Gene Expression of Rat Glutathione S-Transferases

EVIDENCE FOR GENE CONVERSION IN THE EVOLUTION OF THE Ys MULTIGENE FAMILY*

(Received for publication, November 9, 1987, and in revised form, March 25, 1988)

Hui-Chen J. Lai, Biao Qian, Gregory Grove, and Chen-Pei D. Tu†

From the Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

We have characterized a cDNA with complete coding sequence for the rat liver glutathione S-transferase subunit 4 (Yb4) isolated from a constructed λgt10 cDNA library. The nucleotide sequence(s) reported in this paper have been submitted to the GenBankTM/EMBL Data Bank with accession numbers J03814.

The glutathione S-transferases (EC 2.5.1.18) are a family of dimeric proteins that are multifunctional in drug biotransformation and in xenobiotic metabolism (for a review, see Ref. 1). The protein purification and substrate specificity patterns of rat glutathione S-transferases indicate the existence of at least eight classes of subunits: Yb, (Mr = 24,400), Yb (Mr = 25,600), Yb (Mr = 26,300), Yb (Mr = 27,000), Yb (Mr = 28,000), Yb (Mr = 26,300), Yb (Mr = 25,000), and Yb (Mr = 24,000) (2-10).

Analyses of rat glutathione S-transferase cDNAs (11-18) and their genomic DNA hybridization patterns suggest that rat glutathione S-transferases are encoded by a minimum of four separate gene families which constitute a gene superfamily (17, 18). The diverse structures of the glutathione S-transferases are probably required for their multiple functions in the metabolism of a multitude of chemical compounds as well as physiological substrates.

The expression of glutathione S-transferases has been reported to be tissue-specific (5, 6, 9, 19). Rats do not express Yb and Yb subunits in hearts (6, 20, 21) and testes (19), Yb subunits in kidneys (5, 6), or Yb, Yb, and Yb subunits in livers (6, 9, 10, 16). RNA blot hybridization results, with glutathione S-transferase cDNA probes, suggest that regulation of tissue specificity in glutathione S-transferase expression may occur before the maturation of mRNAs (11, 12, 17, 18).

The expression of glutathione S-transferases can be induced. For example, glutathione S-transferase activities and mRNA levels are elevated in rat livers and other tissues upon treatment with phenobarbital, 3-methylcholanthrene, and trans-stilbene oxide (22-25). Using nuclear run-off assays and glutathione S-transferase cDNA probes, Ding et al. (25, 26) have found that transcriptional activation of the glutathione S-transferase genes is responsible for the increased levels of mRNAs.

To begin dissecting the molecular mechanism of glutathione S-transferase tissue-specific expression and induction by xenobiotic compounds, we have analyzed the genomic structures of several glutathione S-transferase genes. In this communication, we report the characterization of two Yb genes, Yb and a new Yb, and the heterospecific expression in Escherichia coli of a complete Yb cDNA.

**EXPERIMENTAL PROCEDURES**

Materials—Nucleotides, dNTPs, and ddNTPs, were obtained from Pharmacia LKB Biotechnology Inc. [α-32P]dATP (specific activity, ~3200 Ci/mmol) and [γ-32P]ATP (specific activity, ~7000 Ci/mmol) were purchased from ICN Pharmaceuticals. [α-32P]dATP was the product of Du Pont-New England Nuclear or Amersham. Restriction endonucleases, DNA polymerase I, Klenow fragment, T4 ligase, and T4 polynucleotide kinase were purchased from New England Biolabs or IBI (New Haven, CT). S1 nuclease and the cDNA synthesis kit were from Boehringer Mannheim. Reverse transcriptase was from Life Science, Inc. (St. Petersburg, FL). The heptadecanoic acid nucleotide (17-mer) (5' CCCAGTGT CATAGGCAT 3') was synthesized by a DNA synthesizer of Applied Biosystems, Inc. The two 12-mers complementary to the λ cohesive end sequences were purchased from New England Biolabs. The λgt10 cDNA library and packaging extracts (Gigapack Gold) were purchased from Stratagene (San Diego, CA).

Construction of λgt10 cDNA Library from Rat Liver Poly(A)+ RNAs—The double-stranded cDNAs were synthesized according to manufacturer’s procedure beginning with ~5 μg of rat liver poly(A)+ RNA. It was size-fractionated on a Sepharose CL-4B column before and after EcoRI linker (GGAATTCC) addition and EcoRI digestions. The cDNAs were ligated to λgt10 arms at 1:1 molar ratio before transfection into E. coli strain C600 hfl according to the manufacturer’s procedure.

Isolation of a Nearly Full-length Glutathione S-Transferase Yb cDNA from the λgt10 cDNA Library—The 17 mer, 5' CCCAGTGT CATAGGCAT 3', derived from the genomic DNA sequence (see below), was 5' end-labeled by T4 polynucleotide kinase and [γ-32P]ATP. Screening of the λgt10 cDNA library was carried out with ~250,000 plaques under conditions described in Maniatis et al. (29). A single positive clone was obtained and designated as λGTR187C. The cDNA insert of λGTR187C (two EcoRI fragments) was excised from the λgt10 arms by partial EcoRI digestion, followed by subclon-
Glutathione S-Transferase Gene Structure and Expression

...purified protein was performed on an LKB 2117 Multiphor system... fraction 2 was identified with one of the Yb genes. The single-stranded DNA probe (198 nucleotides) was hybridized to 5 μg of rat liver poly(A) RNA in 10 μl of hybridization solution (0.5 μM NaCl, 20 mM PIPES (pH 6.4), and 1 mM EDTA) at 37 °C for 3 h. The mixture was then diluted to 100 μl of S1 buffer (350 mM NaCl, 30 mM NaOAc (pH 4.5), and 1 mM ZnSO4) and digested with 300 units of S1 nuclease at 37 °C for 40 min. The protected DNA was purified by ethanol precipitation before gel electrophoresis (15% polyacrylamide gel with 8 μM urea).

**Primer Extension Assay on the Yb2 Gene**—The 17-mer (5' CCCAGTGTGATGAGCCCT 3') was labeled with [γ-32P]ATP and 74 nucleotide kinase (39). Rat liver poly(A) RNA (5–5 μg) was mixed with ~5 pmol of kinase-labeled 17-mer in a 20 μl volume containing 50% formamide, 40 mM PIPES (pH 6.5), 400 mM NaCl, and 1 mM EDTA. The mixture was heated at 65 °C for 1 min followed by incubation at 37 °C for 1 h. The annealed 17-mer, mRNA "complex" was recovered by ethanol precipitation and resuspended in 20 μl of reverse transcriptase buffer (50 mM Tris-HCl (pH 8.3), 10 mM MgCl2, 70 mM KCl, 10 μM dithiothreitol, and 500 μM dNTPs). The reaction was initiated by adding 19 units of reverse transcriptase and incubated at 20 °C for 15 min, 39 °C for 15 min, and 37 °C for 1 h.

**RESULTS AND DISCUSSION**

Characterization of the Yb2 cDNA Clone pGTR187C—DNA sequence analysis of the cDNA insert of pGTR187C revealed that this cDNA contained the complete coding sequences of Yb2 (subunit 4). The sequence started at position 32 in Fig. 1 with CGATG and covered all the other exons and the poly(A) sequence as in the previously described incomplete Yb2 cDNA pGTR187 (17) for a total of 1004 nucleotides of unique sequence.

Characterization of Yb2 Subunit of Rat Glutathione S-Transferee—The isoelectric focusing of purified protein was performed on an LKB 2117 Multipher system using a thin layer PAG plate (pH 3.5–9.5) (LKB) at 3 °C according to manufacturer's instructions. The 3-ml culture was transferred to the PI markers of the calibration kit (Pharmacia). The protein was eluted with a linear NaCl gradient (0 to 500 mM) in 20 mM ethanolamine (pH 9.40). The effluent was monitored at 280 nm and protein-containing fractions were collected in 0.4-ml fractions.

**Restriction Mapping and DNA Sequencing**—The phage DNAs were purified from plate lysates (29) and mapped relative to BarnHI, EcoRI, HindIII, and FnuDI by the methods of Smith and Birnstiel (32) and Ruckwitz et al. (33) in conjunction with Southern hybridization. The Yb cDNA-positive regions of pGTR14-1 and pGTR15-2 were subcloned into pAT153 or pUC18 for DNA preparation, and into M13mp18 or mp19 vectors (34) for sequencing (35, 36).

S1 Mapping of Yb2 Gene's Transcription Initiation Site—The reaction mixture contained template DNA previously annealed to the primer (5' CCCACCTGTGATGAGCCCT 3'), ~7.5 pmol of [α-32P]ATP, three other dNTPs (~640 pmol each), 10 mM Tris-HCl (pH 7.5), 5 mM MgCl2, 7.5 mM dithiothreitol, and 5 units of DNA polymerase I Klenow fragment. The reaction was incubated at room temperature for 20 min followed by a 20-min chase in the presence of 5 mmol of each dNTP. The products were digested with BarnHI and purified by gel electrophoresis under denaturing conditions (3.5% polyacrylamide gel in the presence of 5 μM urea). The 51 mapping conditions were described by Benley and Groudine (38). The single-stranded DNA probe (198 nucleotides) (100,000 cpm) was hybridized to 5 μg of rat liver poly(A) RNA in 10 μl of hybridization solution (0.5 μM NaCl, 20 mM PIPES (pH 6.4), and 1 mM EDTA) at 65 °C for 3 h. The mixture was then diluted to 100 μl of S1 buffer (350 mM NaCl, 30 mM NaOAc (pH 4.5), and 1 mM ZnSO4) and digested with 300 units of S1 nuclease at 37 °C for 40 min. The protected DNA was purified by ethanol precipitation before gel electrophoresis (15% polyacrylamide gel with 8 μM urea).

**Primer Extension Assay on the Yb2 Gene**—The 17-mer (5' CCCAGTGTGATGAGCCCT 3') was labeled with [γ-32P]ATP and 74 nucleotide kinase (39). Rat liver poly(A) RNA (5–5 μg) was mixed with ~5 pmol of kinase-labeled 17-mer in a 20 μl volume containing 50% formamide, 40 mM PIPES (pH 6.5), 400 mM NaCl, and 1 mM EDTA. The mixture was heated at 65 °C for 1 min followed by incubation at 37 °C for 1 h. The annealed 17-mer, mRNA "complex" was recovered by ethanol precipitation and resuspended in 20 μl of reverse transcriptase buffer (50 mM Tris-HCl (pH 8.3), 10 mM MgCl2, 70 mM KCl, 10 μM dithiothreitol, and 500 μM dNTPs). The reaction was initiated by adding 19 units of reverse transcriptase and incubated at 20 °C for 15 min, 39 °C for 15 min, and 37 °C for 1 h.
FIG. 1. Nucleotide sequence of the Yb2 gene. The transcription initiation site is designated as the +1 position. Sequences further upstream are in negative numbers. The important sequences for transcription and poly(A) addition signals (AATAAA) are underlined. The translation initiation codon ATG is located at nucleotides 34–36. The amino acids encoded by the exons are listed above the nucleotide sequence. The stop codon for the translation is indicated by an asterisk.

The noncoding sequence of pGTR187 was used as the probe for the genomic blot hybridization (data not shown). The amino acid sequence of a Yb-sized glutathione S-transferase subunit, designated as subunit 4, has been determined (44). By comparing the deduced amino acid sequence of the anionic Yb2 cDNA clone in XGTR187C to that determined for subunit 4 by protein sequencing, we found only 1 residue, Trp, at amino acid position 147 (Fig. 1), which was incorrectly identified as Ser by Alin et al. (44).

The overall amino acid sequence identity between the anionic Yb2 and the basic Ybl polypeptides is 78% (17). The extent of homology is distributed relatively evenly throughout except for exon 3. Exon 3 encodes a region that is highly conserved between the Yb2 and Ybl polypeptides; there is only one nucleotide substitution between the two DNA segments. We speculate that exon 3 may be crucial for maintaining certain structural or functional aspects common to these two Yb subunits.
Fig. 2. Characterization of rat glutathione S-transferase Yb subunit expressed from E. coli by isoelectric focusing (panel A) and Western blot analysis (panel B). Panel A, lane 1, E. coli-expressed Yb protein purified by Mono Q column chromatography on FPLCTM; lane 2, pl marker proteins (Pharmacia). Panel B, lane 1, rat liver glutathione S-transferases purified by S-hexyl GSH affinity chromatography; lane 2, Yb subunit of rat glutathione S-transferase expressed and purified from E. coli.

Fig. 3. Restriction map and sequencing strategy of the anionic Yb gene in XGTR4-1. The XGTR4-1 insert was mapped relative to EcoRI (E) and HindIII (H) (top of the figure). Fragments that were positive to the pGTR187 cDNA insert by Southern hybridization were mapped further with more restriction endonucleases: BamHI (B), BglII (Bg), PvuII (Pv), PstI (P), and SmaI (S). Exons are represented by rectangles. Open rectangles represent 5' (left) and 3' (right) noncoding sequences. Filled rectangles represent coding sequences. The Charon 30 arms are represented by incomplete open rectangles. The sequencing results were obtained from each restriction site in the direction and extent defined by each arrow. Other sites used are indicated by a lowercase letter in front of the corresponding arrow. They are: r, Rsal; s, Sau3AI; h, HaeIII; d, DraI. The 17-mer has the sequence 5' CCCAGTGTCATAGGCAT 3' which was used as the primer in sequencing across the first BamHI site to the left of the diagram.

The exon-intron structure of the Yb gene is quite different from that of the Ya gene (45) and the Yb gene (46). The Yb gene spans less than 5 kb and splits into eight exons, whereas the Ya gene (~11 kb) and Yp gene (~3 kb) consist of seven exons each.

Transcriptional Initiation Site of the Yb Gene—The initiation site of transcription was determined by S1 nuclease mapping and primer extension experiments (Fig. 4). These two different approaches gave the same conclusion, i.e. that transcription initiates at the A residue which is designated +1 (Fig. 1). The 3' end of the major S1-protected DNA fragment, i.e. the 5' end of the mRNA, corresponds to the T residue which complements the A residue at +1. There is a second band one nucleotide longer in the S1-protected DNA that may result from the partial protection by the capped structure in the glutathione S-transferase poly(A) RNA. The primer-extended DNA is slightly slower in its mobility on the gel because it is phosphorylated at the 5' end. The 17-mer used for primer extension is complementary to nucleotides 34–50 in Fig. 1. The primer extension product was specific for the Yb gene, rather than for the basic Ybl gene (pGTR200), because the size of the 5' noncoding region of the Yb gene as determined by these experiments is at least three nucleotides shorter than the 5' noncoding region of the Yb1 gene (pGTR200) (18). There was no primer extension product for the basic Ybl gene presumably because of a nucleotide mismatch between the 17-mer and the corresponding sequence of the Yb1 gene.

The 5' flanking region of the Yb gene contains two ele-
ments, the CCAAT box (47) and the ATA box (48), commonly found in promoters recognized by eukaryotic RNA polymerase II. There are two copies of CCAAT (−84 to −80 and −76 to −72) in the 5′ flanking region of the Ybl gene. Duplicated CCAAT boxes were also seen in the human γ-globin genes (49). The putative Ybl ATA box, TATCA (−28 to −24), differs from the consensus ATA box (TATAA).

Identification and Characterization of a New Ybl Gene—The 6.2-kb region on XGTR15-2 identified by hybridization to pGTR200 (Ybl) cDNA (18) was subcloned into M13 DNA and pUC18/19 DNA after PstI and BglII digestions for sequence analysis. Analysis of 6210 nucleotides indicated that the exon sequences do not match the Yb3 cDNA (pGTR200) (data not shown). The gene on XGTR15-2 contains no apparent defects and is very similar in structure to the Yb2 gene. The length of the exons are identical between the two genes, predicting a Yb class subunit of 218 amino acids. We designate this new Ybl gene as Ybl4.

The Ybl4 gene has an ATA box sequence TATAAT at 59 nucleotides upstream of the ATG initiation codon. This is comparable to the ATA box sequence in the functional Yb2 gene where the distance is 61 nucleotides (Fig. 1). The CCAAT box sequence, which occurs twice in the Ybl4 gene, is obviously missing in the Ybl gene. The sequence GCAGT, which may be the best fit for the "CCAAT" sequence, exists 73 bp upstream of the TATAA sequence.

We did a quantitative pairwise comparison of the four Ybl subunit sequences (17, 18, 50, 51) using the method of Perler et al. (52) to analyze percentages of amino acid replacement mutations (Table I). The Ybl4 coding sequence is most closely related to the Ybl subunit (40 amino acid replacements or 8.03% divergence on replacement sites).

The close evolutionary relationship between Ybl and Ybl4 is further supported by the obvious homology (~70%) between the two 3′ noncoding sequences (data not shown). None of these DNA segments, however, shares sequence homology with the corresponding regions of Ybl or Yb3 cDNAs (17, 50, 51) either experimentally in DNA hybridization or by computer analysis.

Evolution of the Ybl Multigene Family—The Ybl4 gene is more closely related to the Ybl gene (40 amino acid substitution relationships to the Ybl4 and Ybl5 genes (63 amino acid substitutions) in the coding sequences (Fig. 5). The most obvious regions of sequence divergence lie between residues 104 and 117 and between residues 201 and 218 and are encoded by exons 5 and 8, respectively, in the Ybl2 and Ybl4 genes. These differences may be responsible for the difference in their substrate specificities. Using the constant average rate of molecular evolution for mammalian genes cited by Li et al. (53), we estimate that the gene duplication event leading to multiple Ybl genes may have occurred more than 158 million years ago. The Ybl4 gene may have begun to diverge from the Ybl4 gene around the time of mammalian radiation (80–85 million years ago) (Table I).

Sequence similarity between the Ybl4 and Ybl5 genomic DNAs extends beyond exons and intron-exon junction regions. Matrix analysis revealed extensive homology even in introns, especially introns 3, 4, and 6 and part of intron 2 (Fig. 6). Other regions of intron sequence homology are found near the exon boundaries (Table II).

Conservation of intron sequences is best known from the
Glutathione S-Transferase Gene Structure and Expression

**TABLE II**

*Summary of intron sequence homology between the Yg and Yα genes*

<table>
<thead>
<tr>
<th>DNA location</th>
<th>Length</th>
<th>No. of replacements</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' region of intron 1</td>
<td>28</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>3' region of intron 1</td>
<td>17</td>
<td>3</td>
<td>17.6</td>
</tr>
<tr>
<td>5' region of intron 2</td>
<td>35</td>
<td>10</td>
<td>27.7</td>
</tr>
<tr>
<td>3' region of intron 2</td>
<td>32</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Intron 3</td>
<td>282</td>
<td>16</td>
<td>5.6</td>
</tr>
<tr>
<td>Intron 4</td>
<td>100</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5' region of intron 5</td>
<td>11</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>3' region of intron 5</td>
<td>58</td>
<td>9</td>
<td>15.5</td>
</tr>
<tr>
<td>Intron 6</td>
<td>80</td>
<td>10</td>
<td>12.5</td>
</tr>
<tr>
<td>5' region of intron 7</td>
<td>27</td>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>

* DNA length is relative to the sequence in Yα gene.
* Replacements include nucleotide substitutions, deletions, and insertions.

study of human fetal Gγ- and Aγ-globin genes (54), where these two members of the β-globin gene family share ~98% sequence homology in their two introns. Among other gene systems, the findings of the P450 b/e-related genes (55, 56) may have direct bearings on our present work. Between the two major phenobarbital-inducible rat liver cytochrome P-450 genes (~98% homologous in nucleotide sequence), the distribution of nucleotide substitutions is not random. Among the 36 divergent nucleotides which are located mainly in the 3' halves of these two genes, 14 substitutions reside in two short highly divergent segments encoded by exon 7 in the case of the P-450e gene (54, 55).

Sequences of the 500-bp region encompassing exons 7 and 8 and the intervening intron 7 of the different genes of the cytochrome P-450 b/e family are also homogenized. This is unusual because in general members of multigene families show much higher sequence conservation in exons than in introns. Clustering of nucleotide substitutions also has been observed in the two closely related glutathione S-transferase-Y, subunit cDNAs (11).

In both reports (54–56), gene conversion, a nonreciprocal recombination process, was proposed as the mechanism in maintaining sequence homogeneity. The evidence in the γ-globin genes is particularly strong because the authors found regions of homology between adjacent γ-globin genes on one human chromosome that were not observed on the homologous chromosome for the allelic genes (54).

Gene conversion may be an important mechanism for generating sequence diversity (57–61). The generation of some Ty transposable elements in yeast may result from a gene conversion event (or events) between two different Tys in the gene. Gene conversion may be an important mechanism for generating diversity in the genes encoding xenobiotic enzymes such as the glutathione S-transferases.

Acknowledgments—We thank Dr. S. M. Wang for sharing the rat genomic DNA library, Laura Liermann-Kelly for the initial screening of the Agt10 cDNA library with the oligonucleotide probe, and Eileen McConnell for typing the manuscript.

REFERENCES

Glutathione S-Transferase Gene Structure and Expression