylation and thereby also have a role in gene expression. It is also worth noting that the 83 kDa protein from Aplasia neurons is prominent in both the axon and the nucleus (26). Given that a pathway was recently discovered in neurons that can transport proteins from the axon to the nucleus (27), it will be interesting to know whether this glycoprotein is involved with the transport mechanism.

Cytoskeletal Glycoproteins

Many of the cytosolic O-GlcNAc modified proteins that have been identified are components of the cytoskeleton. Human erythrocyte Band 4.1 (50) was among the first to be characterized. Band 4.1 is involved in maintaining the unique shape of erythrocytes by anchoring actin and spectrin to the cytoplasmic tail of glycophorin. Recently, cytokeratins 13 (15), 8 and 18 (16) were shown to contain O-GlcNAc. Both phosphorylation and O-GlcNAc modification are enriched during mitotic arrest (17). Cytokeratins are a class of intermediate filaments found mostly in epithelial cells (69). Cytokeratin assembly is known to be regulated by phosphorylation (70). The relationship between cytokeratins glycosylation and phosphorylation is not yet well-understood. We recently have shown that neurofilaments from rat and mouse spinal cord are multiply glycosylated. The O-GlcNAc modifications on neurofilaments were localized primarily in the head domain of both neurofilaments M and L (23). The head domain is a region of the proteins that had already been implicated as being required for proper neurofilament assembly by deletion analysis (71,72). Site directed mutagenesis of the sites of O-GlcNAc modification are presently underway to address whether the glycosylation is required for assembly of these filaments.
PHF-Tau is characteristically present in the brain cells of patients with Alzheimer's disease, but not in the brain cells of normal individuals. The finding that Tau is normally glycosylated, suggests that its abnormal phosphorylation in Alzheimer's disease could be related to a defect in addition or removal of O-GlcNAc. Fig. 3 shows a model of normal and PHF-Tau. Fig. 4 illustrates the phosphorylation sites on Tau. Note how they are similar to the O-GlcNAc sites described in Table 2. Mapping of the sites on the Tau protein is underway.

Finally, the cytoskeletal protein talin has also been reported to be modified by O-GlcNAc (38). This important cytosolic protein appears to provide a bridge between the cytoplasmic domain of integrins and the cytoskeleton by interaction with another cytosolic protein, vinculin. Interestingly, non-glycosylated talin derived from platelets does not interact with vinculin, suggesting that O-GlcNAc may be necessary for the interaction of these cytoskeletal components (38).

O-GlcNAc and Protein Synthesis

Recently a 67 kDa O-GlcNAc-modified glycoprotein (p67) has been described that can bind to eIF-2 and protect it from phosphorylation by eIF-2 kinases and thereby maintain protein synthesis initiation (28,73). When p67 was removed from the cell extracts with the lectin, wheat germ agglutinin, or with antibodies to p67, phosphorylation was no longer blocked and the kinases could readily transfer phosphate to the α-subunit of eIF-2. Using antibodies to detect all p67 proteins, and a monoclonal antibody that reacts only with glycosylated p67, it was shown that p67 was first deglycosylated within 10 hrs, then degraded after 16 hr of serum starvation, while the levels of the α and β-subunits of eIF-2 remained constant. It was concluded that p67 protected the α-subunit from
phosphorylation under normal conditions, but during starvation, p67 is rapidly deglycosylated then degraded, which allows the eIF-2 kinases to phosphorylate the eIF-2 and prevent proteins synthesis initiation (74). Reticulocyte lysates contain a 'deglycosylase' (O-GlcNAcase) that remains in a latent form in the presence of hemin, however, when hemin is absent, the deglycosylase is activated to remove O-GlcNAc from p67 and begin the cascade toward inhibition of proteins synthesis (Gupt et al., pres. at Soc. Complex CHO, 1993). These provide direct evidence for a role of O-GlcNAc in protein synthesis initiation.

Viral Proteins

O-GlcNAc-modifications have been demonstrated on a number of viral proteins (Table 1). The fiber proteins of adenovirus is known to form mature trimeric structures that are involved in virus attachment to the host cell surface (75). The O-GlcNAc in the mature trimeric structures is inaccessible to labeling with GT unless the fibers are denatured with detergents, indicating that the O-GlcNAc moieties are buried in the trimeric structures (37). A similar observation has been made for neurofilament assembly, suggesting that O-GlcNAc may also be involved with the formation of these multimeric fiber structures.

O-GlcNAc is also found on the major tegument proteins (the region between the viral capsid and the viral envelope) of human cytomegalovirus (HCMV) (39,40) and baculovirus (41). The function of the basic phosphoprotein (BPP, UL32) of HCMV is presently unknown, but its location in the tegument region of the virus suggests that it might act as a signal for final envelopment of the capsid. O-GlcNAc could be the signal for oligomerization of the BPP or the means of attachment to the cellular compartment used for viral envelopment. The demonstration that gp41 of baculovirus contains O-GlcNAc, made it clear that insect cells are fully capable of adding this modification to proteins. The finding of O-GlcNAc in baculovirus overexpression systems, has proven valuable to the study of the glycosylation of many low abundance proteins that contain O-GlcNAc including transcription factors and oncogenes. Importantly, the BPP from HCMV appears to be glycosylated at the same sites when expressed in insect cell as a recombinant baculovirus or when isolated from HCMV virions, although the stoichiometry at each site appeared to be slightly different (40). These results are also consistent with those found for human cytokeratins 8 and 18 expressed in baculovirus (18).

Fig. 4 Schematic of the Proline-Directed Kinase Sites on the Tau Protein.
Summary and Conclusions

At present, it appears that many of the cell's regulatory and cytoskeletal proteins are transiently glycosylated by O-GlcNAc. Circumstantial evidence suggests that the saccharide is mediating the assembly of these polypeptides into functional complexes. Direct evidence on several proteins suggests that O-GlcNAc plays a reciprocal role to phosphorylation. O-GlcNAc modification of transcription factors and RNA polymerase II appears to directly regulate transcriptional initiation. We have assembled the tools for elucidating the functions of this novel post-translational modification. In the near future detailed data with respect to its functions at the molecular level should be forthcoming.

REFERENCES

Glycosylation of nuclear and cytoplasmic proteins