Frameshift and Nonsense Mutations in a Human Genomic Sequence Homologous to a Murine UDP-Gal:β-d-Gal(1,4)-d-GlcNAc α(1,3)-Galactosyltransferase cDNA*

(Received for publication, January 3, 1990)

Robert D. Larsen‡, Carlos A. Rivera-Marrero§, Linda K. Ernst‡, Richard D. Cummings§, and John B. Lowe††

From the ‡Howard Hughes Medical Institute and the §Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0650 and the †Department of Biochemistry, University of Georgia, Athens, Georgia 30602

We have previously isolated a murine UDP-Gal:β-d-Gal(1,4)-d-GlcNAc α(1,3)-Galactosyltransferase (α(1,3)-GT) cDNA (Larsen, R. D., Rajan, V. P., Ruff, M. M., Kukowska-Latallo, J., Cummings, R. D., and Lowe, J. B. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 8227–8231). This enzyme constructs the terminal α(1,3)-galactosyl linkage within the epitope Galα1→3Gal. This epitope is expressed by New World monkeys and many nonprimate mammals but generally not by Old World primates, anthropoid apes, or man. To investigate the molecular basis for the apparent species-specific absence of this enzyme and its oligosaccharide product, we have sequenced a human genomic DNA fragment homologous to the murine α(1,3)-GT cDNA. This fragment contains a 703-nucleotide region that shares 82% identity with a region of the murine cDNA encoding part of the enzyme’s catalytic domain. The human sequence, however, has suffered deletion of single nucleotides at two separate positions, relative to the murine sequence. These frameshift mutations disrupt the translational reading frame that would otherwise maintain a 76% amino acid sequence identity between the human sequence and the murine α(1,3)-GT. Moreover, nonsense mutations exist within this disrupted reading frame that would truncate the human polypeptide, relative to the murine enzyme. We therefore propose that this human sequence represents a pseudogene and cannot determine expression of Galα1→3Gal epitopes on human cells.

The carbohydrate structure Galα1→3Galβ1→4GlcNac-R is the product of a UDP-Gal:β-d-Gal(1,4)-d-GlcNAc α(1,3)-galactosyltransferase (α(1,3)-GT)* which catalyzes the following reaction:

\[
\text{Galβ1→4GlcNac-R + UDP-Gal} \rightarrow \text{Galα1→3Galβ1→4GlcNac-R}
\]

* This work was supported by the Howard Hughes Medical Institute, and in part by National Institute of Health Grant DK38462 (to J. B. L.) and CA37626 (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.

†† To whom correspondence should be addressed: Howard Hughes Medical Institute, Medical Science Research Bldg., I. Rm. 3510, 1190 W. Medical Center Dr., Ann Arbor, MI 48109-0650. Tel: 313-747-4779.

§ The abbreviations used are: α(1,3)-GT, UDP-Gal:α1→3Gal(1,4)-4GlcNAc α(1,3)-galactosyltransferase; PCR, polymerase chain reaction; kb, kilobase(s); SDS, sodium dodecyl sulfate; bp, base pair(s).

where R is a glycoprotein or glycolipid moiety (1, 2). The product of this reaction, also known as the α-galactosyl epitope, and its corresponding α(1,3)-GT, are distinct from those of the human B blood group system in that the latter involve acceptor and product oligosaccharides whose β-linked galactose is substituted with α-linked fucose (3). Tissues of nonprimate mammals, prosimians, and New World monkeys express abundant α-galactosyl epitopes, as well as a corresponding α(1,3)-GT activity (4, 5). By contrast, the α-galactosyl epitope is generally undetectable in tissues of Old World monkeys, apes, and man, and an α(1,3)-GT activity has not been described in these species. It is interesting that the sera of these latter species contain a potent, naturally-occurring IgG antibody to the α-galactosyl epitope, an antibody not found in the sera of species that do express the epitope (6). These studies have led to the suggestion that suppression of the biosynthesis of the α-galactosyl epitope may have occurred in ancestral Old World primates following the separation of New World and Old World primates approximately 25-35 million years ago, but before the divergence of man, apes, and Old World monkeys (4).

The general absence of α-galactosyl epitopes on human cells is undoubtedly due to a lack of the cognate α(1,3)-GT activity, because human cells express both the donor nucleotide sugar UDP-Gal and N-acetyllactosamine acceptors suitable for galactosylation by exogenous α(1,3)-GT activity (4, 7). Studies using antibody or lectin reagents that recognize the α-galactosyl epitope, however, suggest that trace amounts of these molecules may be detectable on senescent human red blood cells, where they have been proposed to play an important role in the physiologic destruction of these cells (8, 9). Similar studies indicate that α-galactosyl epitopes apparently may be expressed by some malignant human cells (10). Despite the fact that a corresponding α(1,3)-GT activity was not identified in these studies, these and related observations have prompted the speculation that the human genome contains a gene capable of encoding a functional α(1,3)-GT, and that it is expressed in a highly restricted manner in humans (11).

We have recently cloned a functional murine α(1,3)-GT cDNA (12); our preliminary Southern blot studies indicated that sequences homologous to the murine cDNA are present in the human genome. Joziasse et al. (13) have reported similar observations with a bovine α(1,3)-GT cDNA. These results suggested that molecular cloning of these cross-hybridizing sequences and analysis of their structure might provide insight into the molecular basis for the apparent species-specific repression of expression of the α-galactosyl epitope and the corresponding α(1,3)-GT. We report here the identification
of two distinct human DNA sequences that cross-hybridize to a portion of the mouse α(1,3)-GT cDNA encompassed within its catalytic domain. The most strongly hybridizing of these sequences was isolated from a human genomic DNA phage library. Sequence analysis of this segment indicates that it represents a pseudogene and thus does not determine expression of α-galactosyl epitopes by human tissues.

**EXPERIMENTAL PROCEDURES**

**Materials**—DNA restriction and modification enzymes were obtained from Boehringer Mannheim and New England Biolabs. The phage cloning vector X FIIX, the cosmid cloning vector pWE-15, and Gigapak Gold Lambda DNA packaging extracts were from Strатrigen. The plasmid cloning vector pGEM-7zf(−) was purchased from Promega. [α-32P]dCTP (3000 Ci/mmol) was obtained from Amer- sham Corp.

**Construction and Screening of a Human Genomic Library**—High molecular weight human genomic DNA was partially digested with Sau3AI and then size-fractionated by differential centrifugation through a 1.25-5.0 M sodium chloride gradient (14). Fractions with fragment sizes between 9 and 24 kb were pooled, filled in by two nucleotides using the Klenow fragment of DNA polymerase I and dCTP, and then ligated to commercially prepared λ FIX arms. The ligated products were packaged using Gigapak Gold extract. Approximately 1.5 x 10^6 recombinant λ phage were screened by plaque hybridization. Plaque lifts were prepared using nitrocellulose filters (Schleicher & Schuell), and were prehybridized at 42 °C for 4 h in 50% formamide, 5 X SSC, 10X Denhardt's solution (14) and 0.1% SDS. Filters were hybridized for 24 h at 42 °C in the prehybridization solution containing 10% dextran sulfate, 100 µg/ml single stranded salmon sperm DNA, and 1 x 10^6 cpm/ml [32P]-labeled probe. The probe consisted of a 1.5-kb XhoI fragment of the murine α(1,3)-GT cDNA (12), which was labeled with [α-32P]dCTP (15). The filters were rinsed three times for 15 min each at room temperature in 2 X SSC, 0.1% SDS, and then subjected to autoradiography. The blots were washed once at high stringency (0.1 X SSC, 0.1% SDS at 65 °C for 30 min and then subjected to autoradiography. The blots were then washed once at 65 °C for 30 min and exposed to x-ray film.

**RESULTS**

**Mammalian Genomes Contain Sequences Homologous to a Murine α(1,3)-GT cDNA**—Southern blot analysis indicates that the genomes of several mammalian species maintain sequences that cross-hybridize with a segment of the murine α(1,3)-GT cDNA that encodes a portion of the enzyme's catalytic domain (Fig. 1). These sequences were detected in the genomes of lower mammals (pig and hamster), a New World monkey (squirrel monkey), a great ape (chimpanzee), and man. We also have found cross-hybridizing sequences in the macaque genome, an Old World monkey (data not shown). Zojiasse et al. (15) have obtained similar results when they had human genomic DNA with a bovine α(1,3)-GT cDNA.

**Fig. 1. Low stringency Southern blot analysis of mammalian genomic DNA.** High molecular weight genomic DNA was digested with BamHI, electrophoresed through a 0.7% agarose gel, and transferred to a nylon membrane as described under "Experimental Procedures." The probe consisted of a radiolabeled Sac1/XhoI fragment representing the 3' 806 bp of the murine α(1,3)-GT cDNA. The blot was hybridized and washed at low stringency (see "Experimental Procedures") and subjected to autoradiography. Relative mobilities of molecular size markers are indicated at the left in kilobases.
These data indicate that complete deletion of the α(1,3)-GT locus has not occurred in Old World primates, including man, leaving open the possibility that these cross-hybridizing sequences represent functional α(1,3)-GT genes whose expression is repressed in a species-specific manner.

Human Genomic Sequences with Homology to a Murine α(1,3)-GT cDNA Contain Multiple Frameshift and Nonsense Mutations—To further define the molecular basis for the apparent generalized lack of expression of α-galactosidase epitopes in humans, we isolated lambda phages containing human DNA segments that cross-hybridized to the murine α(1,3)-GT cDNA. Five hybridization-positive phages were characterized by restriction site mapping and Southern blot analysis. These were found to have similar restriction maps, and each maintained a 5.5-kb BamHI fragment which hybridized with the murine α(1,3)-GT cDNA insert (data not shown) and which co-migrated with a 5.5-kb cross-hybridizing fragment in BamHI-digested human genomic DNA (see below). This 5.5-kb BamHI fragment was subsequently subcloned from a representative phage into the mammalian expression vector pWE-15 for DNA sequence analysis. This subclone was designated pHG-T-1.

DNA sequence analysis of the region in pHG-T-1 that is homologous to the murine α(1,3)-GT cDNA is presented in Fig. 2. The 805-bp human genomic DNA sequence presented here includes a 703-bp segment which is identical to 578 nucleotides in a corresponding 705-bp region in the murine α(1,3)-GT cDNA, yielding a nucleotide sequence identity between the two segments of 82%. The human sequence also maintains an 80% identity to the corresponding portion of a bovine α(1,3)-GT cDNA (Ref. 13, not shown). This 705-bp homologous segment of mouse α(1,3)-GT cDNA encodes the COOH-terminal 230 amino acids of this enzyme, within its catalytic domain (12). Despite this high degree of nucleotide sequence conservation, the human sequence has accumulated two separate and functionally significant single nucleotide deletions, relative to the murine and bovine sequences. These two deletions, found at bp 823′ and bp 905′ (Fig. 2), each represent a frameshift mutation and disrupt the translational reading frame of the human sequence, relative to the murine and bovine reading frames. This eliminates the potential similarity between the derived human protein sequence and the murine and bovine α(1,3)-GT sequences (distal to the first frameshift mutation. If, however, both frameshift mutations are “suppressed” to yield a “corrected” reading frame (Fig. 2), the derived human protein sequence can be seen to maintain a 76% identity, and a 92% identity, respectively, to the murine and bovine sequences. However, this corrected reading frame contains two separate nonsense mutations, at bp 857′ and 1139′, that would prematurely terminate translation, relative to the conserved position of the authentic murine and bovine termination codons corresponding to bp 1183′ of the human sequence. These sequence features suggest that this human DNA segment cannot encode a functional α(1,3)-GT, and thus represents a pseudogene. This likely is an unprocessed pseudogene since, as shown in Fig. 2, a consensus splice acceptor sequence (23) is found immediately proximal to the beginning of the homology between the murine and human nucleotide sequences. DNA sequence analysis of the remainder of this human fragment indicates that no additional nucleotide sequence similarity exists between it and the murine cDNA sequence (data not shown). This suggests that the remainder of this fragment represents intronic sequence.

Sequence Confirmation by Polymerase Chain Reaction Analysis of Human Genomic DNA—The sequence features described above were identified in a single cloned human genomic DNA segment. We wished to rule out the formal possibility that these represented cloning artifacts or nucleotide sequence polymorphisms. Therefore, we implemented a method based upon the polymerase chain reaction (20) for analysis of multiple alleles representing these sequences. Synthetic oligonucleotides were designed to amplify a 395-bp segment of human genomic DNA that contained both frameshift mutations and the nonsense mutations (Fig. 2 and “Experimental Procedures”). DNA samples prepared from three different individuals were subjected to amplification with these oligonucleotides, and the amplified products from each were gel-purified and subcloned. The relevant region of the inserts in 12 independent subclones from each of the three individuals were sequenced. The nucleotide sequence of all 36 PCR fragment subclones analyzed were identical to that of the homologous segment of pHG-T-1 (data not shown).

We noted that the nonsense mutation between the two frameshift mutations is encompassed by a site for the restriction endonuclease AvrII (Fig. 2). As an independent confirmation of the sequencing results, aliquots of the amplified DNA fragments were digested with this enzyme, fractionated by agarose gel electrophoresis in parallel with undigested fragments, and the fragments detected by Southern blotting with a segment of the murine α(1,3)-GT cDNA (see “Experimental Procedures”). In each instance, the amplified 395-bp fragment from three different individuals could be quantitatively cleaved by AvrII digestion into the predicted component 324- and 71-bp fragments (data not shown). These results indicate that the AvrII site that encompasses a nonsense codon in the corrected translational reading frame is present in both alleles represented by this DNA segment, in three different individuals. Taken together, these results indicate that the sequence data generated from the single cloned human genomic DNA segment faithfully characterize the human genomic sequence represented by that cloned segment, across the region that encompasses the relevant sequence features.

Southern Blot Analysis of Human Genomic DNA and Cloned Human Genomic Sequences Homologous to the Murine α(1,3)-GT cDNA—We wished to confirm that the 5.5-kb BamHI fragment in pHG-T-1 that cross-hybridized with the murine α(1,3)-GT cDNA probe faithfully represents the corresponding genomic BamHI fragment. Moreover, since the DNA sequence analysis of this fragment indicated that it was a pseudogene, we wished to try to identify other cross-hybridizing fragments that might represent an authentic human α(1,3)-GT gene. Therefore, human genomic DNA was subjected to Southern blot analysis, in parallel with plasmid pHG-T-1. To simplify these analyses, we used as a probe an 806-bp ScaI/HindIII fragment isolated from the murine α(1,3)-GT cDNA (Ref. 12 and “Experimental Procedures”). This fragment is colinear with the entire region of homology between the mouse and human sequence in pHG-T-1, but also extends just 53 bp proximal to the 5′ end of the homology between the two sequences. This probe was chosen since it is found within the portion of the enzyme known to encompass the enzyme’s catalytic domain, and thus might represent more highly conserved sequences. Moreover, since it included only 53 bp extraneous to the homology, it would not likely hybridize to fragments containing exonic sequences corresponding to other parts of the murine α(1,3)-GT cDNA. To further simplify these analyses, we used enzymes that do not cut the human DNA fragment in pHG-T-1 within the region of homology to the murine cDNA. Thus, this sequence would be represented by only a single hybridizing fragment after digestion with these enzymes.
Frameshift and Nonsense Mutations in a Human α(1,3)-GT Homologue

Human genomic DNA

Splice acceptor similarity

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Human protein

Human genomic DNA

Murine cDNA

Bovine protein

---

Fig. 2. Comparison of DNA and derived protein sequences of murine α(1,3)-GT cDNA, bovine α(1,3)-GT cDNA, and cloned homologous human genomic sequences. The nucleotide sequence of an 801-bp segment of pHGT-1 (Human genomic DNA) is shown, with the corresponding amino acid sequence derived from the relevant reading frame (Human protein, reading frames denoted by a, b, and c). The nucleotide sequence is numbered in register with the published sequence of the murine α(1,3)-GT cDNA (12), a portion of which is displayed (Murine cDNA) below the human DNA sequence. Vertical lines between the murine and human DNA sequences.
sequences denote nucleotide sequence identity. Human genomic DNA sequences located 5' from bp 492' which exhibit no homology to the murine α(1,3)-GT cDNA, are displayed in lower case letters. A portion of this part of the human sequence which displays strong similarity to the mammalian consensus splice acceptor sequence (25) is double underlined. The predicted amino acid sequences inferred from the nucleotide sequences of the murine α(1,3)-GT cDNA (12) (Murine protein) and the bovine α(1,3)-GT cDNA (13) (Bovine protein) are indicated below the murine nucleotide sequence. Amino acids within these sequences that are identical to the corresponding human amino acid residue are indicated by a hyphen. The two segments of the human DNA sequence used to generate the polymerase chain reaction amplimers are denoted by the stippled underlining. The AurII site used to analyze polymerase chain reaction products is underlined.

Fig. 3 displays the results of this analysis. This blot was probed at low stringency and then consecutively washed at low stringency (panel A), and then high stringency (panel B). After the low stringency wash (Fig. 3A), a simple hybridization pattern is seen in the genomic DNA samples (i.e. one or two relatively intense bands are detected, lanes 1–5), independent of the restriction enzyme(s) used. These include a single prominent BamHI fragment (panel A, lane 1), and two prominent fragments generated by digestion with HindIII (panel A, lane 2), PvuII (panel A, lane 3), BamHI plus HindIII (panel A, lane 4), or BamHI plus PvuII (panel A, lane 5).

DNA was digested with restriction enzymes, electrophoresed through a 0.7% agarose gel, and transferred to a nylon membrane as described in “Experimental Procedures.” The probe consisted of a radiolabeled Scal/XhoI fragment representing the 3' 806 bp of the murine α(1,3)-GT cDNA segment. The relative probe is more similar to the cross-hybridizing fragment in the HindIII genomic segment represented in pHGT-1 than it is to the other fragment. Indeed, following HindIII digestion of human genomic DNA (lane 1) actually represented two fragments. One of these contains one or more HindIII sites, as found in the BamHI fragment in pHGT-1, and represents that fragment, whereas the other member of this doublet is not present in pHGT-1. Analysis of the PvuII and PvuII-BamHI double digests of genomic DNA and pHGT-1 confirms this interpretation. PvuII digestion of genomic DNA (lane 3) yields a strongly hybridizing 5.5-kb fragment, and a more weakly hybridizing 1.7-kb fragment. PvuII-BamHI double digestion of genomic DNA (lane 5) reduces the 5.5-kb BamHI fragment to a single 1.6-kb cross-hybridizing fragment (lane 7). Likewise, a cross-hybridizing 1.6-kb BamHI-HindIII fragment is also present in human genomic DNA (lane 4). However, a prominent 5.5-kb BamHI-HindIII fragment also remains. This fragment is most likely not an artifact of partial digestion with HindIII since it remains despite a severalfold overdigestion with this enzyme. This result suggested that the 5.5-kb BamHI band identified in human genomic DNA (lane 1) actually represented two fragments. One of these contains one or more HindIII sites, as found in the BamHI fragment in pHGT-1, and presents that fragment, whereas the other member of this doublet is not present in pHGT-1. Analysis of the PvuII and PvuII-BamHI double digests of genomic DNA and pHGT-1 confirms this interpretation. PvuII digestion of genomic DNA (lane 3) yields a strongly hybridizing 5.5-kb fragment, and a more weakly hybridizing 1.7-kb fragment. PvuII-BamHI double digestion of genomic DNA (lane 5) reduces the 5.5-kb fragment to 4.3 kb but does not alter the mobility of the more weakly hybridizing 1.7-kb fragment. PvuII-BamHI double digestion of pHGT-1 (lane 8) yields a fragment that co-migrates with the more strongly hybridizing 4.3-kb PvuII-BamHI fragment identified in genomic DNA (lane 5). These results indicate that plasmid pHGT-1 contains one of the two human sequences that cross-hybridize with the Scal/XhoI murine α(1,3)-GT cDNA segment. The relative hybridization intensities observed suggest that the murine probe is more similar to the cross-hybridizing fragment in pHGT-1 than it is to the other fragment. Indeed, following high stringency washes, the weakly hybridizing 5.5-kb BamHI-HindIII and 1.7-kb PvuII-BamHI genomic fragments (panel A, lanes 4 and 5, respectively) not present in pHGT-1 (compare to panel A, lanes 7 and 8) are only faintly visible with the Scal/XhoI probe (panel B, lanes 4 and 5). Likewise, the 7.5-kb HindIII genomic fragment detected at low stringency (panel A, lane 2) is nearly undetectable following high stringency washing (panel B, lane 2), whereas the 1.6-kb HindIII genomic segment represented in pHGT-1 continued to yield a strong signal at the higher stringency (panel B, lane 2). Virtually identical results were obtained when the same blot was probed with a PCR-generated segment of the human

FIG. 3. Southern blot analysis of human genomic DNA and clone pHGT-1 plasmid DNA. High molecular weight human genomic DNA (lanes 1–5) and clone pHGT-1 plasmid DNA (lanes 6–8) were digested with restriction enzymes, electrophoresed through a 0.7% agarose gel, and transferred to a nylon membrane as described in “Experimental Procedures.” The probe consisted of a radiolabeled Scal/XhoI fragment representing the 3’ 806 bp of the murine α(1,3)-GT cDNA. The blot was hybridized at 35 °C in a hybridization solution described in “Experimental Procedures.” Panel A, autoradiograph of the blot after washing at low stringency (2 × SSC, 0.5% SDS at 45 °C for 30 min). Panel B, autoradiograph of the blot after washing at high stringency (0.1 × SSC, 0.1% SDS at 85 °C for 30 min). Lane 1, BamHI; lane 2, HindIII; lane 3, PvuII; lane 4, BamHI-HindIII double digest; lane 5, PvuII-BamHI double digest; lane 6, BamHI; lane 7, BamHI-HindIII double digest (faint bands visible at 1.9 and 4.2 kb represent small amounts of partially digested pHGT-1); lane 8, PvuII-BamHI double digest. Relative mobilities of molecular size markers are indicated at the left in kilobases.
sequence in pHGT-1 that consists exclusively of a sequence comprising the homologous region, using hybridization and wash conditions identical to those described for the mouse probe (data not shown). This excludes the possibility that the most weakly hybridizing fragments detected with the mouse probe represent exonic sequences that hybridize to the 53 base pairs in that probe that are extraneous to the region of homology. Taken together, these results indicate that at least two human genomic DNA sequences exist that exhibit sequence similarity to the murine Scal/Xhol α(1,3)-GT cDNA probe, and that we have cloned and sequenced the one most similar to this mouse α(1,3)-GT cDNA.

**DISCUSSION**

Over the past several years, Galili and his colleagues (4-6, 24-26) have investigated the nature of a relatively abundant, naturally occurring antibody in normal human sera that has an apparent specificity for terminal α-linked galactosyl residues. Hapten inhibition studies and the results of reactivities of these “anti-Gal” antibodies against natural and synthetic oligosaccharides confirm that they recognize terminal, nonsubstituted Galα1-3Gal and Galα1-6Gal linkages, but not Galα1-4Gal or β-galactosyl linkages (24, 26). Based upon the observation that Galα1-3Gal-containing molecules represent the most potent inhibiting hapten for anti-Gal, it has been proposed that this antibody is specific for terminal, unsubstituted Galα1-3Gal linkages (6, 24).

The concept of immune tolerance, as exemplified by the reciprocal relationship between the genetically-determined ability to construct an antigen(s) of the ABO blood group system and the absence of the cognate, naturally occurring ABO isoglutaminins (3), predicts that humans thus do not normally express Galα1-3Gal molecules recognized by this naturally occurring anti-Gal antibody. Indeed, Galili et al. (4, 5, 26) have described numerous examples of a species-dependent reciprocal relationship between the presence of α-galactosyl epitopes on cells or tissues and naturally occurring anti-Gal antibody. In these studies, it has been noted that this antibody is not present in the sera of nonprimate mammals, nor in a variety of New World monkeys, whereas these species each express relatively abundant amounts of an α-galactosyl epitope on their cells. By contrast, anti-Gal antibody is easily detected in the sera of Old World monkeys, anthropoid apes, and man, species that apparently do not normally express α-galactosyl epitopes. This has led to the speculation that some evolutionary event resulted in the species-specific repression of the α(1,3)-GT that constructs Galα1→3Gal linkages (5, 6).

Interestingly, however, these and other investigators find that antibody and lectin reagents capable of detecting α-galactosyl epitopes react weakly with some human cells and tissues (5, 8, 10, 24). These observations have led to the suggestion that the human genome contains a functional α(1,3)-GT gene whose expression is normally subject to stringent regulation (4, 11). It should be noted, however, that in all of these studies the identification of α-galactosyl epitopes as Galα1→3Gal molecules is based entirely upon the specificity of the anti-Gal antibodies or lectins used to detect them. These data are unaccompanied by chemical confirmation of these structures or demonstration of a corresponding α(1,3)-GT activity. Moreover, in just one instance has data been generated by an experimental approach complementary to lectin- or antibody-based methods supporting the suggestion that human cells can express Galα1→3Gal linkages (27). Thus while there is indirect evidence for the expression of Galα1→3Gal linkages by human tissues, we believe that it is not yet compelling. Moreover, the observation that human sera contain significant amounts of naturally occurring antibody directed against the α-galactosyl epitope is consistent with the notion that this structure is not expressed during human ontogeny. Taken together, we believe these considerations create a reasonable uncertainty as to the ability of the human genome to encode a functional α(1,3)-GT.

We and others have recently isolated cloned cDNAs that encode murine (12) and bovine (13) α(1,3)-GTs. Sequence comparisons indicate that they maintain a sequence identity that exceeds 75% throughout most of their coding regions. Our Southern blot analysis of DNA from a variety of other mammalian species, including primates, indicates that each of their genomes contains sequences homologous to these cDNAs. If it is assumed that the murine and bovine enzymes correspond to the putative α(1,3)-GTs whose expression is restricted in Old World monkeys, apes, and man, these data then indicate that large scale deletion of the primate gene homologues is not responsible for the apparent diminished or absent expression in these species.

To further address the molecular basis for the apparent species-specific repression of α-galactosyl epitopes and the corresponding enzyme, we isolated and sequenced a human genomic DNA fragment that is homologous to a murine α(1,3)-GT cDNA. This human DNA sequence is highly homologous to the corresponding region of the murine and bovine cDNA sequences. However, inspection of this sequence indicates that it has suffered deletion of a single nucleotide at two locations, relative to the murine and bovine sequences, that each represent a frameshift “mutation.” The most proximal of these frameshifts, at position 822 within the human sequence, yields a reading frame containing a termination codon at position 973 (Fig. 2). This reading frame predicts a human translation product whose COOH terminus is truncated by 70 amino acids, relative to the corresponding portion of the catalytic domain of the murine and bovine enzymes. Moreover, the 50 amino acids immediately following the frameshift, that are predicted by this reading frame, are markedly different from the corresponding residues in the murine and bovine enzymes within this highly conserved region (not shown). Thus, while it remains possible that this human sequence does encode an enzymatically functional protein, this seems unlikely given the significant differences between this portion of the predicted human protein and its mouse and bovine counterparts.

Furthermore, inspection of this human sequence indicates that it has accumulated two other significant mutations with respect to the murine and bovine cDNA sequences. Thus, in order to maximize the similarity between the protein sequence derived from the human DNA segment and its murine and bovine counterparts, it is necessary to “suppress” two nonsense mutations found in the human DNA sequence, as well as the two frameshift mutations resulting from single nucleotide deletions. Taken together, these analyses are most consistent with the notion that this human genomic DNA segment represents a pseudogene; the presence of a consensus splice acceptor sequence immediately proximal to the region comprising the inter-species homology indicates that it is an unprocessed pseudogene.

These data imply that this human DNA sequence does not represent the gene responsible for “normal” expression of Galα1→3Gal linkages on human red blood cells (5, 26), nor is it responsible for “aberrant” expression of these structures on human lung (27) or mammary (10) carcinoma cells. With regard to aberrant expression of Galα1→3Gal linkages in malignancy, it should be noted that the human tumor cell line MCF7 has been shown by others to apparently express cell
surface α-galactosyl epitopes (10). Using a bovine α(1,3)-GT cDNA probe, however, Joziasse et al. were unable to detect a cross-hybridizing transcript in this cell line, as assayed by Northern blotting of 5 μg of poly(A+) RNA prepared from these cells (13). Given the fact that the sequence of their bovine probe shares 80% sequence identity with the human sequence we present here, it seems unlikely that sequence differences can account for their inability to detect a transcript. We also have been unable to detect homologous transcripts in the MCF7 cell line.*

In order for this sequence to encode a functional α(1,3)-GT, the four mutations described above would presumably have to be corrected by processes involving extensive somatic mutation. Therefore, if human cells do in fact express α-galactosyl structures that represent Galα1-3Gal linkages, it would then seem likely that these structures must be constructed by an enzyme that is the product of another DNA sequence. Such a putative sequence could be either similar to, or completely distinct from, the murine and bovine sequences. In this context, our Southern blot analyses suggest that the human genomic DNA segment we have cloned and sequenced is the one most similar to the 3’ 806 bp of the murine α(1,3)-GT cDNA. However, these analyses also identified an additional sequence in the human genome that is apparently less similar to the 3’ 806 bp of the mouse α(1,3)-GT cDNA. These sequences could represent a functional α(1,3)-GT gene, although their precise nature remains undefined. Neither have our studies eliminated the possibility that the human genome contains a functional α(1,3)-GT gene that is nonhomologous to the murine α(1,3)-GT cDNA. The gene for such an enzyme could potentially be isolated by identifying a human cell line or tissue that expresses α(1,3)-GT activity, constructing a cDNA expression library from mRNA isolated from those cells or tissues, and then screening that library using the cDNA expression cloning system described by Larsen et al. (12).

We are, however, unaware of published evidence documenting the existence of a human α(1,3)-GT activity. Explanations for the apparent presence of the corresponding Galα1-3Gal linkages on human tissues, in the absence of α(1,3)-GT activity, include the possibility that low levels of these molecules might be synthesized by a glycosyltransferase whose major product is distinct from the Galα1→3Gal linkage, but which may also be competent to construct Galα1→3Gal linkages because of a substrate or acceptor “promiscuity.” Observations concerning the substrate requirements of the human B blood group galactosyltransferase provide a precedent for this possibility. It is known, for example, that while the human B blood group galactosyltransferase uses UDP-galactose most efficiently as a substrate in the formation of blood group B linkages, this enzyme is also able to utilize UDP-N-acetylgalactosamine to construct blood group A structures (28). Whether this possibility, or others, represents the mechanism whereby human cells express α-galactosyl epitopes will most likely await determination of the chemical nature of these structures, and identification of corresponding enzymatic activities in human cells or tissues.

Acknowledgments—We thank Christy Arndt for providing technical assistance with DNA sequencing, and express our appreciation to David Ginsburg and Jeff Bonadio for critical review of this manuscript.

Note Added in Proof—Comparisons made between the human sequence reported here and a human UDP-GalNAC:Fucol→2Galα1→3GalNac transferase (blood group A transferase) sequence (Yamaoto, F. I., Macken, J., Tsug, T., White, T., Clewes, H., and Hokomori, S.-I. (1990) J. Biol. Chem. 264, 1146-1151) published after submission of this paper, indicates that they share roughly 55% identity at the nucleic acid level and also maintain similar predicted protein sequences, suggesting an interesting structural and evolutionary relationship between the two classes of enzymes represented by these two sequences.

REFERENCES


C. A. Rivera-Marrero and R. D. Cummings, unpublished data.
Frameshift and nonsense mutations in a human genomic sequence homologous to a murine UDP-Gal:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-galactosyltransferase cDNA.

R D Larsen, C A Rivera-Marrero, L K Ernst, R D Cummings and J B Lowe


Access the most updated version of this article at [http://www.jbc.org/content/265/12/7055](http://www.jbc.org/content/265/12/7055)

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 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/265/12/7055.full.html#ref-list-1](http://www.jbc.org/content/265/12/7055.full.html#ref-list-1)