Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 937–941, February 1986
Biochemistry

Molecular cloning and nucleotide sequence of rat kidney γ-glutamyl transpeptidase cDNA

(plasma membrane proteins/synthetic oligonucleotides)

YANNICK LAPERCHE, FRÉDÉRIQUE BULLE, TSOURIA AISSANI, MARIE-NOËLLE CHOBERT, MARTINE AGGERBECK, JACQUES HANOUNE, AND GEORGES GUELLAËN

Institut National de la Santé et de la Recherche Médicale Unité 99, Hôpital H. Mondor, 94010 Créteil, France

Communicated by Alton Meister, October 7, 1985

ABSTRACT **We have screened a cDNA library (20,000 clones) made from rat kidney poly(A)$^+$ RNA, using an oligonucleotide probe that was a mixture of 14-base DNA oligomers containing all 32 possible sequences coding for residues 32–36 of the γ-glutamyl transpeptidase (EC 2.3.2.2.) heavy chain. We isolated and sequenced two cDNAs corresponding to the mRNA coding for the entire length of the enzyme precursor. The nucleotide sequence that we obtained (2072 bases) reveals an open reading frame of 1707 nucleotides coding for the common precursor of both enzyme subunits. The amino acid sequence begins with the 21 residues located at the NH_2-terminal hydrophobic region of the heavy subunit. We show that this sequence, which is not processed, is the only possible signal peptide in the sequence. Five potential N-glycosylation sites are present in the γ-glutamyl transpeptidase sequence. Using one of the two cDNA clones as probe, a 2.2-kilobase sequence was detected by blot analysis in rat kidney and human fetal liver RNA.**

γ-Glutamyl transpeptidase [(5-glutamyl)-peptide:amino acid 5-glutamyltransferase; EC 2.3.2.2] catalyzes the hydrolysis of γ-glutamyl compounds and the transfer of the γ-glutamyl moiety of glutathione and its derivatives to certain amino acids or dipeptides (1, 2). This heterodimeric enzyme is involved in glutathione metabolism and might play a role in amino acid uptake by cells. The highest activity is found in kidney and in intestinal cells (1, 2). In liver, and in liver-derived cell lines, γ-glutamyl transpeptidase activity is low; its blood level is increased during chronic alcoholism (3). In the liver, the increase in γ-glutamyl transpeptidase has also been recognized as an early marker of various chemically induced tumors (4). Previous reports showed that the γ-glutamyl transpeptidase activity in the liver and in some liver cell lines was increased by alcohol (5) and glucocorticoids (6, 7), and indirect evidence was presented that this effect required RNA synthesis. Both enzyme subunits are encoded by a common mRNA (8, 9). We report here the molecular cloning of DNA complementary to the kidney mRNA. The complete amino acid sequence of the two γ-glutamyl transpeptidase subunits, deduced from the nucleotide sequence, provides definitive information about the mechanism of γ-glutamyl transpeptidase insertion into the plasma membrane.

MATERIALS AND METHODS

Avian myeloblastosis virus reverse transcriptase (RNA-directed DNA polymerase), phage T4 polynucleotide kinase, *Escherichia coli* DNA polymerase I (Klenow fragment), and terminal deoxyribonucleotidyltransferase were obtained from Genofit S.A. (Geneva, Switzerland). Restriction endonucleases were purchased from Genofit S.A., New England Biolabs, or Boehringer Mannheim. Endonuclease S1 and bacterial alkaline phosphatase were from Bethesda Research Laboratories. Radiolabeled nucleotides were obtained from Amersham. Poly(dG)-tailed pBR322 was from New England Nuclear.

Synthesis of a 14-Base Oligodeoxyribonucleotide Probe. Rat kidney γ-glutamyl transpeptidase was purified from the renal brush border membrane after solubilization by papain (10), using 0.4 unit of papain per mg of protein and an incubation of 30 min at 37°C. Homogeneity of the purified enzyme was demonstrated by the presence of two prominent glycosylated bands (M_r 50,000 and 30,000) and the absence of any contaminant on $NaDodSO_4$/polyacrylamide gel electrophoresis (8). The two subunits were purified by high-pressure gel chromatography on a TSK 2000 column under denaturing conditions (11). One nanomole of each subunit was lyophilized and the NH_2-terminal sequence was determined at the Laboratoire d'Immunologie Moléculaire, Université Paris VII, Paris, France. A mixture of tetradecameric oligonucleotides, corresponding to the 32 possible coding sequences for amino acids 32–36 of the enzyme precursor (Fig. 1) and synthesized by Genofit S.A., were 5'-end-labeled to a specific activity of 6×10^6 cpm/mmol with [γ-^{32}P]ATP.

Isolation and Characterization of cDNA Clones. Poly(A)$^+$ RNA from female adult rat kidney was isolated and enriched in sequences coding for γ-glutamyl transpeptidase as described (8). The cDNA was synthesized from the poly(A)$^+$ RNA (5 μg) enriched in γ-glutamyl transpeptidase sequences, according to Land *et al.* (12). The cDNA was introduced by the dG/dC-tailing method at the *Pst* I site of pBR322. The hybrid molecules were used to transform *E. coli* strain MC1061 as described by Hanahan and Meselson (13). Following replica plating, the colonies were lysed and DNA was denatured and fixed on the filters (14). For hybridization with the 14-nucleotide oligomers, the filters were prehybridized at 40°C for 1 hr in 0.9 M NaCl/0.09 M sodium citrate, pH 7/5 mM EDTA/0.1% Ficoll 400/0.1% polyvinyl pyrrolidone/0.1% bovine serum albumin/0.05% sodium pyrophosphate/0.1% $NaDodSO_4$ containing 100 μg of calf liver tRNA per ml. The 5'-labeled tetradecameric oligonucleotide mixture (17 pmol) was added and incubated for 16 hr at 37°C. The filters were washed with 0.9 M NaCl/0.09 M sodium citrate, pH 7/0.05% sodium pyrophosphate/0.1% $NaDodSO_4$ for 30 min at 30°C and then washed at temperatures increasing by 2°C stepwise up to 40°C. After the washings at 36, 38, and 40°C, the signal was checked by exposing the filter to Royal X-Omat AR5 film (Kodak). Hybridization of the nitrocellulose filters to nick-translated cDNA was performed according to Wahl *et al.* (15).

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 pair(s).

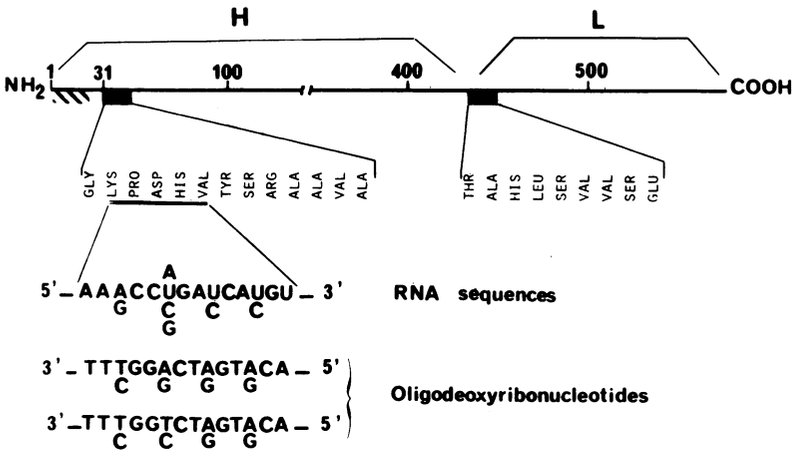

FIG. 1. Partial amino acid sequence of the γ-glutamyl transpeptidase precursor and sequences of the oligodeoxyribonucleotides used to screen the cDNA library. γ-Glutamyl transpeptidase heavy (H) and light (L) subunits are mapped on the γ-glutamyl transpeptidase precursor. Amino acid residues are numbered from the NH_2 terminus. The black boxes represent the sequences that we determined, and the hatched box indicates the sequence of 21 amino acids published by Matsuda *et al.* (19), which corresponds to the NH_2-terminal 21 amino acids of the heavy subunit. The amino acid sequence exhibiting the least codon degeneracy is underlined and the corresponding RNA and DNA sequences are indicated.

Bacterial clones were grown in Luria–Bertani medium supplemented with tetracycline at 15 μg/ml. Plasmid DNA was purified by the alkaline lysis method (16) and further purified on a Sepharose 2B column equilibrated with 10 mM Tris·HCl, pH 8/0.3 M NaCl/1 mM EDTA. Restriction enzyme analyses were performed under the reaction conditions recommended by the suppliers.

Restriction endonuclease-digested DNA fragments were either 5'-end-labeled (5' overhanging or blunt fragments), using T4 polynucleotide kinase and [γ-^{32}P]ATP, or 3'-end-labeled (3' overhanging or blunt fragments), using the 3'-end-nucleotide-labeling kit from Amersham. Labeled ends were separated in acrylamide or agarose gel after restriction endonuclease digestion and were sequenced by the chemical-modification technique of Maxam and Gilbert (17) according to the strategy depicted in Fig. 2.

Poly(A)$^+$ RNAs from rat kidney and from human adult and fetal livers were electrophoresed in 0.8% agarose gel (18), transferred to nitrocellulose, and hybridized to the cDNA probe (15).

RESULTS

NH_2-Terminal Sequence of the Two γ-Glutamyl Transpeptidase Subunits and Synthesis of the 14-Base cDNA Probe. The sequences of 15 and 10 amino acids of the NH_2-terminal parts of the heavy and the light subunit, respectively, were determined by automated microsequence analysis (Fig. 1). The sequence Lys-Pro-Asp-His-Val, underlined in Fig. 1 and corresponding to amino acids 32–36 of the heavy subunit, exhibits minimal codon degeneracy. It could be encoded by 32 different, 14-nucleotide-long RNA sequences. The 32 cDNA sequences were synthesized in two batches of 16 sequences differing from each other by only one base.

Screening of the cDNA Library. The cDNA library (20,000 clones) was screened for γ-glutamyl transpeptidase-specific sequences using the mixture of 14-base oligonucleotides. The four colonies that gave a signal above background were purified and rescreened in an ordered array with the oligonucleotide probe. The purified plasmids exhibited inserts of sizes ranging from 0.3 to 1.4 kilobase pairs (kb). The 1.4-kb insert was nick-translated to a specific activity of 10^7 cpm/μg and used as a probe to rescreen the library. Thirty-six new clones were obtained, and their plasmids contained cDNA ranging from 0.2 to 1.3 kb.

Restriction Maps of the cDNAs 1 and 39. The restriction map of the clones 1 and 39 are shown in Fig. 2. Only the restriction sites that were used for the sequence analysis are presented. The two cDNAs are positioned with respect to each other and to the mRNA of γ-glutamyl transpeptidase. These two cDNAs have a common restriction fragment of 450 base pairs which corresponds to the *Ava* I–*Eco*RI fragment.

Sequence Analysis of the cDNAs 1 and 39. The two cDNAs were sequenced from multiple restriction sites and on both DNA strands according to the strategy outlined in Fig. 2. A nucleotide sequence of 2072 bases was determined (Fig. 3). This sequence contains only one open reading frame long enough to encode a M_r 63,000 peptide. The first ATG codon defines a 227-nucleotide 5' untranslated region in front of an open reading frame of 1707 nucleotides. The 3' untranslated region is 138 nucleotides long and contains the AATAAA polyadenylylation signal 41 nucleotides upstream of the poly(A) tail. In the predicted amino acid sequence, we found three sequences matching perfectly with the γ-glutamyl transpeptidase sequences obtained by protein sequencing: (*i*) positions 31–46 and 380–388, corresponding to the NH_2-terminal sequences that we determined from the heavy and the light subunits (Fig. 1); (*ii*) positions 1–21, corresponding to the 21 amino acid hydrophobic sequence published by Matsuda *et al.* (19) for γ-glutamyl transpeptidase purified by detergent solubilization.

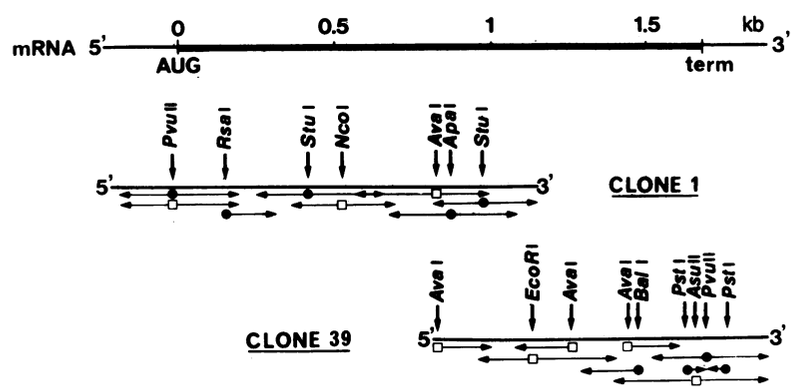

FIG. 2. Restriction map and strategy for sequencing two γ-glutamyl transpeptidase cDNA clones. Nucleotide residues are numbered from 5' to 3' in the message strand. Only the restriction sites used for the sequence analysis are shown. The direction and extent of each sequence determination from each site are shown by horizontal arrows; ● and □ represent 3' and 5' end labeling, respectively.

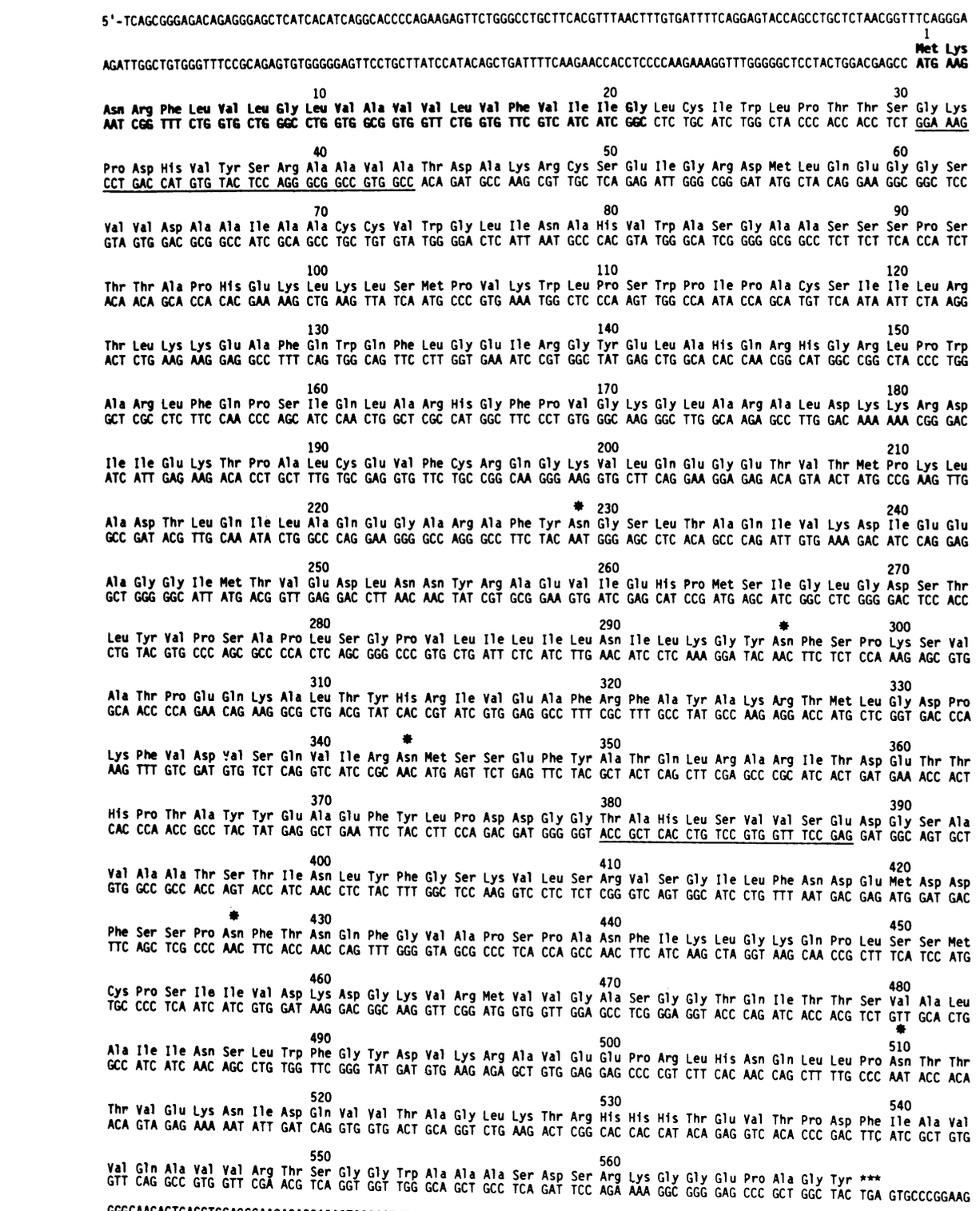

FIG. 3. Nucleotide and derived amino acid sequence of the γ-glutamyl transpeptidase cDNAs. The nucleotide sequence is the composite of the sequence of the γ-glutamyl transpeptidase cDNA clones 1 and 39. The derived amino acid sequence is shown above the corresponding nucleotides. Numbering of the amino acid sequence starts at the initiation codon. Residues 1–21 (boldface) correspond to the amino acid sequence published by Matsuda *et al.* (19). The two amino acid sequences that we determined by microsequence analysis of the NH_2-terminal ends of the γ-glutamyl transpeptidase subunits are underlined, as is the polyadenylylation signal AATAAA. Possible N-glycosylation sites (Asn-Xaa-Ser and Asn-Xaa-Thr) are indicated by asterisks.

The γ-glutamyl transpeptidase precursor encoded by the cDNA sequence reported here would contain 568 amino acids and exhibit a molecular weight of 61,800. The portions corresponding to the heavy and light γ-glutamyl transpeptidase subunits will have molecular weights of 41,766 and 20,052 for 379 and 189 residues, respectively, assuming that no residues are lost during the proteolytic processing of the precursor. The sequences Asn-Xaa-Thr and Asn-Xaa-Ser, which indicate possible N-glycosylation sites (20), were found three times in the heavy subunit and twice in the light subunit (Fig. 3).

Analysis of Rat Kidney and Human Liver RNA. The size of

the kidney poly(A)$^+$ RNA coding for γ-glutamyl transpeptidase is 2.2 kb as estimated from the size of RNA markers run on the same gel (Fig. 4). A band of the same size was also detected in poly(A)$^+$ RNA from human fetal liver, indicating a strong homology between human and rat γ-glutamyl transpeptidase sequences; no signal was detectable in a sample from adult human liver, probably due to the lower level of enzyme expression in the adult tissue (4).

Hydropathy Analysis of the Predicted Amino Acid Sequences of the Two γ-Glutamyl Transpeptidase Subunits. The results of the evaluation of hydropathy by the method of Kyte and Doolittle (21) is illustrated in Fig. 5. Three sequences in the heavy subunit and one in the light subunit can be considered as hydrophobic domains, since they exhibit an hydropathy index greater than 1.6 (dotted line in Fig. 5). Three of these hydrophobic sequences include <10 amino acids, and one (a) located at the NH_2-terminal end of the heavy subunit is 19 amino acids long.

DISCUSSION

The rat kidney cDNA library (20,000 clones) prepared from poly(A)$^+$ RNA enriched in γ-glutamyl transpeptidase sequences contained four positive clones detected on the first screening using the mixture of synthetic oligonucleotides as probe. The nick-translated insert of one of these 4 clones recognized 36 new clones. The mRNA from which the cDNA was reverse-transcribed was enriched 10 times in γ-glutamyl transpeptidase sequences and contained ≈0.5% specific sequences (8). This level is in accord with the percentage of positive clones (0.2%) found in the library. The confirmation that the cloned sequences correspond to γ-glutamyl transpeptidase sequences was obtained from the sequencing data.

In the deduced amino acid sequence, we found several stretches of amino acids that fit exactly with known portions of γ-glutamyl transpeptidase sequences previously established from microsequence analysis of the protein: (*i*) amino acids 1–21 correspond exactly to the sequence obtained by Matsuda *et al.* (19) from the NH_2-terminal end of the heavy subunit purified by membrane solubilization; (*ii*) amino acids 31–43 and 380–388 are the same as those obtained by microsequencing of the two subunits that we purified by papain treatment; (*iii*) the sequences of residues 31–62 and 380–413 were confirmed recently by P. P. Roller *et al.* (ref. 22 and personal communication). Like us, they found also that the papain cleavage site lies between residues 30 and 31, leaving a peptide of 30 amino acids anchored to the membrane. Matsuda *et al.* (19) showed that the heavy γ-glutamyl transpeptidase subunit obtained by papain treatment is 50 residues shorter than the one obtained by membrane solubilization. It is possible that, under their conditions (they used more papain and incubated for a longer period of time), papain cleaved twice, once between residues 30 and 31 and a second time between residues 50 and 51.

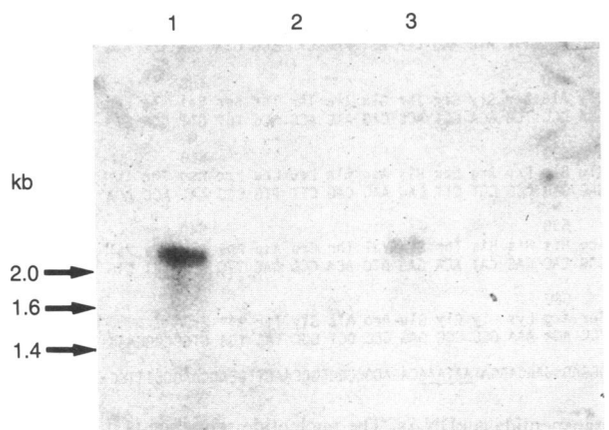

FIG. 4. Rat kidney and human adult and fetal liver mRNAs hybridized with a specific γ-glutamyl transpeptidase cDNA probe. Poly(A)$^+$ RNA from rat kidney (0.25 μg, lane 1), adult human liver (5 μg, lane 2), and fetal human liver (5 μg, lane 3) was electrophoresed in 0.8% agarose, transferred to nitrocellulose, and hybridized with the insert of clone 1 as probe (4×10^7 cpm at a specific activity of 10^8 cpm/μg) according to the conditions described in ref. 15. The blot was washed twice for 5 min at room temperature with 0.15 M NaCl/15 mM sodium citrate, pH 7/0.1% $NaDodSO_4$ and once for 1 hr at 65°C with 15 mM NaCl/1.5 mM sodium citrate, pH 7/0.1% $NaDodSO_4$. The positions of markers run in parallel are indicated at left.

The γ-glutamyl transpeptidase cDNA sequence reported here covers most, if not all, of the mRNA sequence. The size difference between the cloned cDNA sequences (2072 bases) and the kidney poly(A)$^+$ RNA (about 2200 bases, as shown in Fig. 4) may be explained by the absence in the cDNA of sequences complementary to the 150–200 nucleotides of the poly(A) tail present in most eukaryotic mRNA. The first methionine codon in the cDNA sequence is included in the consensus sequence (A/G)NNAUG(A/G) characteristic of an active start codon (23). At the 3′ untranslated region, the recognition sequence AATAAA for the addition of the poly(A) tail is usually found within 11 to 30 bases of the polyadenylylation site (24). In the γ-glutamyl transpeptidase sequence, the distance between the AATAAA sequence and the poly(A) tail is 41 bases. The two sequences CAGAGGC (nucleotides 2031–2037) and GCCTTTG (nucleotides 2057–2063) are, except for one G–T mismatch, inverted repeats that could hybridize to form a loop of 19 nucleotides. This would reduce the distance between the AATAAA signal and the poly(A) tail.

The complete γ-glutamyl transpeptidase sequence definitively confirms that the two γ-glutamyl transpeptidase subunits originate from a common precursor, as previously suggested (8, 9, 25, 26). The sizes of the heavy and light subunits determined from the cDNA are 57 and 21 amino acids shorter, respectively, than the sizes previously published (19), even if we assume that no segments are lost during the processing. The five possible sites for N-glycosylation (Asn-Xaa-Ser or Asn-Xaa-Thr) (20) are located in hydrophilic areas and therefore can be glycosylated very efficiently. The high content of serine, especially in the hydrophilic light subunit, is also favorable to O-glycosylation.

According to Kyte and Doolittle (21), a protein sequence corresponding to a membrane-spanning domain is characterized by a sequence of 18 amino acids exhibiting an average hydropathy index greater than 1.6. It appears that, in the γ-glutamyl transpeptidase sequence reported here, only one domain (Fig. 5, domain a) can meet these criteria. It is located in the NH_2-terminal portion of the γ-glutamyl transpeptidase heavy subunit (amino acids 6–24). This segment is known to remain in the membrane after papain digestion (19). The other hydrophobic peaks (Fig. 5, regions b, c, and d) are too short to span the membrane. Frielle *et al.* (27, 28) showed that the anchoring domain of the enzyme in the membrane comprised two peptides: peptide I, containing 83 residues (M_r 8700), and peptide II, containing 30 residues (M_r 3400). The best fit between our results and those of Frielle *et al.* can be obtained by assuming that peptide II corresponds to the NH_2-terminal hydrophobic sequence which spans the membrane (it has the same molecular weight). The absence of labeling of peptide II by reductive methylation would mean that the NH_2 groups of the methionine and lysine (amino acids 1 and 2) are on the inside of the membrane. This membrane orientation has been proposed for another protein of the kidney microvillar membrane, the dipeptidyl peptidase IV from which a hydrophobic peptide of $M_r \approx 3500$ was identified as a product of papain treatment (29). The existence of an internal signal

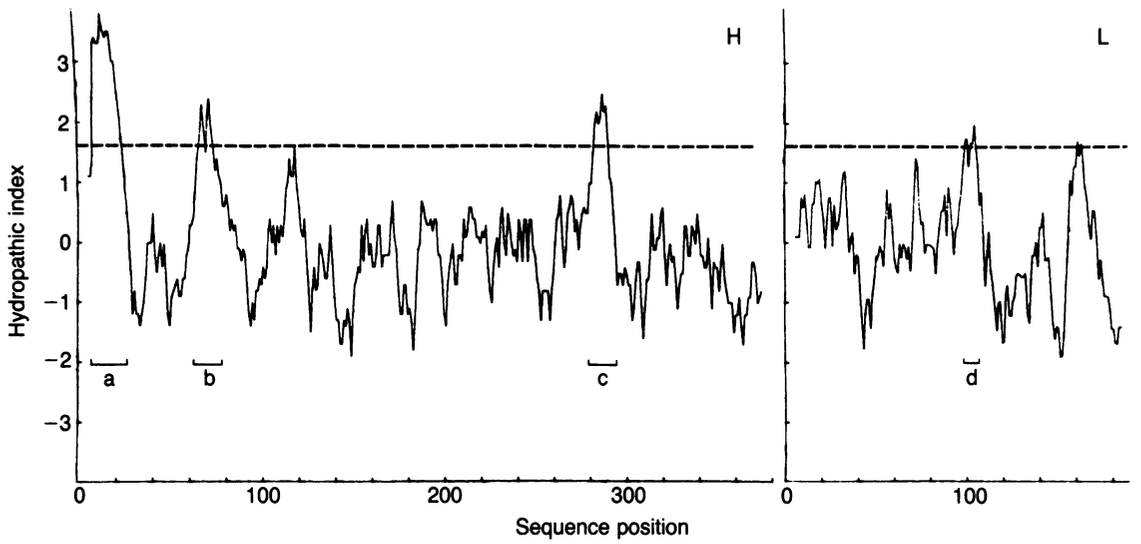

FIG. 5. Plot of the hydropathic index for the heavy and the light subunit of γ-glutamyl transpeptidase. Hydrophobicity values obtained according to Kyte and Doolittle (21) have been plotted with respect to positions of the amino acid sequence for both heavy (H) and light (L) subunits. The values for 11 sequential amino acids have been averaged and assigned to the middle residue of the span. Values for the hydrophobic and hydrophilic regions are above and below zero, respectively. The broken line corresponds to an average hydropathic index of 1.6. Sequences with values above this line (a, b, c, d) are considered hydrophobic regions (21).

peptide as reported for the anion-transport protein of the erythrocyte membrane (30) can be ruled out. Our results demonstrate that γ-glutamyl transpeptidase is inserted into the membrane through an uncleaved NH_2-terminal signal peptide.

Using a rat cDNA probe, we detected human γ-glutamyl transpeptidase mRNA by blot hybridization under conditions that indicate a strong homology between rat and human cDNA. In addition to information on the structure of the protein, the γ-glutamyl transpeptidase cDNA will enable us to investigate the regulation of this protein under various physiological and pathological conditions in rat as well in human tissues.

We thank Drs. L. P. Aggerbeck, A. Dubart, and P. H. Romeo for helpful discussions; Drs. M. Cohen-Solal, G. Craescu, and M. Goossens for their help and free access to their computerized systems; and L. Rosario for secretarial assistance. This work was supported by the Institut National de la Santé et de la Recherche Médicale, by the Délégation Générale à la Recherche Scientifique et Technique, by the Haut Comité d'Etude sur l'Alcoolisme, by the Centre National de la Recherche Scientifique, and by the Université Paris-Val de Marne.

1. Curthoys, N. P. (1983) *Miner. Electr. Metab.* **9**, 236–245.
2. Tate, S. S. & Meister, A. (1981) *Mol. Cell. Biochem.* **39**, 357–368.
3. Rosalki, S. B. & Rau, D. (1972) *Clin. Chim. Acta* **39**, 41–47.
4. Hanigan, M. H. & Pitot, H. C. (1985) *Carcinogenesis* **6**, 165–172.
5. Barouki, R., Chobert, M. N., Finidori, J., Aggerbeck, M., Nalpas, B. & Hanoune, J. (1983) *Hepatology* **3**, 323–329.
6. Billon, M. C., Dupre, G. & Hanoune, J. (1980) *Mol. Cell. Endocrinol.* **18**, 99–108.
7. Barouki, R., Chobert, M. N., Billon, M. C., Finidori, J., Tsapis, R. & Hanoune, J. (1982) *Biochim. Biophys. Acta* **721**, 11–21.
8. Finidori, J., Laperche, Y., Tsapis, R., Barouki, R., Guellaën, G. & Hanoune, J. (1984) *J. Biol. Chem.* **259**, 4687–4690.
9. Barouki, R., Finidori, J., Chobert, M. N., Aggerbeck, M., Laperche, Y. & Hanoune, J. (1984) *J. Biol. Chem.* **259**, 7970–7974.
10. Matsuda, Y., Tsuji, A. & Katanuma, N. (1980) *J. Biochem. (Tokyo)* **87**, 1243–1248.
11. Montelaro, R. C., West, M. & Issel, C. J. (1981) *Anal. Biochem.* **114**, 398–406.
12. Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schütz, G. (1981) *Nucleic Acids Res.* **9**, 2251–2266.
13. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63–67.
14. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3966.
15. Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
16. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
17. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
18. Lehrach, H. D., Diamond, D., Wozney, J. M. & Boedtker, H. (1979) *Biochemistry* **16**, 4743–4749.
19. Matsuda, Y., Tsuji, A. & Katunuma, N. (1983) *J. Biochem. (Tokyo)* **93**, 1427–1433.
20. Hubbard, S. C. & Ivatt, R. J. (1981) *Annu. Rev. Biochem.* **50**, 555–583.
21. Kyte, A. M. & Doolittle, F. (1982) *J. Mol. Biol.* **157**, 105–132.
22. Roller, P. P., Cone, J. L., Glowinski, J. B., Wirth, P. J. & Thorgeirsson, S. S. (1985) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **44**, 1058a.
23. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
24. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
25. Nash, B. & Tate, S. S. (1982) *J. Biol. Chem.* **257**, 585–588.
26. Capraro, M. A. & Hughey, R. P. (1983) *FEBS Lett.* **157**, 139–143.
27. Frielle, T., Brunner, J. & Curthoys, N. P. (1982) *J. Biol. Chem.* **257**, 14979–14982.
28. Frielle, T. & Curthoys, N. P. (1983) *Biochemistry* **22**, 5709–5714.
29. MacNair, R. D. C. & Kenny, A. J. (1979) *Biochem. J.* **179**, 379–395.
30. Braell, W. A. & Lodish, H. F. (1982) *Cell* **28**, 23–31.