Biochemistry

 Volume 42, Number 9 March 11, 2003

Current Topics

Dynamic Nuclear and Cytoplasmic Glycosylation: Enzymes of O-GlcNAc Cycling[†]

Sai Prasad N. Iyer[‡,§] and Gerald W. Hart*[,‡]

Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205-2185, and Graduate Program, Department of Biochemistry and Molecular Genetics, The University of Alabama at Birmingham, Birmingham, Alabama 35294

Received December 3, 2002; Revised Manuscript Received January 9, 2003

Post-Translational Modification of Proteins by O-GlcNAc. For more than the past decade and a half, post-translational modification of Ser/Thr residues of nuclear and cytoplasmic proteins by the O-linked *N*-acetylglucosamine (O-GlcNAc[3]) monosaccharide has been shown to be a ubiquitous and dynamic form of protein modification (extensively reviewed in ref *1*). This post-translational modification, often referred to as O-GlcNAcylation, is found in all higher eukaryotes that have been examined and is dynamic, often in response to extracellular stimuli (*2*). Proteins modified by O-GlcNAc are vast in form and function; they include RNA polymerase II and its associated transcription factors (*3*−*5*), viral (*6*), oncoproteins (*7*), cytoskeletal (*8*, *9*), signal transduction (*10*), and translational machinery (*11*, *12*) proteins, as well as enzymes such as phosphatases (*13*) and kinases (*14*). O-GlcNAcylated proteins have been shown to form reversible multimers in large protein complexes and are also phosphorylated. The sites of O-GlcNAc modification are often the same or adjacent to sites of Ser/Thr phosphorylation, suggesting a role in regulation analogous to or competitive with phosphorylation (*15*). Thus, it is hypothesized that O-GlcNAc regulates the functions of proteins, either exclusively or in concert with phosphorylation. Enzymes that catalyze the cycling of O-GlcNAc onto and off proteins, analogous to those that add and remove phosphates (i.e., kinases and phosphatases), support this hypothesis. A unique nuclear and cytoplasmic glycosyltransferase, O-GlcNAc transferase (OGT), that catalyzes the addition of O-GlcNAc on proteins has been purified (*16*), cloned, and characterized (*14*, *17*−*19*). Likewise, its counterpart, an O-GlcNAc specific *β*-*N*-acetylglucosaminidase known as O-GlcNAcase that removes the monosaccharide, has also been purified, cloned, and characterized (*20*−*22*). Both enzymes are highly conserved throughout evolution from *Caenorhabditis elegans* to *Homo sapiens* (*17*, *18*, *21*). The O-GlcNAc modification was shown to be vital for life since targeted deletion of the OGT gene results in embryonic stem cell lethality in mice (*23*). Other forms of cytosolic glycosylation that may exist include O-fucosylation, O-glucosylation, and mucin type O-glycosylation, and these have been extensively reviewed elsewhere (*65*, *66*).

While much is known about the characteristics of O-GlcNAcylation, the overall and specific function(s) of the modification remain elusive. Since comprehensive reviews on the O-GlcNAc modification and its cycling are present in the literature (*1*, *24*), this review will focus on the enzymes that cycle O-GlcNAc addition and removal (i.e., the OGT and O-GlcNAcase).

O-GlcNAc Transferase—*The Enzyme.* In 1990, an activity assay for monitoring the transfer of GlcNAc from a UDP-GlcNAc sugar nucleotide donor onto peptide substrates was developed by Haltiwanger et al. (*25*). This assay was subsequently used to purify the O-GlcNAc transferase. The O-GlcNAc transferase (uridine diphospho-*N*-acetylglucosamine:polypeptide *β*-*N*-acetylglucosaminyltransferase, EC 2.4.1) or OGT was purified to near homogeneity by Halti-

[†] Supported by NIH Grants CA42486 and HD13563 to G.W.H.

* Corresponding author. Phone: (410) 955-5993. Fax: (410) 614-8804. E-mail: gwhart@jhmi.edu.

[‡] The Johns Hopkins University School of Medicine.

[§] The University of Alabama at Birmingham.

[1] Abbreviations: O-GlcNAc, O-linked *N*-acetylglucosamine; UDP-GlcNAc, uridine diphospho-*N*-acetylglucosamine; OGT, O-GlcNAc transferase; O-GlcNAcase, *β*-*N*-acetylglucosaminidase; TPR, tetratricopeptide repeat; CTD, carboxy terminal domain of RNA polymerase II; GABA, *γ*-amino butyric acid; GRIF-1, $GABA_A$ receptor associated protein; OIP106, OGT interacting protein of 106 kDa.

10.1021/bi020685a CCC: $25.00
Published on Web 01/31/2003

wanger et al. using conventional and affinity chromatography from rat liver (*16*). The rat liver enzyme was found to be a holoenzyme consisting of two 110 kDa subunits and one 78 kDa subunit, migrating with an apparent molecular weight of 340 kDa (*16*), indicating that the enzyme likely existed in a heterotrimer complex. It was found that the OGT had an unusually low apparent K_m (0.545 uM) for UDP-GlcNAc (*16*), with the resulting UDP product strongly inhibiting activity (greater than 50% loss at less than 0.5 uM) (*24*). Radiolabeling studies performed with a noncleavable specific photoaffinity probe revealed the 110 kDa subunit to contain the enzyme's active site. Using conventional cloning techniques, the cDNA for the 110 kDa subunit was cloned and characterized (*17*, *18*).

OGT—Gene and Primary Structure. Cloning and assembly of the cDNA of the 110 kDa OGT subunit from rat liver revealed it to be a polypeptide of 1037 amino acids (*17*). Simultaneously, the *C. elegans* and a partial human OGT sequence was also cloned and published (*18*). At the time of initial publication, the published human OGT sequence was shown to be smaller than the rat or worm sequences, with 921 amino acids. However, recent fine mapping of the OGT gene (*26*) revealed that the human OGT gene is actually 1036 amino acids, indicating that the initial published sequence was a partial one. Comparison of the complete human and rat sequences show that both sequences are virtually identical (>99% identity). Expression of the OGT cDNA in CHO-K1 and HeLa cells resulted in corresponding increases in OGT activity, indicating that the cloned cDNA indeed encoded for the OGT enzyme (*17*, *18*). Interestingly, antibodies raised against the recombinant OGT also reacted with the 78 kDa subunit, indicating that it was either a proteolytic fragment or derived via an alternative start site from the OGT cDNA. While the 110 kDa is ubiquitously expressed in most all tissues that have been examined so far to date, the 78 kDa fragment has limited tissue distribution (*17*), suggesting that there might be a unique relationship between the 110 kDa and the 78 kDa, perhaps in a regulatory manner, analogous to the p110 and p85 subunits of PI-3 kinase. OGT is expressed in high abundance in tissues such as brain (mRNA and protein) and pancreas (mRNA), relative to other tissues that have been examined (*17*, *18*). The enzyme bears post-translational modifications itself; it is tyrosine phosphorylated and is itself modified by O-GlcNAc (*17*).

Analysis of the OGT sequence revealed a very distinct bi-modular nature to the enzyme. The amino terminal half of both rat and human OGT consists of 11.5 tandem tetratricopeptide repeats (TPR), whereas the worm sequence consists of 13 TPRs (*17*, *18*). TPRs or TPR domains, in proteins that contain them, are protein−protein interaction domains (*27*, *28*). The role of the TPR domain of OGT will be reviewed in detail below.

Initially, the carboxy terminal half of the OGT was found to be unique, with no homology to any known protein. Recently, however, alignment of the carboxy termini of OGT sequences from different species predicts that that it adopts a fold that consists of two Rossman-like domains (*29*). Furthermore, a conserved acidic residue in one of the helices of the second domain identifies this region as a putative UDP-GlcNAc binding site (*29*). While there is some preliminary evidence of the carboxy terminus containing the UDP-GlcNAc binding site (Kreppel, L. K., and Hart, G. W., unpublished observations), definitive evidence identifying the exact residue(s) involved in the binding is still lacking. Of note, however, is the observation that deletion of the extreme C-term from residue 944 to 1037 results in a catalytically inactive OGT (*14*, *30*), adding to the notion that the active site indeed resides in the carboxy terminus. Interestingly, these analyses of the carboxy terminus of OGT place it into the superfamily of glycogen phosphorylase/glycosyl transferase (GPGTF) proteins (*29*), some of whose members include glycogen phosphorylase, maltodextrin phosphorylase, and bacterial UDP-GlcNAc 2-epimerase. However, definitive structural evidence such as X-ray crystallography or NMR solution structures remain to be performed to confirm the in silico analyses.

Southern blot analysis of the OGT gene shows it to be highly conserved throughout evolution in higher eukaryotes, from the nematode to man (*17*). While the modification and the enzyme have not been found in *Escherichia coli*, weak homologues have been found in archea bacteria and methanobacteria. While O-GlcNAc modified proteins have been found in the filamentous fungi *Aspergillus oryzae* (*31*), it is not clear whether this activity is due to the cloned OGT (or its homolog) or an unrelated O-GlcNAc transferase. However, to date, no clear OGT homologue has been found in the yeast *Saccharomyces cerevisiae* or *S. pombe*. Thus, while the evidence of the O-GlcNAc modification or the OGT enzyme in yeast is lacking, we cannot disprove its occurrence. It is possible that there may be other monosaccharides, such as O-mannose, that may play the same role as O-GlcNAc does on proteins. Alternatively, it is also entirely plausible that there may be an OGT in yeast that is not related to the cloned OGT on the primary level. There is one example of this in the literature. The α-toxin from *Clostridium novyi* has been shown to be an O-GlcNAc transferase that shows specificity for the Rho subtype proteins (*32*). Analysis of the α-toxin sequence shows very little homology to the cloned OGT sequence; yet it is an OGT. Therefore, there might be a similar situation in yeast *S. cerevisiae* or *S. pombe*.

To date, there has been only one OGT gene that has been described in mammals (*17*, *18*, *26*). The OGT gene was targeted in mouse and deleted using a Cre recombinase based approach (*23*). Deletion of the OGT locus in mice resulted in embryonic lethality, down to the single cell level, indicating that the enzyme is essential for life (*23*). Tissue specific targeted deletion of the OGT locus resulted in developmental ablation of those targeted organs, indicating the further importance of OGT and the O-GlcNAc modification in development (O'Donnell, N., personal communication). The *OGT* gene has been mapped to the X chromosome in mice and humans (Xq13.1) via radiation hybrid PCR and FISH (*23*, *26*). This region is of significance since the loci for X-linked dystonia Parkinsonism (DYT3) also maps to Xq13.1 (*26*). The potential involvement of OGT in neurological disorders is not without precedence since numerous proteins involved in neurological disorders have been shown to be extensively modified by O-GlcNAc, such as the microtubule associated protein tau (*33*) and β-amyloid precursor protein (*34*), both of which have been implicated in the etiology of Alzheimer's disease.

While there is only one reported OGT gene in mammals, two OGT genes have been reported to exist in plants. The *Arabidopsis thaliana* Spindly (SPY) protein is an important

component of the gibberellin (GA) signaling pathway (*35*−*37*), and it has been shown to have a significant level of homology to OGT (*36*, *38*). Mutations in SPY suppress the effects of GA deficiency, and it is thought that SPY negatively regulates GA signaling (*35*−*37*, *39*). Whether SPY is a valid OGT and possesses OGT activity is unclear since the evidence that SPY is an OGT are preliminary (*40*). Interestingly, SPY mutants show reduced levels of O-GlcNAcylated proteins (*41*), indicating that there are possibly additional OGTs in plants. This led to the isolation and cloning of another gene called Secret Agent (SEC), which shows considerable resemblance to both SPY and OGT (*42*). Bacterially expressed and purified recombinant SEC has been shown to be modified by O-GlcNAc, suggesting that it autoglycosylates itself (*42*). However, this may be a result of endogenous *E. coli* glycosyltransferases modifying SEC. Therefore, SEC's ability to O-GlcNAcylate protein or peptide substrates remains to be demonstrated. Interestingly, while mutations in SEC did not exhibit obvious phenotypes, SEC mutants exhibited synthetic lethality when combined with mutations in SPY (*42*), indicating that both proteins have overlapping functions that are indispensable for life in plants, consistent with the role of OGT in animals.

OGT−*TPR Domain*—*Roles in Multimerization, Substrate Specificity, and Interacting Proteins.* TPRs are composed of 34 amino acid repeats that contain the consensus sequence W−L−G−Y−A−F−A−P, which is often loosely conserved (*43*, *44*). These repeats often occur in tandem or separately and modulate protein−protein interactions either in serial or random combinations of TPRs. Different combinations of TPRs can result in different combinations of binding partners, resulting in different functional consequences (*43*, *44*). The TPR domains of protein phosphatase 5 (PP5) has been solved and reveal that each repeat forms a pair of antiparallel α-helices, acting as superhelical structures in protein−protein interactions (*45*).

Systematic deletion of individual sets of TPRs of the OGT have clearly shown that one of the key functions of the TPR domain is to facilitate subunit−subunit interaction (*19*). Recombinant rat OGT, expressed and purified from baculovirus, exhibits identical kinetic parameters and enzymology as the native enzyme (*19*). The recombinant enzyme is a homotrimer, and this trimerization is mediated by the TPR domain. Deletion of the first five and a half TPRs results in the enzyme's inability to assemble into a trimer, indicating that the first six TPRs (Δ6 OGT) contribute to subunit−subunit association (*19*). Deletion of the first three TPRs (Δ3 OGT) results in the enzyme's ability to oligomerize but in a salt dependent manner. Under high salt conditions (1 M NaCl), the Δ 3 OGT enzyme dimerizes, and under physiological conditions (150 mM NaCl), it trimerizes (*19*). Similar results were obtained by Yang et al. (*30*), where they demonstrated that TPRs 2−6 were fully capable of binding full-length OGT.

Interestingly, prevention of the multimerization of OGT by deletion of the first six TPRs did not affect its activity toward certain small peptide substrates. Similarly, both the dimer and the trimer forms of the enzyme exhibited similar kinetic properties (*19*). However, while the activity of the Δ6 OGT toward a 12mer casein kinase II (CK II) acceptor peptide and several other short synthetic peptides was unaffected relative to wild-type full-length enzyme, it's activity was dramatically reduced toward a 10 repeat (70mer) RNA polymerase II carboxy terminal domain (CTD) repeat peptide (CTD_{10}) (*46*). Interestingly, the CTD peptide must be at least 10 repeats long to be a good substrate for OGT; lesser repeats are either not substrates at all ($CTD_{1,2\ or\ 3}$) or are very poor substrates (CTD_5), indicating that the glycosylation of CTD repeats may involve a unique mechanism (*46*). The CK II 12mer peptide is considerably shorter than the CTD_{10} 70mer peptide; therefore, it is possible that larger peptide substrates may require multiple points of contact with multiple TPRs to be efficiently glycosylated by the enzyme. This may hold true for protein substrates as well. Similar studies were performed with the recombinant partial human OGT enzyme expressed in *E. coli* (*14*). However, studies performed in this report are problematic since these authors were unaware that they were working with a mutant form of the enzyme since their original clone lacked two and a half TPRs. Thus, these studies are considered inconclusive. The TPRs may also play a role in mediating OGT's own post-translational modifications as well. While the full-length recombinant OGT (expressed in baculovirus) and Δ6 OGT are tyrosine phosphorylated, the Δ9 OGT TPR deletion is not phosphorylated, indicating that TPRs 7−11 may contain the site of tyrosine phosphorylation (*19*). The glycosylation of the TPR deletion mutants (up to 9 TPRs) is unaffected, and all mutants are efficiently O-GlcNAc modified (*19*). Thus, the multimerization of OGT mediated by the TPRs affect substrate specificity toward certain protein substrates and also affect its post-translational modifications. We have proposed that the specific combinations of TPRs may function as docking sites for protein and large peptide substrates, both targeting and controlling the specificity of the catalytic subunit (Iyer, S. N., and Hart, G. W., unpublished observations, ref *59*).

Another important role of the TPR domain of the OGT is to mediate protein−protein interactions with other interacting proteins. Two recent reports of OGT interacting proteins have been described. Kudlow's group have reported that the co-repressor protein mSin3A interacts with OGT (*30*). It was shown that TPRs 1−6 mediate the binding of OGT to the PAH4 domain of mSin3A (*30*). mSin3A is found complexed with histone deacetylases and is thought to mediate transcriptional repression by recruitment of histone deacetylases (*47*, *48*). This discovery was made partly on the basis that a Gal4-TPR 1−6 fusion protein was able to repress transcription from a Gal4 responsive reporter. By binding to OGT, it is thought to recruit a subset of OGT to mediate repression of target genes, in concert with histone deacetylation, by O-GlcNAc modification of transcription factors and transcriptional machinery proteins, leading to their inactivation. This would result in efficient gene repression and silencing.

We have recently reported the identification and cloning of a novel family of coiled−coil domain proteins that interact with OGT (*59*) isolated by a yeast-two hybrid screen. The members of this family are GRIF-1 ($GABA_A$ associated protein) and OIP106 and their various species orthologs. Both proteins have highly identical amino terminal halves, with loosely conserved carboxy termini. Both proteins interact with the TPR domain of the OGT and bind OGT stoichiometrically (*59*). In addition, both proteins are also modified by O-GlcNAc, indicating that they are also substrates for OGT. However, Nup62, which is heavily modified by

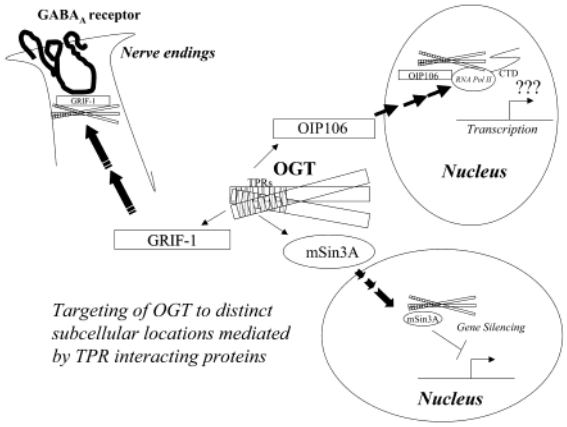

FIGURE 1: Targeting of OGT via TPR interacting proteins**.** By binding to proteins such as mSin3A, OGT would be targeted to the nucleus for gene silencing events, in concert with histone deacetylases. Interactors such as the OIP106/GRIF-1 family of proteins would target OGT to the nucleus and to GABA$_A$ receptor complexes. In the nucleus, OIP106 would target OGT to Pol II containing complexes; however, it is unclear as to what effect this kind of targeting would have on transcription. TPRs 1−6 of the OGT have been implicated in interacting with at least these three known interactors, indicating that these TPRs could mediate these distinct events by forming specific complexes through different combinations of individual sets of repeats.

O-GlcNAc, does not stably interact with OGT, indicating that just being a substrate is not enough of a requirement to interact stably with the enzyme. Using a TPR deletion approach, TPRs 2−6 have been implicated in binding to OIP106 (Iyer, S. N., and Hart, G. W., unpublished observations). GRIF-1 was initially isolated and cloned as a novel protein that associates with the β2 subunit of the GABA$_A$ receptor (*60*) and was shown to be expressed only in excitable tissue such as brain, heart, and muscle. OIP106 displays a more ubiquitous expression pattern (*59*) and localizes to the nucleus in HeLa cells. Both confocal and electron microscopy showed that OIP106 co-localized with a subset of RNA polymerase IIa, and this was confirmed biochemically by showing the existence of an in vivo Pol IIa−OIP106−OGT protein complex (*59*). These findings are significant since it has been shown that RNA Pol II and its associated transcription factors are heavily modified by O-GlcNAc (*3*−*5*). OIP106 may mediate their glycosylation by targeting OGT to such complexes (*59*). GRIF-1 is thought to target OGT to GABA$_A$ receptor complexes in a similar manner.

Thus, in addition to OGT's enzymology, it is clear that the first 6 TPRs also influence its binding partners. We speculate that perhaps the multiple functions attributed to the first 6 TPRs could be facilitated by different combinations of individual sets of TPRs, with one combination of specific TPRs mediating multimerization, whereas another combination mediating interaction with the mSin3A complex and yet another mediating interactions between GRIF-1-GABA$_A$ receptor and OIP106−Pol IIa protein complexes (Figure 1). This has precedence, one example of which are the yeast cell cycle proteins Cdc16, Cdc23, and Cdc27 that all contain TPR domains. They all interact with each other via their TPR domains, and each specific interaction is mediated by specific sets of TPRs (*51*).

One important conclusion that was derived from the TPR deletion studies are that catalytic activity of the OGT is not dependent on the TPRs or its ability to multimerize. This

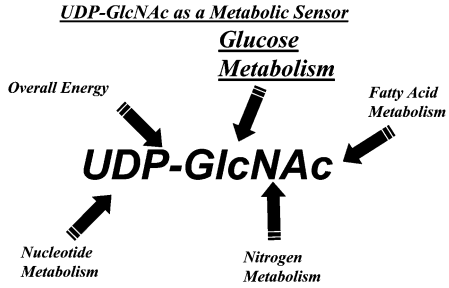

FIGURE 2: UDP-GlcNAc is an ideal metabolic sensor. Levels of UDP-GlcNAc in the cells are dramatically affected by all the indicated factors. Glucose metabolism is highlighted, in the context of this review, since 2−5% of glucose that is uptaken by the cell (depending on cell type) is converted into UDP-GlcNAc via the hexosamine biosynthetic pathway.

implicated the carboxy terminal half of the OGT as containing the active site and the UDP-GlcNAc binding site; a notion has been further supported by the in silico analysis reviewed above. Studies performed with the recombinant rat OGT revealed that the enzyme had three apparent K_m values for UDP-GlcNAc (6, 35, and 217 uM) (*19*). Furthermore, studies performed on synthetic peptide substrates demonstrated that the activity of OGT was clearly dependent on UDP-GlcNAc concentrations, with activity dramatically increasing with concentration (*19*). Further studies performed using the CTD$_{10}$mer as a substrate showed that the enzyme is never saturated with UDP-GlcNAc, even at concentrations as high as the limit of solubility (50 mM) (*49*). Interestingly, analysis of the kinetics reveal that the enzymology of the OGT (full-length and Δ6 OGT) is consistent with a random bi-bi mechanism. Therefore, the observation that it is never saturated with UDP-GlcNAc is very intriguing. It has been suggested that perhaps the enzyme goes through a covalent intermediate (Mildvan, A., personal communication). Regardless, it is apparent that OGT is very sensitive and responsive to even the slightest shifts in the concentrations of UDP-GlcNAc. In the cell, this is a very important physiological consequence since UDP-GlcNAc levels can rise and fall dramatically, especially in response to plasma glucose concentrations. Indeed, as summarized in Figure 2, UDP-GlcNAc is an ideal metabolic sensor. As much as 2−5% of glucose is metabolized to UDP-GlcNAc, depending upon the cell-type, via the hexosamine biosynthetic pathway (*63*, *64*). Furthermore, importantly, recent evidence has shown that elevated O-GlcNAc is a direct cause of insulin resistance, which is the hallmark of type II diabetes (*60*, *61*).

It is clear that the OGT is influenced by a variety of factors, allowing for very complicated and involved mechanisms of regulation. While much has been delineated about the enzyme, it is still not clear as to how the enzyme is exactly regulated. It is unclear as to how the post-translational modifications on the enzyme affect its function. It is likely that the discovery of mSin3A and the OIP106/GRIF-1 protein family as OGT interactors just represent the beginning of a possible long list of as of yet unidentified proteins that regulate OGT's function and localization in a similar manner, by selectively recruiting it to specific locations in the cell or by perhaps modulating its enzymatic activity. This is a reasonable speculation since we have already isolated over 30 unique cDNA clones potentially encoding for different polypeptides in our initial yeast two-hybrid screen (*59*). It

is also unclear as to what the role of the 78 kDa subunit is. Clearly, much work remains to be done.

O-GlcNAcase—The Enzyme. Several years ago, the O-GlcNAcase (*β-N*-acetylglucosaminidase, EC 3.2.1.52) was purified via conventional chromatography from rat spleen (*20*) and characterized. The purified enzyme was designated as a hexosaminidase C, based on its characteristics of having a neutral pH optimum, nucleocytoplasmic distribution, and selectivity for the removal of GlcNAc and not GalNAc (*20*). Subsequently, using a modified protocol based on the original purification scheme, the enzyme was purified from bovine brain, sequenced via mass spectrometry, and its cDNA was cloned from a human brain library (*21*). Several proteins that strongly co-purified with the enzyme were also identified, suggesting that these proteins co-purified by virtue of being strongly associated with O-GlcNAcase. These are the heat shock proteins hsp110 and hsc70, cullin, TATA binding protein 120 (TIP 120) calcineurin, amphiphysin, and dihydropyrimidinase-related protein 2 (DRP-2) (*21*). Subsequent studies ruled out hsp110, hsp70, cullin, and TIP120 as potential interacting proteins (*22*). Studies performed with recombinant *E. coli* expressed enzyme showed that the O-GlcNAcase is monomeric, migrating at its apparent molecular weight of ~130 kDa (*22*). This is significant since native O-GlcNAcase activity migrates at ~600 kDa (*22*), indicating that in the cell, the enzyme may be complexed with some of the above co-purifying proteins.

Confirmation of the cloned cDNA as the O-GlcNAcase enzyme was established by demonstrating that the recombinant protein had very high activity, including the ability to selectively remove GlcNAc, but not GalNAc from glycopeptides (*21*, *22*), with kinetic parameters identical to the native enzyme. Furthermore, overexpression of the cDNA in mammalian cells reduces O-GlcNAc levels on nuclear and cytoplasmic proteins (*22*). O-GlcNAcase is a 916 amino acid protein, with a predicted molecular weight of ~100 kDa, and an apparent molecular weight of 130 kDa. It is expressed in every tissue that has been examined, especially brain (*21*), and is conserved highly throughout evolution, with homologues in *C. elegans* to mammals. The O-GlcNAcase gene localizes to 10q24 (*21*, *50*). This locus maps to the late onset Alzheimer's disease gene (*52*−*54*). Therefore, like the OGT, it would seem that O-GlcNAcase also may directly be involved in neurological disorders.

Surprisingly, the cloned sequence did not show any homology to known glycosidases. Instead, it showed loose homology to a number of proteins, most notably hyaluronidases and acetyltransferases. This is interesting since earlier, the cloned O-GlcNAcase sequences was identified as MGEA5, meningioma expressed antigen 5 protein that induces an immune response in meningioma patients (*50*). Meese and co-workers (*50*) had shown that MGEA5 possessed weak hyaluronidase activity, indicating that O-GlcNAcase has dual functions. Recently, comparison of the O-GlcNAcase sequence to a variety of acetyltransferases places it in the GCN5-related family of acetyltransferases (GNAT) (*55*). It is also predicted that the GNAT homologous region in the C-terminal half contains the O-GlcNAcase activity. The putative weak hyaluronidase activity is predicted to reside in the N-terminal half. Thus, the O-GlcNAcase is predicted to have dual domain structure, separated by a linker sequence in the middle of the molecule.

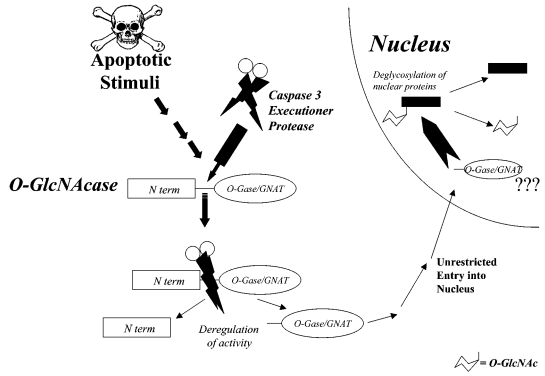

FIGURE 3: Model of O-GlcNAcase deregulation during apoptosis. In the apoptotic pathway, the O-GlcNAcase would get cleaved by the executioner protease caspase-3. The resulting ~65 kDa carboxy terminal fragment would then translocate to the nucleus, and then be free to deglycosylate key apoptotic proteins that may be involved in the pathway. How the putative acetylase activity of the cleaved O-GlcNAcase fragment may contribute to this pathway is unclear at this time (as symbolized by the question marks).

The homology of O-GlcNAcase to hyaluronidases is weak. On the basis of this and the prediction that it is a member of the GNAT superfamily, it is reasonable to suggest that perhaps it also may have acetyltransferase activity as well. This is a testable hypothesis, given that the recombinant O-GlcNAcase has been expressed and purified from *E. coli*. If indeed O-GlcNAcase possesses acetylase activity, this would not only make it a very unique multifunctional enzyme, but it would also fit in with the model of the role of O-GlcNAc in the transcriptional cycle that has been proposed (*46*). It has been shown that the IIA form of RNA polymerase II, found exclusively in preinitiation complexes, is extensively modified by O-GlcNAc in its carboxy terminal domain (CTD) that contains multiple repeats of the heptad sequence YSPTSPS (*3*). During transcriptional elongation, which follows promoter clearance, a set of enzymatic events occur that involve acetylation of histones to loosen and open chromatin and hyperphosphorylation of the CTD. The sites of glycosylation and phosphorylation on the CTD have been shown to be overlapping. Indeed, the presence of either modification blocks the occurrence of the other modification (i.e., glycosylation of the CTD inhibits the activity of the CTD kinase (*46*)), indicating that these modifications are mutually exclusive. Therefore, it can be envisioned that by possessing dual activities of a potential acetylase, as well as removal of the sugar, O-GlcNAcase may function as an integral and novel component in the transcriptional cycle.

O-GlcNAcase—Splice Variants, Localization, and Regulation. The O-GlcNAcase protein is expressed in two forms, the 130 kDa form and a splice variant that migrates at an apparent molecular weight of 75 kDa (*21*, *50*, *56*). Both variants are expressed ubiquitously in various human tissues and throughout mouse development (*56*). The 75 kDa splice variant is a result of an alternative stop codon, resulting in a protein lacking that C-terminal third of the full-length O-GlcNAcase. Meese and co-workers have characterized the 75 kDa splice variant as being a nuclear protein in the glioblastoma cell line TX3868 (*56*). Furthermore, the 130 kDa variant was shown to be mostly cytoplasmic, when overexpressed in mammalian cells (*21*) and in the TX3868 cell line (*56*). These results were further confirmed by performing immunofluorescence with a C-term specific

O-GlcNAcase antibody and localizing the endogenous enzyme predominantly to the cytoplasm in CHO-K1 cells (*22*). Interestingly, mild reactivity in the nucleus at distinct punctate sites was also noticed (*22*), indicating that a fraction of the enzyme localized to the nucleus. This nuclear reactivity is likely not to result from the 75 kDa since the smaller splice variant terminates before the extreme C terminal sequence used to generate the antibody. The biochemical localization correlates well with the O-GlcNAcase activity found in these subcellular fractions; ~90% of the activity resides in the cytoplasm, with ~10% activity found in the nucleus (*22*). It is known that the 75 kDa splice variant does not possess O-GlcNAcase activity (*22*), indicating that the carboxy terminus is necessary for activity. Currently, it is unclear as to what the role of the smaller splice variant is. Several hypotheses can be proposed, including potential roles of the 75 kDa as serving as a regulatory subunit, perhaps by competing for interacting proteins with the larger splice variant.

One of the post-translational modifications that occurs on O-GlcNAcase is cleavage by the apoptotic protease, caspase 3 (*22*). Caspase 3 is the executioner caspase in the apoptotic pathway (*57*). It has been shown that O-GlcNAcase is efficiently cleaved by caspase 3 generating a 65 kDa C term fragment (as ascertained by the C term specific antibody) (*22*). Most interesting is the observation that cleavage of the enzyme had no effect on its activity in vitro. This is not unprecedented since other caspase 3 substrates such as the phosphatase calcineurin (*58*) and PKC δ (*62*) also retain complete activity upon cleavage (*58*). It is reasonable to suggest that perhaps cleavage of the O-GlcNAcase into its individual fragments leads to dys-regulation of its enzymatic activity, a consequence that may be essential for the apoptotic process (see Figure 3). Since the proteolytic fragments are active (presumably the 65 kDa C term fragment), they may be unhindered to deglycosylate key proteins that may play a role during apoptosis.

While much is known about the OGT through structure−function studies that have been performed, much work needs to be done on the O-GlcNAcase. Very little is known about the regulation of the enzyme. Clearly, one area of great interest are the proteins that may interact with the enzyme and potentially regulate it. Another area of great importance is the role of the enzyme in neurological disorders, as well as in cancer and diabetes.

REFERENCES

1. Hart, G. W. (1997) *Annu. Rev. Biochem. 66*, 315.
2. Kearse, K. P., and Hart, G. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1701.
3. Kelly, W. G., Dahmus, M. E., and Hart, G. W. (1993) *J. Biol. Chem. 268*, 10416.
4. Jackson, S. P., and Tjian, R. (1988) *Cell 55*, 125.
5. Jackson, S. P., and Tjian, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1781.
6. Greis, K., Gibson, W., and Hart, G. W. (1994) *J. Virol. 68*, 9339.
7. Chou, T.-Y., Dang, C. V., and Hart, G. W. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 4417.
8. Holt, G. W., Haltiwanger, R. S., Torres, C.-R., and Hart, G. W. (1987) *J. Biol. Chem. 262*, 14847.
9. Dong, D. L.-Y., Xu, Z.-S., Chevrier, M. R., Cotter, R. J., Cleveland, D. W., and Hart, G. W. (1993) *J. Biol. Chem. 268*, 16679.
10. Patti, M. E., Virkamaki, A., Landaker, E. J., Kahn, C. R., and Yki-Jarvinen, H. (1999) *Diabetes 48*, 1562.
11. Chakraborty, A., Saha, D., Bose, A., Chatterjee, M., and Gupta, N. K. (1994) *Biochemistry 33*, 6700.
12. Datta, B., Ray, M. K., Chakrabarti, D., Wylie, D. E., and Gupta, N. K. (1989) *J. Biol. Chem. 264*, 20620.
13. Meikrantz, W., Smith, D. M., Sladicka, M. M., and Schlegel, R. A. (1991) *J. Cell Sci. 98*, 303.
14. Lubas, W. A., and Hanover, J. A. (2000) *J. Biol. Chem. 275*, 10983.
15. Hart, G. W., Greis, K. D., Dong, L. Y., Blomberg, M. A., Chou, T. Y., Jiang, M. S., Roquemore, E. P., Snow, D. M., Kreppel, L. K., Cole, R. N., et al. (1995) *Adv. Exp. Med. Biol. 376*, 115.
16. Haltiwanger, R. S., Blomberg, M. A., and Hart, G. W. (1992) *J. Biol. Chem. 267*, 9005.
17. Kreppel, L. K., Blomberg, M. A., and Hart, G. W. (1997) *J. Biol. Chem. 272*, 9308.
18. Lubas, W. A., Frank, D. W., Krause, M., and Hanover, J. A. (1997) *J. Biol. Chem. 272*, 9316.
19. Kreppel, L. K., and Hart, G. W. (1999) *J. Biol. Chem. 274*, 32015.
20. Dong, L.-Y., and Hart, G. W. (1994) *J. Biol. Chem. 269*, 19321.
21. Gao, Y., Wells, L., Comer, F. I., Parker, G. J., and Hart, G. W. (2001) *J. Biol. Chem. 276*, 9838.
22. Wells, L., Gao Y., Mahoney, J. A., Voseller, K., Chen, C., Rosen, A., and Hart, G. W. (2002) *J. Biol. Chem. 277*, 1755.
23. Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W., and Marth, J. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5735.
24. Wells, L., Vosseller, K., and Hart, G. W. (2001) *Science 291*, 2376.
25. Haltiwanger, R. S., Holt, G. D., and Hart, G. W. (1990) *J. Biol. Chem. 265*, 2563.
26. Nolte, D., and Muller, U. (2002) *Mammalian Genomics 13*, 62.
27. Goebl, M., and Yanagida, M. (1991) *Trends Biochem. Sci. 16*, 173.
28. Lamb, J. R., Tugendreich, S., and Hieter, P. (1995) *Trends Biochem. Sci. 20*, 257.
29. Wrabl, J. O., and Grishin, N. V. (2001) *J. Mol. Biol. 314*, 365.
30. Yang, X., Zhang, F., and Kudlow, J. E. (2002) *Cell 110*, 69.
31. Machida, M., and Jigami, Y. (1994) *Biosci. Biotech. Biochem. 58*, 344.
32. Selzer, J., Hofmann, F., Rex, G., Wilm, M., Mann, M., Just, I., and Aktories, K. (1996) *J. Biol. Chem. 271*, 25173.
33. Arnold, C. S., Johnson, G. V., Cole, R. N., Dong, D., Lee, M., and Hart, G. W. (1996) *J. Biol. Chem. 271*, 28741.
34. Griffith, L. S., Mathes, M., Schmitz, B. (1995) *J. Neurosci. Res. 41*, 270.
35. Jacobsen, S. E., and Olszewski, N. E. (1993) *Plant Cell 5*, 887.
36. Jacobsen, S. E., Binkowski, K. A., and Olszewski, N. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 9292.
37. Sun, T. (2000) *Curr. Opin. Plant. Biol. 3*, 374.
38. Robertson, M., Swain, S. M., Chandler, P. M., and Olszewski, N. E. (1998) *Plant Cell 10*, 995.
39. Swain, S. M., and Olszewski, N. E. (1996) *Plant Physiol. 112*, 11.
40. Thornton, T. M., Kreppel, L. K., Hart, G. W., and Olszewski, N. E. (1999) in *Plant Biotechnology and In Vitro Biology in the 21st Century* (Altman, A., Ziv, M., and Izhar, S., Eds.) p 445, Kluwer Academic Publishers, Dordrecht.
41. Thornton, T. M. (2001) *The Role of O-GlcNAc Modification in Giberrellin Signal Transduction*, Ph.D. Dissertation, University of Minnesota, St. Paul.
42. Hartweck, L. M., Scott, C. L., and Olszewski, N. E. (2002) *Genetics 161*, 1279.
43. Goebl, M., and Yanagida, M. (1991) *Trends Biochem. Sci. 16*, 173.
44. Lamb, J. R., Tugendreich, S., and Hieter, P. (1995) *Trends Biochem. Sci. 20*, 257.
45. Das, A. K., Cohen, P. W., and Barford, D. (1998) *EMBO J. 17*, 1192.
46. Comer, F. I., and Hart, G. W. (2001) *Biochemistry 40*, 7845.
47. Kouzarides, T. (1999) *Curr. Opin. Genet. Dev. 9*, 40.
48. Glass, C. K., and Rosenfeld, M. G. (2000) *Genes Dev. 14*, 121.
49. Comer, F. I. (2000) *Role of O-GlcNAc on the RNA Polymerase II Carboxy-Terminal Domain*: *A Dissertation*, University of Alabama at Birmingham, Birmingham.
50. Heckel, D., Comtesse, N., Brass, N., Blin, N., Zang, K. D., and Meese, E. (1998) *Hum. Mol. Genet. 7*, 1859.
51. Lamb, J. R., Michaud, W. A., Sikorski, R. S., and Hieter, P. A. (1994) *EMBO J. 13*, 4321.
52. Myers, A., Holmans, P., Marshall, H., Kwon, J., Meyer, D., Ramic, D., Shears, S., Booth, J., Crook, R., Abraham, R., Rice, F., Carty, S., Kehoe, P., Jones, L., Owen, M. J., Hardy, J., and Goate, A. M. (2000) *Science 290*, 2304.

53. Bertram, L., Blacker, D., Mullin, K., Keeney, D., Jones, J., Basu, S., Yhu, S., McInnis, M. G., Go, R. C., Vekrellis, K., Selkoe, D. J., Saunders, A. J., and Tanzi, R. E. (2000) *Science 290*, 2302.
54. Nobile, C., and Pitzalis, S. (1999) *Genomics 62*, 86.
55. Schultz, J., and Pils, B. (2002) *FEBS Lett. 529*, 179.
56. Comtesse, N., Maldener E., and Meese, E. (2001) *Biochem. Biophys. Res. Commun. 283*, 634.
57. Wolf, B. B., and Green, D. R. (1999) *J. Biol. Chem. 274*, 20049.
58. Mukerjee, N., McGinnis, K. M., Gnegy, M. E., and Wang, K. K. (2001) *Biochem. Biophys. Res. Commun. 285*, 1192.
59. Iyer, S. N., Akimoto, Y., and Hart, G. W. (2002) *J. Biol. Chem.* 10.1074/jbc.M209384200.
60. Vosseller, K., Wells, L., Lane, M. D., and Hart, G. W. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 5313.
61. McClain, D. A., Lubas, W. A., Cooksey, R. C., Hazel, M., Parker, G. J., Love, D. C., and Hanover, J. A. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 10695.
62. Webb, P. R., Wang, K. Q., Scheel-Toellner, D., Pongracz, J., Salmon, M., and Lord, J. M. (2000) *Apoptosis 5*, 451.
63. Zhivkov, V., Tosheva, R., and Zhivkova, Y. (1975) *Comp. Biochem. Physiol. 51B*, 421.
64. Rosetti, L. (2000) *Endocrinology 141*, 1922.
65. Hart, G. W., Haltiwanger, R. S., Holt, G. D., and Kelly, G. W. (1989) *Annu. Rev. Biochem. 58*, 841.
66. Van den Steen, P., Rudd, P. M., Dwek, R. A., and Opdenakker, G. (1998) *Crit. Rev. Biochem. Mol. Biol. 33*, 151.

BI020685A