Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4720–4724, July 1986
Biochemistry

Cloning and sequencing of cDNA of bovine *N*-acetylglucosamine (β1-4)galactosyltransferase

HISASHI NARIMATSU*, SUDHIR SINHA†, KEITH BREW†, HIROTO OKAYAMA‡, AND PRADMAN K. QASBA*§

*Laboratory of Tumor Immunology and Biology, National Cancer Institute, Bethesda, MD 20205; †Department of Biochemistry, University of Miami, Miami, FL 33124; and ‡Laboratory of Molecular Biology, National Institute of Child Health and Human Development, Bethesda, MD 20205

Communicated by Igor B. Dawid, March 20, 1986

ABSTRACT **Galactosyltransferases constitute a family of enzymes, each member of which transfers galactose from UDPgalactose to a specific acceptor molecule, generating a specific galactose-acceptor linkage. Two synthetic oligonucleotides, 27mer and 21mer, were synthesized, based on the amino acid sequences of two peptides derived from bovine milk *N*-acetylglucosaminide (β1-4)galactosyltransferase (EC 2.4.1.90), and used as hybridization probes to isolate cDNA clones for galactosyltransferase from a bovine mammary gland cDNA library. One of the plasmids, designated pLbGT-1, contains an insert of about 3.7 kilobases that hybridizes to both of the probes and encodes the amino acid sequences of five peptides obtained from bovine milk (β1-4)galactosyltransferase. A second plasmid, designated pLbGT-2, contains an insert of about 4.1 kilobases that hybridizes to only the 27mer and that encodes a polypeptide containing the sequence of the carboxyl-terminal 120 residues identical to the peptide encoded by pLbGT-1; the rest of the protein sequence, however, does not contain known sequences from bovine galactosyltransferase. The two cDNAs contain a 3′-untranslated region of about 2.7 kilobases that includes two copies of the *Alu*-equivalent sequences. pLbGT-1 and pLbGT-2 hybridize to mRNAs of various sizes obtained from the bovine and rat mammary gland and the human mammary tumor cell line MCF-7, with the longest mRNA from each species being around 4.5 kilobases. The results show that pLbGT-1 is a cDNA clone for bovine (β1-4)galactosyltransferase, and pLbGT-2 encodes a protein that is structurally and may be functionally related to transferases.**

Glycosyltransferases are a family of enzymes that are involved in the synthesis and extension of the oligosaccharide chains of glycoproteins and glycolipids (for review, see refs. 1–3). A specific transferase is required for the synthesis of each of the different disaccharide linkages that are known to be present, so that the number of these transferases must be large. Galactosyltransferases, a subgroup of the transferases, have in common the ability to utilize UDPgalactose, but differ in the specificity of the acceptor molecule or residue and the type of linkage formed. The most common carbohydrate sequence, *N*-acetyllactosamine [Gal(β1-4)GlcNAc] or its repeating unit, which is recognized as a developmentally regulated antigenic determinant (for review, see ref. 4), is formed by *N*-acetylglucosamine (β1-4)galactosyltransferase (EC 2.4.1.90; catalytic component of lactose synthase or UDPgalactose:D-glucose 4-β-D-galactosyltransferase; EC 2.4.1.22). This enzyme transfers galactose from UDPgalactose to *N*-acetylglucosamine, generating a $\beta1 \rightarrow 4$ glycosidic linkage. When α-lactalbumin [a protein homologous to the type C lysozymes but with no enzymatic activity of its own (5, 6)] interacts with the galactosyltransferase, the monosaccharide binding property of the enzyme is modified so that glucose, normally a poor acceptor substrate, becomes a good substrate while transfer to glycoproteins and oligosaccharide substrates is inhibited (for review, see refs. 7 and 8). During lactation the protein levels of (β1-4)galactosyltransferase as well as α-lactalbumin are elevated in the mammary gland. Both proteins are found in milk from which they have been purified and characterized (9–12).

Because of the importance of the transferases on whose selective activities the identity of each type of cell and its distinct interaction depend on, an understanding of their nature at the structural and genetic level is of great interest. We report here the molecular cloning of (β1-4)galactosyltransferase and show by the sequence analyses of the cDNA clones that the region of its mRNA coding for the carboxyl-terminal domain is identical with another cDNA clone that may code for another transferase.

MATERIALS AND METHODS

Peptide Sequences from Bovine Galactosyltransferase. Bovine milk galactosyltransferase obtained from Sigma was further purified by gel filtration with Sephadex G-100 followed by affinity chromatography with α-lactalbumin coupled to Sepharose 4B (13). The gel filtration step was essential for removing contaminating immunoglobulins (14). This preparation of the enzyme was acetylated with [^{3}H]acetic anhydride, reduced, carboxymethylated, and subjected to digestion with trypsin as described (15). Peptides were initially subjected to group separation by gel filteration with Sephadex G-50 and equilibrated with 0.1 M ammonium bicarbonate; pools were further fractionated by reverse-phase HPLC using an Aquapore RP-300 column and eluted with linear gradient of 0–80% (vol/vol) acetonitrile containing 0.1% trifluoroacetic acid. Five peptides were subjected to automatic sequence analysis with a Beckman 890c sequencer fitted with a cold trap and microprocessor controller. Phenylthiohydantoin amino acid derivatives were identified by HPLC as described (15).

Construction of cDNA Library. Total RNA was prepared from lactating bovine mammary gland by guanidinium thiocyanate method (16). Poly(A)$^+$ RNA was isolated by adsorption to and elution from oligo(dT)-cellulose (17). A cDNA library was constructed by the method of Okayama–Berg, using 7 μg of poly(A)$^+$ RNA and 1.5 μg of vector/primer DNA (18). Transformation of *Escherichia coli* χ 1776 with the vector-cDNA library gave 3×10^6 transformants resistant to ampicillin. A total plasmid preparation made from this library was used for transfection of *E. coli* MC1061, and a second library of 5×10^5 transformants was prepared and screened with oligonucleotide probes.

Oligonucleotide Probes. Two mixed oligonucleotide probes, 21 and 27 bases in length, were synthesized on a Vega Coder 300 by the phosphoramidite method (19), each con-

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: kb, kilobase(s).
§To whom reprint requests should be addressed.

taining the possible coding sequences for the two peptides of bovine (β1-4)galactosyltransferase, and were generously provided by Ettore Appella. The 21mer was a 128-fold degenerate mixture of the sequence 3' TTDCTYATYCTYATYTTYACY 5', and the 27mer was a 2048-fold degenerate mixture of the sequence 3' CTYTTDTTDTTYCTDGGNTTYGGNGTD 5', where D is T/C, Y is A/G, and N is A/T/G/C. The oligonucleotide mixtures were purified by gel electrophoresis on 20% polyacrylamide/8 M urea gels, and subsequently labeled at the 5' end with [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq) by using T4 polynucleotide kinase (20).

Library Screening. Transformants (6×10^4) were screened by the replica plating method of Hanahan and Meselson (21). Hybridization with the 21mer and the 27mer was carried out at 42°C and 48°C, respectively, in 6× NET [1× NET = 0.15 M NaCl, 0.015 M Tris·HCl (pH 7.5), 0.005 M EDTA, 0.5% Nonidet P-40] (22). Filters were washed at 46°C and 52°C, in 6× SSC, respectively. (1× SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0.)

DNA Sequence Analysis. Plasmid preparations, restriction endonuclease digestions, and end-labeling of fragments were carried out as described (20). Nucleotide sequence analysis was carried out by the Maxam–Gilbert method (23) and by the chain-termination method (24), using restriction fragments of cDNAs subcloned in the phage M13 vectors, mp18 and mp19 (25).

Poly(A)$^+$ RNA Transfer Blot Analysis. Poly(A)$^+$ RNA was isolated from lactating mammary glands from cows, Sprague–Dawley rats, BALB/c mice, and cultured MCF-7 cells. Poly(A)$^+$ RNA samples and marker DNA samples were lyophilized and denatured by heating at 60°C for 10 min in 50% (vol/vol) formamide/6% (vol/vol) formaldehyde/20 mM Mops/5 mM NaOAc/1 mM EDTA and fractionated by electrophoresis in 1% agarose/6% formaldehyde gels (20). The fractionated RNA was transferred to nitrocellulose papers and hybridized for 18 hr to nick-translated ^{32}P-labeled pLbGT-1 cDNA in 50% (vol/vol) formamide/6× SSC/1× Denhardt's solution/1% $NaDodSO_4$/sonicated-denatured salmon sperm DNA (100 μg/ml) at 42°C for the bovine sample and 39°C for the other RNA samples. Blots were washed three times in 2× SSC for 20 min at room temperature. Bovine RNA blots were finally washed three times at 67°C in 0.1× SSC for 15 min. Other RNA blots were washed three times at 0.5× SSC for 15 min at 55°C.

Hybrid Selection, Translation of mRNA, and Immunoprecipitation. Hybrid selection of mRNA was carried out as described (26). Poly(A)$^+$ RNA, 100 μg each from lactating bovine mammary gland and MCF-7 cells, was hybridized with pLbGT-1 DNA filters at 50°C, and 42°C for 8 hr, respectively, in 300 μl of 65% (vol/vol) formamide/0.4 M NaCl/0.2% $NaDodSO_4$/30 mM Pipes, pH 8.5, containing 50 μg of tRNA. Filters were then washed as described (26), except for the MCF-7 sample, which was washed at 55°C. Hybridized RNA was eluted from the filters and translated with micrococcal nuclease-treated rabbit reticulocyte lysate [obtained from Promega Biotec (Madison, WI)] under conditions suggested by the manufacturer, in the presence of [^{35}S]methionine and in a final volume of 50 μl. Aliquots of the resulting labeled products were immunoprecipitated (27) with 10 μl of an ascites fluid containing monoclonal antibody, $4C_6$, $6C_2$, and $11D_1$, to human galactosyltransferase (28). Antigen/antibody complex was bound to Pansorbin (Calbiochem) pretreated with anti-mouse IgG, and the bound complex was eluted as described (26). Proteins synthesized *in vitro* from hybrid-selected mRNA and the immunoprecipitates were analyzed by electrophoresis on $NaDodSO_4$/8.5% polyacrylamide gels.

RESULTS

Peptide Sequences of Bovine (β1-4)Galactosyltransferase. The purified bovine milk galactosyltransferase migrated on $NaDodSO_4$/polyacrylamide gels as a single protein species of M_r 54,000. Since the amino-terminal sequence of bovine galactosyltransferase was found to be heterogeneous, we determined the partial sequences of some of the major peptides obtained by tryptic cleavage of the *N*-acetylated protein. These were the following: peptide 1, Lys-Asp-Tyr-Asp-Tyr-Asn-Cys; peptide 2, Asp-Lys-Lys-Asn-Glu-Pro-Asn-Pro-Gln; peptide 3, Gln-Gln-Leu-Asp-Tyr-Gly-Ile-Tyr; peptide 4, Lys-Val-Ala-Ile-Ile; peptide 5, Met-Asp-Lys. Two mixed oligonucleotide probes, 21 and 27 bases long, were synthesized based on the sequences of peptides 1 and 2, respectively.

Isolation and Characterization of cDNA Clones. About 60,000 transformants, derived from a cDNA library constructed from lactating bovine mammary gland poly(A)$^+$ RNA in the Okayama–Berg vectors, were screened using labeled mixed oligonucleotides, 21mer and 27mer, as hybridization probes. Twelve cDNA clones were identified that sorted into the following three groups: four clones hybridized to both the probes; one clone hybridized only to the 27mer; and seven clones hybridized only to the 21mer probes. The cDNA inserts of the plasmids from the clones of the first and second group had no *Bam*HI endonuclease restriction site. Since the *Bam*HI restriction sites flank the cDNA inserts in the Okayama–Berg vectors, this restriction enzyme was used to isolate the inserts. The sizes of the inserts from the first group varied between 3.4 and 3.7 kilobases (kb), and their restriction endonuclease maps were identical. The longest plasmid was designated as pLbGT-1. The clone of the second group, designated as pLbGT-2, contains a 4.1-kb insert and has, for the most part, a restriction map very similar to that of pLbGT-1, except at the 5' region (Fig. 1). The insert of pLbGT-1 hybridized with the inserts of the plasmid DNAs from the clones of the first and second groups, but not with the inserts of the plasmid DNAs from the clones of third group, with which only the 21mer hybridizes. Subsequent DNA sequence analyses of some of the clones of the third group at the site where the 21mer hybridized showed that they contain one of the primer sequences present in the mixed 21mer but had entirely different DNA sequences compared to pLbGT-1 and pLbGT-2, which were analyzed further in detail.

DNA Sequence Analysis. We confirmed the presence of sequences corresponding to those of the oligonucleotide probes, 21mer and 27mer, in pLbGT-1 by determining the sequence of the relevant region of the cDNA insert. Fig. 2 shows 2 kb of DNA sequence from the 5' end of the 3.7-kb insert of pLbGT-1. In one of the three possible reading frames, the DNA sequence encodes a protein sequence of 328 residues that contains not only the amino acid sequences corresponding to the 21mer and 27mer but also encodes those of the other peptides obtained from bovine milk (β1-4)galactosyltransferase (Fig. 2). The codon specifying the serine at position 328 is followed by the translation termination codon TAG (nucleotides 982–987 in Fig. 2). The encoded protein sequence, which lacks the amino-terminal methionine and, therefore, does not encode the complete sequence of bovine (β1-4)galactosyltransferase, contains two potential *N*-glycosylation sites, ^{17}Asn-Ser-Ser and ^{44}Asn-Leu-Thr, which follow the consensus sequence Asn-Xaa-Ser/Thr (29). The cDNA has, however, an unusually long 3'-noncoding sequence of about 2.7 kb. The DNA sequence after the termination codon TAG, from nucleotide 988 to nucleotide 1452 (Fig. 2) (as well as in pLbGT-2), encodes in the same open reading frame a protein of 155 residues. This DNA sequence, and the sequence in the opposite orientation

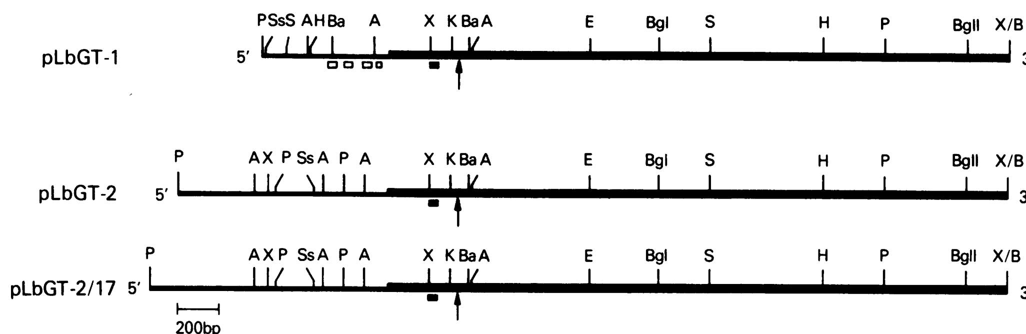

FIG. 1. Restriction endonuclease map of the cDNA inserts of pLbGT-1, pLbGT-2, and pLbGT-2/17. The thick line represents the region common between the three inserts. Open boxes, regions corresponding to the amino acid sequences of the tryptic peptides of (β1-4)galactosyltransferase; solid box, region where the 27mer hybridized. Arrows indicate the site of the translational termination codons. The letters P, Ss, S, A, H, Ba, X, K, E, BgI, BgII, B represent *Pst* I, *Sst* I, *Sma* I, *Ava* II, *Hin*dIII, *Bal* I, *Xho* I, *Kpn* I, *Eco*RI, *Bgl* I, *Bgl* II, and *Bam*HI restriction sites, respectively.

between *Hin*dIII and *Pst* I restriction sites near the 3′ end (not shown), bear similarities to the *Alu* family sequences (30). These sequences, which are about 340 nucleotides long, have a stretch of poly(A) (A_{13} from nucleotide 1310 to 1324) and are flanked by short partial repeats (Fig. 2). Computer-assisted comparison of the DNA sequences with those of the *Alu* family sequences further supported the conclusion that they are *Alu*-equivalent sequences.

The DNA sequence encoding the peptide corresponding to the 27mer is located just adjacent to the common *Xho* I site in pLbGT-1 and pLbGT-2 (Fig. 2, nucleotides 826–852). Comparison of the detailed restriction endonuclease maps of pLbGT-1 and pLbGT-2 showed that a similar sequence of 3.1 kb is present in the two cDNAs, which includes 2.7 kb of the 3′-noncoding region and 360 bases of the coding sequence that encodes the carboxyl-terminal 120 residues of the protein, including the peptide corresponding to 27mer (Fig. 1). To ensure that the pLbGT-2 is not the only clone of this type, we screened 300,000 additional clones from bovine cDNA library of 5×10^5 transformants with the 5′-end probes of pLbGT-1 and pLbGT-2. The ratio of the clones corresponding to these two types remained 4:1 as was observed in the initial screening of 60,000 clones from which pLbGT-1 and pLbGT-2 were isolated. The additional clones of pLbGT-2 type differ in length at the 5′ and 3′ ends, but contain common carboxyl-terminal coding and 3′-noncoding regions. The longest of the type 2 clones, pLbGT-2/17, is shown in Fig. 2.

To determine the exact nucleotide sequence of the region of the DNA where pLbGT-2 sequence diverges from pLbGT-1, relevant fragments of pLbGT-2 near *Xho* I, *Kpn* I, and *Eco*RI sites were subcloned in mp18 and mp19, sequenced, and compared with the DNA sequence of pLbGT-1. The nucleotide sequence of pLbGT-2 was identical with the sequence of pLbGT-1 from nucleotides 624 to 2000, but distinct toward the 5′ end; a comparison of the two sequences at the point of divergence is shown in Fig. 3. The nucleotide

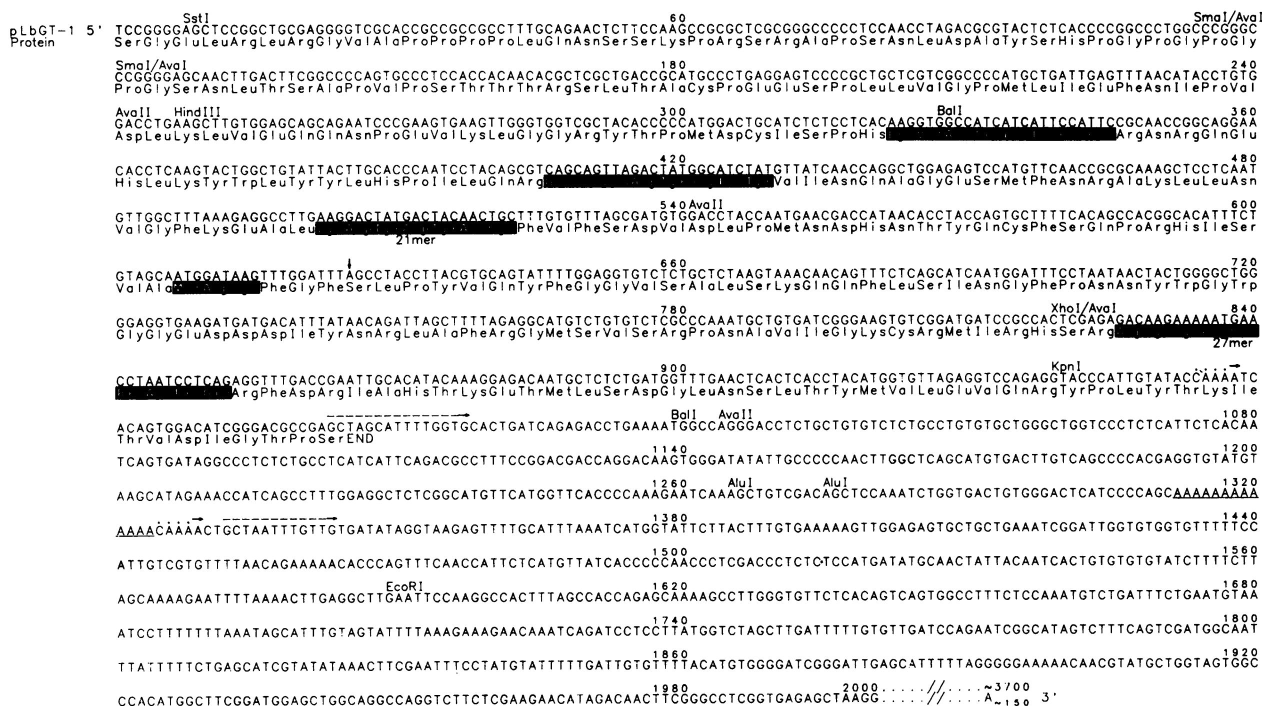

FIG. 2. Partial nucleotide sequence of the 3.7-kb cDNA insert from pLbGT-1 and the predicted partial amino acid sequence of the bovine (β1-4)galactosyltransferase. A translation frame was chosen that aligned the nucleotide sequence with the peptide sequence (black boxes) of (β1-4)galactosyltransferase. In the *Alu*-equivalent sequence, a stretch of A_{13} is underlined, and the partial repeat sequences are shown by dashed lines. The nucleotide sequence of pLbGT-2 was identical with the sequence of pLbGT-1, from nucleotides 625 (indicated by arrow) to 2000, but different toward the 5′ end.

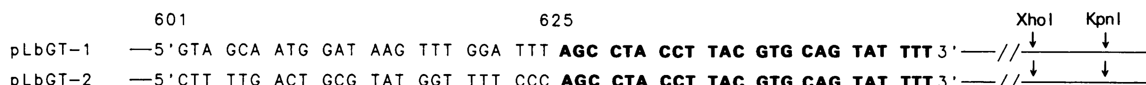

FIG. 3. Nucleotide sequence of pLbGT-1 and pLbGT-2 in the region of divergence. Numbers correspond to the nucleotide numbers of pLbGT-1.

sequence near the restriction sites in the divergent region of pLbGT-2 (Fig. 2) showed no similarities with the sequence of the corresponding region of pLbGT-1. The DNA sequences at various divergent regions encode, in one of the open reading frames, peptide sequences, which are different from those of the protein encoded in the divergent part of pLbGT-1 (data not shown). The inserts of the two plasmids, however, encode protein sequences that have 120 identical residues in the carboxyl-terminal domain. The DNA sequence data show that pLbGT-1 encodes (β1-4)galactosyltransferase, whereas pLbGT-2 encodes a protein that is partially related to this transferase.

RNA Analysis by Blot Hybridization with pLbGT-1. To determine the size of the galactosyltransferase mRNA, total poly(A)$^+$ RNA, extracted from lactating mammary glands from different species and from the human mammary tumor cell line MCF-7, were analyzed in blotting experiments (Fig. 4). The size of the main hybridizable RNA band in both bovine and human species is about 4.5 kb. In bovine RNA two additional hybridizable bands corresponding to 2.2 and 1.8 kb were detected. In the RNA from rat (Fig. 4, lane 3) and mouse (data not shown) lactating mammary gland, two major hybridizable bands were seen, which correspond to mRNAs of about 4.2 and 2.5 kb. These hybridizable bands were clearly visible even after washing the blots under more stringent conditions [0.2× SSC at 55°C (20)]. Beneath these major hybridizable bands a smear was seen in all the samples. These results show that, using bovine cDNA probes, the homologous RNA sequences of multiple sizes are detected in all the other species tested. The longest mRNA in these species is also around 4.5 kb.

Hybrid Selection and Translation. To confirm that the RNA band hybridizing with MCF-7 poly(A)$^+$ RNA is mRNA for (β1-4)galactosyltransferase, we carried out positive translation analyses. Translation products synthesized in a reticulocyte cell-free translational system programmed with the hybrid-selected mRNAs were immunoprecipitated by three different monoclonal antibodies to human (β1-4)galactosyltransferase (28). The translational products and immunoprecipitates were analyzed on NaDodSO$_4$/polyacrylamide gels (Fig. 5). The polypeptides synthesized with the bovine and MCF-7 hybrid-selected mRNAs were similar in size, but not identical. The largest peptide synthesized was about M_r 65,000. The three monoclonal antibodies to human galactosyltransferase, available to us, immunoprecipitated only those polypeptides synthesized with the MCF-7 mRNA samples (Fig. 5, lane 4) but not those synthesized with bovine hybrid-selected mRNA. These results show that pLbGT-1 contains sequences that are homologous to the human galactosyltransferase mRNA.

DISCUSSION

We have initiated the molecular cloning approach to address the question of whether the family of galactosyltransferases have common structural and sequence features despite their varying specificities. In this paper we report the isolation and characterization of two sequence-related cDNA clones, pLbGT-1 and pLbGT-2, and conclude that pLbGT-1 encodes bovine *N*-acetylglucosamine (β1-4)galactosyltransferase. Several lines of evidence support this conclusion: (*i*) The nucleotide sequence of the 2-kb region from the 5′ end of the cDNA insert of pLbGT-1 encodes a protein sequence that includes the amino acid sequences of five peptides from this enzyme (Fig. 2). (*ii*) When pLbGT-1 or its 5′-end fragment was used as a hybridization probe in the RNA blot analysis, multiple mRNA sizes were detected in the poly(A)$^+$ RNA from various species, and among them was an mRNA that was as long as 4.5 kb (Fig. 4). These RNAs were about 10-fold more abundant in the lactating mammary glands of rat and mouse than in liver (data not shown), which is consistent with the observation that lactating mammary glands have high levels of this enzyme. (*iii*) The *in vitro*-translated products coded by human mRNA hybrid-selected with pLbGT-1 were immunologically reactive with monoclonal antibodies (Fig. 4) that recognized human but not the bovine (β1-4)galactosyltransferase epitopes (ref. 28, and unpublished data), providing additional evidence that pLbGT-1 contains sequences

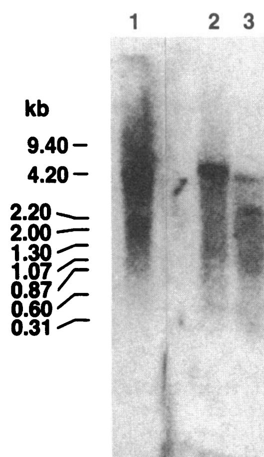

FIG. 4. RNA transfer blot analysis of poly(A)$^+$ RNA from cow (lane 1) and rat (lane 2) lactating mammary glands and from MCF-7 cells (lane 3). RNA was separated on a 1% agarose/6% formaldehyde gel and blotted onto nitrocellulose filter. The filter strips containing the bovine and the remaining samples were separately hybridized to the ^{32}P-labeled pLbGT-1 probe, washed, and exposed to x-ray film.

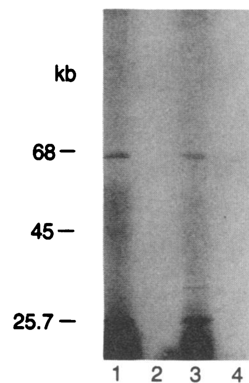

FIG. 5. Autoradiographs of 8.5% polyacrylamide gels showing [^{35}S]methionine-labeled polypeptides synthesized in a rabbit reticulocyte cell-free system programmed with RNA selected by hybridization to pLbGT-1 from the poly(A)$^+$ RNA of bovine lactating mammary gland (lane 1) and of MCF-7 cells (lane 3). Immunoprecipitates of the translation products programmed with the RNA from bovine lactating mammary gland or MCF-7 cells (lanes 2 and 4, respectively).

homologous to the mRNA of human (β1-4)galactosyltransferase.

The largest peptide that was synthesized in an *in vitro* translational system programmed with the hybrid-selected mRNAs was about M_r 65,000 (Fig. 5). These results and the RNA blot analysis indicate that a cDNA insert of about 4.5 kb would be necessary to obtain a full-length clone coding for the largest peptide. The *in vivo*-synthesized (β1-4)galactosyltransferase is a glycoprotein that exists in several molecular forms. The enzyme that can be solubilized from Golgi membranes has an apparent molecular weight of 65,000 (12), whereas those of the soluble forms found in milk and other secretions range between 42,000 and 55,000 (10, 11). Low molecular weight *in vivo* forms may be generated by the proteolytic cleavage of the larger peptide, as proposed earlier (11, 12). However, the presence of multiple mRNAs with variable sizes (4.5, 2.2, and 1.8 kb from cows) and with sequence differences at the 5′ end observed in the present study raises the possibility that some of the smaller size proteins may directly be synthesized from different size mRNAs.

The nucleotide sequence of the 4.1-kb insert of pLbGT-2 at the 5′-end region is different from that of pLbGT-1 but identical between positions 625 and 2000, which includes 360 bases of the coding sequence and 1015 bases of the 3′-noncoding sequence (Fig. 2). The remaining 3′-untranslated region, based on the detailed restriction endonuclease analyses and DNA sequence analyses around various restriction sites, also appears to be identical with that of pLbGT-1. The 3′-noncoding region of both clones has two *Alu*-equivalent sequences present in opposite orientations and an open reading frame encoding a protein of 155 residues. The biological significance of these sequence elements is not clear at this time. The divergent region of pLbGT-2 hybridizes with the RNA that has a size identical to that of pLbGT-1 mRNA, suggesting that this sequence is not a part of any accumulated species of a precursor mRNA that should have a size larger than that of pLbGT-1 mRNA. The coding sequence of pLbGT-2 encodes in one open reading frame a protein sequence that has a 120-residue carboxyl-terminal domain identical with (β1-4)galactosyltransferase. These results and the fact that the pLbGT-2 is not a single clone (since we have isolated additional clones of this type) make it at present less likely that pLbGT-2 is a clone representing a partially spliced mRNA. Nevertheless, the possibility that the divergent region of pLbGT-2 is a part of an intervening sequence has to be left open since the sequence at the junction of the common and the divergent region contains a dinucleotide AG (Fig. 3) that is a part of the consensus sequence for the splice junction (31).

Evidence to date suggests that (β1-4)- and (β1-3)galactosyltransferases may share some common peptides (32). The two blood group transferases, enzyme A (*N*-acetylgalactosaminyltransferase) and enzyme B [(α1-3)galactosyltransferase], which transfer two different sugar moieties to the same blood group substance H [Fuc(α-1-2)Gal(β1-4)GlcNAc-], also have peptides in common (33). pLbGT-2 may be a cDNA clone encoding a different transferase. Our cDNA library is constructed in the pcD expression vector (18), and a full-length cDNA clone after transfection into COS cells has the potential to express its genetic coding sequences (34). By determining the enzymatic activities of the coded proteins, it should be possible to assign a given cDNA clone to a particular transferase. The two mRNAs, corresponding to pLbGT-1 and pLbGT-2, may have been generated by differential splicing (35), by DNA rearrangement in some cell populations [as in immunoglobulins and T-cell receptors (36)], or each coded by a different gene. The results of the preliminary genomic analyses of the bovine and human DNA by the DNA blot analysis, using pLbGT-1 as a hybridization probe, suggest that this DNA does not represent a multigene family. Future analyses of genomic clones and assignment of individual cDNA clones to specific transferases may indicate the mechanism(s) involved in the synthesis of a large number of transferases with some common structures and different enzymatic specificities.

Note Added in Proof. Following submission of this manuscript, a cDNA clone encoding partial bovine galactosyltransferase sequence was reported (37). The protein sequence predicted from this cDNA is same as encoded by pLGT-1 with some differences in the variable region. A cDNA clone with identical sequence in the variable region as reported in ref. 37, among several other clones, has also been isolated by us.

We thank Drs. Teruyuki Kimura and Shioko Kimura for their help with M13 sequencing technique. S.K.S. and K.B. were supported by Grant GM 213-63 from the National Institutes of Health.

1. Beyer, T. A., Sadler, J. E., Rearick, J. I., Paulson, J. C. & Hill, R. L. (1981) *Adv. Enzymol.* **52**, 23–175.
2. Beyer, T. A. & Hill, R. L. (1982) *Glycoconjugates* **3**, 25–45.
3. Roth, S. (1985) *Q. Rev. Biol.* **60**, 145–153.
4. Feizi, T. (1985) *Nature (London)* **314**, 53–57.
5. Brew, K., Vanaman, T. C. & Hill, R. L. (1967) *J. Biol. Chem.* **242**, 3747–3749.
6. Qasba, P. K. & Safaya, S. K. (1984) *Nature (London)* **308**, 377–380.
7. Brew, K. & Hill, R. L. (1975) *Rev. Physiol. Biochem. Pharmacol.* **72**, 105–158.
8. Hill, R. L. & Brew, K. (1975) *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 411–490.
9. Brew, K., Castellino, F. J., Vanaman, J. C. & Hill, R. L. (1970) *J. Biol. Chem.* **245**, 4570–4582.
10. Magee, S. C., Mawal, R. & Ebner, K. E. (1973) *J. Biol. Chem.* **248**, 7565–7569.
11. Powell, J. T. & Brew, K. (1974) *Eur. J. Biochem.* **48**, 217–228.
12. Smith, C. A. & Brew, K. (1977) *J. Biol. Chem.* **252**, 7294–7299.
13. Barker, R., Osten, K. W., Shaper, J. H. & Hill, R. L. (1972) *J. Biol. Chem.* **247**, 7135–7147.
14. Wilson, J. R., Weiser, M. M., Albini, B., Schenck, J. R., Rittenhouse, H. G., Hirata, A. A. & Berger, E. G. (1982) *Biochem. Biophys. Res. Commun.* **105**, 737–744.
15. Sinha, S. & Brew, K. (1981) *J. Biol. Chem.* **256**, 4193–4204.
16. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **24**, 5294–5303.
17. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
18. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289.
19. Adams, S. P., Kavka, K. S., Wiks, E. J., Holder, S. B. & Galupi, G. R. (1983) *J. Am. Chem. Soc.* **105**, 661–663.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
21. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
22. Wallace, B. R., Johnson, T. M., Hirose, T., Miyake, T., Kawashima, H. E. & Itakura, K. (1981) *Nucleic Acids Res.* **9**, 879–893.
23. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
26. Dandekar, A. M. & Qasba, P. K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4853–4857.
27. Lingappa, V. R., Lingappa, J. R., Prasad, R., Ebner, K. E. & Blobel, G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2338–2342.
28. Chatterjee, S. K., Bhattacharya, M. & Barlow, J. J. (1984) *Cancer Res.* **44**, 5725–5732.
29. Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A. & Lennarz, W. J. (1979) *J. Biol. Chem.* **254**, 9747–9754.
30. Schmid, C. W. & Jelnick, W. R. (1982) *Science* **216**, 1065–1070.
31. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
32. Furukawa, K. & Roth, S. (1985) *Biochem. J.* **227**, 573–582.
33. Nagai, M., Muensch, H. & Yoshida, A. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 1441 (abstr.).
34. Yokota, T., Lee, F., Rennick, D., Hall, C., Arai, N., Mosmann, T., Nabel, G., Cantor, H. & Arai, K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1070–1074.
35. de Ferra, F., Engh, H., Hudson, L., Kamholz, J., Puckett, C., Molineaux, S. & Lazzarini, R. A. (1985) *Cell* **43**, 721–727.
36. Lewis, S., Gifford, A. & Baltimore, D. (1985) *Science* **228**, 677–685.
37. Shaper, N. L., Shaper, H. J., Meuth, L. J., Fox, L. J., Chang, H., Kirsch, I. R. & Hollis, G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1573–1577.