Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6674–6678, September 1990
Biochemistry

Molecular cloning, sequence, and expression of a human GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferase cDNA that can form the H blood group antigen

(oligosaccharide biosynthesis/glycosyltransferase/surface antigen/chromosome 19)

ROBERT D. LARSEN*, LINDA K. ERNST*, RAJAN P. NAIR*, AND JOHN B. LOWE*†‡

*Howard Hughes Medical Institute and †Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109-0650

Communicated by Stuart Kornfeld, June 4, 1990

ABSTRACT **We have previously used a gene-transfer scheme to isolate a human genomic DNA fragment that determines expression of a GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferase [α(1,2)FT; EC 2.4.1.69]. Although this fragment determined expression of an α(1,2)FT whose kinetic properties mirror those of the human H blood group α(1,2)FT, their precise nature remained undefined. We describe here the molecular cloning, sequence, and expression of a human cDNA corresponding to these human genomic sequences. When expressed in COS-1 cells, this cDNA directs expression of cell surface H structures and a cognate α(1,2)FT activity with properties analogous to the human H blood group α(1,2)FT. The cDNA sequence predicts a 365-amino acid polypeptide characteristic of a type II transmembrane glycoprotein with a domain structure analogous to that of other glycosyltransferases but without significant primary sequence similarity to these or other known proteins. To directly demonstrate that the cDNA encodes an α(1,2)FT, the COOH-terminal domain predicted to be Golgi-resident was expressed in COS-1 cells as a catalytically active, secreted, and soluble protein A fusion peptide. Southern blot analysis showed that this cDNA identifies DNA sequences syntenic to the human *H* locus on chromosome 19. These results strongly suggest that this cloned α(1,2)FT cDNA represents the product of the human H blood group locus.**

The antigens of the human ABO blood group system are carbohydrate molecules constructed by the sequential action of a series of distinct glycosyltransferases (1, 2). The terminal step in this pathway, catalyzed by the allelic glycosyltransferase products of the ABO locus, requires the expression of a precursor molecule called the H antigen. The blood group H antigen is an oligosaccharide molecule whose expression is normally restricted to the surfaces of human erythrocytes and a variety of epithelial cells, including those that line the gastrointestinal, urinary, and respiratory tracts (1, 3). The H antigen is a fucosylated structure of the form Fucα1-2Galβ-, whose expression is determined by GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferases [α(1,2)FTs; EC 2.4.1.69]. These enzymes catalyze a transglycosylation reaction between their sugar nucleotide substrate GDP-L-fucose and oligosaccharide acceptor substrates with terminal type I (Galβ1-3GlcNAc-) or type II (Galβ1-4GlcNAc-) moieties (1).

Surface-expressed H determinants exhibit precise temporal and spatial changes in their expression patterns during human and murine development (4, 5). The functional significance of these changes is as yet unknown, although evidence suggests that other fucosylated molecules participate in adhesive events during development (6–8). Cloned gene segments that determine H antigen expression represent tools to address this question by genetic approaches that perturb H antigen expression during development. We, therefore, established a gene-transfer approach to isolate human DNA segments that determine expression of cell surface H molecules and their corresponding α(1,2)FTs (9, 10). These experiments yielded a cloned human DNA segment that determines expression of an α(1,2)FT activity when transfected into a mammalian cell line deficient in this enzyme activity. This enzyme activity was kinetically similar to the human H blood group α(1,2)FT but distinct from the human secretor (SE) α(1,2)FT. Although these data were consistent with the hypothesis that this segment represented part or all of the structural gene encoding the H α(1,2)FT, they were consistent also with the possibility that the DNA sequences trans-determined enzyme expression by interaction with an endogenous gene, transcript, or protein. We report here our analysis of a cloned cDNA representing the product of this human genomic DNA segment.§ These data indicate that this segment encodes the human H blood group α(1,2)FT.

MATERIALS AND METHODS

Cell Lines and DNA Samples. DNA from the cell line UV5HL9-5 (11) and from the Chinese hamster ovary hybrid parent were provided by H. Mohrenweiser and K. Tynan (Lawrence Livermore National Laboratory, Livermore, CA). The origins of all other cell lines and conditions for cell culture are as described (9, 10, 12, 13). Genomic DNA samples from a panel of Chinese hamster ovary × human somatic cell hybrids informative for human chromosomes were purchased from BIOS (New Haven, CT).

Isolation of Human α(1,2)FT cDNA Clones. Approximately 1.8×10^6 recombinant clones from an A431 cell cDNA mammalian expression library (13) were screened by colony hybridization using a ^{32}P-labeled (14) 1.2-kilobase (kb) *Hin*fI fragment of pH3.4 (10) as a probe. Filters were hybridized for 18 hr at 42°C in a hybridization solution as described (9, 10), washed, and subjected to autoradiography. Two hybridization-positive colonies were obtained and isolated by two additional rounds of hybridization and colony purification. Preliminary sequence analysis of the inserts in both hybridization-positive cDNA clones indicated that they each were in the anti-sense orientation with respect to the pCDM7 expression vector (15, 16) promoter sequences. The largest insert was, therefore, recloned into pCDM7 in the sense

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: α(1,2)FT, GDP-L-fucose:β-D-galactoside 2-α-L-fucosyltransferase.

‡To whom reprint requests should be addressed at: Howard Hughes Medical Institute, Medical Science Research Building I, Room 3510, 1150 West Medical Center Drive, Ann Arbor, MI 48109-0650.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M35531).

orientation for expression studies, and the resulting plasmid was designated pCDM7-α(1,2)FT.

Flow Cytometry Analysis. COS-1 cells were transfected with plasmid DNAs by using a DEAE-dextran procedure (17) as described (16). Transfected cells were harvested after a 72-hr expression period and stained either with a mouse IgM anti-H monoclonal antibody (10 μg/ml; Chembiomed, Edmonton, AB, Canada) or with a mouse IgM anti-Lewis[a] monoclonal antibody (10 μg/ml; Chembiomed). Cells were then stained with fluorescein-conjugated goat anti-mouse IgM antibody (40 μg/ml; Sigma) and subjected to analysis by flow cytometry (9, 13, 16).

Northern and Southern Blot Analysis. A431 poly(A)$^+$ RNA (10 μg per lane) was subjected to Northern blot analysis as described (16). Genomic DNA (10 μg per lane) was subjected to Southern blot analysis as described (9). Blots were probed with a ^{32}P-labeled (14) 1.2-kb *Hin*fI fragment of pH3.4.

DNA Sequence Analysis. The insert in pCDM7-α(1,2)FT was sequenced by the method of Sanger *et al.* (18) using T7 DNA polymerase (Pharmacia) and 20-mer oligonucleotide primers synthesized according to the sequence of the cDNA insert. Sequence analyses and data base searches were performed using the Microgenie package (Beckman) and the Sequence Analysis software package of the University of Wisconsin Genetics Computer Group (19).

Assay of α(1,2)FT Activity. Cell extracts, conditioned medium from transfected COS-1 cells, and IgG-Sepharose-bound enzyme were prepared and assayed for α(1,2)FT activity by methods described (10, 13). One unit of α(1,2)FT activity is defined as 1 pmol of product formed per hr. The apparent Michaelis constant for the acceptor phenyl β-D-galactoside (20) was determined exactly as described (10).

Construction and Analysis of a Protein A–α(1,2)FT Fusion Vector. A 3196-base-pair *Stu* I–*Xho* I segment of the cDNA insert representing the putative catalytic domain and 3′-untranslated sequences was isolated from pCDM7-α(1,2)FT. This fragment was blunt-ended using the Klenow fragment of DNA polymerase I and ligated to phosphorylated (17) and annealed oligonucleotides (CGGAATTCCCCACATGGCC-TAGG and CCTAGGCCATGTGGGGAATTCCG) designed to reconstruct the coding sequence between the putative transmembrane segment proximal to the *Stu* I site. The ligated fragment was gel-purified, digested with *Eco*RI, and gel-purified again. This *Eco*RI-"linkered" fragment was ligated into the unique *Eco*RI site of pPROTA (21). One plasmid, designated pPROTA-α(1,2)FT$_c$, containing a single insert in the correct orientation, was analyzed by DNA sequencing to confirm the sequence across the vector, linker, and insert junctions. Plasmids pPROTA-α(1,2)FT$_c$, pPROTA, pCDM7-α(1,2)FT, or pCDM7 were transfected into COS-1 cells. After a 72-hr expression period, α(1,2)FT activities in the medium and associated with cells were quantitated as described (10, 13, 16). Affinity chromatography of conditioned medium was performed exactly as described (13, 16).

RESULTS

We have isolated (9, 10) a cloned human genomic DNA restriction fragment whose presence correlates with *de novo* expression of an α(1,2)FT in a set of stably transfected mouse L cells. This fragment determines α(1,2)FT expression in COS-1 cells transfected with a plasmid vector containing these sequences (plasmid pH3.4, ref. 10). The results of these analyses are consistent with the hypothesis that this segment represents a structural gene that encodes the H blood group α(1,2)FT. Nonetheless, these observations are also consistent with the possibility that this segment trans-determines enzyme expression by interaction with an endogenous gene, transcript, or protein. To discriminate between these possibilities and to characterize the nature of the genomic sequences, we first isolated various restriction fragments from the insert in plasmid pH3.4 and tested these for their ability to identify transcripts in the H-expressing stable transfectants and in a human cell line (A431) that also expresses H determinants and a cognate α(1,2)FT (9, 10). We found that a 1.2-kb *Hin*fI restriction fragment identifies a single relatively nonabundant 3.6-kb transcript in A431 cells (Fig. 1). This probe also detects transcripts in the H-expressing mouse L cell transfectants but not in the nontransfected parental L cells (R.D.L. and J.B.L., unpublished data).

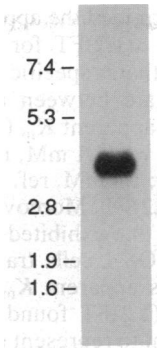

FIG. 1. Northern blot analysis. Poly(A)$^+$ RNA (10 μg) isolated from A431 cells was subjected to Northern blot analysis. The blot was probed with the ^{32}P-labeled 1.2-kb *Hin*fI fragment of pH3.4 (10). The mobilities of RNA molecular size standards, in kb, are indicated at left.

A Cloned cDNA That Directs Expression of Cell Surface H Structures and an α(1,2)FT. We used the 1.2-kb *Hin*fI fragment and colony hybridization to isolate two hybridization-positive cDNA clones from an A431 cell cDNA library (13). To test the cloned cDNAs for their ability to determine expression of surface-localized H antigen and a cognate α(1,2)FT activity, a plasmid was constructed [pCDM7-α(1,2)FT] that consisted of the largest cDNA insert cloned into the mammalian expression vector pCDM7 (15, 16) in the sense orientation with respect to the vector enhancer-promoter sequences. Flow cytometry analysis of COS-1 cells transfected with pCDM7-α(1,2)FT indicates that this cDNA determines expression of cell surface H molecules (Fig. 2). Moreover, COS-1 cells transfected with pCDM7-α(1,2)FT, but not cells transfected with pCDM7, express substantial

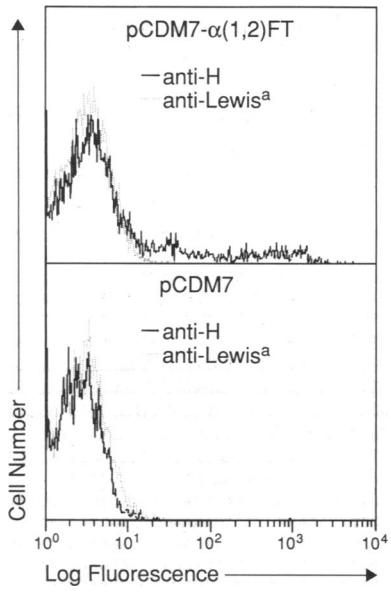

FIG. 2. Flow cytometry analysis of transfected COS-1 cells. COS-1 cells were transfected with plasmid pCDM7-α(1,2)FT (*Upper*) or with the control vector plasmid pCDM7 (*Lower*) and then stained with murine monoclonal IgM antibodies specific for the H antigen (solid lines) or for a negative control antigen (Lewis[a], dotted lines). The cells were then stained with a fluorescein-conjugated goat anti-mouse IgM antibody and subjected to flow cytometry analysis.

quantities of an α(1,2)FT activity. We determined the apparent Michaelis constant exhibited by this α(1,2)FT for an artificial acceptor, phenyl β-D-galactoside, that is specific for this enzyme (20) and that can discriminate between the human H and SE α(1,2)FTs (10, 22). This apparent K_m (2.4 mM) is nearly identical to the apparent K_m we (3.1 mM, ref. 10) and others (4.6 mM and 6.4 mM, ref. 22; 1.4 mM, ref. 23) have determined for the blood group H α(1,2)FT. Moreover, this apparent K_m is also very similar to the one exhibited by the α(1,2)FT in extracts prepared from COS-1 cells transfected with pH3.4 (4.4 mM, ref. 10). This apparent K_m is distinct from the one exhibited by an α(1,2)FT found in human milk (15.1 mM, ref. 10) that is thought to represent the α(1,2)FT encoded by the *SE* locus (22). These data demonstrate that the cDNA in plasmid pCDM7-α(1,2)FT determines expression of an α(1,2)FT whose kinetic properties reflect those exhibited by the human H blood group α(1,2)FT.

The cDNA Sequence Predicts a Type II Transmembrane Glycoprotein. The cDNA insert in pCDM7-α(1,2)FT is 3373 base pairs long (Fig. 3). Its corresponding transcript is 3.6 kb long (Fig. 1), suggesting that this cDNA is virtually full-length. Two potential initiator codons are found within its first 175 nucleotides. Only the second of these, however, is embedded within a sequence context associated with mammalian translation initiation (24). This methionine codon initiates a long open reading frame that predicts a protein of 365 amino acids, with a calculated M_r of 41,249. Hydropathy analysis (25) of the predicted protein sequence indicates that it is a type II transmembrane protein (26), as noted for several other cloned glycosyltransferases (for review, see ref. 27). This topology predicts an 8-residue NH_2-terminal cytosolic domain, a 17-residue hydrophobic transmembrane domain flanked by basic amino acids, and a 340-amino acid COOH-terminal domain that is presumably Golgi-resident and catalytically functional (27). Two potential N-glycosylation sites are found in this latter domain (Fig. 3), suggesting that this sequence, like other glycosyltransferases, may exist as a glycoprotein. No significant similarities were found between this sequence and other sequences in protein or DNA data bases (Protein Identification Resource, release 21.0, and GenBank, release 60.0), with the exception of a 642-base-pair sequence within the 3'-untranslated segment of the cDNA (Fig. 3) that is similar to the human *Alu* consensus sequence (28). Moreover, we identified no significant sequence similarities between this cDNA sequence or its predicted protein sequence and those of other cloned glycosyltransferase cDNAs (13, 16, 29–32).

```
-103 GCCTGGCGTTCCAGGGGCGGCCGGATGTGGCCTGCCTTTGCGGAGGGTGCGCTCCGGCCACGAAAAGCGGACTGTGGATCTGCCACCTGCAAGCAGCTCGGCC

  1  M  W  L  R  S  H  R  Q  L  C  L  A  F  L  L  V  C  V  L  S  V  I  F  F  L  H  I  H  Q  D  S  F  P  H  G  L  G  L  S  I
  1 ATGTGGCTCCGGAGCCATCGTCAGCTCTGCCTGGCCTTCCTGCTAGTCTGTGTCCTCTCTGTAATCTTCTTCCTCCATATCCATCAAGACAGCTTTCCACATGGCCTAGGCCTGTCGATC

 41  L  C  P  D  R  R  L  V  T  P  P  V  A  I  F  C  L  P  G  T  A  M  G  P  Ⓝ  A  S  S  S  C  P  Q  H  P  A  S  L  S  G  T
121 CTGTGTCCAGACCGCCGCCTGGTGACACCCCCAGTGGCCATCTTCTGCCTGCCGGGTACTGCGATGGGCCCCAACGCCTCCTCTTCCTGTCCCCAGCACCCTGCTTCCCTCTCCGGCACC

 81  W  T  V  Y  P  N  G  R  F  G  N  Q  M  G  Q  Y  A  T  L  L  A  L  A  Q  L  N  G  R  R  A  F  I  L  P  A  M  H  A  A  L
241 TGGACTGTCTACCCCAATGGCCGGTTTGGTAATCAGATGGGACAGTATGCCACGCTGCTGGCTCTGGCCCAGCTCAACGGCCGCCGGGCCTTTATCCTGCCTGCCATGCATGCCGCCCTG

121  A  P  V  F  R  I  T  L  P  V  L  A  P  E  V  D  S  R  T  P  W  R  E  L  Q  L  H  D  W  M  S  E  E  Y  A  D  L  R  D  P
361 GCCCCGGTATTCCGCATCACCCTGCCCGTGCTGGCCCCAGAAGTGGACAGCCGCACGCCGTGGCGGGAGCTGCAGCTTCACGACTGGATGTCGGAGGAGTACGCGGACTTGAGAGATCCT

161  F  L  K  L  S  G  F  P  C  S  W  T  F  F  H  H  L  R  E  Q  I  R  R  E  F  T  L  H  D  H  L  R  E  E  A  Q  S  V  L  G
481 TTCCTGAAGCTCTCTGGCTTCCCCTGCTCTTGGACTTTCTTCCACCATCTCCGGGAACAGATCCGCAGAGAGTTCACCCTGCACGACCACCTTCGGGAAGAGGCGCAGAGTGTGCTGGGT

201  Q  L  R  L  G  R  T  G  D  R  P  R  T  F  V  G  V  H  V  R  R  G  D  Y  L  Q  V  M  P  Q  R  W  K  G  V  V  G  D  S  A
601 CAGCTCCGCCTGGGCCGCACAGGGGACCGCCCGCGCACCTTTGTCGGCGTCCACGTGCGCCGTGGGGACTATCTGCAGGTTATGCCTCAGCGCTGGAAGGGTGTGGTGGGCGACAGCGCC

241  Y  L  R  Q  A  M  D  W  F  R  A  R  H  E  A  P  V  F  V  V  T  S  N  G  M  E  W  C  K  E  N  I  D  T  S  Q  G  D  V  T
721 TACCTCCGGCAGGCCATGGACTGGTTCCGGGCACGGCACGAAGCCCCCGTTTTCGTGGTCACCAGCAACGGCATGGAGTGGTGTAAAGAAAACATCGACACCTCCCAGGGCGATGTGACG

281  F  A  G  D  G  Q  E  A  T  P  W  K  D  F  A  L  L  T  Q  C  N  H  T  I  M  T  I  G  T  F  G  F  W  A  A  Y  L  A  G  G
841 TTTGCTGGCGATGGACAGGAGGCTACACCGTGGAAAGACTTTGCCCTGCTCACACAGTGCAACCACACCATTATGACCATTGGCACCTTCGGCTTCTGGGCTGCCTACCTGGCTGGCGGA

321  D  T  V  Y  L  A  Ⓝ  F  T  L  P  D  S  E  F  L  K  I  F  K  P  E  A  A  F  L  P  E  W  V  G  I  N  A  D  L  S  P  L  W
961 GACACTGTCTACCTGGCCAACTTCACCCTGCCAGACTCTGAGTTCCTGAAGATCTTTAAGCCGGAGGCGGCCTTCCTGCCCGAGTGGGTGGGCATTAATGCAGACTTGTCTCCACTCTGG

361  T  L  A  K  P  *
1081 ACATTGGCTAAGCCTTGAGAGCCAGGGAGACTTTCTGAAGTAGCCTGATCTTTCTAGAGCCAGCAGTACGTGGCTTCAGAGGCCTGGCATCTTCTGGAGAAGCTTGTGGTGTTCCTGAAG

1201 CAAATGGGTGCCCGTATCCAGAGTGATTCTAGTTGGGAGAGTTGGAGAGAAGGGGGACGTTTCTGGAACTGTCTGAATATTCTAGAACTAGCAAAACATCTTTTCCTGATGGCTGGCAGG
1321 CAGTTCTAGAAGCCACAGTGCCCACCTGCTCTTCCCAGCCCATATCTACAGTACTTCCAGATGGCTGCCCCCAGGAATGGGGAACTCTCCCTCTGGTCTACTCTAGAAGAGGGGTTACTT
1441 CTCCCCTGGGTCCTCCAAAGACTGAAGGAGCATATGATTGCTCCAGAGCAAGCATTCACCAAGTCCCCTTCTGTGTTTCTGGAGTGATTCTAGAGGGAGACTTGTTCTAGAGAGGACCAG
1561 GTTTGATGCCTGTGAAGAACCCTGCAGGGCCCTTATGGACAGGATGGGGTTCTGGAAATCCAGATAACTAAGGTGAAGAATCTTTTTAGTTTTTTTTTTTTTTTTTGGAGACAGGGTCT
1681 CGCTCTGTTGCCCAGGCTGGAGTGCAGTGGCGTGATCTTGGCTCACTGCAACTTCCGCCTCCTGTGTTCAAGCGATTCTCCTGTCTCAGCCTCCTGAGTAGATGGGACTACAGGCACAGG
1801 CCATTATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCGGGATGGTCTCGATCTCCTGACCTTGTCATCCACCTGTCTTGGCCTCCCAAAGTGCTGGGA
1921 TTACTGGCATGAGCCACTGTGCCCAGCCCGGATATTTTTTTTAATTATTTATTTATTTATTTATTGAGACGGAGTCTTGCTCTGTAGCCCAGGCCAGAGTGCAGTGGCGCGATCT
2041 CAGCTCACTGCAAGCTCTGCCTCCCGGGTTCATGCCATTCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCGCCCGCCACCACGCCCGGCTAATTTTTTTTGTATTTTTAGTAGAGA
2161 CGGGGTTTCATCGTGTTAACCAGGATGGTCTCGATCTCCTGACCTCGTGATCTGCCCACCTCGGCCTCCCACAGTGCTGGGATTACCGGCGTGAGCCACCATGCCTGGCCCGGATAATTT
2281 TTTTTAATTTTTGTAGAGACGAGGTCTTGTGATATTGCCCAGGCTGTTCTTCAACTCCTGGGCTCAAGCAGTCCTCCCACCTTGGCCTCCCAGAATGCTGGGTTTATAGATGTGAGCCAG
2401 CACACCGGGCCAAGTGAAGAATCTAATGAATGTGCAACCTAATTGTAGCATCTAATGAATGTTCCACCATTGCTGGAAAAATTGAGATGGAAAACAAACCATCTCTAGTTGGCCAGCGTC
2521 TTGCTCTGTTCACAGTCTCTGGAAAAGCTGGGGTAGTTGGTGAGCAGAGCGGGACTCTGTCCAACAAGCCCCACAGCCCCTCAAAGACTTTTTTTTGTTTGTTTTGAGCAGACAGGCTAA
2641 AATGTGAACGTGGGGTGAGGGATCACTGCCAAAATGGTACAGCTTCTGGAGCAGAACTTTCCAGGGATCCAGGGACACTTTTTTTTAAAGCTCATAAACTGCCAAGAGCTCCATATATTG
2761 GGTGTGAGTTCAGGTTGCCTCTCACAATGAAGGAAGTTGGTCTTTGTCTGCAGGTGGGCTGCTGAGGGTCTGGGATCTGTTTTCTGGAAGTGTGCAGGTATAAACACACCCTCTGTGCTT
2881 GTGACAAACTGGCAGGTACCGTGCTCATTGCTAACCACTGTCTGTCCCTGAACTCCCAGAACCACTACATCTGGCTTTGGGCAGGTCTGAGATAAAACGATCTAAAGGTAGGCAGACCCT
3001 GGACCCAGCCTCAGATCCAGGCAGGAGCACGAGGTCTGGCCAAGGTGGACGGGGTTGTCGAGATCTCAGGAGCCCCTTGCTGTTTTTTGGAGGGTGAAAGAAGAAACCTTAAACATAGTC
3121 AGCTCTGATCACATCCCCTGTCTACTCATCCAGACCCCATGCCTGTAGGCTTATCAGGGAGTTACAGTTACAATTGTTACAGTACTGTTCCCAACTCAGCTGCCACGGGTGAGAGAGCAG
3241 GAGGTATGAATTAAAAGTCTACAGCACTAA
```

FIG. 3. DNA and derived polypeptide sequence of the cDNA insert in pCDM7-(1,2)FT. The amino acid sequence is shown in single-letter code. The hydrophobic segment representing the putative transmembrane domain is double underlined. Asparagine residues that represent potential N-glycosylation sites are circled. The two copies of a sequence homologous to the human *Alu* consensus sequence are underlined. Not shown are 16 additional deoxyadenine residues found at the 3' end of the insert that represent a portion of the transcript's poly(A) tail.

The Protein Encoded by the cDNA Is an α(1,2)FT. The results of the expression experiments presented above, when considered together with the domain structure predicted by the cDNA sequence, are consistent with the presumption that it encodes an α(1,2)FT. Nonetheless, we wished to directly confirm this and thus exclude the possibility that it instead encodes a molecule that trans-determines this enzyme activity. We, therefore, fused the putative catalytic domain of the predicted protein to a secreted form of the IgG-binding domain of *Staphylococcus aureus* protein A in the mammalian expression vector pPROTA (21), to yield the vector pPROTA-α(1,2)FT$_c$ (Fig. 4). By analogy to similar constructs we have prepared with other cloned glycosyltransferases (13, 16), we expected that, if the cDNA sequence actually encodes an α(1,2)FT, then plasmid pPROTA-α(1,2)FT$_c$ would generate a secreted, soluble, and affinity-purifiable α(1,2)FT. Indeed, conditioned medium prepared from a plate of COS-1 cells transfected with pPROTA-α(1,2)FT$_c$ contained a total of 5790 units of α(1,2)FT activity, whereas a total of 1485 units were found to be cell-associated. Moreover, virtually 100% of the released α(1,2)FT activity was specifically retained by IgG-Sepharose, and most could be recovered after exhaustive washing of this matrix (Table 1). By contrast, we found that most of the activity in COS-1 cells transfected with pCDM7-α(1,2)FT was cell-associated (3450 units), with only trace amounts of activity in the conditioned medium prepared from these cells (≈80 units). Virtually none of this latter activity bound to either matrix (Table 1). Extracts prepared from COS-1 cells transfected with vector pCDM7 or vector pPROTA did not contain any detectable cell-associated or released α(1,2)FT activity. These data demonstrate that the cDNA insert in pCDM7-α(1,2)FT encodes an α(1,2)FT and that information sufficient to generate a catalytically active α(1,2)FT is encompassed within the 333 amino acids distal to the putative transmembrane segment.

The cDNA Corresponds to Genomic Sequences Syntenic to the H Locus on Human Chromosome 19. Genetic evidence indicates that expression of the human H α(1,2)FT is determined by a locus on chromosome 19 (33, 34). By using the 1.2-kb *Hin*fI probe, we identified a cross-hybridizing 6.5-kb *Eco*RI restriction fragment in the genome of the Chinese hamster ovary × human somatic cell hybrid line UV5HL9-5 (Fig. 5, lane 1) that contains human chromosome 19 as its only detectable human DNA (11). This fragment comigrates with a 6.5-kb *Eco*RI restriction fragment detectable in human genomic DNA (Fig. 5, lane 3) but absent from the hybrid parent Chinese hamster ovary cell line (Fig. 5, lane 2). The assignment of these sequences to human chromosome 19 was independently confirmed by Southern blot analysis of a pair of karyotypically stable (35) mouse 3T3 × human somatic cell hybrids (KLEJ-47 and KLEJ-47/P1, ref. 12) that differ only in their human chromosome 19 complement (data not shown). These results were also confirmed by Southern blot analysis of a commercial panel of Chinese hamster ovary × human somatic cell hybrid DNAs (BIOS) (data not shown). These observations support the results of the transfection experiments indicating that the cloned cDNA encodes the human H blood group α(1,2)FT.

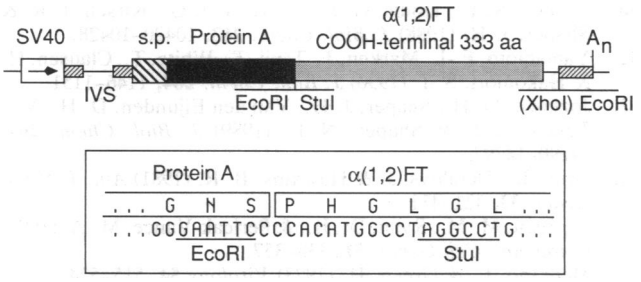

FIG. 4. Protein A–α(1,2)FT fusion vector. The vector pPROTA-α(1,2)FT$_c$ contains amino acids 33–365, representing the putative α(1,2)FT catalytic domain encoded by pCDM7-α(1,2)FT, fused in-frame with the IgG binding domain of *S. aureus* protein A. SV40, simian virus 40 early gene promoter sequences. Sequences denoted by ▨ indicate segments of the vector derived from rabbit β-globin sequences including an intervening sequence (IVS) and a polyadenylylation signal (An). s.p., Transin signal peptide. The *Xho* I (destroyed during the construction, in parentheses) and *Stu* I restriction sites used to isolate the catalytic domain from pCDM7-α(1,2)FT are depicted below the vector cartoon. The DNA sequence and the derived amino acid sequence across the protein A–α(1,2)FT junction are shown in the inset. The *Eco*RI and *Stu* I sites derived from the synthetic linker are underlined.

Table 1. Affinity chromatography of α(1,2)FT activity released from transfected COS-1 cells

Vector	α(1,2)FT activity, units					
	IgG-Sepharose			Sepharose		
	Applied	Spn	Bound	Applied	Spn	Bound
pCDM7-α(1,2)FT	≈30	≈50	<1	≈30	≈80	<1
pPROTA-α(1,2)FT$_c$	2316	<1	1464	2316	2136	<1

Conditioned medium from COS-1 cells transfected with pCDM7-α(1,2)FT or with pPROTA-α(1,2)FT$_c$ was chromatographed on IgG-Sepharose or Sepharose. Unbound (Spn) and matrix-retained materials (Bound) were assayed for α(1,2)FT activity (10, 13, 16).

Our previous observations indicated that the 3.4-kb *Eco*RI fragment in the plasmid pH3.4 (10) and detected in the genomes of H-expressing mouse L cell transfectants (9) was responsible for determining α(1,2)FT expression. Sequence analysis of this fragment and of the 6.5-kb *Eco*RI fragment identified in these Southern blot experiments indicates that the 3.4-kb segment is encompassed within the 6.5-kb human *Eco*RI fragment, which was apparently truncated at a position on the 3′ side of the coding sequences during the transfection process (R.D.L., L.K.E., and J.B.L., unpublished data).

DISCUSSION

Genetic and biochemical evidence indicates that the human genome encodes at least two discrete α(1,2)FT activities thought to represent the products of two distinct loci (*H* and *SE*) closely linked on human chromosome 19 (33, 34). A third distinct α(1,2)FT activity may also be expressed by human cells (36). Isolation of cloned genes or cDNAs encoding these molecules has not been possible because these enzymes are found in small amounts and are difficult to purify. The isolation of the α(1,2)FT cDNA described here was made possible by a gene-transfer approach (9, 10) designed to isolate genes that determine α(1,2)FT expression without the need to first purify the enzyme. Although it remains to be demonstrated by formal linkage analysis that this cDNA represents the human H blood group locus, we nonetheless

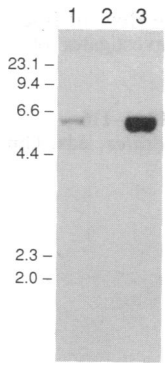

FIG. 5. Southern blot analysis of somatic cell hybrids. Genomic DNA samples prepared from various cell lines were digested with *Eco*RI and subjected to Southern blot analysis. The blot was probed with the ^{32}P-labeled 1.2-kb *Hin*fI fragment of pH3.4 (10). Mobilities of DNA molecular size standards, in kb, are indicated at left. Lanes: 1, somatic cell hybrid line UV5HL9-5; 2, Chinese hamster ovary cell parent of UV5HL9-5 hybrid; 3, human peripheral blood leukocytes.

believe the kinetic analyses reported here and elsewhere (10) plus the chromosomal localization studies provide very strong support for this assignment. Structural and functional analyses of null alleles isolated from rare H-negative individuals (Bombay and para-Bombay phenotypes, ref. 1) should also contribute to our understanding of this gene.

It appears that, in general, glycosyltransferases exist as Golgi-resident membrane-anchored molecules as well as secreted, soluble, and catalytically active forms thought to be derived from the membrane-bound precursors by intracellular proteolytic cleavage (27, 29). Our transfection studies using the cloned α(1,2)FT cDNA indicate, however, that only trace amounts of α(1,2)FT activity are released from COS-1 cells. This observation differs from our results with two other cloned glycosyltransferase cDNAs (13, 16) that determine significant quantities of released soluble enzyme activities when transfected into COS-1 cells. Apparent lack of α(1,2)FT release by transfected COS-1 cells is also at odds with the observation that the H blood group α(1,2)FT can generally be detected in human serum (10, 22, 23). Resolution of these apparent discrepancies will await biosynthetic studies designed to establish the structure(s) of polypeptides (catalytically active or not) encoded by transfected glycosyltransferase cDNAs and subsequently retained or released from the transfected cells.

The cDNA sequence predicts a type II transmembrane glycoprotein whose domain structure appears to be topologically and functionally identical to other cloned glycosyltransferases (13, 16, 27). However, we found no significant primary sequence similarities between this fucosyltransferase and other glycosyltransferase sequences, including those that utilize identical oligosaccharide acceptor molecules [α(1,3)galactosyltransferase, refs. 16 and 32; α(2,6)-sialyltransferase, ref. 29] or sugar nucleotide substrates [human α(1,3/1,4)FT, ref. 13]. These observations are in keeping with other glycosyltransferase sequence comparisons (29–32) as well as our analyses (13, 16) and suggest that the structural basis for substrate recognition by glycosyltransferases is not necessarily predicated upon generic protein domains with specificity for distinct oligosaccharide acceptors or nucleotide sugar substrates. Indeed, we have noted (37) substantial primary sequence similarity between a murine α(1,3)galactosyltransferase (16) and a human α(1,3)*N*-acetylgalactosaminyltransferase (31) that exhibit distinct nucleotide sugar and oligosaccharide acceptor substrate requirements. Nevertheless, low-stringency Southern blot analyses using the α(1,2)FT cDNA described here and other cloned glycosyltransferase sequences (J.B.L., unpublished data) suggest that structural similarities may exist within distinct classes of glycosyltransferases. The outcome of cloning experiments designed to determine the structures and test the function(s) of such cross-hybridizing sequences should determine whether this is indeed the case.

We thank Jeff Leiden for his review of this manuscript, Craig Thompson for helpful comments, and Brian Seed for the gift of the plasmid pCDM7. This work was supported by the Howard Hughes Medical Institute. Dr. Lowe is an Assistant Investigator of the Howard Hughes Medical Institute.

1. Watkins, W. M. (1980) *Adv. Hum. Genet.* **10**, 1–116.
2. Sadler, J. E. (1984) in *Biology of Carbohydrates*, eds. Ginsburg, V. & Robbins, P. W. (Wiley, New York), Vol. 2, pp. 199–213.
3. Szulman, A. E. (1962) *J. Exp. Med.* **115**, 977–996.
4. Szulman, A. E. (1964) *J. Exp. Med.* **119**, 503–523.
5. Fenderson, B. A., Holmes, E. H., Fukushi, Y. & Hakomori, S.-I. (1986) *Dev. Biol.* **114**, 12–21.
6. Fenderson, B. A., Zehavi, U. & Hakomori, S.-I. (1984) *J. Exp. Med.* **160**, 1591–1596.
7. Bird, J. M. & Kimber, S. J. (1984) *Dev. Biol.* **104**, 449–460.
8. Eggens, I., Fenderson, B., Toyokuni, T., Dean, B., Stroud, M. & Hakomori, S.-I. (1989) *J. Biol. Chem.* **264**, 9476–9484.
9. Ernst, L. K., Rajan, V. P., Larsen, R. D., Ruff, M. M. & Lowe, J. B. (1989) *J. Biol. Chem.* **264**, 3436–3447.
10. Rajan, V. P., Larsen, R. D., Ajmera, S., Ernst, L. K. & Lowe, J. B. (1989) *J. Biol. Chem.* **264**, 11158–11167.
11. Thompson, L. H., Bachinski, L. L., Stallings, R. L., Dolf, G., Weber, C. A., Westerveld, A. & Siciliano, M. J. (1989) *Genomics* **5**, 670–679.
12. Miller, D. A., Miller, O. J., Dev, V. G., Hashmi, S., Tantravahi, R., Medrano, L. & Green, H. (1974) *Cell* **1**, 167–173.
13. Kukowska-Latallo, J. F., Larsen, R. D., Nair, R. P. & Lowe, J. B. (1990) *Genes Dev.*, in press.
14. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
15. Seed, B. (1987) *Nature (London)* **329**, 840–842.
16. Larsen, R. D., Rajan, V. P., Ruff, M. M., Kukowska-Latallo, J., Cummings, R. D. & Lowe, J. B. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8227–8231.
17. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Methods in Molecular Biology* (Elsevier, New York).
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
20. Chester, M. A., Yates, A. D. & Watkins, W. M. (1976) *Eur. J. Biochem.* **69**, 583–593.
21. Sanchez-Lopez, R., Nicholson, R., Gesnel, M.-C., Matrison, L. M. & Breathnach, R. (1988) *J. Biol. Chem.* **263**, 11892–11899.
22. Kumazaki, T. & Yoshida, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4193–4197.
23. Le Pendu, J., Cartron, J. P., Lemieux, R. U. & Oriol, R. (1985) *Am. J. Hum. Genet.* **37**, 749–760.
24. Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241.
25. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
26. Wickner, W. T. & Lodish, H. F. (1985) *Science* **230**, 400–407.
27. Paulson, J. C. & Colley, K. J. (1989) *J. Biol. Chem.* **264**, 17615–17618.
28. Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S. & Matsubara, K. (1987) *Gene* **53**, 1–10.
29. Weinstein, J., Lee, E. U., McEntee, K., Lai, P.-H. & Paulson, J. C. (1987) *J. Biol. Chem.* **262**, 17735–17743.
30. Shaper, N. L., Hollis, G. F., Douglas, J. G., Kirsch, I. R. & Shaper, J. H. (1988) *J. Biol. Chem.* **263**, 10420–10428.
31. Yamamoto, F.-I., Marken, J., Tsuji, T., White, T., Clausen, H. & Hakomori, S.-I. (1990) *J. Biol. Chem.* **264**, 1146–1151.
32. Joziasse, D. H., Shaper, J. H., Van den Eijunden, D. H., Van Tunen, A. J. & Shaper, N. L. (1989) *J. Biol. Chem.* **264**, 14290–14297.
33. Oriol, R., Danilovs, J. & Hawkins, B. R. (1981) *Am. J. Hum. Genet.* **33**, 421–431.
34. Le Beau, M. M., Ryan, D., Jr., & Pericak-Vance, M. A. (1989) *Cytogenet. Cell Genet.* **51**, 338–357.
35. Medrano, L. & Green, H. (1973) *Virology* **54**, 515–524.
36. Blaszczyk-Thurin, M., Sarnesto, A., Thurin, J., Hindsgaul, O. & Koprowski, H. (1988) *Biochem. Biophys. Res. Commun.* **151**, 100–108.
37. Larsen, R. D., Rivera-Marrero, C. A., Ernst, L. K., Cummings, R. D. & Lowe, J. B. (1990) *J. Biol. Chem.* **265**, 7055–7061.