Isolation, Characterization, and Expression of a cDNA Encoding N-Acetylgalcosaminyltransferase V*

(Received for publication, February 16, 1993, and in revised form, April 12, 1993)

Mohamed Shoreibah‡, Guang-Shing Perng§, Beverly Adler‡, Jasminder Weinstein‡, Ria Basu‡, Rod Cupple‡, Duanzhi Wen‡, Jeffrey K. Brown‡, Phillip Buckhaults‡, Nevis Fregien‡, and Michael Pierce‡

From the ‡Department of Biochemistry and Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, Xamgen, Inc. Thousand Oaks, California 91320, and the §Department of Cell Biology and Anatomy, University of Miami School of Medicine, Miami, Florida 33101

A cDNA clone for the complete coding sequence for α-1,3(6)-mannosylglycoprotein β-1,6-N-acetylgalcosaminyltransferase V (GlcNAc-T V) was isolated and expressed in COS-7 cells. Degenerate oligonucleotide primers for polymerase chain reaction were synthesized based on the amino acid sequence of three tryptic peptides isolated from affinity-purified GlcNAc-T V. Polymerase chain reaction amplimers were isolated from rat and mouse mRNA. A cDNA encoding full-length enzyme was isolated from a rat 1 cell (EJ-ras-transformed) library and sequenced. Transient expression of this clone in COS-7 cells, followed by enzymatic activity assays, demonstrated that this cDNA sequence encodes GlcNAc-T V. Northern analysis of rat kidney mRNA revealed a single band corresponding to a length of about 7 kilobases. Sequence analysis of the cDNA clone demonstrated an open reading frame that encoded a type II membrane protein of 740 amino acids.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M14284.

The abbreviations used are: GlcNAc-T V, α-1,3(6)-mannosylglycoprotein β-1,6-N-acetylgalcosaminyltransferase V (GlcNAc-T V)*; β(1,6)-linkage to α(1,6)-linked mannos of biantennary N-linked oligosaccharides (1). This enzyme is of particular interest because its activity in many cell types has been shown to significantly increase after transformation by diverse tumor genes (4-6) and v-src; and v-src. In addition, changes in its activity appear to correlate with changes in the metastatic potential of several cell lines (7, 8). The branch synthesized by GlcNAc-

α-1,3(6)-Mannosylglycoprotein β-1,6-N-acetylgalcosaminyltransferase V (GlcNAc-T V)* is a Golgi enzyme that catalyzes the transfer of N-acetylgalosamnine (GlcNAc) in β(1,6)-linkage to the α(1,6)-linked mannos of biantennary N-linked oligosaccharides (1). This enzyme is of particular interest because its activity in many cell types has been shown to significantly increase after transformation by diverse tumor viruses (2, 3) and isolated oncogenes, including several ras genes (4-6) and v-src. In addition, changes in its activity appear to correlate with changes in the metastatic potential of several cell lines (7, 8). The branch synthesized by GlcNAc-

* This research was supported in part by National Cancer Institute Grant CA 49926 (to N. F.). 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.

EXPERIMENTAL PROCEDURES

Preparative Purification of GlcNAc-T V, Proteolytic Digestion, and Microsequencing—GlcNAc-T V was purified essentially as described previously (18). Briefly, eluants from multiple purifications using the synthetic oligosaccharide affinity column containing about 30 μg of protein in a volume of 150 ml were concentrated by precipitation overnight in 7 volumes of methanol at —20°C. The precipitate was centrifuged at 10,000 × g at 4°C for 2 h, and the resulting pellet was resuspended in 100 μl of 0.1% SDS at room temperature. This sample was then applied to an Applied Biosystems high performance electrophoresis chromatography (model 230A) to purify the 75- and 69-kDa bands representing the GlcNAc-T V specific activity over endogenous levels, demonstrating that the protein it encodes is catalytically active. A fragment of the mouse cDNA specifically hybridized to a rat kidney mRNA with a size of about 7 kb.

TV expresses in most cells the majority of the N-linked poly-
N-acetyllactosamine (polyLacNAc) sequences (3, 9); there-
fore, the regulation of GlcNAc-T V activity in large part can control the expression of polyLacNAc on cell surfaces. This is important because polyLacNAc sequences have been implicated in mediating cell adhesion (10, 11). These cell surface changes involving increases in the β(1,6) branch and polyLacNAc expression after transformation help to explain at least in part many seminal observations by Warren, Glick, Buck, and other investigators (12, 13) who have documented changes in N-linked oligosaccharides after malignant trans-
formation.

GlcNAc-T V is expressed in low copy number in many tissues and cultured cell lines, although its specific activity varies considerably between tissues (14, 15). Two lectin-variant cells are known that have specific lesions in the expres-
sion of GlcNAc-T V activity, the mouse lymphoma BW5147 PHAR and Chinese hamster ovary Lec 4 variants (1, 16). Another Chinese hamster ovary variant, Lec 4A, expresses activity, but its GlcNAc-T V has altered kinetic properties and is mislocalized, causing a lectin-resistant phenotype similar to that of the Lec 4 cells (17). Recently, we purified the enzyme over 400,000-fold to homogeneity from rat kidney acetone powder (18).

In order to study in detail GlcNAc-T V and the regulation of its activity, we have isolated a cDNA that contains the entire coding sequence for GlcNAc-T V. The sequence encodes a type II membrane protein of 740 amino acids. When transiently expressed in COS-7 cells, the cDNA produced a 16-fold increase in GlcNAc-T V specific activity over endogenous levels, demonstrating that the protein it encodes is catalytically active. A fragment of the mouse cDNA specifically hybridized to a rat kidney mRNA with a size of about 7 kb.

EXPERIMENTAL PROCEDURES

Preparative Purification of GlcNAc-T V, Proteolytic Digestion, and Microsequencing—GlcNAc-T V was purified essentially as described previously (18). Briefly, eluants from multiple purifications using the synthetic oligosaccharide affinity column containing about 30 μg of protein in a volume of 150 ml were concentrated by precipitation overnight in 7 volumes of methanol at —20°C. The precipitate was centrifuged at 10,000 × g at 4°C for 2 h, and the resulting pellet was resuspended in 100 μl of 0.1% SDS at room temperature. This sample was then applied to an Applied Biosystems high performance electrophoresis chromatography (model 230A) to purify the 75- and 69-kDa bands representing the GlcNAc-T V specific activity over endogenous levels, demonstrating that the protein it encodes is catalytically active. A fragment of the mouse cDNA specifically hybridized to a rat kidney mRNA with a size of about 7 kb.

**
Microchemistry Facility where it was reduced with dithiothreitol, alkylated with iodoacetamide, resuspended in urea, and then subjected to trypsin digestion. The sample was then chromatographed by using reverse phase HPLC (Vydac C18, 2.1 x 150 mm), and the amino acid sequence of three peaks was determined. These peaks were chosen for gas-phase Edman degradation based on the magnitude of their absorbances at 210 and 280 nm and on their clear separation from other peaks.

cDNA Libraries and Screening Procedures—Three cDNA libraries were screened for the isolation of the GlcNAc-T V cDNA sequences: an oligod(T) primed mouse cDNA library was constructed from poly(A)^+ RNA from BW5147 cells by using the Librainer 1 kit (Litrigen, Inc.) that uses the vector pcDNA; a random primed cDNA library was constructed from mouse 3T3 cells by primer(A)^+ RNA and cloned into X-ZAP II (Stratagene) (19); and a rat 1-EJ cDNA library (20) was cloned into pSPORT-1 plasmid vector (Life Technologies, Inc., Bethesda, MD). The libraries were screened by conventional filter hybridization techniques (21) by using either synchrotron-labeled probes and autoradiography or fluorescent-labeled probes and immunological detection (enhanced chemiluminescence 3'-oligolabeling and detection kit, Amersham Corp.). In some cases the libraries were divided into pools and screened for the presence of GlcNAc-T V sequences by PCR analysis, prior to hybridization.

Oligonucleotide Synthesis, Reverse Transcriptase-PCR, and DNA Sequencing—Oligodeoxynucleotides were synthesized on an Applied Biosystems automated DNA synthesizer. PCR reactions were done using Thermus aquaticus (Tag) DNA polymerase (Perkin-Elmer Cetus or Promega) utilizing manufacturer-supplied buffers. Oligodeoxynucleotides for reverse transcriptase-PCR were synthesized from either total cytoplasmic RNA or poly(A)^+ RNA by Moloney murine leukemia virus-reverse transcriptase (Life Technologies, Inc.) by using the supplied buffer and oligod(T) (total RNA) or random hexanucleotides (poly(A)^+ RNA) as a primer. Sequences were determined by using the Applied Biosystems, Inc. automated DNA sequencer (Applied Biosystems 373A) and following the manufacturer’s instructions.

Transient Expression of Rat GlcNAc-T V in COS-7 Cells—The entire cDNA insert from one rat GlcNAc-T V clone was ligated into pCT-2 plasmid expression vector (20). COS-7 cells (American Type Culture Collection, CRL 1651, Rockville, MD) were transfected with the pCT-2 plasmid alone or with pCT-2 plasmid containing the rat cDNA by electroporation as follows. 4 x 10^6 cells in 0.5 ml of Dulbecco’s modified Eagle’s medium (Life Technologies, Inc.) and 7.5% fetal bovine serum (Bocknek, Toronto, Canada) were transferred to a 0.4-cm cuvette and mixed with 10 µg of plasmid DNA in 10 µl of water. Electroporation was performed at room temperature at 1600 V and 25 microfarads by using a Gene Pulser apparatus (Bio-Rad) with the pulse controller unit set at 200 ohms. The cells were then transferred immediately 40 ml of Dulbecco’s modified Eagle’s medium, 7.5% fetal bovine serum, transferred to 100-mm culture dishes. After a 17-h incubation at 37 °C, the medium was replaced, and incubation was continued for an additional 51 or 75 h. The cells were rinsed with PBS, harvested by scraping, washed again with PBS, and centrifuged to pellet the cells. After the PBS had been aspirated, the cell pellet was subjected to quick freezing by immersing the tube in liquid nitrogen and was kept frozen on dry ice until resuscitated for assay.

GlcNAc-T V Enzymatic Assays—The pellets of electroporated COS-7 cells were thawed at 37 °C and assayed by radiochemical (22) and enzyme-linked immunosorbent assays (23). The radiochemical assay mixture contained the following reagents that were dried under vacuum in a 1.5-ml microcentrifuge tube: 10^6 cpm of UDP-[H] GlcNAc (25 cpm/µmol) (American Radiolabeled Chemicals), 1 mM synthetic oligosaccharide acceptor (generously provided by Dr. O. Hindsgaul, University of Alberta), and 10 mM EDTA in total volume of 0.01 ml. Radiolabeled product was isolated and quantitated as described (23). All assays were conducted in duplicate for three time points, and the results were averaged. The synthetic trisaccharide acceptor used was octyl-β-D-glucopyranosyl-(1→2)-octyl-β-D-glucopyranosyl-(1→6)-octyl-β-D-glucopyranoside.

Analysis of Rat Kidney mRNA—Ten µg of rat kidney poly(A)^+ RNA (Clontech) was subjected to electrophoresis in a 1% agarose gel and blotted to a Zeta-Probe GT membrane. A BglII-HindIII fragment of murine GlcNAc-T V cDNA was radiolabeled and used as a probe. The blot was visualized by using a Molecular Dynamics phosphorImager. A laser printer output was generated and photographed.

RESULTS AND DISCUSSION

The strategy employed to isolate the GlcNAc-T V cDNA was to obtain amino acid sequence data from peptide fragments generated by trypsin digestion of highly purified GlcNAc-T V and then to utilize these data to design degenerate oligonucleotides to be used as primers for the reverse transcriptase-PCR analysis of rat and mouse mRNAs. First, microgram quantities of GlcNAc-T V were purified from rat kidney acetone powder by affinity chromatography as described (18) and were further purified by preparative SDS-polyacrylamide gel electrophoresis with direct elution of samples to ensure purity after scale-up of the purification procedure and to maximize recovery. Fractions containing protein species corresponding to both the 75- and 69-kDa forms of the enzyme were pooled and digested with immobilized trypsin; the resulting peptides were fractionated by reverse-phase HPLC. Three peptides were selected for N-terminal amino acid sequence determination as follows: Peptide 1, Asn-Thr-Asp-Phe-Phe-Ile-Gly-Lys-Pro-Thr-Leu-Arg; Peptide 2, Ala-Ile-Leu-Asn-Gln-Ile-Glu-Pro-Tyr-Met-Pro-Tyr-Glu-Ph-Phe; Peptide 3, Val-Leu-Asp-Ser-Phe-Gly-Thr-Glu-Pro-Glu-Phe-Arn.

Oligonucleotides were designed from a region within Peptides 1 and 2, based on the least amount of degeneracy. Sequences were chosen from the central region of the peptide sequence when possible, since this would permit confirmation of any cDNA clones obtained by PCR by comparison of amino acids encoded by sequences adjacent to primer sequences with the known, peptide-derived amino acid sequences. Within these oligonucleotides, inosines were substituted at positions of four-base degeneracy. The selected oligonucleotides are shown as follows: GNT-V 1, 5'-AYACIGAYTTYTTYTAT (Y/A)GOGIAARCCNAC; GNT-V 1o (antisense), 5'-CGIT GYTITCC(R/T)ATRAAARTGCTRTT; GNT-V 2, 5'-AT(Y/A)GARCCITAYTGCCTAFYAGARYrrTYAC; GNT-V 2o (antisense), 5'-TCRTAIGGCG ATRTAAIYGCYT(R/T)ATYTTYT.

Since the relative positions of the two peptides in protein sequence were not known, oligonucleotides were synthesizing both sense (GNT-V 1 and 2) and antisense (GNT-V 1o and 2o sequences). This permitted PCR testing for the GlcNAc-T V mRNA in either of the possible orientations, with only one of two combinations of oligonucleotides capable of giving an amplification product.

The initial analysis used both total and poly(A)^+ RNAs from two cell lines known to express GlcNAc-T V enzyme activity, mouse BW5147 cells (1) and rat 13762 mammary adenocarcinoma cells (24). The RNAs were reverse transcribed into cDNA by reverse transcriptase primed by random hexanucleotides and then used in PCR reactions using the two possible combinations of degenerate primers, GNT-V 1 and 2o or GNT-V 1o and 2 primers. These results revealed that both poly(A)^+ RNAs give a 178-bp amplification product using the GNT-V 1 and 2o primers but no detectable bands from the other set of primers. Both amplifications products were cloned and sequenced. Their sequences showed that they contained a short open reading frame. Translation of the sequences revealed that the amplification produce encoded all seven of the amino acids predicted by the peptide sequence, but not included in the primer sequences, as well as a Lys at the presumptive tryptic cleavage site.

The PCR amplification product was used to screen two mouse cDNA libraries in order to isolate larger cDNA sequences. Screening approximately 10^9 clones from each library resulted in the isolation of two overlapping cDNA clones. Sequence analysis of these cDNAs demonstrated that they
cDNA Encoding GlcNAc-T V

contained an open reading frame encoding the C-terminal end of the protein, representing about 60% of the length of the polypeptides observed after SDS-PAGE (18). The remaining coding sequence was isolated from a Rat 1-EJ cDNA library by anchored primer PCR using a 5’-sense primer derived from the cDNA library cloning vector, pSPORT-1, and a 3’-antisense primer designed from the original 178-bp PCR amplification product. The PCR priming using these two oligomers would identify the 5’-end of the cDNA clones. Using this strategy, an approximately 1.8-kb PCR product was obtained. Sequence analysis of this fragment revealed overlap with the two partial mouse cDNA clones and completed the open reading frame region. The rat PCR amplification product contained the putative ATG translation initiation codon and approximately 300 bp of 5’-untranslated region. An oligonucleotide, 35 to 15 bp upstream from the ATG start codon, was used to screen the Rat 1-EJ cDNA library and resulted in the isolation of a plasmid containing a 4.8-kb cDNA insert. Restriction digestion and nucleotide sequence analysis demonstrated that the rat cDNA clone contained approximately 300 bp of 5’-untranslated sequence, a 2220-base open reading frame, and approximately 2300 bases of apparent 3’-untranslated region (Fig. 1).

The initiation of translation most likely occurs at the ATG at 299, since this codon is 9 bases after the TAG stop codon at 290. The open reading frame encodes a 740-amino acid protein with a calculated molecular mass of 84,561 Da. The amino acid sequences of all three peptide fragments are located within the predicted amino acid sequence encoded in the open reading frame and were identical to the original sequences. These are Peptide 1, amino acids 546-557; Peptide 2, amino acids 592-607; and Peptide 3, amino acids 375-386 (Fig. 1). Hydrophobicity analysis of the amino acid sequence (25) indicated a putative transmembrane region near the N-terminus, amino acids 14-30, as expected for a type II membrane protein similar to other lumenal Golgi enzymes. There are six consensus sites for N-linked glycosylation. Comparison of the sequences in the mouse and rat clones showed that they were greater than 95% homologous at the nucleotide level with only one difference in the C-terminal 445 amino acids, which results in a T residue at position 679 in the mouse sequence as compared with the I residue in rat GlcNAc-T V. When the rat GlcNAc-T V sequence was compared with those of other cloned glycosyltransferases, in particular GlcNAc-T I, GlcNAc-T II, and the “core 2” GlcNAc-T, no significant degree of similarity could be observed.

The entire cDNA insert of approximately 4.8 kb was subcloned into the pJT-2 expression vector. This vector, containing the sequence of the putative GlcNAc-T V clone, was transfected into COS-7 cells. Identical cultures of cells were transfected with vector alone as a control. The cells were harvested after 68 or 92 h and assayed for GlcNAc-T V activity by using the synthetic trisaccharide acceptor specific for GlcNAc-T V, UDP-[3H]GlcNAc, and standard assay conditions (14). EDTA was included in the incubation to provide additional evidence for GlcNAc-T V activity since GlcNAc-T V and very few other glycosyltransferases are active in vitro in the absence of exogenous cation. The results from this experiment, shown in Fig. 2, demonstrated that the level of

Fig. 1. Nucleotide and deduced amino acid sequence of rat GlcNAc-T V cDNA. Both strands of rat GlcNAc-T V cDNA were sequenced, and the nucleotide and deduced amino acid compositions are presented. The 5’-untranslated region consists of nucleotides 1–298. The methionine-specifying initiation codon (ATG) starts at nucleotide 299, and the amino acid coding region spans 2220 bases, encoding a 740-amino acid protein. The putative transmembrane domain (amino acids 14–30) is doubly underlined. Tryptic peptides are singly underlined. Consensus sites for N-linked glycosylation are indicated by asterisks. The first 103 bases of the 3’-untranslated region are also presented.
GlcNAc-T V in the cells transfected with the putative GlcNAc-T V sequence expressed activity at levels 16-fold higher than the cells that received vector alone. The specific activity of the population of cells expressing the GlcNAc-T V cDNA was 624 pmol/mg/h. Similar results were obtained from the transfected cell cultures incubated for 92 h. Moreover, assay of the same samples by using an enzyme-linked immunosorbent-based assay for GlcNAc-T V (23), which utilizes specific antibody to detect GlcNAc-T V oligosaccharide product, also demonstrated a large increase in activity in the cells transfected with GlcNAc-T V cDNA (data not shown).

Translation of the open reading frame encoded by the GlcNAc-T V cDNA predicts a protein of about 84.6 kDa. This is larger than the size observed for affinity-purified rat kidney GlcNAc-T V, which reveals a doublet of bands at 69 and 75 kDa after SDS-PAGE, suggesting that proteolysis of the kidney enzyme occurs during purification. Preliminary results suggest that when the affinity purification of the rat kidney enzyme is performed rapidly and in the presence of 5-fold greater levels of a mixture of protease inhibitors, an additional band at about 95 kDa is observed on silver-stained SDS-PAGE gels. The 11-kDa difference between the calculated molecular mass and the species with the largest observed molecular mass is most likely due to N-linked glycosylation since there are six of these potential sites in the sequence. An equivalent difference between calculated and observed molecular mass has been observed for GlcNAc-T III (26) and other glycosyltransferases. The cleavage of the intact enzyme to soluble, lower molecular mass species may reflect a normal proteolytic event that allows the enzyme to be released from cells, as has been observed for several other glycosyltransferases (27, 28). Soluble GlcNAc-T V activity has been detected in human serum (23).

A fragment from the mouse cDNA was radiolabeled and hybridized to a Northern blot containing poly(A)+ RNA from rat kidney. In order to gain enhanced sensitivity, the blot was visualized with a phosphorimaging system, and the results of this experiment are shown in Fig. 3. The cDNA hybridizes to a moderately broad band about 7 kb in size, suggesting that there may be more than one GlcNAc-T V transcript in rat kidney. Similar hybridization patterns have been observed in other rat and mouse tissues and cell lines. Moreover, these results demonstrate that less than 40% of this large mRNA encodes protein.

In conclusion, we have isolated a cDNA containing the entire GlcNAc-T V coding sequence that expresses catalytically active enzyme after transient transfection into COS-7 cells. This enzyme is by far the largest luminal Golgi glycosyltransferase to be reported since other cloned glycosyltransferases encode proteins with calculated molecular masses ranging from 40 to 60 kDa (27, 28). GlcNAc-T V was present in very low copy number in the cDNA libraries that contained it, and its message is large with a high degree of noncoding potential.

Acknowledgments—We thank Dr. Kelley Moremen for advice and assistance; Irene Margitich for technical assistance; and Dr. William S. Laine for expertise in peptide cleavage, HPLC, and microsequencing at the Harvard Microchemistry Facility.
cDNA Encoding GlcNAc-T V