Molecular Cloning and Functional Expression of Two Splice Forms of Human N-Acetylglucosamine-1-phosphodiester α-N-Acetylglucosaminidase*

(Received for publication, June 22, 1999, and in revised form, September 10, 1999)

Rosalind Kornfeld‡§, Ming Bao¶, Kevin Brewer¶, Carolyn Noll¶, and William Canfield¶

From the ¶Department of Medicine, Washington University School of Medicine, St. Louis, Missouri 63110 and ¶¶W. K. Warren Medical Research Institute and the Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104

We have isolated and sequenced human cDNA and mouse genomic DNA clones encoding N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase (phosphodiester α-GlcNAcase) which catalyzes the second step in the synthesis of the mannose 6-phosphate recognition signal on lysosomal enzymes. The gene is organized into 10 exons. The protein sequence encoded by the clones shows 80% identity between human and mouse phosphodiester α-GlcNAcase and no homology to other known proteins. It predicts a type I membrane-spanning glycoprotein of 514 amino acids containing a 24-amino acid signal sequence, a luminal domain of 422 residues with six potential N-linked glycosylation sites, a single 27-residue transmembrane region, and a 41-residue cytoplasmic tail that contains both a tyrosine-based and an NPF internalization motif. Human brain expressed sequence tags lack a 102-base pair region present in human liver cDNA that corresponds to exon 8 in the genomic DNA and probably arises via alternative splicing. COS cells transfected with the human cDNA expressed 50–100-fold increases in phosphodiester α-GlcNAcase activity proving that the cDNA encodes the active enzyme. In contrast to glycosyltransferases and other glycoprotein-processing enzymes in the Golgi apparatus, which to date are all type II membrane-spanning proteins (7), the phosphodiester α-GlcNAcase is composed of identical type I membrane-spanning subunits.

EXPERIMENTAL PROCEDURES

Materials—The enzymes were obtained from the following suppliers: Taq DNA polymerase, EcoRI, and HindIII from Promega; BamHI from New England BioLabs; and T4 DNA ligase from Life Technologies, Inc. Transformation-competent DH5α cells were from Life Technologies, Inc. The protease inhibitor mixture (1000 ×) was prepared by combining antipain, chymostatin, leupeptin, and pepstatin A, all from Sigma, at 1 mg/ml each.

Preparation and Sequencing of Peptides from Pure Bovine Phosphodiester α-GlcNAcase—Bovine phosphodiester α-GlcNAcase was purified and subjected to amino-terminal sequencing by Edman degradation as described previously (6). In addition, to obtain internal peptide sequence, the pure enzyme was subjected to trypsin digestion in solution, and the digest was fractionated by reverse phase microbore high performance liquid chromatography (HPLC) on a C4 reversed-phase column (Eclipse, 7.8 × 150 mm, Waters) eluted with 0-40% acetonitrile in 0.1% formic acid at a flow rate of 0.3 ml/min. Fractions were collected and subsequently analyzed by MALDI-TOF mass spectrometry as described previously (6). The abbreviations used are: phosphodiester α-GlcNAcase, N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase; PAGE, polyacrylamide gel electrophoresis; EST, expressed sequence tag; PCR, polymerase chain reaction; TGN trans-Golgi network; bp, base pair; kb, kilobase pair.

* This work was supported in part by National Institutes of Health Grant CA08759 and American Heart Association, Oklahoma Affiliate, Grant 970805S. 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 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF187072 (human) and AF187073 (mouse).

§ To whom correspondence should be addressed: Division of Hematology, Washington University School of Medicine, 660 S. Euclid Ave., Box 8125, St. Louis, MO 63110. Tel.: 314-362-8835; Fax: 314-362-8826.

---

The biosynthesis of the mannose 6-phosphate recognition signal on the oligosaccharides of lysosomal acid hydrolases occurs in the Golgi apparatus and is catalyzed by the sequential action of two enzymes. The first step is the addition of N-acetylglucosamine-1-P to the C-6 hydroxyl group on selected mannose residues in the high mannose oligosaccharides of other known proteins. It predicts a type I membrane-spanning glycoprotein of 514 amino acids containing a 24-amino acid signal sequence, a luminal domain of 422 residues with six potential N-linked glycosylation sites, a single 27-residue transmembrane region, and a 41-residue cytoplasmic tail that contains both a tyrosine-based and an NPF internalization motif. Human brain expressed sequence tags lack a 102-base pair region present in human liver cDNA that corresponds to exon 8 in the genomic DNA and probably arises via alternative splicing. COS cells transfected with the human cDNA expressed 50–100-fold increases in phosphodiester α-GlcNAcase activity proving that the cDNA encodes the active enzyme. In contrast to glycosyltransferases and other glycoprotein-processing enzymes in the Golgi apparatus, which to date are all type II membrane-spanning proteins (7), the phosphodiester α-GlcNAcase is composed of identical type I membrane-spanning subunits.

EXPERIMENTAL PROCEDURES

Materials—The enzymes were obtained from the following suppliers: Taq DNA polymerase, EcoRI, and HindIII from Promega; BamHI from New England BioLabs; and T4 DNA ligase from Life Technologies, Inc. Transformation-competent DH5α cells were from Life Technologies, Inc. The protease inhibitor mixture (1000 ×) was prepared by combining antipain, chymostatin, leupeptin, and pepstatin A, all from Sigma, at 1 mg/ml each.

Preparation and Sequencing of Peptides from Pure Bovine Phosphodiester α-GlcNAcase—Bovine phosphodiester α-GlcNAcase was purified and subjected to amino-terminal sequencing by Edman degradation as described previously (6). In addition, to obtain internal peptide sequence, the pure enzyme was subjected to trypsin digestion in solution, and the digest was fractionated by reverse phase microbore high performance liquid chromatography (HPLC) on a C4 reversed-phase column (Eclipse, 7.8 × 150 mm, Waters) eluted with 0-40% acetonitrile in 0.1% formic acid at a flow rate of 0.3 ml/min. Fractions were collected and subsequently analyzed by MALDI-TOF mass spectrometry as described previously (6). The abbreviations used are: phosphodiester α-GlcNAcase, N-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase; PAGE, polyacrylamide gel electrophoresis; EST, expressed sequence tag; PCR, polymerase chain reaction; TGN trans-Golgi network; bp, base pair; kb, kilobase pair.

---

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF187072 (human) and AF187073 (mouse).

§ To whom correspondence should be addressed: Division of Hematology, Washington University School of Medicine, 660 S. Euclid Ave., Box 8125, St. Louis, MO 63110. Tel.: 314-362-8835; Fax: 314-362-8826.
Cloning and Expression of Human Phosphodiester α-GlCNacase

Isolation of cDNAs Encoding Human Phosphodiester α-GlCNacase—
The amino acid sequences of the three bovine phosphodiester α-GlCNacase peptides were used to BLAST search the non-redundant GenBank EST data base (8), and four clones were found that had sequence homologous to tryptic peptide 3 (see Table 1) at their 5’ ends. These EST clones from the I.M.A.G.E. Consortium (9) were all obtained from ATCC and included three human infant brain clones: I.M.A.G.E. Consortium clone ID 43150 (GB R60451, ATCC 367524), I.M.A.G.E. Consortium clone ID 28009 (GB R13596, ATCC 354452), and I.M.A.G.E. Consortium clone ID 23953 (GB T77682, ATCC 329860) as well as one mouse embryo clone, I.M.A.G.E. Consortium clone ID 354478 (GB W45789, ATCC 811592). Sequencing of the four cDNAs using the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction mixture (Perkin-Elmer) showed that all three human clones were identical except for small variations in length at the 3’ and 5’ ends. The sequence of the mouse embryo clone was homologous to the human infant brain clones except that it contained a 102-base pair insert not present in the latter clones.

The plasmid DNA of the human infant brain clone (I.M.A.G.E. ID 43150) which had an insert of about 1.4 kb was digested with HindIII to liberate a 700-bp fragment from its extreme 5’ end. The DNA fragment was gel-purified, and 50 ng were labeled with 100 μCi of [α-32P]dCTP using the High Prime DNA labeling kit (Roche Biochemicals) according to the manufacturer’s instructions. The labeled probe was used to screen a human liver 5’-Stratch plus cDNA library (CLONTECH HL 50220) in the λ TriplEx-phenagmid vector. Eight strongly positive colonies were detected in the primary screen of 108 plaque-forming units carried out as described by the manufacturer. In the secondary screen of these 8, strongly positive colonies were seen on 7 plates, and a number of separate colonies from each positive plate were cloned. The λ clones were converted to pTriplEx plasmids making use of the Cre-Lox recombinase feature of the vector in the host cells Escherichia coli BM 258, supplied by CLONTECH. The plasmid DNAs were subjected to restriction digestion to excise the cloned inserts. The inserts that also contained sequences of the EST varied in size from 1000 to 2200 bp. Clone 6.5, an example of the largest, was taken for the rest of the analysis.

Northern Blot—The BamHI-HindIII fragment of clone 6.5, gel-purified and labeled with [α-32P]dCTP as above, was used to probe a human multiple tissue Northern blot (CLONTECH) using the Express Hyb and protocol supplied by the manufacturer.

Isolation of Mouse Phosphodiester α-GlCNacase Genomic DNA—A mouse genomic PI clone that was not included in the human EST data base was used to generate a 5’-terminal fragment of clone 6.5 with a start codon. The 5’ sense primer contained 14 bp of un unmatched sequence encoding an EcoRI site, a Kozak consensus sequence (10), and the Met codon, followed by the first 17 bp of the clone 6.5 sequence, and was 5’-GGATTCCACATGGGACCTCCAGGGTGTCG-3’. The 3’-anti-sense primer corresponded to nucleotides 520–538 downstream of a BamHI site at nucleotide 512 (in the full-length sequence shown in Fig. 1) and was 5’-TGCAGGTCCGCTCCGG-3’. The PCR was performed with Taq polymerase in PCR buffer containing 1.3 mM betaine and 1% MeSO using clone 6.5 DNA as template for 25 cycles at an annealing temperature of 45 °C. The PCR product of approximately 500 bp was gel-purified using the QIAquick Gel Extraction Kit (Qiagen, Inc.) and digested with EcoRI and BamHI to generate a fragment, also gel-purified, for ligation into the expression vector pcDNA3.1(−) obtained from Invitrogen. The vector was also digested with EcoRI and BamHI to provide an acceptor site in its multiple cloning site, and after gel purification it was treated with calf intestine alkaline phosphatase (New England Biolabs) and subjected to gel purification. The ligation was performed with T4 DNA ligase, and the reaction mixture was used to transform DH5α competent cells that had been propagated in Luria-Bertani (LB) top broth ultracentrifuged. A number of clones were found that also contained sequences of the EST varied in size from 1000 to 2200 bp as described previously (12), except the concentration of [3H]GlcNAc-phosphomannose-α-methyl was reduced to 0.5 mM. One unit is defined as cleavage of 1 nmol of substrate per h at 37 °C.

Protein Determination—Protein concentration was measured using the Micro BCA assay (Pierce) standardized with bovine serum albumin (13).

Protein expression—All the fractions were assayed for phosphodiester α-GlCNac activity and protein content.

Phosphodiester α-GlCNacase Assay—Enzyme assays and synthesis of [3H]GlcNAc-phosphomannose-α-methyl were performed as described previously (12), except the concentration of [3H]GlcNAc-phosphomannose-α-methyl was reduced to 0.5 mM. One unit is defined as cleavage of 1 nmol of substrate per h at 37 °C.
RESULTS

Cloning of the cDNA Encoding Human Phosphodiester α-GlcNAcase—Affinity purified, homogeneous bovine phosphodiester α-GlcNAcase was subjected to amino-terminal amino acid sequencing as described previously (6). The pure enzyme was also subjected to trypsin digestion and high pressure liquid chromatography to generate two internal tryptic peptides which were sequenced. The amino acid sequences of these three peptides are shown in Table I. The protein, nucleotide, and EST data bases were searched for sequences that matched these peptide sequences, and several human and mouse ESTs were found that had the sequence of peptide 3 at their amino termini. Three human infant brain EST clones and one mouse embryo clone were obtained from ATCC and sequenced by us. The three human clones were all identical except for total length at their 3’ ends and virtually identical to the mouse clone, except that the mouse EST contained a 102-bp region that was absent from all three human brain ESTs. An EcoRI-HindIII fragment of about 700 bp was excised from the human cDNA clone (ATCC 367524) and used to probe a human liver cDNA library directionally cloned in λ TriplEx vector (CLONTECH). Of the positive clones isolated from the library and converted to plasmids (pTriplEx), the largest (2200 bp) was represented by clone 6.5 which was used for the rest of the analysis. The nucleotide sequence of clone 6.5 cDNA encoding phosphodiester α-GlcNAcase and the corresponding deduced amino acid sequence are shown in Fig. 1 which also shows, in italics, the 5’-untranslated sequence and Met encoding ATG derived from the recently deposited human genomic DNA sequence. The cDNA clone has been completely sequenced on both strands and is a novel sequence that predicts a mature protein of about 50 kDa which is in agreement with the size of the deglycosylated mature bovine liver phosphodiester α-GlcNAcase. There is a unique BamHI site at base 512 and a unique HindIII site at base 1581. The schematic diagram of the amino acid sequence and the Kyte-Doolittle hydrophilicity plot shown in Fig. 2 highlight some of the features of the phosphodiester α-GlcNAcase structure. All three bovine peptide sequences (1, 2, and 3) were found. Although the sequences of peptides 2 and 3 in the human are 100% identical to the bovine sequences, the amino-terminal peptide in human is only 67% identical to the bovine sequence. The human liver clone contains the 102-base pair insert that has the characteristics of an alternatively spliced segment that was missing in the human brain EST. The hydrophilicity plot indicates the presence of a hydrophobic membrane-spanning region from amino acids 448 to 474 and another hydrophobic region from amino acid 8 to 26 which fits the motif for a signal sequence, and there is a likely signal sequence cleavage site between Gly-25 and Leu-26 (14). There are six Asn-X-Ser/Thr potential N-linked glycosylation sites, one of which is within the 102-bp insert. All of these sites are amino-terminal of the putative trans-membrane region. These features indicate that the phosphodiester α-GlcNAcase is a type I membrane-spanning glycoprotein with the amino terminus in the lumen of the Golgi and the carboxyl terminus in the cytosol. This orientation is different from that of other glycosyltransferases and glycosidases involved in glycoprotein processing, which to date have been shown to be type II membrane-spanning proteins (7). Interestingly, there is a potential

---

TABLE I

| Peptide     | Sequence                                      |
|-------------|-----------------------------------------------|
| 1 Amino-terminal | DCTRVAHGRLHESWP                                 |
| 2 Tryptic  | PAAQTAGAHRPSVRTV                               |
| 3 Tryptic  | RDGTVTVSLEEVLTDEN                              |
| 4 N-terminal | GINLWMAEFLK                                    |

Cloning and Expression of Human Phosphodiester α-GlcNAcase

Fig. 1. Nucleotide and deduced amino acid sequence of human liver phosphodiester α-GlcNAcase. The italicized 5’-untranslated sequence and Met encoding ATG were derived from the recently deposited human genomic DNA sequence. The rest of the sequence, shown in regular type, is from clone 6.5. The signal sequence is indicated by dotted underline and the TM domain by solid underline. The potential N-linked glycosylation sites are double underscore. The potential internalization signals YHPL and NPFKD in the cytosolic domain are noted with --- and dashed underline, respectively. The possible polyadenylation signals AATAAA have a bold underline.
tyrosine-based internalization signal (Y<sup>488</sup>HPL<sup>491</sup>) in the cytoplasmic tail of the phosphodiester α-GlcNAcase sequence that suggests that the enzyme may travel to the TGN (where it becomes sialylated) or even to the plasma membrane before being retrieved to its site of action in the cis/medial-Golgi. A second potential retrieval signal in the phosphodiester α-GlcNAcase sequence is the carboxyl-terminal NPFKD. In yeast the sequence NPFKD has been shown to act as an endocytosis signal (15), and more recently (16–20) peptides and proteins containing the NPF motif have been shown to interact with the Eps15 homology domain (EH domain). The cDNA of clone 6.5 shown is missing the initiation codon and 5'-untranslated region.

**Size of the Human mRNA and Organization of the Mouse and Human Genes for Phosphodiester α-GlcNAcase**—The human cDNA clone 6.5 is about 2.2 kb in size, and a fragment of it was labeled with <sup>32</sup>P and used to probe a human multiple tissue Northern blot (CLONTECH) as shown in Fig. 3. An mRNA of approximately 2.4 kb was detected for all tissues except brain where the band was somewhat smaller (2.3 kb), consistent with the fact that the ESTs isolated from human brain are missing a 102-bp segment present in the human liver cDNA of clone 6.5. In addition, another fainter mRNA band at about 3.5 kb is present in the liver.

A mouse genomic Pl clone was obtained that was positive when probed with <sup>32</sup>P-labeled DNA fragments of the human clone 6.5 cDNA. Sequencing of this Pl clone revealed the intron/exon organization of most of the mouse gene, which is presented schematically in Fig. 4. The intron/exon borders were also presented in Table II. The Pl clone was 9.8 kb in length and contained a large intron at the 5’ end followed by what is designated as exon 2 which encodes the mouse equivalent of amino acids 30–179 in the human phosphodiester α-GlcNAcase. Thus this genomic Pl clone is missing at least one 5’ exon that encodes the first 29 amino acids or more. The 102-base pair insert that is missing from the human brain ESTs exactly comprises exon 8, indicating that those mRNAs are the result of alternative splicing. Exon 10 encodes the trans-membrane domain and the cytoplasmic tail through the stop codon, followed by a 3’-untranslated region of at least 535 bp which occurs at the 3’ end of the Pl clone. While this manuscript was in preparation, a routine search of the non-redundant data base with the human cDNA sequence revealed that a recently deposited 160 kb of human genomic sequence (AC007011 from Los Alamos Laboratory) contained a 10-kb sequence, and ATG to encode methionine in frame just before the first nucleotide of clone 6.5 and inserted the construct into the expression vector pcDNA3.1(−) (Invitrogen) as described under “Experimental Procedures.” Given the sequence we know from the human genomic DNA, it is now evident that our construct, in fact, encodes the entire phosphodiester α-GlcNAcase sequence for expression. A deletion mutant of clone 6.5 missing the 102-bp insert (Δ102-bp variant) was also constructed and put into pcDNA3.1(−). When these expression plasmids were transfected into COS cells and the cells were harvested 48 h after replacing the transfecting DNA with full medium, the results shown in Table III were obtained. Dupli-cates plates of cells expressing the full-length and the Δ102-bp variant as well as mock-transfected cells were lysed by sonication and subjected to high speed centrifugation to sediment a membrane pellet that was separated from the supernatant cytosol. The membranes were solubilized in detergent, and both fractions were assayed for phosphodiester α-GlcNAcase activity. Between 70 and 75% of the enzyme activity in the transfected cells was in the membrane pellet, and 78% of the endogenous COS cell phosphodiester α-GlcNAcase was in the membrane pellet. Table III shows that the solubilized mem-

![Fig. 2. Domain structure and hydrophilicity plot of the cloned coding region of human phosphodiester α-GlcNAcase from the cDNA of clone 6.5. The upper diagram shows the signal peptide (β) at the amino-terminal region and the TM region (β) at the carboxyl-terminal region. The sequences corresponding to the bovine peptides in Table I are shown (−); the 162-bp insert is indicated (β) as are the six potential N-linked glycosylation sites (β). The * indicates the location of the potential tyrosine based internalization signal in the cytosolic domain. The lower diagram is the Kyte-Doolittle hydrophilicity plot. a.a., amino acid.](image)

![Fig. 3. Human multiple tissue Northern blot probed with human cDNA of phosphodiester α-GlcNAcase. mRNA from H, heart; B, brain; Pl, placenta; Lu, lung; Li, liver; Sk, skeletal muscle; K, kidney; and Pa, pancreas were present on the blot from CLONTECH.](image)
branes of COS cells have a level of endogenous phosphodiester α-GlcNAcase activity (21.5 nmol/h/mg protein) that is 4 times higher than that of bovine liver membranes (5.2 nmol/h/mg) (6). However, transfection with the full-length human liver cDNA resulted in an average increase of 50-fold in enzyme-specific activity, and transfection with the D102-bp variant cDNA caused an average increase of 90-fold in phosphodiester α-GlcNAcase activity. This result shows that the 102-bp region is dispensable for enzymatic activity.

An aliquot of each solubilized membrane extract was subjected to reducing SDS-PAGE, and the gel was blotted onto nitrocellulose that was probed with an antibody raised against a peptide in the amino terminus of the bovine liver phosphodiester α-GlcNAcase. The Western blot shown in Fig. 6 reveals several things about the human liver enzyme expressed in COS cells as follows: 1) the full-length protein has a molecular weight of about 77,000 which is consistent with a polypeptide of 490 amino acids (after cleavage of the signal peptide) bearing 6 N-linked oligosaccharides and is similar in size to the bovine liver enzyme on SDS-PAGE (68–72 kDa); 2) the Δ102-bp variant has a molecular weight of about 69,000 (about 8000 smaller than the full-length) consistent with its missing 102 bp or about 3700 molecular weight of peptide and a single N-linked oligosaccharide; and 3) both human proteins cross-react with the antipeptide antibody despite the fact that the peptide to which it was raised (amino acids 3–15 of the bovine amino-terminal peptide in Table I) differs in three of the 13 amino acids from the human sequence. Endogenous COS cell phosphodiester α-GlcNAcase, in contrast, does not cross-react with the antibody. In an effort to evaluate whether the Δ102-bp variant protein really had a higher intrinsic enzymatic activity than the full-length protein (as opposed to being expressed at higher copy number per cell), we quantitated the amount of antigen protein in the blots in Fig. 6 using laser densitometry to integrate the volume of each band on the blot. These values appear in the third column in Table III and show that the Δ102-bp variant protein has, on average, about one-half the phosphodiester α-GlcNAcase activity per antigen unit as the full-length protein. Both membrane extracts were analyzed for their Km values for the artificial substrate [3H]GlcNAc-phosphomannose and gave values (full-length, 0.4 mM and Δ102-bp variant, 0.43 mM) comparable with that of the pure bovine liver enzyme (0.49 mM) (6). The phosphodiester α-GlcNAcase activity in both membrane extracts showed a broad pH optimum between pH 5 and pH 7, again comparable to the bovine enzyme. The full-length protein eluted from a Superose 6 gel filtration column in the same position as the purified bovine liver enzyme, indicating that the expressed human enzyme is a homotetramer (data not shown).

A truncated form of the modified human cDNA was con-

![Fig. 4. Genomic organization of phosphodiester α-GlcNAcase. Upper, schematic presentation of the intron/exon organization of the human genomic DNA for phosphodiester α-GlcNAcase. Lower, schematic presentation of the intron/exon organization of the mouse PI genomic DNA clone. UTR, untranslated region.](image)

**Table II**

| Exon sequences are shown in uppercase letters and introns in lowercase letters. The size of introns and exons are given in kilobases or base pairs. |
|---|
| **Donor** | **Acceptor** | **Intron size** | **Exon size** | **Numbers** |
|---|---|---|---|---|
| **Human** | | | | |
| TCGGG gt gagt. . . .cccga ag GCCCT | 0.10 | 456 | 1 |
| ACCGG gt gagg. . . .gttcc ag GTACC | 1.49 | 140 | 2 |
| GACAG gt tgtat. . .ccttc ag GTTCC | 1.25 | 109 | 3 |
| CGTGG gt gatctc. . .gcgcg ag CACCA | 1.38 | 129 | 4 |
| CACTG gt aagca. . .ctcac ag CAGGG | 0.68 | 206 | 5 |
| GAGAGA gt gagt. . .cccgc ag CGGCG | 0.09 | 48 | 6 |
| TGAAG gt aggag. . .ticaat ag AGTGT | 0.47 | 102 | 7 |
| CAGAG gt accta. . .gttac ag TAAAG | 0.08 | 64 | 8 |
| ACCAG gt aggtg. . .ctgccg ag GACCCG | 1.45 | 840 | 9 |
| **Mouse** | | | | |
| ACCGG gt gagg. . . .cttccg ag GTTCC | 1.83 | 140 | 10 |
| GACAG gt caag. . . .ccttcg ag GTTCT | 1.32 | 109 | 11 |
| CGTGG gt gatctc. . .tggccg ag CACCA | 0.20 | 129 | 12 |
| CACTG gt aagaa. . .cttttc ag CACCA | 0.68 | 206 | 13 |
| AGAGA gt gagt. . . .cttcag gc CACCA | 0.11 | 48 | 14 |
| TGAAG gt aggag. . . .ctcagc gc AGTGT | 0.75 | 102 | 15 |
| CCAGG gt atgcgg. . . .ctgccg gc TGAAG | 0.09 | 64 | 16 |
| ACCAG gt aagtg. . . .atgcgc gc GACCA | 1.48 | 736 | 17 |

208^**et al.**
cells transfected with full-length phosphodiester α-GlcNAcase cDNA, the truncated phosphodiester α-GlcNAcase was found secreted in the medium. In the experiment shown in Table IV, the COS cells were incubated in serum-free medium for 24 h following removal of the transfecting cDNA, and the medium and cells were harvested for enzyme assay. The endogenous COS cell phosphodiester α-GlcNAcase activity is predominantly (92%) found in the cell membrane fraction as is that encoded by the full-length human cDNA (67%), but the truncated human phosphodiester α-GlcNAcase was predominantly secreted into the medium (75%).

**DISCUSSION**

In these studies we have isolated the cDNA that encodes human phosphodiester α-GlcNAcase, which catalyzes the second step in the formation of the mannose 6-phosphate recognition signal on lysosomal enzymes. Although our cDNA clone is missing the initiator methionine and 5′-untranslated region, the human phosphodiester α-GlcNAcase genomic sequence, recently deposited in the database as part of a large cosmid sequence, has provided the missing sequence. This human genomic DNA was derived from human chromosome 16, and interestingly the gene for the α-subunit of GlcNAc phosphotransferase also is on chromosome 16. In parallel we have sequenced a P1 clone of mouse genomic DNA that encodes most of the sequence of mouse phosphodiester α-GlcNAcase. The genomic organization of the mouse and human genes is the same, with both containing a single exon (exon 8) of 102 bp that is missing in human infant brain ESTs. This splice variant may be brain-specific since the Northern blot of mRNA from various human tissues revealed that brain mRNA for phosphodiester α-GlcNAcase is about 2.3 kb in size in contrast to the 2.4-kb mRNA of other tissues. The amino acid sequences encoded by the human and mouse phosphodiester α-GlcNAcase cDNAs are 80% identical.

The nucleotide sequence of the coding portion of the human genomic DNA for phosphodiester α-GlcNAcase is identical to that of the human liver clone 6.5 cDNA except for a single base pair (bp 1394) in the full-length sequence, which is T in clone 6.5 and C in the genomic. This changes the Ile residue at amino acid 465 in clone 6.5 to a Thr residue in the genomic sequence. Interestingly, the mouse sequence encodes the Thr residue as does a human heart EST cDNA (clone 3NHCO336), whereas the human infant brain ESTs encode the Ile residue. These results suggest that the occurrence of T or C at base pair 1394 may represent a polymorphism and not sequencing errors.

Phosphodiester α-GlcNAcase acts in the Golgi (21) and the bovine enzyme is a membrane-spanning tetramer composed of two disulfide linked dimers containing 68-kDa N-glycosylated monomers (6). The structural features revealed by the amino acid sequence show that the human phosphodiester α-GlcNAcase contains a hydrophobic signal sequence and signal peptide cleavage site at the NH2 terminus and another hydrophobic transmembrane region near the COOH terminus.

---

**Table IV**

| Sample       | Medium | Activity of cells |
|--------------|--------|-------------------|
|              |        | Soluble | Membrane |
|              |        | nmol/h (%) | total      |
| Mock-transfected | 0.25 (4) | 44 | 10.4 |
| Full-length   | 36 (12) | 64 (21) | 197 (67) |
| W450Stop     | 128 (75) | 12 (7) | 29 (18) |

*Average of values for duplicate samples.

---

**FIG. 5.** Comparison of human and mouse phosphodiester α-GlcNAcase sequences. The amino acids identical in both sequences are boxed.

**TABLE III**

Expression of human cDNA in COS cells induces phosphodiester α-GlcNAcase activity

| Sample       | Enzyme specific activity | Fold increase in specific activity | Enzyme specific activity |
|--------------|--------------------------|-----------------------------------|--------------------------|
|              | nmol/h/mg                |                                   | nmol/h/antigen unit*     |
| Full-length  | 932                      | 44                                | 10.4                     |
| Full-length  | 1251                     | 58                                | 15.7                     |
| Δ102-bp variant | 1752                    | 82                                | 6.0                      |
| Δ102-bp variant | 2124                    | 99                                | 6.9                      |
| Mock-transfected | 20                      | 1                                 | 1                       |
| Mock-transfected | 23                      | 1                                 | 1                       |

*Antigen unit is an arbitrary number derived from densitometry of the bands on the Western blot in Fig. 6.

---

**FIG. 6.** Western blot of membrane extracts from COS cells expressing phosphodiester α-GlcNAcase. Aliquots of the extracts in the experiment shown in Table III were subjected to SDS-PAGE, blotted to nitrocellulose, and probed with anti-peptide antibody to the bovine enzyme.

---

2 W. Canfield, unpublished observations.
cates that the enzyme is a type I membrane-spanning protein with the NH₂ terminus in the lumen of the Golgi and the COOH terminus in the cytosol. This orientation is opposite to that of other cloned glycosyltransferases and glycosidases of the oligosaccharide processing pathway which are type II membrane-spanning proteins (7). There are six potential N-linked oligosaccharide sites in the luminal domain and, as revealed by a ProfileScan, a carboxyl-terminal cystine knot profile (22) (amino acids 307–390) including an epidermal growth factor-like domain (amino acids 362–389) occurs just prior to the 102-bp insert which also encodes a very cysteine-rich domain. These features indicate that the disulfide-bonded dimers of phosphodiesterase α-GlcNAcase are probably stabilized by a number of S–S bridges. This conclusion is also supported by the fact that on reducing SDS-PAGE the purified bovine liver enzyme showed some dimer band even after boiling for 5 min in 5% β-mercaptoethanol/SDS sample buffer (6).

It is very interesting that the carboxyl-terminal cytosolic tail contains a potential tyrosine-based internalization signal YHPL that fits the consensus YXXΦ (Φ is a bulky hydrophobic amino acid) first described by Canfield et al. (23) for the mannose 6-phosphate receptor and subsequently found for a number of other receptors that undergo endocytosis in coated pits from the plasma membrane prior to entry into the endosomal compartment (24). The trans-Golgi membrane marker protein TGN 38 (rodent) (25) and human TGN 46 (26) also contain such a tyrosine-based signal (YQRL) which is essential for their retrieval from the plasma membrane to the TGN. This raises the intriguing possibility that phosphodiesterase α-GlcNAcase travels to the plasma membrane and is retrieved to the Golgi apparatus. Such a traffic pattern is unusual for a Golgi-processing enzyme, but we already have evidence that the bovine enzyme is sialylated (6), a modification believed to occur solely in the TGN. The human phosphodiesterase α-GlcNAcase contains another potential endocytosis signal, the NPFKD sequence at its COOH terminus, which may also play a role in its intracellular trafficking. Tan et al. (15) have shown that the yeast type I integral membrane protein Kex2p, which resides in a late Golgi compartment, contains the endocytosis signal NPFXD in its cytoplasmic tail. When fused to a truncated form of the α-factor receptor Ste2p, the cytoplasmic tail of Kex2p mediated α-factor endocytosis that was dependent on the sequence NPFXD as demonstrated by alanine-scanning mutagenesis of the sequence. The endocytosis motif was active in both its normal internal location as well as at the COOH terminus of the cytoplasmic tail. Subsequently, Salcini et al. (16) showed that the EH (Eps15 homology) domain involved in protein-protein interactions binds in vitro to peptides containing an NPF motif. They also isolated a number of proteins that interacted with EH domains and found that all contained NPF motifs responsible for the binding. The direct interaction of the NPF motif with a binding pocket in the EH₂ domain of Eps15 was shown by de Beer et al. (17) who solved the three-dimensional structure of the EH₂ domain using heteronuclear magnetic resonance spectroscopy. Others have examined both the interaction of specific NPF-containing proteins with EH domains (18) and the peptide recognition specificity of EH domains from a variety of proteins (19). Most recently Yamahai et al. (20) have isolated a new adaptor protein they named intersectin because it contains two EH domains and five SH3 domains and thus can potentially bring together EH and SH3 domain-binding proteins in a macromolecular complex that is part of the endocytic machinery. Both EH domains of intersectin were shown to interact with NPF-containing peptides as well as the mouse RAB protein which contains four NPF motifs including a COOH-terminal NPF motif. By using glutathione S-

transferease fusion proteins of RAB and RAB without the COOH-terminal NPF motif, they showed that only the former fusion protein interacted with the intersectin EH domains, indicating that in this case only the NPF motif at the COOH terminus can bind the EH domain. Furthermore, GST-TNPF and GST-TNPF-LA could bind the EH domain but GST-TNPF-LAA could not, further emphasizing the importance of the carboxylate group in the interaction. The COOH-terminal NPFKD sequence, conserved between human and mouse phosphodiesterase α-GlcNAcase, is therefore an attractive candidate for either an endocytosis signal acting at the plasma membrane or a retrieval signal acting at the TGN to return the enzyme to the cis/medial-Golgi. One could imagine it interacting with EH-containing cystolic proteins at either site which could initiate formation of intracellular trafficking vesicles.

When we inserted the complete coding sequence of human phosphodiesterase α-GlcNAcase in an expression vector (pcDNA 3.1(−)h-Po-G) and expressed this construct in COS cells, the membrane extracts from the cells expressed over 50 times the endogenous level of enzyme activity. The plasmid DNA encoding the full-length protein produced an enzyme that had the appropriate mobility on SDS-PAGE (77 kDa) for a 490-amino acid mature protein with 6 N-linked oligosaccharides. This was determined by Western blotting with an antibody raised to a 13-amino acid sequence from the amino-terminal peptide 1 of the purified bovine phosphodiesterase α-GlcNAcase as described previously (6). It should be pointed out that the sequence of bovine peptide 1 starts 24 amino acids downstream of the signal peptide cleavage site (following an arginine residue) in the human enzyme sequence. Thus the mature bovine liver enzyme as isolated had undergone an additional proteolytic clip that may account for the fact that on SDS-PAGE it had a mobility corresponding to 68–72 kDa. We also expressed two mutant constructs, one missing the 102-bp insert (Δ102-bp variant) and the other missing the transmembrane and cytosolic tail domains. In both cases, good expression and high enzyme activity were obtained, and the distribution of the activity indicated the following: 1) that the 102-bp insert is not required for activity or retention in the cell membrane, and 2) that the transmembrane and cytosolic tail are not required for activity but are required for retention in the membrane since that truncated enzyme was recovered in the medium. A number of membrane-spanning enzymes of the Golgi require their transmembrane and cytosolic domains. In both cases, good expression and high enzyme activity were obtained, and the distribution of the activity indicated the following: 1) that the 102-bp insert is not required for activity or retention in the cell membrane, and 2) that the transmembrane and cytosolic tail are not required for activity but are required for retention in the membrane since that truncated enzyme was recovered in the medium. A number of membrane-spanning enzymes of the Golgi require their transmembrane domains for retention, but there are some in which the luminal stem region plays a role in Golgi retention (see Ref. 7). Our finding with the truncated expressed human phosphodiesterase α-GlcNAcase was not surprising since Lee and Pierce (27) have described a soluble form of the enzyme in human serum.

Our future studies will focus on analyzing the role of the various domains of expressed human phosphodiesterase α-GlcNAcase on the intracellular trafficking of the enzyme. We will also explore what role the alternatively spliced 102-bp exon 8 plays in the phosphodiesterase α-GlcNAcase structure and function.

Acknowledgments—R. Kornfeld and C. Noll especially thank Monita Wilson and Marina Kisseleva for their expert advice and assistance in library screening and cDNA cloning. W. Canfield gratefully thanks H. Pan and B. Roe for their help in sequencing the mouse genomic DNA.

REFERENCES
1. Kornfeld, S. (1987) FASEB J. 1, 462–468
2. Kornfeld, S., and Mellman, I. (1989) Annu. Rev. Cell Biol. 5, 483–525
3. Dahms, N. M., Lobel, P., and Kornfeld, S. (1989) J. Biol. Chem. 264, 12115–12118
4. Mullis, K. G., Huynh, M., and Kornfeld, R. H. (1994) J. Biol. Chem. 269, 1718–1726
5. Mullis, K. G., and Kornfeld, R. H. (1994) J. Biol. Chem. 269, 1727–1732
6. Kornfeld, R., Bao, M., Brewer, K., Noll, C., and Canfield, W. M. (1998) J. Biol. Chem. 273, 23203–23210
7. Colley, R. J. (1997) Glycobiology 7, 1–13
Cloning and Expression of Human Phosphodiesterase α-GlcNAcase

8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
9. Lennon, G. G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) Genomics 33, 151–152
10. Kozak, M. (1986) Cell 44, 283–292
11. Horton, R. M. (1977) Methods Mol. Biol. 67, 141–149
12. Mullis, K. G., and Ketchum, C. M. (1992) Anal. Biochem. 205, 200–207
13. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Anal. Biochem. 150, 76–85
14. Nielsen, H., Engelbrecht, J., Brunck, S., and von Heijne, G. (1997) Protein Eng. 10, 1–6
15. Tan, P. K., Howard, J. P., and Payne, G. S. (1996) J. Cell Biol. 135, 1789–1800
16. Salcini, A. E., Confalonieri, S., Deria, M., Santolini, E., Tassi, E., Minenkova, O., Cesareni, G., Pelicci, P. G., and Di Fiore, P. P. (1997) Genes Dev. 11, 2239–2249
17. de Beer, T., Carter, R. E., Lobel-Rice, K. E., Sorkin, A., and Overduin, M. (1998) Science 281, 1357–1360
18. McPherson, P. S., de Heuvel, E., Phillipie, J., Wang, W., Sengar, A., and Egan, S. (1998) Biochem. Biophys. Res. Commun. 244, 701–705
19. Paoluzi, S., Castagnoli, L., Laura, I., Salcini, A. E., Coda, L., Free S., Confalonieri, S., Pelicci, P. G., Di Fiore, P. P., and Cesareni, G. (1998) EMBO J. 17, 6541–6550
20. Yamabhai, M., Hoffman, N. G., Hardisson, N. L., McPherson, P. S., Castagnoli, L., Cesareni, G., and Kay, E. R. (1998) J. Biol. Chem. 273, 31401–31407
21. Varki, A., and Kornfeld, S. (1981) J. Biol. Chem. 256, 9937–9943
22. Bork, P. (1995) FEBS Lett. 327, 125–130
23. Canfield, W. M., Johnson, K. F., Ye, R. D., Gregory, W., and Kornfeld, S. (1991) J. Biol. Chem. 266, 5682–5688
24. Trowbridge, I. S., Collawn, J. F., and Hopkins, C. R. (1993) Annu. Rev. Cell Biol. 9, 129–161
25. Humphrey, J. S., Peters, P. J., Yuan, L. C., and Bonifacino, J. S. (1993) J. Cell Biol. 120, 1123–1135
26. Kain, R., Angata, K., Kerjaschki, D., and Fukuda, M. (1998) J. Biol. Chem. 273, 981–988
27. Lee, J. K., and Pierce, M. (1995) Arch. Biochem. Biophys. 319, 413–425