

Comparative Analysis of Retroviral and Native Promoters Driving Expression of β 1,3-Galactosyltransferase β 3Gal-T5 in Human and Mouse Tissues*

Received for publication, July 13, 2006, and in revised form, November 15, 2006. Published, JBC Papers in Press, November 15, 2006, DOI 10.1074/jbc.M606666200

Lydia Mare¹ and Marco Trinchera²

From the Department of Biomedical Sciences Experimental and Clinical, University of Insubria Medical School, 21100 Varese, Italy

β 1,3-Galactosyltransferase β 3Gal-T5 is highly expressed in the colons of humans and certain primates due to a retroviral long terminal repeat (LTR) acting as a strong promoter. Because this promoter is inactive in other human tissues or mice, we attempted to understand how adoption of a retrotransposon allowed the gene to acquire tissue-specific expression. We identified three novel 5'-UTRs of β 3Gal-T5 mRNA, types A, B, and C, and found widespread expression of the type A transcript at much lower levels than the LTR transcript, the expression of which is restricted to organs of the gastrointestinal tract. Expression of the type C 5'-UTR transcript was mostly restricted to the ileum, where it was expressed at high levels. We cloned the 5'-flanking regions of both types A and B 5'-UTRs, found deletion constructs functionally active as promoters, and identified CCAAT-binding factor (CBF) and hepatocyte nuclear factor 1 (HNF-1) as the principal nuclear factors controlling the promoters of types A and B 5'-UTR transcripts, respectively. The CCAAT-binding factor binding site and the entire downstream sequence driving the expression of type A transcripts in humans are structurally and functionally conserved in mice, where they constitute a unique β 3Gal-T5 promoter that appears to be the ancestral promoter of the gene. The HNF-1 binding motif of the second human promoter is identical to the HNF-1/Cdx binding motif of the LTR promoter but is in the antisense orientation, resulting in much lower binding affinity and promoter strength. These data may explain the successful insertion of the transposon during evolution.

β 1,3-Galactosyltransferase 5 (β 3Gal-T5)³ is a late-acting glycosyltransferase responsible for the synthesis of type 1 chain carbohydrates, including Lewis antigens, in human epithelial

cells (1, 2). β 3Gal-T5 is highly expressed in human colon mucosa, down-regulated in colon cancers (3), and large amounts of its reaction products are found in the bile and small intestine (4, 5). In some colon cancer cells, the first exon (exon 1) in the 5'-UTR of the β 3Gal-T5 transcript is under the control of a promoter located \sim 150 bp upstream and regulated primarily by the transcription factors HNF-1 and Cdx (6). Interestingly, the entire exon 1, as well as the HNF-1/Cdx binding motif, belongs to a retroviral long terminal repeat (LTR) sequence that is the dominant promoter for the β 3Gal-T5 gene in the colon but is inactive in other tissues where β 3Gal-T5 is expressed (7).

Transposable elements within eukaryotic genomes from plants (8) to mammals (9) greatly affect gene expression both at the protein (10) and regulatory sequence (11) levels. The most common transposable elements in the human genome are of retroviral origin (retroelements). Retroelements often directly donate transcriptional regulatory signals that may be either excluded from some genes or taken up by others. The former group probably includes highly conserved genes with basic functions, whereas the latter likely includes gene classes that have recently expanded, such as those involved in cellular responses to external stimuli (12). The glycosylation machinery includes gene products from both classes; core enzymes probably belong to the former class (13), whereas terminal glycosyltransferases to the latter (14). In addition to β 3Gal-T5, a retroviral LTR has been detected in the 5'-flanking region of two other glycosyltransferases, a fucosyltransferase and a sialyltransferase (12), but little is known about the promoter activity.

With respect to β 3Gal-T5, the same investigators found an alternative 5'-UTR that is expressed in brain whose first exon, exon -3, is located $>$ 100 kb upstream of the coding sequence (7). This 5'-UTR also contains two additional exons, -2 and -1, not present in the LTR-derived 5'-UTR. More recently, they also demonstrated that the β 3Gal-T5 LTR promoter is active in certain non-human primates but not in the mouse genome, thus dating the transposon insertion to \sim 25–30 million years ago (15).

It is not clear whether the brain 5'-UTR transcript is tissue-specific, whether other 5'-UTR promoters for the β 3Gal-T5 gene exist in humans, whether the mouse ortholog of the gene has one or more 5'-UTRs, or whether the mouse ortholog shares promoters with the human transcripts. Assuming that the LTR sequence was active at the time of insertion (15), understanding these aspects would allow us to understand the

* This work was supported by grants from Ministero dell'Università e della Ricerca (PRIN 2004) and from the University of Insubria (FAR 2003–2005) (to M. T.). 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) DQ645732–DQ645738.

¹ Supported by a fellowship from the University of Insubria. Present address: DBCM, University of Lausanne, Rue du Bugnon 9, CH-1005 Lausanne, Switzerland.

² To whom correspondence should be addressed: DSBS, via JH Dunant 5, 21100, Varese, Italy. Tel.: 39-0332-397-160; Fax: 39-0332-397-119; E-mail: marco.trinchera@uninsubria.it.

³ The abbreviations used are: β 3Gal-T5, β 1,3-galactosyltransferase 5; CBF, CCAAT-binding factor; HNF-1, hepatocyte nuclear factor 1; LTR, long terminal repeat; RT, reverse transcription; RACE, rapid amplification of cDNA ends; EMSA, electrophoretic mobility shift assay.

Multiple β 3Gal-T5 Promoters

putative context of β 3Gal-T5 expression at the time of LTR insertion, which would thereby facilitate understanding of why the insertion was tolerated and what stabilized the mobile element. The present paper is aimed at identifying the native β 3Gal-T5 promoter(s) present in the human genome, with particular attention to that (if any) common to the mouse genome and to detect features in native promoter(s) that may explain stabilization of the retroviral promoter. To address these issues, we studied the 5'-UTR of β 3Gal-T5 transcripts through rapid amplification of cDNA ends (RACE) analysis of RNA isolated from different human sources and determined their relative expression in various tissues and cell lines using competitive reverse transcription (RT)-PCR. We then focused on the 5'-flanking region of novel 5'-UTR transcripts, trying to identify sequences active as promoters, using the luciferase reporter assay and then binding sites for transcription factors through electrophoretic mobility shift assay (EMSA). We also compared the sequences of the novel 5'-flanking regions with that flanking the mouse β 3Gal-T5 transcript, looking for possible homologies. Because we found a putative β 3Gal-T5 promoter apparently conserved in both species as well as a human sequence poorly active as a promoter but able to bind the same transcription factors as the LTR promoter, we studied such sequences in detail. The results suggest that we found the ancestral mammalian promoter of the gene as well as an apparently defective promoter probably related to the successful insertion of the transposable element in some primates.

EXPERIMENTAL PROCEDURES

Cell Lines, Tissues, and RNAs—Human breast cancer cell lines MCF-7 and MDA-MB-231 (a gift of Dr. Fabio Dall'Olio, University of Bologna, Italy) and the pancreas carcinoma cell line Capan-2 (American Type Culture Collection HTB 80) were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum, 100 units/ml penicillin, 1 mg/ml streptomycin, and 2 mM L-glutamine. Human bile duct carcinoma cells HuCC-T1 (Japanese Collection of Research Bioresources 0425) were cultured in RPMI 1640 medium containing the same supplements as above. Human colon adenocarcinoma cells Caco-2 and SW-1116 and human gastric carcinoma cells MKN-45 were cultured as previously described (3, 16). Human ileum and colon samples were collected at surgery and kindly donated by Dr. Filippo Mare (Hospital di Circolo, Varese, Italy). Total RNA from human tissues was obtained from Ambion (mammary gland), Stratagene (colon, stomach, duodenum, pancreas, mammary gland, thymus, and trachea), or Clontech (stomach, pancreas, and mammary gland). Mouse total RNA was purchased from Clontech. Total and poly(A)⁺ RNAs were prepared from human tissues or cell lines as previously reported (3, 16).

5'-RACE—RNA ligase-mediated RACE was performed essentially according to the manufacturer's instructions with the FirstChoice RLM RACE kit (Ambion). Each reaction, containing RNase inhibitor, was treated with phenol/chloroform and then mixed with 1 μ l of 20 mg/ml glycogen (Fermentas), and the RNA was precipitated with sodium acetate/ethanol. Five μ g of total RNA or 0.125 μ g of poly(A)⁺ RNA was used as the starting material. A tobacco acid phosphatase reaction was

performed on one-half of the calf intestinal phosphatase-treated RNA, and the whole amount was ligated to an RNA adaptor. One-fourth of the obtained material was then used for individual RTs in a final volume of 5.0 μ l using either avian myeloblastosis virus or Moloney murine leukemia virus reverse transcriptase (GE Healthcare). Moloney murine leukemia virus RTs were incubated 20 min at 37 °C, 40 min at 42 °C, and 20 min at 45 °C. Avian myeloblastosis virus RTs were incubated 40 min at 42 °C, 20 min at 45 °C, and 20 min at 50 °C. The coding sequence-specific RT primer was 5'-AGGAGGAAGAAT-GTC-3', and the exon -1-specific RT primer was 5'-CTTAC-TAGCCAATAAC-3'. Amplifications (35 cycles) were performed in 25 μ l of LA *Taq* (Takara) according to the manufacturer's recommendations, with 4 μ l of the RTs as the template. Nested PCRs (30 cycles) were performed using 0.4–2.0 μ l of the first amplifications and inner primers designed to contain restriction sites. Amplified fragments were column-cleaned, digested with appropriate restriction enzymes, gel-isolated, and cloned into pGI3 vector for sequencing.

RT-PCR—Competitive RT-PCR was performed as previously described (3, 16). The human and mouse β -actin and human β 3Gal-T5 coding sequence competitors and oligonucleotide primers used were those previously described (3). To quantify the different β 3Gal-T5 5'-UTRs, a competitor was prepared by subcloning the entire sequence of the short splicing variant of the type A 5'-UTR in the HindIII-XhoI sites of pCDM8 vector. Subsequently, a double-stranded oligonucleotide (see below) with NdeI/HindIII ends was ligated into the corresponding cohesive ends of the construct. The oligonucleotide NdeI/HindIII ends are underlined; also shown are primer sequences for exons 1C (bold) and 1B (italic): 5'-TATGAAGGGCATTGGAGACCCAGGGACCGCTGCCCATTACAGTGACTCACA (upper strand) and 5'-AGCTTGTGAGTCACGTAATGGGCATGGCGGTCCCTGGGTCTCCAATGCCCTTCA (lower strand). The resultant construct was truncated by PmlI digestion and self-ligated. The LTR competitor was prepared by PmlI digestion and self-ligation of a plasmid with the LTR transcript cloned in the pCDM8 vector. Other exon-specific upper strand primers were 5'-CTGCCGGGGCTGCCCCGCGC (exon 1A), 5'-CCTGCCCTTGGACATCAGAGCTGC (exon 1), and 5'-CAAACCAGAGGTTCCTCTTACCAGC (exon 3); the lower strand primer for the coding sequence was 5'-GAATGGTGGGTACCTGTCCCACGG. Mouse β 3GalT5 cDNA was obtained by PCR and cloned in the pCDM8 vector using an upper strand primer that annealed to the mouse first exon (5'-CGTCCCCTGCCCTGTCCCTTCTC) and a lower strand primer that annealed to the end of the coding sequence (5'-GGTCTAGGTGCCCCAGGGTCC). The upper strand primer for the mouse ortholog of exon 3 was 5'-GTGGACCTCACTACCGGCTGC.

Luciferase Assay—A cDNA encompassing the 5'-flanking region of β 3Gal-T5 at positions -808 to +148 relative to the beginning of exon 1A was amplified by PCR (30 cycles) using LA *Taq* and GC II buffer (Takara) following the manufacturer's recommendations. The 25- μ l reaction contained 150 ng of human placenta genomic DNA as the template and primers

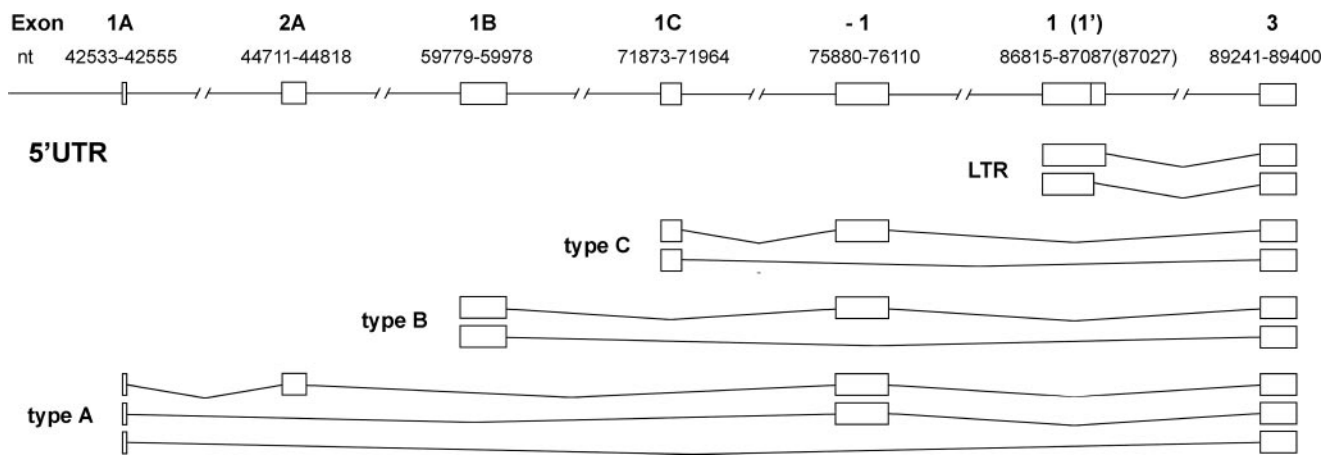


FIGURE 1. **Genomic structure of the human β Gal-T5 gene.** The positions of the exons in the context of human chromosome 21 are shown following the nucleotide (nt) numeration of the available sequence (GenBank™ accession number AF064860). The structure of three novel 5'-UTRs of β Gal-T5 mRNA with their splice variants is presented together with the reported LTR-derived 5'-UTR.

having restriction sites at their 5'-ends. The resultant fragment was cloned in the corresponding sites of the vector pGL3 (Promega) upstream of the firefly luciferase coding sequence. Deletion constructs were prepared by PCR (25 cycles) under similar conditions with the proper primer pairs but using the plasmid pGL3-808+148 (5 ng/reaction) as the template. The 5'-flanking region of the mouse ortholog was prepared similarly using corresponding primers. Site-directed mutagenesis of the CCAAT-binding factor (CBF) sites was performed using *Pfu* Turbo DNA polymerase (Stratagene) under the conditions described in the QuikChange mutagenesis kit protocol. A cDNA encompassing the 5'-flanking region of β Gal-T5 from positions -1011 to +20 from the beginning of exon 1B as well as the cognate deletion constructs were obtained and cloned in pGL3 using the same approach but under standard LA *Taq* reaction conditions.

For transfection, host cells were plated 20 h in advance in 96-well plates with 50,000 cells/well in 0.1 ml of culture medium. Transfection solutions were prepared by mixing 300 ng of test DNA with 15 ng of the *Renilla* luciferase expression vector pRL-CMV (Promega) in 25 μ l of serum-free medium and then adding 25 μ l of serum-free medium containing 0.9 μ l of Lipofectamine (Invitrogen). Liposomes were allowed to form 20 min at room temperature. Cells were washed twice with serum-free medium, fed with transfection solution, and incubated under regular growing conditions for 3 h. After the addition of 0.1 ml of standard medium, the incubation was continued for 20 h. The cells were then washed with phosphate-buffered saline and lysed with 40 μ l of Passive Lysis buffer (Promega). The luciferase activity of lysate aliquots was assayed using the dual luciferase reporter system (Promega).

EMSA—EMSA were performed with the Lightshift chemiluminescence kit (Pierce) following the manufacturer's recommendations, but the binding reaction volume was scaled down to 10 μ l. In the case of G/C-rich sequences, the binding reactions contained 50 mM KCl, 5.0% glycerol, 5.0 ng/ μ l poly(dI·dC), and no detergent or MgCl₂. Nuclear extracts were prepared using the NE-PER extraction kit (Pierce), and 1 μ l of extract (3–4 μ g of protein) was added to each binding reaction.

Anti-HNF-1 (Santa Cruz Biotechnology catalog number sc-8986X) or anti-CBF-B (catalog number sc-10779X), at 1 μ l/reaction, was incubated 30 min on ice in the binding reaction before adding the probes. Competitor consensus sequences for CBF (catalog number sc-2591), Ets-1 (sc-2549), and Oct-1 (sc-2506) were double-stranded oligonucleotides whose sequences corresponded to those suggested by Santa Cruz Biotechnology, as indicated. HNF-1 competitor DNA was the 29-bp SIF3 sequence previously reported (17). DNA probes were biotinylated using the biotin 3'-end labeling Kit (Pierce). DNA/protein complexes were separated by 5% native PAGE, transferred to a nylon membrane, cross-linked under UV light, and detected according to the kit protocol.

RESULTS

Identification of Different 5'-UTRs in β Gal-T5 Transcripts—To identify novel 5'-UTRs, we first performed 5'-RACE analysis using ileum RNA and an RT primer specific for the coding sequence and found a single major band of ~650 bp. Sequencing of the cloned fragment revealed exon 3, exon -1, and a novel exon of 92 bp, named exon 1C that was located ~4 kb upstream of exon -1 in chromosome 21 (Fig. 1). RT-PCR analysis with an upper strand primer specific to this novel exon and a lower strand primer localized in the coding sequence yielded no amplification using either mammary gland or MKN-45 gastric carcinoma cell RNAs. We could, however, successfully amplify products in these reactions using an upper strand primer specific to exon -1 and the same lower strand primer. We then performed 5'-RACE analysis starting from these same RNAs using a primer that annealed to exon -1 for RT. We obtained two major bands of ~250 and ~350 bp from mammary gland RNA and two bands of ~250 and ~400 bp from MKN-45 RNA. Sequencing of the fragments cloned from the mammary gland revealed a short novel sequence of 23 bp (exon 1A) joined to exon -1 and a longer one having the same 23-bp sequence at the 5'-end and an additional 108-bp sequence (exon 2A) in between. In the genome, they indicate a transcription start site located in a G/C-rich region ~33 kb from exon-1 (Fig. 1).

Multiple β Gal-T5 Promoters

Sequencing of the fragments cloned from MKN-45 cells also indicated the presence of exon 1A and, in addition, a novel sequence of 200 bp (exon 1B) joined to exon -1, suggesting another transcription start site located ~16 kb upstream of exon -1 (Fig. 1). RT-PCR performed with primers specific to the novel exons and subsequent sequencing of the products enabled us to delineate the exon structure of the β Gal-T5 5'-UTRs (Fig. 1). Type A 5'-UTRs begin with exon 1A and comprise two main splice variants, a longer one that includes exon -1 and a shorter one without it. A third very minor type A variant that includes exon 2A is the longest and is almost undetectable in competitive experiments. Type B and C 5'-UTR transcripts begin with exon 1B and 1C, respectively, and both comprise the two main splice variants described above. Exon 3 is present at the 3'-end of all of the novel 5'-UTR-containing transcripts, as is true for the LTR- and brain-derived 5'-UTR mRNAs (1, 7).

Distribution of β Gal-T5 Transcripts with Different 5'-UTRs in Various Human Tissues and Cell Lines—We amplified β Gal-T5 starting from normalized cDNA prepared from different sources using a common lower strand primer that annealed to the coding sequence and different upper strand primers specific to exons 1 (LTR), 1A, 1B, 1C, -3, and 3. Amplification with the exon -3-specific primer yielded no products regardless of the cDNA (data not shown). The LTR-derived 5'-UTR mRNA predominated in the colon, duodenum, stomach, and pancreas, where the actual expression levels ranged from 0.3 to 10 fg/pg of β -actin (Fig. 2). These tissues also contained types A and B 5'-UTR transcripts but at much lower levels. On the other hand, type A 5'-UTR transcripts were the predominant β Gal-T5 mRNA in the mammary gland, thymus, and trachea, which lacked the LTR transcript. The actual expression levels of type A transcripts in such tissues ranged from 0.1 to 0.2 fg/pg of β -actin, suggesting that they may be under the control of a promoter weaker than the LTR promoter. The type C 5'-UTR-containing mRNA predominated in the ileum (15 fg/pg of β -actin) and was relevant in only one other tissue, the colon. By contrast, type B 5'-UTR transcripts were not the major species in any of the normal tissues tested. The actual expression levels of type B transcripts in any tissue were <0.15 fg/pg of β -actin, suggesting that they may be under the control of a very weak promoter. In some gastrointestinal cancer cell lines, type B transcripts were expressed at significant levels, and in the case of SW-1116 (colon), Capan-2 (pancreas), and HuCC-T1 (bile duct) cells, LTR transcript expression was low or absent. Analysis of additional independent RNA samples provided substantial sample-to-sample reproducibility; nevertheless, type C transcripts were present in one stomach and colon sample but not in the other (Fig. 2, boxed panel). Moreover, the total amount of β Gal-T5 transcript measured in the sample pairs varied from 20 to 150% (Fig. 2), as previously reported in human colon (3, 6).

Identification of Promoter Activity in the 5'-Flanking Regions of β Gal-T5 Transcripts with Types A or B 5'-UTRs—We focused on types A and B 5'-UTR transcripts because the tissue distribution and expression levels made them candidates to be under the control of native promoters. On the other side, no cell line is currently available as a model for further studies of type C

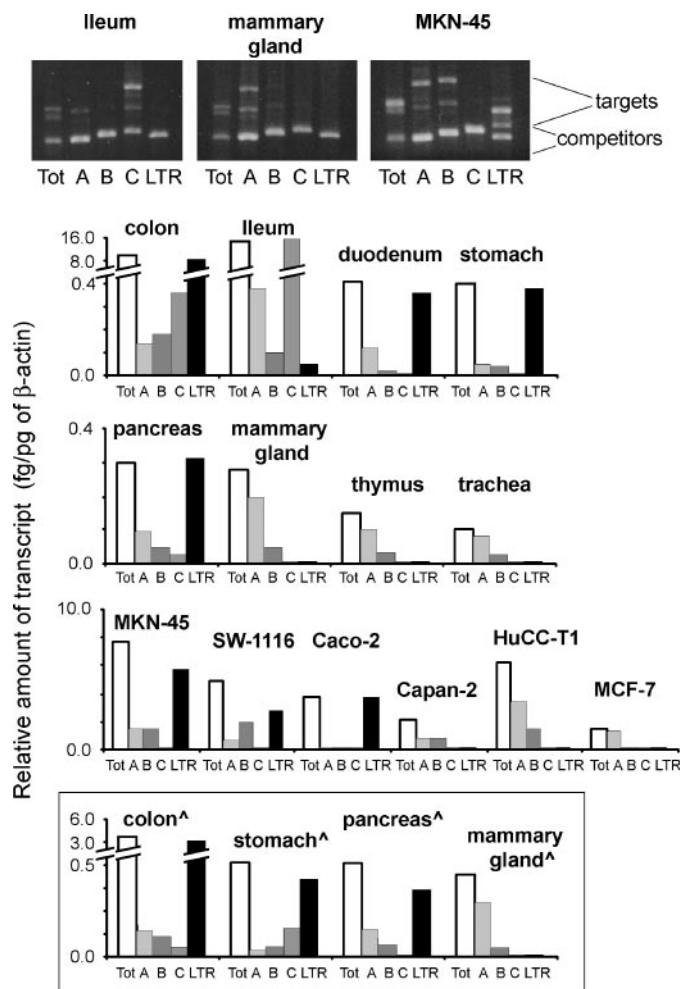


FIGURE 2. Quantitative expression of different β Gal-T5 5'-UTR transcripts in various human tissues and cell lines. RNA extracted from single tissue samples and cell lines was reverse-transcribed, and normalized amounts of the resultant first strand cDNA were mixed with competitor (truncated) cDNAs and subjected to PCR using primers specific to the following β Gal-T5 exons: 3 (total transcript, *Tot*), 1A, 1B, 1C, and 1 (*LTR*). Twenty percent of each amplification reaction was electrophoresed in a 1% agarose gel and visualized by ethidium bromide staining. Three relevant examples are shown in the upper panels. Target doublets are due to the splice variants indicated in Fig. 1. Quantification was performed by densitometric scanning of gel images. The amounts of amplified target cDNAs, presented in the bar graphs, were calculated from their respective standard curves and normalized to the amounts calculated for β -actin. The boxed panel shows the more relevant differences obtained analyzing independent RNA samples derived from the corresponding tissues (marked with a caret). The results are the means for two independent quantifications. Experimental variation was always <15% of the mean values.

transcripts. To identify such putative promoters, we cloned 800–1000-bp fragments of each 5'-flanking region in the luciferase reporter vector pGL3, prepared successive deletion mutants, and studied their ability to drive luciferase expression upon transfection into the appropriate cells. We used either MKN-45 cells, which express the LTR, types A and B 5'-UTR transcripts, or as a control, MDA-MB-231 cells, which do not express these transcripts. In the case of the 5'-flanking region of exon 1A (Fig. 3A), luciferase activity was evident when the whole fragment was assayed in MKN-45 cells (-808 to +20). Inclusion of the intron sequence did not effect this activity (-808 to +148), whereas cleavage of the first 47 bp upstream of the exon improved activity (-223 to -48). Activity was main-

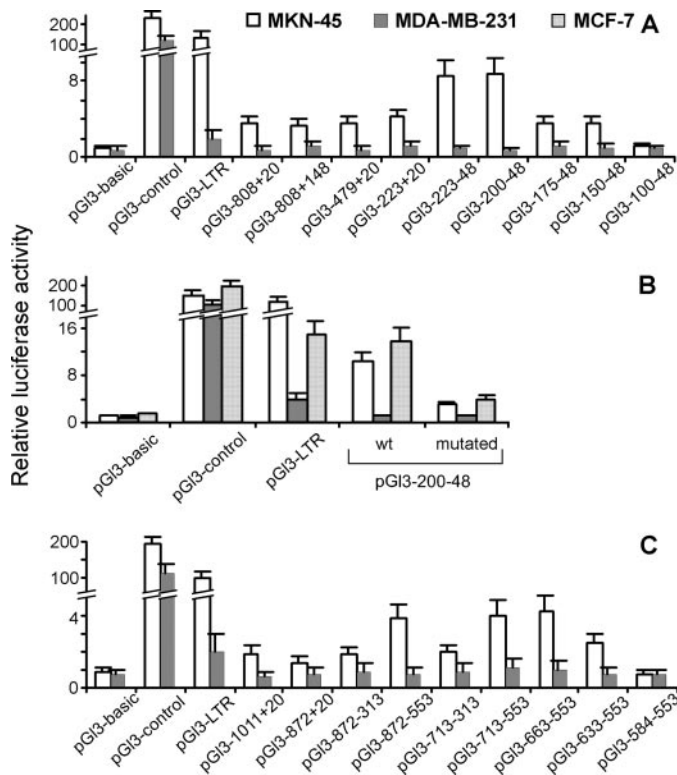


FIGURE 3. Promoter activity of the genomic sequences flanking exons 1A and 1B. The indicated DNA fragments flanking exon 1A (A and B) or 1B (C) were cloned in the pGI3 vector carrying the firefly luciferase gene and transfected in MKN-45 cells that express all three types of transcripts (types A and B both 1.5 fg/pg of β -actin; LTR 5.9 fg/pg of β -actin) or in MDA-MB-231 cells that express no β 3Gal-T5 transcripts together with a *Renilla* luciferase reporter expression plasmid. Firefly luciferase activity was measured 24 h later and expressed relative to the *Renilla* luciferase activity determined for each sample. pGI3-basic, lacking any promoter sequence, and pGI3-control, containing the SV-40 promoter, were used as negative and positive controls, respectively. pGI3-LTR, having the sequence -148 to -28 flanking exon 1 and known as the LTR promoter, is shown for comparison. The constructs representing the putative promoter sequence of type A 5'-UTR transcripts and its CBF site-directed mutant were also transfected in MCF-7 cells (B), a mammary gland cell line expressing mainly type A 5'-UTR transcripts (1.3 fg/pg of β -actin) and no detectable LTR transcripts. The mutated CBF sequence has CACGCGTCA instead of the wild type (wt) CACCTATCA. Values are the means \pm S.D. for at least two experiments in duplicate.

tained with fragment -200 to -48 (~ 9 -fold over background), was drastically reduced with fragments -175 to -48 and -150 to -48 (~ 3.5 -fold over background), and was lost with fragment -100 to -48 , suggesting that the sequences from -200 to -175 and from -150 to -100 may contain binding sites for relevant transcription factors. The highest luciferase activity measured was ~ 10 -fold lower than that of the LTR promoter, whereas the expression levels of type A transcripts in MKN-45 cells were only ~ 4 -fold less than that of LTR transcripts (Fig. 2). To further assess this aspect, we performed additional reporter assays in MCF-7 cells, which mainly express type A 5'-UTR transcripts and no detectable LTR transcripts, as shown in Fig. 2. The luciferase activity driven by the LTR promoter was found to be dramatically decreased in MCF-7 cells compared with MKN-45 cells, but was still ~ 11 -fold higher than the background, whereas the activity of the -200 to -48 fragment flanking the type A 5'-UTR was maintained at ~ 10 -fold over the background in these cells (Fig. 3B). These results suggest

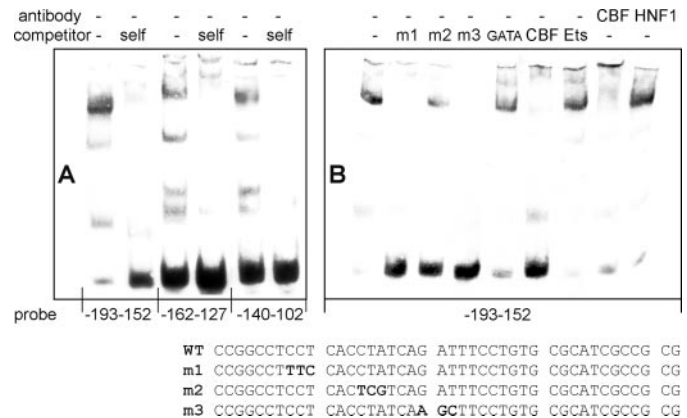


FIGURE 4. Characterization of the promoter driving type A β 3Gal-T5 transcripts. A, EMSAs were performed by incubating biotinylated oligonucleotide probes with MKN-45 nuclear protein extracts in the presence of a 200-fold molar excess of unlabeled oligonucleotides (competitor), and the resultant complexes were separated by 5% native PAGE and visualized by chemiluminescence. Oligonucleotide probe sequences are numbered from the transcription initiation site (+1). B, competitive EMSAs were performed with probe -193 to -152 in the presence of mutated oligonucleotides or consensus sequences of candidate transcription factors as competitors or specific antibodies. The mutated oligonucleotide sequences are shown at the bottom.

that the LTR construct *per se* is more highly active than the type A promoter in reporter systems.

For the type B 5'-UTR, the reporter activity of the whole fragment (-1011 to $+20$) was detectable but very low in MKN-45 cells (Fig. 3C). Cleavage of the 3'-end up to nucleotide -553 was found to be necessary to stimulate luciferase activity (compare -872 to $+20$ and -872 to -313 with -872 to -553 and also -772 to -313 with -772 to -553), suggesting the presence of an inhibitory region in this part of the sequence (see last paragraph). However, the most active fragment (-663 to -553) generated only ~ 4.5 -fold higher luciferase activity than the background, even though the expression levels of type B 5'-UTR transcripts are similar to those of type A in MKN-45 cells. Fragments that included nucleotides up to -633 at the 5'-end had activity, whereas those shorter did not, suggesting that the sequence from -633 to -553 may contain stimulatory binding regions.

Characterization of the Promoter-driving Expression of Type A 5'-UTR Transcripts—To characterize the candidate promoter of the type A 5'-UTR transcripts, we prepared four overlapping DNA probes, each one ~ 40 bp long, together representing the sequence from -215 to -102 from the transcription initiation site. We used these probes in EMSAs with nuclear extracts prepared from MKN-45 cells. A major complex was found using probe -193 to -152 , and other minor complexes were evident with probes -162 to -127 and -140 to -102 (Fig. 4A). Each complex appeared to be specific, because it disappeared in the presence of molar excess of the unlabeled probe. A search for potential transcription factor binding sites in the -193 to -152 sequence using the TRANSFAC data base suggested high affinity sites for transcription factors GATA-1 and Ets-1 and a lower affinity site for CBF. In competition EMSAs with three substitution mutants of the probe (Fig. 4B), one (m2), potentially affecting both GATA-1 and CBF binding, failed to compete for complex formation. The

Multiple $\beta 3\text{Gal-T5}$ Promoters

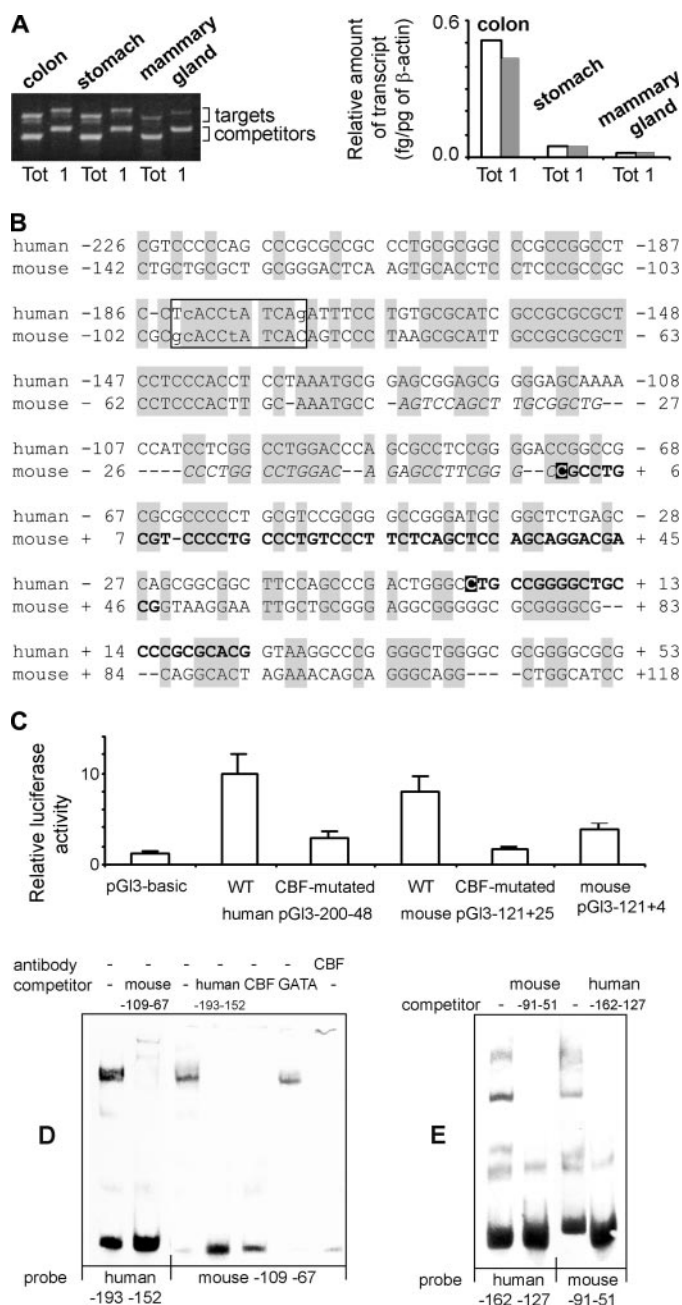


FIGURE 5. Expression of $\beta 3\text{Gal-T5}$ in mouse tissues is under the control of a promoter structurally and functionally similar to the human promoter driving type A transcripts. *A*, quantitative expression of total (*Tot*) $\beta 3\text{Gal-T5}$ transcript versus transcripts originating at exon 1 (1) was determined in three mouse tissues by competitive RT-PCR, as in Fig. 2. The amounts of amplified target cDNAs, presented in the *right panel*, were calculated from their respective standard curves and normalized to the amounts calculated for β -actin. *B*, ClustalW alignment of genomic regions flanking human type A and mouse $\beta 3\text{Gal-T5}$ transcription initiation sites. Human chromosome 21 sequence from GenBankTM accession number AF064860 (nucleotides 37,000–47,000) and mouse chromosome 16 sequence from GenBankTM accession number AC165170 (nucleotides 78,000–88,000) were used as input sequences. The first exons in the 5'-UTRs are bold, and translation initiation sites (+1) are shown in reverse type (white on black background). The mouse-extended first exon is italicized, the CBF binding site is boxed, and differences from the consensus sequence are in lower case. *C*, the mouse genomic sequence homologous to the human promoter of type A transcripts also contains a functional CBF binding site. The indicated DNA fragments flanking mouse exon 1 were cloned in the pGI3 vector, transfected in MKN-45 cells, and luciferase activity was measured and expressed as in Fig. 3. Mutated CBF sequences for both mouse and human have CACGCTCA instead of the wild type (WT)

CBF competitor prevented complex formation, whereas competitors with GATA or Ets family consensus sequences did not affect the complex. In addition, preincubation of the binding reaction with an antibody specific to CBF yielded a supershifted complex, indicating that CBF is involved in complex formation. On this basis, we prepared a CBF site-directed mutant of the pGI3-200-48 construct and found that the luciferase activity measured upon transfection in MKN-45 or MCF-7 cells is drastically reduced with respect to the wild type construct (Fig. 3*B*). Altogether, the data confirm that the sequence is actually a $\beta 3\text{Gal-T5}$ native promoter driving expression of the type A 5'-UTR transcripts.

Comparison of the Human and Mouse $\beta 3\text{Gal-T5}$ Genes—The 5'-UTR of the mouse $\beta 3\text{Gal-T5}$ transcript is reported to consist of two exons (GenBankTM accession number NM_033149). The first one, rich in G/C, is at least 47 bp long and may contain an additional 43 bp at the 5'-end (15). The second exon is 177 bp and appears homologous to the human exon 3. By competitive RT-PCR, we found that the amount of transcript originating at exon 1 corresponds to the total amount of transcript expressed in mouse colon, stomach, and mammary gland, suggesting that only one promoter is used in mouse tissues (Fig. 5*A*). The relative expression level in the colon was much higher than in the stomach or the mammary gland, as found in human tissues, but the actual amount, relative to β -actin, was much lower than in the human colon. We compared 10 kb of mouse genomic sequence from chromosome 16 (18), ~5 kb upstream and 5 kb downstream from the transcription initiation, with the corresponding human sequences flanking the three novel transcription initiation sites deduced from exons 1A, 1B, and 1C. No significant alignment was found in the latter two cases, whereas a surprisingly high level of homology was found between the mouse sequence and human sequence flanking exon 1A, mostly in the region just upstream of the exon (Fig. 5*B*). In particular, the CBF binding site and the downstream sequence responsible for the other complexes detected by EMSA, are almost completely conserved. To assess whether they are functional, we cloned the sequence in the pGI3 vector and transfected the construct in MKN-45 cells. Constructs with mutated CBF sites were also transfected as controls. The luciferase activities measured for the human and mouse sequences were very similar and were similarly reduced by the CBF mutation (Fig. 5*C*). Moreover, in competitive EMSAs, the mouse sequence containing the putative CBF site competed with the corresponding human probe, and when used as a probe itself, generated an identical complex that was competed out by the human sequence, the CBF consensus sequence, or anti-CBF (Fig. 5*D*). Similarly, the downstream sequence of the mouse promoter prevented complex formation with the downstream sequence of the human promoter. When labeled and used as a

CACCTATCA. pGI3-200-48, containing the human promoter driving type A transcripts, was mutated identically and used as a control. pGI3-basic, lacking any promoter sequence, is shown as a reference. *D*, competitive EMSAs were performed with human probe -193 to -152 in the presence of mouse oligonucleotide -109 to -67 and vice versa in the presence or absence of the CBF consensus sequence or GATA consensus sequence or anti-CBF. *E*, competitive EMSAs were performed with human probe -162 to -127 in the presence of mouse oligonucleotide -91 to -51 and vice versa.

probe, the mouse sequence generated three complexes that were apparently identical to the human complexes and whose formation was prevented by a molar excess of the human sequence (Fig. 5E). These data indicate that the promoter driving the mouse and the human type A transcripts is conserved through evolution, suggesting that it is the ancestral β Gal-T5 promoter.

Characterization of the Promoter Driving Expression of Type B 5'-UTR Transcripts—Because the type A promoter of β Gal-T5 has no apparent features that can explain the insertion and stabilization of the LTR promoter, we tried to characterize the putative promoter region of the type B 5'-UTR transcripts of the gene, despite its low levels of activity in luciferase reporter experiments. To this end, EMSAs were performed using three overlapping DNA probes covering nucleotides -643 to -500 from the origin of exon 1B. A major complex and a very minor one were detected with probe -614 to -574 (Fig. 6A); both were inhibited by a molar excess of the probe and by the shorter sequence -606 to -582 (Fig. 6B; WT, wild type). Among six substitution mutants of fragment -606 to -582 , m5 failed to inhibit both complexes and m6 failed to inhibit the minor one only (Fig. 6B). A TRANSFAC data base search using both the parental and mutated sequences suggested the presence of recognition sites for transcription factors Oct-1 and HNF-1. In competitive EMSAs (Fig. 6C), formation of the major complex was prevented by the addition of the HNF-1 binding sequence of the SIF3 oligonucleotide as well as by an antibody specific to HNF-1, indicating the presence of an HNF-1 binding site, as reported for the LTR promoter (6, 7). To directly compare these two promoters, competitive EMSAs were performed using both sequences (Fig. 6D). The LTR sequence completely abolished both the HNF-1 complex and the minor complex, suggesting that the binding is due to a Cdx binding site (6). Identical complexes were found with the LTR probe, but only the major complex was inhibited by a 200-fold molar excess of the -614 to -574 sequence, suggesting a lower affinity of the -614 to -574 sequence for the minor complex. By varying the competitor concentrations, we found that the relative affinity of the -614 to -574 fragment for the HNF-1 site of SIF3 was also much lower than that of the LTR sequence (Fig. 6E). A comparison of these two sequences (Fig. 6F) further revealed a string of nine identical nucleotides that includes the HNF-1 binding site and the overlapping sequence suggested to be a Cdx binding site (6). This same sequence is also present in the chimpanzee ortholog located in chromosome 22 (19). In the SIF3 and LTR sequences, the HNF-1/Cdx binding sequence is found in the sense orientation, whereas in both the -614 to -572 fragment and in the chimpanzee ortholog, this element is present in the antisense orientation. Taken together, our present data support the hypothesis that the sequence represents the promoter for the type B 5'-UTR transcripts, suggesting that the low luciferase activity associated with this fragment in our reporter experiments may be due to technical constraints probably related to the truncation of the 3'-region. To evaluate the possibility that this 3'-region may bind inhibitory factors, as suggested by luciferase experiments, we prepared five overlapping DNA probes of ~ 40 bp in length covering the sequence from -494 to -314 from the

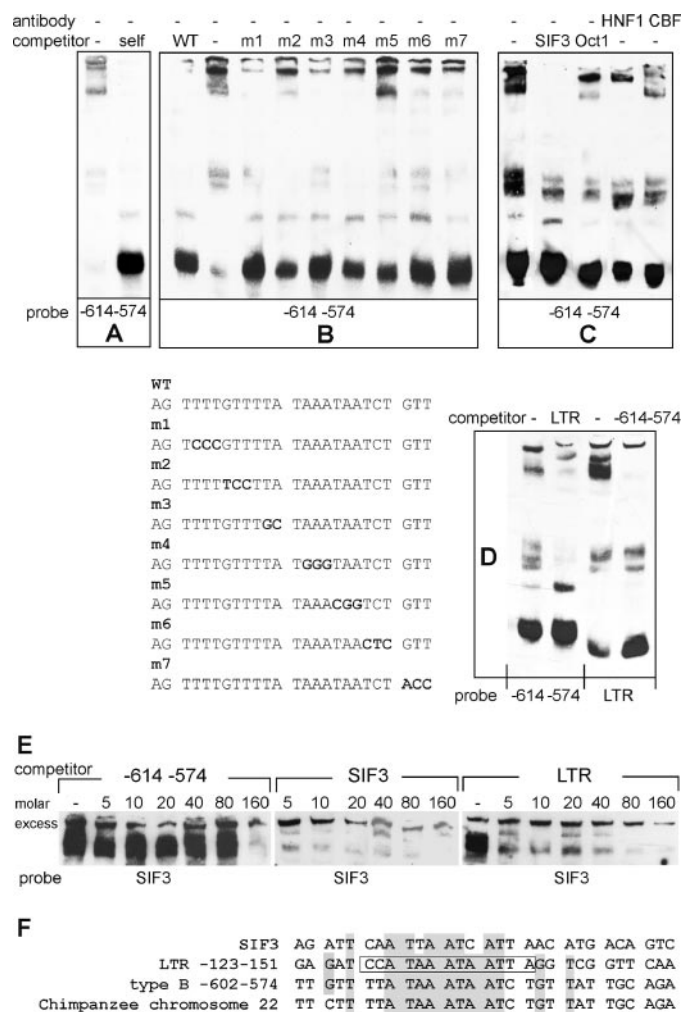


FIGURE 6. Characterization of the promoter driving type B β Gal-T5 transcripts and comparison with the LTR promoter. A, EMSAs were performed as described in the legend to Fig. 4. B, competitive EMSAs were performed with probe -614 to -574 in the presence of oligonucleotide -606 to -582 and its mutants. The bands at the top of the gel are unspecific. Mutated oligonucleotide sequences are shown at the bottom of the panel. C, competitive EMSAs were performed with probe -614 to -574 in the presence of consensus sequences of candidate transcription factors or specific antibodies. SIF3 is an HNF-1-binding oligonucleotide. D, competitive EMSAs were performed with probe -614 to -574 in the presence of the LTR oligonucleotide or vice versa. E, competitive EMSAs were performed with SIF3 as the probe in the presence of different concentrations of oligonucleotides -614 to -574 , SIF3, or LTR. F, comparison of the SIF3 oligonucleotide sequence (antisense), the corresponding sequence of the LTR (antisense), the corresponding sequence of the -614 to -574 oligonucleotide (sense), and the corresponding sequence of the chimpanzee ortholog (sense) (GenBank™ accession number AL954227, nucleotides 55926–55954). The HNF-1/Cdx binding motif of the LTR sequence (see Ref. 6) is boxed.

transcription initiation site. EMSA performed with these probes and nuclear extracts of MKN-45 cells provided many specific spots, including a major complex with probe -353 to -314 (Fig. 7A). The addition of a molar excess of the -353 to -330 fragment did not inhibit this complex efficiently but only another minor complex formed with probe -353 to -314 (Fig. 7B). The major complex could be competed out using fragment -337 to -314 , and substitution mutations of this fragment affected competition (Fig. 7B). The data indicate that the 3'-region of the putative type B promoter specifically binds putative regulatory factors in MKN-45 cells.

Multiple β 3Gal-T5 Promoters

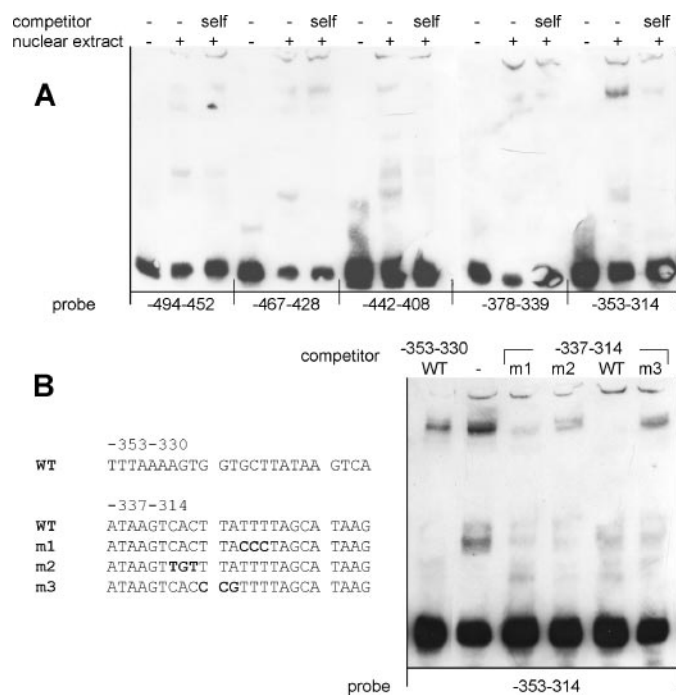


FIGURE 7. Binding of putative regulatory factors to the 3'-region of the type B promoter. *A*, EMSAs were performed as in Fig. 4. *B*, competitive EMSAs were performed with probe -353 to -314 in the presence of oligonucleotide -353 to -330 or oligonucleotide -337 to -314 and its mutants. Mutated oligonucleotide sequences are shown at the left of the panel.

DISCUSSION

We find from our current data that multiple promoters are present in the human β 3Gal-T5 gene. Our quantitative analysis of the expression of four 5'-UTR transcripts suggests that they account for the large majority of the β 3Gal-T5 mRNA molecules expressed in the tissues that we studied. However, because the total β 3Gal-T5 mRNAs detected is not always equal to the sum of the individual 5'-UTR transcripts and also because of inherent experimental error in these analyses, we cannot rule out the possibility that additional 5'-UTR transcripts are present in some tissues driven by alternative cognate promoters.

The LTR and type C promoters, both active in the gastrointestinal tract, appear as strong promoters; on the other side, the ubiquitous type B promoter and the brain-specific promoter appear to be very weak promoters, whose physiological relevance is uncertain. The type A promoter, which is also quite weak, is indeed responsible for β 3Gal-T5 expression in various tissues, such as the mammary gland, thymus, and trachea. This type A promoter is mostly responsive to CBF stimulation and is conserved in mouse, where it acts as a unique β 3Gal-T5 promoter. This ancestral promoter provides for low β 3Gal-T5 expression (0.1–0.4 fg/pg of β -actin) in both humans and mice but has substantial gastrointestinal specificity in mice, as previously suggested (15). The intrinsic weakness of such a promoter is explained by the relevant divergence from the CBF consensus sequence. In fact, it has been reported that any variant of the CCAAT core sequence, including CCTAT, strongly reduces the binding affinity (20). These structural and functional features are reflected in model cell systems, where luciferase activity driven by a type A promoter construct is much lower than

the SV-40 promoter control. A common role of β 3Gal-T5 in some general mammalian cell function may be proposed when it is transcribed under the control of this promoter. Little information is currently available concerning type 1 chain expression in the mouse (21, 22), where α 1,4 fucosylation does not occur (23). Moreover, it is not presently clear how β 3Gal-T5 predominates in the mouse gastrointestinal tract, as it happens in the human counterpart, simply using the conserved CBF-dependent promoter.

On the other hand, the LTR promoter provides humans and some primates with not only gastrointestinal tissue specificity but also high expression levels (>10 fg/pg of β -actin in the colon). In the human ileum, however, high expression levels and possibly specificity is determined by another promoter not yet characterized, leading to the proposal of a distinct role for β 3Gal-T5 in intestinal cell-specific functions (6). The intrinsic strength of the LTR promoter is also consistent with the conserved and proper orientation of both the HNF-1 and Cdx binding sites (6, 24) and is recapitulated by our *in vitro* experiments showing very high luciferase activity for the LTR reporter construct in model cells. However, nothing is currently apparent with regards to the ancestral type A promoter that can explain the insertion and stabilization of the LTR retrotransposon. On this basis, we attempted to characterize the putative type B promoter, even though it has activity in certain human gastrointestinal cell lines but has only very weak activity in the normal tissues we tested. We found that the luciferase activity driven by the type B promoter construct was lower than expected in MKN-45 cells, as they express types A and B 5'-UTR transcripts at equivalent levels. However, to measure luciferase activity, it was necessary to remove a 552-bp segment from the 3'-end of the sequence, and this caused the possible deletion of specific binding sites for inhibitory factors. In fact, we found that this region binds various potential regulatory factors in MKN-45 cells. We speculate also that this truncation physically affects the ability of the promoter to drive luciferase activity *in vitro*. Moreover, additional binding sites for stimulatory factors may also have been present in this deleted region, affecting the activity measured for the type B reporter construct. Continuing on from these hypotheses, we pursued the characterization of the type B putative promoter and did find that this promoter and the LTR promoter share an almost identical HNF-1/Cdx binding motif but placed in the opposite orientation. Because such a perfect sequence match probably helped stabilize the element at the time of its transposition, a suggestive hypothesis is that the isolated sequence represents a remnant of a true native promoter that was relevant at the time of the LTR insertion but has since been overwhelmed by the mobile element. The present disposition of binding sites for stimulatory and inhibitory factors may not allow for full expression of the promoter in a reporter system.

Because other human (but not mouse) genes have been reported to have an LTR with or without similar expression patterns (12), it will be interesting to verify whether the context of the LTR insertion in these cases mirrors that here proposed for the β 3Gal-T5 gene.

We also found that the LTR transcript is poorly expressed or not expressed at all in some gastrointestinal cancer cell lines. In

some of these lines, the type B 5'-UTR transcript is very detectable, indicating that its promoter, with an antisense HNF-1/Cdx motif, is active to some extent. Given the dependence of the two promoters on the same transcription factors (but the higher affinity and strength of the LTR promoter), we hypothesize that an unknown strong inhibitory mechanism specifically affects the LTR promoter in cancer cells. This is in good agreement with the finding that the negative regulation of β 3Gal-T5 in colon cancers (3, 6) exceeds the simple reduction of HNF-1 and Cdx binding (6) and represents another source of regulatory variations potentially introduced by retroelements.

Using exon -3-specific primer, we were unable to detect β 3Gal-T5 by PCR in any cDNA we tested, suggesting that such exon belongs to a brain-specific transcript placed under the control of a distinct promoter yet unknown. Because the expression levels of β 3Gal-T5 in brain are extremely low (7), such promoter is presumably very weak. In fact, other β 1,3-galactosyltransferases, such as β 3Gal-T1 (25-27) and β 3Gal-T2 (28), are expected to play a major role in tissues and cells of neuroectodermal origin. In conclusion, the strict conservation of the ancestral promoter from mice to humans, together with the adoption of multiple promoters, including the strong LTR promoter during evolution, suggest that more work is needed to understand the contribution of β 3Gal-T5 and its reaction products in specific cell functions.

Acknowledgments—We thank Dr. Gabri Brambilla (University of Milan) for help and Dr. John Lowe (Case Western Reserve University, Cleveland, OH) for critical reading of the manuscript.

REFERENCES

1. Isshiki, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M., Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T., and Narimatsu, H. (1999) *J. Biol. Chem.* **274**, 12499–12507
2. Zhou, D., Berger, E. G., and Hennet, T. (1999) *Eur. J. Biochem.* **263**, 571–576
3. Salvini, R., Bardoni, A., Valli, M., and Trinchera, M. (2001) *J. Biol. Chem.* **276**, 3564–3573
4. Bäckström, D., Karlsson, N., and Hansson, G. C. (1994) *J. Biol. Chem.* **269**, 14430–14437
5. Angstrom, J., Larsson, T., Hansson, G. C., Karlsson, K. A., and Henry, S. (2004) *Glycobiology* **14**, 1–12
6. Isshiki, S., Kudo, T., Nishihara, S., Ikehara, Y., Togayachi, A., Furuya, A., Shitara, K., Kubota, T., Watanabe, M., Kitajima, M., and Narimatsu, H. (2003) *J. Biol. Chem.* **278**, 36611–36620
7. Dunn, C. A., Medstrand, P., and Mager, D. L. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12841–12846
8. Feschotte, C., Jiang, N., and Wessler, S. R. (2002) *Nat. Rev. Genet.* **3**, 329–341
9. Britten, R. J. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9374–9377
10. Britten, R. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1798–1803
11. Jordan, I. K., Rogozin, I. B., Glazko, G. V., and Koonin, E. V. (2003) *Trends Genet.* **19**, 68–72
12. van de Lagemaat, L. N., Landry, J. R., Mager, D. L., and Medstrand, P. (2003) *Trends Genet.* **19**, 530–536
13. Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R., and Robbins, P. W. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1548–1553
14. Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. (2005) *Glycobiology* **15**, 805–817
15. Dunn, C. A., van de Lagemaat, L. N., Baillie, G. J., and Mager, D. L. (2005) *Gene (Amst.)* **364**, 2–12
16. Mare, L., and Trinchera, M. (2004) *Eur. J. Biochem.* **271**, 186–194
17. Wu, G. D., Chen, L., Forslund, K., and Traber, P. G. (1994) *J. Biol. Chem.* **269**, 17080–17085
18. Pletcher, M. T., Wiltshire, T., Cabin, D. E., Villanueva, M., and Reeves, R. H. (2001) *Genomics* **15**, 45–54
19. Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T. D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., Kube, M., Taenzer, S., Galgoczy, P., Platzer, M., Scharfe, M., Nordsiek, G., Blocker, H., Hellmann, I., Khaitovich, P., Paabo, S., Reinhardt, R., Zheng, H. J., Zhang, X. L., Zhu, G. F., Wang, B. F., Fu, G., Ren, S. X., Zhao, G. P., Chen, Z., Lee, Y. S., Cheong, J. E., Choi, S. H., Wu, K. M., Liu, T. T., Hsiao, K. J., Tsai, S. F., Kim, C. G., Oota, S., Kitano, T., Kohara, Y., Saitou, N., Park, H. S., Wang, S. Y., Yaspo, M. L., and Sakaki, Y. (2004) *Nature* **429**, 382–388
20. Mantovani, R. (1998) *Nucleic Acids Res.* **26**, 1135–1143
21. Bry, L., Falk, P. G., and Gordon, J. L. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1161–1166
22. Zhou, D., Henion, T., Jungalwala, F. B., Berger, E. G., and Hennet, T. J. (2000) *J. Biol. Chem.* **275**, 22631–22634
23. Dupuy, F., Germot, A., Marenda, M., Oriol, R., Blancher, A., Julien, R., and Maftah, A. (2002) *Mol. Biol. Evol.* **19**, 815–824
24. Suh, E., Chen, L., Taylor, J., and Traber, P. G. (1994) *Mol. Cell. Biol.* **11**, 7340–7351
25. Amado, M., Almeida, R., Carneiro, F., Levery, S. B., Holmes, E. H., Nomoto, M., Hollingsworth, M. A., Hassan, H., Schwientek, T., Nielsen, P. A., Bennet, E. P., and Clausen, H. (1998) *J. Biol. Chem.* **273**, 12770–12778
26. Bardoni, A., Valli, M., and Trinchera, M. (1999) *FEBS Lett.* **451**, 75–80
27. Zhu, D., Shen, A., Wang, Y., Gu, X., and Gu, J. (2003) *FEBS Lett.* **538**, 163–167
28. Kolbinger, F., Streiff, M. B., and Katopodis, A. G. (1998) *J. Biol. Chem.* **273**, 433–440