0026-895X/97/060958-08$3.00/0

MOLECULAR PHARMACOLOGY, **52:**958–965 (1997).

Characterization of a Human Glutathione *S*-Transferase μ Cluster Containing a Duplicated *GSTM1* Gene that Causes Ultrarapid Enzyme Activity

ROMAN A. McLELLAN, MIKAEL OSCARSON, ANNA-KARIN ALEXANDRIE, JANERIC SEIDEGÅRD, DAVID A. PRICE EVANS, AGNETA RANNUG, and MAGNUS INGELMAN-SUNDBERG

Division of Molecular Toxicology (R.A.M., M.O., M.I.-S.) and Unit of Genetic Toxicology (A.-K.A., A.R.), Institute of Environmental Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden, National Institute for Working Life, 171 84 Solna, Sweden (A.-K.A., A.R.), The Wallenberg Laboratory and Astra Draco AB, 221 00 Lund, Sweden (J.S.), and C 123 Riyadh Armed Forces Hospital, Riyadh 11159, Kingdom of Saudi Arabia (D.A.P.E.)

Received July 16, 1997; Accepted August 26, 1997

SUMMARY

The μ class glutathione *S*-transferase gene *GSTM1* is polymorphic in humans, with approximately half of the Caucasian population being homozygous deleted for this gene. GSTM1 enzyme deficiency has been suggested to predispose people to lung and bladder cancer. Some people in a Saudi Arabian population, however, have been described previously with ultrarapid GSTM1 enzyme activity. Here we have evaluated the molecular genetic basis for this observation. Genomic DNA from two Saudi Arabian subjects exhibiting ultrarapid enzyme activity and from 13 Swedish subjects having null, one, or two *GSTM1* genes were subjected to restriction fragment length polymorphism analysis using the restriction enzymes *Eco*RI, *Eco*RV, and *Hin*dIII and combinations thereof. Hybridization was carried out using a full-length GSTM1 cDNA or the 5′ and 3′ parts of the cDNA. The restriction mapping data revealed the presence of a *GST* μ cluster with two *GSTM1* genes in tandem situated between the *GSTM2* and *GSTM5* genes. A quantitative multiplex polymerase chain reaction method, which simultaneously amplified a fragment of the *GSTM1* gene and the β-globin gene, was developed, and the genomic *GSTM1* copy number was determined from the *GSTM1*/β-globin ratio. This method clearly separated *GSTM1* +/− subjects (ratios between 0.4 and 0.7) from *GSTM1* +/+ subjects (ratios between 0.8 and 1.2). The two Saudi Arabians with ultrarapid GSTM1 activities had ratios of approximately 1.5, indicating that they carried three *GSTM1* genes. These results demonstrate the existence of a novel μ class *GST* cluster containing a duplicated active *GSTM1* gene causing ultrarapid enzyme activity.

The GSTs are a family of enzymes involved in xenobiotic detoxification by means of conjugating glutathione to the electrophilic center of the compound. GSTs contribute in the protection against a broad range of compounds including carcinogens, pesticides, antitumor agents, and environmental pollutants (1). The mammalian cytosolic GSTs have been grouped into six classes based on amino acid sequence similarity and antibody cross-reactivity, which are designated α, μ, π, σ, θ, and κ. One member of the *GST* μ class, *GSTM1*, is polymorphic in humans and three alleles have been described: *GSTM1*A*, *GSTM1*B,* and *GSTM1*O* (2–4). The *GSTM1*A* and **B* alleles differ only by a K172N amino acid exchange and seem to be functionally identical (5). The *GSTM1*O* allele has been shown to be the result of a deletion of the entire *GSTM1* gene (6). Approximately half of the Caucasian population are homozygous deleted for this allele and fail to express the protein (7). As the GSTM1 enzyme is effective at detoxifying some carcinogenic epoxides, including the highly carcinogenic diol-epoxide intermediate from benzo[*a*]pyrene (8), much importance has been placed on the *GSTM1*O*/*O* (*GSTM1* −/−) genotype. It has been suggested that homozygous deletion of the *GSTM1* gene is associated with an increased risk of developing some types of lung cancers, in particular adenocarcinomas (9, 10) and squamous cell carcinomas (11, 12). There are numerous studies that have also demonstrated a significant association between subjects lacking GSTM1 activity and the risk of developing bladder cancer (13, 14), adenocarcinoma of the stomach and colon (15, 16), and pituitary adenomas (17). Although this association has not been observed in some investigations (16, 18) and is probably confounded by other factors (1, 19), it seems that the combination of cigarette smoking and lack of GSTM1 activity constitutes an increased risk for developing certain cancers.

In humans, the *GSTM1* gene is situated in the *GST* μ cluster, which has been localized to chromosome one in the

This work was supported by grants from the European Union, Biomed 2, the Swedish Medical Research Council, and the Research and Ethical Committee, Riyadh Armed Forces Hospital.

ABBREVIATIONS: GST, glutathione *S*-transferase; RFLP, restriction fragment length polymorphism; PCR, polymerase chain reaction; kb, kilobase pair(s); bp, base pair(s).

region 1p13.3 (20, 21). The cluster contains four other μ class genes in addition to *GSTM1*, namely *GSTM2*, *M3*, *M4*, and *M5* (20). A detailed physical map of the cluster has been described showing the *GSTM1* gene to be situated downstream of the *GSTM4* and *GSTM2* genes and upstream of *GSTM5* (22). The *GSTM1*O* deletion is postulated to be caused by an unequal crossing over event between sequences about 5 kb downstream from *GSTM2* and *GSTM1* (22), resulting in the deletion of the entire *GSTM1* gene and spanning a region of approximately 18 kb. Evidence suggests that at least one other μ class *GST* gene or pseudogene exists and is found on chromosome 3, probably in the region 3p24–3pter (20, 23). With the exception of *GSTM1*, none of the other known *GST* μ genes have been found to be deleted in humans. The only other polymorphisms that have been identified in these genes are a *Hin*dIII RFLP, which probably occurs in the *GSTM5* gene (20) and a 3-base deletion in intron 6 of *GSTM3* (24).

In a study examining the potential association of the GSTM1 null phenotype and coronary atherosclerosis development among Saudi Arabians and Filipinos, it was observed that a small percentage of the subjects expressed very high GSTM1 activities when compared with activities expected for a *GSTM1* +/− or +/+ genotype (25). To determine the molecular genetic basis for this ultrarapid GSTM1 activity, restriction mapping and multiplex PCR were used to analyze the *GST* μ cluster in genomic DNA from Swedish and Saudi Arabian subjects having variable levels of enzyme activity, as measured using the substrate *trans*-stilbene oxide. The results presented in this study describe the identification and characterization of a duplicated *GSTM1* gene carried by the people who displayed ultrarapid GSTM1 activity.

Materials and Methods

Nomenclature. This report uses the nomenclature system recommended for human GSTs (26).

Samples and GSTM1 activity. Genomic DNA samples were obtained from Saudi Arabian subjects who participated in a GSTM1 phenotyping study (25) and Swedish subjects involved in a lung cancer susceptibility study (10). These people were phenotyped for GSTM1 activity in whole blood using *trans*-stilbene oxide as the substrate according to the assay procedure described previously (27). In particular, two people were selected from the Saudi Arabian population because they displayed very high GSTM1 activities. Other Saudi Arabian and Swedish control samples were chosen for analysis based on their GSTM1 activities, including those with phenotypes suggesting *GSTM1* +/−, +/+, or −/− genotypes. The present study was approved by the ethical committee at Karolinska Institutet.

Genomic RFLP analysis. After restriction enzyme digestion, samples (2 μg per lane) were subjected to gel electrophoresis at 0.8 V/cm for a period of 24 hr (*Eco*RI and *Hin*dIII digests) or 6 days (*Eco*RV), using 0.8% or 0.7% agarose gels, respectively. The DNA was then transferred to Qiabrane membranes (Qiagen GmbH, Hilden, Germany). A full-length human GSTM1*B cDNA probe described previously (6) was used for Southern blot analysis, which also cross-hybridizes with *GSTM2*, *M4*, and *M5* but not *GSTM3* (22). To construct a 5′ half and a 3′ half of the full-length probe, the GSTM1 cDNA was cut at a *Bgl*II restriction site located in the beginning of exon 8 (6). All hybridizations and washes were done at 61°.

Multiplex PCR analysis. The oligonucleotide primers selected for multiplex PCR included the primers 5′-CTGGATTGTAGCAGATCATGC-3′ and 5′-CTCCTGATTATGACAGAAGCC-3′, which amplify a 625-bp fragment of the human *GSTM1* gene (28). A 268-bp fragment from the human β-globin gene was amplified simultaneously using the primers 5′-CAACTTCATCCACGTTCACC-3′ and 5′-GAAGAGCCAAGGACAGGTAC-3′ (29). Approximately 125 ng of genomic DNA from each sample were added to a reaction tube containing 0.5 μM of each primer, 1 × reaction buffer IV (Advanced Biotechnologies, Surrey, UK), 0.25 mM of each dNTP, 1.6 mM $MgCl_2$ and 1.2 units of *Taq* DNA polymerase (Advanced Biotechnologies, Surrey, UK) in a total reaction volume of 30 μl. PCR amplification was achieved using a Perkin Elmer Cetus thermal cycler (Norwalk, CT) with a 94° initial denaturation for 5 min, followed by 23 cycles of 94° for 48 sec, 60° for 48 sec, 72° for 1.5 min, and a final extension of 72° for 5 min. To determine the optimal cycle number, a *GSTM1* +/+ sample was amplified for increasing cycles to a maximum number of 32. When quantified and plotted against cycle number, it was established that 23 cycles of PCR amplification fell in the exponential phase of the reaction and yet was easily visible on ethidium bromide stained gels. Samples were analyzed by running 12 μl of the product on a 1.5% agarose gel containing ethidium bromide and photographed under UV light using Polaroid type 665 film. Quantification of PCR band and Southern blot fragment intensities was performed using a personal densitometer (Molecular Dynamics, Sunnyvale, CA).

Results

Southern blot analysis of the *GST* μ cluster from people displaying very high GSTM1 activity. The fragments observed in *Eco*RI Southern blot analysis have been assigned to the known μ class *GST* genes that can cross-hybridize with the full-length GSTM1 probe used (22). The 8.0-kb fragment contains the *GSTM1* gene and is absent in people who are homozygous deleted for the gene (6). Genomic DNA samples from 13 Swedish control subjects with variable GSTM1 enzyme activities indicating the presence of null, one, or two *GSTM1* genes (Table 1) were subjected to *Eco*RI RFLP analysis to determine the actual genotypes (Fig. 1). The Southern blot revealed the absence of the 8.0-kb fragment in one subject (S5) with no GSTM1 activity. The other

TABLE 1

GSTM1 activities and genotypes in Swedish control subjects and two Saudi Arabians with ultrarapid GSTM1 activity

GSTM1 activities were obtained from previous studies (10, 25) and determined in whole blood using *trans*-stilbene oxide as the substrate. Activities <600 indicated a *GSTM1* −/− genotype, 600–2500 a +/− genotype, and >2500 a +/+ genotype. *GSTM1/GSTM4* ratios were calculated from the *Eco*RI Southern blot as described in Fig. 1 and under Materials and Methods. *GSTM1* genotypes were determined from the *GSTM1/GSTM4* ratios, whereby a ratio of 0.5–0.8 represented a *GSTM1* +/− genotype, 1.2–1.5 a +/+ genotype, and 1.7–2.0 a +/++ genotype.

Sample	GSTM1 activity	*GSTM1/GSTM4* ratio	*GSTM1* genotype
	pmol/min/ml of whole blood		
A1	5900	1.69	+/++
S1	2600	1.23	+/+
S2	2400	0.56	+/−
S3	2300	1.25	+/+
S4	1300	0.67	+/−
S5	300	0	−/−
S6	1600	0.72	+/−
S7	4400	1.38	+/+
S8	1500	0.61	+/−
S9	400	0	−/−
S10	1400	0.77	+/−
S11	3200	1.41	+/+
S12	1500	0.67	+/−
A2	7500	1.96	+/++
S13	2500	0.63	+/−

S, Swede; A, Saudi Arabian.

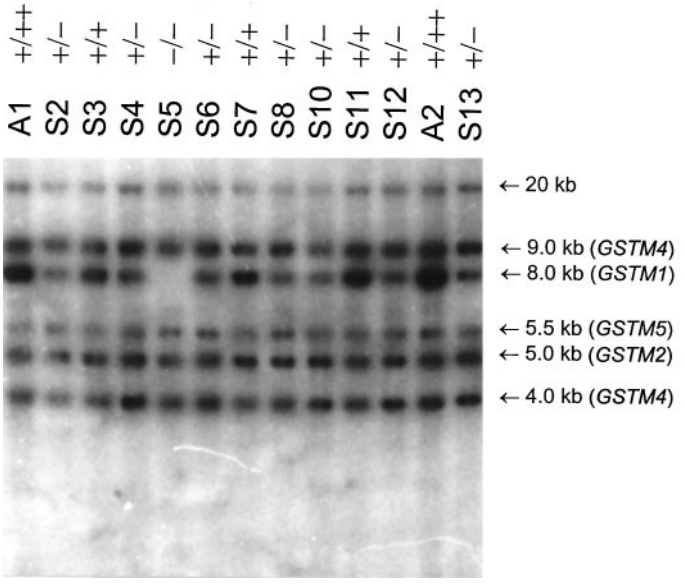

Fig. 1. *Eco*RI RFLP analysis of the μ class *GST* genes from genomic DNA samples. Hybridization was done with a human GSTM1 cDNA probe. The lengths of the fragments are indicated and the *GST* μ gene found on each fragment. DNA samples from 11 Swedish (*S*) people and two Saudi Arabian subjects with ultrarapid GSTM1 activity (*A1* and *A2*) are shown. The *GSTM1* genotypes of the people are also indicated; + *GSTM1* gene; −, the absence of a gene.

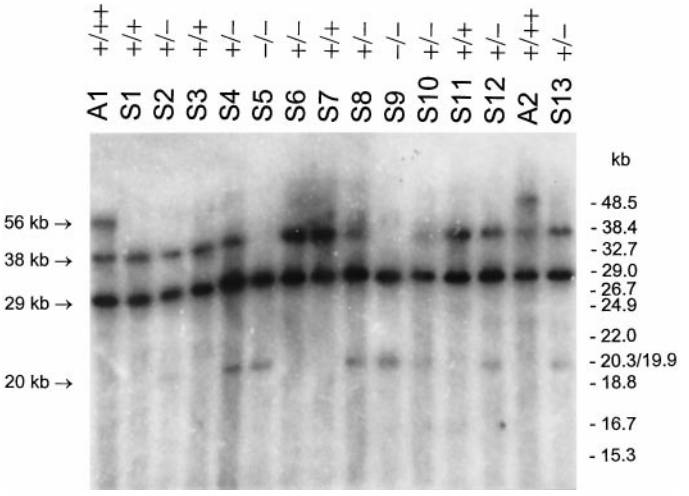

Fig. 2. *Eco*RV RFLP analysis of the μ class *GST* genes from genomic DNA samples. Hybridization was done with a human GSTM1 cDNA probe. The approximate lengths of the fragments are indicated on the left-hand side and the positions of DNA molecular weight Marker XV (Boehringer Mannheim, Mannheim, Germany) fragments are shown on the right-hand side. DNA samples from 13 Swedish people (*S1–S13*) and two Saudi Arabian subjects with ultrarapid GSTM1 activity (*A1* and *A2*) are shown. The *GSTM1* genotypes of the people are also indicated; + *GSTM1* gene; −, the absence of a gene.

samples carried the 8.0-kb *Eco*RI fragment. The intensity of this fragment was compared with that of the 9.0-kb fragment (*GSTM4*) used as an internal standard, and a ratio was calculated from the densities of the two fragments to determine the *GSTM1* genotype of the people. As shown in Table 1, the homozygous and heterozygous carriers were clearly separated into two groups whereby *GSTM1* +/− subjects gave ratios between 0.5 and 0.8 and those with a *GSTM1* +/+ genotype between 1.2 and 1.5. It was also observed that the two Saudi Arabian people with very high GSTM1 activities (A1 and A2) had 8.0-kb *Eco*RI fragments approximately 30 and 50% more intense than the mean of the four *GSTM1* +/+ samples (see Fig. 1), suggesting the presence of an extra *GSTM1* gene in these people.

To further characterize the locus from subjects having ultrarapid enzyme activity, Southern blot analysis using the restriction enzyme *Eco*RV was performed as shown in Fig. 2. Using the same full-length GSTM1 probe that was employed for *Eco*RI RFLP analysis, a 29-kb fragment was observed in all samples. A 38-kb fragment was seen in *GSTM1* +/+ and +/− subjects but was absent in subjects homozygous for a gene deletion, indicating that the *GSTM1* gene was located on this fragment. Samples homozygous or heterozygous deleted for the *GSTM1* gene demonstrated a 20-kb fragment. An additional fragment of approximately 56 kb was present only in the two Saudi Arabians with ultrarapid GSTM1 activity. Previous data have shown the absence of *Eco*RV restriction sites located in the complete sequences of the *GSTM1* and *M4* genes (30, 31) or in the cDNA from the *GSTM2* gene (32). The only reported *Eco*RV restriction site has been found in the 3′-noncoding region of the cDNA from *GSTM5* (33). The extra fragment of 56 kb was thus suggestive of a *GST* μ cluster containing two *GSTM1* genes on the fragment (Fig. 2).

Determination of the approximate locations of the *Eco*RV restriction sites in the *GST* μ cluster. To characterize the cluster containing two *GSTM1* genes with respect to the location of the *Eco*RV restriction sites, genomic DNA from two Swedes, one being *GSTM1* +/+ (S1) and the other *GSTM1* −/− (S9) were digested with selected combinations of restriction enzymes and hybridized with the GSTM1 probe. When the combination of *Eco*RV and *Eco*RI restriction enzymes were used on the two genomic DNA samples, the previously assigned *Eco*RI fragment of 20 kb (6) was lost in both cases, and the 9.0-kb fragment appeared with only half of the intensity (Fig. 3A). Hybridization with the 3′ half of the GSTM1 probe (6) (see Materials and Methods) caused the 9.0-kb fragment to disappear from samples cleaved with *Eco*RI and *Eco*RV, indicating that it was part of the original 20-kb fragment and that the 9.0-kb *Eco*RI fragment was cut by *Eco*RV (Fig. 3A). An additional fragment of approximately 5 kb was also observed in samples that were digested with the combination of enzymes when hybridized with the full-length or the 5′ half of the probe. It represents one of the two fragments created by *Eco*RV digestion of the 9.0-kb *Eco*RI fragment, which contains the *GSTM4* gene.

The two samples were also digested with *Hin*dIII alone and with a combination of *Eco*RV and *Hin*dIII (Fig. 3B). The fragments obtained after *Hin*dIII digestion of genomic DNA have been assigned to the μ class *GST* genes (22). In particular three 5.5-kb fragments are observed, one of which contains the 5′ end of *GSTM4* and another the 3′ end of *GSTM2*. As shown in Fig. 3B, the 5.5-kb fragment, which was not shortened using both *Eco*RV and *Hin*dIII, hybridized with the 5′ half and not the 3′ half of the probe, showing that it contained the 5′ part of the *GSTM4* gene and had no *Eco*RV site. One of the 5.5-kb *Hin*dIII fragments was shortened to 3.8 kb when the two samples were digested with the combination of *Eco*RV and *Hin*dIII. The 3.8-kb fragment hybridized with the 3′ half of the probe and not with the 5′ half,

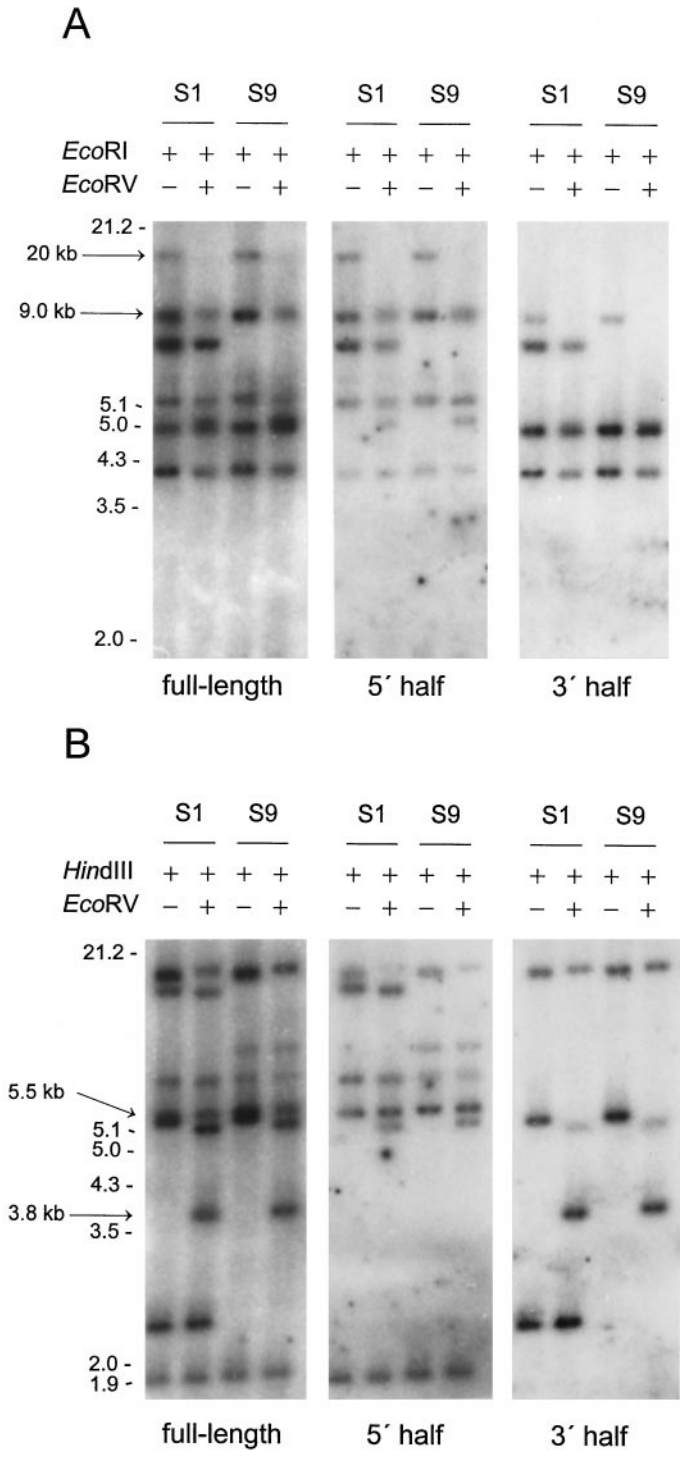

Fig. 3. RFLP analysis of the μ class *GST* genes using different combinations of restriction enzymes. Genomic DNA samples from a *GSTM1* +/+ (*S1*) and a *GSTM1* −/− (*S9*) subject were used. A, Samples were digested with *Eco*RI alone and in combination with *Eco*RV as indicated. The blot was hybridized with the full-length GSTM1 cDNA probe and also the 5′ and 3′ parts of the same probe (see Materials and Methods). The fragment sizes of a molecular weight marker (*Hin*dIII/*Eco*RI digest of λ DNA) are shown in kilobases. B, Samples were digested with *Hin*dIII alone and in combination with *Eco*RV as indicated. The blot was hybridized with the full-length GSTM1 cDNA probe and also the 5′ and 3′ parts. The fragment sizes of a molecular weight marker (*Hin*dIII/*Eco*RI digest of λ DNA) are shown in kilobases.

indicating that it was created by *Eco*RV digestion of the 5.5-kb fragment that contained the 3′ end of the *GSTM2* gene. The other 5.5-kb fragment that was shortened to 5.1 kb using the enzyme combination probably contains another μ class *GST* gene. Based on the *GST* μ cluster map (22) and the present Southern blot results, maps of clusters containing a duplicated *GSTM1* gene, one *GSTM1* gene, and a deleted *GSTM1* gene were determined and are shown in Fig. 4, A and B.

Determination of *GSTM1* genomic copy number from DNA samples. A quantitative multiplex PCR method was used to simultaneously amplify a fragment of the *GSTM1* gene and the β-globin gene from genomic DNA. By calculating the densitometric ratio between the intensity of the *GSTM1* band and the constant β-globin band, the *GSTM1* copy number carried by an individual could be determined. Genomic DNA from the Swedish control subjects, which had been analyzed for GSTM1 activity and genotyped by *Eco*RI RFLP densitometric analysis, was amplified and subsequently quantified to determine the *GSTM1* copy number. A representative gel showing the PCR products is shown in Fig. 5. The two Saudi Arabians with ultrarapid enzyme activity had *GSTM1* bands that were much more intense than that of *GSTM1* +/+ subjects. Using the *GSTM1*/β-globin ratios calculated from the control samples, a ratio falling between 0.40 and 0.70 was defined as representing a *GSTM1* +/− genotype and one between 0.80 and 1.20 indicated a *GSTM1* +/+ genotype. This approach allowed for the clear separation between the two different genotypes (Table 2 and Fig. 6A). A good correlation between the *GSTM1* copy number determined by multiplex PCR and the assigned *Eco*RI RFLP genotype was observed (Table 2). The *GSTM1*/β-globin ratios for the ultrarapid Saudi Arabian subjects were approximately 1.5, higher than the range representing a *GSTM1* +/+ genotype and indicative of three *GSTM1* genes (Table 2 and Fig. 6A).

Twenty additional Saudi Arabian people with phenotypes suggesting *GSTM1* +/− or +/+ genotypes were analyzed for *GSTM1* copy number. This was done to ensure that the comparison between the two ultrarapid Saudi Arabians and the Swedish samples was not influenced by factors such as interethnic differences in primer binding sequences. The *GSTM1*/β-globin ratios for the subjects are shown in Fig. 6B. No differences were observed in the calculated ratios between the two populations and all Saudi Arabian samples fell into the ranges indicative of a *GSTM1* +/− or +/+ genotype.

Discussion

The results presented indicate strongly that the molecular genetic basis for ultrarapid GSTM1 enzyme activity, demonstrated previously among Saudi Arabians, is caused by the presence of a *GST* μ cluster containing two functional *GSTM1* genes in tandem. This conclusion is based on several observations: (i) *trans*-stilbene oxide is a substrate specific for GSTM1, and this compound was used in the phenotyping study; (ii) the 8.0-kb *Eco*RI fragment contains the entire *GSTM1* gene and was more intense in subjects with the duplicated *GSTM1* gene; (iii) the 56-kb *Eco*RV fragment corresponds to a haplotype that has been generated by the addition of the 18-kb fragment lost in *GSTM1* deleted haplotypes; (iv) restriction mapping using various combinations of enzymes showed that ultrarapid activity

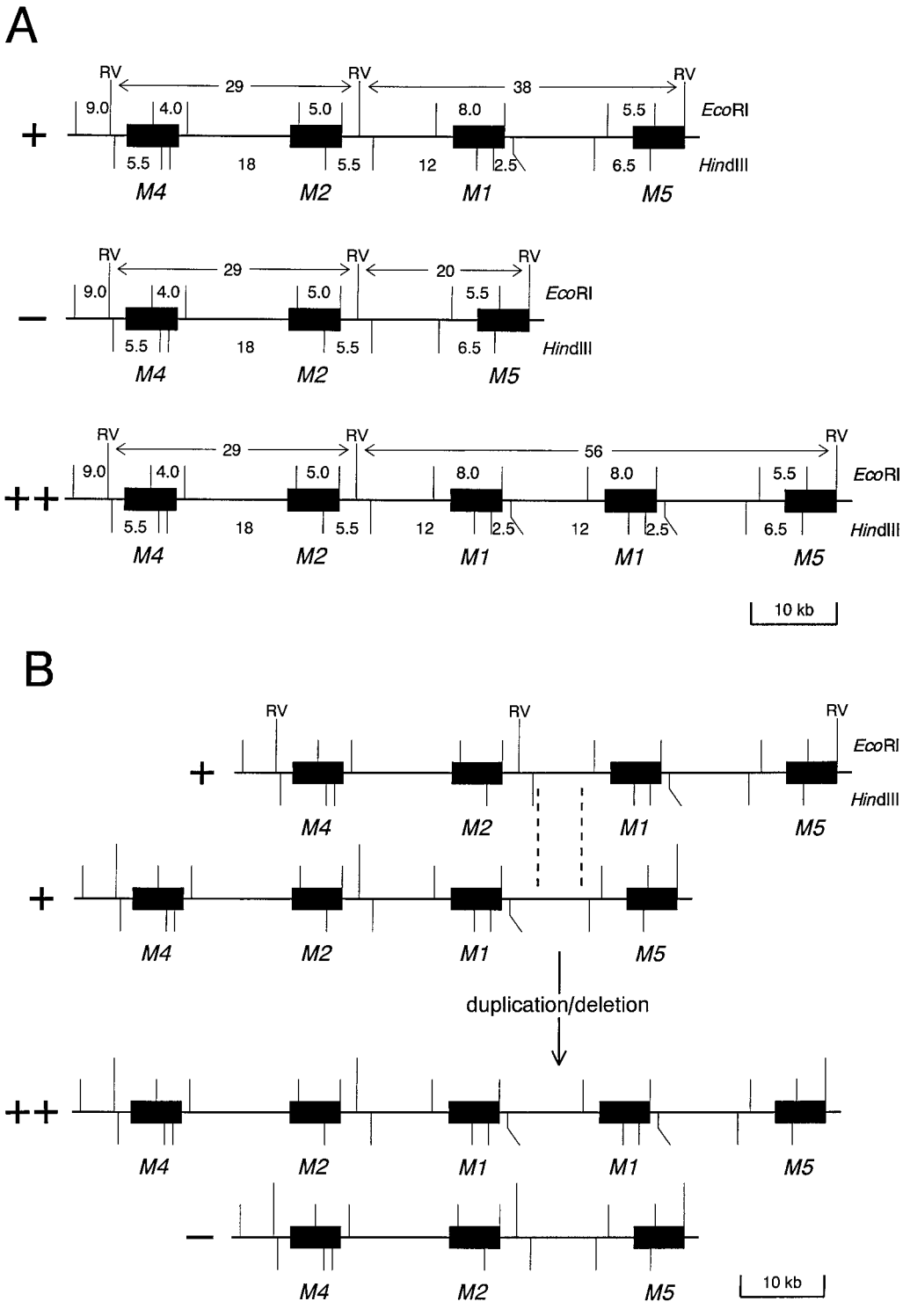

Fig. 4. Schematic illustration of the human *GST* μ cluster containing a duplicated *GSTM1* gene. A, Maps of *GST* μ clusters containing one *GSTM1* gene (+), no *M1* gene (−), and two *M1* genes (++). The relative positions of the four μ class genes located in the cluster have been determined previously. *Eco*RI, *Hin*dIII, and *Eco*RV (*RV*) restriction sites giving rise to the fragments observed during Southern blot analysis are indicated. B, Hypothetical mechanism for the generation of a *GST* μ cluster containing a duplicated *GSTM1* gene caused by an unequal crossover event between two homologous sequences.

subjects carried a *GST* μ cluster with two *GSTM1* genes in tandem; and (v) the PCR primers used in the multiplex PCR reactions were specific for the *GSTM1* gene and gave ratios indicative of an extra gene in both subjects having ultrarapid GSTM1 activity.

Quantification of *GSTM1* copy number using genomic DNA from the subjects was performed using a multiplex PCR method. The coamplification of the β-globin fragment with the variable *GSTM1* fragment allowed the determination of the *GSTM1* copy number carried by the individual and also acted as an internal amplification control for the PCR reaction. It was found that the calculated number of *GSTM1* copies determined by this method corresponded with the genotypes determined by *Eco*RI RFLP analysis. The two Saudi Arabian people with ultrarapid enzyme activities had ratios indicative of three *GSTM1* genes, supporting the Southern blot data that revealed the presence of a duplicated *GSTM1* gene. The multiplex PCR approach proved to be a simple and convenient way of determining *GSTM1* genotype and genomic copy number and could be applied easily in the

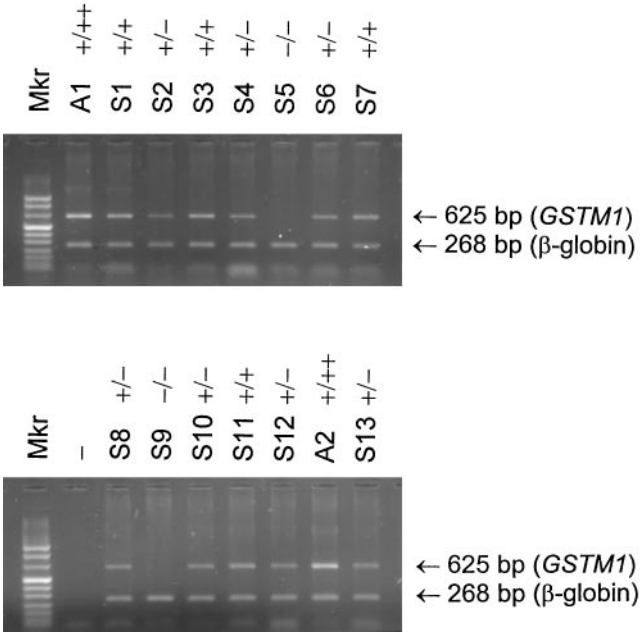

Fig. 5. Multiplex PCR analysis of genomic DNA samples for *GSTM1* copy number. A representative gel showing fragments amplified simultaneously from genomic DNA in a multiplex PCR. Thirteen Swedish samples (*S1–S13*) and the two Saudi Arabian subjects with ultrarapid GSTM1 activity (*A1* and *A2*) were amplified. The longer fragment (625 bp) was amplified from the *GSTM1* gene, and the 268-bp fragment was from the β-globin gene. Homozygous *GSTM1* deleted subjects (*S5* and *S9*) failed to amplify the *GSTM1* fragment. The molecular weight Marker VIII (Boehringer Mannheim) was run with the PCR products and also a negative control containing no template DNA (−).

TABLE 2

Determination of the *GSTM1* copy number using a multiplex PCR method

GSTM1 genotypes were determined from the *GSTM1/GSTM4* ratios as described. *GSTM1*/β-globin ratios between 0.40 and 0.70 were representative of a *GSTM1* +/− genotype, 0.80 and 1.20 a +/+ genotype, and 1.30 and 1.60 a +/++ genotype.

Sample	*GSTM1* genotype	*GSTM1*/β-globin ratio mean ± SD
A1	+/++	1.59 ± 0.13
S1	+/+	1.17 ± 0.03
S2	+/−	0.63 ± 0.03
S3	+/+	1.15 ± 0.03
S4	+/−	0.68 ± 0.04
S5	−/−	0
S6	+/−	0.70 ± 0.05
S7	+/+	1.13 ± 0.08
S8	+/−	0.54 ± 0.10
S9	−/−	0
S10	+/−	0.68 ± 0.05
S11	+/+	0.97 ± 0.10
S12	+/−	0.66 ± 0.02
A2	+/++	1.47 ± 0.02
S13	+/−	0.60 ± 0.10

S, Swede; A, Saudi Arabian.

routine analysis of *GSTM1* copy number in molecular epidemiology studies as well as for the analysis of any other gene of interest.

The generation of a *GST* μ cluster containing a duplicated *GSTM1* gene probably involved an unequal recombination between two homologous but nonallelic sequences flanking the *GSTM1* gene. This may have occurred because of a chromosomal misalignment followed by an unequal crossing over

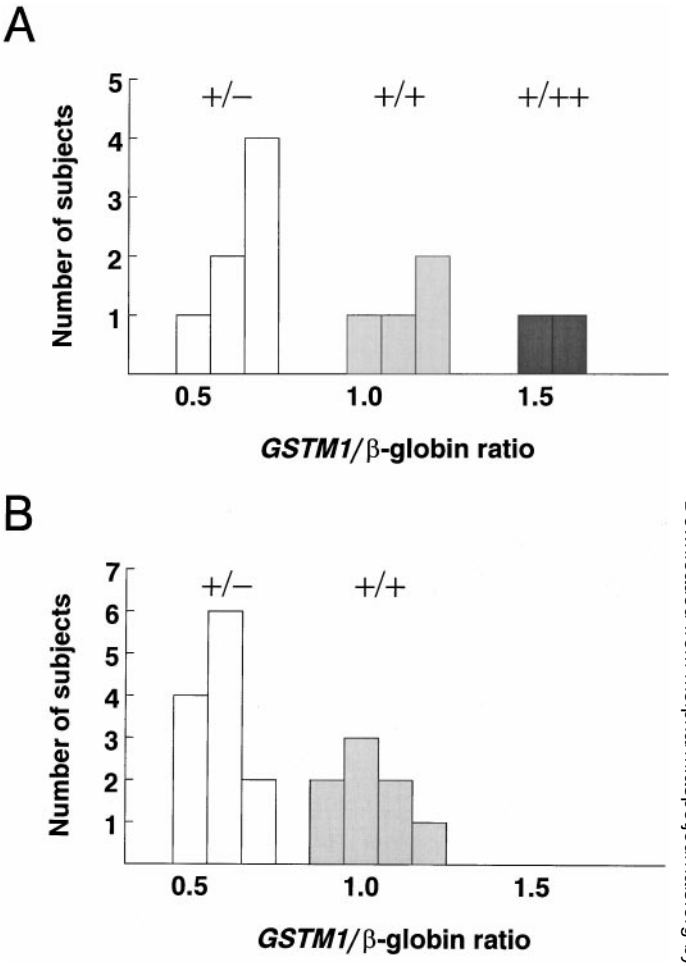

Fig. 6. Evaluation of *GSTM1*/β-globin ratios from multiplex PCR analysis in relation to *GSTM1* genotype. The ratio between the variable *GSTM1* fragment and the constant two copy β-globin fragment was used to determine the *GSTM1* copy number. A, The results from 13 Swedish subjects who had been genotyped by *Eco*RI RFLP analysis (Table 1) are shown; *open bars*, *GSTM1* +/− genotype; *light gray bars*, *GSTM1* +/+ genotype. *Dark gray bars,* the Saudi Arabians with ultrarapid GSTM1 activity (*A1* and *A2*) gave ratios that indicated three *GSTM1* genes (+/++). The results are the mean of three independent experiments. B, The distribution of *GSTM1*/β-globin ratios from Saudi Arabian people with GSTM1 phenotypes that suggest *GSTM1* +/− or +/+ genotypes are shown. Of the 20 Saudi Arabian subjects that were analyzed, all had ratios indicative of a *GSTM1* +/− genotype (*open bars*) or a *GSTM1* +/+ genotype (*light gray bars*). The results are the mean of at least three independent experiments.

event or by unequal crossing over between sister chromatids. The outcome of such an event would be the generation of a μ cluster containing two *GSTM1* genes in tandem on one cluster and a deleted *GSTM1* gene on the other (Fig. 4B). The break points of the *GSTM1* deletion have been reported to be about 5 kb downstream from the *GSTM2* gene and 5 kb downstream from the *GSTM1* gene (22), resulting in a deletion spanning approximately 18 kb. *Eco*RV RFLP analysis showed that a cluster containing one *GSTM1* gene gave a 38-kb fragment; therefore, a cluster consisting of two *GSTM1* genes would be expected to be roughly 18 kb longer because of the unequal crossover event. Both Saudi Arabians with ultrarapid GSTM1 activity carried an *Eco*RV fragment of about 56 kb, which was the expected length of a fragment consisting of two *GSTM1* genes in tandem.

A potential outcome of having extra *GSTM1* genes could be an increased protective effect against some carcinogenic chemicals because of the more rapid detoxification of these compounds. One carcinogen found in cigarette smoke is benzo[*a*]pyrene, which can be metabolized by phase I enzymes to a diol-epoxide, one of its most carcinogenic forms. This carcinogen can be further conjugated with glutathione by GSTM1, resulting in a deactivated compound (8). Therefore, people carrying extra *GSTM1* genes have the ability to rapidly conjugate and inactivate compounds that are substrates for this enzyme, including potential carcinogens.

Excluding the human immunoglobulin genes, which are subject to complex somatic recombination, reports describing inherited duplication events of an entire gene causing a phenotypic change are uncommon (34–36). One well studied example of this involves *CYP2D6*, a cytochrome P450 enzyme involved in the metabolism of many clinically used drugs and implicated as a risk factor for cancer (19). People carrying duplicated, multiduplicated, and amplified (13-fold) functional *CYP2D6* genes have been shown (34, 37–39). These subjects display ultrarapid metabolism of probe drugs for this enzyme and require higher than normal doses of drugs to achieve therapeutic levels, indicating that the extra gene(s) result in higher levels of protein to be expressed (34, 37). Also in the present case, the duplicated *GSTM1* gene seemed to be functional and a gene dosage effect was observed in the subjects with this genotype.

In the Saudi Arabian population phenotyped, 56% of the subjects were homozygous for the *GSTM1* deleted allele (25), which is roughly the same frequency as reported in several other studies in Caucasian populations. Because over half of the Saudi Arabian population were homozygous for the *GSTM1*O* allele it would be reasonable, based on the unequal crossover event as the cause for this genotype, to expect a high frequency of people carrying duplicated *GSTM1* genes. This was not the case, as only 3% of the Saudi population had very high GSTM1 activities, some of which probably carried duplicated *GSTM1* genes. However, because of the high frequency of the *GSTM1*O* allele in the population, it is possible that some people had a ++/− genotype, which would have been included in the +/+ phenotypic group. The low frequency of potential duplicated *GSTM1* genes is in contrast to the high frequency (21%) of people from the same Saudi Arabian population who were found to carry duplicated *CYP2D6* genes (39). It has been postulated that the *CYP2D6* gene has undergone a selective environmental pressure in some populations, including the Saudi Arabians (40). The high frequency of the *GSTM1*O* allele suggests that this gene has not encountered strong environmental selection pressure during evolution, although the rapidly expanding number of compounds found in the environment make it likely that the GSTM1 enzyme might be increasingly involved in chemical detoxification.

In conclusion, this study identified a duplicated functional *GSTM1* gene as the genetic basis for the ultrarapid enzyme activity observed in some Saudi Arabian subjects. The finding that this polymorphic detoxification enzyme has undergone a stable duplication event implicates that other enzymes involved in the metabolism of foreign compounds may also be found in multiple copies and cause phenotypic alterations. The global distribution of people carrying duplicated or multiduplicated *GSTM1* genes and the consequences of multiple *GSTM1* gene copies for environmentally related diseases remain to be determined.

References

1. Hayes, J. D., and D. J. Pulford. The glutathione *S*-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.* **30:** 445–600 (1995).
2. Board, P. G. Biochemical genetics of glutathione-*S*-transferase in man. *Am. J. Hum. Genet.* **33:**36–43 (1981).
3. DeJong, J. L., C. M. Chang, J. Whang-Peng, T. Knutsen, and C. P. Tu. The human liver glutathione *S*-transferase gene superfamily: expression and chromosome mapping of an H_b subunit cDNA. *Nucleic Acids Res.* **16:**8541–8554 (1988).
4. Seidegård, J., and R. W. Pero. The hereditary transmission of high glutathione transferase activity towards *trans*-stilbene oxide in human mononuclear leukocytes. *Hum. Genet.* **69:**66–68 (1985).
5. Widersten, M., W. R. Pearson, A. Engström, and B. Mannervik. Heterologous expression of the allelic variant μ-class glutathione transferases μ and ψ. *Biochem. J.* **276:**519–524 (1991).
6. Seidegård, J., W. R. Vorachek, R. W. Pero, and W. R. Pearson. Hereditary differences in the expression of the human glutathione transferase active on *trans*-stilbene oxide are due to a gene deletion. *Proc. Natl. Acad. Sci. USA* **85:**7293–7297 (1988).
7. Board, P., M. Coggan, P. Johnston, V. Ross, T. Suzuki, and G. Webb. Genetic heterogeneity of the human glutathione transferases: a complex of gene families. *Pharmacol. Ther.* **48:**357–369 (1990).
8. Warholm, M., C. Guthenberg, and B. Mannervik. Molecular and catalytic properties of glutathione transferase μ from human liver: an enzyme efficiently conjugating epoxides. *Biochemistry* **22:**3610–3617 (1983).
9. Seidegård, J., R. W. Pero, M. M. Markowitz, G. Roush, D. G. Miller, and E. J. Beattie. Isoenzyme(s) of glutathione transferase (class Mu) as a marker for the susceptibility to lung cancer: a follow up study. *Carcinogenesis* **11:**33–36 (1990).
10. Alexandrie, A. K., M. Ingelman-Sundberg, J. Seidegård, G. Tornling, and A. Rannug. Genetic susceptibility to lung cancer with special emphasis on *CYP1A1* and *GSTM1*: a study on host factors in relation to age at onset, gender and histological cancer types. *Carcinogenesis* **15:**1785–1790 (1994).
11. Hirvonen, A., K. Husgafvel-Pursiainen, S. Anttila, and H. Vainio. The *GSTM1* null genotype as a potential risk modifier for squamous cell carcinoma of the lung. *Carcinogenesis* **14:**1479–1481 (1993).
12. Kihara, M., M. Kihara, and K. Noda. Lung cancer risk of GSTM1 null genotype is dependent on the extent of tobacco smoke exposure. *Carcinogenesis* **15:**415–418 (1994).
13. Bell, D. A., J. A. Taylor, D. F. Paulson, C. N. Robertson, J. L. Mohler, and G. W. Lucier. Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen-metabolism gene glutathione *S*-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. *J. Natl. Cancer Inst.* **85:**1159–1164 (1993).
14. Brockmöller, J., R. Kerb, N. Drakoulis, B. Staffeldt, and I. Roots. Glutathione *S*-transferase M1, and its variants A and B as host factors of bladder cancer susceptibility: a case-control study. *Cancer Res.* **54:**4103–4111 (1994).
15. Strange, R. C., B. Matharoo, G. C. Faulder, P. Jones, W. Cotton, J. B. Elder, and M. Deakin. The human glutathione *S*-transferases: a case-control study of the incidence of the GST1 0 phenotype in patients with adenocarcinoma. *Carcinogenesis* **12:**25–28 (1991).
16. Zhong, S., A. H. Wyllie, D. Barnes, C. R. Wolf, and N. K. Spurr. Relationship between the GSTM1 genetic polymorphism and susceptibility to bladder, breast and colon cancer. *Carcinogenesis* **14:**1821–1824 (1993).
17. Fryer, A. A., L. Zhao, J. Alldersea, M. D. Boggild, C. W. Perrett, R. N. Clayton, P. W. Jones, and R. C. Strange. The glutathione *S*-transferases: polymerase chain reaction studies on the frequency of the GSTM1 0 genotype in patients with pituitary adenomas. *Carcinogenesis* **14:**563–566 (1993).
18. Brockmöller, J., R. Kerb, N. Drakoulis, M. Nitz, and I. Roots. Genotype and phenotype of glutathione *S*-transferase class μ isoenzymes μ and ψ in lung cancer patients and controls. *Cancer Res.* **53:**1004–1011 (1993).
19. Derrico, A., E. Taioli, X. Chen, and P. Vineis. Genetic metabolic polymorphisms and the risk of cancer— a review of the literature. *Biomarkers* **1:**149–173 (1996).
20. Pearson, W. R., W. R. Vorachek, S. J. Xu, R. Berger, I. Hart, D. Vannais, and D. Patterson. Identification of class-mu glutathione transferase genes *GSTM1-GSTM5* on human chromosome 1p13. *Am. J. Hum. Genet.* **53:** 220–233 (1993).
21. Ross, V. L., P. G. Board, and G. C. Webb. Chromosomal mapping of the human Mu class glutathione *S*-transferases to 1p13. *Genomics* **18:**87–91 (1993).
22. Xu, S.-J., and W. R. Pearson. Characterization of the polymorphic human class mu GSTM1–0 deletion, in *Glutathione S-Transferases: Structure, Function and Clinical Implications* (N. P. E. Vermeulen, G. J. Mulder, H. Nieuwenhuyse, W. H. M. Peters, and P. J. Van Bladerer, eds.). Taylor & Francis Ltd., London, 227–238 (1996).
23. Islam, M. Q., A. Platz, J. Szpirer, C. Szpirer, G. Levan, and B. Mannervik.

Chromosomal localization of human glutathione transferase genes of classes alpha, mu and pi. *Hum. Genet.* **82:**338–342 (1989).
24. Inskip, A., J. Elexperu-Camiruaga, N. Buxton, P. S. Dias, J. MacIntosh, D. Campbell, P. W. Jones, L. Yengi, J. A. Talbot, R. C. Strange, and A. A. Fryer. Identification of polymorphism at the glutathione S-transferase, GSTM3 locus: evidence for linkage with *GSTM1**A. *Biochem. J.* **312:**713–716 (1995).
25. Evans, D. A. P., J. Seidegård, and N. Narayanan. The GSTM1 genetic polymorphism in healthy Saudi Arabians and Filipinos, and Saudi Arabians with coronary atherosclerosis. *Pharmacogenetics* **6:**365–367 (1996).
26. Mannervik, B., Y. C. Awasthi, P. G. Board, J. D. Hayes, C. Di Ilio, B. Ketterer, I. Listowsky, R. Morgenstern, M. Muramatsu, W. R. Pearson, C. B. Pickett, K. Sato, M. Widersten, and C. R. Wolf. Nomenclature for human glutathione transferases. *Biochem. J.* **282:**305–306 (1992).
27. Seidegård, J., J. W. De Pierre, W. Birberg, A. Pilotti, and R. W. Pero. Characterization of soluble glutathione transferase activity in resting mononuclear leukocytes from human blood. *Biochem. Pharmacol.* **33:** 3053–3058 (1984).
28. Brockmöller, J., D. Gross, R. Kerb, N. Drakoulis, and I. Roots. Correlation between *trans*-stilbene oxide-glutathione conjugation activity and the deletion mutation in the glutathione *S*-transferase class Mu gene detected by polymerase chain reaction. *Biochem. Pharmacol.* **43:**647–650 (1992).
29. Bauer, H. M., Y. Ting, C. E. Greer, J. C. Chambers, C. J. Tashiro, J. Chimera, A. Reingold, and M. M. Manos. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *JAMA* **265:**472–477 (1991).
30. Zhong, S., N. K. Spurr, J. D. Hayes, and C. R. Wolf. Deduced amino acid sequence, gene structure and chromosomal location of a novel human class Mu glutathione *S*-transferase, GSTM4. *Biochem. J.* **291:**41–50 (1993).
31. Comstock, K. E., K. J. Johnson, D. Rifenbery, and W. D. Henner. Isolation and analysis of the gene and cDNA for a human Mu class glutathione *S*-transferase, GSTM4. *J. Biol. Chem.* **268:**16958–16965 (1993).
32. Vorachek, W. R., W. R. Pearson, and G. S. Rule. Cloning, expression, and characterization of a class-mu glutathione transferase from human muscle, the product of the *GST4* locus. *Proc. Natl. Acad. Sci. USA* **88:**4443–4447 (1991).
33. Takahashi, Y., E. A. Campbell, Y. Hirata, T. Takayama, and I. Listowsky. A basis for differentiating among the multiple human Mu-glutathione *S*-transferases and molecular cloning of brain *GSTM5*. *J. Biol. Chem.* **268:**8893–8898 (1993).
34. Bertilsson, L., M.-L. Dahl, F. Sjöqvist, A. Åberg-Wistedt, M. Humble, I. Johansson, E. Lundqvist, and M. Ingelman-Sundberg. Molecular basis for rational megaprescribing in ultrarapid hydroxylators of debrisoquine. *Lancet* **341:**63 (1993).
35. Kubota, T., S. Saitoh, T. Matsumoto, K. Narahara, Y. Fukushima, Y. Jinno, and N. Niikawa. Excess functional copy of allele at chromosomal region 11p15 may cause Wiedemann-Beckwith (EMG) syndrome. *Am. J. Med. Genet.* **49:**378–383 (1994).
36. Mori, Y., Y. Miura, H. Takeuchi, Y. Igarashi, J. Sugiura, H. Saito, and Y. Oiso. Gene amplification as a cause of inherited thyroxine-binding globulin excess in two Japanese families. *J. Clin. Endocrinol. Metab.* **80:**3758–3762 (1995).
37. Johansson, I., E. Lundqvist, L. Bertilsson, M.-L. Dahl, F. Sjöqvist, and M. Ingelman-Sundberg. Inherited amplification of an active gene in the cytochrome P450 *CYP2D* locus as a cause of ultrarapid metabolism of debrisoquine. *Proc. Natl. Acad. Sci. USA* **90:**11825–11829 (1993).
38. Aklillu, E., I. Persson, L. Bertilsson, I. Johansson, F. Rodrigues, and M. Ingelman-Sundberg. Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional *CYP2D6* alleles. *J. Pharmacol. Exp. Ther.* **278:**441–446 (1996).
39. McLellan, R. A., M. Oscarson, J. Seidegård, D. A. P. Evans, and M. Ingelman-Sundberg. Frequent occurrence of *CYP2D6* gene duplication in Saudi Arabians. *Pharmacogenetics* **7:**187–191 (1997).
40. Ingelman-Sundberg, M. Genetic polymorphism of drug metabolizing enzymes. Implications for toxicity of drugs and other xenobiotics. *Arch. Toxicol. Suppl.* **19:**3–13 (1997).

Send reprint requests to: Dr. Magnus Ingelman-Sundberg, Division of Molecular Toxicology, Institute of Environmental Medicine, Karolinska Institutet, Box 210, S-171 77 Stockholm, Sweden. E-mail: magnus.ingelman-sundberg@imm.ki.se