Cloning and Expression of Human Core 1 β1,3-Galactosyltransferase*

Received for publication, September 19, 2001, and in revised form, October 24, 2001 Published, JBC Papers in Press, October 24, 2001, DOI 10.1074/jbc.M109060200

Tongzhong Ju‡§, Kevin Brewer‡, Anil D’Souza‡, Richard D. Cummings§¶, and William M. Canfield‡§**‡‡

From the ‡W. K. Warren Medical Research Institute, the Departments of §§Medicine and ¶¶Biochemistry and Molecular Biology, the $Oklahoma Center for Medical Glycobiology, and the University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104

The common core 1 O-glycan structure Galβ1→3GalNAc-R is the precursor for many extended mucin-type O-glycan structures in animal cell surface and secreted glycoproteins. Core 1 is synthesized by the transfer of Gal from UDP-Gal to GalNAc1-R by core 1 β3-galactosyltransferase (core 1 β3-Gal-T). Amino acid sequences from purified rat core 1 β3-Gal-T (Ju, T., Cummings, R. D., and Canfield, W. M. (2002) J. Biol. Chem. 277, 169–177) were used to identify the core 1 β3-Gal-T sequences in the human expressed sequence tag data bases. A 1794-bp human core 1 β3-Gal-T cDNA sequence was determined by sequencing the expressed sequence tag and performing 5′-rapid amplification of cDNA ends. The core 1 β3-Gal-T predicts a 368-amino acid type II transmembrane protein. Expression of both the full-length and epitope-tagged soluble forms of the putative enzyme in human 293T cells generated core 1 β3-Gal-T activity that transferred galactose from UDP-Gal to GalNAc1-O-phenyl, and a synthetic glycopeptide with Thr-linked GalNAc and the product was shown to have the core 1 structure. Northern analysis demonstrated widespread expression of core 1 β3-Gal-T in tissues with a predominance in kidney, heart, placenta, and liver. Highly homologous cDNAs were identified and cloned from rat, mouse, Drosophila melanogaster, and Caenorhabditis elegans, suggesting that the enzyme is widely distributed in metazoans. The core 1 β3-Gal-T sequence has minimal homology with conserved sequences found in previously described β3-galactosyltransferases, suggesting this enzyme is only distantly related to the known β3-galactosyltransferase family.

O-Glycosylation is an important post-translational modification found on many secreted and membrane-bounded glycoproteins. Mucin-type O-linked glycans extending from GalNAc1-O-Ser/Thr are synthesized in the Golgi apparatus by the stepwise transfer of sugars from sugar nucleotide donors to acceptor substrates through the sequential action of glycosyltransferases (1). The mucin-type O-glycans are classified according to the presence of a series of core structures. At least eight different O-glycans core structures have been described so far, with core 1 and 2 structures being the most common in non-mucin glycoproteins and core 1–4 being the most common in mucins (2).

The core 1 structure Galβ1→3GalNAc-R is synthesized from GalNAc1-R by the action of core 1 UDP-galactose:GalNAc-α-R β1,3-galactosyltransferase (core 1 β3-Gal-T,1 EC 2.4.1.122), where R is Ser/Thr. The core 1 β3-Gal-T activity was first described in porcine and ovine submaxillary glands (3) and has been found in most cell types and mammalian tissues (reviewed in Refs. 1 and 4). It is likely, however, that the activity of core 1 β3-Gal-T is regulated during cell differentiation and development. The core 1 structure is an intermediate in the synthesis of most extended O-glycans and may be further modified by the addition of GlcNAc or sialic acid. Interestingly, the unsubstituted core 1 disaccharide Galβ1→3GalNAc1-R is recognized as a cancer-associated antigen (T- or Thomsen-Friedenreich antigen) (5). The T-antigen is found in several types of cancers, including breast, bladder, and colon carcinomas.

Alterations in core 1 β3-Gal-T activity may play an important role in other diseases, such as Tn syndrome and IgA nephropathy (6–11). It is thought that these diseases may result from somatic mutations in core 1 β3-Gal-T, thereby causing expression of the truncated GalNAc-α-R (known as the Tn antigen). The Tn antigen is recognized by a naturally occurring Tn antibody in adult serum and in Tn syndrome results in polyagglutinability of blood cells (6, 12) or persistent mixed field polyaagglutination (13). A similar mechanism may underlie IgA nephropathy (14). The hinge region of IgA1 contains at least five O-glycosylation sites normally modified by the core 1 structure or its sialylated derivative (15). In the absence of core 1 β3-Gal-T, reduced terminal galactosylation of O-glycans may lead to a defective clearance of IgA, resulting in mesangial deposition and glomerular injury. Although these diseases might be predicted to result from deficiency of core 1 β3-Gal-T, the exact pathogenesis is unknown, because the key enzyme, core 1 β3-Gal-T, has not been amenable to molecular analysis.

* This work was supported in part by National Institutes of Health Grant RO1 AI48075-01 (to R. D. C.). 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) AF155582, AF157963, AF157962, AF295963, AC905592, AJ192943.

¶ To whom correspondence may be addressed: Department of Biochemistry and Molecular Biology, Univ. of Oklahoma Health Sciences Center, 975 N.E. 10th St., BRC Rm. 417, Oklahoma City, OK 73104. Tel.: 405-271-2481; Fax: 405-271-3910; E-mail: richard-cummings@ouhsc.edu.

†† Ts whom correspondence may be addressed: Novazyme Pharmaceuticals, Inc., 800 Research Pkwy., Ste. 200, Oklahoma City, OK 73104. Tel.: 405-271-8144; Fax: 405-271-1030; E-mail: wcanfield@novazyme.com.

1 The abbreviations used are: Core 1 β3-Gal-T, Core 1 UDP-galactose: N-acetylgalactosamine-α-R β3-galactosyltransferase; EST, expressed sequence tag; RACE, rapid amplification of cDNA ends; HPC4, monoclonal antibody to human protein C; AP1, adaptor primer 1 5′-CCATCTAATACGACTCACTATAGGGCTCGAGCGGC-3′; AP2, nested adaptor primer 5′-ACTCACTATAGGGCTCGAGCGGC-3′; CHO, Chinese hamster ovary; MES, 4-morpholinoethanesulfonic acid; MOPS, 4-morpholinopropane-sulfonic acid.

This paper is available on line at www.jbc.org
In the present study, we describe the cloning and expression of a human cDNA encoding core 1 β-Gal-T, using the amino acid sequence of the purified enzyme from rat liver, as described in the accompanying article (see Ju, et al. (33)). The human cDNA was then used to identify homologs in rat, mouse, Drosophila melanogaster and Caenorhabditis elegans. The core 1 β-Gal-T sequences described here have minimal homology to previously described β-galactosyltransferases, suggesting that core 1 β-Gal-T is a distinctly related member of this large enzyme family.

**EXPERIMENTAL PROCEDURES**

**Materials**—Rat and human multiple tissue Northern blots, Marathon ready cDNA from human, rat, and mouse, and Marathon RACE kits were obtained from CLONTECH (Palo Alto, CA). GalNAc1-O-phenyl, UDP-Gal, phenylmethylsulfonyl fluoride, benzamidine HCl, leupeptin, and Perfection™ hybridization buffer were obtained from Sigma. UDP-Gal-[32P] (40–60 Ci/mmol) was obtained from American Radiolabeled Chemicals, Inc. (St. Louis, MO). Dulbecco’s modified Eagle’s media and gentamicin sulfate were obtained from Invitrogen. SDS-PAGE gels were obtained from NOVEX, San Diego, CA. Cell lines 293T, CHO K1 and Lecl-CHO were obtained from American Type Culture Collection (Manassas, VA). Sep-Pak C18 cartridges was obtained from Waters Associates (Milford, MA). Nitrocellulose membrane was obtained from Bio-Rad. ESTs were obtained from Genome Systems, St. Louis, MO. A monoclonal antibody (antiIGG) was purchased from Charles Eson, Oklahoma Medical Research Foundation (Oklahoma City, OK). Restriction enzymes were obtained from New England Bioslabs, Beverly, MA. Vectors PCR 2.1™ and pcDNA 3.1(+) were obtained from Invitrogen (Carlsbad, CA). FuGene™, Thermus aquaticus DNA polymerase, and O-glycanase were obtained from Roche Molecular Biochemicals. The vector pcDNA 3.1(+)→TH that contains a human transferrin signal peptide and an HPC4 epitope was a gift of Dr. Ray Reitzas, Oklahoma Medical Research Foundation (Oklahoma City, OK). QUIgene Plasmid Purification kit and QIAquick Gel Extraction kits were obtained from Qiagen, Inc. (Valencia, CA). HighSignal West Pico Chemiluminescent Substrate, BCA protein assay kit, and UltraLink™ QIAgene Plasmid Purification kit and QIAquick Gel Extraction kits were obtained from CLONTECH (Palo Alto, CA). GalNAc3-epitope was purified from human plasma using the protocol and cultured in Dulbecco’s modified Eagle’s media containing 10% fetal bovine serum and were transfected with the soluble expression vector using FuGENE™ 6, as described above for the human 293T cells. The transfected cells were selected with 600 µg/ml gentamicin sulfate in the same media. After 2–3 weeks of selection, the stably expressing cell lines were obtained, and media from the cells were collected.

**Construction of an Expression Vector Encoding Wild-type Human Core 1 β-Gal-T**—In 293T cells expressing the core 1 β-Gal-T to characterize the possible post-translational modifications of the enzyme, we used both CHO K1 and Lecl-CHO cells. These cells were cultured in Dulbecco’s modified Eagle’s media containing 10% fetal bovine serum and were transfected with the soluble expression vector using FuGENE™ 6, as described above for the human 293T cells. The transfected cells were selected with 600 µg/ml gentamicin sulfate in the same media. After 2–3 weeks of selection, the stably expressing cell lines were obtained, and media from the cells were collected.

**Capture of Soluble Core 1 β-Gal-T on HPC4-Ultradinx—**HPC4-Ultradinx (100 µl) equilibrated with 50 mM Tris-HCl, pH 7.2, 100 mM NaCl, 1 mM CaCl2 was incubated with 500 µl of media containing human 293T cells expressing the HPC4 epitope-tagged core 1 β-Gal-T for 2 h at 4°C. The beads were collected by centrifugation (2,000 × g, 2 min) and washed three times with 500 µl of 50 mM Tris-HCl, pH 7.4, 1 mM NaCl, and 1 mM CaCl2. The beads were then washed once with equilibration buffer and directly assayed for core 1 β-Gal-T using GalNAc1-O-phenyl as the acceptor.

**Western Blot of Human HPC4 Epitope-tagged Soluble Core 1 β-Gal-T Expressed in Different Mammalian Cell Lines—**Twenty-five µl of media from transiently transfected 293T, stably transfected CHO K1, and stably transfected Lecl-CHO cells were electrophoresed on an SDS-PAGE (4–20%) and transferred to a nitrocellulose membrane. After blocking with 5% milk, the membrane was incubated with 10% of HPC4- monoclonal antibody (IgG1) at room temperature for 1 h. The membrane was then washed twice with a low salt solution (25 mM Tris-HCl, pH 7.4, containing 1 mM CaCl2 and 150 mM NaCl) and then washed three times with a high salt solution (25 mM Tris-HCl, pH 7.4, containing 1 mM CaCl2 and 500 mM NaCl). The membrane was then incubated with horseradish peroxidase-conjugated, goat anti-mouse IgG at room temperature for 1 h, washed as above with low and high salt buffers, and then incubated with 6 ml of HighSignal West Pico Chemiluminescent Substrate at room temperature for 1 min. The blot was exposed to a BioMax film (Kodak) for 30 s, and the film was developed.

**Identification of the Product Generated by the HPC4 Epitope-tagged Core 1 β-Gal-T**—The HPC4 epitope-tagged core 1 β-Gal-T in 40 µl of media from the stably transfected CHO-K1 cell line was captured by incubation with 200 µl of HPC4-Ultradinx beads at 4°C for 3 h. After washing with 50 ml of low salt solution and 50 ml of high salt solution (see above), 50 µl of the beads was incubated with 100 µg of PSGL-1 glycopeptide (4-GP-1) in 100 µl containing 100 mM MES, pH 6.5, 500 µM UDP-[3H]Gal (150,000 cpm), 20 mM MgCl2, and 2 mM MgATP. As a control, a reaction was carried out except that pure rat liver core 1 β-Gal-T (2 nmol/h) was used as the enzyme source instead of protein captured onto beads, and 0.2% of Triton X-100 was added. After a 16-h incubation at 37°C, the reaction mixtures were loaded onto C18 cartridges (500 mg) following the standard core 1 β-Gal-T assay (33). The bound material was eluted with 2.5 ml of methanol, and a 0.125-ml portion was mixed with 5 ml of Scintiverse-BD and radioactivity determined in a liquid scintillation counter.
The residual 2.375 ml from each sample was dried in a speed-vac concentrator. The dried material was redissolved in 200 μl of 50 mM Tris-HCl, pH 7.0, containing 150 mM NaCl. Twenty μl of the redissolved material was treated with 5 milliliters of G-lysogelose at 37 °C for 24 h. One-half of this reaction mixture was then loaded onto C18 Cartridge (50 μg). The cartridges were washed with 10 ml of water and bound material was eluted with 2 ml of methanol. The eluates were mixed with 18 ml of Scintiverse-BD and counted.

Northern Blot Analysis—A human multiple tissue Northern blot was probed with a 32P-labeled 147-bp EcoRI fragment corresponding to bp 179–325 of the human core 1 β3-Gal-T cDNA. The probe was random hexamer-labeled to 1 × 106 cpm/μg DNA (17, 18) and hybridized using standard techniques (19) overnight at 65 °C in PerfectHybl solution. After hybridization, the blots were washed twice in 2× SSC 0.1% SDS at room temperature for 20 min and 0.1× SSC 0.1% SDS at 65 °C twice for 30 min, and then autoradiographed.

Isolation of Mouse cDNA Encoding Core 1 β3-Gal-T—A mouse EST (AI060759) was identified by a BlastN search using the rat core 1 β3-Gal-T. The 426-bp EST was sequenced and found to be incomplete at the 3′ end. The cDNA was completed by nested 5′-RACE using CANF254 (5′-TGAGCACTTGCTGAACTTCCAAC-3′) and AP1 as primers and rat liver marathon-ready cDNA as template. PCR was performed for 35 cycles at 94 °C, 1 min; 94 °C, 30 s; 53 °C, 1 min; 72 °C, 2 min for 35 cycles. The second round reaction was performed with primers CANF312 (5′-AGCACGGATACATGAGTGGAGGAG-3′) and AP2, and an additional round of the first reaction as template was incubated at 94 °C, 30 s; 53 °C, 1 min; 70 °C, 1 min for 25 cycles followed by 72 °C for 10 min. The 682 bp band was excised, cloned into PCR 2.1, and sequenced.

Isolation of C. elegans Core 1 β3-Gal-T cDNA—BlastP searching using the human core 1 β3-Gal-T protein sequence demonstrated the C. elegans gene C38H2.2 was highly homologous. Core 1 β3-Gal-T was found using its cDNA with 1.28 kb as predicted. For isolation of its cDNA, two PCRs covering both 5′ and 3′ ends with 170-bp overlap were carried out using primers based on the cDNA from C38H2.2 using C. elegans cDNA library. The 5′ end PCR primer set utilized the forward primer 5′-CCACCATGAGGAGCAGGTTTG-3′ and the reverse primer 5′-CCACCATGAGGAGCAGGTTTG-3′. The derived amino-terminal sequence indicated only the initiator methionine had been removed, consistent with a type 1 transmembrane topology. The determined amino-terminal sequence of the predicted translation of the EST clone C41813 sequence was sequenced (930 bp) and found to be incomplete at the 3′ end. The 5′ portion of the cDNA was amplified by 5′-RACE and sequenced as described under “Experimental Procedures.” The 1794-bp cDNA sequence predicts an open reading frame of 1089 bp, predicting a protein of 363 amino acids (Fig. 1A). The human mRNA sequence has been deposited in GenBank™ as accession number AF155582.) Analysis of the hydropathy plot indicated no consensus sites for N-glycosylation.

Construction of Plasmids Encoding Wild-type and an HPC4 Epitope-tagged Soluble Form of Core 1 β3-Gal-T—To express the human wild-type core 1 β3-Gal-T, the Kozak sequence was optimized by the insertion of a -CCACC- sequence before the initiator ATG (22). The modified cDNA was cloned into the mammalian expression vector pcDNA3.1(−) generating the plasmid designated phCore1-WT. We also constructed a second plasmid designated phCore1-SF that encoded an HPC4 epitope-tagged soluble form of core 1 β3-Gal-T in which the cDNA sequences encoding the cytoplasmic and transmembrane domains were replaced by the human transferrin signal peptide and the HPC4 epitope as described under “Experimental Procedures.”

Expression of phCore1-WT and phCore1-SF—To confirm the cloned cDNA-encoded core 1 β3-Gal-T, the cDNA was expressed by transient transfection of the human embryonic kidney cell line, 293T. Cells were transfected with phCore1-WT, phCore1-SF, or the empty vector, pcDNA3.1(−), as described under “Experimental Procedures.” Cell monolayers were harvested at 24, 48, or 72 h after transfection; the cells were disrupted with Triton X-100, and core 1 β3-Gal-T activity was measured toward the acceptor GalNAc 1-phenyl. The β3-Gal-T activity was modestly and progressively increased when compared with cells transfected with empty vector yielding an ∼3.3-fold increase at 72 h (Fig. 2). The modest increases were in part because the parental cell line expresses significant core 1 β3-Gal-T activity. To exploit the observation that little core 1 β3-Gal-T activity was present in tissue culture media, a plasmid expressing a soluble form of the enzyme lacking its cytoplasmic and transmembrane domains was constructed and expressed in the same cell line. Core 1 β3-Gal-T activity in the media from these transfected cells was ∼37-fold above control media from cells transfected with the empty vector after 72 h (Fig. 3A), further supporting the possibility that this cDNA encodes core 1 β3-Gal-T.

To demonstrate conclusively the cloned cDNA represents the structural gene for core 1 β3-Gal-T rather than an activator or co-factor, the cDNA was modified to direct the synthesis of an epitope-tagged protein. Plasmid phCore1-SF was engineered to modify the amino terminus of core 1 β3-Gal-T with the 12-amino acid epitope for the monoclonal antibody HPC4 (23, 24). Incubation of 293T cell media following transfection with phCore1-SF with HPC4-UltraLink resulted in capture of the core 1 β3-Gal-T activity on beads containing HPC4 and depletion of...
the activity from the media (Fig. 3B). Direct assay of the HPC4-UltraLink for core 1 \( \beta \)-Gal-T activity demonstrated the enzyme was active when bound to the antibody.

To confirm the isolation, purity, and size of the recombinant core 1 \( \beta \)-Gal-T containing the HPC4 epitope was expressed in human 293T cells, wild-type CHO cells (CHO-K-1), and a mutant CHO cell line (CHO Lec1), which lacks N-acetylglucosaminyltransferase I and therefore exclusively synthesizes high mannos-type N-glycan and no complex- or hybrid-type N-glycans containing sialic acid or galactose. The HPC4-tagged recombinant protein in media from these cells was analyzed by Western blot (Fig. 4A). The enzyme from all three cell lines migrated as a defined band at \( \sim 39 \) kDa (Fig. 4B), which is in agreement with the predicted size of the truncated, soluble protein containing an HPC4 epitope with the calculated size of 38,726.7 daltons. The sequence of the vertebrate core 1 \( \beta \)-Gal-T predicts an absence of N-glycosylation sequons. Thus, the data in Fig. 4A confirm the absence of N-glycosylation on the protein and suggest that there are few, if any, other types of post-translational modifications affecting the size of the recombinant protein. It should be noted that the results shown in Fig. 4A represent the size of the recombinant, soluble core 1 \( \beta \)-Gal-T in reducing SDS-PAGE. In non-reducing SDS-PAGE the recombinant protein behaved as a species of \( \sim 80 \) kDa, consistent with its occurrence as a disulfide-bonded dimer (data not shown).

Because this recombinant is truncated and therefore lacks the transmembrane domain and the two Cys residues in the transmembrane domain, the results suggest that disulfide-based dimerization of the recombinant protein does not require those Cys residues.

**Fig. 1.** cDNA and deduced protein sequence of the human core 1 \( \beta \)-Gal-T. A, the nucleotide and deduced amino acid sequence of human core 1 \( \beta \)-Gal-T is shown. The putative transmembrane domain is singly underlined. The portions of the sequence that correspond to the identified amino-terminal and internal tryptic peptide of the purified rat liver core 1 \( \beta \)-Gal-T (33) are indicated by the double underlining with two changes Asn → Asp and Gln → Glu. (The sequence of rat protein in Fig. 7 has the exact sequence of these peptides.) B, hydrophilicity plot of human core 1 \( \beta \)-Gal-T. Hydrophilicity, as determined by the method of Kyte and Doolittle (21), is plotted versus the amino acid sequence of human core 1 \( \beta \)-Gal-T. A window size of 7 was used.

**Fig. 2.** Expression of wild-type core 1 \( \beta \)-Gal-T. Human 293T cells were transfected with an expression plasmid containing the wild-type full-length core 1 \( \beta \)-Gal-T cDNA or a control vector. Cells were harvested at 24, 48, and 72 h and assayed for the expression of core 1 \( \beta \)-Gal-T activity.
Identification of the Product of HPC4 Epitope-tagged Core 1 3-Gal-T

To demonstrate that epitope-tagged core 1 3-Gal-T synthesizes core 1 O-glycan, a PSGL-1 glycopeptide modeled after the extreme amino terminus of the human P-selectin glycoprotein ligand (PSGL-1) designated 4-GP-1 was utilized. The glycopeptide 4-GP-1 has the sequence Glu-Tyr-Glu-Tyr-Leu-Asp-Tyr-Asp-Phe-Leu-Pro-Glu-(GalNAc/H9251-H9252)4-Thr)-Glu-Pro-Pro-Glu-Met with a single GalNAc-substituted Thr, as indicated. 4-GP-1 was incubated with HPC4 beads captured epitope-tagged core 1 3-Gal-T and UDP-[3H]Gal in the absence of Triton X-100. 4-GP-1 was 3H-galactosylated 100% by the captured core 1 3-Gal-T, and activity was equivalent to that obtained with the purified rat liver core 1 3-Gal-T (Fig. 5). After treatment of the products of both enzyme preparations with O-glycanase, the product was quantitatively cleaved from the peptide, and only background radioactivity was re-bound to the C18 column (Fig. 5). These results demonstrate that the O-glycan products generated by the HPC4 epitope-tagged core 1 3-Gal-T and the purified rat liver core 1 3-Gal-T were sensitive to O-glycanase and represent the disaccharide Galβ1–3GalNAc (Fig. 5).

Expression Pattern of Core 1 3-Gal-T in Human Tissues

To determine the tissues in which core 1 3-Gal-T is expressed, a multiple tissue Northern blot was probed with a random hexamer-labeled 147-bp probe derived from the core 1 3-Gal-T cDNA. Two different transcripts of 7.0 and 2.0 kb...
were identified, with the 2.0-kb transcript predominating (Fig. 6A). Expression was especially high in heart, kidney, liver, and placenta, moderate in brain and skeletal muscle, and relatively low in other tissues examined.

The 2.0-kb message agrees well with the predicted size of a polyadenylated 1794-bp transcript. The 7.0-kb message could represent an alternatively spliced message or a partially spliced polyadenylated species. High stringency washing failed to displace the probe with the 7.0-kb transcript, strongly suggesting this message contains the same sequences (not shown).

**Identification of Putative Core 1 β3-Gal-T Sequences in Other Species**—Putative core 1 β3-Gal-T sequences from rat (Rattus norvegicus), mouse (Mus musculus), fruit fly (D. melanogaster), and nematode (C. elegans) were identified by data base searching.

The 930-bp rat EST clone AI059600 identified above was sequenced and found to be incomplete at the 3′ end. The remaining 3′ sequence was cloned by nested 3′-RACE as described under “Experimental Procedures” (25). The sequence predicts a protein of 363 amino acids that is 89.8% identical to the human sequence (Fig. 7). (The rat mRNA sequence has been deposited in GenBank™ as accession number AF157963.)

The nematode genome was searched by BlastP using the C38H2.2 gene product, C38H2.2, with homology to the human gene (Fig. 7). To confirm that the predicted C. elegans gene is actually transcribed, two separate PCRs were performed for the 5′- and 3′-halves of the predicted cDNA using C. elegans cDNA library of mixed developmental stages as a template. As described in detail under “Experimental Procedures,” the sequence of the PCR product from the 3′-half (744 bp) perfectly matched the predicted sequence from the gene C38H2.2. By contrast, the PCR product from 5′-half of the cDNA generated...
a product with an unexpected size of ~650 bp instead of the expected 704 bp. However, the 3' portion (520 bp) of the 650-bp product did match perfectly within the expected 704-bp product, but the remaining 130 bp did not match with any sequence.

We then used the matching 520-bp sequence for BlastN searching in the data base, and we identified a *C. elegans* EST that contained a sequence corresponding to correct 5' end of the cDNA. With this information PCR was then performed on the *C. elegans* cDNA library as a template using 5' primer and 3' primer corresponding to the predicted full-length cDNA with 1,170 bp that include the stop TAA. The sequence of the PCR product was obtained, and the translated sequence is shown in FIG. 7.

**FIG. 7.** Comparison of human, rat, mouse, fly, and nematode core 1 β3-Gal-Ts. ClustalW alignment of the putative protein sequences. Identical residues are indicated by boxes. Gaps in the sequences are indicated by hyphens. Sequences were identified as described under “Experimental Procedures.”
Fig. 7. The presence of this sequence in a cDNA library strongly suggests the predicted gene is actually expressed. The C. elegans core 1 β3-Gal-T is predicted to contain 389 amino acids (Fig. 7). Overall, the protein is 42.7% identical to the human protein. (The C. elegans mRNA sequence has been deposited in GenBank™ as accession number AF269063.)

Homologous sequences from D. melanogaster were identified by data base searching, but they were not physically cloned. Two highly homologous Drosophila sequences were identified by data base searching. The Drosophila core 1 β3-Gal-T was identified by BlastP searching using the human protein sequence. A predicted protein of 388 amino acids contained in gbaAF52723.1 was identified that was 43% identical to the human protein sequence. A second homologous Drosophila sequence was similarly identified as the CG8708 gene product (gbAF59121.1). Analysis of the predicted protein suggested the carboxyl-terminal portion of this protein was incorrectly predicted. The correct sequence was identified by searching the genomic sequence. This sequence predicted a protein of 366 amino acids that is 41.3% identical to the human sequence. The two Drosophila sequences are very similar at the amino terminus but are significantly different at the carboxy terminal, suggesting that the two proteins may differ in the catalytic activity or substrate interactions.

Genomic Structure of Human Core 1 β3-Gal-T—The cDNA sequence of human core 1 β3-Gal-T was used for a BlastNr search (non-redundant) which identified the Homo sapiens PAC (P-1-derived artificial chromosome) clone DJ0733B09 from 7p14-p13 (GenBank™ accession number AC005532). The sequence of this clone predicts a gene containing 115,954 bp. Analysis of its sequence compared with the cDNA sequences of human core 1 β3-Gal-T shows that the human core 1 β3-Gal-T gene contains 3 exons (Fig. 8). Exon I is derived from 64,041 to 64,260 with a size of 220 bp. Exon II is from 67,976 to 68,643 with a size of 668 bp. Exon III is from 73,245 to 73,448 with a size of 204 bp.

DISCUSSION

In the present study, we used the amino acid sequence of the purified rat liver core 1 β3-Gal-T to clone and express the cDNA encoding human core 1 β3-Gal-T. Core 1 β3-Gal-T is a key glycosyltransferase catalyzing the formation of the core 1 structure Galβ1→3GalNAc1-R, which can be further elongated to form complex O-glycans. The deduced primary structure of the core 1 β3-Gal-T predicts a type II transmembrane topology similar to previously described glycosyltransferases. Expression of an epitope-tagged version lacking the putative transmembrane domain generated a soluble epitope-tagged enzyme with a size nearly identical to that predicted by the sequence of the cDNA, confirming the type II topology.

To confirm that the gene identified as encoding the human core 1 β3-Gal-T actually encodes a catalytically active enzyme, we assayed the full-length and HPC4 epitope-tagged soluble forms of the enzyme for core 1 β3-Gal-T activity. Treatment of the reaction products with O-glycosidase, which is specific for the core 1 disaccharide structure, quantitatively released the disaccharide. These results demonstrate that the HPC4 epitope-tagged soluble form of the enzyme is active and generates the expected product.

Interestingly, the human core 1 β3-Gal-T lacks N-glycosyla-
AJ132443). Both the human core 1 β3-Gal-T described here and the β3-Gal-T8 are encoded by a gene located at 7p14-p13. However, the β3-Gal-T8 gene is predicted to contain only two exons, whereas the human core 1 β3-Gal-T gene contains three exons. This difference may arise because the β3-Gal-T8 sequence described to date is incomplete at the 3’ end, and some sequence reported for the protein is predicted from what we have determined to be derived from an intron between exons 2 and 3. The putative β3-Gal-T8 mRNA is predicted to encode a protein containing only 309 amino acids with 4 Cys residues in the luminal domain. The core 1 β3-Gal-T, by contrast, contains 363 amino acids with 6 Cys residues in the luminal domain. The uncertainty regarding the 3’ sequence of the β3-Gal-T8 probably led to the conclusion that the gene has only two exons. However, it is likely that the β3-Gal-T8 gene and the gene we describe to encode the core 1 β3-Gal-T are the same genes.

Comparison of the human core 1 β3-Gal-T sequence with that of the previously described (30) β3-galactosyltransferase family demonstrates no overall apparent homology. However, sequences similar to four of the six identified conserved motifs in the β3-galactosyltransferase family can be identified in the core 1 β3-Gal-T sequence, namely the VEKTW, DXDXF, GXYGV(1)XS, and DLXXG motifs (Fig. 7). All eight of the previously described β3-galactosyltransferases each contain all six of these motifs. Interestingly, the core 1 β3-Gal-T sequence does not share the Cys residues that are found in all members of the β3-galactosyltransferase family. No homology with the seven-member β4-galactosyltransferase family could be demonstrated. These results suggest that the core 1 β3-Gal-T identified in this study may be distantly related to the β3-galactosyltransferase family.

In marked contrast to the previously described β-galactosyltransferase families encompassing many highly homologous genes, only a single human core 1 β3-Gal-T was identified by sequence homology searching. Given the completeness of the data bases, it seems unlikely an additional human homolog of the core 1 β3-Gal-T will be identified.

It has been postulated that Tn syndrome and IgA nephropathy result from a stem cell mutation leading to deficiency of core 1 β3-Gal-T (6–11). These diseases could result from mutations in the structural gene for the core 1 β3-Gal-T or could be due to repression of the gene expression (11). Cloned T-cells derived from patients with Tn syndrome lack core 1 β3-Gal-T, but activity could be elevated by treatment of cells with 5-azacytidine or sodium n-butyrate, both inducers of gene expression (11). IgA from patients with IgA nephropathy are deficient in the core 1 O-glycan structure in the hinge region, and expression of Tn antigen (31) and core 1 β3-Gal-T is reduced ~30% in lymphocytes from patients with IgA nephropathy (9). However, the hinge region of the IgA may be underglycosylated through lack of the core N-acetylglactosamine, and this may also contribute to antigenic recognition of the altered IgA (32).

The existence of Tn syndrome and IgA nephropathy, both of which have been proposed to result from somatic mutations in core 1 β3-Gal-T, is consistent with the hypothesis that only a single gene encoding the core 1 β3-Gal-T is found in humans. Further studies, now in progress, will be required to determine whether mutations in the structural gene encoding the core 1 β3-Gal-T or dysregulation of the expression and/or subcellular localization of the enzyme are responsible for Tn syndrome or IgA nephropathy.

Acknowledgments—We thank the Molecular Biology Core Facility of the W. K. Warren Medical Research Foundation for protein sequencing and oligonucleotide synthesis, the DNA sequencing facility at the Oklahoma Medical Research Foundation, and Michel Arcade for manuscript preparation.

REFERENCES
Cloning and Expression of Human Core 1 β1,3-Galactosyltransferase
Tongzhong Ju, Kevin Brewer, Anil D'Souza, Richard D. Cummings and William M. Canfield

doi: 10.1074/jbc.M109060200 originally published online October 24, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M109060200

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

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

This article cites 30 references, 11 of which can be accessed free at http://www.jbc.org/content/277/1/178.full.html#ref-list-1