Functional Expression of O-linked GlcNAc Transferase
DOMAIN STRUCTURE AND SUBSTRATE SPECIFICITY*
(Received for publication, August 11, 1999, and in revised form, January 21, 2000)

William A. Lubas and John A. Hanover‡

From the Laboratory of Cell Biochemistry and Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

O-GlcNAc transferase (OGT) modifies nuclear pore proteins and transcription factors. In Arabidopsis, the OGT homolog participates in the gibberellin signaling pathway. We and others have proposed that mammalian OGT is the terminal step in a glucose-sensitive signal transduction pathway that becomes deregulated in insulin resistance. To facilitate mutational analysis of OGT in the absence of competing endogenous activity, we expressed the 103-kDa human OGT in Escherichia coli. Kinetic parameters for the purified recombinant enzyme (Km = 1.2 μM for Nup 62; Km = 0.5 μM for UDP-GlcNAc) are nearly identical to purified mammalian OGT. Deletions in the highly conserved C terminus result in a complete loss of activity. The N-terminal tetratricopeptide repeat domain is required for optimal recognition of substrates. Removal of the first three tetratricopeptide repeats greatly reduces the O-GlcNAc addition to macromolecular substrates. However, this altered enzyme retains full activity against appropriate synthetic peptides. Autoglycosylation of OGT is augmented when the first six tetratricopeptide repeats are removed showing that these repeats are not required for catalysis. Given its proposed role in modulating insulin action, OGT may modify kinases involved in this signaling cascade. Among the many kinases tested, OGT glycosylates glycogen synthase kinase-3 and casein kinase II, two enzymes critical in the regulation of glycogen synthesis.

One ubiquitous form of cytoplasmic and nuclear glycosylation is the addition of a single N-acetylglucosamine in O-glycosidic linkage to serine or threonine residues (1). O-GlcNAc-modified proteins are typically found in large macromolecular complexes suggesting a role in protein-protein interactions. O-GlcNAc addition is a dynamic process; the half-life of O-GlcNAc residues is significantly shorter than the half-lives of the protein modified (1–3). Changes in cell physiology may be accompanied by transient and dramatic alteration in O-GlcNAc levels (4). Such rapid changes have led to proposals that posttranslational modification by O-GlcNAc may be analogous to protein phosphorylation. Indeed, in most cases, proteins bearing O-GlcNAc are also found to be phosphorylated (1, 5). The relationship between O-GlcNAc and phosphorylation is obviously complex and may be a key to understanding the function of O-GlcNAc addition.

Another clue to understanding the role of O-GlcNAc in cell physiology derives from in vitro studies on the role of the hexosamine biosynthetic pathway in insulin resistance. The hexosamine biosynthetic pathway mediates the metabolic interconversion of intracellular glucose to UDP-GlcNAc, the substrate for OGT. The work of a number of investigators has implicated the hexosamine biosynthetic pathway in insulin resistance (6–9). Because OGT uses UDP-GlcNAc as a substrate, it is likely that the OGT participates in some aspect of mammalian insulin resistance (10, 11).

A final clue was the identification of a homolog of OGT in the plant Arabidopsis. This protein (spindly) corresponds to the protein product of the spy locus responsible for regulating the gibberellin signaling pathway (12). Mutations at this locus yield a phenotype exhibiting constitutive activation of the gibberellin pathway (13). Although similar mutations have yet to be identified in mammalian OGT, this enzyme is also likely to serve a signaling role. The enzyme responsible for O-linked GlcNac addition, UDP-N-acetylglucosamine:polypeptide β-N-acetylglucosaminyl transferase (O-GlcNAc transferase, OGT)1 was initially isolated from rat liver as a heterotrimer of 350 kDa, containing two 110-kDa subunits and one 78-kDa subunit (14). In rabbit red blood cells, however, only the 110-kDa subunit was routinely observed (10). A polyclonal antibody to the 110-kDa rat enzyme recognized both the 110- and 78-kDa subunits by Western blot analysis (15), suggesting that both proteins had common peptide epitopes. The larger subunit has recently been cloned from humans, nematodes (10), and rats (15). Overexpression of the human cDNA, encoding the 103-kDa subunit, in HeLa cells increased O-GlcNAc transferase activity. However, because mammalian cells have endogenous activity, the possibility that the cloned protein was indirectly stimulating endogenous activity could not be rigorously excluded (10, 15). Here we show that expression of the human OGT cDNA in Escherichia coli, an organism that has no endogenous activity, results in an active enzyme capable of modifying recombinant substrates in vitro. This argues that other subunits are not essential for enzyme activity, although other regulatory subunits may exist in vivo.

The cDNAs from humans, rats, and nematodes encode for proteins of molecular mass 103, 116, and 128 kDa, respectively. There are 13 tandem tetratricopeptide repeats (TPRs) in the nematode protein compared with 11 in rat and 9 in human. The TPR is a 34-amino acid repeat containing the 8 following residues (16, 17). A recent crystal structure suggests that each degenerate repeat forms a pair of antiparallel α-helices resulting in a superhelical structure believed to be the site of protein-protein interaction (17). Here we show that the TPR domain in human OGT is impor-

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1794 solely to indicate this fact.

‡ To whom all correspondence should be addressed. Tel.: (301) 496-0943; Fax: (301) 496-9431.

1 The abbreviations used are: OGT, O-GlcNAc transferase; UDP-N-acetylglucosamine:polypeptide β-N-acetylglucosaminyl transferase; TPR, tetratricopeptide repeat; bp, base pair; PAGE, polyacrylamide gel electrophoresis; HPLC, high pressure liquid chromatography; GSK, glycogen synthase kinase; Mes, 4-morpholineethanesulfonic acid.
tant for glycosylation of proteins such as Nup 62 and casein kinase II as well as peptides such as YSDDSPSTST. Yet the first six TPRs are not absolutely required for catalytic activity. The recombinant enzyme is also shown to be capable of autoglycosylation, and removal of the first six TPRs increases this activity.

The number of proteins known to be modified by OGT is steadily increasing since it was identified on cytoplasmic and nuclear proteins (18, 19). In this study we show that two enzymes that play a key role in glycogen metabolism, glycogen synthase kinase-3 and casein kinase II, can be modified with recombinant OGT in vitro. These data are consistent with a role of OGT in modulating insulin resistance in mammals.

**EXPERIMENTAL PROCEDURES**

**Cloning of Human O-GlcNAc Transferase into the pET 32 Expression Vector—**Human OGT, clone Lv4F, (10) was subcloned into pET 32B (Novagen) using the EcoRV and NcoI restrictions sites to make clone A. Clone B was isolated by excising the 441-bp BglII-BgII fragment from clone A and religating with T4 DNA ligase. Clone C was generated by excising the 726-bp EcoRV-NsiI fragment and religating using a linker made from the following primers (p)ACAGGAAGCTCTGATGCA and (p)TCAGAGCCTTCTGTC. Clone D was isolated by subcloning the EcoRI-NcoI fragment from Lv4F into pET 32C (Novagen). Clone E was isolated by excising the 1819-bp AflII-NcoI fragment from Clone A, filling in the 5′-overhangs with Klenow and using T4 DNA ligase for blunt-end ligation. Clone F was isolated by excising the 1212-bp ExpI-ExpI fragment from Clone A and religating with T4 DNA ligase. Clone G was isolated by excising the 341-bp XhoI-XhoI fragment from clone A and religating with T4 DNA ligase. All clones were sequenced through the cloning sites to confirm the sequence.

**Purification of Human OGT Expressed in E. coli—** Cultures of BL21 (DE3) cells containing the pET 32 human OGT clone expression vector were grown overnight at room temperature at 220 rpm in LB media (Digene) supplemented with carbencillin (50 μg/ml). Cells were centrifuged at 3000 rpm for 5 min in a Beckman GS-6R centrifuge. The pellet was freeze-thawed and resuspended in 1/20 of the original volume in 1 mg/ml lysozyme, 20 mM Tris-HCl, pH 7.4, 2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 1 μM chymostatin, 1 μM leupeptin, 1 μM pepstatin, 0.1% Triton X-100. The lysozyme digestion was performed for 15 min. The lysate was sonicated on ice (2 × 10 s, 70 °C).

The peptide glycosylation assay was performed as described previously (20) again shows that OGT glycosylation is quite specific. The enzyme does not modify any soluble E. coli proteins (Fig. 1A, lane 3) under conditions in which Nup 62 is highly modified (Fig. 1A, lane 4). Because O-GlcNac glycosylation can readily be observed in bacteria transformed with OGT cDNA, it appears that this enzyme is sufficient for activity, and other mammalian subunits are not essential.

Subcloning of OGT into the pET 32 vector produces a fusion protein containing an N-terminal epitope tag permitting easy purification. A soluble 114-kDa fusion protein can be highly purified in a single step using S-protein beads as shown in Fig. 1B, lane 3. When incubated with UDP-[3H]GlcNAc, the recombinant OGT adds [3H]GlcNAc to Nup 62 in a reaction linearly dependent upon enzyme concentration and time (Fig. 2 and data not shown). The reaction is completed inhibited by UDP (Fig. 2, lane 6) as had been previously described for the mammalian enzyme (14).

The purified recombinant enzyme exhibits a specific activity and kinetic parameters similar to the enzyme isolated from rabbit blood (10) and rat liver (14) (Fig. 3). These quantitative data strongly suggest that the addition of the N-terminal epitope tag did not significantly effect enzyme activity.

**Glycosylation of In Vitro Substrates by OGT—** In addition to Nup 62 we tested to see if OGT could modify other peptide substrates and proteins. Fig. 4A again shows that OGT glycosylation is quite specific. The enzyme can modify a peptide YSDDSPSTST, which had previously been shown to be a substrate for the rat enzyme, but does not modify another similar peptide YSDDSTST, which was shown not serve as an acceptor (21).

Given this specificity, we asked if any of the kinases known to be involved in insulin signaling might be substrates for OGT. As Fig. 4B demonstrates, recombinant OGT can modify two mammalian kinases, casein kinase II and GSK-3β, but not cAMP-dependent protein kinase A, casein kinase I, mitogen-activated protein kinase (extracellular signal-regulated kinase 2), or calmodulin-dependent protein kinase II. Both casein kinase II and GSK-3β phosphorylate glycogen synthase and modulate its activity.

**Deletion Analysis of Human OGT—** Starting with the active...
recombinant enzyme, deletion mutants of OGT were made as shown in Fig. 5A to characterize the domains of OGT that are responsible for catalysis. Human OGT has only 9 tetra-tricopeptide repeats compared with rat, which has 11, and nematode, which has 13. N-terminal deletions removing the first 3 repeats (clone B), the first 6 repeats (clone C), or the entire TPR domain (clone D), and three C-terminal deletions labeled clones E, F, and G were constructed as shown in Fig. 5A. The recombinant deletion mutations were purified using S-Tag chromatography and probed with S-protein-horseradish peroxidase reductase conjugate to confirm that the predicted single full-length species were obtained for each clone (Fig. 5B).

Each of these mutants was analyzed for the ability to modify Nup 62, casein kinase II, and the peptide YSDSPSTST (Table I). Deletion of the first 3 tetra-tricopeptide repeats in clone B results in a significant loss of activity (58%) for Nup 62, whereas removal of the first 6 tetra-tricopeptide repeats in clone C results in an almost complete loss of activity toward Nup 62 as a substrate. C-terminal deletions, clones E, F, and G, had no measurable activity suggesting that this part of the protein is essential for catalysis. This is consistent with the high level of conservation seen between human and rat cDNA in the C-terminal region. There is 99% identity at the protein level in the C-terminal 872 amino acids of rat and human OGT. Casein kinase II, which was a much poorer substrate than Nup 62, appeared to be modified only by the full-length enzyme clone A. In contrast, the peptide YSDSPSTST was preferentially modified by clone B lacking the first three tetra-tricopeptide repeats, suggesting the first three tetra-tricopeptide repeats were partially inhibiting its glycosylation. Taken together these data suggest that the tetra-tricopeptide repeats contribute to the substrate specificity of OGT. Although all of the tetra-tricopeptide repeats are not essential for catalytic activity, they do appear to be necessary for optimal glycosylation of full-length proteins.

In analyzing the glycosylation products made by the various forms of the enzyme (Fig. 5C), we noted that clone C preferentially catalyzes self-glycosylation. We then measured autoglycosylation of clones A, B, and C in the absence of an exogenous substrate (Fig. 6B). Although comparable levels of A, B, and C were used (Fig. 6A), clones A and B still had significantly lower levels of autoglycosylation than clone C (Fig. 6B). We also used galactosyltransferase labeling to identify terminal O-GlcNAc residues on the purified recombinant enzymes (Fig. 6C) to see if OGT had already partially modified itself in E. coli. If clones A and B were already highly modified in E. coli, this could explain why they were poor substrates for autoglycosylation in the in vitro assays. The galactosyltransferase labeling, however, showed that clone C also had the highest level of autoglycosylation in E. coli. This confirms that clones A and B are poorer substrates for autoglycosylation than clone C. Taken together, these data suggest that removal of the first 6 TPRs interferes with the recognition of Nup 62 as a substrate and augments autoglycosylation.

**DISCUSSION**

**OGT and Signal Transduction**—Evidence is accumulating that O-GlcNAc addition can act as a regulatory modification involved in signal transduction pathways. In plants, the OGT homolog spindly (spy) is involved in suppression of the gibberellin signaling pathway (12). Furthermore, a large number of nuclear and cytoplasmic proteins have been previously identified as substrates for OGT. Many of these proteins are directly involved in signal transduction. Examples include RNA polymerase II and the transcription factors, Ap1, Sp1, serum response factor, hepatocyte nuclear factor, and the pancreatic-specific transcription factor. Also modified are the estrogen receptor, oncogenes such as c-myec, v-erb, SV40 large T antigen, and the tumor suppressor gene, p53. It has been shown that glycosylation of p53 in its C-terminal regulatory domain enhances its DNA binding activity (22), whereas glycosylation of an Sp1-derived peptide inhibits its binding to the TATA-binding protein-associated factor 110 (23). In this paper we have shown that recombinant human OGT expressed in E. coli is fully active and capable of modifying recombinant proteins in vitro. This should make it possible to clarify the role that glycosylation plays in signal transduction. The ability to modify recombinant proteins should make it easier to study the effect of O-GlcNAcylation on protein-protein and protein-DNA interactions in vitro.
Our data suggest that OGT glycosylation is quite specific. We were unable to detect any glycosylation of bacterial proteins isolated along with recombinant OGT in the total supernatant fraction from *E. coli*. However, autoglycosylation of OGT was observed. Because catalytically active OGT can be isolated from *E. coli*, which has no endogenous O-GlcNAc activity, autoglycosylation is not likely to be essential for enzyme activity. However, autoglycosylation may play an important role in the regulation of enzyme activity or substrate specificity *in vivo*. We are presently carrying out a more detailed mutational analysis of OGT to identify residues required for catalytic activity and autoglycosylation.
Functional Expression of Recombinant O-GlcNAc Transferase

The O-GlcNAc transferase activity of deletion mutants

<table>
<thead>
<tr>
<th>OGT clone</th>
<th>Nup 62</th>
<th>Casein kinase II</th>
<th>YSDSPSTST</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>100</td>
<td>74</td>
</tr>
<tr>
<td>B</td>
<td>42</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE I

**Fig. 5.** Deletion analysis of human O-GlcNAc transferase. A, deletion mutants of human O-GlcNAc transferase. Clones A–G were prepared as described under “Experimental Procedures.” A diagrammatic representation of the structural domains of the enzyme is shown. Construction of the fusion protein with the N-terminal epitope tag (S-Tag) removes the first 14 amino acids from the full-length cDNA Ll4F to make clone A. Deletions in clones B–G are depicted by a gap in the open reading compared with clone A. The TPR are designated by the stippled boxes. The proposed nuclear localization sequence (NLS) is shown to divide the TPR and catalytic domains. B, S-Tag horseradish peroxidase detection of affinity purified deletion mutants. S-Tag beads were used to purify the O-GlcNAc transferase deletion mutants, clones A–G, as described under “Experimental Procedures.” The amount of mutant OGT bound to the beads was confirmed by Western blot analysis using the S-Tag horseradish peroxidase Lumiblot Kit. C, glycosylation by O-GlcNAc transferase deletion mutants using Nup 62 as substrate. S-Tag-purified O-GlcNAc transferase and clones A–G were incubated with Nup 62 (2.3 μg) for 1.5 h at 37 °C at 220 rpm in a final volume of 40 μl under standard assay conditions. Proteins were separated by SDS-PAGE, and the incorporation of [3H]GlcNAc was visualized by fluorography. The migration profile of molecular mass standards is shown to the left. Glycosylation of Nup 62 and autoglycosylation of OGT is shown by the arrows to the right.

When the substrate Nup 62 is added as a substrate, it is quite efficiently recognized and modified by recombinant OGT with kinetic properties nearly identical to the native enzyme. The data we have presented strongly argue that the TPR domain is required for the interaction with protein substrates such as Nup 62. Although TPR motifs follow a general consensus sequence, the individual motifs in a domain are unique and may be involved in specific protein-protein interactions. In the case of San6p, a TPR-containing protein responsible for glucose repression in yeast, only 4 of the 10 TPRs are essential and distinct subsets of TPR motifs have been shown to be responsible for repression of different classes of genes (24). TPRs 1–3 associate with the co-repressor Tup1; TPRs 4–7 are involved in repression of oxygen-regulated genes; whereas TPRs 8–10 appear to be required for repression of glucose-regulated genes.

Similarly in the case of the peroxisomal import protein PAS8p it appears that TPRs 1–3 specifically interact with the peroxisomal targeting signal (25).

**Fig. 6.** Autoglycosylation of O-GlcNAc transferase. A, Coomassie Brilliant Blue staining of purified deletion mutants of human O-GlcNAc transferase. Clones A, B, and C were S-Tag-purified as described under “Experimental Procedures.” The migration profile of molecular mass standards is shown to the left. B, incorporation of [3H]GlcNAc to purified OGT in vitro. S-Tag-purified O-GlcNAc transferase, clones A, B, and C were incubated with 1 μCi of UDP-[3H]GlcNAc (35 Ci/mmol, DuPont) for 1.5 h under standard assay conditions in a final volume of 40 μl. Proteins were separated by SDS-PAGE, and the incorporation of [3H]GlcNAc was visualized by fluorography. C, incorporation of [3H]Gal to purified OGT in vitro. S-Tag-purified O-GlcNAc transferase, clones A, B, and C were incubated with 2 μCi of UDP-[3H]Gal (30 Ci/mmol, DuPont) for 2 h in 50 mM Mes-HCl, pH 7.3, 2.5 mM MnCl2, and 5 milliunits of galactosyltransferase (Sigma) in a final volume of 20 μl. Proteins were separated by SDS-PAGE, and the incorporation of [3H]Gal was visualized by fluorography.

Based on these precedents, we sought to identify those tetratricopeptide repeats important for substrate recognition. Our data suggest that optimal binding of Nup 62 and casein kinase II appears to require the intact TPR domain. Although the first 3 tetratricopeptide repeats are not absolutely necessary for glycosylation of Nup 62, they do appear to contribute to substrate recognition. In contrast the first three tetratricopeptide repeats are not required for glycosylation of the peptide YSDSPSTST and in fact may inhibit the access of the peptide to the...
catalytic site. Interestingly, removal of the first six tetratricopeptide repeats severely inhibits OGT glycosylation of exogenous substrates, yet this enzyme is still catalytically active and undergoes autoglycosylation more readily, suggesting that the first six tetratricopeptide repeats are not essential for catalysis. Taken together these data suggest that the TPR domain is important in substrate recognition, and it should be possible to alter OGT substrate specificity by interfering with TPR substrate binding. This could provide a useful means of selectively inhibiting glycosylation of specific substrates.

Identification of Casein Kinase II and Glycogen Synthase Kinase-3 as Substrates for OGT—Recombinant OGT was shown to specifically modify two purified recombinant kinases GSK-3 and casein kinase II, two enzymes critical in the regulation of glycogen synthesis. Other kinases in the insulin-signaling cascade such as protein kinase A, casein kinase I, mitogen-activated protein kinase (extracellular signal-regulated kinase-2), and calmodulin-dependent protein kinase II were not found to be substrates for OGT in vitro. OGT glycosylation of kinases such as GSK-3 and casein kinase II suggests a potential mechanism for cross-talk between glycosylation and the insulin-mediated phosphorylation cascade. Both of these kinases are involved in phosphorylating glycogen synthase, the rate-limiting step in glycogen synthesis. Casein kinase II was originally isolated as GSK-5, and phosphorylation of glycogen synthase by GSK-3 requires prior phosphorylation by casein kinase II. GSK-3 has been shown to inhibit both glycogen synthase and insulin receptor substrate-2 (26).

OGT may also be directly implicated in glucose-induced insulin resistance mediated via the hexosamine biosynthetic pathway. It has been shown that hyperglycemia can lead to 2–4-fold increases in UDP-GlcNAc levels (27, 28) and results in increased levels of O-GlcNAc-modified proteins (27, 29). This is consistent with the $K_m$ of OGT being in the physiologic range and responding to dynamic changes in intracellular glucose. As expected therefore, the $K_m$ for the nucleocyttoplasmic enzyme UDP-GlcNac is similar to that found for UDP-GlcNAc and UDP-Glc transporters present in microsomal vesicles exposed to the cytoplasm (30). Thus, by responding to physiologic changes in cytoplasmic UDP-GlcNAc levels, OGT glycosylation of cytoplasmic and nuclear proteins may act as an intracellular glucose sensor. Glycosylation of transcription factors, kinases, or vesicle-associated proteins in response to changes in intracellular glucose concentration would provide a potential mechanism for insulin resistance.

During final revision of this manuscript, a report appeared describing the expression and characterization of OGT in baculovirus. In agreement with our findings, these authors present data suggesting that the C terminus is likely to contain a catalytic domain (31). Using synthetic peptides as substrates, they conclude that OGT has three distinct values for the $K_m$ of UDP-GlcNAc. Although use of synthetic peptide substrates may provide some useful information about binding at the catalytic site of OGT, it is doubtful that synthetic peptides with millimolar $K_m$ values are binding in the same manner as physiologic protein substrates such as Nup 62 ($K_m = 1.2 \mu M$). It is possible that larger protein substrates may interact with the TPR domain of OGT and then all potential glycosylation sites in close proximity to the C-terminal catalytic site would be glycosylated. This is similar to the mechanism for protein phosphorylation by certain kinases in which the docking site is important for specificity but does not undergo phosphorylation itself.

In our studies, we chose to focus upon E. coli as an expression system after our original observation that proteins expressed in the baculovirus system are efficiently glycosylated by the endogenous insect OGT (32). Expression of OGT in the baculovirus system can, therefore, produce mixed heteromers confounding the ability to accurately characterize the activity of homotrimers of the recombinant enzyme. Because E. coli lacks an endogenous activity, the expression system we have described should facilitate additional mutational analysis not possible in a system with high levels of endogenous activity. Expression in E. coli also allows for production of very large amounts of recombinant material required for solving the three-dimensional structure of OGT as well as subsequent structure-function studies. We are currently pursuing these studies.

REFERENCES


