The α and β-subunits of the Human UDP-N-acetylglucosamine: Lysosomal Enzyme Phosphotransferase Are Encoded by a Single cDNA

Received for publication, August 16, 2005, and in revised form, August 24, 2005. Published, JBC Papers in Press, August 24, 2005, DOI 10.1074/jbc.M509008200

Mariko Kudo1, Ming Bao1, Anil D’Souza2, Fu Ying2, Huaqin Pan1, Bruce A. Roe2, and William M. Canfield1

From the 1Genzyme Corp., Oklahoma 73104, 2W. K. Warren Medical Research Institute and the Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, 3System Biology, Berlex Biosciences, Richmond, California 94804-0099, 4Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104, and 5Department of Chemistry and Biochemistry, The University of Oklahoma, Norman, Oklahoma 73019, and 6Center for Integrated Fungal Research, North Carolina State University, Raleigh, North Carolina 27606

Lysosomal enzymes are targeted to the lysosome through binding to mannose 6-phosphate receptors because their glycans are modified with mannose 6-phosphate. This modification is catalyzed by UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase). Bovine GlcNAc-phosphotransferase was isolated using monoclonal antibody affinity chromatography, and an αββγγ-subunit structure was proposed. Although cDNA encoding the γ-subunit has been described, cDNAs for the α- and β-subunits have not. Using partial amino acid sequences from the bovine α- and β-subunits, we have isolated a human cDNA that encodes both the α- and β-subunits. Both subunits contain a single predicted membrane-spanning domain. The α- and β-subunits appear to be generated by a proteolytic cleavage at the Lys298-Asp299 bond. Transfection of 293T cells with the α/β-subunits-precursor cDNA with or without the γ-subunit cDNA results in a 3.6- or 17-fold increase in GlcNAc-phosphotransferase activity in cell lysates, suggesting that the precursor cDNA contains the catalytic domain. The sequence lacks significant similarity with any described vertebrate enzyme except for two Notch-like repeats in the body affinity chromatography from lactating bovine mammary glands and found to contain three distinct polypeptides. Based on the presence of disulfide-linked αα and γγ dimers and quantitative protein sequencing, an αββγγ-subunit structure with a molecular mass of 540 kDa was proposed (6). cDNA and gene encoding the γ-subunit have been isolated and characterized (7), but the cDNA(s) or gene(s) for the α- and β-subunits has not been reported.

Lysosomal storage diseases are caused by a genetic deficiency of one or more lysosomal enzymes. Enzyme replacement therapy, the therapeutically favorable strategy for many lysosomal storage diseases (8, 9). The cation-independent M6P receptor is widely expressed on the surface of many cells and can be used for efficient receptor-mediated endocytosis of therapeutic enzymes (10–12). However, the enzyme must be modified with sufficient M6P to bind to the receptor with the appropriate affinity. Access to the cDNA(s) for the α- and β-subunits may allow reconstitution of phosphorylation in vitro as well as other strategies to improve the quality of lysosomal enzyme therapeutics.

GlcNAc-phosphotransferase is absent or unregulated in a group of diseases of lysosomal targeting termed mucolipidoses (13–17). GlcNAc-phosphotransferase activity is absent in mucolipidosis II (ML II; MIM 252500, l-cell disease), it is reduced in ML IIIA (MIM 252600; pseudo-Hurler polydystrophy), and in ML IIIC (MIM 252605, mucolipidosis III,McGready et al. 1997) it is normal. Lysosomal enzymes are targeted to the lysosome through binding to mannose 6-phosphate receptors because their glycans are modified with mannose 6-phosphate. This modification is catalyzed by UDP-N-acetylglucosamine:lysosomal enzyme N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase). Bovine GlcNAc-phosphotransferase was isolated using monoclonal antibody affinity chromatography, and an αββγγ-subunit structure was proposed. Although cDNA encoding the γ-subunit has been described, cDNAs for the α- and β-subunits have not. Using partial amino acid sequences from the bovine α- and β-subunits, we have isolated a human cDNA that encodes both the α- and β-subunits. Both subunits contain a single predicted membrane-spanning domain. The α- and β-subunits appear to be generated by a proteolytic cleavage at the Lys298-Asp299 bond. Transfection of 293T cells with the α/β-subunits-precursor cDNA with or without the γ-subunit cDNA results in a 3.6- or 17-fold increase in GlcNAc-phosphotransferase activity in cell lysates, suggesting that the precursor cDNA contains the catalytic domain. The sequence lacks significant similarity with any described vertebrate enzyme except for two Notch-like repeats in the body affinity chromatography from lactating bovine mammary glands and found to contain three distinct polypeptides. Based on the presence of disulfide-linked αα and γγ dimers and quantitative protein sequencing, an αββγγ-subunit structure with a molecular mass of 540 kDa was proposed (6). cDNA and gene encoding the γ-subunit have been isolated and characterized (7), but the cDNA(s) or gene(s) for the α- and β-subunits has not been reported.

Lysosomal storage diseases are caused by a genetic deficiency of one or more lysosomal enzymes. Enzyme replacement therapy, the therapeutically favorable strategy for many lysosomal storage diseases (8, 9). The cation-independent M6P receptor is widely expressed on the surface of many cells and can be used for efficient receptor-mediated endocytosis of therapeutic enzymes (10–12). However, the enzyme must be modified with sufficient M6P to bind to the receptor with the appropriate affinity. Access to the cDNA(s) for the α- and β-subunits may allow reconstitution of phosphorylation in vitro as well as other strategies to improve the quality of lysosomal enzyme therapeutics.

GlcNAc-phosphotransferase is absent or unregulated in a group of diseases of lysosomal targeting termed mucolipidoses (13–17). GlcNAc-phosphotransferase activity is absent in mucolipidosis II (ML II; MIM 252500, l-cell disease), it is reduced in ML IIIA (MIM 252600; pseudo-Hurler polydystrophy), and in ML IIIC (MIM 252605, mucolipidosis III,McGready et al. 1997) it is normal.
variant form) it is unregulated. Mutations in the \( \gamma \)-subunit gene have been demonstrated to be the molecular basis of MLIIC (7, 18). Because cells from ML IIIC patients retain GlcNAc-phosphotransferase activity, whereas the substrate recognition is unregulated, the catalytic domain of the enzyme is predicted to reside on \( \alpha \)- and/or \( \beta \)-subunits of the enzyme. Cloning of the human \( \alpha \)- and \( \beta \)-subunit gene will allow us to understand the genetic basis of ML II and ML IIIA.

In this paper we describe the cloning of a single cDNA and gene that encodes a precursor of the \( \alpha \)- and \( \beta \)-subunits of human GlcNAc-phosphotransferase (gene symbol, GNPTAB). Transfection of the \( \alpha \)/\( \beta \)-subunits precursor cDNA with or without co-transfection of the regulatory \( \gamma \)-subunit cDNA resulted in an increase in GlcNAc-phosphotransferase activity on low molecular weight substrate. This demonstrates that the \( \alpha \)/\( \beta \)-subunits precursor cDNA encodes the catalytic domain of the enzyme.

**EXPERIMENTAL PROCEDURES**

Oligonucleotides, cDNA Libraries, Clones, Plasmid Vectors—Oligonucleotide primers were synthesized at the Molecular Biology Resource Facility at University Oklahoma Health Sciences Center. Mouse liver Marathon Ready™ cDNA and human whole brain Marathon-Ready™ cDNA were purchased from Clontech (Palo Alto, CA). A human placental cDNA library in Agt10 (catalog 77399) was purchased from ATCC (Manassas, VA). SuperScript™ human brain cDNA library was purchased from Invitrogen. pCR2.1, pUC19, pcDNA 3.1(–), and pcDNA6/V5/His-A were purchased from Invitrogen.

Amino Terminal Sequencing of Protein and Peptides—Bovine GlcNAc-phosphotransferase was isolated as previously described (19) and subjected to non-reducing SDS-PAGE (20). The gel was either electroblotted onto a polyvinylidene fluoride membrane (21) or subjected to in-gel Lys-C digestion. The protein bands on the polyvinylidene fluoride membrane was stained for 1 min with 0.1% Coomassie Blue in 10% acetic acid, 50% methanol, and excised. Protein bands in the gel were stained, excised, and subjected to in-gel reduction, S-Carboxamidomethylation, and Lys-C digestion in 0.1 M Tris-HCl, pH 8.0, at 37 °C overnight. The peptides were resolved by reverse phase chromatography on a C18 column equilibrated with 0.1% trifluoroacetic acid and developed with a linear gradient in acetonitrile. Individual peaks were examined by matrix-assisted laser desorption/ionization-mass spectrometry to identify a peak containing a single mass. Strategies for peak selection, reverse phase selection, and Edman microsequencing has been described (22).

Proteins on the polyvinylidene fluoride membrane and Lys-C-derived peptides were subjected to automated Edman degradation on an Applied Biosystems, Inc. (Foster City, CA) model 492 protein sequencer at the Molecular Resource Facility in the University of Oklahoma Health Sciences Center.

3'-RACE for the Mouse \( \beta \)-subunit—A primer 5'-gacccctgagaaagggc-3' that spans 441–463 nucleotides upstream from the first amino acid of the mouse \( \beta \)-subunit in a mouse cDNA clone (GenBank™ accession number L36434) was used to amplify the 3'-end of the cDNA from mouse liver Marathon-Ready cDNA by nested 3'-RACE. Primers AP1 5'-catctaatgctggcagagc-3' and AP2 5'-acactataggctggcagagc-3' were used as reverse primers for the first and second PCR, respectively. The cycling parameters for the first PCR were 30 cycles of denaturation at 94 °C for 30 s and annealing/extension at 68 °C for 4 min. The second PCR used the same parameters for 20 cycles. A 3.2-kb product was gel-purified, subcloned into pCR2.1, and sequenced on both strands.

Screening of Human Placental cDNA Library and 3'-RACE for Human GlcNAc-phosphotransferase cDNA—A 1.1-kb fragment was amplified from the partial mouse \( \beta \)-subunit cDNA by PCR using a forward primer 5'-agtggggcaggctggcagagc-3' and a reverse primer 5'-aatctgctggcagagc-3' and used to screen a size-selected human placental cDNA library in Agt10 under reduced stringency (55 °C, 2× SSC (1× SSC = 0.15 M NaCl and 0.015 M sodium citrate)). Inserts from positive clones were subcloned into pCR2.1 using Agt10 LD-Insert Screening Amplimer Set (Clontech, Palo Alto, CA) and sequenced. The remaining portion of the cDNA was cloned by a combination of a walking strategy and RACE. The 3'-end of the cDNA was amplified from human whole brain Marathon-Ready™ cDNA by nested 3'-RACE, first using primers 5'-agtggggcaggctggcagagc-3' and AP1 5'-catctaatgctggcagagc-3', then using primers 5'-cccaaaagctctcggcagc-3' and AP2 5'-acactataggctggcagagc-3'. The cycling parameters for the first PCR were 30 cycles of denaturation at 94 °C for 30 s and annealing/extension at 68 °C for 4 min. The second PCR used the same parameters for 20 cycles. A 3.4-kb product was gel-purified, subcloned into pCR2.1, and sequenced on both strands.

Construction of the Full-length cDNA Encoding Human GlcNAc-phosphotransferase \( \alpha \)-\( \beta \)-subunits Precursor—A full-length human precursor cDNA was constructed by splicing three pieces of cDNA, the 1118-nucleotide fragment of I.M.A.G.E. Consortium clone 682681 (GenBank™ accession number AA204698), a 1703-nucleotide fragment from the placental clone, and a 2698-nucleotide fragment from the 3'-RACE product by ligation at two BbsI restriction sites and subcloning into pUC19. The full-length cDNA (5.6 kb) was excised and subcloned into the NotI site of pcDNA 3.1(–). Instability of a 5'-adenosine repeat (nucleotides 2925–2929) was identified as an occasional problem that results in deletion of 1 adenosine, introducing a frameshift. This problem was prevented by replacing a 215-bp nucleotide fragment between the MfII and DraIII sites with an MfII- and DraIII-cleaved PCR product. The product was amplified by using the cDNA as a template with a forward superprimer 5'-gaagcataagctgcaggttaaaggtaggtcgtg-3' and a reverse primer 5'-actgtcatcttacaggtta-3'. The sequence aaaaag was replaced by aagaa (underlined in the forward primer sequence) that does not alter the amino acid sequence.

Construction of Expression Plasmids for the Precursor—To generate expression plasmids for the precursor, the following modifications were made. The 3'-untranslated region was removed by further subcloning an ~3.8-kb Nhel-BgIII fragment from the cDNA between Nhel and BamHI sites of pcDNA6/V5/His-A. The sequence around the initiation codon was modified in an attempt to improve expression. A ~1.1-kb fragment between the Nhel site in the polylinker and the Xhol site in the cDNA was replaced by Nhel- and Xhol-cleaved PCR products amplified using various forward primers described below and a reverse primer, 5'-ctaaatgttcagagctgc-3'. Forward primers were 5'-ctggcggcggctc-3' and 5'-catctaatgctggcagagc-3'. The cycling parameters were 35 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 79 °C for 1 min. The product was used as a probe for screening a human genomic BAC library by Genome Systems (St. Louis, MO). Four genomic clones were identified, and BAC 14951 was isolated and sequenced as described previously (23–25). Briefly, fragments in the 1–3-kbp range were randomly sheared from the purified BAC DNA (50...
µg), blunt-ended, and subcloned into Smal site of pUC18. A random library was generated by transforming Escherichia coli strain XL1BlueMRF’ (Stratagene, La Jolla, CA) by electroporation. ~1200 colonies were picked from each transformation, and the sequencing templates were isolated by a cleared lysate-based protocol from the cultured library (24). Sequencing was performed as previously described (25). After base-calling with the ABI analysis software, the analyzed data were transferred to a Sun workstation cluster for base-calling and assembly using Phred and Phraps programs (26, 27), respectively. Overlapping sequences and contigs were analyzed using Consed (28). Gap closure and proofreading was performed using either custom primer-walking or PCR amplification of the region corresponding to the gap in the sequence followed by subcloning into pUC18 and cycle-sequencing with the universal pUC primers. In some instances additional synthetic custom primers were necessary for at least 3-fold coverage for each base.

Northern Blot Analysis—A cDNA fragment corresponding to nucleotides 1252–3356 of the cDNA was excised by Xhol and EcoRV from the plasmid containing the full-length cDNA. The cDNA was labeled with [γ-32P]dCTP using random hexamer priming with a Klenow fragment of DNA polymerase I (Readiprime II random prime labeling System, Amersham Biosciences) and used to probe a human MTN (multi-tissue northern) blot from BD Bioscience following the manufacturer’s recommendation, except that all the washing steps were performed at 65 °C. The membrane was exposed to a BIOMAX MS x-ray film (Eastman Kodak Co.) with an intensifying screen for 16 h at –80 °C.

Expression of the Precursor in 293T Cells—The human embryonic kidney cell line 293T was grown in Dulbecco’s modified Eagle’s medium with 10% (v/v) fetal bovine serum at 37 °C and 5% CO2, 95% air. Cells were transfected with empty vector or vector containing cDNA using FuGene-6 (Roche Applied Science). Cells were harvested by scraping at 72 h post-transfection and incubated in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 2 mM MgCl2) on ice for 30 min (14). The lysate was clarified by centrifugation at 40,000 g at 4 °C for 20 min, and the supernatant was assayed for GlcNAc-phosphotransferase activity as described below and protein concentration by Bradford assay (Pierce).

GlcNAc-phosphotransferase Assay—GlcNAc-phosphate transferred to α-methylmannoside was determined using [β-32P]UDP-GlcNAc as described previously (29). One unit of GlcNAc-phosphotransferase activity is defined as 1 pmol of GlcNAc-phosphate transferred to α-methylmannoside/h in a reaction containing 150 µM UDP-GlcNAc and 100 mM α-methylmannoside.

Alignments of Deduced Protein Sequences—GlcNAc-phosphotransferase sequences derived from databases were assembled and translated. Deduced protein sequences were aligned using ClustalW (30) as implemented by MacVector™ 7.2 (Accelrys, San Diego, CA).

RESULTS

Bovine GlcNAc-phosphotransferase Protein Sequence—Bovine GlcNAc-phosphotransferase was purified 488,000-fold to apparent homogeneity as previously described (19), subjected to non-reduced SDS-PAGE, and transferred to a polyvinylidene fluoride membrane. The bands corresponding to the α-subunit dimer (320 kDa) and β-subunit (56 kDa) were excised and subjected to amino-terminal sequencing. Duplicate samples were subjected to in situ Lys-C digestion and fractionated by reverse phase high performance liquid chromatography, and several isolated peptides were subjected to amino-terminal microsequencing. The α-subunit bands generated a single amino-terminal sequence of MLKLLLRQRTY and internal peptide sequences of VPM-LVLDXAXPTXVXLK and DLPLOYPSLSASDVFNVAKPR. The β-subunit bands generated an amino-terminal sequence of DTFAFDSSLYNKILNSKFGTSRKVPWV and internal peptide sequences of ILNSK, TSHK, FGFTSR, and SLVTNCKPVTDK. Two of the internal sequences, ILNSK and FGFTSR, overlap with the amino-terminal sequence.

Molecular Cloning of a Partial Mouse GlcNAc-phosphotransferase β-subunit cDNA—Bovine GlcNAc-phosphotransferase β-subunit peptide sequences were used to search the EST data base using the TBLASTN algorithm (31). The search identified a mouse cDNA (GenBank™ accession number L36434) that was previously incorrectly annotated as a basic domain-leucine zipper transcription factor (32). The 1846-bp murine clone contained highly homologous sequences to the determined β-subunit amino-terminal and internal peptide amino acid sequences. The amino-terminal sequence of the bovine β-subunit began at nucleotide 799 of the mouse cDNA. The 798-bp sequence 5’ to this revealed an in-frame open reading frame that lacked evidence for either a signal peptide or an initiator methionine, suggesting the β-subunit might be derived from a larger precursor. The murine clone was extended by 3’-RACE as described under "Experimental Procedures,” yielding a ~4600-bp partial cDNA. Attempts to extend the 5’-sequence by 5’-RACE, library screening, and data base searching were unsuccessful.

Molecular Cloning and Characterization of Human GlcNAc-phosphotransferase α- and β-subunits Precursor cDNA—A 1.1-kb fragment from the mouse cDNA was used to screen a human placental cDNA library in Agt10 under reduced stringency. Several positive clones were isolated, and the clone with the largest insert (3.4 kb) was sequenced. The remaining portion of the cDNA was cloned by a combination of a walking strategy, 3’-RACE, and data base searching. The placental sequence was used to design nested 3’-RACE primers that allowed the isolation of a 3’-RACE product from human brain Marathon-Ready™ cDNA. The 3426-nucleotide 3’-RACE product contained a 1761-nucleotide coding region and a 1665-nucleotide 3’-untranslated region. The placental sequence was also used to search the EST data base using the BLASTN algorithm (31). The search identified an L.M.A.G.E. clone 682681 (GenBank™ Accession Number AA204698) which contained a 1271-bp insert composed of a 164-nucleotide 5’-untranslated region and 1106-nucleotide coding region that overlapped the 5’ portion of the placental sequence. Together, these strategies allowed the isolation of a full-length cDNA that spanned 5597 bp. The cDNA sequence for human GlcNAc-phosphotransferase has been deposited to GenBank™ and the accession number is AY667932.

Deduced Structure of GlcNAc-phosphotransferase α/β-subunits Precursor—The nucleotide and deduced amino acid sequences of human GlcNAc-phosphotransferase α/β-subunits precursor are shown in Fig. 1. The 5597-bp cDNA is predicted to encode a protein of 1256 amino acids. The cDNA appears to encode an precursor with the α-subunit in the 5’-position and the β-subunit in the 3’-position. Sequences highly homologous to all the amino-terminal and internal-peptide sequences of both the α- and β-subunit of bovine GlcNAc-phosphotransferase are represented in the predicted human protein sequence (89 and 100% identity in the α- and β-subunits, respectively). The precursor protein has a predicted molecular mass of 143,582 Da, of which 104,706 and 38,894 Da are for the α- and β-subunits, respectively. The sequence contains 17 consensus sites for N-glycosylation in the α-subunit and 3 in the β-subunit. It also contains 22 cysteine residues, 18 in the α-subunit and 4 in the β-subunit.

The sequence surrounding the proposed initiation codon (ggtgcatga) has a guanosine in position –3 and a cytidine in position +4, yielding a non-preferred translation initiation sequence (33). There is an
FIGURE 1. Nucleotide and deduced amino acid sequences of the human GlcNAc-phosphotransferase αβ-subunits precursor. The predicted protein sequence is shown below the DNA sequence. The putative transmembrane domains are underlined. Sequences homologous to the experimentally identified bovine α- and β-subunit amino acid sequences are indicated by dashed underlines. The amino terminus of the β-subunit is indicated by an arrow. The 20 potential sites for N-linked glycosylation are indicated by asterisks. A n in-frame stop codon upstream of the initiation codon is double-underlined. A potential polyadenylation signal is shown by dotted underline.
FIGURE 1 — continued

GlcNAc-phosphotransferase α/β-subunits Precursor cDNA
in-frame stop codon 129 nucleotides upstream from the initiation codon. The cDNA has a single potential polyadenylation signal (ATAAAA) 13–18 nucleotides upstream from the poly(A) tail, although the signal was not the more commonly found AATAAA or ATTAAA.

Membrane Topology of GlcNAc-phosphotransferase α- and β-subunits—Examination of a Kyte-Doolittle hydrophilicity plot (34) generated by MacVector™ 7.2 (Accelrys, San Diego, CA) of the predicted GlcNAc-phosphotransferase precursor sequence (Fig. 2) reveals prominent hydrophobic segments of 24 and 26 residues near the amino terminus and the carboxyl terminus, respectively. Because the amino terminus of the deduced human α-subunit sequence was similar to the amino terminus of the bovine α-subunit sequence, the α-subunit contains a signal anchor. The hydrophobic segment in the carboxyl terminus is followed by a cluster of basic residues. This suggests that the GlcNAc-phosphotransferase α- and β-subunits have type I and II topology, respectively.

Expression of GlcNAc-phosphotransferase α/β-subunits precursor mRNA in Human Tissues—Northern blot analysis (Fig. 3) revealed a single transcript in all human tissues examined (Fig. 3). The ~6.6-kb mRNA was slightly larger than the 5.6-kb cDNA isolated. This may be due to an unusually long poly(A) tail or an incomplete 5′-untranslated region. Longer autoradiography revealed a transcript of the same size in kidney, which is not visible in Fig. 3 (not shown). There was no evidence for individual α- or β-subunits RNA transcripts or alternative splicing.

Expression of Human GlcNAc-phosphotransferase in Mammalian Cells—A set of expression plasmids for human GlcNAc-phosphotransferase α/β-subunits precursor was constructed in pcDNA6/V5/His-A. The cDNA for the precursor was altered in an attempt to increase the expression level by replacing the native sequence surrounding the initiation codon (ggggtgatgc) with a preferred Kozak consensus sequence (33). The preferred consensus sequence (gccaccatgg) that contains guanosine at position +4 required changing the second amino acid from leucine (ctg) to glycine (ggg), alanine (gcg), or valine (gtg). The control plasmid with leucine had the sequence gccaccatgc, where the initiation codon is underlined.

When 293T cells were transfected with the precursor expression plasmids, GlcNAc-phosphotransferase activity was increased at 72 h, 12–17-fold compared with endogenous level in mock-transfected cell lysate (Fig. 4A). Interestingly, the highest activity was obtained when the second amino acid remained leucine. This suggested that leucine is preferred over other amino acids that better satisfy the Kozak consensus features at position +4.
Co-transfection of the precursor expression plasmid (Leu+) with the γ-subunit expression plasmid resulted in a 3.6-fold increase in GlcNAc-phosphotransferase activity when compared with the mock transfection (Fig. 4B). For comparison, when the precursor expression plasmid was co-transfected with empty vector, a 5.7-fold increase was observed. Similar results were obtained with the Glyγ expression plasmid, but the overall expression level was lower. The reduced activity on α-methylnosside by co-transfection of the regulatory γ-subunit is consistent with the role of this subunit in lowering the $K_m$ for lysosomal enzymes.

Isolation and Sequencing of a Gene Encoding the Human GlcNAc-phosphotransferase α/β-subunits Precursor—Human BAC 1495113 was isolated using a 107-bp probe that corresponded to nucleotides 181–287 in the cDNA as described under “Experimental Procedures” and contained a 177.4-kb insert encoding the full-length precursor gene as diagrammed in Fig. 5A. The sequence for BAC14951 was deposited in GenBank and given accession number AC005409. This BAC maps to chromosome 12q23 by Mega Blast (35). The gene spans 85.3 kb and is distributed in 21 exons. The detailed structure of the gene including the sequence of each intron-exon junction, size, and location in the chromosome framework are provided in Supplemental Fig. S1.

Inspection of the gene sequence showed that Intron 1 is unusually large (33.8 kb); however, no known gene or pseudogene was found in that region. The nucleotide sequences at the 5′-donor and 3′-acceptor sites of all the introns conform to the GT...AG rule (36). There are only three nucleotides in the coding region of the human GlcNAc-phosphotransferase precursor cDNA that did not match the genomic sequence. One apparent polymorphism was identified (g2096a) that does not change the protein sequence. Two other nucleotide substitutions (g1338a and t2866a) result in amino acid substitutions, V329A and L901Q, respectively.

GlcNAc-phosphotransferases in Other Organisms—Comparison of the GlcNAc phosphotransferase α/β-subunits precursor cDNA sequence with available databases allowed the identification of apparent GlcNAc phosphotransferases in a variety of organisms. Putative full-length sequences were identified in Mus musculus, Dictyostelium discoideum, and Drosophila melanogaster. The deduced protein sequences for these sequences are aligned with the human sequence in supplemental Fig. S2. Additional incomplete sequences were identified in Rattus norvegicus, Bos taurus, Gallus gallus, and Danio rerio; the deduced partial protein sequences from these organisms are aligned in supplemental Fig. S3. Two of the three unique bovine β-subunit protein sequences determined were found in the deduced partial bovine sequence. The third β-subunit sequence and the α-subunit sequences were from regions not covered by the partial sequence. Inspection of the aligned vertebrate sequences in supplemental Fig. S3 demonstrates that the proteins are highly homologous. Between human and mouse α-subunit sequences, amino acids 1–195 define a region with a remarkable 94% sequence identity. A second region from amino acids 207–740 demonstrates more than 88% sequence identity. A region of less homology then extends from amino acid 741 to the carboxyl terminus of the α-subunit at amino acid 928, except from a short region from amino acid 876 to 928 with a >95% sequence identity. The carboxyl-terminal portion of the α-subunit sequence demonstrates the least similarity, and this region is absent from the D. discoideum and D. melanogaster sequences (Supplemental Figs. S2 and S3). The β-subunit sequences are also highly homologous, ranging from 97% for the mouse-human comparison to 92% identity for the chicken-human comparison.

Sequence Motifs in GlcNAc-Phosphotransferase—Comparison of the deduced GlcNAc-phosphotransferase sequences with the sequence databases identified two sequence motifs in the α-subunit (Fig. 6). A 113-amino acid sequence extending from amino acids 321 to 432 has similarity to a group of bacterial capsular polymerases. The human sequence demonstrates 46% sequence identity and 21% similarity with Neisseria meningitidis XcbA over this region (supplemental Fig. S4A). The aligned sequences are compared in Fig. 6. The α-subunit also contains two Notch-like repeats (pfam 00066.11) that span amino acids 433–469 and 500–536. Although the overall similarity to the Notch repeat consensus sequence is only 47 and 46% for repeats one and two, respectively, the positions of the six cysteines satisfy the Notch consensus for spacing. Twelve of the 18 cysteines in the α-subunit are found in the Notch-like repeats (Fig. S4B).

**DISCUSSION**

We isolated a single cDNA and gene encoding both the α and β subunits of human GlcNAc-phosphotransferase. Using purified bovine GlcNAc-phosphotransferase, we identified three different subunits and proposed that they were in an αβγγ configuration (6). The model for the 540-kDa complex was based on the presence of disulfide-linked α2 and γ2 dimers. The stoichiometry of the β-subunits was determined by quantitative protein sequencing that demonstrated a 1:1 molar ratio of β-subunit to γ-subunit. The identification of a single cDNA encoding the α and β subunits suggests that these two subunits are derived by proteolytic processing of a polyprotein precursor. This finding further supports the model containing two β-subunits per α2 dimers. The
GlcNAc-phosphotransferase α/β-subunits Precursor cDNA

Because the γ-subunit sequence does not predict a transmembrane domain, the retention of GlcNAc-phosphotransferase in the endoplasmic reticulum and cis-Golgi may be controlled by the transmembrane domains in the α- and β-subunits. The process of assembling the γ-subunit with the α- and β-subunits is still unclear. Although the γ-subunit showed limited similarity to endoplasmic reticulum glucosidase II β-subunit, no homology was found between the α- or β-subunits of GlcNAc-phosphotransferase and endoplasmic reticulum glucosidase II α-subunit.

Another sequence homology identified was two Notch-like repeats between amino acids 433–469 and 500–536 (Fig. 6 and supplemental Fig. S4B). Unlike Notch proteins, where there are three tandem repeats, only two repeats were found in the α-subunit of GlcNAc-phosphotransferase. These two Notch-like repeats were separated by ~30 amino acids, whereas no such separation is found in Notch proteins. In Notch proteins each repeat contains six cysteines that form three disulfide bonds between the first and fifth, the second and fourth, and the third and sixth cysteines (pfam00661.11, Notch). Notch repeats are reported to be involved in negative regulation of Notch proteins (40, 41). The role of the Notch-like repeats in GlcNAc-phosphotransferase, if any, is unknown.

Of the 18 cysteines in the α-subunit sequence, 12 are likely to be involved in Notch-related disulfide bonds. Two of the other 6 cysteines are in the amino-terminal membrane anchor domain, and 1 at position 562 is not conserved in the mouse homologue. Therefore, cysteines at position 70, 128, and/or 133 appear most likely to be involved in the observed intermolecular disulfide linkage(s) between the two α-subunits. The β-subunit contains 4 cysteines, all of which are presumed to be in intramolecular disulfide linkages, since the β-subunits are not disulfide-linked to each other or another subunit.

Sequences homologous to human GlcNAc-phosphotransferase α/β-subunits precursor protein were found both in vertebrates (supplemental Fig. S3) and invertebrates (supplemental Fig. S2). In vertebrates, an apparent full-length sequence was identified in mouse, and partial sequences were identified in rat, cow, chicken, and zebrafish. These vertebrate sequences are highly homologous to each other and to the human sequence. In invertebrates, two putative full-length sequences were found in slime mold (D. discoideum) and fruit fly (D. melanogaster). Interestingly, these invertebrate sequences have shorter α-subunits than human or mouse GlcNAc-phosphotransferase, lacking the carboxyl-terminal portion (Fig. 6 and supplemental Fig. S2). However, the sequences similar to bacterial capsular polymerase and Notch-like repeats are present (Fig. 6). The vertebrate and invertebrate β-subunits sequences are more similar (supplemental Fig. S2). It remains to be demonstrated that these similar sequence represent invertebrate GlcNAc-phosphotransferase.

The gene for GlcNAc-phosphotransferase precursor was localized to human chromosome 12q. This is in contrast to an abstract suggesting the gene for ML II (I-cell disease) resides at 4q21–4q23 (42). We found no evidence for an assignment to chromosome 4 and believe conclusion of the previous report is in error. Recently, we reported a mutation in the precursor gene in an ML IIIA (pseudo-Hurler polydystrophy) patient who had a very low level of GlcNAc-phosphotransferase activity (43). Mutations in the precursor gene have been identified in additional ML II and ML IIIA patients.

In summary, we cloned the cDNA and gene encoding the precursor of GlcNAc-phosphotransferase α/β-subunits. This completes the cloning of both the cDNA and genomic DNA of all the three subunits of

---

3 M. Kudo, M. Brem, and W. M. Canfield, manuscript in preparation.
GlcNAc-phosphotransferase. This allows the expression of recombinant GlcNAc-phosphotransferase and \textit{in vitro} phosphorylation of lysosomal hydrolase as we report later.\textsuperscript{4} Additionally, the genomic structure allows us to dissect the relationships between the complex and related disorders of lysosomal targeting, mucolipidosis II (I-cell disease), IIIA (pseudo-Hurler polydystrophy), and IIIC.

REFERENCES

1. Reitman, M. L., and Kornfeld, S. (1981) \textit{J. Biol. Chem.} 256, 11977–11980
2. Reitman, M. L., and Kornfeld, S. (1981) \textit{J. Biol. Chem.} 256, 4275–4281
3. Waheed, A., Hasilik, A., and von Figura, K. (1982) \textit{J. Biol. Chem.} 257, 12322–12331
4. Varki, A., and Kornfeld, S. (1981) \textit{J. Biol. Chem.} 256, 9937–9943
5. Varki, A. P., Reitman, M. L., and Kornfeld, S. (1981) \textit{J. Biol. Chem.} 256, 7773–7777
6. Bao, M., Booth, J. L., Elmendorf, B. J., and Canfield, W. M. (1996) \textit{J. Biol. Chem.} 271, 31446–31451
7. Raas-Rothschild, A., Cormier-Daire, V., Bao, M., Genin, E., Salomon, R., Brewer, K., Zeigler, M., Mandel, H., Toth, S., Roe, B., Nagaraju, K., and Plotz, P. (2003) \textit{Am. J. Hum. Genet.} 73, 1574–1579
8. Grabowski, G. A., Barton, N. W., Pastores, G., Dambrosia, J. M., Banerjee, T. K., McKee, M. A., Parker, C., Schifflmann, R., Hill, S. C., and Brady, R. O. (1995) \textit{Ann. Intern. Med.} 122, 33–39
9. Raben, N., Danon, M., Gilbert, A., Dwivedi, S., Collins, B., Thurberg, B., Mattaliano, R., Nagaraju, K., and Plotz, P. (2003) \textit{Mol. Genet. Metab.} 80, 159–169
10. Lobel, P., Fujimoto, K., Ye, R. D., Griffiths, G., and Kornfeld, S. (1989) \textit{Cell} 57, 787–796
11. Canfield, W. M., Johnson, K. F., Ye, R. D., Gregory, W., and Kornfeld, S. (1991) \textit{J. Biol. Chem.} 266, 5682–5688
12. Jadot, M., Canfield, W. M., Gregory, W., and Kornfeld, S. (1992) \textit{J. Biol. Chem.} 267, 11069–11077
13. Hasilik, A., Waheed, A., and von Figura, K. (1981) \textit{Biochem. Biophys. Res. Commun.} 98, 761–767
14. Varki, A. P., Reitman, M. L., and Kornfeld, S. (1981) \textit{Proc. Natl. Acad. Sci. U. S. A.} 78, 7773–7777
15. Varki, A., Reitman, M. L., Vannier, A., Kornfeld, S., Grubb, J. H., and Sly, W. S. (1982) \textit{Am. J. Hum. Genet.} 34, 717–729
16. Waheed, A., Pohlmann, R., Hasilik, A., von Figura, K., van Elsen, A., and Leroy, J. (1982) \textit{Biochem. Biophys. Res. Commun.} 105, 1052–1058
17. Reitman, M. L., Varki, A., and Kornfeld, S. (1981) \textit{J. Clin. Investig.} 67, 1574–1579

\textsuperscript{4} M. Kudo and W. M. Canfield, manuscript in preparation.
The α- and β-subunits of the Human UDP-N-acetylglucosamine: Lysosomal Enzyme Phosphotransferase Are Encoded by a Single cDNA
Mariko Kudo, Ming Bao, Anil D’Souza, Fu Ying, Huaqin Pan, Bruce A. Roe and William M. Canfield

J. Biol. Chem. 2005, 280:36141-36149.
doi: 10.1074/jbc.M509008200 originally published online August 24, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M509008200

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

Supplemental material:
http://www.jbc.org/content/suppl/2005/09/22/M509008200.DC1

This article cites 42 references, 17 of which can be accessed free at
http://www.jbc.org/content/280/43/36141.full.html#ref-list-1
VOLUME 280 (2005) PAGES 36141–36149

The α- and β-subunits of the human UDP-N-acetylgalactosamine:lysosomal enzyme N-acetylgalactosamine-1-phosphotransferase are encoded by a single cDNA.

Mariko Kudo, Ming Bao, Anil D’Souza, Fu Ying, Huaqin Pan, Bruce A. Roe, and William M. Canfield

PAGE 36141:

“N-acetylgalactosamine-1-" was inadvertently omitted from the full name of the enzyme in the title. The correct title is shown above.