

Structural analysis of a rat liver glutathione S-transferase Ya gene

(gene structure/promoter/heteroduplex analysis/primer extension)

CLAUDIA A. TELAKOWSKI-HOPKINS, GAIL S. ROTHKOPF, AND CECIL B. PICKETT

Department of Molecular Pharmacology and Biochemistry, Merck Sharp & Dohme Research Laboratories, Rahway, NJ 07065-0900

Communicated by P. Roy Vagelos, August 28, 1986

ABSTRACT We have isolated and characterized a complete structural gene encoding a rat liver glutathione S-transferase (glutathione transferase; EC 2.5.1.18) Ya subunit. The gene spans \approx 11 kilobases and is comprised of seven exons separated by six introns. A sequence similar to the Goldberg-Hogness promoter ("TATA" box), TATTA, is located 32 base pairs upstream from the transcription initiation site. Exons 2 and 4 of the glutathione S-transferase gene encode amino acid sequences of the Ya subunit that are highly conserved in the Yc subunit, whereas exons 3 and 5 encode amino acids that are divergent in the Yc subunit. These data suggest that exons 2 and 4 may encode domains of the Ya subunits that have similar structural or functional properties to the corresponding domains in the Yc subunit (e.g., glutathione binding site), whereas exons 3 and 5 may encode domains of the Ya subunit that have unique structural or functional properties to the corresponding domains in the Yc subunit (e.g., substrate binding site).

The glutathione S-transferases (glutathione transferase; EC 2.5.1.18) are a family of enzymes that catalyze the conjugation of glutathione to a variety of electrophilic ligands. In addition, the transferases bind heme and bilirubin as well as various exogenous hydrophobic compounds with high affinity (1-3). The enzymes are comprised of binary combinations of at least seven major subunits, Y α , Ya, Yb₁, Yb₂, Yc, Yn, and Yp, which are electrophoretically distinguishable on one-dimensional NaDodSO₄/polyacrylamide gels (4-7). Peptide mapping experiments and carboxyl-terminal sequence analysis have indicated that Ya and Yc subunits are comprised of a mixture of at least two microheterogeneous polypeptides (8-10).

Our laboratory has reported the construction of Ya, Yc, Yb₁, and Yb₂ cDNA clones (11-14). We have used these clones in RNA blot hybridization and nuclear run-off assays to demonstrate that the rat liver glutathione S-transferase genes are transcriptionally activated by phenobarbital and 3-methylcholanthrene (11, 14, 15). DNA sequence analysis of the four cDNA clones indicates that the Ya and Yc genes are members of the same gene family, whereas the Yb₁ and Yb₂ genes are members of a second glutathione S-transferase gene family. Genomic blots using the Ya and Yc cDNAs as probes suggest the presence of at least five to seven Ya and/or Yc genes in the rat (16).

In the present study, we have isolated a complete structural gene encoding a Ya subunit of the rat liver glutathione S-transferases. The Ya structural gene spans \approx 11 kilobases (kb) and is comprised of seven exons separated by six introns. Nucleotide sequence analysis of the exons of the glutathione S-transferase structural gene indicates that the sequence is equivalent to the Ya cDNA clone, pGTB38, which has been isolated and characterized by our laboratory (11).

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.

MATERIALS AND METHODS

Screening of Rat Genomic Library. One million plaques from a rat *Hae* III genomic library (gift of T. D. Sargent, R. B. Wallace, and J. Bonner) were screened with a modification of the Benton and Davis procedure allowing *in situ* amplification of plaques on nitrocellulose filters (17). The rat genomic library was constructed by partial digestion of high molecular weight DNA with *Hae* III and insertion into λ Charon 4A. Plaque hybridization was performed essentially as described by Maniatis *et al.* (18).

Subcloning of Genomic Fragments. *Eco*RI fragments of λ 45-15 were mixed with linearized, dephosphorylated pBR325 at a final DNA concentration of 0.4 μ g/30- μ l reaction volume. T4 ligase was added and the reaction mixture was incubated at 15°C for 16 hr. *Escherichia coli* HB101 was then transformed with ligated DNAs using the CaCl₂ procedure (19) and transformants were plated out on L plates containing tetracycline. Recombinants were selected on the basis of their sensitivity to chloramphenicol.

Restriction Mapping of Genomic Subclones. Restriction maps of genomic subclones were constructed by the method of Smith and Birnstiel (20) using 5' and 3' end-labeled fragments.

In Vitro Labeling. DNA was labeled *in vitro* either with [α -³²P]dCTP or [α -³²P]dATP by nick-translation, at the 5' end with [³²P]ATP by T4 polynucleotide kinase, or at the 3' end with [³²P]ddATP using deoxynucleotidyltransferase.

Nucleotide Sequence Analysis. The chemical method of Maxam and Gilbert was used for DNA sequence analysis. Appropriate restriction fragments were 5' end-labeled with polynucleotide kinase or 3' end-labeled using terminal transferase and used for DNA sequence analysis. In some cases it was necessary to sequence fragments using the M13 cloning and dideoxynucleotide sequencing method (21).

Primer Extension Analysis. An oligonucleotide primer (5' GTGCTTCACTGTCTAGCGAG 3'), which represents a 20-base-pair (bp) fragment complementary to the 5' untranslated region of the Ya mRNA, was synthesized and hybridized to 26 μ g of rat liver mRNA. The primer was extended using reverse transcriptase (22) and electrophoresed on a 15% polyacrylamide/8 M urea sequencing gel. Alongside the extended primer, we ran a sequence mix of M13mp19 obtained with the dideoxy chain-termination method. The mp19 sequence provided us with the appropriate size markers.

Heteroduplex Analysis. Heteroduplex analysis between the linearized Ya cDNA clone pGTB38 and the genomic clone λ GTB45-15 was performed as described by Yamada *et al.* (23).

RESULTS

Isolation of a Rat Glutathione S-Transferase Structural Gene. In a previous study from our laboratory, we reported

Abbreviations: bp, base pair(s); kb, kilobase(s).

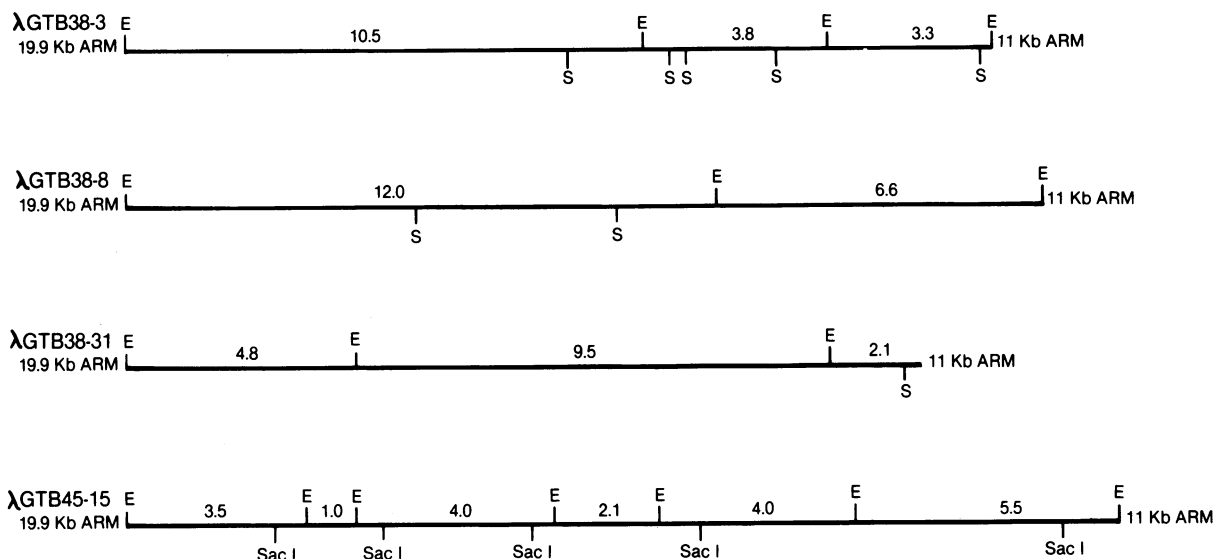


FIG. 1. Restriction endonuclease map of λGTB45-15; restriction endonuclease maps of genomic clones (λGTB38-3, λGTB38-8, and λGTB38-31) isolated from the rat *EcoRI* library and the restriction endonuclease map of genomic clone λGTB45-15 isolated from the rat *Hae* III library. Restriction sites are shown for *EcoRI* (E), *Sma* I (S), and *Sac* I.

the isolation of three unique genomic fragments (λGTB38-8, λGTB38-31, λGTB38-3) from an *EcoRI* rat liver library that hybridized to our Ya cDNA clones (16). However, further examination of these clones revealed that none contained a full-length glutathione *S*-transferase structural gene. Subsequent screening of a *Hae* III library identified three additional clones. Two of the clones overlapped λGTB38-3, providing a 1.1-kb extension in the 3' direction (16). The third genomic clone, λGTB45-15, had a unique restriction endonuclease map and did not overlap any of the other genomic clones (Fig. 1). Southern blots of the isolated phage DNA from λGTB45-15 revealed hybridization to the 5' and 3' regions of our Ya cDNA clones (data not shown). These data suggested that λGTB45-15 most likely represented a full-length glutathione *S*-transferase Ya structural gene.

DNA Sequence and Heteroduplex Analysis of λGTB45-15. The rat liver glutathione *S*-transferase Ya structural gene is contained within five *EcoRI* fragments of λGTB45-15. All five genomic fragments, which ranged in size from 5.5 kb to 1.0 kb, were subcloned into pBR325 and subjected to restriction endonuclease mapping. Regions of the subcloned fragments that hybridized to the cDNA clones were sequenced by the Maxam-Gilbert chemical sequencing procedure. The exon-intron structure of the Ya gene is presented in Fig. 2. The gene spans ≈11 kb and is separated into seven exons by six introns. The nucleotide sequences of each exon and all exon-intron junctions are presented in Fig. 3. Exon 1 is only 43 bp in length and encodes the 5' untranslated region of the mRNA from nucleotides 1 to 43. Exon 2 is 109 bp in length and encodes amino acids 1-29 of the Ya subunit and nucleotides 44-65 of the 5' untranslated region of the mRNA. Exon

3 is 52 bp and encodes amino acids 30-46 of the Ya protein. Exons 4 and 5 are 133 bp and 142 bp in length and encode amino acids 47-91 and 92-139, respectively. Exons 6 and 7 are 132 and 234 bp in length, respectively. Exon 6 encodes amino acids 140-183 of the Ya polypeptide, whereas exon 7 encodes amino acids 184-222 of the Ya polypeptide and 121 bp of the 3' untranslated region of the Ya mRNA.

Interestingly, the Ya amino acid sequence encoded by exon 3 represents the region that is the most divergent between the Ya and Yc polypeptides. There is only a 36% amino acid sequence homology between the Ya and Yc polypeptides in this region despite an overall identity of 66% between the two polypeptides. Similarly, exon 5 encodes an amino acid sequence that is very divergent between the Ya and Yc polypeptides. There is only a 51% amino acid sequence identity between the Ya and Yc polypeptides in this region.

Exons 2 and 4 encode amino acid residues of the Ya subunit that are the most highly conserved in the Yc polypeptide. The amino acid sequences of the Ya polypeptide encoded by exons 2 and 4 are 86% and 91% identical, respectively, to the corresponding sequences in the Yc polypeptide. Exons 6 and 7 encode regions of the Ya polypeptide that have an amino acid sequence identity that is similar to the overall sequence homology of the Ya and Yc polypeptide (≈66%).

The approximate sizes of the introns in the glutathione *S*-transferase structural gene were determined by a combination of heteroduplex analysis and restriction endonuclease mapping. The heteroduplex formed between λGTB45-15 and linearized pGTB38 is presented in Fig. 4. Since the Ya cDNA clone pGTB38 has only 17 bp of exon 1, no stable hybrid was

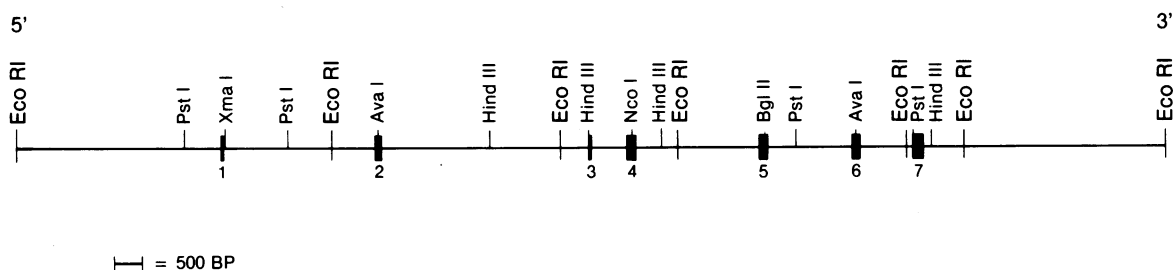


FIG. 2. Exon-intron structure of the rat liver glutathione *S*-transferase Ya gene. All exons are indicated by black boxes. The 5' to 3' orientation of the gene is indicated on the restriction endonuclease map.

atcaacgaagtctggaactcttgactctatgggtctctgtgggaaggctgttccctattggccccacacctgggt a a
 -32 +1 Exon 1
 gaattgtcaccatattaaagtggcgtgcacactcctctggagctGGAGTTGGGAGCTGAGTGGAGAAGAAGCCACGATGC
 TCGTAGTcagtagtactctcttttacaaccc-----2.3 Kb-----gccttactcaacacactccatagc
 acaggctccccgaactgatctctgcctttctctcctagACAGTGAAGCACAGTTGCTGCTATG TCT GGG AAG CCA
 M S G K P
 V L H Y F N A R G R M E C I R W L L A A
 Exon 2
 GTG CTT CAC TAC TTC AAT GCC CGG GGC AGA ATG GAG TGC ATC CGG TGG CTC CTG GCT GCA
 A G V E F E E
 GCA GGA GTG GAG gtagggtggf-----3.5 Kb-----tattttctgctttccag TTT GAA GA
 K L I Q S P E D L E K L K K
 Exon 3
 G AAG CTT ATA CAG AGT CCA GAA GAC TTG GAA AAG CTA AAG AAA Ggtaataccaagt-----
 D G N L M F D Q V P H V E I
 --0.65 Kb-----tttcagagAC GGG AAT TTG ATG TTT GAC CAA GTG CCC ATG GTG GAG AT
 D G M K L A Q T R A I L N Y I A T K Y D
 Exon 4
 T GAC GGG ATG AAG CTG GCA CAG ACC AGA GCC ATC CTC AAC TAC ATC GCC ACC AAA TAT GA
 L Y G K D M K E R A L
 C CTC TAT GGG AAG GAC ATG AAG GAG AGA GCC CTgtacgatgtttctgttcttacctcagtggaagc
 aagggaaggtttgggtccctcctcagtgagcaagagagtgccagagtttgggtcaggagcagg-----2.1
 I D M Y S E G I L
 Kb-----cccttgacaatttggaattttatctttcagG ATT GAC ATG TAT TCA GAG GGT ATT TT
 D L T E M I I Q L V I C P P D Q R E A K
 Exon 5
 A GAT CTG ACT GAA ATG ATT ATC CAA TTG GTA ATA TGT CCC CCA GAC CAA AGA GAA GCC AA
 T A L A K D R T K N R Y L P A F E K
 G ACC GCC TTG GCA AAA GAC AGG ACC AAA AAC CGG TAC TTG CCT GCC TTT GAA AAG gtaagt
 aggtccatgaagtctggggac-----1.5 Kb-----agggtgtcgccttgagagacgtggggagagt
 V L K S H G Q
 gggggccccggcccgactgtttctgttcataatcagctcttttatttttgcagGTG TTG AAG AGC CAT GGC CAA
 D Y L V G N R L T R V D I H L L E L L L
 Exon 6
 GAC TAC CTT GTA GGT AAC AGG CTG ACC CGG GTA GAC ATC CAC CTG CTG GAA CTT CTC CTC
 Y V E E F D A S L L T S F P L L K
 TAT GTT GAA GAG TTT GAT GCC AGC CTT CTG ACC TCT TTC CCT CTG AAG gtgagaccacc
 tcagagagggcaccacacactgcctctgacagtaacaatggtgggcccattggg-----0.80 Kb-----g
 A F K S R I S
 ggtgtgcacatgagtagtactaccagctctcaagctgtgtttctggattacag GCC TTC AAG AGC AGA ATC AGC
 S L P N V K K F L Q P G S Q R K P A M D
 AGC CTC CCC AAT GTG AAG AAG TTC CTG CAG CCT GGC AGT CAG AGA AAG CCA GCC ATG GAT
 Exon 7
 A K Q I E E A R K V F K F
 GCA AAA CAA ATC GAA GAA GCA AGG AAG GTT TTC AAG TTT TAG CGAGCTGCACCTGCCAATTTCTT
 GTAATCCAGGCTCTGATGTTTTGCAAAAATGAGAAGCAATGTTGATCTCGGCTATTTTGCATATAAATAAATAAGAAC
 AAATGGA

FIG. 3. DNA sequence analysis of the glutathione S-transferase Ya structural gene. The nucleotide sequences of all exons and portions of each intron were sequenced by the Maxam-Gilbert chemical sequencing procedure. Nucleotides in the exons are denoted by uppercase letters, whereas nucleotides in the introns are denoted by lowercase letters. The underlined sequence AATAAA represents the poly(A) signal.

formed. Consequently, intron 1 cannot be visualized. The other five introns are marked by numbers, with intron 2 being the largest. The approximate sizes of the six introns along with the exons are presented in Table 1. All splicing junctions satisfy the canonical GT/AG rule (24).

Determination of the Transcriptional Initiation Site. The start of initiation of transcription was determined by primer extension experiments. Since the Ya and Yc mRNAs share significant sequence homology throughout much of their sequences, we synthesized an oligonucleotide probe complementary to nucleotides 36-55 of the 5' untranslated region of the Ya mRNA. This region was chosen because the Ya and Yc mRNAs are divergent in this region. Poly(A) RNA from rat liver was hybridized to the oligonucleotide and the primer was extended using reverse transcriptase. The extended primer was electrophoresed alongside a sequencing gel of mp19 as a size marker (Fig. 5). The length of the extended primer was 55 nucleotides, which corresponds to the guanine marked as +1 in Fig. 3 of the 5' flanking region of the Ya gene. There is an (A+T)-rich region, TATTAATA, 32 bp from the

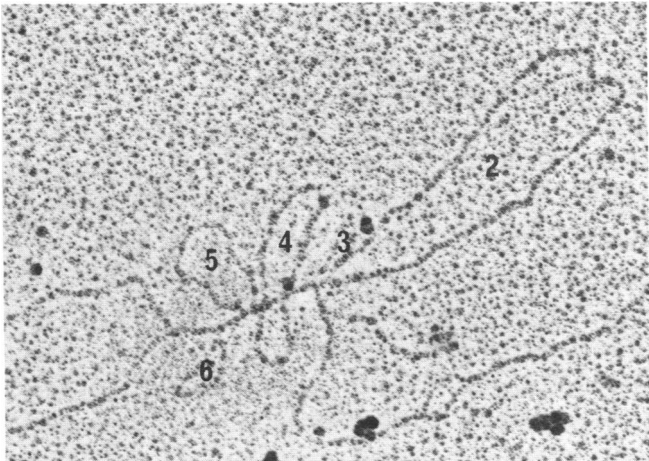


FIG. 4. Heteroduplex analysis of the glutathione S-transferase Ya genomic clone. Electron microscopy of the hybrid between lambdaGTB45-15 and linearized pGTB38. All introns are identified by numbers.

initiation codon. This (A+T)-rich sequence most likely represents the Goldberg-Hogness promoter, the "TATA" box (25).

DISCUSSION

The rat liver glutathione S-transferase Ya-Yc gene family is comprised of at least five to seven genes (16). We have isolated from an *EcoRI* and *Hae III* library four distinct genomic fragments that hybridize to our Ya cDNA clone, pGTB38 (ref. 16, and present study). Southern blots of rat genomic DNA indicate the presence of sequences not represented by the four genomic fragments. Therefore, not all glutathione S-transferase Ya-Yc structural genes have been isolated from the two libraries. The glutathione S-transferase structural gene isolated in this study corresponds to a Ya gene. The gene spans 11 kb and is separated into seven exons by six introns.

Although the precise number of Ya or Yc genes is unknown, evidence for the existence of multiple Ya genes does exist. Our laboratory (11) and Tu's laboratory (26) have constructed and characterized two different cDNA clones, pGTB38 and pGTR261, that are complementary to Ya mRNAs. These two cDNA clones differ by 15 nucleotides, which translate into eight amino acid differences. The 3' untranslated regions of the two cDNA clones are divergent, which support the notion that the two cDNAs are derived from separate genes. Since five different laboratories have isolated overlapping clones corresponding to either pGTB38 or pGTR261 (11, 26-30), the nucleotide differences that exist between the clones are not due to either cloning artifacts or DNA sequencing mistakes.

Table 1. Size of exons and introns in the glutathione S-transferase Ya gene

Exon	Sequence analysis	Intron	Restriction mapping	Electron microscopy*
1	43	1	2350	—
2	109	2	3500	3650
3	52	3	650	640 ± 150
4	133	4	2100	1889 ± 349
5	142	5	1500	1421 ± 337
6	132	6	800	642 ± 184
7	234			

Sizes are given in bp. *Double-stranded circular simian virus 40 DNA was included as an internal standard.

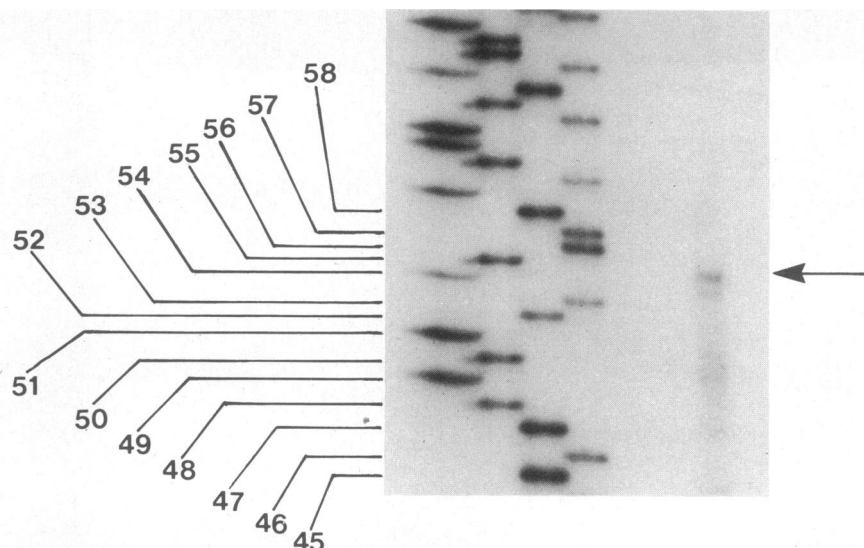


FIG. 5. Determination of the start of transcription. The start of transcriptional initiation was determined by primer extension experiments. The DNA sequence corresponds to M13mp19. The nucleotides of the sequence are indicated by number in order to provide a size standard. The arrow corresponds to the primer extended product.

Wang *et al.* (31) have prepared monoclonal antibodies to the glutathione *S*-transferase that recognize a Ya subunit of a YaYa dimer but not the Ya subunit of a YaYc heterodimer. These data indicate the existence of structurally distinct Ya subunits. Although the protein encoded by λ GTB45-15 is a Ya subunit, we do not know if it corresponds to the Ya subunit in the YaYc heterodimer or a Ya subunit in the YaYa dimer.

A comparison of the amino acid sequences encoded by the seven exons of the Ya gene with the corresponding regions in the Yc polypeptide has revealed some interesting findings. The amino acids encoded by exons 2 and 4 are highly conserved in the Ya and Yc subunits. These findings suggest these two exons encode domains of the subunits that have similar structural or functional properties. For example, both subunits have a glutathione binding site. In marked contrast to exons 2 and 4 are exons 3 and 5, which encode regions of the Ya subunit that are very divergent from the Yc subunit. Interestingly, one of the most distinctive differences between Ya homodimers and Yc homodimers is their substrate specificity towards Δ^5 -androstene-3,17 dione and cumene hydroperoxide. The Ya homodimer has high activity toward Δ^5 -androstene-3,17 dione, whereas the Yc homodimer has high activity toward cumene hydroperoxide (32). It is tempting to speculate that the amino acids encoded by exon 3 or 5 may form part of the substrate binding site.

We have demonstrated that various xenobiotics transcriptionally activate members of the Ya gene family leading to an accumulation of Ya mRNA (11, 15). The isolation and characterization of the glutathione *S*-transferase Ya gene will allow a detailed analysis of the promoter-regulatory region of the gene. Construction of the Ya gene promoter-regulatory region to suitable vectors and subsequent transfections into appropriate cell lines will allow the elucidation of cis-acting regulatory elements on the gene as well as trans-acting factors that regulate these elements. The exon-intron structure of the Ya gene will also provide a basis to compare the evolutionary relationship between the various glutathione *S*-transferase gene families.

We thank Mrs. Joan Kiliyanski for her assistance in the preparation of this manuscript.

1. Arias, I. M., Fleischner, G., Kirsch, R., Mishkin, S. & Gatmaitan, Z. (1976) in *Glutathione: Metabolism and Function*, eds. Arias, I. M. & Jakoby, W. (Raven, New York), pp. 175-188.

2. Litwack, G., Ketterer, B. & Arias, I. M. (1971) *Nature (London)* **234**, 466-467.
3. Jakoby, W. B. & Habig, W. H. (1980) in *Enzymatic Basis of Detoxification*, ed. Jakoby, W. B. (Academic, New York), Vol. 2, pp. 63-94.
4. Bass, N. M., Kirsch, R. E., Tuff, S. A., Marks, I. & Saunders, S. J. (1977) *Biochim. Biophys. Acta* **492**, 163-175.
5. Reddy, C. C., Li, N.-Q. & Tu, C.-P. D. (1984) *Biochem. Biophys. Res. Commun.* **121**, 1014-1020.
6. Hayes, J. D. (1984) *Biochem. J.* **224**, 839-852.
7. Satoh, K., Kitahara, A., Soma, Y., Inaba, Y., Hatayama, I. & Sato, K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3964-3968.
8. Beale, D., Ketterer, B., Carne, T., Meyer, D. & Taylor, J. B. (1982) *Eur. J. Biochem.* **126**, 459-463.
9. Boyer, T. D., Kenney, W. C. & Zakim, D. (1983) *Biochem. Pharmacol.* **32**, 1843-1850.
10. Ketterer, B., Beale, D., Taylor, J. B. & Meyer, D. J. (1983) *Biochem. Soc. Trans.* **11**, 466-467.
11. Pickett, C. B., Telakowski-Hopkins, C. A., Ding, G. J.-F., Argenbright, L. & Lu, A. Y. H. (1984) *J. Biol. Chem.* **259**, 5182-5188.
12. Telakowski-Hopkins, C. A., Rodkey, J. A., Bennett, C. D., Lu, A. Y. H. & Pickett, C. B. (1985) *J. Biol. Chem.* **260**, 5820-5825.
13. Ding, G. J.-F., Lu, A. Y. H. & Pickett, C. B. (1985) *J. Biol. Chem.* **260**, 13268-13271.
14. Ding, G. J.-F., Ding, V. D.-H., Rodkey, J. A., Bennett, C. D., Lu, A. Y. H. & Pickett, C. B. (1986) *J. Biol. Chem.* **261**, 7952-7957.
15. Ding, V. D.-H. & Pickett, C. B. (1985) *Arch. Biochem. Biophys.* **240**, 553-559.
16. Rothkopf, G. S., Telakowski-Hopkins, C. A., Stotish, R. L. & Pickett, C. B. (1986) *Biochemistry* **25**, 993-1002.
17. Woo, S. L. C. (1979) *Methods Enzymol.* **68**, 389-395.
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
19. Peacock, S. L., McIver, C. M. & Monahan, J. J. (1981) *Biochim. Biophys. Acta* **655**, 243-250.
20. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387-2398.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
22. Gubler, U. & Hoffman, B. S. (1983) *Gene* **25**, 263-269.
23. Yamada, Y., Mudry, M., Sullivan, M. & de Crombrughe, B. (1983) *J. Biol. Chem.* **258**, 2758-2761.
24. Sharp, P. A. (1981) *Cell* **23**, 643-646.
25. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349-383.
26. Tu, C.-P. D., Lai, H.-C. J., Li, N.-Q., Weiss, M. J. & Reddy,

- C. C. (1984) *J. Biol. Chem.* **259**, 9434–9439.
27. Kalinyak, J. E. & Taylor, J. M. (1982) *J. Biol. Chem.* **257**, 523–530.
28. Tu, C.-P. D., Weiss, M. J., Karakawa, W. K. & Reddy, C. C. (1982) *Nucleic Acids Res.* **10**, 5407–5419.
29. Daniel, V., Sarid, S., Bar-Nun, S. & Litwack, G. (1983) *Arch. Biochem. Biophys.* **227**, 266–271.
30. Taylor, J. B., Craig, R. K., Beale, D. & Ketterer, B. (1984) *Biochem. J.* **219**, 223–231.
31. Wang, I. Y., Tung, E., Wang, A.-C., Argenbright, L., Wang, R., Pickett, C. B. & Lu, A. Y. H. (1986) *Arch. Biochem. Biophys.* **245**, 543–547.
32. Mannervik, B. (1985) *Adv. Enzymol. Relat. Areas Mol. Biol.* **57**, 357–417.