Dynamic O-Glycosylation of Nuclear and Cytosolic Proteins: Cloning and Characterization of a Neutral, Cytosolic β-N-acetylglucosaminidase from Human Brain*

Yuan Gao¹, Lance Wells¹, Frank I. Comer¹, Glendon J. Parker² and Gerald W. Hart¹,³

¹ Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205-2185
² Division of Endocrinology, Metabolism and Diabetes, Department of Internal Medicine, The University of Utah, Salt Lake City, Utah 84132
³ To whom correspondence should be addressed. Fax: (410) 614 8804. Phone: (410) 614 5993. E-mail: gwhart@jhmi.edu

* Supported by NIH grant HD13563.

Running title: Cloning of O-GlcNAcase from Human Brain
Abstract

Dynamic modification of cytoplasmic and nuclear proteins by $O$-linked $N$-acetylglucosamine ($O$-GlcNAc) on Ser/Thr residues is ubiquitous in higher eukaryotes, and is analogous to protein phosphorylation. The enzyme for the addition of this modification, $O$-GlcNAc transferase, has been cloned from several species. Here, we have cloned a human brain $O$-GlcNAcase that cleaves $O$-GlcNAc off proteins. The cloned cDNA encodes a polypeptide of 916 amino acids with a predicted molecular weight of 103 kD and a $pI$ value of 4.63, but the protein migrates as a 130 kD band on SDS-PAGE. The cloned $O$-GlcNAcase has a pH optimum of 5.5-7.0, and is inhibited by GlcNAc but not by GalNAc. $pNP-\beta$-GlcNAc, but not $pNP-\beta$-GalNAc or $pNP-\alpha$-GlcNAc, is a substrate. The cloned enzyme cleaves GlcNAc, but not GalNAc, from glycopeptides. Cell fractionation suggests that the overexpressed protein is mostly localized in the cytoplasm. It therefore has all the expected characteristics of $O$-GlcNAcase and is distinct from lysosomal hexosaminidases. Northern blots show that the transcript is expressed in every human tissue examined but is the highest in the brain, placenta and pancreas. An understanding of $O$-GlcNAc dynamics and $O$-GlcNAcase may be key to elucidating the relationships between $O$-phosphate and $O$-GlcNAc, and to the understanding of the molecular mechanisms of diseases such as diabetes, cancer and neurodegeneration.

Keywords: $N$-acetylglucosamine, $O$-GlcNAc, $O$-GlcNAcase, $\beta$-hexosaminidase, $\beta$-glucosaminidase, proteomics
Abbreviations used: BSA, bovine serum album; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; O-GlcNAc, O-linked N-acetylglucosamine; OGT, O-GlcNAc transferase; O-GlcNAcase, N-acetyl-β-D-glucosaminidase; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PMSF, phenylmethylsulfonyl fluoride; pNP, p-nitrophenyl; PUGNAc, O-(2-acetamido-2-deoxy-D-glucopyranosylidene)-amino-N-phenylcarbamate; SDS, sodium dodecyl sulfate.
INTRODUCTION

Since the description of O-linked N-acetylglucosamine (O-GlcNAc) as an abundant modification in murine lymphocytes (1), a myriad of cytoplasmic and nuclear proteins in all metazoans have been found to carry this modification. Such proteins cover a broad range, including many transcription factors, RNA polymerase II, oncogenes, nuclear pore proteins, viral proteins and tumor repressors (for details, see 2, 3 and citations within). Unlike classical O- or N-linked protein glycosylations, the O-GlcNAc modification involves only a single GlcNAc moiety linked to the hydroxyl group of Ser/Thr residues, generally is not elongated, and is found exclusively in the cytoplasm and nucleoplasm.

Protein O-GlcNAcylation is highly dynamic, and the cycle of addition/removal of the sugar moiety is rapid, analogous to protein phosphorylation/dephosphorylation catalyzed by kinases and phosphatases (2). Indeed, existing evidence suggests that this modification has a ‘yin-yang’ relationship with protein phosphorylation in some cases (4). Many O-GlcNAcylation sites have been mapped to phosphorylation sites or adjacent sites (5, 6, 7). Such spatial localization indicates that O-GlcNAc may regulate the target protein by competing with protein kinases (4). Recent studies using phosphatase and kinase inhibitors have provided direct evidence for a general reciprocal relationship between O-GlcNAcylation and phosphorylation on some proteins (8, 9).

O-GlcNAcylation appears to be involved in gene transcription. Most transcription factors examined so far, including Sp1, AP1, AP2, AP4 (10), serum response factor (11), the estrogen receptor (7, 12), the insulin promoter factor-1 (IPF-1) and peroxisome proliferator-activated receptor-γ (PPAR-γ), as well as RNA polymerase II (13) and chromatin (14) are O-GlcNAcylated. O-GlcNAcylation of Sp1 appears to enhance its activity in transcription and
conversely, blocking the GlcNAc residues with lectin wheat germ agglutinin (WGA) suppresses the transcriptional activity (10). O-GlcNAcylation of Sp1 also controls its degradation by the proteosome (15). Hyperglycemia induced superoxide production increases Sp1 glycosylation resulting in the activation of genes that contribute to the pathogenesis of diabetes (16).

O-GlcNAc transferase (OGT), which transfers GlcNAc from the donor substrate UDP-GlcNAc to target proteins, has been purified and cloned from several species including human, rat and C. elegans (17, 18, 19). It does not share any significant homology with any other known proteins including glycosyltransferases, and is highly conserved from C. elegans to human. Disruption of the ogt gene is lethal in mouse embryonic stem cells, further underlining the importance of O-GlcNAc modification in cellular functions (20). O-GlcNAcase, the enzyme that removes O-GlcNAc from such proteins, was purified several years ago from rat spleen (21). It is a neutral cytosolic β-glucosaminidase or hexosaminidase C (EC 3.2.1.52). To further study the function of this modification, we have now extensively purified O-GlcNAcase from bovine brain, sequenced the protein by mass spectrometry and cloned the cDNA. The O-GlcNAcase is evolutionarily conserved, distinct from lysosomal acidic hexosaminidases A and B. The recombinant protein has all the expected characteristics of O-GlcNAcase, including the ability to cleave O-GlcNAc from glycopeptides.
EXPERIMENTAL PROCEDURES

Purification of O-GlcNAcase from Bovine Brain- All chromatographic materials were purchased from Amersham-Pharmacia Biotech (Piscataway, NJ). Purification procedure is substantially modified from that of Dong and Hart (21). All steps were conducted at 4°C or on ice.

Step 1. Tissue Homogenization- Three bovine brains (~1 kg) were frozen in liquid N2 and shipped on dry ice from Pel Freez Biologicals (Rogers, AR) and stored at -80°C until use. The brains were smashed into smaller pieces and homogenized in 5 volumes (v/w) of homogenization buffer (20 mM sodium phosphate, pH 7.5, 15 mM 2-mecaptoethanol, 10 mM MgCl2, 1 mM PMSF, 1 mM EDTA) in a Hamilton Beach blender with 5x20 s bursts. The homogenate was centrifuged at 18,000g for 30 min. The pellet was discarded and the cytosolic supernatant was pooled in a 5 L beaker.

Step 2. Ammonium Sulfate Precipitation- The cytosolic supernatant was subjected to 30-50% ammonium sulfate precipitation. The pellet was resuspended in 500 mL of Buffer A (20 mM sodium phosphate, pH 7.5, 5 mM 2-mercaptoethanol), and centrifuged to clarify the solution. The solution was then thoroughly dialyzed against buffer A and centrifuged again to eliminate any insoluble materials that had resulted from dialysis.

Step 3. DE52 Cellulose Ion Exchange Chromatography- The dialyzed sample was loaded onto a DE52 column (900 mL bed vol) at a flow rate of 2 mL/min using a peristaltic pump. After washing the column with 3 L of buffer A, bound proteins were eluted with a linear gradient of 0-1 M NaCl in 4 L of buffer A at a flow rate of 4 mL/min. The protein profile was monitored by absorbance at 280 nm. Fractions (16 mL) enriched in O-GlcNAcase activity were pooled.
**Step 4. Concanavalin A-Sepharose 4B Chromatography** - MgCl$_2$ was added to the pooled fractions at a final concentration of 1 mM. The preparation was then applied to a Con A column (60 mL) equilibrated in Con A buffer (20 mM sodium phosphate, pH 7.5, 5 mM 2-mercaptoethanol, 150 mM NaCl, 1 mM MgCl$_2$). The column was washed with 200 mL of Con A buffer. The flow through and the wash were combined.

**Step 5. Affinity Blue A Chromatography** - The enzyme solution from Step 4 was concentrated by 60% ammonium sulfate precipitation, dialyzed and applied three times to a Blue A sepharose column (25 mL) equilibrated in Buffer A. Again the activity was present in the flow through fraction. The protein was pooled and clarified by centrifugation.

**Step 6. Re-chromatography on DE52 Column** - The sample from Step 5 was injected to the DE52 cellulose column (same size as above) and protein was eluted with a linear gradient of 50-350 mM NaCl in 4 L of buffer A. Activity was recovered as in Step 3 and precipitated with 60% ammonium sulfate. The pellet was resuspended in 20 mL of Mono-Q Buffer (20 mM Tris pH 7.5, 5 mM 2-mercaptoethanol, 10% glycerol, 1 mM EDTA plus protease inhibitor cocktail (Pics 1 & 2) (22) and 1 mM PMSF, dialyzed and clarified by centrifugation.

**Step 7. Native Polyacrylamide Gel Electrophoresis** - Native PAGE was performed using a preparative Prepcell apparatus (Bio-Rad. Hercules, CA). The sample from Step 6 was divided into 3 equal volumes (45 mg protein each) and loaded batch-wise onto a 6% native polyacrylamide gel (5 cm long separating gel). The gel was run for 24 h at 12 watts constant power. Protein was eluted in Mono-Q buffer at a flow rate of 0.75 mL/min. Five min fractions were collected and assayed for protein content and enzyme activity.

**Step 8. Mono-Q Chromatography** - The O-GlcNAcase containing fractions from each native PAGE run was resolved on a Mono-Q column (HR10/10) with a linear gradient of 0-500
mM NaCl in 450 mM Mono-Q buffer at a flow rate of 3 mL/min. Fractions (4.5 mL) rich in O-GlcNAcase activity were pooled and then separated for a second time on the Mono-Q column. The final preparation was concentrated using Millipore concentrators to a final volume of 0.4 mL. Glycerol (40% final) and 1 mM PMSF and protease inhibitor cocktail were added to the preparation. The enzyme was stored at -20°C.

**Identification of Proteins by Mass Spectrometry** - The final preparation from Step 8 was separated by 10% SDS-PAGE and stained with Coomassie G-250 or with silver. The desired protein bands were excised individually, reduced and alkylated, and in-gel digested with modified trypsin (Worthington, Freehold, NJ) as described (23). The tryptic peptides from each protein were analyzed by capillary reversed phase HPLC with in-line tandem ms/ms on a Finnigan LCQ. Proteins were identified by the SEQUEST algorithm with sequencing at least seven tryptic peptides for each protein (24).

**Cloning of O-GlcNAcase** - A putative O-GlcNAcase with a theoretical length of 916 amino acids in human was identified by the above proteomic approach. A cDNA fragment, KIAA0679 (Genbank accession no: AB014579), which contains the coding sequence for 767 amino acids of the C-terminus of the human O-GlcNAcase and a 2.0 kb 3'-untranslated region, was obtained from the Kazusa DNA Research Institute, Japan, in the vector pBluescript. The coding sequence of this fragment was subsequently transferred to pcDNA3.1His A using Xho I and Xba I, which were located within the polycloning cloning site of pBluescript and in the 3’-untranslated region of the cDNA, respectively. The missing 5’-end fragment of the full length coding cDNA (447 bp) was amplified by PCR from a human brain Marathon cDNA library (Clontech, Palo Alto, CA) using the forward primer GGATGGTGAGAAGGAGAGTCAAGCGAC and the reverse primer
TAGAAACCTCTTCGATGGACTCTACTGG. The forward primer sequence was based on published data (25) and the reverse primer was located in the KIAA0679 clone. PCR conditions were 94°C, 30 s, 63°C, 30 s and 72°C, 3 min for 30-35 cycles. A second round of PCR using the first PCR product as template and a forward primer incorporating a Not I site (CCGGGCGGCCGGATGGTGCAGAAGGAGAG) and the same reverse primer was performed. The product was digested with Not I and HindIII (unique site in the PCR product), and ligated in-frame into the pcDNA3.1His A-KIAA0679 construct. This gave rise to a full length cDNA in the vector pcDNA3.1His A. The final construct was sequenced.

**O-GlcNAcase Assays**- Unless stated otherwise, O-GlcNAcase activity was assayed as described in 50 mM sodium cacodylate, pH 6.5, 0.3% BSA, 2 mM pNP-β-GlcNAc, 50 mM GalNAc (21). Purified bovine kidney lysosomal β-hexosaminidase (Roche. Indianapolis, IN) activity was assayed in citrate phosphate buffer, pH 4.5, 0.3% BSA, 2 mM pNP-β-GlcNAc. To test the ability of recombinant O-GlcNAcase to cleave O-GlcNAc from glycopeptides, two glycopeptides, CTD-GlcNAc [N-YSPTS(GlcNAc)PSK-C] or CTD-GalNAc [N-YSPTS(GalNAc)PSK-C], were synthesized by standard Fmoc chemistry. The peptides were purified on a C18 column under reversed phase HPLC conditions, and used as a substrate for cloned O-GlcNAcase. The reaction products were analysed by Matrix Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF).

**Overexpression and Purification of O-GlcNAcase from Cos-7 Cells**- Plasmid of pcDNA3.1His A containing the full length O-GlcNAcase cDNA was prepared using Qiagen kit (Qiagen. Valencia, CA). Transfection was mediated by Lipofectamine Plus (Life Technologies. Gaithersburg, MD) using 50-90% confluent Cos-7 cells. Cells were harvested two days post transfection and sonicated for 2 x 12 s in 20 mM Tris (pH 7.5), 10% glycerol, 150 mM NaCl, 1
mM DTT, 0.1 mM EDTA, 1 mM PMSF and protease inhibitor cocktail, and clarified by centrifugation. For characterizations, the recombinant protein was purified over a nickel affinity column.

**Cell Fractionation and Western Blot**- After transfection with O-GlcNAcase, Cos-7 cells were separated into cytoplasmic and nuclear fractions as described (26) with one modification: the 25-50 µl nuclear pellet was carefully washed in 500 µl of hypotonic buffer A to minimize cross contamination. Immunoblot analysis was performed using antibodies recognizing the nuclear protein retinoblastoma (Rb) (Santa Cruz, Santa Cruz, CA), cytoplasmic protein α-tubulin (Sigma, St. Louis, MO), or with anti-Xpress antibody which is specific for the sequence DLYDDDDK located at the N-terminus of the overexpressed O-GlcNAcase fusion protein (Invitrogen, Carlsbad, CA).

**Northern Blot**- Northern blot analyses were performed on a human Multiple Tissue Northern (MTN) blot (Clontech) using the manufacturer’s protocol. To prepare an O-GlcNAcase specific probe, the full length coding sequence (2.75 kb) was amplified by PCR and labeled by random primer using [α-32P]dCTP (Stratagene, La Jolla, CA). After stripping in 0.5% SDS at 100°C for 10 min, the blot was re-probed for β-actin.
RESULTS

Native Polyacrylamide Gel Electrophoresis Aides in O-GlcNAcase Purification from Bovine Brain- Historically, O-GlcNAcase, or neutral hexosaminidase C, has been difficult to purify. For example, an early report described a purification of only 25-40 fold from bovine brain despite the extensive use of chromatographic steps (27). While in rat brain, the enzyme has been purified over 2,000 fold to a major band (28). More recently, renewed effort has gone into its purification from rat spleen and bovine brain in the pursuit to cloning the cDNA (21, 29).

We have taken an approach, partly based on published literature (21) but yet incorporating some novel steps in the purification of O-GlcNAcase from bovine brain. Notably, we have discovered that the enzymatic activity survives the harsh conditions of native PAGE (high pH and high ionic strength), and migrates more slowly (Rf = 0.28 in 6% native gel) than most other proteins in the gel (data not shown). This property allows the effective separation of O-GlcNAcase from other proteins with higher Rf values on a preparative scale native gel (Fig. 1b). This step, in conjunction with other chromatographic steps outlined in the protocol, purified the protein approximately 1,500 fold, with a specific activity of 1,840 nmol/min/mg protein.

The final preparation still shows seven well defined bands on SDS-PAGE following silver staining, even after extensive purification (Fig. 2). We do not understand the basis for this difficulty, but we have observed that the peaks for O-GlcNAcase activity are very broad throughout the purification procedure (Fig. 1). One example of this is illustrated in the Mono-Q step, where the general protein peaks are sharp but yet the activity peak spreads over 50 mL (Fig. 1c). We have also tried ion exchange on Superose Q and hydroxyapatite columns or hydrophobic interaction chromatography on a phenyl sepharose column. They, too, give poor separations or
the enzyme binds very tightly to phenyl sepharose resulting in >50% activity loss (data not shown).

Mass Spectrometry Identifies O-GlcNAcase on SDS-PAGE Gel- The seven bands on the silver stained SDS-PAGE gel were excised individually, digested with trypsin and sequenced by electrospray ms/ms. The fragmentation data were used to search protein and DNA databases. This approach identified six proteins with known functions in six of the bands (Fig. 2). Another protein, which runs as a 130 kD band on SDS-PAGE, is a hypothetical protein without any clearly defined functions (KIAA0679). Blast searches indicated that the hypothetical protein shared significant homology with a protein from *C. elegans* called ‘Similar to Hyaluronoglucosaminidase’ (AAA68333.1). Since hyaluronoglucosaminidase degrades hyaluronic acid which is a GlcNAcβ1-4GlcUA polymer, it was possible that a hyaluronoglucosaminidase may share some homology with O-GlcNAcase. Furthermore, careful comparisons of O-GlcNAcase activity and protein patterns on the SDS gels of different pools during the purification procedure indicated that this protein was one of only two bands that corresponded with activity (The other band was #1 in Fig. 2, data not shown). We therefore hypothesized that this may be the O-GlcNAcase, and cloned the cDNA. Further characterization of the expression product of the cDNA confirmed that band 2 on Fig. 2 was, indeed, O-GlcNAcase (see below).

O-GlcNAcase is Unique and Conserved during Evolution- Blast searches of databases reveal that O-GlcNAcase is conserved in higher eukaryotic species and the homologue is absent in yeast or prokaryotes. The sequences and alignment of O-GlcNAcase from human, *C. elegans* and *Drosophila* are shown in Fig. 3. In a pair-wise alignment, the human sequence shares 55% and 43% homology with that of *Drosophila* and *C. elegans*, respectively, while *Drosophila* and
C. elegans are 43% similar. Close inspection of the sequences indicate that the N-terminal ~400 and the C-terminal ~350 amino acids in the human sequence is conserved to a higher degree. These two domains are separated by a highly variable region of ~150 amino acids. Another feature is that most of the aromatic residues are conserved among the species. For example, out of the 13 Trp residues found in the human sequence, 9 are invariant in Drosophila and C. elegans, two are conservative (substituted by Tyr or Phe), and only two are variable.

The O-GlcNAcase sequence is conserved at a strikingly higher level in mammals. Four overlapping EST sequences from cow, which cover 46% of the human protein, show that these two species are 100% identical in these regions (BE481597, BE588694, BF043559 and AW463869). Five EST entries for mouse, most of which are overlapping, show that human and mouse are 97.8% identical (AW907793, AW324047, AI530529, AW762257 and AA240394). In the case of zebrafish, two overlapping EST sequences covering 33.8% of the human protein indicate that zebrafish and human are 85% identical and 92% similar (AI882982 and AI722710).

Apart from the above described homologs, O-GlcNAcase does not show significant homology with any other proteins, including known glycosidases. Short stretches of ~200 amino acids of the polypeptide do show loose homology to a number of proteins such as hyaluronidase (AAA23259.1), a putative acetyltransferase (AL158057), eukaryotic translation elongation factor-1γ (Z11531, S26649) and the 11-1 polypeptide (X07453, S00485). Sequence analyses by a computer program PSORT II (http://psort.nibb.ac.jp) show that O-GlcNAcase does not possess any known signal peptides, domains or motifs. The analyses do, however, suggest that the endogenous protein is localized in the cytoplasm \( (p=0.522) \) and the nucleus \( (p=0.391) \).
Overexpression of O-GlcNAcase in Cos-7 Cells- To ascertain that the cloned cDNA indeed encoded O-GlcNAcase, we subcloned the entire coding region in-frame into the mammalian expression vector pcDNA3.1His and overexpressed for activity in Cos-7 cells. Transient transfection resulted in a 6-fold increase in O-GlcNAcase activity over endogenous activity in the cells (Fig. 4a). After nickel affinity purification, the activity from the O-GlcNAcase transfected cells was 230 nmol/min/mg protein, but was not detectable from control transfected cells (Fig. 4b). These data show that the activity is due to overexpression from the plasmid. Fig. 4c shows that a distinct band of the correct molecular weight (135 kD) was isolated after nickel purification from transfected cells. This band was immunoreactive the Xpress antibody, which was specific for a peptide sequence in the overexpressed protein.

Recombinant O-GlcNAcase has Distinct Properties from Lysosomal β-hexosaminidase-

We further characterized the properties of the cloned O-GlcNAcase and compared them with those of lysosomal β-hexosaminidase purified from bovine kidney. As expected, the lysosomal β-hexosaminidase had an acidic pH optimum (pH 3.5-5.5) with little activity at pH 7.0 or above (Fig. 5a). On the other hand, the cloned O-GlcNAcase had a pH optimum of 5.7-7.0, and retained significant activity (~30%) at pH 7-8. This pH profile is consistent with the expected localization of O-GlcNAcase in the cytoplasm and the nucleus.

The two enzymes also responded differently to inhibitors. GalNAc, a widely used inhibitor of acidic β-hexosaminidase, inhibited the lysosomal enzyme 50% at 5.0 mM and 88% at 50 mM. The cloned O-GlcNAcase was not inhibited at all by GalNAc up to 50 mM (Fig. 5b). GlcNAc and its synthetic analogue PUGNAC inhibited both enzymes but were more potent with the O-GlcNAcase (Fig. 5 c, d).
Recombinant O-GlcNAcase Shows Strict Substrate Specificity for β-linked GlcNAc- O-GlcNAcase also differed from lysosomal β-hexosaminidase in substrate requirements. In the in vitro assays, purified recombinant O-GlcNAcase cleaved only pNP-β-GlcNAc, but not pNP-β-GalNAc or pNP-α-GlcNAc (Fig. 6). The activity using the latter two compounds as substrates was not detectable. This substrate specificity was in contrast to the lysosomal β-hexosaminidase, which also cleaved pNP-β-GalNAc, albeit with slightly lower efficiency compared to pNP-β-GlcNAc.

O-GlcNAcase Cleaves O-GlcNAc from Glycopeptides- A true O-GlcNAcase should cleave O-GlcNAc attached to proteins or peptides. We synthesized two glycopeptides containing one repeat of the C-terminal domain of RNA polymerase II linked to β-GlcNAc or α-GalNAc through the hydroxyl group of a serine residue [CTD-GlcNAc or CTD-GalNAc]. The design of these glycopeptides is based on earlier information that this serine residue is glycosylated in vivo (13). These peptides were tested as substrates for the purified recombinant O-GlcNAcase and the product peptides were analyzed by Matrix Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF). Cleavage of GlcNAc or GalNAc from the peptides should result in a down shift of 203 in the molecular weight (the mass of GlcNAc or GalNAc minus 18). O-GlcNAcase did not cleave GalNAc from the peptide (Fig. 7a and b), but did successfully cleave GlcNAc from the peptide, as judged by the expected shift in molecular weight (1069.8 to 866.6, 1091.8 to 888.6) (Fig. 7 c and d).

Overexpressed O-GlcNAcase is Localized to the Cytoplasm- As stated earlier, sequence analyses suggest that O-GlcNAcase is localized in the cytoplasm and nucleus. This is consistent with the localization of the O-GlcNAc modification. To obtain direct evidence on its localization, we performed cellular fractionation and assayed O-GlcNAcase activity in the
cytoplasm and the nucleus. The data show that in non-transfected cells, O-GlcNAcase activity was distributed in both the cytoplasm and the nucleus. However, when O-GlcNAcase was overexpressed, it was predominantly found in the cytoplasm (Fig. 8a). We also probed for its localization by Western blots (Fig. 8b). Retinoblastoma protein (Rb) and α-tubulin, which were exclusively localized in the nucleus and cytoplasm, respectively, were used as markers. In agreement with activity assays, overexpressed O-GlcNAcase protein was only detected in the cytoplasm of overexpressed cells.

**O-GlcNAcase Has One Transcript in Human Tissues** - To estimate the number of transcripts of O-GlcNAcase and their levels of expression, we performed Northern analyses using a human multiple tissue blot consisting of RNA from eight different tissues. Labeled PCR product of the entire coding region of the O-GlcNAcase cDNA was used as a probe. The Northern blot analysis showed only one transcript of ~5.5 kb for O-GlcNAcae (Fig. 9). Exposing the film for extended time (32 h) did not reveal any additional bands (data not shown). The gene was expressed in every tissue on the blot, but was the highest in the brain, followed by placenta and pancreas. Lung and liver had the lowest expression (Fig. 9). This pattern of expression largely agrees with that of the ogt gene. ogt is expressed the highest in the pancreas, followed by heart, brain and placenta, but the lowest in lung, liver and kidney (18). Unlike ogt gene which is on the X chromosome (20), the O-GlcNAcase gene is localized on chromosome 10 (25, http://www.kazusa.or.jp/huge/gfpage/KIAA0679/).
DISCUSSION

Two categories of β-hexosaminidases are known to exist in eukaryotic cells. One category, which comprises the A, B, I, S isozymes, are exclusively localized in the lysosomes and are responsible for the degradation of complex glycans. This group of hexosaminidases, particularly the A and B isozymes, have been extensively studied because their deficiency leads to Tay-Sachs and Sandhoff diseases (30, 31). These enzymes are characterized by their acidic pH optima, inhibition by both GlcNAc and GalNAc, the ability to use both artificial glucosaminide and galactosaminide as substrates, and their thermostability (32). The second category, on the other hand, consists of two neutral hexosaminidases, the GlcNAc-specific glucosaminidase (hexosaminidase C) and the GalNAc-specific galactosaminidase (hexosaminidase D) (28, 33, 34). In contrast to lysosomal hexosaminidases, they reside in the cytosol, have neutral pH optima, and are heat labile. Despite the wide occurrence of these neutral isoforms in tissues, their natural substrates have not been previously identified (28).

The discovery of O-GlcNAc provides clues to the likely physiological function of neutral glucosaminidase (hexosaminidase C) (2). This modification consists of the addition of a single GlcNAc moiety to the hydroxyl group of Ser/Thr residues (2). The enzyme responsible for GlcNAc addition, OGT, was identified and cloned several years ago (17, 18, 19). On the other hand, the identity of the enzyme for the removal of O-GlcNAc from proteins, O-GlcNAcase, has remained obscure. Judged from the known properties of hexosaminidase C, we speculate that it might in fact be the O-GlcNAcase. We thus have extensively purified hexosaminidase C from bovine brains. Proteomic analysis allowed us to identify its DNA sequence. The sequence was originally isolated during screening of a meningioma expression library (25). Based on its initial characterization and sequence homology with an unidentified protein called ‘Similar to
Hyaluronoglucosaminidase’ from C. elegans, the authors suggested that it was a new type of hyaluronidase. We have here demonstrated that the cloned enzyme has all the expected properties of O-GlcNAcase. Most importantly, it specifically cleaves O-GlcNAc, but not O-GalNAc from glycopeptides. Based on this information, we conclude that we have cloned hexosaminidase C, which we have called O-GlcNAcase herein.

The recombinant protein differs from acidic lysosomal hexosaminidase in a number of ways. It has a neutral, instead of acidic, pH optimum. This pH optimum is expected due to its physiological functions in the cytoplasm and nucleus. The protein displays strict substrate requirement in the linkage of the sugar and is also sensitive to the C4 orientation of the sugar. As such, it hydrolyzes only β-linked GlcNAc but is totally ineffective towards α-linked GlcNAc or β-linked GalNAc. This substrate specificity is in contrast with that of lysosomal hexosaminidases, which hydrolyze both GlcNAc and GalNAc substrates. It also determines that GlcNAc, but not GalNAc, is a competitive inhibitor of O-GlcNAcase. Unlike lysosomal hetero-oligomeric hexosaminidases A (αβαβ) or B (2αβαβ) (32), the cloned human O-GlcNAcase has only one polypeptide of 916 amino acids with a predicted molecular weight of 103 kD. The apparent size on SDS-PAGE is 130 kD, which is identical to the glucosaminidase purified from rat brain (28). The oligomeric status of the hexosaminidase C is still not known, as gel filtration, SDS-PAGE and sucrose density gradient centrifugation give inconsistent estimations of its native molecular size (28).

O-GlcNAcase is a unique protein with no other obvious family members. It is conserved in evolution down to the nematode C. elegans, and shows striking homology in mammals. It does not, however, possess any known signal peptides, domains or motifs. Consistent with the literature that antibodies that recognize lysosomal hexosaminidases A or B do not cross react
with hexosaminidase C (35), O-GlcNAcase does not share any significant homology with these enzymes, or with any other known proteins.

The predicted localization of O-GlcNAcase by sequence analysis is consistent with the literature that O-GlcNAc modification and OGT are found both in the cytoplasm and nucleus (19, 36). It is therefore somewhat unexpected to find that overexpressed O-GlcNAcase in Cos-7 cells is almost exclusively localized to the cytoplasm (Fig. 8). Nevertheless, it is not uncommon that overexpressed proteins often form aggregates in the cytoplasm, and can not be correctly translocated. We cannot exclude the possibility of an unidentified isoform of O-GlcNAcase that is specifically localized to the nucleus. We are currently generating an O-GlcNAcase specific polyclonal antibody based on its protein sequence. Immunofluorescence microscopic studies will determine the relative distribution of the endogenous protein.

The cloning of O-GlcNAcase provides a valuable tool to further study the biological function of O-GlcNAc modification and may help to understand the mechanisms of a number of prevalent diseases such as the Alzheimer’s and diabetes. Recent data suggest that many proteins in the brain, including Synapsin I and neurofilaments (6, 37), are O-glycosylated. Some of the glycosylated proteins may be involved in the development of the Alzheimer’s diseases. For example, Tau, whose hyperphosphorylation leads to the formation of neurofibrillary tangles in the neurons of Alzheimer’s brains, is extensively O-glycosylated (38). The β-amyloid precursor protein, which gives rise to the neurotoxic β-amyloid peptide in Alzheimer’s brains, is also O-GlcNAcylated (39). Furthermore, Alzheimer’s disease has been correlated to the glycosylation of at least one protein, the chathrin assembly protein-3 (40). Continuing efforts in several laboratories are underway to understand the role of O-GlcNAc in the development of such neurodegenerative diseases. In addition, strong evidence suggests that O-GlcNAc is involved in
the development of insulin resistance in diabetes mellitus (4, 41). Infusion of glucosamine or the overexpression of glutamine:fructose-6-phosphate amidotransferase (GFAT) in animal models, both of which increase cellular UDP-GlcNAc levels through the hexosamine synthetic pathway (42, 43), leads to insulin resistance (44, 45). The cloning of $O$-GlcNAcase will not only allow the selective disruption of the gene, or tissue specific overexpression of the protein in animal models, but will also make it possible to directly regulate and monitor $O$-GlcNAc modified proteins, and thus facilitate the understanding of the role of this modification in the development of these diseases.
Acknowledgment: The authors would like to thank the Kazusa DNA Research Institute, Japan for providing the KIAA0679 cDNA, and Dr. N. Zachara for reading the manuscript.
Footnote:

5 Y. Gao and G.W. Hart, manuscript in preparation
Figures and Legends

Fig. 1. Purification steps of O-GlcNAcase from bovine brain. The cytosolic fraction of bovine brain proteins were sequentially purified by precipitation with 30-50% ammonium sulfate, DE52 ion exchange, Con A and Blue A affinity columns, then re-separated on DE52 cellulose ion exchange chromatography (a), native polyacrylamide gel electrophoresis (b) and Mono-Q ion exchange chromatography (c). The protein profiles in a and c were continuously monitored by absorbance at 280 nm with a UV detector during purification on an FPLC system, and are presented by solid lines without symbols. In b, the protein contents were manually determined by Bio-Rad assay (OD595) for each fraction after the protein was eluted from the native gel. So it is presented with symbols. See experimental procedures for details. ●, protein profile; ▲, O-GlcNAcase activity.

Fig. 2. Silver staining and protein identification by mass spectrometry. The final purification was separated on 10% SDS-PAGE and silver stained. Each protein band was individually excised, digested with modified trypsin and the resultant peptides were separated and sequenced by LC-ms/ms. Seven to 22 peptides were identified for each protein.

Fig. 3. The predicted amino acid sequence of O-GlcNAcase and its alignment with hypothetical proteins in C. elegans (`Similar to Hyaluronoglucosaminidase`. AAA68333.1) and Drosophila (AAF55867.1). The alignment was done with Clustal W in MacVector. Identical and conservative residues are indicated by asterisks and dots, respectively. The underlined sequences
are the typtic peptides that are identified by LCQ-ms/ms. Bold and underlined letters indicate conserved aromatic residues in the three species.

Fig. 4. Overexpression of O-GlcNAcase in Cos-7 cells. Cos-7 cells were transiently transfected with the vector pcDNA3.1His alone or with pcDNA3.1His containing O-GlcNAcase cDNA using Lipofectamine Plus. 48 h post transfection, soluble proteins were isolated from the cells. (a), O-GlcNAcase activity in total cell extract. The activity in control transfected cells is arbitrarily set as 1 fold. (b), O-GlcNAcase activity in control and transfected cells after nickel purification. (c), Western blot and Coomassie G-250 staining of nickel affinity purified O-GlcNAcase from transfected cells. The primary antibody was anti-Xpress which recognized a peptide sequence (DLYDDDDK) present in the vector pcDNA3.1His. Arrow indicates the band of O-GlcNAcase.

Fig. 5. Some key properties of the cloned O-GlcNAcase. A lysosomal β-hexosaminidase purified from bovine kidney was used as a control enzyme. (a), pH optima of the two enzymes. Inhibition studies by GalNAc (b), GlcNAc (c) and PUGNAc (d). In b, c and d, the activity was assayed at pH 4.5 for the lysosomal hexosaminidase and pH 6.5 for the O-GlcNAcase. The highest activity at optimal pH (a) or in the absence of any inhibitor (b, c, d) is arbitrarily set as 100%. ●, O-GlcNAcase; ▽, lysosomal β-hexosaminidase.

Fig. 6. Substrate specificity of Cloned O-GlcNAcase. pNP-β-GlcNAc, pNP-β-GalNAc and pNP-α-GlcNAc (all 2 mM) were tested as substrates for purified O-GlcNAcase from transfected Cos-7 cells. The lysosomal hexosaminidase was used as a control. The assays were done at pH 4.5 for
the lysosomal enzyme and at pH 6.5 for \( O\text{-GlcNAcase} \). The activity using \( p\text{NP-\( \beta \text{-GlcNAc} \) as substrate is set as 100\%}.

Fig. 7. Recombinant \( O\text{-GlcNAcase} \) cleaves \( O\text{-GlcNAc} \), but not \( O\text{-GalNAc} \), from glycopeptides. Synthetic peptides, CTD-GlcNAc \([N\text{-YSPTS(GlcNAc)PSK-C}]\) or CTD-GalNAc \([N\text{-YSPTS(GalNAc)PSK-C}]\) were tested as substrates for purified recombinant \( O\text{-GlcNAcase} \). The reactions containing 0.1 mM peptide were incubated at pH 6.5, 37\(^\circ\)C overnight. The mock reactions were done using nickel purification from non-transfected Cos-7 cells. The peptides were then cleaned up by zip tips and analyzed by MALDI-TOF. The values 1091.7 (or 1091.8) and 1069.7 (or 1069.8) represented molecular weight of the \( \text{Na}^+ \) and \( \text{H}^+ \) form of the peptides, respectively. After the GlcNAc was cleaved in CTD-GlcNAc peptide by \( O\text{-GlcNAcase} \), the mass expectedly shifted down to 888.6 and 866.6, respectively. The numbers on the right hand side of each spectrum (2744, 2930, 5158 and 2100) are the total ion counts recorded by the detector in MALDI-TOF analysis.

Fig. 8. Overexpressed \( O\text{-GlcNAcase} \) is distributed in the cytoplasm. Cos-7 cells were transfected with \( O\text{-GlcNAcase} \). Two days post transfection, the cells were fractionated into cytoplasmic or nuclear fractions. (a), \( O\text{-GlcNAcase} \) activity (\( n=3 \)). Assays were done with 100 mM GalNAc as inhibitor. (b), Western blots with anti-Xpress (for recombinant \( O\text{-GlcNAcase} \) detection), anti-\( \alpha \)-tubulin (cytoplasmic marker) and anti-Retinoblastoma (Rb) (nuclear marker).
Fig. 9. O-GlcNAcase transcript is expressed in every human tissue examined but is the highest in the brain, placenta and pancreas. (a), a human multiple tissue blot was probed with [α-32P] dCTP labeled full length O-GlcNAcase coding sequence and exposed to film for 18 h. (b), after stripping, the blot was re-probed for β-actin and exposed to film for 7 h.
REFERENCES


   *Biochem. J.* **151**, 257-261


   *J. Biol. Chem.* **271**, 28741-28744


   *Metabolism* **47**, 449-455

   (1999) *Endocrinol.* **140**, 1151-1157

44. Hebert, L. F. Jr, Daniels, M. C., Zhou, J., Crook, E. D., Turner, R. L., Simmons, S. T.,

    Clin. Invest.* **96**, 2792-2801
Fig. 1
Electrospray Identification of Proteins

<table>
<thead>
<tr>
<th>#</th>
<th>Protein Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TATA binding protein-interacting-protein (TIP-120, 135 kD)</td>
</tr>
<tr>
<td>2</td>
<td>O-GlcNAcase (KIAA0679)</td>
</tr>
<tr>
<td>3</td>
<td>Heat shock protein 110</td>
</tr>
<tr>
<td>4</td>
<td>Cullin</td>
</tr>
<tr>
<td>5</td>
<td>DNA K-type molecular chaperone hsc70</td>
</tr>
<tr>
<td>6</td>
<td>Amphiphysin</td>
</tr>
<tr>
<td>7</td>
<td>Dihydropyrimidinase related protein-2</td>
</tr>
</tbody>
</table>
Fig. 4

a. Total Cell Extract

b. After Purification

c. Coomassie Staining and Western Blot
Fig. 5
Fig. 6
Fig. 7

a, CTD-GalNAc, Mock

b, CTD-GalNAc + O-GlcNAcase

c, CTD-GlcNAc, Mock

d, CTD-GlcNAc + O-GlcNAcase

Mass (m/z)

% Intensity
Fig. 8

a

<table>
<thead>
<tr>
<th>Transfection</th>
<th>Cytoplasmic</th>
<th>Nuclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0.98</td>
<td>0.39</td>
</tr>
<tr>
<td>+</td>
<td>15.71</td>
<td>0.82</td>
</tr>
</tbody>
</table>

b

<table>
<thead>
<tr>
<th>Kd</th>
<th>Cytoplasmic</th>
<th>Nuclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Transfection
- +
Fig. 9

a, O-GlcNAcase

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart</td>
<td>9.5</td>
</tr>
<tr>
<td>Brain</td>
<td>7.5</td>
</tr>
<tr>
<td>Placenta</td>
<td>4.4</td>
</tr>
<tr>
<td>Lung</td>
<td>2.4</td>
</tr>
<tr>
<td>Liver</td>
<td>1.35</td>
</tr>
<tr>
<td>S. muscle</td>
<td>0.24</td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
</tr>
<tr>
<td>Pancreas</td>
<td></td>
</tr>
</tbody>
</table>

b, β-actin
Dynamic O-glycosylation of nuclear and cytosolic proteins: Cloning and characterization of a neutral, cytosolic β-N-acetylglucosaminidase from human brain
Yuan Gao, Lance Wells, Frank I. Comer, Glendon J. Parker and Gerald W. Hart

Access the most updated version of this article at doi: 10.1074/jbc.M010420200

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 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2001/01/08/jbc.M010420200.citation.full.html#ref-list-1