Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 8840–8844, December 1988
Biochemistry

Cloning and nucleotide sequence of human γ-glutamyl transpeptidase

(glutathione metabolism/chromosome 22/plasma membrane proteins)

EWA RAJPERT-DE MEYTS*†, NORA HEISTERKAMP*, AND JOHN GROFFEN*‡

*Section of Molecular Genetics, Division of Medical Genetics, and †Laboratory of Biochemical Toxicology, Division of Neonatology and Pulmonary Diseases, Childrens Hospital of Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027

Communicated by Rachmiel Levine, August 11, 1988

ABSTRACT We have identified the gene for human γ-glutamyl transpeptidase [GGT; glutamine:D-glutamyl-peptide 5-glutamyltransferase (also called γ-glutamyltransferase), EC 2.3.2.2] in a BCR gene-related region located in band q11→qter of chromosome 22. Two cDNAs complementary to the GGT mRNA have been isolated from a human placental library constructed in phage λgt11. The largest cDNA has a size of 2535 base pairs (bp) and an open reading frame of 1707 nucleotides encoding 569 amino acids. By using a probe corresponding to this cDNA, a mRNA of ≈2.4 kilobases was detected by RNA blot-hybridization analysis in mouse kidney RNA. The GGT precursor encoded by the coding sequence would have an estimated M_r of 61,400. We compared our nucleotide and deduced amino acid sequences with the published results of rat kidney cDNAs. The human and rat amino acid sequences are similar; however, a considerable discrepancy in nucleotide sequence was found within a 180-bp fragment of the heavy chain, resulting in a completely different amino acid sequence for this region. In addition, the 5′ untranslated sequence of the human cDNA (669 bp) is substantially larger than that determined in the rat cDNA (227 bp). Our results may be valuable for further studies on the protein structure of human GGT as well as studies on the regulation of the enzyme.

γ-Glutamyl transpeptidase [GGT; glutamine:D-glutamyl-peptide 5-glutamyltransferase (also called γ-glutamyltransferase), EC 2.3.2.2] is a plasma membrane-bound enzyme of major importance in the metabolism of glutathione. Glutathione is a multifunctional tripeptide [L-γ-glutamyl-L-cysteinylglycine) involved in such important biological functions as transport and biosynthesis of amino acids and peptides, metabolism of certain endogenous compounds such as prostaglandins and leukotrienes, metabolism of drugs and toxic substances, and protection of cells against oxidant injury and radiation (1). GGT catalyzes the hydrolysis and transfer of the γ-glutamyl moiety from glutathione and other γ-glutamyl compounds to acceptors (amino acids, peptides) (1, 2).

GGT is present on the surface of the epithelial cells involved in secretory or absorbtive functions. The highest activity of the enzyme has been found in kidney, pancreas and intestine, whereas a lower activity is present in liver and in lung (2, 3). In hepatocytes and in liver-derived cell lines, the regulation of GGT activity has been extensively studied because of its increase in liver diseases (4) and in several chemically induced tumors, in which the enzyme can be used as a marker of the early stages of tumorigenesis (5, 6). The activity of the enzyme in the liver can be also stimulated by ethanol (7) and glucocorticoids (8, 9), and data suggest that increased RNA synthesis is required for this effect.

Highly purified enzyme preparations have been obtained from kidney and liver of several mammalian species. The rat and human kidney GGTs have been the most thoroughly studied to date. The enzyme is a heterodimer composed of two glycosylated subunits (2, 10). The heavy subunit (M_r, 50,000–62,000) anchors the molecule to the membrane through its hydrophobic amino-terminal portion. The light subunit (M_r = 22,000–30,000) with the putative γ-glutamyl binding region is noncovalently associated with the heavy chain (11, 12). The differences in relative molecular weights are most likely due to glycosylation of these proteins. Both human and rat renal GGT deglycosylated subunits exhibit a M_r of 42,000 and 19,000, respectively (10). The two subunits originate from a common, single-chain precursor ($M_r \approx$ 78,000), cotranslationally inserted in the cell membrane and core-glycosylated (12–17). The rat kidney mRNA coding for the propeptide has been isolated and translated *in vitro* (17); corresponding cDNAs have been cloned, and their nucleotide sequence has been determined (18, 19). Recently, a mutant *Escherichia coli* strain with high GGT activity has been obtained by using recombinant techniques and the purified *E. coli* enzyme has been crystallized (20). The chromosomal location of the human gene for GGT has been assigned to band q11 of chromosome 22 (21) using a rat GGT cDNA probe.

Although considerable progress has been made in elucidating the metabolism of glutathione and the action of GGT and other enzymes involved in the glutathione cycle, many questions remain. To determine the human GGT structure, membrane integration, its expression and induction in various types of cells, and its role in glutathione antioxidant function, we are investigating this enzyme at the molecular level. In this study we present the molecular cloning of a GGT cDNA from a human placental library and its complete nucleotide and deduced amino acid sequences.§ The human cDNA sequence is quite similar to the rat kidney cDNA; however, we found a considerable difference within a 180-base-pair (bp) sequence of the heavy chain, resulting in an amino acid sequence unlike that published for rat GGT.

MATERIALS AND METHODS

Enzymes and Reagents. Phage T4 DNA ligase, the *E. coli* large fragment of DNA polymerase I (Klenow fragment), and restriction endonucleases were purchased from Bethesda Research Laboratories and used according to the suppliers' recommendations. Radiolabeled nucleotides including Adenosine 5′-[α-thio]triphosphate were from New England Nu-

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: GGT, γ-glutamyl transpeptidase (γ-glutamyltransferase).

‡To whom reprint requests should be addressed.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04131).

clear. Deoxy- and dideoxynucleotides were purchased from Pharmacia.

Isolation of cDNA Clones. Repeat-free probes isolated from the 5′ end of cosmid JG-15 (37) were used to screen a human placental cDNA library in λgt11 (Clontech, Palo Alto, CA) as described (22). One positive clone, GT, was subcloned into pUC8 and analyzed in detail by restriction enzyme mapping. An 0.4-kilobase (kb) *Taq* I probe from the 5′ end of this cDNA was utilized to isolate an additional, larger cDNA (clone 31-1) from the same library.

Nucleotide Sequence Determination. DNA fragments of the cDNA clones obtained by restriction enzymes digestion were subcloned into M13mp18 and M13mp19 vectors (23). The nucleotide sequence of the single-stranded DNA fragments was determined by the dideoxynucleotide chain-termination method as described by Sanger *et al.* (24) with deoxyadenosine 5′-[α-thio]triphosphate and 8% methylenebispolyacrylamide/7 M urea gels for electrophoresis. The complete sequence was determined in both directions.

RNA Blot-Hybridization (Northern) Analysis. RNAs isolated from different mouse tissues were electrophoresed on an 0.8% agarose gel, transferred to nitrocellulose, and hybridized as described (25).

RESULTS

Isolation of the GGT cDNAs. Previously, a number of cosmids have been isolated that contain segments hybridizing to a *BCR* gene probe; the *BCR* gene, which encodes a protein of unknown function is located on chromosome 22 and is involved in translocations occurring in chronic myelogenous leukemia and in Philadelphia (Ph′)-positive acute lymphoblastic leukemia (26, 27). *BCR* gene-related sequences have been characterized by restriction enzyme mapping and hybridization: one of these, represented by cosmid JG-15 (Fig. 1) contains sequences homologous only to the 3′ coding region of the *BCR* gene and appears to represent a 5′ truncated copy of this gene. To investigate whether the 5′ region of cosmid JG-15 could contain any additional coding sequences, repeat-free probes prepared from this region were used to screen a human placental cDNA library. One positive clone, GT, was obtained in this manner. The cDNA hybridized to different restriction enzyme fragments within the 5′ end of cosmid JG-15, indicating an exon–intron organization (see Fig. 1); partial sequence analysis of this cDNA revealed close homology to a rat GGT cDNA previously isolated by others (18, 19). To characterize the human *GGT* gene, a second cDNA was isolated. This cDNA, clone 31-1, is 2.6 kb in length; clone GT with a size of 1.8 kb (see Fig. 2) lacks the 5′-terminal portion including the ATG initiation codon. Clone 31-1 contains a larger poly(A) tail than does clone GT (about 80 bp), in addition to 24 nucleotides preceding the poly(A) tail.

Nucleotide and Amino Acid Sequence Analysis of the cDNAs. The cDNA clones were sequenced after digestion with endonucleases according to the strategy shown in Fig. 2. The sequence of clone GT was completely concordant with that

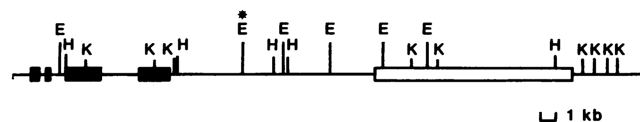

FIG. 1. Restriction enzyme map of cosmid JG-15. The open bar represents sequences hybridizing to a *BCR* gene cDNA probe; the solid bar indicates regions hybridizing to the *GGT* gene cDNA clone GT. Restriction enzymes are *Eco*RI (E), *Hin*dIII (H), and *Kpn* I (K). E* denotes a polymorphic *Eco*RI restriction enzyme site. Both the *GGT* gene and the *BCR* gene-homologous segments are oriented in a 5′-to-3′ fashion.

of clone 31-1. Clone 31-1 consists of 2535 nucleotides (Fig. 3), including an open reading frame of 1707 bp, a termination codon TGA, a 5′ untranslated region of 669 nucleotides, and a 3′ flanking region of 159 nucleotides followed by a poly(A) tail. The 5′ untranslated sequence of 669 nucleotides is considerably longer than that reported for the longest rat cDNA (227 nucleotides, ref. 18). However, at least four identical stretches of nucleotides, varying in length from 6 to 16 bp, are present in the rat and human 5′ untranslated GGT sequence. The AATAAA polyadenylylation sequence is 42 nucleotides upstream of the poly(A) tail in clone 31-1 and 18 nucleotides in the shorter clone GT. The first ATG resulting in the longest open reading frame is part of the sequence GCCATGA, which matches the consensus sequence ($^{A}_{G}$N-NAUG$^{A}_{G}$) for the initiation codon (28). The deduced polypeptide would be 569 amino acids in length, having an estimated molecular weight of 61,400. In analogy to the human and rat kidney sequences partially determined by a protein microsequence analysis (18, 19, 29–31), we predict that the heavy subunit would consist of 380 residues (M_r, 41,650) and the light subunit of 189 residues (M_r, 19,750).

We found seven possible N-glycosylation sites (Asn-Xaa-Ser or Asn-Xaa-Thr; ref. 32), six in the heavy chain and one in the light chain (Fig. 3). Four of these are in similar locations to the sites determined in the rat sequence (18, 19). A comparison of the complete deduced amino acid sequence of the human cDNA to that previously published for rat is shown in Fig. 4. The human sequence has one deletion (residue 76) when compared to that of rat, while the rat sequence has two deletions (residues 35 and 71); as a net result, the human sequence is one amino acid longer than that of rat. Between residues 81 and 135 (80 and 134 in the rat), we found a substantially different amino acid sequence because of a change in the nucleotide reading frame caused by single discrepancies in the nucleotide sequence. In comparison with both published sequences (18, 19), we found a deletion of two nucleotides (adenosine and thymidine at positions 898 and 1075, respectively). Furthermore, the human GGT nucleotide sequence contains additional nucleotides (cytidine at position 883 and adenosine at position 913) that are present in one but not in the other published rat cDNA sequences (18, 19). This results in a completely different reading frame between nucleotides 912 and 1075 (Fig. 4). To confirm the nucleotide sequence of this region of the human *GGT*, both DNAs were sequenced three times; in addition, partial genomic cDNA sequencing was performed (data not shown). In all instances, we could unambiguously determine the nucleotide sequence as shown in Fig. 3.

Analysis of Hydropathy and Possible Transmembrane Location of Human GGT. The estimation of hydropathy is shown in Fig. 5. Based on a hydropathy index of >1.6 (33), three domains in the heavy subunit and one in the light one are hydrophobic. Transmembrane segments were predicted by the ALOM program developed by Minoru Kaneshisa (National Cancer Institute, Bethesda, MD) and adapted by Carolyn Bucholtz (Commonwealth Scientific and Industrial Research Organization, Sydney, Australia). The most plausible transmembrane segment with a probability of 10^6:1 consists of 17 amino acids (residues 5–21; maximum extension, 1–31), the second transmembrane segment (probability 5:1) may consist of the residues 62–78 (maximum extension, 61–81).

Northern Blot Analysis. To analyze expression of GGT, total RNA was isolated from mouse kidney, pancreas, liver, spleen, and thymus. A mRNA of ≈2.4 kb is visible in kidney upon hybridization with a probe from the GGT cDNA (Fig. 6). A longer exposure of the autoradiogram revealed the presence of a low level of GGT mRNA in liver. These results are in complete concordance with the results obtained by Laperche *et al.* (18) using a rat GGT cDNA probe.

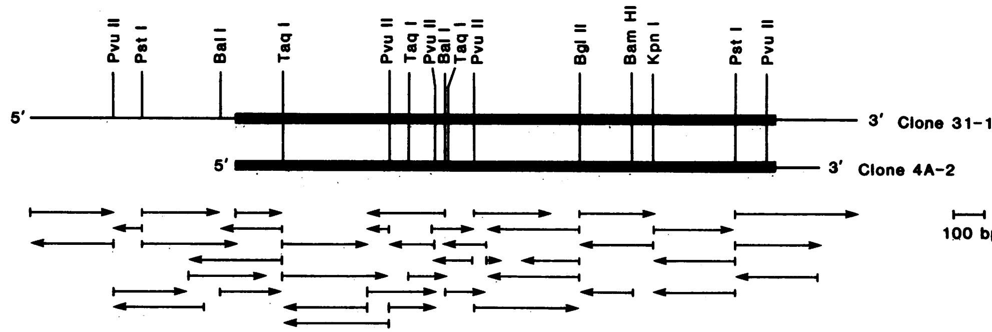

FIG. 2. Restriction map and sequencing strategy of two GGT cDNA inserts. The solid bar indicates the coding region; thin lines at the 5′ and 3′ ends represent 5′ and 3′ untranslated regions. Horizontal arrows indicate the direction and extent of each sequence determination. Only those restriction enzyme sites used for the sequence analysis are shown.

DISCUSSION

In the course of our studies on specific rearrangements of chromosome 22 associated with chronic myelogenous leukemia and Philadelphia-positive acute lymphoblastic leukemia, we isolated a human *GGT* gene by virtue of the fact that it is linked to a *BCR* gene-related segment; this segment is located on chromosome 22 (37). Thus, the location of the *GGT* gene is in agreement with the mapping by Bulle *et al.* (21), who localized *GGT* to the interface of band q11→q12 of human chromosome 22 by using *in situ* hybridization of human metaphase chromosomes to a rat cDNA probe. In addition, Bulle *et al.* (21) observed two regions on human chromosome 22 hybridizing in a specific manner to the rat cDNA probe (q11.1→q11.2 and q13.1), suggesting the presence of more than one *GGT* gene (38). Our studies on the genomic organization of the *GGT* gene support the hypothesis that there may be at least two *GGT* genes (unpublished results). The broad substrate specificity of GGT would be consistent

```
GCAGACCGGGCGTCGGGTGAGCCCAGAAGTGAGAGCAGTTGCCTGTGCCCCAGTGCTGTGTGACCCAGAGGCGCCGCTCAGCCTCTCTGAGCTGGTGGACATCATAGCGGGTGCAGCCCA   120

AAACTGTCTTCTGAGGAAGAGGTGCTCTCCTGGGCCCCCACTGTCCCCAGGCCTCAGATGCTGCATATTGCTGGCCCAGCTCCTGAGGCTCTTTGGACCTCCAGGAATCGGTGTCTCTAT   240

CAGGAACCCTTAACCCTGACCCGGACTCCCAGCTGGGACCCGGGGTGTCGGAGTGGCAAGAGCGCTGTCAGGCCTGGTGAAGGGTGTGAGCTGTCCAACGGGGCAGGGAGGAGGCAGGGC   360

CTGTTCTGCAGTTGGACAGACAGAGCCCTCTAGCTGCTTTCTGGAAGACTGAAGGGCAGCAAGGCAAGTGAGGTGCTGCCGTCATCCAGGCTGCACAGTTCAGTGATTTGCCTGAGGCCC   480

CACAGCAGGAGAACGAGAAGGCTGCCTGATCAGAGAGTCCCTGAAGAAGATTCTGTGGCTACAGGCTTCAGCAGAGTGTGAGGGAGACCCCGGTTATTTCCTCAGCTATTTCCACCAAAT   600

CCTCCTGTCTTTCGTGGCCAACACCCCAGGCAAGGCTTGGGGCCCCCGTCTGCTGCTGGACGCAGAGCCATGAAGAAGAAGTTAGTGGTGCTGGGCCTGCTGGCCGTGGTCCTGGTGCTG   720
                                                                       M  K  K  K  L  V  V  L  G  L  L  A  V  V  L  V  L
                                                                       1                 10

GTCATTGTCGGCCTCTGTCTCTGGCTGCCCTCAGCCTCCAAGGAACCTGACAACCATGTGTACACCAGGGCTGCCGTGGCCGCGGATGCCAAGCAGTGCTCGAAGATTGGGAGGGATGCA   840
 V  I  V  G  L  C  L  W  L  P  S  A  S  K  E  P  D  N  H  V  Y  T  R  A  A  V  A  A  D  A  K  Q  C  S  K  I  G  R  D  A
    20                            30                            40                            50
                                                                                                                     *
CTGCGGGACGGTGGCTCTGCGGTGGATGCAGCCATTGCAGCCCTGTTGTGTGTGGGGCTCATGAATGCCCACAGCATGGGCATCGGGGGTGGCCTCTTCCTCACCATCTACAACAGCACC   960
 L  R  D  G  G  S  A  V  D  A  A  I  A  A  L  L  C  V  G  L  M  N  A  H  S  M  G  I  G  G  G  L  F  L  T  I  Y  N  S  T
    60                            70                            80                            90
                                                                  *
ACACGAAAAGCTGAGGTCATCAACGCCCGCGAGGTGGCCCCCAGGCTGGCCTTTGCCACCATGTTCAACAGCTCGGAGCAGTCCCAGAAGGGGGGGCTGTCGGTGGCGGTGCCTGGGGAG   1080
 T  R  K  A  E  V  I  N  A  R  E  V  A  P  R  L  A  F  A  T  M  F  N  S  S  E  Q  S  Q  K  G  G  L  S  V  A  V  P  G  E
    100                           110                           120                           130

ATCCGAGGCTATGAGCTGGCACACCAGCGGCATGGGCGGCTGCCCTGGGCTCGCCTCTTCCAGCCCAGCATCCAGCTGGCCCGCCAGGGCTTCCCCGTGGGCAAGGGCTTGGCGGCAGCC   1200
 I  R  G  Y  E  L  A  H  Q  R  H  G  R  L  P  W  A  R  L  F  Q  P  S  I  Q  L  A  R  Q  G  F  P  V  G  K  G  L  A  A  A
    140                           150                           160                           170

CTGGAAAACAAGCGGACCGTCATCGAGCAGCAGCCTGTCTTGTGTGAGGTGTTCTGCCGGGATAGAAAGGTGCTTCGGGAGGGGGAGAGACTGACCCTGCCGCAGCTGGCTGACACCTAC   1320
 L  E  N  K  R  T  V  I  E  Q  Q  P  V  L  C  E  V  F  C  R  D  R  K  V  L  R  E  G  E  R  L  T  L  P  Q  L  A  D  T  Y
    180                           190                           200                           210
                                    *
GAGACGCTGGCCATCGAGGGTGCCCAGGCCTTCTACAACGGCAGCCTCACGGCCCAGATTGTGAAGGACATCCAGGCGGCCGGGGGCATTGTGACAGCTGAGGACCTGAACAACTACCGT   1440
 E  T  L  A  I  E  G  A  Q  A  F  Y  N  G  S  L  T  A  Q  I  V  K  D  I  Q  A  A  G  G  I  V  T  A  E  D  L  N  N  Y  R
    220                           230                           240                           250
                        *                                                                                                *
GCTGAGCTGATCGAGCACCCGCTGAACATCAGCCTGGGAGACGCGGTGCTGTACATGCCCAGTGCGCCGCTCAGCGGGCCCGTGCTGGCCCTCATCCTCAACATCCTCAAAGGGTACAAC   1560
 A  E  L  I  E  H  P  L  N  I  S  L  G  D  A  V  L  Y  M  P  S  A  P  L  S  G  P  V  L  A  L  I  L  N  I  L  K  G  Y  N
    260                           270                           280                           290

TTCTCCCGGGAGAGCGTGGAGAGCCCCGAGCAGAAGGGCCTGACGTACCACCGCATCGTAGAGGCTTTCCGGTTTGCCTACGCCAAGAGGACCCTGCTTGGGGACCCCAAGTTTGTGGAT   1680
 F  S  R  E  S  V  E  S  P  E  Q  K  G  L  T  Y  H  R  I  V  E  A  F  R  F  A  Y  A  K  R  T  L  L  G  D  P  K  F  V  D
    300                           310                           320                           330
                   *
GTGACTGAGGTGGTCCGCAACATGACCTCCGAGTTCTTCGCTGCCCAGCTCCGGGCCCAGATCTCTGACGACACCACTCACCCGATCTCCTACTACAAGCCCGAGTTCTACACGCCGGAT   1800
 V  T  E  V  V  R  N  M  T  S  E  F  F  A  A  Q  L  R  A  Q  I  S  D  D  T  T  H  P  I  S  Y  Y  K  P  E  F  Y  T  P  D
    340                           350                           360                           370

GACGGGGGCACTGCTCACCTGTCTGTCGTCGCAGAGGACGGCAGTGCTGTGTCCGCCACCAGCACCATCAACCTCTACTTTGGCTCCAAGGTCCGCTCCCCGGTCAGCGGGATCCTGTTC   1920
 D  G  G  T  A  H  L  S  V  V  A  E  D  G  S  A  V  S  A  T  S  T  I  N  L  Y  F  G  S  K  V  R  S  P  V  S  G  I  L  F
    380                           390                           400                           410

AATAATGAAATGGA[illegible]GACTTCAGCTCTCCCAGCATCACCAACGAGTTTGGGGTACCCCCCTCACCTGCCAATTTCATCCAGCCAGGGAAGCAGCCGCTCTCGTCCATGTGCCCGACGATC   2040
 N  N  E  M  [illegible]  D  F  S  S  P  S  I  T  N  E  F  G  V  P  P  S  P  A  N  F  I  Q  P  G  K  Q  P  L  S  S  M  C  P  T  I
    420                           430                           440                           450

ATGGTGGGCCAGGACGGCCAGGTCCGGATGGTGGTGGGAGCTGCTGGGGGCACACAGATCACCACGGCCACTGCACTGGCCATCATCTACAACCTCTGGTTCGGCTATGACGTGAAGCGG   2160
 M  V  G  Q  D  G  Q  V  R  M  V  V  G  A  A  G  G  T  Q  I  T  T  A  T  A  L  A  I  I  Y  N  L  W  F  G  Y  D  V  K  R
    460                           470                           480                           490
                                    *
GCCGTGGAGGAGCCCCGGCTGCACAACCAGCTTCTGCCCAACGTCACGACAGTGGAGAGAAACATTGACCAGGCAGTGACTGCAGCCCTGGAGACCCGGCACCATCACACCCAGATCGCG   2280
 A  V  E  E  P  R  L  H  N  Q  L  L  P  N  V  T  T  V  E  R  N  I  D  Q  A  V  T  A  A  L  E  T  R  H  H  H  T  Q  I  A
    500                           510                           520                           530

TCCACCTTCATCGCTGTGGTGCAAGCCATCGTCCGCACGGCTGGTGGCTGGGCAGCTGCCTCGGACTCCAGGAAAGGCGGGGAGCCTGCCGGCTACTGAGTGCTCCAGGAGGACAAGGCT   2400
 S  T  F  I  A  V  V  Q  A  I  V  R  T  A  G  G  W  A  A  A  S  D  S  R  K  G  G  E  P  A  G  Y
    540                           550                           560

GACAAGCAATCCAGGGACAAGATACTCACCAGGACCAGGAAGGGGACTCTGGGGGACCGGCTTCCCCTGTGAGCAGCAGAGCAGCACAATAAATGAGGCCACTGTGCCAGGCTCAGGTGG   2520

CCTCCTGGCCTTTCCAAAAAAAAAAAAAAAAAAAA
```

FIG. 3. Complete nucleotide and deduced amino acid sequence of GGT cDNA clone 31-1. Numbering of the amino acid sequence (below the corresponding codons) starts at the initiation codon. Possible N-glycosylation sites are indicated by asterisks. The potential transmembrane regions are underlined.

```
1     MKKKLVVLGLLAVVLVLVIVGLCLWLPSASKEPDNHVYTRAAVAADAKQCSKIGRDALRD    HUMAN
      ::    :::: ::::: ::.::: :::  :  :: ::: ::::: ::: :: :::: :
1'    MKNRFLVLGLVAVVLVFVIIGLCIWLPTTSGKPD-HVYSRAAVATDAKRCSEIGRDMLQE    RAT
1"    ---                                                             RAT (19)

61    GGSAVDAAIAALLCV-GLMNAHSMGIGGGLFLTIYNSTTRKAEVINAREVAPRLAFATMF    HUMAN
      ::: :::::::  :: :: : :     :               :           :
60'   GGSVVDAAIAA-CCVWGLINAHVWASGAASSSPSTTAPHEKLKLSMPVKWLPSWPIPACS    RAT
60"       GPSQ        SLMPT                                G          RAT (19)

120   NSSEQSQKGGLSVAVPGEIRGYELAHQRHGRLPWARLFQPSIQLARQGFPVGKGLAAALE    HUMAN
               :         :::::::::::::::::::::::::::: ::::::::: ::
119'  IILRTLKKEAFQWQFLGEIRGYELAHQRHGRLPWARLFQPSIQLARHGFPVGKGLARALD    RAT

180   NKRTVIEQQPVLCEVFCRDRKVLREGERLTLPQLADTYETLAIEGAQAFYNGSLTAQIVK    HUMAN
       ::  ::  : :::::::  ::: :::  : : ::::   :: ::: ::::::::::::::
179'  KKRDIIEKTPALCEVFCRQGKVLQEGETVTMPKLADTLQILAQEGARAFYNGSLTAQIVK    RAT

240   DIQAAGGIVTAEDLNNYRAELIEHPLNISLGDAVLYMPSAPLSGPVLALILNILKGYNFS    HUMAN
      ::: :::: : ::::::::: ::::   : :::   :: ::::::::::: ::::::::::::
239'  DIQEAGGIMTVEDLNNYRAEVIEHPMSIGLGDSTLYVPSAPLSGPVLILILNILKGYNFS    RAT

300   RESVESPEQKGLTYHRIVEAFRFAYAKRTLLGDPKFVDVTEVVRNMTSEFFAAQLRAQIS    HUMAN
        ::  :::: ::::::::::::::::::: :::::::::  : ::: ::: : :::: :
299'  PKSVATPEQKALTYHRIVEAFRFAYAKRTMLGDPKFVDVSQVIRNMSSEFYATQLRARIT    RAT

                         *
360   DDTTHPISYYKPEFYTPDDGGTAHLSVVAEDGSAVSATSTINLYFGSKVRSPVSGILFNN    HUMAN
      : ::::  ::  ::: ::::::::::::: :::::: :::::::::::::: : :::::::
359'  DETTHPTAYYEAEFYLPDDGGTAHLSVVSEDGSAVAATSTINLYFGSKVLSRVSGILFND    RAT
359"                                       M                V         RAT (19)

420   EMDDFSSPSITNEFGVPPSPANFIQPGKQPLSSMCPTIMVGQDGQVRMVVGAAGGTQITT    HUMAN
      ::::::::  :: ::: ::::::: ::::::::::: : :  :: ::::::: :::::::
419'  EMDDFSSPNFTNQFGVAPSPANFIKPGKQPLSSMCPSIIVDKDGKVRMVVGASGGTQITT    RAT

480   ATALAIIYNLWFGYDVKRAVEEPRLHNQLLPNVTTVERNIDQAVTAALETRHHHTQIAST    HUMAN
         :::::  ::::::::::::::::::::::::: :::: :::: ::: : ::::::
489'  SVALAIINSLWFGYDVKRAVEEPRLHNQLLPNTTTVEKNIDQVVTAGLKTRHHHTEVTPD    RAT

540   FIAVVQAIVRTAGGWAAASDSRKGGEPAGY                                  HUMAN
      ::::::: ::: :::::::::::::::::::
539'  FIAVVQAVVRTSGGWAAASDSRKGGEPAGY                                  RAT
```

FIG. 4. Comparison of human and rat GGT deduced amino acid sequences. We compared our sequence to the amino acid sequences published by Laperche et al. (ref. 18; the line indicated by "Rat"). From the second sequence published by Coloma and Pitot (19), only the residues differing from those of the first rat sequence are shown [the line indicated with "Rat (19)"]. The predicted amino-terminal end of the light subunit is indicated by an asterisk. The underlined parts of the rat sequence refer to fragments of rat GGT previously analyzed by protein sequencing.

with some variation in different forms of the enzyme. It has been suggested (34) that the enzyme that catalyzes the conversion of leukotriene C_4 to leukotriene D_4 may be different from that which transfers the γ-glutamyl moiety from unconjugated glutathione. Although a brief report by Orning and Hammarstrom (35) showed that the Michaelis constant values (K_m) for leukotriene C_4 and glutathione hydrolysis were similar in a crude preparation from porcine kidney, this does not rule out differences in other tissues or in localization within the cell.

The size of rat kidney GGT mRNA (18, 19) and human fetal liver GGT mRNA (18) has been estimated to be 2.2 kb when using rat cDNA probes. In mouse kidney we detected a mRNA of similar size, 2.4 kb, using the human GGT cDNA as a probe. This further confirms that our cDNA corresponds indeed to the human mRNA for GGT. Our data show that the

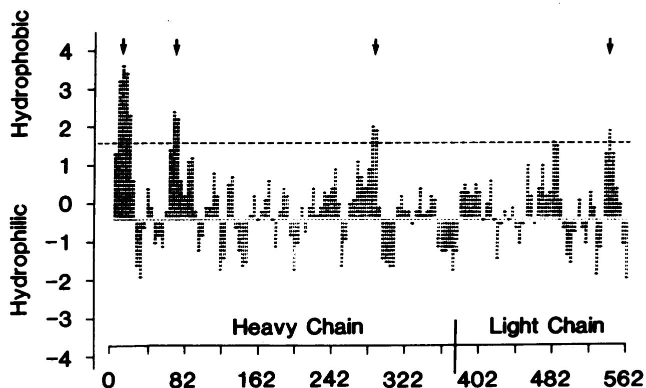

FIG. 5. Hydropathy plot of the deduced amino acid sequence of GGT. The broken line indicates an average hydropathic index of 1.6 (25). The fragments considered hydrophobic are indicated by arrows.

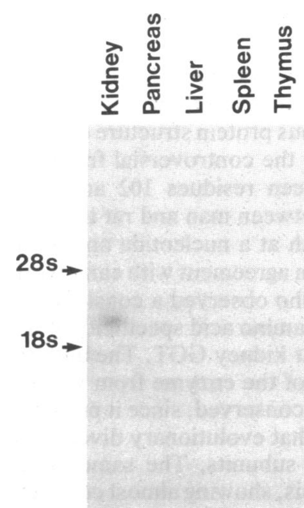

FIG. 6. Northern blot analysis of GGT and mRNA expression. Five micrograms of total RNA isolated from different mouse tissues was loaded in each lane. The positions of the ribosomal RNAs, 28S and 18S, are indicated at the left.

size of the human placental GGT cDNA is 2.6 kb. This difference is essentially due to the longer 5′ untranslated region (227 nucleotides in the rat kidney cDNA; 669 nucleotides in the human placental cDNA), although minor differences can be explained by the size of the poly(A) tail included in the cDNA. Several explanations are possible for this discrepancy. Since our cDNA was isolated from a human library constructed from placental RNA and the rat cDNAs were isolated from a kidney cDNA library, tissue-specific differences in mRNA length may exist. Alternatively, minor differences in mRNA size between human and rat GGT may have been undetected on a Northern blot. It is also possible that the rat cDNA clones of 2.2 kb were incomplete at the 5′ end. Finally, the differences observed in length may be due to transcription from different GGT loci in the human genome. Whatever the case, the substantial length of the 5′ untranslated sequence present in the human cDNA isolated by us may indicate a regulatory function for this region in translation; the presence of domains conserved between human and rat in the 5′ untranslated region supports this hypothesis.

The cDNA sequence reported here allows an estimate of molecular weights of the heavy and light GGT subunits to be 41,650 and 19,750, respectively. This is in complete concordance with the results obtained by Tate and Khadse (10), who showed by using specific antibodies that the deglycosylated subunits of purified human kidney GGT exhibit M_r values of about 42,000 and 19,000, respectively.

Several fragments of the rat and human GGT protein sequence have been reported. Matsuda et al. (29) determined the amino-terminal sequences of both heavy and light subunits of rat GGT, which then were confirmed by several authors using the microsequence analysis of the protein (18, 19, 31). Recently, the N-terminal amino acid sequences of the human renal GGT subunits have been published (30). The sequences of these regions are in complete concordance with the corresponding segments of the amino acid sequence deduced by us. There is also a high degree of similarity between the deduced amino acid sequences of human and rat GGT. Generally, the amino acid sequence in both human and rat GGT is highly conserved, with the exception of the region between residues 81 and 135 in the heavy subunit. As most of the amino acid changes in this region are caused by a change of the reading frame, we repeated the sequencing of the controversial cDNA fragment three times in both orientations and of both isolated cDNAs. In addition, we compared this sequence with a corresponding genomic DNA fragment (not shown). In conclusion, we feel confident that

the sequence of this region as shown in Fig. 3 has no ambiguities.

Several possibilities may be considered to explain the discrepancy found. There may be an evolutionary difference in the amino acid sequence and thus protein structure of GGT between human and rat species; the controversial fragment includes a region located between residues 102 and 127, which is much less conserved between man and rat than the other parts of the sequence, both at a nucleotide and at an amino acid level. This would be in agreement with early work of Tate and Ross in 1977 (36), who observed a considerable difference between the acceptor amino acid specificity of the heavy subunits of human and rat kidney GGT. The authors suggested that the light subunit of the enzyme from various mammalian species may be very conserved, since it probably contains the active region, and that evolutionary divergence may be confined to the heavy subunits. The same group recently confirmed this hypothesis, showing almost complete homology in the 23 amino-terminal residues of the light subunits of rat, dog, bovine, and human GGT (30). Alternatively, it is possible that different transcripts and proteins are produced by multiple GGT loci within one species. Our data, in concordance with those obtained by others, support the presence of a hydrophobic domain in the amino-terminal part of the heavy chain subunit at residues 6–22. Evaluation of the hydrophobicity (see Fig. 5), in combination with the analysis of the statistical probability of a transmembrane location for certain fragments, might indicate, although with significantly smaller probability, the existence of an additional putative integral domain between residues 62 and 78 in the heavy subunit. However, it has been found that papain cleaves the enzyme from its membrane-integrated domain in human kidney at residue 30 or 31 (30) and in rat kidney at residue 30 (18). In the light of these results, the presence of the second integral domain is unlikely. It has been proposed (18) that the amino-terminal amino acids of rat GGT are located on the inside of the cytoplasmic membrane. Our results support this hypothesis. The most amino-terminal part consisting of methionine, three lysines, and leucine could be intracellular, followed by a hydrophobic domain of 18–24 amino acids spanning the cytoplasmic membrane. The elucidation of structural organization of the GGT molecule will require further investigation, especially considering the possibility of organ or tissue-specific heterogeneity of the enzyme.

We thank Drs. Henry J. Forman and John A. Barranger for critical review of the manuscript, Ms. Karen Betz and Olga Torrejon for typing of the manuscript, and Miguel Brizuela and Eva Knoppel for excellent technical assistance. This work was supported in part by National Institutes of Health Grant HL 37556 to H. J. Forman.

1. Meister, A. & Anderson, E. M. (1983) *Annu. Rev. Biochem.* **52**, 711–760.
2. Tate, S. S. & Meister, A. (1982) *Mol. Cell. Biochem.* **39**, 357–368.
3. Curthoys, N. P. (1983) *Miner. Electrolyte Metab.* **9**, 236–329.
4. Rosalki, S. B. & Rau, D. (1972) *Clin. Chim. Acta* **39**, 41–47.
5. Hanigan, M. H. & Pitot, H. C. (1985) *Carcinogenesis* **6**, 165–172.
6. Sawabu, N., Nagaku, M., Ozaki, K., Wakabayashi, T., Toya, D., Hattori, N. & Ishii, M. (1983) *Cancer* **51**, 327–331.
7. Barouki, R., Chobert, M. N., Finidori, J., Aggerbeck, M., Nalpas, B. & Hanoune, J. (1983) *Hepatology* **3**, 323–329.
8. Billon, M. C., Dupre, G. & Hanoune, J. (1980) *Mol. Cell. Endocrinol.* **18**, 99–108.
9. Barouki, R., Chobert, M. N., Billon, M. C., Finidori, J., Tsapis, R. & Hanoune, J. (1982) *Biochim. Biophys. Acta* **721**, 11–21.
10. Tate, S. S. & Khadse, V. (1986) *Biochem. Biophys. Res. Commun.* **141**, 1189–1194.
11. Hughey, R. P., Coyle, P. J. & Curthoys, N. P. (1979) *J. Biol. Chem.* **254**, 1124–1128.
12. Tsuji, A., Matsuda, Y. & Katunuma, N. (1980) *J. Biochem. (Tokyo)* **87**, 1567–1571.
13. Finidori, J., Laperche, Y., Haguenauer-Tsapis, R., Barouki, R., Guellaen, G. & Hanoune, J. (1984) *J. Biol. Chem.* **259**, 4687–4690.
14. Barouki, R., Finidori, J., Chobert, M. N., Aggerbeck, M., Laperche, Y. & Hanoune, J. (1984) *J. Biol. Chem.* **259**, 7970–7974.
15. Nash, B. & Tate, S. S. (1982) *J. Biol. Chem.* **257**, 585–588.
16. Capraro, M. A. & Hughey, R. P. (1983) *FEBS Lett.* **157**, 139–143.
17. Nash, B. & Tate, S. S. (1984) *J. Biol. Chem.* **259**, 678–685.
18. Laperche, Y., Bulle, F., Aissani, T., Chobert, M. N., Aggerbeck, M., Hanoune, J. & Guellaen, G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 937–941.
19. Coloma, J. & Pitot, H. C. (1986) *Nucleic Acids Res.* **14**, 1393–1403.
20. Suzuki, H., Kumagai, H., Echigo, T. & Tochikura, T. (1988) *Biochem. Biophys. Res. Commun.* **150**, 33–38.
21. Bulle, F., Mattei, M. G., Siegrist, S., Pawlak, A., Passage, E., Chobert, M. N., Laperche, Y. & Guellaen, G. (1987) *Hum. Genet.* **76**, 283–286.
22. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. & Grosveld, G. (1985) *Nature (London)* **314**, 758–761.
23. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Grosveld, G., Verwoerd, T., van Agthoven, T., de Klein, A., Heisterkamp, N., Stam, K. & Groffen, J. (1986) *Mol. Cell. Biol.* **6**, 607–616.
26. Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosveld, G. (1984) *Cell* **36**, 93–99.
27. Hermans, A., Heisterkamp, N., von Lindern, M., van Baal, S., Meijer, D., van der Plas, D., Wiedemann, L. M., Groffen, J., Bootsma, D. & Grosveld, G. (1987) *Cell* **51**, 33–40.
28. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
29. Matsuda, Y., Tsuji, A. & Katunuma, N. (1983) *J. Biochem. (Tokyo)* **93**, 1427–1433.
30. Tate, S. S., Khadse, V. & Wellner, D. (1988) *Arch. Biochem. Biophys.* **262**, 397–408.
31. Roller, P. P., Cone, J. L., Glowinski, J. B., Wirth, P. J. & Thorgeirsson, S. S. (1985) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **44**, 1058 (abstr.).
32. Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A. & Lennarz, W. J. (1979) *J. Biol. Chem.* **254**, 9747–9753.
33. Kyte, A. M. & Doolittle, F. (1982) *J. Mol. Biol.* **157**, 105–132.
34. Bach, M. K. (1984) in *The Leukotrienes: Chemistry and Biology*, eds. Chakrin, L. W. & Bailey, D. M. (Academic, New York), pp. 163–194.
35. Orning, L. & Hammarstrom, S. (1982) *Biochem. Biophys. Res. Commun.* **106**, 1304–1309.
36. Tate, S. S. & Ross, M. E. (1977) *J. Biol. Chem.* **252**, 6042–6045.
37. Heisterkamp, N. & Groffen, J. (1988) *Nucleic Acids Res.* **16**, 8045–8056.
38. Pawlak, A., Lahuna, O., Bulle, F., Suzuki, A., Ferry, N., Siegrist, S., Chikhi, N., Chobert, M. N., Guellaen, G. & Laperche, Y. (1988) *J. Biol. Chem.* **263**, 9913–9916.