The Structure of the Rat Glutathione S-Transferase P Gene and Related Pseudogenes*

Akihiko Okuda, Masaharu Sakai, and Masami Muramatsu
From the Department of Biochemistry, The University of Tokyo Faculty of Medicine, Hongo, Bunkyo-ku, Tokyo 113, Japan

(Received for publication, August 18, 1986)

We have isolated the rat placental-type glutathione S-transferase (GST-P) gene from a λ phage library using GST-P cDNA clone, pGP5 (Sugioka, Y., Kano, T., Okuda, A., Sakai, M., Kitagawa, T., and Muramatsu, M. (1985) Nucleic Acids Res. 13, 6049–6057), as a probe. The rat GST-P gene is about 3 kilobase pairs long and contains 7 exons and 6 introns, encoding the same GST-P protein specified by pGP5. The cap site maps 79 nucleotides upstream from the translation initiation site. The canonical promoter “TATA” box was found 27 base pairs upstream from the putative cap site. Two hundred base pairs upstream from the cap site are rich in G + C residues (61%), and the hexanucleotide sequence 5’-GGGCGG-3’ is found at position -47 to -42. We have also isolated several processed-type pseudogenes which were presumably originated by reverse transcription followed by insertion at target sites.

In the course of an attempt to identify a specific pattern of gene expression during chemical hepatocarcinogenesis, we came up with a protein that increased dramatically in precancerous liver cells as seen by an O’Farrell-type two-dimensional gel electrophoresis (1). This protein, designated originally as p26-6.9 by its molecular weight and pi, was found to be identical with the placental-type glutathione S-transferase(GST-P) reported by Sato et al. (1–3). The glutathione S-transferases are a group of dimeric multifunctional proteins in drug biotransformation and xenobiotics metabolism (4–6). GST-P was first found in placenta, but later also in kidney, lung, and testis (3). The concentrations in these normal cells, however, are much lower than in precancerous and cancerous liver (2, 3). GST-P protein which is hardly detectable in normal rat liver becomes constitutively expressed in hyperplastic nodules and hepatocellular carcinomas at concentrations nearly two orders of magnitude higher in every foci examined irrespective of the kind of carcinogen used (1–3). This extremely high coincidence between the hepatocarcinogenesis and the derepression of this protein prompted us to study the induction mechanisms of this enzyme during cancerous changes of the liver cells.

* This work was supported in part by grants from the Ministry of Education, Science and Culture, especially for Special Project Research, Cancer Bioscience, and from Foundation for Promotion of Cancer Research, Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1724 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J02900.

The abbreviations used are: GST-P, glutathione S-transferase P; kb, kilobases; bp, base pair; SDS, sodium dodecyl sulfate.

We have recently isolated a cDNA clone, pGP5, complementary to GST-P mRNA and determined the primary structure of this protein (7). We have also demonstrated that the dramatic increase in the enzyme activity and the protein of GST-P parallels with the amount of GST-P mRNA (7). The obvious next step is to isolate the genomic clones of this gene and study their structure and regulation. We report here the isolation and characterization of several genomic clones that hybridize to pGP5. We show that one of these clones is most likely the normal gene encoding GST-P and others are processed-type pseudogenes by several criteria.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes, polynucleotide kinase, T4 DNA ligase, DNA polymerase (Klenov fragment), bacterial alkaline phosphatase, and nuclease S1 were purchased from Takara Shuzo (Kyoto, Japan), Sankyo (Tokyo, Japan), and Bethesda Research Laboratories and used according to the manufacturers’ specifications. [γ-32P]ATP (specific activity 7000 Ci/mmol) and [α-32P]dCTP (specific activity 3000 Ci/mmol) were from New England Nuclear. Nick translation kit was purchased from Amersham Corp. A partial HaelI + Alul rat genomic DNA library cloned into Charon 4A was a generous gift of Thomas Sargent, National Institutes of Health.

Screening of the Rat Genomic Library—A rat genomic library was screened with 32P-labeled pGP5 insert (7). Hybridization was carried out in 1 M NaCl, 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 10 × Denhardt’s solution (1 × Denhardt’s solution = 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll), 0.1% SDS, 50 μg/ml denatured salmon sperm DNA, and 1.2 × 106 dpm of 32P-labeled DNA probe/filter at 65 °C for 16 h. Filters were washed twice in 2 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS at room temperature, and then twice in 0.2 × SSC, 0.1% SDS at 65 °C for 30 min each time.

DNA Blot Hybridization—DNA isolated from Sprague-Dawley rat liver was digested by restriction enzymes, electrophoresed on a 0.8% agarose gel. The gel was soaked in 0.25 M HCl for 15 min at room temperature and transferred to a nitrocellulose filter after alkali treatment and neutralization (8). Hybridization and washing were carried out as described for the plaque hybridization.

DNA Sequencing—Fragments of DNA digested with various restriction endonucleases were ligated to the replicative form of M13 mp10 or 11 cleaved with restriction enzymes which produce the complementary ends for fragments (9). Competent Escherichia coli cells, strain JM105, were transfected with the ligated DNA, and β-galactosidase negative plaques were screened. Single-stranded phage DNA were prepared from these clones and used as template for the chain terminator sequencing procedure of Sanger et al. (10).

Preparation of RNA and 5’ End Analysis—Total cellular RNA was prepared from an acetyaminofluorene-induced rat hepatoma using the guanidine thiocyanate extraction method (11). The location of the cap site of GST-P mRNA was determined by nuclease S1 protection mapping according to Berk and Sharp (12). A HindIII fragment (101 bp) of GST-P genomic clone corresponding to 5’ end was labeled at the 5’ end with [γ-32P]ATP, denatured with 0.3 M NaOH, and loaded on an 8% polyacrylamide strand separation gel. Both strand-separated fragments were recovered from the gel and hybridized with 50 μg each of total hepatoma RNA in 0.5 M NaCl, 50 mM Tris-HCl, pH 8.0, 1 mM EDTA for 3 h at 65 °C. The reaction mixture was then diluted 10-fold in ice-cold S1 buffer (0.25 M NaCl, 3858
0.03 M sodium acetate, pH 4.6, 1 mM ZnSO$_4$, 100 μg/ml denatured salmon sperm DNA) containing 2000 units/ml of nuclease S1. The S1 digestion was performed at 30°C for 60 min. After the reaction, DNA was precipitated by ethanol and analyzed on an 8 M urea, 10% polyacrylamide gel (13).

RESULTS AND DISCUSSION

Southern Blot Analysis of Rat Genomic DNA—To estimate the number of GST-P genes in the rat genome, we carried out Southern blot hybridization with probes I and II which cover the amino- and carboxyl-terminal portions of GST-P, respectively (Fig. 1A). These probes were hybridized to rat DNA digested with restriction endonuclease, SacI. Fig. 1B shows that five SacI fragments (5.9, 5.6, 3.8, 3.6, and 0.95 kb) of rat DNA hybridize to both probes I and II, suggesting that these fragments contain the entire GST-P coding sequence. Five bands common to both probes were also detected with other enzymes, such as BamHI or BglII (data not shown). These results suggest that there are at least five gene sequences homologous to the GST-P mRNA in the rat genome.

Isolation of Genomic Clones—We have screened $3.0 \times 10^6$ plaques from Charon 4A genomic library with a cDNA clone, pGP5 (7), as a probe. Seven positive plaques were detected. Two of them contained an identical insert. Six nonidentical clones were designated as ChGSTP11, 12, 22, 32, 62, and 71 and were further characterized. The restriction maps of these clones are shown in Fig. 2. Also shown in these maps are the regions that hybridized to pGP5. From the results of restriction mapping analysis, ChGSTP22 insert seems to be included in ChGSTP32, while the other clones show cleavage maps completely different from each other. As discussed below, ChGSTP22 and 32 contain an active gene and the other clones contain processed-type pseudogenes.

Sequence Analysis of the Active Gene—3.8-kb BamHI fragment of ChGSTP22 containing the entire coding region was electrophoretically isolated for DNA sequencing. Several different restriction enzymes were used to yield small overlapping pieces of this fragment. These were then individually cloned into various M13 vectors and sequenced according to the methods of Sanger et al. (10). All the sequence data were compiled to construct a stretch of DNA with the aid of a computer program. This procedure yielded 3236 bp of a genomic DNA sequence (Fig. 3). We conclude that this DNA sequence represents an active GST-P gene of the rat by the following criteria. First, all of the GST-P mRNA sequence is

![Diagram](https://example.com/diagram.png)

**Fig. 1.** A, schematic representation of pGP5-cDNA probes used for genomic Southern blot analysis. Solid and open boxes indicate the amino acid coding and 5' and 3' noncoding regions, respectively. Solid lines represent pUC8 sequences. cDNA probes I and II were prepared by digestion of pGP5 (7) with the indicated restriction endonucleases and isolation of appropriate fragments from preparative agarose gel using DE81 paper. Restriction enzymes are abbreviated as follows: Bg, BglII; Bsa, BstEI; E, EcorI; H, HindIII; S1, Sall; B, Southern blot hybridization patterns of rat DNA with GST-P cDNA fragments. Total DNA was extracted from a Sprague-Dawley rat liver and cleaved to completion by SacI. The digest (15 μg per lane) was electrophoresed on a 0.8% agarose gel and then transferred to a nitrocellulose filter (8). Hybridization with nick-translated DNA probe I and II (indicated on the top of each lane) and washing were performed as described under "Experimental Procedures."

**Fig. 2.** A, Restriction maps of six recombinant phages containing GST-P mRNA sequence. The solid boxes indicate the regions complementary to pGP5; i.e. GST-P mRNA sequence. Probes A, B, and C are unique in the genome and used to determine which genomic fragment detected in Fig. 1 corresponds to each phage clone. Restriction enzymes are abbreviated as follows: B, BamHI; H, HindIII; E, EcoRI; S, SacI; B, detailed restriction map and organization of the active GST-P gene in the 3.8-kb BamHI fragment of ChGSTP22. Restriction sites are abbreviated as in A and as follows: A, AccI; K, KpnI; P, PstI; Sm, SmaI. Solid boxes indicate exons having GST-P mRNA sequence and introns and flanking DNA are indicated by open boxes. TATA box (T-A-T-A-A), translation initiation codon (ATG), and poly(A) addition signal (A-A-T-A-A-A) are indicated by arrows.
**Glutathione S-Transferase P Gene**

Upstream from the transcription start site. Amino acids of the GST-P protein are indicated underlined. Negative numbers indicate nucleotide positions below the start site. The TATA box and poly(A) addition sequence are boxed. The hexanucleotide sequence 5'-GGCGGG-3' is underlined. GT and AG residues present at 5' and 3' boundaries are in double underlines.

---

**FIG. 3. Nucleotide sequence of the rat GST-P gene.**

Amino acids of the GST-P protein are indicated underlined. Negative numbers indicate nucleotide positions below the start site. The TATA box and poly(A) addition sequence are boxed. The hexanucleotide sequence 5'-GGCGGG-3' is underlined. GT and AG residues present at 5' and 3' boundaries are in double underlines.
found in this genomic DNA sequence except for one nucleotide substitution. The nucleotide G at number 6 of cDNA (the first letter of the initiation codon being number 1) is changed to A at the corresponding position of genomic DNA (a dot is placed in Fig. 3). We regard the reason for this difference as the putative cap site, the sequence TATAA-, the expected initiation codon, is most likely the nucleotide to be capped in this mRNA. In addition, the homologous sequence ends with the poly(A)-like stretch at the end of the 3'-untranslated sequence which is possibly derived from poly(A) addition. 

**Analysis of Other Clones Representing Four Pseudogenes—** Since multiple loci were suggested by the genomic Southern blot analyses, we decided to analyze more clones that hybridized with pGP5 in order to characterize the nature of these loci.

ChGSTP11: 9.0-kb EcoRI fragment was subcloned in plasmid, pUC8. As hybridizing regions lay in 3.8-kb BamHI fragment, we isolated this fragment for DNA sequencing. Several different restriction enzymes were used to yield small overlapping pieces of this fragment. These fragments were then individually cloned into M13 mp10 or 11 vectors and sequenced by the Sanger's dideoxynucleotide sequencing method (9, 10). Sequence data shown in Fig. 5 confirm that ChGSTP11 is a pseudogene of the processed type, a type that has been proposed to be originated from reverse transcription of mRNA and reinsertion of a cDNA into the genome. It has no intron, carries a poly(A)-like stretch at the end of the 3'-untranslated sequence which is possibly derived from poly(A) of the mRNA. In addition, the homologous sequence ends through comparison of GC boxes of many Sp1-responsive promoters (20).
precisely at the limits of mRNA, and bordered by an almost complete direct repeat, which is presumably originated from a staggered break at the insertion site. Upstream and downstream sequences of pGP5-homologous region of ChGSTP11 are completely different from corresponding regions of the active gene. Although ChGSTP11 is highly homologous (91%) to the cDNA clone, this gene cannot be functional. This clone has an opal stop codon in the coding region of this pseudogene.

ChGSTP12: Southern blotting of EcoRI digest of this clone revealed that the 2.0-kb EcoRI fragment contains the entire hybridizing region. This fragment was isolated and subcloned in pUC8. There is one HindIII site in this 2.0-kb EcoRI fragment. We sequenced from this HindIII site to both directions about 350 bp each. Sequencing data confirmed HindIII site of this fragment to correspond to that of pGP5. There is a poly(A)-like stretch from 273 bp downstream of the HindIII site, and no intron was present in sequenced 700 bp. Homology to pGP5 is about 90%, but several stop codons were found in every reading frame (data not shown).

ChGSTP62: Hybridizing 3.7-kb EcoRI fragment was subcloned in pUC8. This fragment has the HindIII site in it. We again sequenced from HindIII site to both directions up to 350 bp each (data not shown). This clone also has many features of a pseudogene of processed-type, i.e. it has a poly(A)-like sequence and no intron. Homology of this clone to pGP5 is 91%. ChGSTP71 is also found to be a processed-type pseudogene by sequencing analysis (data not shown).

Homology of this clone to pGP5 is 78%.

Relationship of Cloned Genes to Genomic Southern Blot Bands—Genomic Southern blot analyses show that at least five gene sequences exist in the rat genome that are closely related to pGP5 sequence. The existence of several loci which probably corresponds to the 3.8-kb SacI fragment detected in genomic Southern blot in Fig. 1 (data not shown). We also searched for a unique portion in ChGSTP71 and found that 0.45-kb HindIII/SacI fragment, probe C in Fig. 2, was unique in the genome. Probe C was hybridized to 14-kb SacI fragment of rat DNA (data not shown). This band was not detected in Fig. 1, probably because the homology of ChGSTP71 to pGP5 was significantly lower than that of other clones. From these restriction maps and Southern blot analyses, it is clear that four clones are now obtained out of five loci detected in genomic Southern blot analysis on total rat DNA (Fig. 1). Hybridization signals of the band comprising ChGSTP22 and 32, which are now identified as the active gene, are the strongest of all and those of ChGSTP11, 12, and 62, which are determined to be pseudogenes, are less strong. The signal of 0.95-kb SacI fragment is the weakest among five fragments detected in the genomic Southern blot, and a clone containing this fragment has not been obtained so far. Weakness of the signal of this fragment may well be due to the high degree of mismatches of the sequence. We have a clone, ChGSTP71, which does not correspond to any fragment detected in genomic Southern blot. Its hybridization signal was rather weak compared to other clones, suggesting that the locus corresponding to this clone may not be detected in genomic Southern blot analysis with cDNA. Although it is difficult to determine the exact number of GST-P gene family, these results strongly suggest that only one active gene of GST-P exists in rat genome. Further support to this conclusion is obtained in the following way. Total rat DNA was restricted with BamHI, EcoRI, or SacI, and Southern blot hybridization was performed with 0.6-kb Smal fragment (probe A in Fig. 2) in the fifth intron of ChGSTP22 as a probe. A single band appeared for rat DNA digested with any of these restriction enzymes (3.8-kb BamHI, 4.8-kb EcoRI, 3.6-kb SacI fragments) (data not shown). These bands are predicted from the restriction map of ChGSTP22, and the intensity of the signal corresponded to that expected from a single copy gene.

The mechanism by which dramatic increase of GST-P protein occurs in hyperplastic nodule and hepatocellular carcinoma is presently unknown. Northern blot analysis shows that the content of GST-P mRNA in normal liver, hyperplastic nodules, and hepatocellular carcinoma is proportional to this protein content, showing a two orders of magnitude increase in the former two. Northern blot analysis also shows that the content of GST-P mRNA in normal liver, hyperplastic nodules, and hepatocellular carcinoma is proportional to this protein content.
increase in the last two tissues (1, 7). This result suggests that the increase of the GST-P enzyme is at least partly controlled at the transcriptional level. Further use of the genomic clone in in vivo transfection and in vitro transcriptional systems will certainly help elucidate the induction mechanism of GST-P during hepatocarcinogenesis.

REFERENCES
1. Sugioka, Y., Fujii-Kuriyama, Y., Kitagawa, T., and Muramatsu, M. (1985) Cancer Res. 45, 365-378
2. Sato, K., Kitahara, A., Satoh, K., Ishikawa, T., Tatematsu, M., and Ito, N. (1984) Gann 75, 199-202
3. Satoh, K., Kitahara, A., Soma, Y., Inaba, Y., Hatayama, I., and Sato, K. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 3964-3968
4. Jakoby, W. B. (1978) Adv. Enzymol. 46, 383-414
5. Chasseaud, L. F. (1979) Adv. Cancer Res. 29, 175-273
6. Booth, J., Boyland, E., and Sims, P. (1961) Biochem. J. 79, 516-524
7. Sugioka, Y., Kano, T., Okuda, A., Sakai, M., Kitagawa, T., and Muramatsu, M. (1985) Nucleic Acids Res. 13, 6049-6057
8. Southern, E. (1975) J. Mol. Biol. 98, 503-517
9. Messing, J., Crea, R., and Seeburg, P. H. (1977) Nucleic Acids Res. 9, 309-321
10. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
12. Berk, A. J., and Sharp, P. A. (1977) Cell 12, 721-732
13. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564
14. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
15. Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L., and Luskey, K. L. (1984) Cell 38, 275-285
16. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L., and Weissman, S. M. (1978) Science 200, 494-502
17. Fiers, W., Contreras, R., Haegeman, G., Rogievs, R., van de Voorde, A., van Heuverswyn, H., van Harreweghe, J., Volckaert, G., and Ysehaert, M. (1978) Nature 273, 115-120
18. Dynan, W. S., and Tjian, R. (1983) Cell 35, 79-87
19. Dynan, W. S., and Tjian, R. (1983) Cell 32, 669-680
20. Kadonaga, J. P., Jones, K. A., and Tjian, R. (1986) Trends Biochem. 11, 29-23