Vol. 139, No. 1, 1986
August 29, 1986

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 163-168

ISOLATION OF A cDNA CODING FOR HUMAN GALACTOSYLTRANSFERASE

Hubert E. Appert*[1], Thomas J. Rutherford[1,2], George E. Tarr[3], Jon S. Wiest[2], Neil R. Thomford[1] and D. James McCorquodale[2]

[1]Medical College of Ohio at Toledo, Department of Surgery

[2]Medical College of Ohio at Toledo, Department of Biochemistry
C.S. 10008, Toledo, Ohio 43699

[3]The University of Michigan, Department of Biochemistry
Ann Arbor, Michigan 48109

Received July 17, 1986

Human milk galactosyltransferase (EC 2.4.1.22) was purified to homogeneity using affinity chromatography. Edman degradation was used to determine the amino acid sequences of eight peptide fragments isolated from the purified enzyme. A 60-mer "optimal" oligonucleotide probe that corresponded to the amino acid sequence of one of the galactosyltransferase peptide fragments was constructed and used to screen a λgt10 cDNA library. Two hybridization-positive recombinant phages, each with a 1.7 Kbp insert, were detected among 3 x 10^6 recombinant λgt10 phages. Sequencing of one of the cDNA inserts revealed a 783 bp galactosyltransferase coding sequence. The remainder of the sequence corresponded to the 3'-region of the mRNA downstream from the termination codon.

Glycosylation reactions are needed to confer biologic activity to a number of different protein and lipid molecules. Epidermal growth factor receptor requires post-translational glycosylation in order to bind epidermal growth factor (11). Tunicamycin, a glycosylation inhibitor, decreases the ability of insulin and growth hormone receptors to bind their respective ligands (5). A number of genetically controlled physiologic and pathologic conditions such as embryonic development and malignant transformation are associated with changing patterns of glycoconjugate synthesis (2, 10).

*To whom correspondence should be addressed.

Abbreviations: HPLC, High performance liquid chromatography. SSC, Sodium chloride/Sodium citrate buffer (7); SSPE, Sodium chloride/Sodium phosphate monobasic/EDTA buffer; SDS, Sodium dodecylsulfate; CNBr, Cyanogen bromide; DTT, Dithiothreitol.

0006-291X/86 $1.50

Little progress has been made in understanding the genetic regulation of glycoconjugate synthesis because gene probes are not available for that purpose. This investigation was consequently initiated to clone and to sequence the human galactosyltransferase cDNA in order to provide a DNA probe to study the organization of the galactosyltransferase gene and to study the expression of that gene at the mRNA level under various physiological and pathologic conditions. A human liver cDNA library was screened with a 60-mer "optimal" oligonucleotide probe whose sequence was predicted from the amino acid sequence of one of the CNBr cleavage fragments from purified galactosyltransferase, and two hybridization-positive clones were isolated.

METHODS

Purification of Galactosyltransferase

Galactosyltransferase was purified from human milk, as described previously (1) using an N-acetylglucosamine affinity column and successive passages through two α-lactalbumin Sepharose columns.

Amino Acid Sequencing of Galactosyltransferase

The affinity purified galactosyltransferase was reduced and alkylated, repurified by HPLC, and then cleaved with CNBr. The CNBr cleavage fragments were separated by reverse phase HPLC. Fragments up to 20 amino acid residues in length were sequenced manually as described by Tarr (12), and longer fragments were sequenced using the Applied Biosystems 470A Protein Sequencer. The PTH amino acids obtained from Edman degradation were analyzed by HPLC.

Construction and Radioactive Labeling of the Oligonucleotide Probe

The nucleotide sequence of a 60-mer "optimal" probe (Fig. 1a lower line) was deduced from the amino acid sequence of the CNBr cleavage fragment shown in Fig. 1a (upper line) using the method proposed by Lathe (6) to select "optimal" codons. The 60-mer probe was constructed by first synthesizing two 36-mer oligonucleotides with 12 complementary nucleotides at their 3' ends as

a

```
ProHisLysValAlaIleIleIleProPheArgAsnArgGlnGluHisLeuLysTyrTrp
CCCCACAAGGTGGCCATCATCATCCCATTCCGGAACCGGCAGGAGCATCTGAAGTACTGG
```

b

```
1st Strand (coding)
5-CCCCACAAGGTGGCCATCATCATCCCATTCCGGAAC-3'
                    3'-GGTAAGGCCTTGGCCGTCCTCGTAGACTTCATGACC-5'
                                        2nd Strand (non-coding)
```

Fig. 1. a, Amino acid sequence (upper line) of galactosyltransferase CNBr fragment and corresponding predicted nucleotide sequence (lower line) used to design 60-mer probe. b, Two chemically synthesized 36 residue-long coding (upper) and non-coding (lower) strands with 12 complementary nucleotides on their 3' ends. Synthesis, hybridization and fill-in reaction required for probe construction are described in text.

shown in Fig. 1b. The oligonucleotide synthesis was performed by the phosphoramidite method (3) using an Applied Biosystems DNA Synthesizer. The oligonucleotides were then purified by preparative polyacrylamide gel electrophoresis followed by gel filtration using Sephadex G-50.

Hybridizaton of the 36-mer oligonucleotides was conducted in a 10 µl volume containing 1X polymerase buffer (40 mM potassium phosphate, pH 7.4, 1 mM 2-mercaptoethanol, 6.2 mM $MgCl_2$), and 0.5 µg each of the two oligonucleotides. The solution of the two 36-mer oligonucleotides was heated to 90°C for 5 minutes, and cooled to room temperature over a period of 45 minutes. The fill-in reaction was performed in a final volume of 30 µl containing 10 µl of the hybridized oligonucleotide solution, 1 mM DTT, 1 unit of DNA polymerase I, and 250 µCi each of ^{32}P-labeled deoxyadenosine, deoxyguanosine, deoxythymidine and deoxycytidine triphosphates. The reaction mixture was incubated at room temperature for 10 minutes, and was chased for 10 minute at 25°C with 1 nanomole of each of the four non-radioactive deoxynucleoside triphosphates. The reaction was stopped by addition of 3 µl of 0.1 M EDTA (pH 7.5), diluted to 500 µl with a buffer containing 10 mM Tris-Cl, 5 mM NaCl, and 0.1 mM EDTA, pH 8.0, and applied to a 1 x 30 cm G-50 Sephadex column previously equilibrated with the same buffer. The probe eluted immediately following the break-through volume, and the peak radioactive fractions containing the probe were used without further preparation.

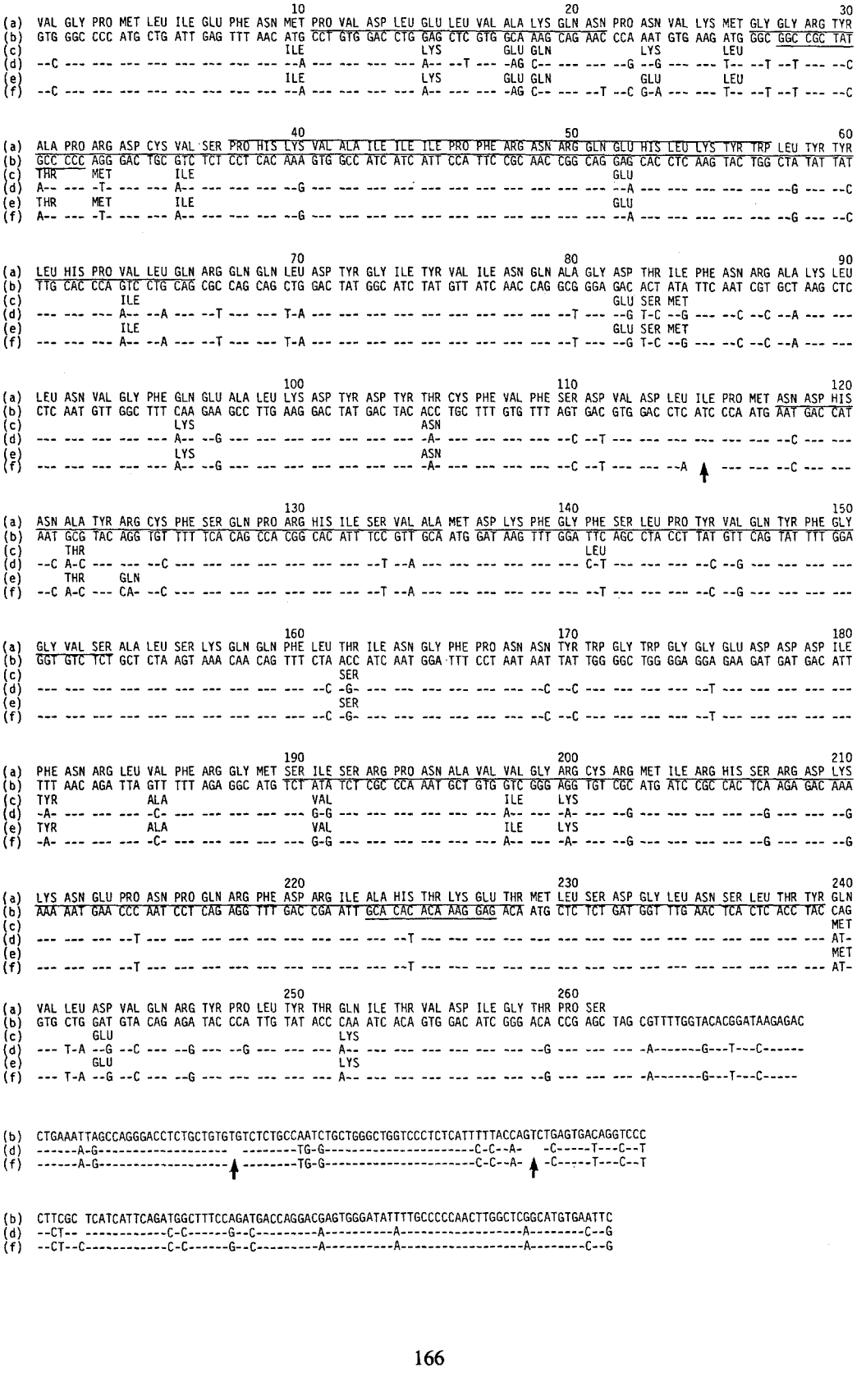
Screening of the λgt10 cDNA library

A λgt10 library, prepared from human liver mRNA as described previously (4, 13), was the generous gift of Axel Ullrich and Lisa Coussens, Genentech, Inc., South San Francisco, CA. The library was plated on Escherichia coli C600 Hfl at a density of 200,000 plaques per 23 x 23 cm plate for the initial screening. Subsequent screening was conducted on 100 mm diameter plates. Duplicate nitrocellulose lifts were made and hybridization with the 60-mer probe was conducted at 42°C for 14 hours in a solution containing 50% formamide, 2X Denhardt's solution, 5X SSPE, 100 µg/ml sonicated, denatured Salmon sperm DNA. The nitrocellulose lifts were washed with 2 x SSC containing 0.1% SDS for 30 minutes at room temperature, followed by a second wash at 37°C for 2 hours, and a final wash at 42°C for 1 hour. The hybridized filters were exposed to X-ray film for 2 days at -20°C. Plaques displaying positive signals on both lifts were selected for purification, which was accomplished by the fourth round of selection.

RESULTS AND DISCUSSION

Library Screening and cDNA Sequencing

Initial screening of 3 x 10^6 plaques with the 60-mer probe detected 2 hybridization-positive recombinant phages that were subsequently plaque-purified. Excision of cDNA from either of the two plaque-purified phages with EcoRI yielded three fragments, which were 1.0 and 0.35, and 0.4 Kbp long, for a total of 1.75 Kbp, as determined by electrophoresis in 0.8% agarose gels. Sequencing of the 1 Kb EcoRI fragment revealed a segment of cDNA, 982 base pairs in length, with an open reading frame containing 783 nucleotides that translated into an amino acid sequence which included 7 of the 8 partial amino acid sequences determined by Edman degradation of the



CNBr cleavage products (Fig. 2). The 8th sequence, which was shown previously (1) to be the N-terminal end of the soluble form of the enzyme, did not correspond to any nucleotide sequence in the cDNA isolate. Thus, the open reading frame codes for most of the amino acid sequence of the soluble form of the enzyme, but not the putative N-terminal membrane-bound portion, nor the translational initiation codon. A comparison of the determined cDNA sequence with the DNA sequence of the oligonucleotide probe, which was predicted according to the method of Lathe (6), revealed that the probe had 6 mismatches (90% homology) with the longest stretch of perfect matches being 14 base pairs in length (Fig. 3). This investigation confirms the observation of previous investigators (14, 15) that the use of long synthetic "optimal" probes for screening cDNA libraries is effective. A comparison with two recently published sequences for bovine galactosyltransferase cDNA (8, 9) indicates an 89% homology to that of human galactosyltransferase cDNA as shown in Fig. 2.

The amino acid sequences of CNBr cleavage fragments of the soluble form of human milk galactosyltransferase were 100% homologous with corresponding amino acid sequences defined by the nucleotide sequence of galactosyltransferase cDNA prepared from human liver mRNA. Since liver galactosyltransferase mRNA probably codes for the membrane-bound form of the enzyme, it appears that one species of mRNA codes for both the soluble and membrane bound forms of galactosyltransferase. This is consistent with the concept

Fig. 2. Nucleotide sequence of human galactosyltransferase cDNA (line b) and corresponding predicted amino acid sequence (line a). Amino acids determined by sequential Edman degradation are underlined. Nucleotide sequence of bovine cDNA (line c) and predicted amino acid sequence (line d) of bovine galactosyltransferase reported by Shaper et al. (9). Nucleotide sequence of bovine cDNA (line e) and predicted amino acid sequence (line f) of bovine galactosyltransferase reported by Narimatsu et al. (8). Places where bovine amino acid and nucleotide sequences differ from corresponding human sequences are indicated in lines c, d, e, and f. Dashes show where sequences reported by Shaper et al. and Narimatsu et al. are identical to human sequences. Additional nucleotides detected in the human nucleotide sequence compared to bovine sequence are indicated by arrows. Amino acids 38-57 (overlined) were used to predict the nucleotide sequence of the 60-mer probe. Sequencing of the 1 Kbp fragment of isolated human galactosyltransferase cDNA began with the universal M13 primer in the non-coding region. The dideoxy sequencing reaction was continued into the coding region by sequential priming with synthetic oligonucleotides at the places where the nucleotide sequence is underlined.

```
CCTCACAAAGTGGCCATCATCATTCCATTCCGCAACCGGCAGGAGCACCTCAAGTACTGG
  *     *           *       *           * *
CCCCACAAGGTGGCCATCATCATCCCATTCCGGAACCGGCAGGAGCATCTGAAGTACTGG
```

Fig. 3. Upper line, determined nucleotide sequence of human galactosyltransferase cDNA corresponding to region of probe. Lower line, nucleotide sequence of probe deduced from determined amino acid sequence and from Lathe's (6) optimization of codons. Asterisks indicate mismatches.

that soluble galactosyltransferase is a proteolytic product of the bound form of the enzyme.

ACKNOWLEDGMENTS

Supported by an ACS grant and Biomedical Research Grants (NIH) No. 2 S07 RR05700-15, -16 and AM 27982. The authors gratefully acknowledge the use of the Harold A. McMaster Recombinant DNA laboratory and thank Ms. S. Kinker for obtaining the human milk samples.

REFERENCES

1. Appert, H. E., Rutherford, T. J., Tarr, G. E., Thomford, N. R. and McCorquodale, D.J., (1986) Biochem. Biophys. Res. Comm, in Press.
2. Hakomori, S. and Kannagi, R., (1983) J. Nat. Cancer. Soc. 71, 231-251.
3. Heron, E. S. Am. Biotech. Lab. (1984) 2, 52-59.
4. Huynh, T. V., Young, R. A. and Davis, R. W. (1985) DNA Cloning. Vol. 1, pp. 49-78, IRL Press, Oxford.
5. Keefer, L. M. and DeMeyis, P (1981) Biochem, Biophys. Res. Comm. 22-29.
6. Lathe, R (1985) J. Mol. Biol. 183, 1-12.
7. Maniatis, T., Fritsch, E. F. and Sambrook (1983) Cold Spring Harbor Laboratory.
8. Narimatsu, H., Sinha, S., Brew, K., Okayama, H., Quasba, P. K. (1986) Proc. Natl. Acad. Sci. 83, 4720-4724.
9. Shaper, N. L., Shaper, J. H., Meuth, J. L., Fox, J. L., Chang, H., Kirsch, I. R., and Hollis, G. F. (1986) Proc. Natl. Acad. Sci. 83, 1573-1577.
10. Shur, B. D. Mol. Cell. Biochem (1984) 61, 143-158.
11. Slieker, L. J. and Lane, M. D. (1985) J. Biol. Chem. 260, 687-690.
12. Tarr, G. E. (1986) Microcharacterization of Polypeptides, pp. 155-194, Humana Press, Inc., Clifton, N.J.
13. Ullrich, A (1984) Nature. 307, 521-527.
14. Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y., Tsubokawa, M., Mason, A., Seeburg, P.H., Grunfeld, C., Rosen, O.M. and Ramachandran, (1985) J. Nature. 313, 756-761.
15. Ullrich, A., Berman, C. H., Dull, T. J., Gray, A., and Lee, J. M., (1984) EMBO J. 3, 361-364.