Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7962–7966, November 1987
Cell Biology

cDNA cloning for a bile canaliculus domain-specific membrane glycoprotein of rat hepatocytes

(λgt11 cDNA cloning/membrane glycoprotein)

WANJIN HONG AND DARRELL DOYLE

Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260

Communicated by Robert T. Schimke, July 27, 1987

ABSTRACT **Hepatocytes are polarized cells with distinct sinusoidal, bile canalicular, and basolateral plasma membrane domains. Each domain contains proteins that are specific for it. We have isolated three cDNA clones encoding a rat liver bile canaliculus domain-specific glycoprotein with M_r 110,000 (gp110) by immunologically screening a rat kidney λgt11 cDNA library with a rabbit polyclonal antiserum directed against purified gp110. The authenticity of these clones was verified as follows. (*i*) The antiserum recognizes specifically isopropyl β-D-thiogalactoside-induced fusion proteins on electrophoretic transfer blots of total lysogen lysates containing these cDNA clones. (*ii*) Antibodies epitope-selected by these clones are able to interact with gp110 on electrophoretic transfer blots. (*iii*) The amino acid sequencing derived from the DNA sequence was confirmed by amino acid sequencing of a tryptic peptide of gp110. Rescreening of the same library with the cDNA clones identified a full-length cDNA clone for this glycoprotein. Sequence analysis indicates that the N-linked carbohydrate chains are concentrated on the N-terminal part of this highly glycosylated protein.**

The plasma membranes of polarized cells, such as hepatocytes and epithelial cells of intestine and kidney tubules, are differentiated into morphologically and functionally distinct domains. The hepatocyte, for example, has a sinusoidal domain facing the circulation, a basolateral domain involved in cell–cell and cell–substratum interactions, and a bile canalicular domain comprised of the opposed membranes of two hepatocytes and involved in the transport of bile salts across the plasma membrane. Much attention has been devoted to the mechanism for the establishment and maintenance of these domains, especially of how domain-specific proteins are sorted into their specific localization during their biosyntheses. Many domain-specific membrane proteins have been identified biochemically and immunohistologically in polarized cells. In hepatocytes, the receptor for asialoglycoproteins has been localized to the sinusoidal domain, whereas leucine aminopeptidase and dipeptidyl peptidase IV have been localized specifically to the bile canalicular domain (1–4). Farquhar and coworkers have identified a few proteins that are localized on the apical plasmalemma domain of epithelial cells of kidney tubules. Interestingly, these apical membrane proteins are localized to specific microdomains of the apical membrane (5, 6). Similar studies have been done on the epithelial cells of intestine (7, 8).

We have localized three membrane glycoproteins to the bile canalicular domain of rat hepatocytes (9). We have characterized in more detail one of these proteins, a glycoprotein of M_r 110,000, denoted gp110 here. This protein may be involved in the transport and secretion of organic anions, including bile acids, into bile in liver (10, 11). Although gp110 was identified first in rat liver, our recent studies show that gp110 is expressed at a higher level in kidney than in liver (unpublished data). By immunologically screening a λgt11 rat kidney cDNA library, we have isolated three cDNA clones for gp110. A full-length cDNA clone was isolated* after rescreening the same library with the cDNA clones. In this report, we describe the purification of, the preparation of antiserum to, and the isolation and verification of cDNA clones for the bile canalicular domain-specific membrane glycoprotein.

MATERIALS AND METHODS

Purification of gp110. A membrane fraction denoted N_2 enriched in plasma membrane derived from the bile canaliculus was isolated as described (9). The membrane glycoproteins were purified by wheat germ agglutinin (WGA) affinity chromatography after extraction of the total membrane proteins from the N_2 membrane fraction with 10 mM Tris·HCl, pH 8.0/1% deoxycholate/1 mM phenylmethylsulfonyl fluoride. gp110 was then purified to homogeneity by preparative $NaDodSO_4$/PAGE from the WGA affinity-purified membrane glycoproteins as described (12).

Preparation and Characterization of Antiserum to gp110. A rabbit was immunized with 50–100 μg of gp110 in complete Freund's adjuvant and given booster injections 2 weeks later with the same amount of gp110 in incomplete Freund's adjuvant. After the booster injection, the rabbit was bled once a week for 6 weeks. The first two bleedings were used for the studies reported in this paper. The specificity of the antiserum was assessed by probing a total liver cell extract in 1% deoxycholate by electrophoretic transfer blot with the antiserum as described (13). All electrophoretic transfer blot analyses described here use the same detection method as described below for the isolation of cDNA clones.

Trypsin Digestion and Partial Amino Acid Sequencing. About 500 pmol of gp110 was digested with 1 μg of L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin (Sigma) in 100 mM NH_4HCO_3 with 2 mM $CaCl_2$ at 37°C for 24 hr. The tryptic peptides were then separated by reversed-phase HPLC on a C_{18} column (Beckman HPLC system) with a linear 60-min gradient from 0 to 100% solvent B. Solvent A was water with 0.1% trifluoroacetic acid. Solvent B was 60% acetonitrile in water with 0.1% trifluoroacetic acid. Selected peaks were then sequenced on a 470A gas-phase protein sequencer from Applied Biosystems (Foster City, CA).

Isolation and Characterization of cDNA Clones for gp110 with Antiserum. A Sprague–Dawley rat kidney λgt11 cDNA

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.

Abbreviations: WGA, wheat germ agglutinin; IPTG, isopropyl β-D-thiogalactoside.

*This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02997).

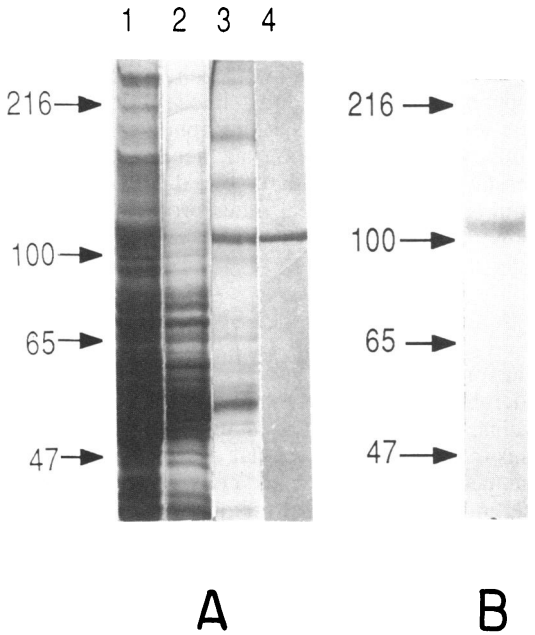

FIG. 1. (*A*) Purification of gp110. Total N_2 membrane proteins extracted with 1% deoxycholate (lane 1), WGA-Sepharose flow through (lane 2), WGA-Sepharose elution (lane 3), and purified gp110 (lane 4) were resolved by 7% $NaDodSO_4$/PAGE and stained with Coomassie blue. (*B*) Characterization of antiserum to gp110 by electrophoretic transfer blot analysis. A total hepatocyte lysate was run on a 7% $NaDodSO_4$/polyacrylamide gel, transferred to a nitrocellulose filter, and probed with the antiserum. Molecular weights are shown as $M_r \times 10^{-3}$.

library (Clontech, Palo Alto, CA) was screened with 1:100 diluted antiserum to gp110 (preabsorbed with Y1090 bacterial lysate) as described (14, 15). Lysogenization of Y1089 with cDNA clones and λgt11 and the analyses of the fusion proteins were as described (16). Selection and characterization of antibodies by the expressed fusion proteins were also performed as described (17–19).

Isolation of a Full-Length cDNA Clone for gp110. After sequencing λC1, -2, and -3, we found that they overlap each other and that λC3 has more 5′ sequence than λC1 and -2. A 285-base-pair (bp) 5′ *Eco*RI–*Pst* I fragment of λC3 was purified (20), oligolabeled (21, 22), and used as probe for rescreening the same λgt11 library (20). A cDNA clone with a large insert was identified as full length.

DNA Sequencing. The cDNA inserts were subcloned into M13mp19. A sequential series of overlapping deletions were constructed (23) and sequenced (24, 25). Both strands were sequenced.

RESULTS

Purification of gp110 and Preparation of Antiserum to It. A three-step scheme was used for the purification of gp110. First, a membrane fraction denoted N_2 was isolated by a combination of differential and discontinuous sucrose gradient centrifugation (9). This membrane fraction is enriched in membrane derived from the bile canaliculus (9). The membrane glycoproteins were purified by WGA affinity chromatography after extraction of total membrane proteins with deoxycholate. As seen in Fig. 1*A* (compare lanes 1, 2, and 3), this lectin affinity purification step removes much protein that is not glycosylated, leaving a few distinct glycoproteins. gp110 is the major glycoprotein after WGA affinity chromatography, as judged by Coomassie blue staining (Fig. 1*A*, lane 3). gp110 was then further purified by preparative $NaDodSO_4$/PAGE (Fig. 1*A*, lane 4). The purified gp110 was used to raise antibodies in a rabbit. The specificity of the antiserum was checked by electrophoretic transfer blot analysis. When probing a total lysate of hepatocytes, the antiserum recognizes specifically gp110 (Fig. 1*B*). When a [^{35}S]methionine-labeled hepatocyte lysate was immunoprecipitated with the antiserum, gp110 was specifically precipitated (38). The antiserum also stained specifically the bile canaliculus on rat liver tissue sections, as assayed by indirect immunofluorescence staining (38).

Immunological Isolation and Characterization of cDNA Clones for gp110. Our recent studies have shown that gp110 is expressed at a higher level in kidney than in liver (unpublished data). Therefore, we screened a λgt11 rat kidney cDNA library with the antiserum raised to gp110. By screening 1.0×10^6 independent clones, three positive clones with inserts of about 2.2 kilobases (kb), 1.8 kb, and 2.0 kb, respectively, were isolated after three consecutive rescreenings. They were named λC1, λC2, and λC3, respectively. The fusion proteins produced by these cDNA clones were analyzed by electrophoretic transfer blot, as shown in Fig. 2. IPTG induced these clones and an unrelated clone, λCX, to express fusion proteins (Fig. 2*A*, lanes 2, 4, 6, and 8) that have higher molecular weights than β-galactosidase induced from λgt11 (Fig. 2*A*, lane 10). When a duplicate gel of Fig. 2*A* was probed with the antiserum after transferring to a nitrocellu-

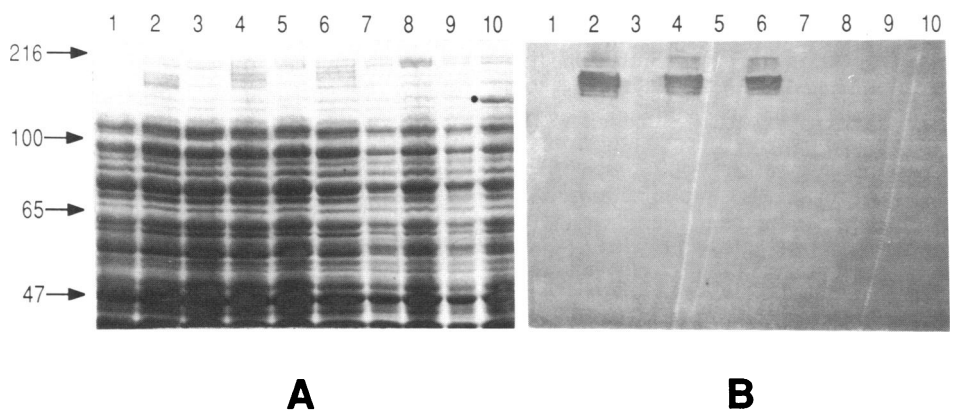

FIG. 2. Immunoblot analysis of fusion proteins encoded by the gp110 cDNA clones. Bacterial lysates were prepared from the lysogens of λC1 (lanes 1 and 2), λC2 (lanes 3 and 4), λC3 (lanes 5 and 6), λCX (lanes 7 and 8), and λgt11 (lanes 9 and 10). After inactivation of the temperature-sensitive repressor at 45°C, the lysogen cultures were either induced with isopropyl β-D-thiogalactoside (IPTG) (lanes 2, 4, 6, 8, and 10) or not induced (lanes 1, 3, 5, 7, and 9). Samples were resolved on 7% $NaDodSO_4$/polyacrylamide gels in duplicate. One gel was stained with Coomassie blue (*A*) and the other was transferred to a nitrocellulose filter and probed with the antiserum (*B*). The dot indicates the β-galactosidase. Molecular weights are shown as $M_r \times 10^{-3}$.

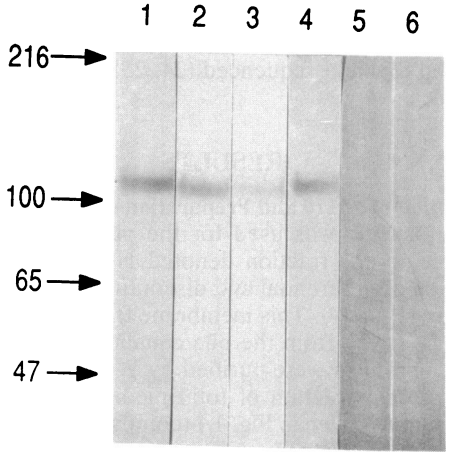

FIG. 3. Analysis of recombinant cDNA clones by epitope selection. λC1, λC2, λC3, λCX, and λgt11 were used to select antibodies, which were then used to probe gp110 on electrophoretic transfer blots. The antibodies used are as follows: crude antiserum (lane 1) and epitope-selected antibodies from λC1 (lane 2), λC2 (lane 3), λC3 (lane 4), λCX (lane 5), and λgt11 (lane 6). Molecular weights are shown as $M_r \times 10^{-3}$.

lose filter, only the fusion proteins induced from λC1, λC2, and λC3 by IPTG can be specifically recognized (Fig. 2*B*, lanes 2, 4, and 6). Neither the fusion protein induced from λCX (Fig. 2*B*, lane 8) nor the β-galactosidase induced from λgt11 (Fig. 2*B*, lane 10) showed any reaction with the antiserum. The fusion proteins produced by λC1, λC2, and λC3 migrate as multiple bands on NaDodSO$_4$/PAGE. This may be due to premature termination of transcription and/or translation of the fusion genes. Further characterization of these cDNA clones was done by epitope selection analyses, as shown in Fig. 3. Like the antiserum to gp110 (Fig. 3, lane 1), antibodies selected by the fusion proteins specified by λC1, λC2, and λC3 cDNA clones recognize gp110 on electrophoretic transfer blot (Fig. 3, lanes 2, 3, and 4, respectively). Antibodies selected by λCX and λgt11 clones do not interact with the gp110 by electrophoretic transfer blot analysis (Fig. 3, lanes 5 and 6). These results taken together suggest strongly that λC1, λC2, and λC3 are cDNA clones for gp110.

Isolation of a Full-Length cDNA Clone and Sequencing Analysis. The three cDNA clones were subcloned into M13mp19. A series of overlapping deletions were produced and sequenced. The three cDNA clones overlap each other and altogether give a total sequence of 2389 bp, as shown in Fig. 4. The 285-bp 5′ *Eco*RI–*Pst* I fragment of λC3 was purified and used to rescreen the same λgt11 library. One full-length cDNA clone with an insert of about 3.2 kb (λC4) was identified and sequenced as above. The full-length cDNA sequence and its derived amino acid sequence are shown in Fig. 4. The 5′ noncoding region consists of 28 nucleotides. From the first ATG codon to the TGA termination codon (nucleotides 29–2404), an open reading frame with 2376 nucleotides coding for 792 amino acids was identified. The 3′ noncoding region contains 810 nucleotides. The sequence context of the first ATG is in good agreement with the proposed initiation codon consensus sequence (26). The amino acid sequence of a tryptic peptide of gp110 is found in the open reading frame, which confirmed the open reading frame. Eight potential sites for N-linked glycosylation are present in the deduced amino acid sequence; seven of them are located on the N-terminal end of the glycoprotein. Our biochemical characterization of gp110 also suggests that there are about eight N-linked carbohydrate chains attached to each gp110 molecule (38, 39). These studies indicate that most or all of the N-linked glycosylation sites are used and that the carbohydrate chains are concentrated on the N-terminal part of this highly glycosylated protein. The first 28 amino acids at the N-terminal region were identified as a hydrophobic domain and most likely represent the signal peptide sequence. There is a hydrophobic domain on the C terminus, which may be the transmembrane domain of this membrane glycoprotein. Search of a data base using FASTP program (27) did not reveal other proteins with significant homology to gp110.

DISCUSSION

Protein sorting is an area in cell biology currently receiving considerable attention. Recent studies have focused on unraveling the molecular mechanisms for directing different proteins to their different cellular destinations, including the plasma membrane. In polarized cells, the plasma membrane is differentiated into distinct membrane domains (1–9). Proteins that are specific for specific domains are being identified biochemically and immunologically and by using the techniques of molecular biology (1–9, 28–33). cDNA clones for human and rat liver gap junction proteins have recently been isolated (29, 30). cDNAs for asialoglycoprotein receptor, composed of a set of sinusoidal domain-specific glycoproteins, have also been cloned for human and rat liver (31–33). Molecular cloning of domain-specific proteins in combination with expression of the cloned cDNAs in cell culture system, like the Madin–Darby canine kidney polarized cell line system (34–37), may provide some important insight into the molecular basis for protein sorting to specific cell-surface domains.

We report here the cDNA cloning for a glycoprotein of M_r 110,000 that is specifically localized in the bile canaliculus of rat liver. This protein is expressed with polarity in other tissues, including kidney and small intestine (unpublished data). Expression of gp110 with polarity in several polarized tissues makes it an excellent marker for studies of protein sorting and for examining the mechanism for the establishment and maintenance of polarity during development. Our studies show that expression of gp110 in liver is developmentally regulated with an onset of expression around birth (unpublished data). Recent studies suggest that this protein may be involved in secretion of bile salts across the bile canalicular membrane (10, 11). To our knowledge, cDNA cloned for a bile canalicular domain-specific protein has not been reported previously. Initial studies with these cDNA clones using RNA transfer blot analysis indicate that the mRNA coding for gp110 is about 3.0–3.5 kb in size and that the mRNA level for gp110 is higher in kidney than in liver, which is consistent with our observation that gp110 is present at higher level in rat kidney than in rat liver, as assessed by electrophoretic transfer blot analysis (manuscript in preparation). The amino acid sequence derived from the cDNA sequence indicates that the N-linked carbohydrate chains are concentrated on the N-terminal part of the polypeptide and implies that most of the polypeptide is extracellular with a small cytoplasmic domain. Further molecular biological studies utilizing these cDNA clones should reveal important aspects about the mechanism for sorting of proteins to the bile canalicular membrane domain of hepatocytes.

We thank E. Hou and Drs. A. V. Le, M. Diamond, B. Brown, J. Petell, J. Sawyer, G. Snyder, and M. Cennerazzo for their useful assistance and suggestions. We also thank B. Beccue for her excellent secretarial assistance. This work was supported by Grants CA38773 and GM34032 to D.D. from the National Institutes of Health.

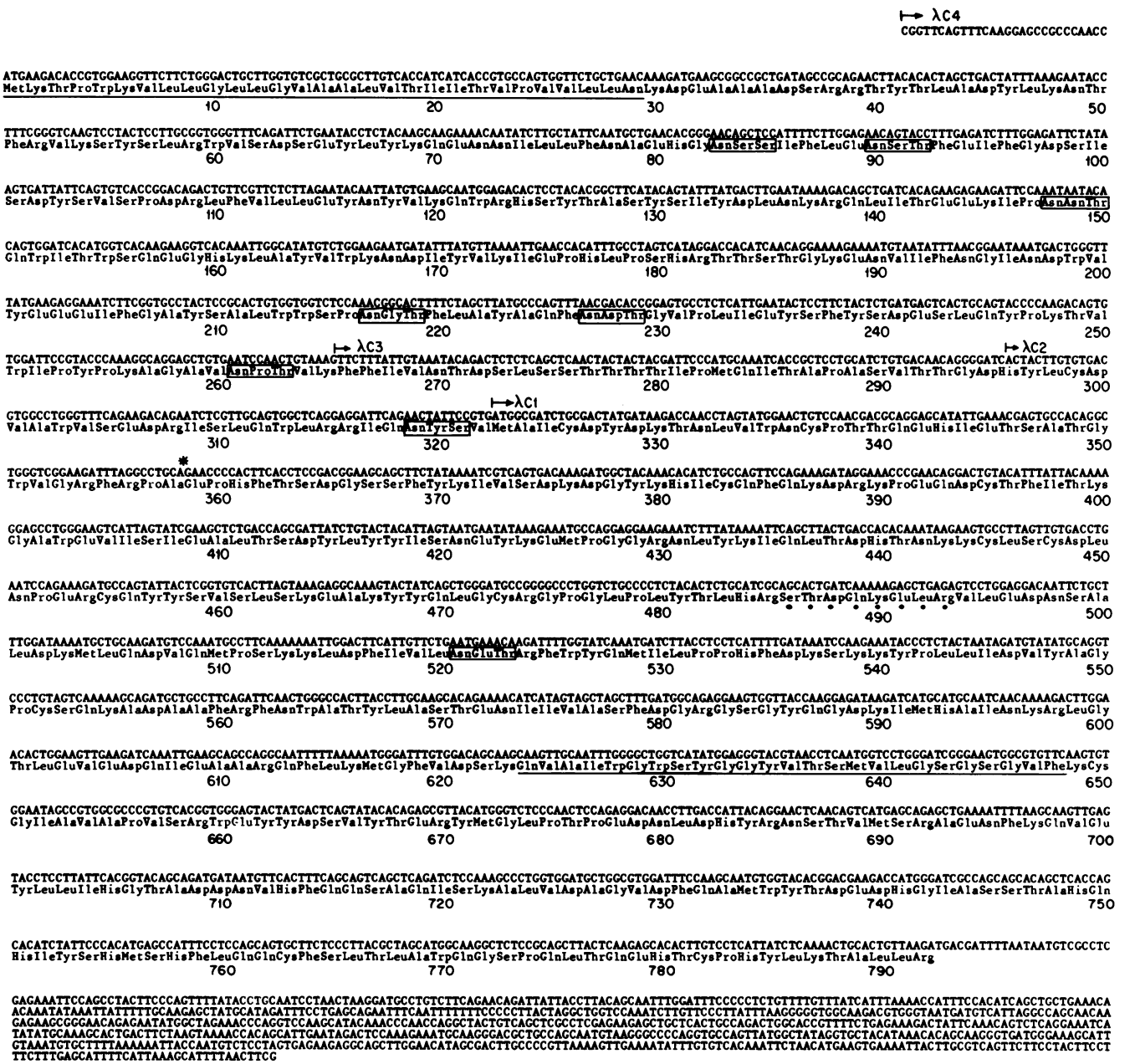

FIG. 4. Nucleotide sequence and deduced amino acid sequence of cDNA clones for rat gp110. The dotted line indicates the sequence that was confirmed by amino acid sequencing of a tryptic peptide of gp110. Boxes indicate the potential sites for N-linked glycosylation. The potential signal peptide and transmembrane domains are underlined. The starting nucleotide for each clone is indicated as ⊢. The *Pst* I restriction site in λC3 is indicated as *.

1. Roman, L. M. & Hubbard, A. L. (1983) *J. Cell Biol.* **96,** 1548–1558.
2. Hubbard, A. L., Bartles, J. R. & Braiterman, L. T. (1985) *J. Cell Biol.* **100,** 1115–1125.
3. Bartles, J. R., Braiteman, L. T. & Hubbard, A. L. (1985) *J. Cell Biol.* **100,** 1126–1138.
4. Bartles, J. R., Braiteman, L. T. & Hubbard, A. L. (1985) *J. Biol. Chem.* **260,** 12792–12802.
5. Kerjaschki, D., Noronha-Blob, L., Sacktor, B. & Farquhar, M. G. (1984) *J. Cell Biol.* **98,** 1505–1513.
6. Rodman, J. S., Seidman, L. & Farquhar, M. G. (1986) *J. Cell Biol.* **102,** 77–87.
7. Quaroni, A. (1984) *Prog. Cancer Res. Ther.* **29,** 267–293.
8. Quaroni, A. & Isselbacher, K. J. (1985) *Dev. Biol.* **111,** 267–279.
9. Cook, J., Hou, E., Hou, Y., Cairo, A. & Doyle, D. (1983) *J. Cell Biol.* **97,** 1823–1833.
10. Meier, P. J., Ruetz, S., Hugentobler, G., Fricker, G. & Kurtz, G. (1985) *Hepatology* **5,** 958.
11. Landmann, L., Ruetz, S. & Meier, P. J. (1986) *J. Cell Biol.* **103,** 466 (abstr.).
12. Hunkapiller, M. W., Lujan, E., Ostander, F. & Hood, L. (1983) *Methods Enzymol.* **91,** 227–236.
13. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. (1984) *Gene Anal. Tech.* **1,** 3–8.
14. Huynh, T. V., Young, R. A. & David, R. W. (1985) in *DNA Cloning Technique: A Practical Approach*, ed. Gover, D. (IRL, Oxford), Vol. 1, pp. 49–78.
15. Young, R. A., Bloom, B. R., Grosskinsky, C. M., Ivanyi, J., Thomas, D. & David, R. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2583–2587.
16. Govindan, M. V., Devic, M., Green, S., Gronemeyer, H. & Chambon, P. (1985) *Nucleic Acids Res.* **13,** 8293–8304.
17. Weinberger, C., Hollenberg, S. M., Ong, E. S., Harmon, J. M., Brower, S. T., Cidlowski, J., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Science* **228,** 740–742.
18. Williams, B. D., Mitchell, D. R. & Rosenbaum, J. L. (1986) *J. Cell Biol.* **103,** 1–11.

19. Lewis, S. A., Villasante, A., Sherline, P. & Cowan, N. J. (1986) *J. Cell Biol.* **102,** 2098–2105.
20. Davis, L., Dibner, M. & Battey, J. (1986) *Basic Methods in Molecular Cloning* (Elsevier, New York).
21. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
22. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
23. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13,** 31–40.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
25. Williams, S. A., Slatko, B. E., Moran, L. S. & DeSimone, S. M. (1986) *BioTechniques* **4,** 138–147.
26. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
27. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
28. Hunziker, W., Spiess, M., Semenza, G. & Lodish, H. F. (1986) *Cell* **46,** 227–234.
29. Paul, D. L. (1986) *J. Cell Biol.* **103,** 123–134.
30. Kumar, N. M. & Gilula, N. B. (1986) *J. Cell Biol.* **103,** 767–776.
31. Spiess, M., Schwartz, A. L. & Lodish, H. F. (1985) *J. Biol. Chem.* **260,** 1979–1982.
32. Spiess, M. & Lodish, H. F. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6465–6469.
33. Holland, E. C., Leung, J. O. & Drickamer, K. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7338–7342.
34. Fuller, S. D., Bravo, R. & Simons, K. (1985) *EMBO J.* **4,** 297–307.
35. Misek, D. E., Bard, E. & Rodriguez-Boulan, E. (1984) *Cell* **39,** 537–546.
36. Stephens, E. B., Compans, R. W., Earl, P. & Moss, B. (1986) *EMBO J.* **5,** 237–245.
37. Caplan, M. J., Anderson, H. C., Palade, G. E. & Jamieson, J. D. (1986) *Cell* **46,** 623–631.
38. Petell, J. K., Diamond, M., Hong, W., Bujanover, Y., Amarri, S., Pittschieler, K. & Doyle, D. (1987) *J. Biol. Chem.* **262,** in press.
39. Diamond, M., Petell, J. K. & Doyle, D. (1987) *J. Biol. Chem.* **262,** in press.