O-Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats*

(Received for publication, November 22, 1996, and in revised form, December 13, 1996)

William A. Lubas, David W. Frank, Michael Krause, and John A. Hanover‡
From the Laboratory of Cell Biochemistry and Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

O-Linked GlcNAc addition and phosphorylation may compete for sites on nuclear pore proteins and transcription factors. We sequenced O-linked GlcNAc transferase from rabbit blood and identified the homologous Caenorhabditis elegans transferase gene on chromosome III. We then isolated C. elegans and human cDNAs encoding the transferase. The enzymes from the two species appear to be highly conserved; both contain multiple tetratricopeptide repeats and nuclear localization sequences. The C. elegans transferase accumulated in the nucleus and in perinuclear aggregates in overexpressing transgenic lines. O-Linked GlcNAc transferase activity was also elevated in HeLa cells transfected with the human cDNA. At least four human transcripts were observed in the tissues examined ranging in size from 4.4 to 9.3 kilobase pairs. The two largest transcripts (7.9 and 9.3 kilobase pairs) were enriched at least 12-fold in the pancreas. Based on its substrate specificity and molecular features, we propose that O-linked GlcNAc transferase is part of a glucose-responsive pathway previously implicated in the pathogenesis of diabetes mellitus.

Over the last 10 years, a novel post-translational modification involving the addition of a single N-acetylglucosamine in O-glycosidic linkage to serine or threonine residues on cytoplasmic and nuclear proteins has been identified (1, 2). We became interested in this form of glycosylation when it was found to modify a group of nuclear pore proteins, which have since been molecularly characterized (2–4). In addition to nuclear pore proteins, O-linked GlcNAc modifies a large number of polypeptides in multimeric structures including RNA pol II transcription complexes and p67/eIF-2-initiation factor in the translation machinery.

Although the addition of O-linked GlcNAc to proteins is formally a glycosyltransferase reaction, it is quite distinct from other forms of glycosylation. The addition occurs in the cytosol and nucleus, unlike other glycosylation reactions, which are restricted to the endomembrane system of the cell (2, 3). Because it must function in the cytoplasm where UDP-GlcNAc levels are lower, it has a much lower $K_m$ with respect to this substrate than is usually observed for glycosyltransferases. In many ways O-linked GlcNAc addition is analogous to protein phosphorylation. The enzyme which catalyzes O-linked GlcNAc addition, uridine diphospho-N-acetylglucosamine:polypeptide $\beta$-N-acetylglucosaminyl transferase (O-GlcNAc transferase, OGT), has been shown to recognize a large number of phosphoproteins, some of which play a direct role in signal transduction. In the case of RNA-polymerase II, these modifications seem to be mutually exclusive so that while the glycosylated enzyme is necessary for assembly of the preinitiation complex, subsequent deglycosylation and phosphorylation are necessary for transition to the elongation complex (5). In the case of other substrates, like neurofilaments (6) or the nuclear pore proteins Nup62, -97, and -200 (7), it appears that phosphorylation and glycosylation can coexist on the same molecule. The role of protein phosphorylation as a regulatory mechanism for signal transduction in eukaryotic cells was originally identified in studies over 40 years ago on glycogen phosphorylase, an enzyme involved in carbohydrate metabolism (8). It is likely that O-linked GlcNAc addition to proteins in the cytoplasm and nucleus is also highly regulated. Since both phosphorylation and glycosylation compete for similar serine or threonine residues, it is possible that the two processes could be directly competing for sites, or they may alter the substrate specificity of nearby sites by steric or electrostatic effects.

No strict consensus sequence for O-linked GlcNAc addition has so far been identified, although most glycosylation sites occur nearby proline or valine residues and typically in stretches rich in serine or threonine residues. A subset of glycosylation sites is located near acidic amino acid residues (9, 10). These glycosylation sites are similar to phosphorylation sites for several protein kinases (11). We have previously shown that OGT has a much higher affinity for the recombinant nucleoprotein, Nup62, than any synthetic peptide, suggesting that the enzyme may recognize other parts of the protein and not just a specific consensus sequence (10).

Although OGT has been purified from several different sources (10, 12), it has not been molecularly cloned. Here we describe the purification of rabbit OGT and use of this sequence data to clone the Caenorhabditis elegans and human enzymes. The cloned enzymes are highly conserved; both contain multiple tandem tetratricopeptide repeats (TPR) and putative nuclear localization sequences. HeLa cells transiently transfected with the human enzyme had elevated OGT activity. Polyclonal antiserum prepared against the recombinant OGT was used to localize the enzyme to the nucleus and perinuclear aggregates in transgenic C. elegans embryos. Human OGT transcripts were observed in all tissues examined, but the highest levels of expression were observed in the pancreas.

**EXPERIMENTAL PROCEDURES**

Purification of Rabbit O-GlcNAc Transferase—Fresh rabbit blood (4 liters), anticoagulated with EDTA (Pel-Freez), was pelleted in a GS3 rotor at 2,000 × g for 5 min. The red blood cells were washed three times...
with an isonicotinic acid solution (140 mM NaCl, 5 mM KCl, 1.5 mM magnesium acetate) and collected after centrifugation at 20,000 \times g for 5 min for the first two washes and 5,000 \times g for 10 min after the final wash. Hypotonic lysis was performed using an equal volume of ice cold water containing the following protease inhibitors (Boehringer Mannheim): leupeptin, 100 \mu g/ml pepstatin, 10 \mu g/ml leupeptin, 0.1% aprotinin, and 2 mM EDTA. The lysate was pelleted at 10,000 \times g for 40 min in a GSA rotor. The soluble fraction was made 30% saturated with ammonium sulfate by adding a stock of 100% saturated ammonium sulfate equilibrated at 4 °C slowly over 1 h and stirring the solution an additional 2 h at 4 °C.

The precipitate was collected after centrifugation at 100,000 \times g for 20 min in a GSA rotor and resuspended in 15–20 ml of 50 mM Tris–HCl, pH 7.4, 2 mM MgCl₂ using a Dounce homogenizer. The insoluble material was removed by centrifugation at 20,000 \times g for 20 min in a SS34 rotor. The soluble fraction from the 30% ammonium sulfate precipitation was loaded onto a 15-ml phenyl-Sepharose column (Pharmacia Biotech Inc.), washed with 100 ml of 10 mM Tris–HCl, pH 7.5, 100 mM ammonium sulfate, and eluted with 40 ml of 10 mM Tris–HCl, pH 7.5, 60% ethylene glycol. All chromatography buffers also contained the following protease inhibitors: 0.1% aprotinin, 10 \mu g/ml leupeptin, 10 \mu g/ml pepstatin, 0.1 mM phenylmethylsulfonyl fluoride, and all procedures were performed at 4 °C. The active fractions (15–20 ml) were pooled, passed through a 0.45-μm Millex-HA filter, and loaded onto a Mono Q HR 5/5 column (Pharmacia Biotech) equilibrated with 50 mM Tris–HCl, pH 7.5, 12.5 mM MgCl₂, 20% glycerol, 2 mM EDTA using a Pharmacia FPLC system. The column was washed with 30 ml of the equilibration buffer and then eluted with a linear gradient from 0 to 300 mM NaCl in 50 ml of equilibration buffer at a flow rate of 1 ml/min. The active fractions (8–10 ml) were pooled and concentrated to a final volume of 0.3 ml using a Centricon-30 microcentrutor (Amicon) and loaded in 0.15-mI aliquots onto a Superox 6 FPLC column equilibrated with 50 mM Tris-HCl, pH 7.5, 12.5 mM MgCl₂, 20% glycerol, 2 mM EDTA, 100 mM NaCl. The column was run at a flow rate of 0.15 ml/min, and 0.6 ml fractions were collected. Protein was calculated using the BCA reagent (Pierce) using bovine serum albumin as a standard. O-GlcNAc transferase activity was measured using recombinant Nub26 to nitrocellulose membranes as described previously (10) or in a modification of the method using recombinant Nub26 bound to StrepII polystyrene scintillation strips (Wallac). A typical purification results in a 30,000-fold purification and a 1–2% yield. Purified O-GlcNAc transferase was subject to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and the 10-kDa dimer was cut out and sent to the William M. Keck Foundation at Yale University for in gel trypsin digestion, high pressure liquid chromatography purification, and cDNA sequencing.

Cloning of the C. elegans O-GlcNAc Transferase—Polymerase chain reaction (PCR) primers GTTTGTTAATCGGAAGAACATCG and ATCGAAAATCTCCTGGCCTCTT were made to amplify a 195-base pair fragment from the cDNA clone yk13e2 (Fig. 2A). This PCR fragment was gel-purified and used to probe a ZAP C. elegans cdNA library (10^9 units/ml) (15). 140,000 clones were screened in this manner, and only 1 positive clone was identified. The insert from this clone was cloned into pGem and pET 32 using EcoRI, SalI, and NotI partial digest, 4.25 kb was cloned into the NcoI and SacI restriction sites of the vector. Transgenic strains were generated by microinjection using the pRF4 plasmid as a marker to identify transformed animals (15). Test plasmid constructs were injected with combinations of pRF4 DNA at 50 ng/μl each.

Overexpression of the OGT gene was achieved by transformation of N2 animals with derivatives of the heat shock promoter vectors (pPD49.75 and pPD49.83) (15) in which the full-length C. elegans OGT cDNA (NcoI–SacI partial digested, 4.25 kb) was cloned into the NcoI and SacI restriction sites of the vector. The transgenic animals were heat-shocked at 33 °C for 2–4 h in order to induce production of fusion proteins driven by heat shock promoters. Overexpressed OGT was detected by immunoblotting using a anti-OGT guinea pig antibody raised against the recombinant protein made in E. coli.

Concluding the Human O-GlcNAc Transferase—Taking advantage of the published sequence of a human expressed sequence tag, accession number R75943 (Fig. 2B), two oligonucleotide primers (GGGTTTTCGACACGTTAAGAG and ACATTCTGAGGGCTTTCC) were constructed and used to screen superscript human brain and liver cdNA libraries using the Genetrapper cdNA-positive selection system. The first primer was biotinylated at the 5′ end with biotin-14-dCTP using terminal deoxynucleotidyl transferase and used to screen single-stranded human liver or brain cdNA libraries. Hybrids between the biotinylated oligonucleotide and the cdNA libraries were captured on streptavidin-coated paramagnetic beads and retrieved using a magnet. The captured ssDNA was separated from the biotinylated primer, re-paired to double-stranded DNA using the second oligonucleotide primer, and transformed into ElectroMAX DH10B cells. There were a total of 48 liver and 53 brain clones identified on the initial screen. These clones were then rescreened by hybridization with the full-length human placenta expressed sequence tag, accession number R75943, and 40/48 liver and 42/53 brain clones were found to be positive. The insert size was estimated by restriction digestion with SalI and NorI. All liver clones were > 2.5 kb and brain clones > 3 kb were screened by in vitro transcription. The largest cDNA clone was 4.4 kb and was a protein of about 100 kDa formed by 6 different liver and 2 brain clones (data not shown). DNA sequencing showed that they were all overlapping clones of the same gene with variable 5′ and 3′ untranslated regions. The full-length clone Lv4F was fully sequenced and is reported here (GenBank™ accession number U77413).

Filter Hybridization—Human, rabbit, rat, and mouse genomic DNA (Clontech) was digested with wheat germ extract system. The digested DNA was fractionated in agarose gel, transferred to nylon membranes (GeneScreen Plus, DuPont) by capillary action. The blot was hybridized in 1% bovine serum albumin, 0.5 M NaPO₄, pH 7, 1 mM EDTA, 7% sodium dodecyl sulfate, 100 μg/ml denatured salmon testis DNA at 55 °C for 1 h and then hybridized overnight at 55 °C with the gel-purified, radioiodinated 3 kb NorI/SalI fragment from human liver clone Lv4F. The blot was washed twice for 15 min with 0.5% bovine serum albumin, 5% sodium dodecyl sulfate, 40 mM NaPO₄, pH 7, 1 mM EDTA at 55 °C; two times for 15 min with 1% bovine serum albumin, 40 mM NaPO₄, pH 7, 1 mM EDTA at 55 °C; and once with 0.2 × SSPE (30 mM NaCl, 2 mM NaPO₄, pH 7.4, 0.2 mM EDTA) at 55 °C for 15 min. It was exposed to Kodak Bio-Max MR film for 1–7 days at −70 °C. A human multiple tissue Northern blot (Clontech) was prehybridized as described previously for the Southern blot except that all incubations were performed at 65 °C.

Expression of Human O-GlcNAc Transferase in Transiently Transfected HEK Cells—Human OGT clone Lv4F was introduced to HeLa cells by lipid-mediated transfection or by electroporation. For the lipid-mediated method, 10^7 cells were plated per well in six-well plates in Dulbecco's modified Eagle medium, 10% fetal bovine serum for 14–18 h prior to transfection. The transfection was carried out in Opti-MEM (Life Technologies, Inc.). The plasmid pCEC-OGT/Lv4F (0.1 μg) was mixed with 4 μl of Lipofectin reagent (Life Technologies, Inc.) and applied to the cells according to the manufacturer's recommendations. Control cells were transfected with plasmid bearing no insert. Electroporation of HeLa cells was performed in Opti-MEM in cell suspensions.
RESULTS

Isolation and Purification of Mammalian OGT—A mammalian OGT was purified from rabbit blood using a modification of previously described methods (10, 12). The purified enzyme shown in Fig. 1 contains two polypeptides at 110 and 78 kDa. Recovery of the 78-kDa band was variable between preparations, thus preventing isolation of sufficient amounts for further analysis. Proteolytic fingerprinting of both polypeptides suggested they are related. The 78-kDa band may be a proteolytic product of the larger 110-kDa band or the product of a second translation start site. The 110-kDa band was subjected to tryptic digestion and microsequencing; two peptides were isolated from the rabbit OGT, initially identified, a 20-mer peptide, \(\text{XXX}-\text{SQLT}-(\text{C})\text{LG(C)LELIAK}\), and a 17-mer, \(\text{XXX}-\text{SQLT}-(\text{C})\text{LG(C)LELIAK}\). These fragments by trypsin digestion. Fig. 2A shows the structure of the gene and localizes the tryptic peptides to the 8th and 14th exons in the \(\text{C. elegans}\) gene. Two human expressed sequence tags (accession numbers R75943 and R76782), showing greater than 60% identity to the \(\text{C. elegans}\) gene K04G7.3 were also identified and found to match perfectly the 17-mer rabbit OGT tryptic peptide (Fig. 2B).

Cloning of the cDNA Encoding \(\text{C. elegans}\) OGT—The OGT cDNA was isolated using a combination of phage library screening and polymerase chain reaction (see "Experimental Procedures"). The final clone (ZAP-\text{CeOGT} accession number U77412) was sequenced and found to match perfectly the 17-mer rabbit OGT cDNA clone yk15c2 is shown. The partial amino acid sequences of the two tryptic peptides isolated from the rabbit OGT, XVTLDPNFLDAYINLGNVLK and XXXSQLT-(C)LG(C)LELIAK, are compared with the sequence predicted from R04G7.3 open reading frame in exons 8 and 14. The cysteine residues are tentative assignments because they could not be distinguished from glutamine residues, which would comigrate in the amino acid profile. B, the 3.1-kb full-length sequence of human liver OGT clone Lv4F (accession number U77413) is shown. The partial amino acid sequences of the two tryptic peptides isolated from the rabbit OGT, XVTLDPNFLDAYINLGNVLK and XXXSQLT-(C)LG(C)LELIAK, are compared with clone Lv4F and to two human expressed tags (accession numbers R75943 and R76782).

Cloning of the Human OGT—The human OGT was isolated
using primers constructed from the sequence of the human expressed sequence tag (accession number R75943). These primers were used to screen superscript human brain and liver cDNA libraries using the Genetrapper cDNA-positive selection system (see "Experimental Procedures"). From 101 initial clones, a total of 8 full-length clones were identified. Six of these clones were obtained from liver and 2 from brain libraries; all had overlapping 5′ and 3′ untranslated sequences.

When translated in vitro, each of these full-length clones produced a polypeptide of approximately 100 kDa and variable amounts of a smaller 70-kDa species. The 70-kDa species, resulting from an alternative translation start, may be related to the 78-kDa species that has been observed with purified OGT preparations (Fig. 1). One of the liver cDNA clones, Lv4F (accession number U77413), was fully sequenced (Fig. 5) and found to encode an open reading frame with 68% identity with the C. elegans K40G7.3 gene product over the C-terminal 872 amino acids (Fig. 4A). The human cDNA open reading frame encodes a shorter protein (103 kDa) containing only the last nine TPR sequences found in C. elegans (Fig. 4B). While this is consistent with the observed size of the in vitro translation product, it is likely that post-translational modification of the...
enzyme occurs since the human OGT translated in reticulocyte lysate was slightly larger than the product seen from wheat germ extract (data not shown). This behavior has been previously observed for proteins modified by O-linked GlcNAc (4).

To examine the properties of the protein, we expressed OGT in E. coli as described under “Experimental Procedures.” Attempts to purify fully functional native enzyme from the bacteria were unsuccessful since the enzyme aggregated into inclusion bodies and became insoluble. The enzyme could be solubilized in 6 M urea and purified by His-Tag chromatography. This recombinant form of OGT was used to raise polyclonal antisera in guinea pigs for immunodetection (see below).

Transgenic C. elegans Lines Overexpressing OGT—To examine the localization of the OGT in C. elegans, several transgenic lines were produced that overexpress enzyme under control of heat shock promoters. When induced to overexpress by heat shock, the C. elegans OGT was readily detected by immunoblotting using the guinea pig antiserum (Fig. 6A).

**FIG. 4.** Protein sequence comparison of C. elegans and human O-GlcNAc transferase deduced from the isolated cDNAs. A, sequence alignment of C. elegans and human O-GlcNAc transferases. Identical amino acids matches are boxed and shaded. Similar amino acids are shaded only. The peptide sequences (Peptide seq) corresponding to the partial amino acid sequences of the tryptic peptides XVTLDPNFLDAYINLGNVLK and XXXSQLTG(C)LG(C)LELIAK are underlined. The cysteine residues are tentative assignments because they could not be distinguished from glutamine residues, which would comigrate in the amino acid profile. The putative nuclear localization signal (NLS) is underlined. Sequence data are derived from the cDNAs (accession numbers U77412 and U77413) reported here. B, schematic diagram showing the relative sizes of the C. elegans and human OGT, as well as the location of TPR repeats and putative nuclear localization signal (NLS).
pattern (Fig. 6, top panel). In embryos overexpressing the enzyme (Fig. 6, lower panels), OGT was found within the nucleus in the gut suggesting that the nuclear localization sequence in C. elegans OGT is functional. In other regions of the embryos, the overexpressed OGT exhibited a distinct perinuclear localization (small arrows). This was particularly striking in the neurons. Similar localization was observed in all of the lines produced, although the tissue distribution was somewhat dependent upon the heat shock promoter used as has been previously reported (15). Thus, the enzyme was found in both the nucleus and the cytoplasm, depending on the tissue overexpressing the cDNA.

**Functional Expression of OGT in HeLa Cells—** The full-length human cDNA (clone Lv4F) was cloned into the pECE vector downstream of the SV40 promoter. HeLa cell cultures were transiently transfected with vector alone or the vector containing the clone Lv4F open reading frame. Cells were harvested at 24 h because the transfected cells did not survive well during prolonged incubations suggesting the gene may be toxic to the cells. Toxicity had also been observed in experiments where the gene was overexpressed in transgenic C. elegans.2 Up to a 3-fold increase in enzyme activity relative to background activity was observed using two different transfection procedures (Fig. 7).

**Conservation and Tissue Distribution of OGT—** Hybridization of the human liver OGT cDNA (clone Lv4F) to genomic DNA digested with EcoRI from several different species is shown in Fig. 8. The Southern analysis identifies a single large fragment in human, whereas several smaller fragments are observed in rabbit, rat, and mouse genomic DNA. The high degree of conservation observed is not surprising since C. elegans and human OGT cDNAs were found to be so similar. Several additional sequences in the data base searches were found to be related to the C. elegans and human sequences, including sequences from schistosomes and rice.

2 W. A. Lubas, D. W. Frank, M. Krause, and J. A. Hanover, unpublished observations.
To examine the relative abundance of the human OGT mRNA in various adult human tissues, a Northern blot analysis was performed (Fig. 9). The human clone Lv4F probe identifies four distinct bands at 9.3, 7.9, 6.3, and 4.4 kb, which are present in different amounts in various human tissues. Skeletal muscle and heart exhibited a relative enrichment of the 6.3-kb species, while all transcripts were at low relative abundance in the kidney and lung. The pancreas, where the two largest species (9.3 and 7.9 kb) were most abundant, showed the highest level of expression. A Northern blot analysis of the same blot with β-actin cDNA confirmed that similar levels of mRNA were loaded in each lane.

**DISCUSSION**

**OGT Is Encoded by a Previously Unidentified, Conserved Gene in C. elegans**—Here we describe the first molecular characterization of a mammalian OGT. The OGT was purified using recombinant rat nuclear pore protein, Nup62, as substrate. The enzyme was found to be nearly identical to an open reading frame purposed in 1985, which was the C. elegans open reading frame encodes a larger than 9.3 kb, which is the same size as the human gene (Fig. 4B). OGT is a 34-amino acid repeated motif having the following 8 loosely conserved residues: -W-LG-Y-A-F-A-P- (19). Over the TPR region the identity between human and C. elegans OGT is even more striking with 87% identity on the amino acid level. TPR motifs are typically arranged as tandem arrays as seen here for OGT. The motif has been identified in a wide variety of proteins in multimeric complexes involved in neurogenesis, cell cycle control, transcription, and peroxisomal transport (20). Several proteins involved in these processes are also known substrates for OGT. These include neurofilaments, tau, synapsins, RNA polymerase II, Sp1, c-Fos, c-Jun, c-Myc, v-Erb A, and the estrogen receptor (9). It is believed that TPR motifs can interact with each other mediating intra- and intermolecular protein-protein interactions. The motifs have also been suggested to play a role in targeting proteins to their proper intracellular localization (21). It is therefore likely that the TPR domain of OGT plays a role in targeting the enzyme to its many sites of action.

Evidence from database searches and our Southern hybridization results (Fig. 8) suggest that OGT is highly conserved among eukaryotic species. Interestingly, no clear-cut structural homolog emerged from an examination of the Saccharomyces cerevisiae genome. Three genes emerged as having some structural similarities to OGT. These proteins are TPR proteins involved in cell cycle control and transcription: Cdc616, Cdc27, and SSN6. SSN6 is very similar to TPR repeats to those of OGT and exhibits 42% similarity and 20% identity to both human and C. elegans OGT throughout its entire length. SSN6 is of particular interest since it is known to be involved in transcriptional repression in response to glucose; this is one of the proposed roles of OGT in mammalian cells (see below).

**OGT Contains a Bipartite Nuclear Localization Sequence Which Can Function to Direct Nuclear Localization**—Both the protein. The C. elegans open reading frame encodes a larger protein with 13 tandem TPR motifs compared with 9 for the human gene (Fig. 4B). TPR is a 34-amino acid repeated motif having the following 8 loosely conserved residues: -W-LG-Y-A-F-A-P- (19). Over the TPR region the identity between human and C. elegans OGT is even more striking with 87% identity on the amino acid level. TPR motifs are typically arranged as tandem arrays as seen here for OGT. The motif has been identified in a wide variety of proteins in multimeric complexes involved in neurogenesis, cell cycle control, transcription, and peroxisomal transport (20). Several proteins involved in these processes are also known substrates for OGT. These include neurofilaments, tau, synapsins, RNA polymerase II, Sp1, c-Fos, c-Jun, c-Myc, v-Erb A, and the estrogen receptor (9). It is believed that TPR motifs can interact with each other mediating intra- and intermolecular protein-protein interactions. The motifs have also been suggested to play a role in targeting proteins to their proper intracellular localization (21). It is therefore likely that the TPR domain of OGT plays a role in targeting the enzyme to its many sites of action.

Evidence from database searches and our Southern hybridization results (Fig. 8) suggest that OGT is highly conserved among eukaryotic species. Interestingly, no clear-cut structural homolog emerged from an examination of the Saccharomyces cerevisiae genome. Three genes emerged as having some structural similarities to OGT. These proteins are TPR proteins involved in cell cycle control and transcription: Cdc616, Cdc27, and SSN6. SSN6 is very similar to TPR repeats to those of OGT and exhibits 42% similarity and 20% identity to both human and C. elegans OGT throughout its entire length. SSN6 is of particular interest since it is known to be involved in transcriptional repression in response to glucose; this is one of the proposed roles of OGT in mammalian cells (see below).

**OGT Contains a Bipartite Nuclear Localization Sequence Which Can Function to Direct Nuclear Localization**—Both the protein. The C. elegans open reading frame encodes a larger protein with 13 tandem TPR motifs compared with 9 for the human gene (Fig. 4B). TPR is a 34-amino acid repeated motif having the following 8 loosely conserved residues: -W-LG-Y-A-F-A-P- (19). Over the TPR region the identity between human and C. elegans OGT is even more striking with 87% identity on the amino acid level. TPR motifs are typically arranged as tandem arrays as seen here for OGT. The motif has been identified in a wide variety of proteins in multimeric complexes involved in neurogenesis, cell cycle control, transcription, and peroxisomal transport (20). Several proteins involved in these processes are also known substrates for OGT. These include neurofilaments, tau, synapsins, RNA polymerase II, Sp1, c-Fos, c-Jun, c-Myc, v-Erb A, and the estrogen receptor (9). It is believed that TPR motifs can interact with each other mediating intra- and intermolecular protein-protein interactions. The motifs have also been suggested to play a role in targeting proteins to their proper intracellular localization (21). It is therefore likely that the TPR domain of OGT plays a role in targeting the enzyme to its many sites of action.

Evidence from database searches and our Southern hybridization results (Fig. 8) suggest that OGT is highly conserved among eukaryotic species. Interestingly, no clear-cut structural homolog emerged from an examination of the Saccharomyces cerevisiae genome. Three genes emerged as having some structural similarities to OGT. These proteins are TPR proteins involved in cell cycle control and transcription: Cdc616, Cdc27, and SSN6. SSN6 is very similar to TPR repeats to those of OGT and exhibits 42% similarity and 20% identity to both human and C. elegans OGT throughout its entire length. SSN6 is of particular interest since it is known to be involved in transcriptional repression in response to glucose; this is one of the proposed roles of OGT in mammalian cells (see below).
human and C. elegans OGTs have putative nuclear localization sequences immediately after the last TPR motif, although the human sequence has a single amino acid change rendering it less prototypic (Fig. 4A). Indirect immunofluorescence of the native enzyme in C. elegans embryos (Fig. 6) shows a nuclear and punctate perinuclear staining. Overexpression of the recombinant enzyme in C. elegans transgenic lines using heat shock promoters shows an accumulation of the enzyme in the nucleus and cytoplasmic aggregates. The transgenic lines expressed OGT in both the gut and neural tissues where the heat shock promoters preferentially drive expression of exogenous genes. It is likely that both the nuclear localization sequence and TPR repeats play a role in maintaining the steady-state localization of OGT observed in C. elegans transgenic lines.

**Overexpression of the OGT cDNA Is Sufficient to Increase Enzyme Activity in Human Cells**—Our studies suggest that expression of human OGT cDNA is sufficient to elevate the levels of OGT activity in a transfected HeLa cell population. Under conditions in which approximately 10–20% of the cells were expressing the exogenous cDNA, the enzyme activity was elevated 2–3-fold. Overexpression of the enzyme was apparently toxic, since expressing cells showed a much reduced growth rate compared with controls. The protein is also likely to interact with cellular proteins, which could alter its activity or modify it. Post-translational modification of the OGT, possibly by phosphorylation, could alter the level of glycosylation of cytoplasmic and nuclear proteins.

**Although Ubiquitously Expressed in Adult Human Tissues, OGT Is Enriched in the Pancreas**—The Northern analysis distinguishes four distinct OGT transcripts at 9.3, 7.9, 6.3, and 4.4 kb (Fig. 9). The signal in the pancreas is over 12-fold higher than seen in the lung and kidney. There appears to be a tissue-specific distribution of these different bands. The largest signals at 9.3 and 7.9 kb are most abundant in the pancreas and placenta, while the 6.3-kb transcript is the major signal seen in the other tissues. It is not known at this time if the multiple transcripts represent the transcription of different genes or alternative splicing and processing of the same gene. The large size of the mRNA transcripts compared with the
isolated clones and open reading frame of the gene presumably corresponds to extensive 5′ and 3′ untranslated sequences. This has been observed for a number of glycosyltransferases (22). The role of these larger regions of untranslated mRNA is not known, but it may be important in regulation of these genes. The human clones identified here also show variation in the polyadenylation signal, which could partially explain the different size of the messages.

The Hexosamine Biosynthetic Pathway and Glucose Metabolism—The hexosamine biosynthetic pathway is responsible for the synthesis of cytoplasmic UDP-GlcNAc utilized by OGT. Normally 2–3% of incoming glucose fluxes through this pathway (23). Increased glucose flux through the hexosamine biosynthetic pathway, caused by hyperglycemia, has been shown to mediate insulin resistance (23–28). The hexosamine biosynthetic pathway, by controlling intracellular UDP-GlcNAc concentrations may be acting in peripheral tissues as a glucose sensor, which is reflected in substrate-driven glucose homeostasis. Further suggesting a possible role in insulin secretion and glucose metabolism.

Similarity between Phosphorylation Sites and Sites of O-Linked GlcNAc Addition—O-Linked GlcNAc modifies many phosphoproteins, which are components of multimeric complexes. The sites modified by O-linked GlcNAc often resemble phosphorylation sites, leading to the suggestion that the modifications may compete for substrate in these polypeptides (9). In general, the sites modified by OGT very closely resemble those of the glycogen synthase kinases (GSK-3, casein kinase 2). The role of these large regions of untranslated mRNA is not known, but it may be important in regulation of these genes. The human clones identified here also show variation in the polyadenylation signal, which could partially explain the different size of the messages.

REFERENCES

1. Torres, C.-R., and Hart, G. W. (1984) J. Biol. Chem. 259, 3308–3317
2. Hanover, J. A., Cohen, C. K., Willingham, M. C., and Park, M. K. (1987) J. Biol. Chem. 262, 9887–9894
3. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, D. J., Gerace, L., and Hart, G. W. (1987) J. Cell Biol. 104, 1157–1164
4. Starr, C. M., and Hanover, J. A. (1990) J. Biol. Chem. 265, 6868–6873
5. Kelly, W. G., Dahmus, M. E., and Hart, G. W. (1993) J. Biol. Chem. 268, 10416–10424
6. Deng, D. Y., Xu, Z. S., Hart, G. W., and Cleveland, D. W. (1996) J. Biol. Chem. 271, 20845–20852
7. Macauly, C., Meier, E., and Forbes, D. J. (1995) J. Biol. Chem. 270, 254–262
8. Fisher, E. H., and Krebs, E. G. (1955) J. Biol. Chem. 216, 121–132
9. Hart, G. W., Gross, K. D., Deng, L. Y., Blomberg, M. A., Chou, T. Y., Jiang, M. S., Roquemore, E. P., Snow, D. M., Kreggel, K. L., Cole, R. N., Conner, F. I., Arnold, C. S., and Hayes, B. K. (1995) Adv. Exp. Med. Biol. 376, 115–123
10. Lobas, W. A., Smith, M., Starr, C. M., and Hanover, J. A. (1995) Biochemistry 34, 1686–1694
11. Kennelly, P. J., and Krebs, E. G. (1991) J. Biol. Chem. 266, 15555–15558
12. Haltiwanger, R. S., Blomberg, M. A., and Hart, G. W. (1992) J. Biol. Chem. 267, 9055–9063
13. Barstead, R. J., and Waterston, R. H. (1989) J. Biol. Chem. 264, 10177–10185
14. Krause, M., Fire, A., Harrison, S. W., Priess, J., and Weintraub, H. (1996) Cell 83, 907–919
15. Mello, C., and Fire, A. (1995) Methods Cell Biol. 45, 481–482
16. Favello, A., Hillier, L., and Wilson, R. K. (1995) Methods Cell Biol. 48, 551–569
17. Nyame, K., Cummings, R. D., and Damian, R. T. (1987) J. Biol. Chem. 262, 7990–7995
18. Heese-Peck, A., Cole, R. N., Barksenious, O. N., Hart, G. W., and Raikhel, N. V. (1995) Plant Cell 7, 1459–1471
19. Lath, R. V., Zengelreich, S., and Herder, P. (1995) Trends Biochem. Sci. 20, 257–259
20. Goebl, M., and Yanagida, M. (1991) Trends Biochem. Sci. 16, 173–177
21. Ownes-Grillie, J. K., Czar, M. J., Hutchison, R. A., Hoffmann, K., Perdew, G. H., and Pratt, W. B. (1996) J. Biol. Chem. 271, 13468–13475
22. Homa, F. L., Hollander, T., Lehman, D. J., Thomsen, D. R., and Elhammer, A. P. (1993) J. Biol. Chem. 268, 12609–12616
23. Marshall, S., Barote, V., and Traxinger, R. R. (1991) J. Biol. Chem. 266, 4706–4712
24. Rossetti, L., Hawkins, M., Chen, H., Gindi, J., and Barzilai, N. (1995) J. Clin. Invest. 96, 1320–1326
25. Daniels, M. C., Kansal, P., Smith, T. M., Paterson, A. J., Kudlow, J. E., and McClain, D. A. (1993) Mol. Endocrinol. 7, 1041–1048
26. Crook, E. D., Zhou, J., Daniels, M., Neidigh, J. L., and McClain, D. A. (1995) Diabetes 44, 315–320
27. Hebert, L. J. F., Daniels, M. C., Zhou, J., Crook, E. D., Turner, R. L., Simmons, S. T., Neidigh, J. L., Zhu, J. S., Baron, A. D., and McClain, D. A. (1996) J. Clin. Invest. 98, 930–936
28. Sayeski, P. P., and Kudlow, J. E. (1996) J. Biol. Chem. 271, 15237–15243
29. Balkan, B., and Dunning, B. E. (1994) Diabetes 43, 1173–1179
30. Zhou, Y., Deng, C. V., and Hart, G. W. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 4417–4421
31. Woodward, J. R. (1991) Trends Biochem. Sci. 16, 177–181
32. Stambolic, V., and Woodgett, J. R. (1994) Biochem. J. 303, Part 3, 701–704
33. Pyte, S. E., Hughes, K., Nikolakakis, E., Pulverer, B. J., and Woodgett, J. R. (1992) Biochim. Biophys. Acta 1114, 147–162
34. Towle, H. C. (1995) J. Biol. Chem. 270, 23235–23238
O-Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats

William A. Lubas, David W. Frank, Michael Krause and John A. Hanover

J. Biol. Chem. 1997, 272:9316-9324.
doi: 10.1074/jbc.272.14.9316

Access the most updated version of this article at http://www.jbc.org/content/272/14/9316

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 34 references, 22 of which can be accessed free at http://www.jbc.org/content/272/14/9316.full.html#ref-list-1