Archives of Biochemistry and Biophysics
Vol. 361, No. 1, January 1, pp. 85–93, 1999
Article ID abbi.1998.0964, available online at http://www.idealibrary.com on IDEAL®

# Distinctive Structure of the Human *GSTM3* Gene—Inverted Orientation Relative to the Mu Class Glutathione Transferase Gene Cluster

Yury V. Patskovsky,*[,1] Ming-Qian Huang,†[,1] Tetsuji Takayama,‡ Irving Listowsky,* and William R. Pearson†[,2]

*Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461; †Department of Biochemistry, University of Virginia, Charlottesville, Virginia 22908; and ‡Department of Internal Medicine, Sapporo Medical College, Sapporo, Japan*

Received June 22, 1998, and in revised form September 14, 1998

**The sequence and exon–intron structure of the human class mu *GSTM3* glutathione transferase gene and its orientation with respect to the remainder of the human class mu *GSTM* gene cluster were determined. The *GSTM3* gene is 2847 bp long and is thus considerably shorter than the other class mu genes in the cluster, which range in size from 5325 to 7212 bp. Outside the protein-coding region, the *GSTM3* gene does not share significant sequence similarity with other class mu glutathione transferase genes. Identification of overlapping cosmid clones that span the region between *GSTM5*, the next nearest glutathione transferase gene, and *GSTM3* showed that the two genes are about 20,000 bp apart. PCR primers developed from sequences 3′-downstream from the *GSTM5* gene were used to identify clones containing the *GSTM3* gene. Amplification with these primers showed that the orientation of the *GSTM3* gene is 5′-GSTM5-3′–3′-GSTM3-5′. Long-range PCR reactions confirmed this orientation both in the GSTM-YAC2 YAC clone, which contains the five class mu glutathione transferase genes on chromosome 1, and in human DNA. This tail-to-tail orientation is consistent with an evolutionary model of class mu glutathione transferase divergence from a pair of tail-to-tail "M1-like" and "M3-like" class mu glutathione transferase genes that was present at the mammalian radiation to the current organization of multiple head-to-tail M1-like genes tail-to-tail with a single M3-like gene with distinct structural properties and expression patterns.**



***Key Words:*** **GSTM3; class mu glutathione transferase; gene duplication; evolution.**

The glutathione transferases (GST, EC 2.5.1.18)[3] are a large family of phase II detoxication enzymes that catalyze conjugation reactions between glutathione and a wide variety of electrophilic compounds, including chemical carcinogens produced by the cytochrome P450 mixed-function oxygenases (1, 2). Mammalian soluble glutathione transferases have been assigned in a species-independent manner to alpha, mu, pi, and theta classes (3); in addition, two novel classes have recently been identified (4). Five human class mu genes—*GSTM1*, *GSTM2*, *GSTM3*, *GSTM4*, and *GSTM5*—were mapped to band 1p13.3 on chromosome 1 by *in situ* hybridization (5) and FISH with GSTM-YAC2, a YAC clone that contains these five class mu genes (6). More recently, a physical map for four of the five class mu genes—*GSTM1*, *GSTM2*, *GSTM4*, and *GSTM5*—showed that they are arrayed in a head-to-tail orientation in the order 5′-M4–M2–M1–M5-3′ (7). Taken together with earlier analysis of the GSTM-YAC2 YAC clone (6, 8), the *GSTM3* gene is predicted to be downstream from the 3′-end of the *GSTM5* gene, but its exact location and orientation are not known.

The human class mu glutathione transferase gene cluster is of particular interest because about 50% of the population is homozygous deleted for *GSTM1*. The

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed at Department of Biochemistry, Box 440 Jordan Hall, University of Virginia, Charlottesville, VA 22908. Fax: (804) 924-5069. E-mail: wrp@virginia.edu.

[3] Abbreviation used: GST, glutathione transferase.

0003-9861/99 $30.00

*GSTM1* deletion has been associated with increased risk of lung, bladder, and skin cancer (9), but in some cases the association is weak (10). The *GSTM1* deletion is the result of a homologous unequal crossing-over between two 2.3-kb repeated regions in the intergenic regions that flank the *GSTM1* gene (7). Sequences downstream of the deletion are undisturbed from about 10 kb upstream of the 5′-end of the *GSTM5* gene; thus, the *GSTM3* gene, which is even further downstream, should not be directly affected by the deletion. It was therefore unexpected that GSTM3 protein levels were reported to be reduced in lungs of individuals that are *GSTM1*-null (11), although the effect was not seen in brain (12). However, studies of GSTs from multiple human tissue specimens showed no correlation between *GSTM3* expression and the *GSTM1*-null genotype (13).

The protein encoded by the *GSTM3* gene is unusual. It is clearly a member of the mu class (it shares about 70% amino acid sequence identity with other class mu proteins and less than 35% sequence identity with the alpha, pi, or theta classes), but it is the most distantly related member of that family in mammals. hGSTM1, hGSTM2, hGSTM4, and hGSTM5 share more than 84% amino acid sequence identity with one another and more than 80% sequence identity with most rodent and rabbit class mu enzymes. In contrast, hGSTM3 shares only about 70% identity with most class mu protein sequences. "M3-like" sequences have been identified in rodents (GTM5_MOUSE,[4] 89% identical to GTM3_HUMAN, Ref. 14; and rGSTM5, Ref. 15). The proteins in the M3-like subclass exhibit atypical catalytic and structural properties (15). Moreover, these types of glutathione transferases are selectively expressed in both rodent and human germ cells (J. D. Rowe *et al.*, unpublished) suggesting that they may be a functionally distinct subfamily of class mu glutathione transferases. Thus, in mammals, there appears to be a broad division of class mu glutathione transferases which may be categorized as "M1-like" (human, hGSTM1, hGSTM2, hGSTM4, and hGSTM5; mouse, GTM1_MOUSE, GTM2_MOUSE, and GTM3_MOUSE) and M3-like (human, hGSTM3; mouse, GTM5_MOUSE; and rat, rGSTM5). Comparison of the hGSTM1 and hGSTM3 protein sequences with those in other vertebrates suggests that this division occurred after the mammalian line diverged from birds; the chicken class mu glutathione transferase (GTM2_CHICK) is approximately equally distant from the M1-like (ranging from 64 to 67% identity) and the M3-like (64 to 66% identity) sequences. An evolutionary tree summarizing these relationships is shown in Fig. 1.

*GSTM3* contains a polymorphic three-base deletion in intron 6 that has been associated with the *GSTM1*a allele (16). This deletion removes a potential binding site for the YY1 transcription factor and it has been suggested that it may modify the risks associated with the *GSTM1*-0 deletion and a cytochrome P450 CYP1A1 m1/m1 polymorphism in cases of multiple basal cell carcinomas (17). Thus, it is important to determine the physical linkage between *GSTM3* and the other class mu glutathione transferase genes on chromosome 1p13.3. In this report, we show that the *GSTM3* gene is considerably shorter than the other M1-like class mu genes, and that it is transcribed in the reverse orientation to *GSTM1, GSTM2, GSTM4,* and *GSTM5.*

## MATERIALS AND METHODS

*Reagents.* Reagent grade chemicals were obtained from Sigma and Fisher, restriction enzymes from Gibco Life Sciences or New England Biolabs (Beverly, MA), and DNA size markers and Klenow fragment from Gibco Life Sciences. DNA *Taq* polymerase was obtained from Promega (Madison, WI). GeneAmp XL PCR kits from Perkin–Elmer were used for long-range polymerase chain reactions (PCR). PCR primers and other oligonucleotides were synthesized by Operon (Alameda, CA) or synthesized using an Applied Biosystems Model 380-A DNA synthesizer from Perkin–Elmer Cetus (Emeryville, CA). [$\alpha$-$^{32}$P]dCTP and [$\gamma$-$^{32}$P]ATP were obtained from ICN Pharmaceuticals (Costa Mesa, CA). Zetabind blotting nylon membrane was obtained from Cuno Inc. (Meriden, CT). The materials used in the construction of lambda FIX II genomic library and cosmid library, as well as T3 and T7 primers, were purchased from Stratagene (La Jolla, CA). Nitrocellulose membranes for colony screening were obtained from Schleicher & Schuell (Keene, NH). The QIAGEN Plasmid Mega kit and the lambda DNA purification kit were purchased from QIAGEN Inc. (Chatsworth, CA).

*Nucleic acids methods.* Southern blotting was done as described (18) with slight modifications. Restriction enzyme digestions were performed as suggested by the manufacturers. Digested DNA was phenol extracted, ethanol precipitated, and electrophoresed in on a 0.8% agarose gel and transferred to nylon membranes as described (19). The hybridization DNA probes were prepared by random-priming labeling (20); hybridizations were performed according to Church and Gilbert (21) at 65°C for 24 h followed by washing at 65°C.

*Human genomic library screening.* Procedures were performed as described (18). A human leukocyte lambda FIX II genomic library was propagated on *Escherichia coli* strain LE392 (Stratagene), and infected bacterial cells were plated in 0.7% soft agarose at a density of about $3–5 \times 10^4$ plaques per 15-cm petri dish. Plaques were lifted onto nitrocellulose filters and initially screened with [$\alpha$-$^{32}$P]dCTP-radiolabeled HTGT-6, a *GSTM3*-cDNA probe (22). Eight positive plaques were isolated from approximately $10^6$ phage plaques (20 dishes) and were purified by two additional rounds of screening. Two [$\alpha$-$^{32}$P]dCTP-radiolabeled oligonucleotides were used as hybridization probes to test for the presence of exons 1–8 of the entire *GSTM3* gene sequence as follows: oligo I, 5′-CGTCTATGGTTCTCGGGTACTGG-3′ (from *GSTM3* exon 1), and oligo II, 5′-AACCGGAGTTCGGAAACCGTCCTGTGTA-3′ (exon 8). Two phage clones positively cross-hybridized with oligo I and did not hybridize with oligo II; and one clone cross-hybridized with oligo II, but did not hybridize with oligo I.

*Phage DNA sequencing.* DNA was purified from phage plate lysates as described (18) and purified using a QIAGEN Lambda DNA purification kit. After restriction mapping, inserts from two recombinant phage that contained overlapping parts of the *GSTM3* gene were sequenced. DNA sequences were determined for both strands of the inserts with oligonucleotide primers generated on the basis of the *GSTM3* cDNA sequence. An ABI377 Applied Biosystems automated

[4] Protein sequence names are taken from SwissProt.

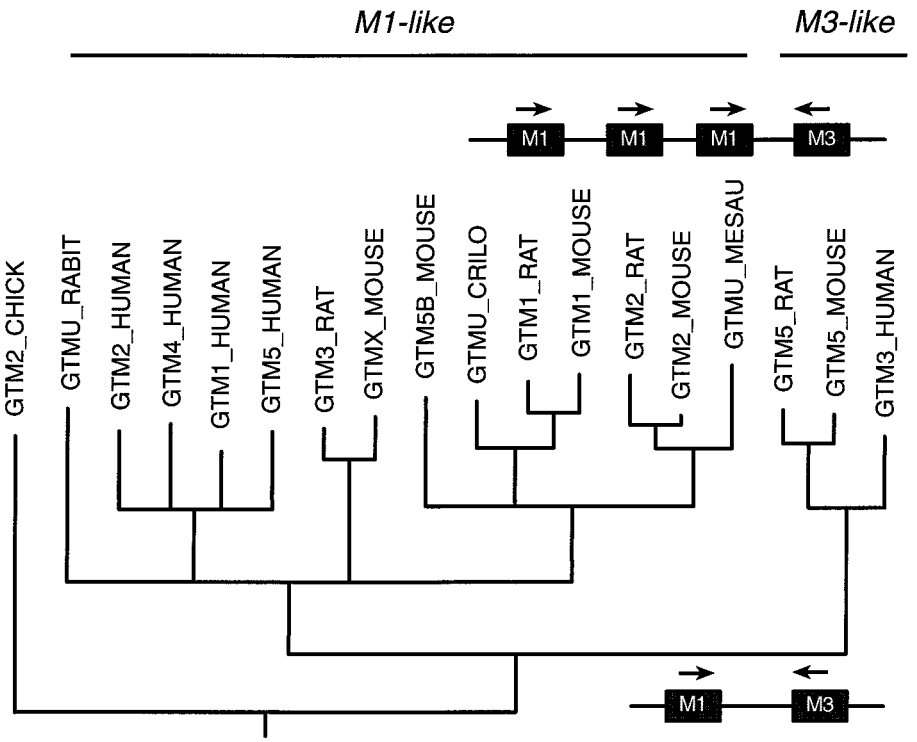


**FIG. 1.** Evolutionary relationships between vertebrate class mu glutathione transferases. Protein sequence data were taken from SwissProt (release 34), and SwissProt names are used, with the exception of GTMX_MOUSE, which was assembled from EST sequences, and GTM5B_MOUSE, which was described in Ref. 36. The GTM5_RAT sequence was obtained from Ref. 15. Protein sequences were multiply aligned with ClustalW (24), and the aligned amino acids were used to direct an alignment of the encoding mRNA sequences. The tree shown was constructed using the PHYLIP dnaml program (25). The same tree topology is found using aligned cDNA sequences and the dnapars program and protein distances with the fitch program. Nodes were combined if they appeared in <80% of trees found in 50 different orderings of the dnaml program. Proposed structures for the ancestral tail-to-tail "M1-like:M3-like" gene pair and the current organization are also shown.

sequencer (Perkin–Elmer) was used. After partial sequencing and restriction mapping, restriction fragments containing portions of the *GSTM3* cDNA sequence were subcloned into pUC18 and sequenced using common pUC oligonucleotide primers and primers generated from determined gene sequences. Oligonucleotides used for this procedure were provided by Perkin–Elmer.

*Identification and characterization of GSTM-YAC2 cosmids.* A cosmid library from GSTM-YAC2 (7) was screened as described (7, 18) using probes from exon 8 of *GSTM5* and cosmid end sequences or from exons 5–8 (primers P9 and P10, Table I) of *GSTM3*. Restriction mapping by partial digestion and indirect end-mapping with T3 and T7 oligonucleotides was performed as described (7). Sequencing from

**TABLE I**

PCR Primers Used for Cosmid Identification and Orientation

| | PCR primers: (all 5′ → 3′) | | Location |
|---|---|---|---|
| P1 | m5-2t3a | ACTGTAATGTGTCCTTGACC | cGTM5-2, T3 end, plus |
| P2 | m5-2t3b | GAAAAAGGAGAGGAGAATTG | cGTM5-2, T3 end, minus |
| P3 | m5-699 | CCGAGGTCTTTTGTTTGGAA | *GSTM5*, exon 8, forward |
| P4 | m5-1143r | AAGGGAGCCTCAGGGAATAA | *GSTM5*, exon 8, reverse |
| P5 | m5-1500r | GAGAAGTGTCATGACGAA | *GSTM5*, exon 8, reverse |
| P6 | m5-5t7b | TAGAAAAGCAAGTCTTAGCC | cGTM5-5, T7 end, minus |
| P7 | m5-4t7a | TTGCCAGCCAGACCAAATTG | cGTM5-4, T7 end, plus |
| P8 | m5-4t7b | CCAAGTATCAAACCAGCAAGGG | cGTM5-4, T7 end, minus |
| P9 | m3-338 | GCACACAACTGATAAGGC | *GSTM3*, exon 5, forward |
| P10 | m3-607r | TTGCAGAACTGATCAGAC | *GSTM3*, exon 8, reverse |
| P11 | m3-1200 | CTGGGTGAGGATAACACAAG | *GSTM3*, exon 8, forward |
| | T3 | AATTAACCCTCACTAAAGGG | Cosmid T3 primer |
| | T7 | GTAATACGACTCACTATAGGGC | Cosmid T7 primer |

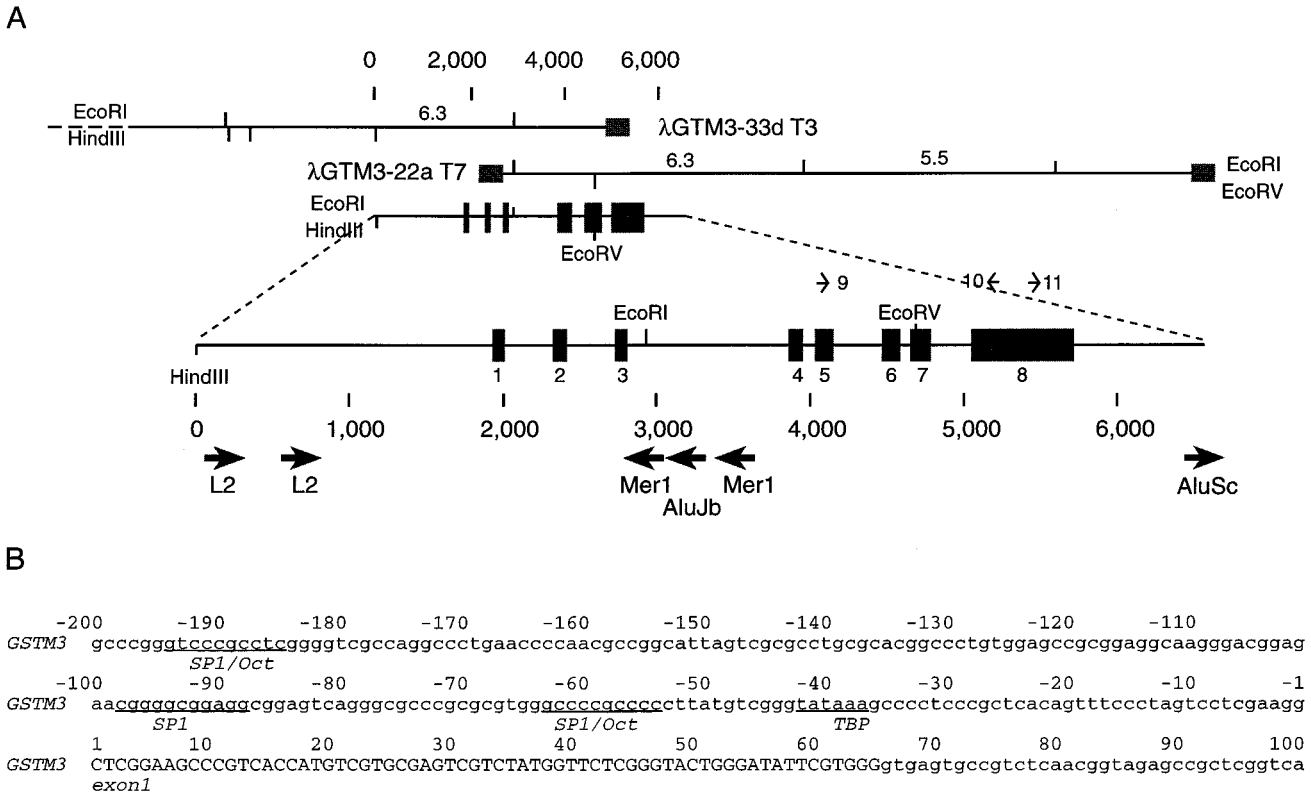


**FIG. 2.** Physical map of the *GSTM3* gene and flanking sequences. (A) Partial restriction maps of the λGTM3-22a and λGTM3-33d phage inserts used to determine the sequence of the *GSTM3* gene. Also shown are the exon locations in *GSTM3*, the location and orientation of the M3-specific PCR primers m3-338 (P9), m3-607r (P10), and m3-1200 (P11), and the location and orientation of six repeated sequence elements. (B) The promoter proximal sequence and first 100 bp of GSTM3. A TATAAA box and potential Sp-1 and octamer binding factor binding sites are underlined.

the ends of the cosmid clone inserts was performed at University of Virginia Sequencing Center.

*PCR and PCR cloning.* PCR reactions were performed in a Perkin–Elmer Cetus DNA Thermal Cycler 480. In addition to the template DNA, the reaction contained 100 $\mu$l of 50 mM KCl, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 1.5 $\mu$M $MgCl_2$, 200 $\mu$M of dNTP, 1 $\mu$M of each primer, and 2.5 units of *Taq* polymerase. All PCR reactions used a "hot start" procedure with Ampliwax from Perkin–Elmer. Samples were amplified for 35 cycles with denaturation for 1 min at 94°C, reannealing for 2 min at 55°C, and chain elongation for 3 min at 72°C. Long-range PCR was done following the instructions for the GeneAmp XL PCR kit, with one cycle of 1 min at 94°C, 16 cycles including 30 s at 94°C and 13 min at 64°C per cycle, and then 12 cycles, each of which has 30 s at 94°C and 13 min at 64°C with an increment of 15 s per cycle. For all PCR reactions, the final cycle was followed by incubation at 72°C for 10 min. PCR primers used are shown in Table I.

*Sequence analysis.* General sequence analysis employed the GCG suite of programs (version 9.0; Genetics Computer Group, Madison, WI). Protein sequences were extracted from SwissProt, release 34, and updates (23). Multiple sequence alignments were performed using the ClustalW program (24). Phylogenetic analyses were performed using programs from the PHYLIP package (25). Sequence similarity searches were performed using FASTA (26). Searches for repeated sequence elements were performed at the `repeatmasker` Worldwide Web site (27). Transcription factor binding sites were identified using the TESS server `http://agave.humgen.upenn.edu/tess/index.html` (28).

Phylogenetic analysis was performed using the PHYLIP v. 3.52 package (25). Class mu protein sequences were aligned using the ClustalW (24) program; the aligned class mu sequences contained no internal gaps. Evolutionary trees were constructed from protein distances calculated with the `protdist` program or using DNA parsimony (`dnapars`) or maximum likelihood (`dnaml`).

## RESULTS

### *Identification and Characterization of the Human GSTM3 Gene*

The complete sequence of *GSTM3* was determined from two overlapping genomic lambda clones. (The sequence has been deposited in the GenBank database under Accession No. AF043105.) These clones were also used to physically map approximately 30 kb of chromosomal DNA surrounding *GSTM3* (Fig. 2). The two phage clones overlap in a 2.1-kb region that contains exons 3–7 of *GSTM3*. The overlapping nucleotide sequences were determined for both genomic DNA clones; no differences were found between them. No other M3 homologous sequences were detected by Southern blot hybridization in the 25-kb chromosomal DNA fragment that is located around the M3 gene.

**TABLE II**

Exon/Intron Structure of Human Class Mu GST Genes

| | Length (bp) | | *GSTM3* | |
|---|---|---|---|---|
| | *GSTM3* | M1,M2,M4,M5 | 5′ (start) | 3′ (end) |
| Exon 1 | 64 | 51 | TCGGA | GTGGG |
| Intron 1 | 332 | 261–287 | gtgag | cccag |
| Exon 2 | 76 | 76 | CTGGC | GGAAG |
| Intron 2 | 339 | 426–429 | gtaat | tacag |
| Exon 3 | 65 | 65 | CTCCT | CTAAT |
| Intron 3 | 1062 | 302,312 | gtaag | ttcag |
| Exon 4 | 82 | 82 | CTGCC | CATGT |
| Intron 4 | 93 | 96,101 | gtgag | tacag |
| Exon 5 | 101 | 101 | GTGGT | ACCAC |
| Intron 5 | 339 | 942–1742 | gtgag | ctcag |
| Exon 6 | 96 | 96 | GAAAA | AAAAG |
| Intron 6 | 88 | 88–92 | gtagg | tctag |
| Exon 7 | 111 | 111 | CTCAC | TTGAG |
| Intron 7 | 285 | 2055–3164 | gtgac | ttcag |
| Exon 8 | 672 | 535–624 | GCTTT | TGATG |

The exon–intron junctions of the *GSTM3* gene were determined by alignment of the *GSTM3* genomic DNA sequence with the previously published *GSTM3* cDNA sequences (22, Table II). Like the other class mu glutathione transferase genes, *GSTM3* has eight exons and seven introns. All the internal exons (2–7) are the same length as those of the other human class mu glutathione transferase genes. Exons 1 and 8 are slightly longer, as expected from the extended N- and C-termini (four and three additional amino acid residues, respectively) of the hGSTM3 subunit. There are two striking differences in intron length between *GSTM3* and the other class mu glutathione transferase genes (*GSTM1*, *GSTM2*, *GSTM4*, and *GSTM5*) on chromosome 1. In the other human class mu genes, intron 3 is either 302 or 312 bp; in *GSTM3* it is much longer (1062 bp, Table II). All of the additional sequence in intron 3 of *GSTM3* can be accounted for by the presence of three repeated sequence elements, two MER1-type elements (183 and 299 bp) and an AluJb element (301 bp, Ref. 29) (27). More striking are the length differences in intron 7; the other class mu genes have very long intron 7s [ranging in length from 2055 bp (*GSTM4*) to 3164 bp (*GSTM2*)] that are relatively repeat-free; intron 7 in *GSTM3* is only 285 bp long.

While the *GSTM1*, *GSTM2*, *GSTM4*, and *GSTM5* glutathione transferase genes share extensive DNA sequence identity both in protein coding and noncoding regions (30, manuscript in preparation), the *GSTM3* gene shares statistically significant sequence similarity with the other human class mu genes only in the protein coding sequence [ranging from 68% (*GSTM2*) to 72% (*GSTM4*) protein sequence identity; 71% (*GSTM2*) to 77% (*GSTM5*) cDNA sequence identity]. No statistically significant similarities were observed in comparisons of noncoding M3 cDNA sequence (exon 8) with other class mu cDNAs.

The *GSTM3* genomic sequence contains a typical complement of human repeated sequence elements, which are indicated in Fig. 2. The `repeatmasker` program (27) identified two MER1-type elements and an AluJb (intron 3), portions of two L2 LINEs, an L1 LINE, and an AluSc SINE (Fig. 2A).

DNA sequences in the 5′-flanking region of *GSTM3* share no significant similarity with any of the other class mu genes on chromosome 1. The promoter region of M3 (Fig. 2B) is the only human mu class glutathione transferase gene that contains a TATA box (at the distance about −41 bp from the transcription start). The *GSTM3* promoter also contains GC boxes (SP1 elements); these elements may be responsible for the

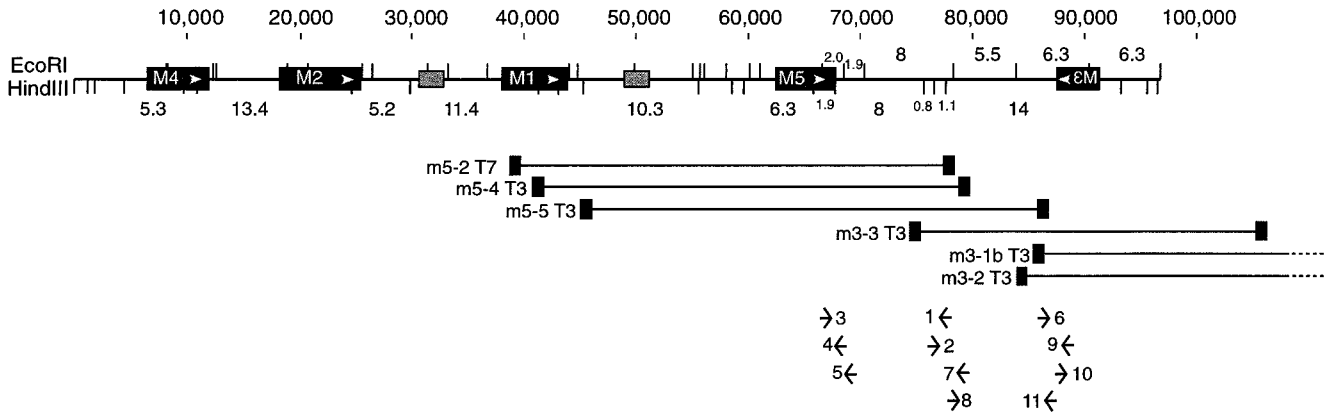


**FIG. 3.** Physical map of the human class mu gene cluster. Cosmids and primers used for linking *GSTM5* and *GSTM3* are shown. Restriction fragment sites and gene locations and orientations for *GSTM1*, *GSTM2*, *GSTM4*, and *GSTM5* are taken from Ref. 7; restriction site locations to the right of *GSTM5* were determined in this report. Locations and orientations of the cGTM5-2, cGTM5-4, cGTM5-5, cGTM3-1b, cGTM3-2, and cGTM3-3 cosmid inserts are shown, and PCR primers used to determine the orientation of *GSTM3* with respect to *GSTM5* are also shown. PCR primer numbers are shown in Table I.

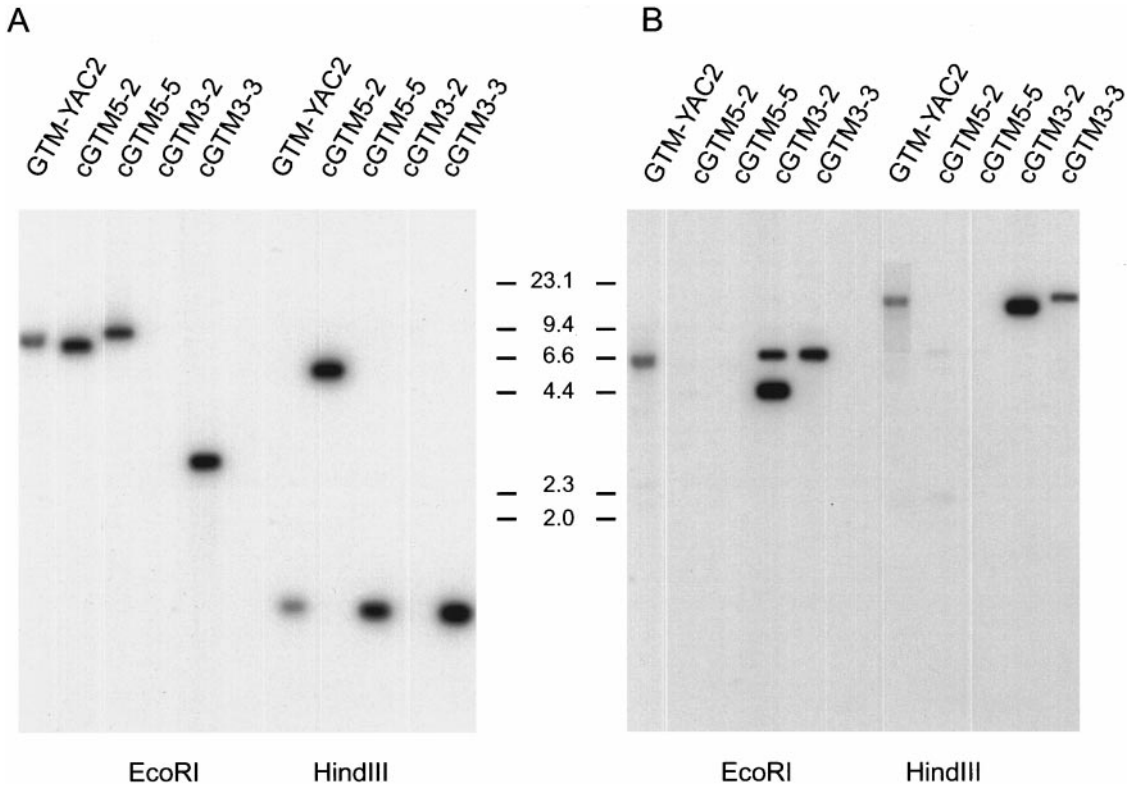


**FIG. 4.** Characterization of cosmid clones between GSTM5 and GSTM3. Southern blot hybridization of *Eco*RI and *Hin*dIII digests of GSTM-YAC2 DNA and DNAs from cGTM5-2, cGTM5-5, cGTM3-2, and cGTM3-3. (A) Hybridization with a 400-bp PCR probe cM5-2T3 from the end of cGTM5-2 (Fig. 3). (B) Hybridization with the 1.3-kb *GSTM3* cDNA probe (22). Positions of DNA size standards are shown.

low, but detectable level of GST expression in most human tissues (13).

### *Physical Linkage of GSTM3 and GSTM5*

Because of reports linking the polymorphic *GSTM1* deletion with differences in *GSTM3* expression (11, 12), we sought to determine the distance between the *GSTM3* gene and the *GSTM5* gene, which is the nearest class mu glutathione transferase gene in the cluster on chromosome 1 (6, 7). To identify clones that spanned the region between *GSTM5* and *GSTM3*, a cosmid library constructed from GSTM-YAC2, a YAC clone from chromosome 1 that contains all five class mu glutathione transferase genes on chromosome 1 (6, 7), was screened. Cosmid clones containing exon 8 of *GSTM5* were first identified using a probe from PCR primers P3 and P4 (Table I, Fig. 3). PCR primers P1 and P2 were constructed from the end sequence of one of the *GSTM5* exon 8-containing cosmid clones, cGTM5-2; P1 and P2 were then used to amplify probe cM5-2T3, which was used to rescan the GSTM-YAC2 cosmid library. One of the clones identified in this screen, cGTM3-3, also contains sequences from the *GSTM3* gene (it hybridized with a probe from primers P9 and P10). Cosmids cGTM3-1b and cGTM3-2, which contain *GSTM3* (Fig. 3), were isolated using a *GSTM3* probe (exons 5–8). Restriction maps of cosmids from this region were determined by partial restriction digestion and indirect end-labeling (Fig. 3, partial restriction mapping data are not shown).

### *GSTM3 Is Oriented Opposite to M1, M2, M4, and M5*

One end of the cGTM3-1b clone contains sequences from exon 8 of *GSTM3*, and one end of cGTM3-2 lies within 2 kb of *GSTM3* exon 8. Since these cosmids have ~45-kb inserts, it was surprising that these cosmids failed to cross-hybridize with a probe from the *GSTM5* 3′-flanking sequences in cGTM5-2 (cM5-2T3, Fig. 3). While this result might imply that *GSTM3* is far downstream from *GSTM5*, isolation of the cGTM3-3 cosmid, which contains sequences both from cGTM5-2 and *GSTM3*, suggested otherwise. We then tested the hypothesis that cGTM3-1b and cGTM3-2 might not contain cGTM5-2 sequences because the *GSTM3* gene was oriented with its 3′-end (exon 8) nearest *GSTM5* (Fig. 3).

We confirmed that the physical map was correct and that the cGTM5 and cGTM3 cosmids were not rear-

**TABLE III**

Summary of PCR Amplifications between *GSTM5* and *GSTM3*

| | | | GSTM cosmids | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Primer1 | Primer2 | Size (kb) | M5-2 | M5-4 | M5-5 | M3-2 | M3-3 | YAC2 | Human |
| (P2) m5-2t3b | (P7) m5-4t7a | 1.8 | | + | | | + | + | + |
| | (P1) m5-2t3a | 0.4 | + | + | + | | + | + | |
| | (P10) m3-607r | | | | | | − | − | |
| (P3) m5-699 | (P5) m5-1500r | 0.85 | + | + | + | | | | |
| | (P1) m5-2t3a | 10.0 | + | + | + | | | + | |
| | (P2) m5-2t3b | | − | | | | | − | |
| | (P7) m5-4t7a | 11.3 | | | | | | + | + |
| (P11) m3-1200 | (P6) m5-5t7b | 1.7 | | | | + | + | + | + |
| | (P8) m5-4t7b | 8.7 | | | | | + | + | + |
| (P9) m3-338 | (P6) m5-5t7b | 3.0 | | | | + | + | + | |
| | (P10) m3-607r | 1.0 | | | | + | + | + | |
| T3 | (P1) m5-2t3a | 1.9 | | | | | + | | |
| | (P2) m5-2t3b | | | − | − | | | | |
| | (P10) m3-607r | | | | | | − | | |
| | (P11) m3-1200 | 12.0 | | | | | + | | |
| | | 1.9 | | | | + | | | |
| T7 | (P2) m5-2t3b | 2.1 | | + | | | | | |
| | (P6) m5-5t7b | 0.2 | | | + | | | | |
| | (P3) m5-699 | 11.5 | | + | | | | | |
| | (P9) m3-338 | | | | | | − | | |
| | (P11) m3-1200 | | | | | | − | | |

*Note.* Results of PCR amplifications using the indicated primers (Table I) on DNA samples in each column. M5-2, M5-4, etc. refer to the cosmid clones cGTM5-2, cGTM5-4, etc. YAC2 refers to GSTM-YAC2 DNA. Human DNA was from a *GSTM1*-null individual. +, A band of the indicated size was observed. −, A PCR reaction was performed and no band was seen. Blank entries indicate that the primers were not used on the DNA sample.

ranged by hybridizing *Eco*RI and *Hin*dIII restriction digests of GSTM-YAC2 and selected cosmids with a PCR product amplified from P1 and P2 (probe cM5-2T3, Fig. 4). As expected (Fig. 3), the cM5-2T3 probe hybridizes with cGTM5 cosmids and cGTM3-3; cM5-2T3 does not hybridize with cGTM3-1b and cGTM3-2, which contain complete *GSTM3* genes but very little 3′-flanking sequence. (The 8.0-kb *Eco*RI band in cGTM3-3 is truncated to 3 kb by the cloning vector; the 8.0-kb *Eco*RI band in cGTM5-2 is slightly truncated by the vector. Likewise, the 5.5-kb band in the cGTM5-2 *Hin*dIII lane of Fig. 4A is 4.5 kb larger than the corresponding genomic band because of the cloning vector.) Figure 4B shows that a M3 cDNA probe hybridizes only with the cGTM3 cosmids and that the 6.3-kb *Eco*RI fragment in cGTM3-3 is actually a doublet, which is separated in cGTM3-2 because of a truncation.

The relative orientations of *GSTM5* and *GSTM3* were confirmed by performing PCR reactions between primers in *GSTM3* and in the 3′-flanking region of cGTM5-2, cGTM5-4, and cGTM5-5. Table III summarizes the results of PCR reactions performed with various primer pairs on the cGTM5 and cGTM3 cosmids. These data, together with the physical map of the region (Fig. 3), show that the 3′-end of the *GSTM5* gene is about 20 kb upstream from the 3′-end of the *GSTM3* gene. We were unable to span this distance using long-range PCR in a single reaction, but shorter reactions from the 3′-end of *GSTM5* (primer P3, exon 8, forward) to the end of the cGTM5-4 insert (P7) and from the end of cGTM5-4 (P8) to the 3′-end of *GSTM3* (P11, exon 8, forward with respect to *GSTM3*) produced the expected bands (Fig. 5). The data in Fig. 3 predict that the *GSTM5* (P3)–cGTM5-4 (P7) reaction will produce an 11.3-kb band, which *Eco*RI will digest into 7.9-, 1.9-, 1.0-, and 0.4-kb bands (the 0.4-kb band is too light to be seen in Fig. 5) and *Hin*dIII will cut into 8.0-, 1.1- (2), 0.8-, and 0.2-kb fragments, while the cGTM5-4 (P8)–*GSTM3* (P11) reaction produces an 8.7-kb band which does not contain a *Hin*dIII site and is cut into 5.0- and 3.7-kb fragments by *Eco*RI. These predicted fragments were obtained (Fig. 5), confirming the tail to tail organization of the *GSTM5* and *GSTM3* genes, not only in our cosmid clones but also in GSTM-YAC2 and in human genomic DNA.

## DISCUSSION

We have cloned and sequenced the *GSTM3* gene, and, using overlapping cosmid clones, shown that *GSTM3* is oriented tail-to-tail with respect to the ad-

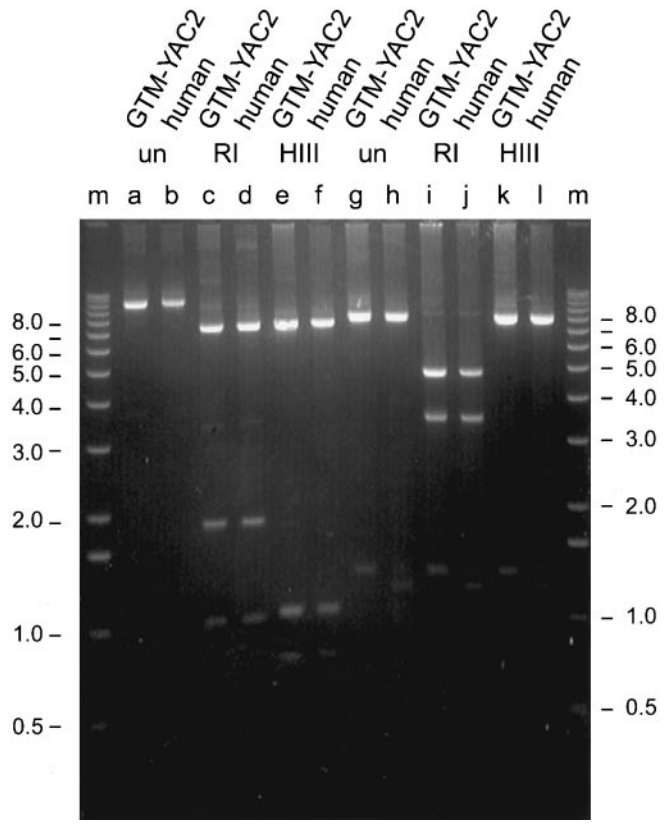


**FIG. 5.** Long-range PCR of sequences between *GSTM5* and *GSTM3*. Long-range PCR (see Materials and Methods) was used to demonstrate the orientation of the *GSTM5* and *GSTM3* genes in GSTM-YAC2 and human DNA samples using sequences from the end of cGTM5-4 as an intermediate. PCR was performed using the m5-699 (P3)/m5-4t7a (P7) primer pair (lanes a–f) or the m3-1200 (P11)/m5-4t7b (P8) primer pair (lanes g–l). The sizes of the undigested product (un) and *Eco*RI-digested (RI) and *Hin*dIII-digested (HIII) products are shown.

jacent *GSTM5* gene. This orientation has been confirmed by overlapping long-range PCR on both GSTM-YAC2 and human genomic DNA.

It has been recognized for some time that the *GSTM3* gene is more distantly related to other human and rodent class mu glutathione transferase genes (Fig. 1, Refs. 8 and 22), and it has been suggested that the rodent and human M1-like genes diverged after the mammalian radiation (8). Both the size of the *GSTM3* gene and its location and orientation with respect to the M1-like genes on chromosome 1 (Figs. 2 and 3; Table II) are consistent with this perspective. The schematic gene diagrams in Fig. 1 suggest the transition from the pair of tail-to-tail M1-like and M3-like genes to the current gene structure. Because of the limited amount of sequence data from nonmammalian vertebrates (only one class mu sequence, which encodes GTM2_CHICK, is available in the vertebrate portion of Genbank) it is currently impossible to determine when the M1-like, M3-like duplication occurred that produced the two genes in a tail-to-tail orientation. The presence of both types of genes in rodents and humans suggests that this early class mu duplication occurred either before or early in the mammalian radiation. Over the past 80 million years since then, multiple paralogous M1-like genes appeared in a head-to-head orientation, as shown in Fig. 1. In contrast, there is no evidence for any subsequent *GSTM3* gene duplications. DNA parsimony analysis suggests that the *GSTM3* gene may be older than the M1-like genes. However, protein parsimony, protein distance, and the presence of an additional four residues at the N-terminus of the M3 genes, place the M1-like and M3 genes equally distant from the chicken class mu gene.

In humans, the evidence is quite strong that there is only one M3-like gene; all of the bands seen in Southern blots of human DNA with a *GSTM3* cDNA probe can be assigned to the *GSTM3* gene on chromosome 1 shown in Fig. 2 (hybridization data not shown). In addition, there is no evidence for additional M3-like proteins in rodents or other mammals (15). While we do not know why the primordial *GSTM1* gene duplicated and *GSTM3* did not, we note that once the *GSTM1* duplicated the first time, the pair of head-to-tail genes would greatly increase the likelihood of additional duplication through unequal crossing-over events. That both humans and rodents have several M1-type genes and a single *GSTM3* suggests an initial *GSTM1* duplication before the divergence of the rodent and primate lines.

While there is considerable evidence for gene conversion and other interactions among the tandemly arranged head-to-tail M1-like genes on chromosome 1, the tail-to-tail orientation of *GSTM3* with respect to the M1-like genes makes it unlikely that significant interactions have occurred between *GSTM3* and the M1-like genes. Likewise, the putative promoter of the *GSTM3* gene is located almost 40,000 bp from the 3′-boundary of the *GSTM1*-0 deletion, reducing the likelihood that the *GSTM1*-0 deletion has any direct effect on transcription from *GSTM3*. Moreover because of very high identity shared between the 3′-downstream sequences of *GSTM2* and *GSTM1* (7), the *GSTM1*-0 deletion produces a *GSTM2–GSTM5* intergenic region that is almost identical to the *GSTM1–GSTM5* region in individuals carrying a non-null *GSTM1* allele. Thus, correlations between the *GSTM1* deletion and *GSTM3* expression are expected to be weak.

The human class-mu glutathione transferase gene cluster appears to have a dynamic and complex history. One member of the cluster, *GSTM1*, is deleted at about 70% of human loci (31); this deletion has been shown to result from a homologous unequal crossing-over event (7) for which the reciprocal gene duplication has re-

cently been found (32). Rodent gene-mapping experiments suggest a similar class mu cluster on the syntenic regions of rat chromosome 2 and mouse chromosome 3 (33, 34); we predict that this cluster will contain both M1-like and M3-like genes in a tail-to-tail configuration. While rodent class mu genes are dramatically induced by anticarcinogenic antioxidants (2), similar induction has not been observed in humans (35). Comparison of the human and rodent glutathione transferase class mu gene clusters may provide new insights into the regulation of class mu glutathione transferase genes.

## ACKNOWLEDGMENTS

This research was supported by a grant from the American Cancer Society (CN-92) to W.R.P. and a grant from the National Institutes of Health (CA42448) to I.L.

## REFERENCES

1. Ketterer, B., and Christodoulides, L. G. (1994) *Adv. Pharmacol.* **27,** 37–69.
2. Hayes, J. D., and Pulford, D. J. (1995) *Crit. Rev. Biochem. Mol. Biol.* **30,** 445–600.
3. Mannervik, B., Awasthi, Y. C., Board, P. G., Hayes, J. D., Di Ilio, C., Ketterer, B., Listowsky, I., Morgenstern, R., Muramatsu, M., Pearson, W. R., Pickett, C. B., Sato, K., Widersten, M., and Wolf, C. R. (1992) *Biochem. J.* **282,** 305–308.
4. Board, P. G., Baker, R. T., Chelvanayagam, G., and Jermin, L. S. (1997) *Biochem. J.* **328,** 929–935.
5. Ross, V. L., Board, P. G., and Webb, G. C. (1993) *Genomics* **18,** 87–91.
6. Pearson, W. R., Vorachek, W. R., Xu, S., Berger, R., Hart, I., Vannais, D., and Patterson, D. (1993) *Am. J. Hum. Genet.* **53,** 220–233.
7. Xu, S., Wang, Y., Roe, B., and Pearson, W. R. (1998) *J. Biol. Chem.* **273,** 3517–3527.
8. Xu, S., and Pearson, W. R. (1996) *in* Proceedings of the International ISSX-Workshop on Glutathione *S*-Transferases (Vermeulen, N. P. E., Mulder, G. J., Nieuwenhuyse, H., Peters, W. H. M., and van Bladeren, P. J., Eds.), pp. 227–238, Taylor and Francis, Basingstoke, UK.
9. Smith, G., Stanley, L. A., Sim, E., Strange, R. C., and Wolf, C. R. (1995) *Cancer Surveys* **25,** 27–65.
10. London, S. J., Daly, A. K., Cooper, J., Navidi, W. C., Carpenter, C. L., and Idle, J. R. (1995) *J. Natl. Cancer Inst.* **87,** 1246–1253.
11. Nakajima, T., Elovaara, E., Anttila, S., Hirvonen, A., Camus, A. M., Hayes, J. D., Ketterer, B., and Vainio, H. (1995) *Carcinogenesis* **16,** 707–711.
12. Hand, P. A., Inskip, A., Gilford, J., Alldersea, J., Elexpuru-Camiruaga, J., Hayes, J. D., Jones, P. W., Strange, R. C., and Fryer, A. A. (1996) *Carcinogenesis* **17,** 1919–1922.
13. Rowe, J. D., Nieves, E., and Listowsky, I. (1997) *Biochem. J.* **325,** 481–486.
14. Fulcher, K. D., Welch, J. E., Klapper, D. G., O'Brien, D. A., and Eddy, E. M. (1995) *Mol. Reprod. Dev.* **42,** 415–424.
15. Rowe, J. D., Patskovsky, Y. V., Patskovska, L. N., Novikova, E., and Listowsky, I. (1998) *J. Biol. Chem.* **273,** 9593–9601.
16. Inskip, A., Elexperu-Camiruaga, J., Buxton, N., Dias, P. S., MacIntosh, J., Campbell, D., Jones, P. W., Yengi, L., Talbot, J. A., Strange, R. C., and Fryer, A. A. (1995) *Biochem. J.* **312,** 713–716.
17. Yengi, L., Inskip, A., Gilford, J., Alldersea, J., Bailey, L., Smith, A., Lear, J. T., Heagerty, A. H., Bowers, B., Hand, P., Hayes, J. D., Jones, P. W., Strange, R. C., and Fryer, A. A. (1996) *Cancer Res.* **56,** 1974–1977.
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Pearson, W. R., Reinhart, J., Sisk, S., Anderson, K., and Adler, P. N. (1988) *J. Biol. Chem.* **263,** 13324–13332.
20. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
21. Church, G., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
22. Campbell, E., Takahashi, Y., Abramovitz, M., Peretz, M., and Listowsky, I. (1990) *J. Biol. Chem.* **265,** 9188–9193.
23. Bairoch, A., and Apweiler, R. (1996) *Nucleic Acids Res.* **24,** 21–25.
24. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
25. Felsenstein, J. (1989) *Cladistics* **5,** 164–166.
26. Pearson, W. R. (1990) *Methods Enzymol.* **183,** 63–98.
27. Smit, A. F. A., and Green, P., http://ftp.genome.washington.edu/RM/RepeatMasker.html and personal communication, 1997.
28. Schug, J., and Overton, G. C. (1997) TESS: Transcription Element Search Software on the WWW, Technical Report CBIL-TR-1997-1001-v0.0, Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.
29. Batzer, M. A., Deininger, P. L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C. M., Schmid, C. W., Zietkiewicz, E., and Zuckerkandl, E. (1996) *J. Mol. Evol.* **42,** 3–6.
30. Vorachek, W. R., Pearson, W. R., and Rule, G. S. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4443–4447.
31. Suzuki, T., Coggan, M., Shaw, D. C., and Board, P. G. (1987) *Ann. Hum. Genet.* **51,** 95–106.
32. Mclellan, R. A., Oscarson, M., Alexandrie, A. K., Seidegard, J., Evans, D. A. P., Rannug, A., and Ingelman-Sundberg, M. (1997) *Mol. Pharmacol.* **52,** 958–965.
33. Klinga-Levan, K., Andersson, A., Hanson, C., Ridderstrom, M., Stenberg, G., Mannervik, B., Vajdy, M., Szpirer, J., Szpirer, C., and Levan, G. (1993) *Hereditas* **119,** 285–296.
34. Muramatsu, Y., Yamada, T., Agui, T., Yamada, J., Serikawa, T., and Matsumoto, K. (1993) *Cytogenet. Cell Genet.* **63,** 141–143.
35. Morel, F., Fardel, O., Meyer, D. J., Langouet, S., Gilmore, K. S., Meunier, B., Tu, C. P., Kensler, T. W., Ketterer, B., and Guillouzo, A. (1993) *Cancer Res.* **53,** 231–234.
36. De Bruin, W. C., Te Morsche, R. H., Wagenmans, M. J., Alferink, J. C., Townsend, A. J., Wieringa, B., and Peters, W. H. (1998) *Biochem. J.* **330,** 623–626.