Gene structure, expression and chromosomal localization of murine Theta class glutathione transferase mGSTT1-1

Angela T. WHITTINGTON*, Vanicha VICHAI†, Graham C. WEBB‡, Rohan T. BAKER*, William R. PEARSON† and Philip G. BOARD*[1]

*Molecular Genetics Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra, Australian Capital Territory, 2601 Australia, †Department of Biochemistry, University of Virginia, Charlottesville, VA 22908, U.S.A., and ‡Department of Obstetrics and Gynaecology, The Queen Elizabeth Hospital, Woodville, South Australia, 5011 Australia

We have isolated and characterized a cDNA and partial gene encoding a murine subfamily 1 Theta class glutathione transferase (GST). The cDNA derived from mouse *GSTT1* has an open reading frame of 720 bp encoding a peptide of 240 amino acids with a calculated molecular mass of 27356 Da. The encoded protein shares only 51 % deduced amino acid sequence identity with mouse GSTT2, but greater than 80 % deduced amino acid sequence identity with rat GSTT1 and human GSTT1. Mouse GSTT1-1 was expressed in *Escherichia coli* as an N-terminal 6 × histidine-tagged protein and purified using immobilized-metal affinity chromatography on nickel-agarose. The yield of the purified recombinant protein from *E. coli* cultures was approx. 14 mg/l. Recombinant mouse GSTT1-1 was catalytically active towards 1,2-epoxy-3-(*p*-nitrophenoxy)propane, 4-nitrobenzyl chloride and dichloromethane. Low activity towards 1-menaphthyl sulphate and 1-chloro-2,4-dinitrobenzene was detected, whereas mouse GSTT1-1 was inactive towards ethacrynic acid. Recombinant mouse GSTT1-1 exhibited glutathione peroxidase activity towards cumene hydroperoxide and t-butyl hydroperoxide, but was inactive towards a range of secondary lipid-peroxidation products, such as the *trans*-alk-2-enals and *trans*,*trans*-alka-2,4-dienals. Mouse GSTT1 mRNA is most abundant in mouse liver and kidney, with some expression in intestinal mucosa. Mouse GSTT1 mRNA is induced in liver by phenobarbital, but not by butylated hydroxyanisole, β-naptho-flavone or isosafrole. The structure of mouse *GSTT1* is conserved with that of the subfamily 2 Theta class GST genes mouse *GSTT2* and rat *GSTT2*, comprising five exons interrupted by four introns. The mouse *GSTT1* gene was found, by *in situ* hybridization, to be clustered with mouse *GSTT2* on chromosome 10 at bands B5–C1. This region is syntenic with the location of the human Theta class GSTs clustered on chromosome 22q11.2. Similarity searches of a mouse-expressed sequence tag database suggest that there may be two additional members of the Theta class that share 70 % and 88 % protein sequence identity with mouse GSTT1, but less than 55 % sequence identity with mouse GSTT2.

Key words: gene organization, glutathione S-transferases, detoxification.

INTRODUCTION

The glutathione transferases (GSTs; EC: 2.5.1.18) are a family of multifunctional isoenzymes. They play a major role in cellular detoxication by catalysing the conjugation of GSH to a variety of electrophilic compounds, including mutagens, carcinogens and some therapeutic agents. Based on sequence similarities, mammalian cytosolic GSTs have been grouped into at least six classes called Alpha, Mu, Pi [1], Theta [2], Sigma [3] and Zeta [4].

Theta class GSTs are distinguished from other classes by their failure to bind to immobilized GSH affinity matrices and their negligible activity towards the model GST substrate 1-chloro-2,4-dinitrobenzene (CDNB). The Theta class can be divided into two orthologous subfamilies in humans [2,5–8], rats [2,9–13] and mice [14–17]. Subfamily 1 isoenzymes (GSTT1-1) are typically active towards dichloromethane (DCM) and 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP) while subfamily 2 isoenzymes (GSTT2-2) are active towards 1-menaphthyl sulphate (MS). The rat and mouse isoenzymes (rGSTT2-2 and mGSTT2-2 respectively) are also known to detoxify reactive sulphate esters derived from carcinogenic arylmethanols [9,17]. The combined activities of the Theta class GST subfamilies may thus play a critical role in detoxifying many mutagens and carcinogens not metabolized by members of the other GST classes.

Although the gene structures of *mGSTT2* [14] and *rGSTT2* [18] have been reported, nothing has been published about the structure of subfamily 1 Theta class GST genes. The gene structures of both subfamily 1 and 2 Theta class GSTs provide a good starting point for understanding the evolutionary history of this family and of GSTs from other classes, as well as for understanding the molecular mechanisms involved in regulating the expression of these important isoenzymes. In this study, we report the characterization of a cDNA encoding mGSTT1 and its heterologous expression in *Escherichia coli*. In addition we have studied the distribution of *mGSTT1* mRNA expression in mouse tissues, its induction in mouse liver by xenobiotics, and the structure of its corresponding gene, *mGSTT1*. We have also mapped the location of *mGSTT1* and found that it is clustered with *mGSTT2* on chromosome 10 at bands B5–C1.

EXPERIMENTAL

Materials

Analytical-grade reagents used in culture media and buffers were supplied by Difco Laboratories (Detroit, MI, U.S.A.), Sigma Chemical Co. (St. Louis, MO, U.S.A.) and Ajax Chemical Co.

Abbreviations used: GST, glutathione transferase (EC: 2.5.1.18); Ni–NTA, nickel–nitrilotriacetic acid; RT-PCR, reverse transcriptase PCR; EST, expressed sequence tag; BLAST, basic local alignment search tool; BHA, butylated hydroxyanisole; CDNB, 1-chloro-2,4-dinitrobenzene; DCNB, 1,2-dichloro-4-nitrobenzene; EPNP, 1,2-epoxy-3-(*p*-nitrophenoxy)propane; 4-NBC, 4-nitrobenzylchloride; 4-NPA, 4-nitrophenylacetate; EA, ethacrynic acid; G3PDH, glyceraldehyde 3-phosphate dehydrogenase; MS, 1-menaphthyl sulphate; NBD-Cl, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole; DCM, dichloromethane; 6 × His, polyhistidine tag.

[1] To whom correspondence should be addressed (e-mail Philip.Board@anu.edu.au).

(Sydney, Australia). A Lambda FIX™II mouse (129SV) liver genomic library and a Lambda ZAP® mouse (B6/CBA) liver cDNA library were both purchased from Stratagene (La Jolla, CA, U.S.A.). Hybond-N^+ nylon and Hybond-C^+ nitrocellulose membranes and filters, [α-^{33}P]dATP, tritiated deoxynucleotides and a random-primer labelling kit were all purchased from Amersham (Amersham, Bucks., U.K.). The [α-^{32}P]dATP was purchased from Bresatec (Adelaide, Australia). Restriction endonucleases and their buffers were purchased from Boehringer Mannheim (Mannheim, Germany), Pharmacia Biotech (Uppsala, Sweden) or Progen Industries (Dara, Australia). Oligonucleotide primers were synthesized on an Applied Biosystems 394B DNA/RNA synthesizer (Applied Biosystems Inc., Foster City, CA, U.S.A.). Ilford L4 nuclear emulsion was used for autoradiography of *in situ* hybridization slides. Compounds used in enzyme activity assays were purchased from Sigma Chemical Co, Aldrich Chemical Co. (Milwaukee, WI, U.S.A.), Fluka (Buchs, Switzerland), Merck (Darmstadt, Germany) and Ajax Chemical Co. MS was synthesized by the method of Clapp and Young [19]. Glutathione reductase and GSH were obtained from Sigma Chemical Co. NADPH was purchased from Boehringer Mannheim. The QIA*express* system and nickel–nitrilotriacetic acid (Ni–NTA) resin were obtained from QIAGEN (Hilden, Germany).

Screening of mouse liver cDNA and genomic libraries

Library screenings were conducted using the filter hybridization method of Benton and Davis [20] as described elsewhere [14]. A mouse mGSTT1 reverse-transcriptase (RT)-PCR product was amplified using primers designed from rGSTT1 [11]. The forward primer, GAT CTG CTG TCG CAG CCC, matches rGSTT1 exactly and matches rGSTT2 with a single mismatch; the reverse primer, CAG CCT GGG ACG CCC TTC A, matches rGSTT1. One of the liver RT-PCR products amplified shared 92 % deduced amino acid sequence identity with rGSTT1, and less than 60 % identity with rGSTT2 over amino acid residues 14–185. This clone was used to screen the mouse liver cDNA library, producing the cDNA clone pMT1a that was used in further studies. The genomic library has been previously screened using a cDNA encoding rGSTT2, which coincidently isolated a clone (called *λMT1*) with sequence similarity to rGSTT1 and human GSTT1 (hGSTT1) [14]. Liquid-culture phage lysates were used to prepare DNA from *λMT1*, whereas hybridizing cDNA clones were subjected to *in vivo* excision protocols outlined in the library supplier's instructions to rescue pBSK- double-stranded phagemids containing cloned cDNA inserts from the lambda ZAP vector.

Cloning and sequencing

The cDNA clone pMT1a was digested with restriction endonucleases to generate smaller overlapping fragments that were subcloned into pUC vectors and sequenced on both strands by the dideoxy chain-termination method [21] using a Sequenase™ Version 2.0 DNA sequencing kit (Amersham). DNA prepared from the genomic clone *λMT1* was subjected to digestion with various restriction endonucleases (singly or in pairs) and Southern blot analysis [22] using both the RT-PCR product and pMT1a insert as probes. Hybridizing fragments were cloned into pUC vectors for further subcloning and sequencing. In some cases, intron sizes were estimated by PCR using oligonucleotide primers designed from flanking exon sequences. PCR mixtures and conditions were those recommended by the Expand™ Long Template PCR System Manual (Boehringer Mannheim) with modifications allowing for primer annealing temperatures. PCR products were sequenced using a Thermo Sequenase™ cycle sequencing kit (Amersham).

Computer software for DNA analysis

DNA restriction maps and sequence translations were generated using DNA Strider™ Version 1.2 (Commissariat a l'Energie Atomique, France). DNA and protein sequence alignments and comparisons were conducted using the Bestfit program of the GCG software (Version 8.0, 1994, Genetics Computer Group, Madison, WI, U.S.A.) as provided by the Australian National Genome Information Service (ANGIS). Database searches of expressed sequence tags (ESTs; generated by partial sequencing of random cDNA clones and assignment to gene families based on sequence similarities) were conducted using the basic local alignment search tool (BLAST) as provided by the National Centre for Biotechnology Information (NCBI) (located at worldwide website http;//www.ncbi.nlm.nih.gov/BLAST/). Searches of the mouse EST sequence database were performed using the TFASTY program [23].

Tissue distribution and induction of mGSTT1 mRNA expression

Four- to five-week-old female CD-1 mice were obtained from Charles River Breeding Laboratories, Wilmington, MA, U.S.A. and acclimatized for 2 weeks before administration of butylated hydroxyanisole (BHA). BHA-induced mice were fed a powdered diet *ad libitum* that was supplemented with 0.75 % (w/w) BHA. For some experiments, mice were injected intraperitoneally with β-naphthoflavone (125 mg/kg in corn oil), phenobarbital (100 mg/kg in 0.85 % NaCl) or isosafrole (125 mg/kg in corn oil). Animals were injected twice, 48 h apart, and killed 24 h after the second injection [24]. Tissues were isolated and stored as described [24], total RNA was isolated with guanidinium isothiocyanate homogenization and CsCl centrifugation, and mRNA expression was measured by Northern blot hybridization of formamide-denatured RNA on 2.2 M formaldehyde/1.2 % agarose gels. To examine mRNA integrity, membranes were reprobed with mouse glyceraldehyde 3-phosphate dehydrogenase (G3PDH) cDNA prepared using RT-PCR with primers obtained from Clontech Laboratories (Palo Alto, CA, U.S.A.).

Southern blot analysis

Genomic DNA was prepared from strain C57BL female mouse liver, and after digestion to completion with various restriction endonucleases it was electrophoresed on a 1 % agarose gel and transferred to a nylon membrane according to the method of Southern [22]. One sample set was probed with a 700 bp *Bst*YI–*Aat*II fragment derived from pMT1a encoding deduced amino acid residues 7–240 of mGSTT1. The duplicate samples were probed with the mGSTT2 probe previously described [14].

In situ hybridization

Specific details of mouse chromosome preparation and *in situ* hybridization methods have been described [25]. The 700 bp *Bst*YI–*Aat*II fragment described above was labelled with [^{3}H]dATP, [^{3}H]dTTP and [^{3}H]dCTP to a specific radioactivity of 3.4×10^7 c.p.m./μg and hybridized to chromosome preparations from Balb/c and C57BL mice. After autoradiography for 25–28 days, the slides were developed and stained to give G-banded chromosomes, allowing the identification of individual chromosomes. Silver grains in the emulsion were then scored on to idiograms of G-banded mouse chromosomes [26].

Construction of an expression plasmid

A 1083 bp fragment containing the entire coding and 3′ non-coding regions of mGSTT1 was PCR-amplified from pMT1a using a forward primer mT1ExA (5′ ATG CGG ATC CGT TCT GGA GCT GTA CCT G 3′) and a universal T7 primer (5′ GTA ATA CGA CTC ACT ATA GGG C 3′) (Pharmacia Biotech). Primer mT1ExA contains a *Bam*HI site used to replace the initiating methionine residue (ATG) of mGSTT1 and permit in-frame cloning into the pQE30 vector (QIAGEN) with the 6 × histidine (6 × His) residue tag at the N-terminus. PCR was performed using a capillary thermal cycler (Corbett Research, Sydney, Australia) in a 20 μl reaction mix containing 200 ng of template, 200 μM dNTPs, 1.5 mM $MgCl_2$, 10 pmol of each primer, 1 × reaction buffer and 1.5 units of *Taq* DNA polymerase. PCR conditions were as follows: a hot start (95 °C/1 min); 20 cycles of 95 °C/20 s, 56 °C/20 s, 72 °C/40 s; 10 cycles of 95 °C/20 s, 62 °C/20 s, 72 °C/40 s and an extra extension cycle of 72 °C/2 min. To generate an expression plasmid, the amplified product was digested with *Bam*HI and *Hin*dIII, ligated into the pQE30 vector and transformed into *E. coli* M15[pREP4] host cells. Selected recombinants were completely sequenced on both DNA strands to detect any errors introduced into the mGSTT1 coding sequence during the amplification of the cDNA. A clone termed pQEMT1 had no changes to the sequence and was used for further investigations.

Expression of recombinant mGSTT1-1

A culture of pQEMT1 transformed into *E. coli* M15[pREP4] host cells was grown in standard Luria broth supplemented with 100 μg/ml ampicillin and 25 μg/ml kanamycin at 37 °C until an A_{600} of 0.8–0.9 was reached, at which time protein expression was induced by adding isopropyl thio-β-D-galactoside to a final concentration of 0.1 mM. The incubation was continued for 16 h at 37 °C and the bacterial cells were subsequently harvested by centrifugation at 4000 ***g*** for 20 min at 4 °C and stored at −20 °C overnight. After thawing on ice, the cells were resuspended in 30 ml of buffer A (50 mM sodium phosphate/300 mM NaCl, pH 6.0) and completely lysed by passage through a Ribi cell fractionator model RF-1 (Sorvall, Newtown, VA, U.S.A.). The lysate was centrifuged at 10000 ***g*** for 30 min at 4 °C to pellet cellular debris and the supernatant was immediately used for recombinant mGSTT1-1 purification.

Purification of recombinant mGSTT1-1

All purification procedures were conducted at 4 °C. Cleared supernatant was diluted to 50 ml with buffer A and imidazole (pH 6.0) was added to a final concentration of 10 mM. This sample was mixed with 5 ml of a 50% slurry of Ni-NTA resin pre-equilibrated with buffer A and incubated for 1 h on a rotary mixer. The resin was collected by centrifugation and washed twice with 20 vols. of buffer A by centrifugation, followed by a more copious wash on a vacuum scintered glass funnel with 250 ml of buffer A. The resin was packed under gravity into a small column and washed through with a further 10 bed vols. of buffer A. The column was then developed with three concentrations (50, 100 and 500 mM) of imidazole (pH 6.0) in buffer A. Fractions containing recombinant mGSTT1-1 were identified by SDS/PAGE screening [27], pooled and dialysed for 16–20 h against 5 litres of buffer B (10 mM Tris/HCl/1 mM EDTA/0.5 mM 2-mercaptoethanol, pH 8.0). Any minor contaminating proteins were removed by applying samples to a Mono Q® HR 5/5 FPLC column (Pharmacia Biotech) pre-equilibrated with buffer B. This column was developed by elution with buffer B (at a flow rate of 1 ml/min for 20 min) followed by two gradients of 0–150 mM and 150 mM–1 M NaCl in buffer B at a flow rate of 1 ml/min for 20 min and 15 min respectively. Fractions were monitored for recombinant mGSTT1-1 by measuring A_{280} and activity towards dichloromethane (DCM). Fractions that showed activity were pooled and protein concentrations were determined [28].

Enzyme activity assays

All enzyme activity assays were performed at 37 °C. Assay mixtures minus recombinant mGSTT1-1 served as controls. Activities toward CDNB, 1,2-dichloro-4-nitrobenzene (DCNB), EPNP, 4-nitrobenzyl chloride (4-NBC), 4-nitrophenylacetate (4-NPA) and ethacrynic acid (EA) were measured according to the methods described by Habig et al. [29] and Mannervik and Widersten [30]. Activity towards MS was determined by the method of Gillham [31] and glutathione peroxidase activity was measured according to the method of Lawrence and Burk [32]. Assays using *trans*-alk-2-enals and *trans*,*trans*-alka-2,4-dienals have been described [33]. The conjugation of GSH to 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) was measured by the method of Ricci et al. [34]. The activity of recombinant mGSTT1-1 towards DCM was determined by the formation of formaldehyde as follows: 10 mM GSH, 2 μg of isoenzyme, 40 mM DCM (from a 1.6 M stock prepared in ethanol) and 0.1 M potassium phosphate (pH 7.4) to a final volume of 1 ml were incubated for 1 h at 37 °C. The reaction was stopped by adding 200 μl of 10% trichloroacetic acid to precipitate the protein. After centrifugation, the supernatant was analysed for formaldehyde by the method of Nash [35] using known concentrations of formaldehyde as standards.

RESULTS

Characterization of a cDNA encoding mGSTT1

Sequencing of clone pMT1a isolated from the mouse liver cDNA library revealed that it encodes a mouse Theta class GST with high sequence identity to rGSTT1 and hGSTT1. The cDNA, called MT1a, is 1053 bp long and contains 101 bp of 5′ non-coding sequence, an open reading frame of 720 bp and a 3′ non-coding region of 132 bp (Figure 1) The consensus poly(A) addition signal (AATAAA) is found 197 bp after the stop codon followed 21 bp downstream by a poly(A) tail. The 5′ non-coding sequence may be a hybrid of two cDNA species resulting from a cloning artefact since the −101 to −36 bp region of this cDNA is almost identical with unpublished and unrelated EST sequences that encode a protein similar to the human serum albumin precursor. The remaining 5′ non-coding sequence (−35 to −1 bp), however, is highly similar to that in several unpublished Theta class subfamily 1 ESTs isolated from mouse embryonic (GenBank accession number AA048417) and mammary gland (GenBank accession number AA457839) tissues. The most 5′ nucleotide of the 5′ non-coding sequences in these ESTs falls within the −26 to −10 bp region of MT1a.

The nucleotide sequence of MT1a is very similar to another cDNA clone reported by Mainwaring et al. [16] and published during the course of this study. The nucleotide sequences of these two cDNAs differ at six positions, three of which lead to two changes in the deduced amino acid sequence. These differences probably arise from strain variation, since the cDNA isolated by Mainwaring et al. [16] was from strain B6C3F1 compared with strain C57 black/6xCBA used in the present study.

```
                                        ggccagttacagaagtggttcccgccgcctatgcc   -1

ATGGTTCTGGAGCTGTACCTGGATCTGCTGTCGCAGCCCTGTCGCGCCATTTATATCTTC   60
M  V  L  E  L  Y  L  D  L  L  S  Q  P  C  R  A  I  Y  I  F     20

                                                    ><
GCCAAGAAGAACAATATCCCGTTCCAGATGCACACGGTGGAGCTGCGCAAGGGTGAGCAC  120
A  K  K  N  N  I  P  F  Q  M  H  T  V  E  L  R  K  G  E  H     40

CTCAGCGATGCGTTTGCCCGGGTGAACCCCATGAAGAGGGTACCAGCCATGATGGATGGT  180
L  S  D  A  F  A  R  V  N  P  M  K  R  V  P  A  M  M  D  G     60

                ><
GGCTTCACCCTGTGTGAGAGTGTGGCTATCTTGCTCTACCTGGCACACAAGTATAAGGTT  240
G  F  T  L  C  E  S  V  A  I  L  L  Y  L  A  H  K  Y  K  V     80

CCTGACCACTGGTACCCCCAAGACCTGCAGGCTCGTGCTCGTGTAGACGAGTACCTGGCA  300
P  D  H  W  Y  P  Q  D  L  Q  A  R  A  R  V  D  E  Y  L  A    100

                                                    ><
TGGCAGCATACGGGCCTTCGGAGAAGCTGCCTCAGGGCCCTGTGGCATAAGGTGATGTTC  360
W  Q  H  T  G  L  R  R  S  C  L  R  A  L  W  H  K  V  M  F    120

CCTGTTTTCCTTGGTGAGCAAATACCTCCTGAAACACTGGCAGCCACGTTGGCAGAACTG  420
P  V  F  L  G  E  Q  I  P  P  E  T  L  A  A  T  L  A  E  L    140

GATGTTAACCTACAGGTGCTTGAAGACAAGTTCCTCCAGGACAAAGACTTCCTTGTTGGG  480
D  V  N  L  Q  V  L  E  D  K  F  L  Q  D  K  D  F  L  V  G    160

                                                    ><
CCCCACATCTCCCTGGCCGACTTGGTGGCCATCACAGAGCTGATGCATCCTGTAGGTGGT  540
P  H  I  S  L  A  D  L  V  A  I  T  E  L  M  H  P  V  G  G    180

GGCTGCCCAGTCTTTGAAGGGCATCCCAGGCTGGCTGCATGGTACCAGCGAGTGGAGGCA  600
G  C  P  V  F  E  G  H  P  R  L  A  A  W  Y  Q  R  V  E  A    200

GCTGTGGGGAAGGACCTCTTCCGGGAAGCCCATGAAGTCATCCTGAAGGTGAAGGACTGT  660
A  V  G  K  D  L  F  R  E  A  H  E  V  I  L  K  V  K  D  C    220

CCCCCTGCTGACCTCATCATAAAGCAGAAGCTGATGCCCAGAGTGCTGGCAATGATCCAG  720
P  P  A  D  L  I  I  K  Q  K  L  M  P  R  V  L  A  M  I  Q    240

TGAcgtcagacgcttcatccctgcaccagctgcggcagttcacaactcacagttcacaat  780
***
tcacaacttgcattacactggggatgcacctcactcctcacaccatgatggctctcaggc  840
tcagctccatcttcacccttttttctgtgatagccacacatgacctcaactataatcacta  900
ctttcccctgagtctgggtaataaactggggcttgatttgggctttaaaga(n)         952
```

Figure 1 Nucleotide and deduced amino acid sequences of MT1a

Reversed arrowheads show the positions of splice sites determined from the partial *mGSTT1* gene sequence. The stop codon is marked by three asterisks, the poly(A) addition signal is underlined and '$a_{(n)}$' denotes the poly(A) tail. Nucleotides and deduced amino acid residues that differ from the cDNA isolated by Mainwaring et al. [16] are highlighted in bold.

```
MT1a      MVLELYLDLLSQPCRAIYIFAKKNNIPFQMHTVELRKGEHLSDAFARVNP   50
rGSTT1    ..........................................Q...      50
hGSTT1    .G.....V.....D...ELRI.D.I..Q.....Q...              50

mGSTT2    MGLELYLDLLSQPSRAVYIFAKKNGIPFQTRTVDILKGQHMSEQFSQVNC   50
rGSTT2    ..........................L....L....L.........      50
hGSTT2    .....F...V.....LEL....LV...K.KE.L.I.S              50

MT1a      MKRVPAMMDGGFTLCESVAILLYLAHKYKVPDHWYPQDLQARARVDEYLA   100
rGSTT1    ..K....K.....................................       100
hGSTT1    L.K...LK..D...T.....TR......Y.....                 100

mGSTT2    LNKVPVLKDGSFVLTESTAILIYLSSKYQVADHWYPADLQARAQVHEYLG   100
rGSTT2    .K..........................................       100
hGSTT2    .G.L.T....D.I....S.......C...TP.....S......R.....   100

MT1a      WQHTGLRRSCLRALWHKVMFPVFLGEQIPPETLAATLAELDVNLQVLEDK   150
rGSTT1    ....T.....T.............R..M......D....V.....Q      150
hGSTT1    ....T.................GPVS.Q...........T..L....      150

mGSTT2    WHADNIRGTFGVLLWTKVLGPLIGV-QVPQEKVERNRDRMVLVLQQLEDK  149
rGSTT2    .......................-..E.......NS...A..R.....     149
hGSTT2    ....C......IP..VQ........-..E.......TA.DQA..W.....  149

MT1a      FLQDKDFLVGPHISLADLVAITELMHPVGGGCPVFEGHPRLAAWYQRVEA  200
rGSTT1    ................V..................R.......R....      200
hGSTT1    ...N.A..T..................CA..Q....R.K..T.R.....   200

mGSTT2    FLRDRAFLVGQQVTLADLMSLEELMQPVALGYNLFEGRPQLTAWRERVEA  199
rGSTT2    .......IA..............I......C...................   199
hGSTT2    ..G..P..A..........A.............E......R.A...G....  199

MT1a      AVGKDLFREAHEVIL-----KVKDCPPADLIIKQKLMPRVLAMIQ       240
rGSTT1    .......L.....------..R......PV..........T...        240
hGSTT1    ...E...Q.....------..A..F....PT.......W.....R       240

mGSTT2    FLGAELYQEAHSTILSILGQAAKKMLPVPPPEVHASMQLRIARIP       244
rGSTT2    ......C....NP.M.V.......T.......A....M.......       244
hGSTT2    ......C.....I.....E.....T..T.S..AYQA.L.......       244
```

Figure 2 Comparison of the deduced amino acid sequences of mammalian Theta class GSTs

The deduced amino acid sequence of the mouse Theta class GST subunit encoded by MT1a is compared with those of mGSTT2 and other rat and human Theta class GST subunits. The subunits are shown grouped into subfamilies and identical amino acid residues within a subfamily are indicated by dots. Asterisks show amino acid residue differences between the mouse subunits. The conserved serine residue at position 11 is highlighted in bold and conserved G-site residues are shown in the shaded boxes. The alignment is based on that reported by Chelvanayagam et al. [44].

Properties of the encoded protein

The open reading frame of MT1a encodes a subunit of 240 amino acids which is identical in length with rGSTT1 and hGSTT1. The calculated molecular mass of the subunit encoded by MT1a (27356 Da) is very similar to the molecular masses of rGSTT2 (26000–27311 Da; [9,12]) and hGSTT2 (25100–27489 Da; [5,7]), and slightly higher than those estimated for the mGSTT1 [15] and mGSTT2 [17] subunits reported recently. The discrepancies between the calculated and estimated molecular masses of mGSTT1 probably reflect migrational variation of the subunits in SDS/PAGE, however they fall within the expected experimental range.

The deduced amino acid sequence of MT1a shares only 51.1% amino acid identity with that of mGSTT2 and a higher amino acid identity with rGSTT1 (92.5%) and hGSTT1 (80.8%) (Figure 2 and Table 1). These similarities suggest that the murine Theta class is divided into two subfamilies whose members have orthologues in the rat and human. Thus, in accordance with the guidelines of the published consensus nomenclature [36], the subunit encoded by MT1a will be referred to as mGSTT1.

Expression and purification of recombinant mGSTT1-1

Expression of the mGSTT1 cDNA with an N-terminal histidine tag using the QIA*express* system allowed the rapid purification of recombinant mGSTT1-1 using immobilized-metal affinity chromatography. The 6×His tagged recombinant mGSTT1-1 was eluted from the Ni–NTA resin with between 100 and 500 mM imidazole (pH 6.0), along with some very minor contaminating proteins of sizes between 14 and 20 kDa and around 40 kDa. These minor contaminants were removed using a Mono Q column where the contaminants bound to the column and the 6×His tagged protein passed straight through. The purity of recombinant mGSTT1-1 during the different purification stages was analysed by 12% SDS/PAGE and is shown in Figure 3. A 14 mg yield of recombinant mGSTT1-1 was obtained from the 1 litre culture of pQEMT1.

Purified recombinant mGSTT1-1 has a molecular mass of approx. 26 kDa (Figure 3), which agrees with the value of 27356 Da calculated from the cDNA sequence and a previously reported estimate of 25000 Da [15]. In addition, the specific activity of recombinant mGSTT1-1 towards DCM (see below) is very similar to that reported for mGSTT1-1 purified from liver cytosol [15], and the N-terminal deduced amino acid sequence of mGSTT1 is identical with that provided by Mainwaring et al. [15] for mGSTT1-1 purified from tissue. These features confirm that the mGSTT1 cDNA clone used in these studies and the one reported by Mainwaring et al. [16] encodes the mGSTT1-1 isoenzyme.

Table 1 Amino acid sequence identity (%) between mammalian Theta class GSTs

Deduced amino acid sequences of the mouse subunits encoded by MT1a and mGSTT2 are compared with those of subunits rGSTT1 [11], rGSTT2 [12], hGSTT1 [6] and hGSTT2 [7] using the Bestfit program as described in the text.

	Sequence identity (%)					
	Subfamily 1			Subfamily 2		
	MT1a	rGSTT1	hGSTT1	mGSTT2	rGSTT2	hGSTT2
MT1a	100	92.5	80.8	51.1	52.1	50.8
rGSTT1		100	79.6	50.6	51.5	50.2
hGSTT1			100	54.8	55.5	55.0
mGSTT2				100	91.4	77.5
rGSTT2					100	78.3
hGSTT2						100

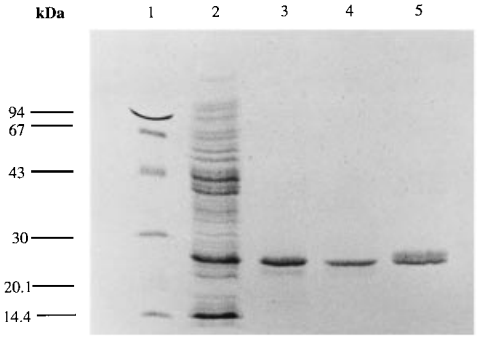

Figure 3 Purification of recombinant mGSTT1-1

The purity of recombinant mGSTT1-1 was analysed by 12% SDS/PAGE. Lane 1, molecular-mass markers; lane 2, crude lysate of bacteria expressing recombinant mGSTT1-1; lane 3, Ni–NTA-purified recombinant mGSTT1-1; lane 4, Mono Q-purified recombinant mGSTT1-1; lane 5, Ni–NTA-purified recombinant hGSTT2-2.

Lability of recombinant mGSTT1-1

The specific activity of recombinant mGSTT1-1 towards some known subfamily 1 Theta class GST substrates (EPNP, DCM, 4-NBC, cumene hydroperoxide [2]) was measured throughout the purification procedures and it was noticed that the specific activity of the Ni–NTA-purified isoenzyme was two to four times higher than that of the purer Mono Q preparation for each substrate. After storage of each preparation at 4 °C for 5 days in buffer B, the specific activity of the Mono Q preparation with these substrates further decreased by 50% or more, whereas that of the Ni–NTA preparation remained stable. Thus it appears that the recombinant isoenzyme became destabilized after passage through the Mono Q column. This is in contrast to N-terminal 6 × His tagged recombinant hGSTT2-2 which appeared to be stable in buffer B after Mono Q purification [37]. Hiratsuka et al. [17] purified stable mGSTT2-2 from mouse liver cytosol but reported their failure to isolate mGSTT1-1 from the same source, noting an unstable nature of this isoenzyme throughout purification. In a more recent report, Mainwaring et al. [15] found that mGSTT1-1 isolated from mouse liver cytosol was unstable unless 0.5 mM GSH and at least 10% (v/v) glycerol was included in buffers for the chromatofocusing and hydrophobic-interaction purification steps.

Table 2 Specific activity of recombinant mGSTT1-1 towards various substrates

Activities are expressed as μmol/min per mg of protein and all values are the means ± S.D. of four measurements. ND, not detectable, –, not measured (due to low activities observed in the Ni–NTA measurements) and * denotes measurement after 13 days storage in buffer B at 4 °C.

	Ni–NTA-purified		Mono Q-purified	
Substrate	Fresh	5 Days	Fresh	5 Days
CDNB	0.03 ± 0.01	–	–	–
DCNB	ND	–	–	–
EA	ND	–	–	–
EPNP	90.6 ± 16.2	97.2 ± 5.8	20.6 ± 4.0	6.8 ± 1.7
NBD-Cl	0.03 ± 0.01	–	–	–
4-NBC	8.2 ± 1.0	9.8 ± 2.3	3.4 ± 0.4	0.2 ± 0.1
4-NPA	0.09 ± 0.01	–	–	–
MS	0.021 ± 0.004	–	–	–
Cumene hydroperoxide	2.9 ± 0.1	1.7 ± 0.3	1.7 ± 0.1	1.0 ± 0.1
t-Butyl hydroperoxide	1.7 ± 0.2	1.4 ± 0.2	0.58 ± 0.02	0.37 ± 0.05
DCM	4.6 ± 0.2	4.9 ± 0.1*	2.7 ± 0.5	0.7 ± 0.2*
Hexa-2,4-dienal	ND	–	–	–
trans-Hex-2-enal	ND	–	–	–
trans-Oct-2-enal	ND	–	–	–
trans-Non-2-enal	ND	–	–	–
trans,*trans*-Hepta-2,4-dienal	ND	–	–	–
trans,*trans*-Nona-2,4-dienal	ND	–	–	–
trans,*trans*-Deca-2,4-dienal	ND	–	–	–

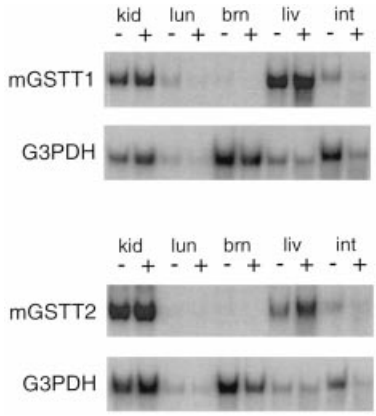

Figure 4 Tissue distribution of *mGSTT1* and *mGSTT2* expression and induction

A 5 μg portion of total cellular RNA from kidney (kid), lung (lun), brain (brn), liver (liv) and intestinal mucosa (int) from animals fed a normal (−) or BHA-supplemented diet (+) was denatured at 65 °C for 15 min in 57% formamide/7% formaldehyde and electrophoresed on a 1.2% agarose gel in 2.2 M formaldehyde and transferred to a nylon membrane. Hybridizations with mGSTT1, mGSTT2 or G3PDH cDNA probes were incubated and washed at 65 °C. G3PDH was used to re-probe the nylon membranes after mGSTT1 and mGSTT2 hybridization.

Specific activities of recombinant mGSTT1-1

Since only a low background of very minor contaminating proteins was present in the apparently stable Ni–NTA preparation, the specific activities measured at this purification stage are considered here to be the most reliable characteristics of recombinant mGSTT1-1. As evident in Table 2, known substrates of the human and rat subfamily 1 Theta class GSTs are also good substrates for the orthologous mouse isoenzyme. The activity of

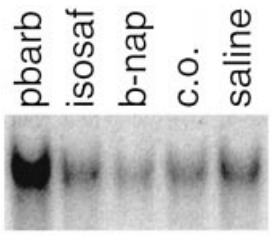

Figure 5 Xenobiotic induction of mGSTT1 mRNAs

RNA samples from livers of mice injected with phenobarbital (pbarb), β-naphthoflavone (b-nap), isosafrole (isosaf), corn oil (c.o.) or saline were denatured, electrophoresed and hybridized with mGSTT1 cDNA clone insert as in Figure 4.

recombinant mGSTT1-1 towards DCM is in general agreement with that reported for mGSTT1-1 [15] and rGSTT1-1 [2] purified from liver cytosol and is considerably higher than that recently reported for recombinant hGSTT1-1 [38]. Recombinant mGST-T1-1 is also active towards EPNP and 4-NBC and showed glutathione peroxidase activity towards cumene hydroperoxide and t-butyl hydroperoxide. Low but detectable levels of activity were observed towards CDNB, NBD-Cl, 4-NPA and MS, whereas no activity was detected towards DCNB, EA, *trans*-alk-2-enals or *trans*,*trans*-alka-2,4-dienals.

Expression of mGSTT1 and mGSTT2 mRNAs

We examined the expression of murine Theta class mRNAs using Northern blot hybridization to total RNA isolated from brain, intestinal mucosa, kidney, liver and lung (Figure 4). *mGSTT1* and *mGSTT2* mRNAs are abundantly expressed in liver and kidney and at lower levels in intestinal mucosa. Some expression is detectable in lung, but expression in brain is quite low. Unlike Mu and Alpha class GST mRNAs [39], Theta class GST mRNA expression is not increased significantly with administration of the dietary antioxidant BHA. We also examined expression of *mGSTT1* mRNA in liver tissue after administration of phenobarbital, β-napthoflavone and isosafrole (Figure 5); a modest (4-fold) induction of *mGSTT1* (and *mGSTT2*, results not shown) can be seen with phenobarbital, but not with either isosafrole or β-napthoflavone.

Structure of the *mGSTT1* gene

In our previous study, a single genomic clone called $\lambda MT1$ was isolated from a genomic DNA library using rGSTT2 as a probe [14]. Southern blot analysis of this clone using rGSTT2 as a probe identified two hybridizing *Eco*RI fragments that were subcloned into pUC vectors to generate the clones pT1G1 (0.9 kb fragment) and pT1G2 (2.2 kb fragment). Clone pT1G1 was found to encode amino acid residues 1–37 of *mGSTT1* flanked by a 5′ non-coding sequence as well as a 3′ acceptor splice site and intronic sequence. The isolation of clone $\lambda MT1$ using an rGSTT2 cDNA probe was surprising, since various reports have suggested that there is no cross-hybridization between Theta class subfamilies [6,7,11,14]. Comparison of the pT1G1 sequence with that of mGSTT1 and to the *mGSTT2* structure [14] indicated that the region of highest similarity represents the coding region of exon 1 of the murine Theta class GST gene, *mGSTT1*. Clone pT1G2 was partially sequenced and found to contain exon 2 (deduced amino acid

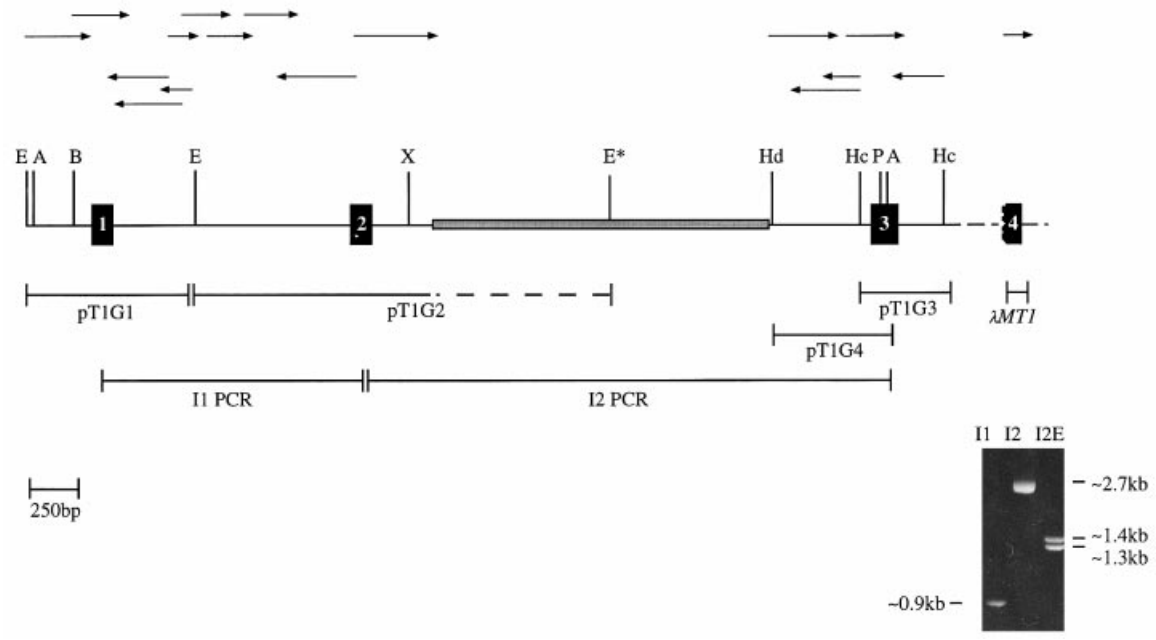

Figure 6 Restriction map of the partial *mGSTT1* gene

Solid boxes represent coding regions (exons numbered 1–4). Solid lines show available intron sequence data while the broken lines indicate unknown intronic sequence in these regions. The grey shaded area represents the remaining region of intron 2 to be sequenced as estimated by PCR analysis. Sites are E = *Eco*RI, A = *Acc*I, B = *Bam*HI, X = *Xba*I, Hd = *Hin*dIII, Hc = *Hin*cII and P = *Pst*I. Arrows above the restriction map depict the sequencing strategy on each DNA strand. Sequences obtained from the clones and PCR products described in the text are indicated below the restriction map. Inset: digestion of the I2 PCR product indicates that it contains an *Eco*RI site (lane I2E inset) that is likely to be the 3′ *Eco*RI site (E*) of clone pT1G2 (the extent of un-sequenced region of this clone is indicated by the broken line).

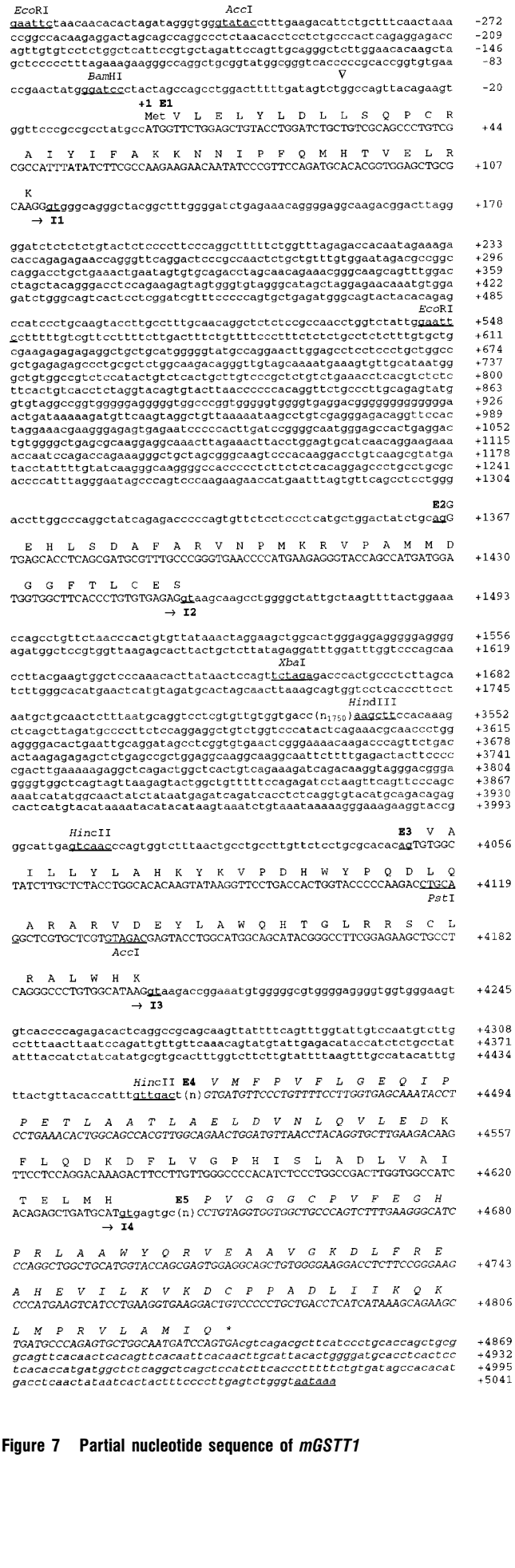

Figure 7 Partial nucleotide sequence of *mGSTT1*

residues 38–67) of *mGSTT1*, with flanking 5′ donor and 3′ acceptor splice sites as well as intronic sequences. PCR-amplification of the intronic sequence between exons 1 and 2 showed that the 0.9 kb and 2.2 kb *Eco*RI fragments are contiguous (Figure 6).

Additional hybridizing fragments were also detected by Southern blot analysis of *λMT1* using the 700 bp *Bst*YI–*Aat*II mGSTT1 probe. Fragments of approx. 0.5 kb (*Hin*cII) and 0.6 kb (*Hin*cII–*Acc*I) were both cloned into pUC vectors, generating clones pT1G3 and pT1G4 respectively. Sequence analysis of pT1G3 showed that it encoded deduced amino acid residues 68–117 (exon 3) of *mGSTT1* flanked by 5′ acceptor and 3′ donor splice sites as well as intronic sequences. Clone pT1G4 overlapped clone pT1G3, encoding deduced amino acid residues 68–94, followed immediately by the *Acc*I cloning site with the 5′ intronic sequence extending approx. 460 bp upstream of that in clone pT1G3 (Figure 6). PCR amplification of the sequence between exons 2 and 3 indicated that intron 2 is approx. 2.7 kb long. This PCR product was only partially sequenced to confirm the intron–exon boundaries, however digestion with *Eco*RI yielded two fragments of similar size (1.3 and 1.4 kb). The *Eco*RI site in this PCR product is likely to be the *Eco*RI cloning site located downstream of exon 2 in clone pT1G2. This places clones pT1G3 and pT1G4 approx. 0.8 kb downstream of this *Eco*RI site (Figure 6).

Further Southern blot analysis of *λMT1*, using a series of C-terminal-region probes generated by restriction endonuclease digestion of mGSTT1, failed to identify any fragments encoding amino acid residues 118–240 in this genomic clone. Direct sequencing of total *λMT1* DNA with an exon 4 forward-sequencing primer (mT1ex4F), however, confirmed that amino acid residues 151–176 of *mGSTT1* are encoded in this clone and that a 3′ acceptor splice site occurs immediately after amino acid residue 176. Unfortunately, further efforts to locate the sequence encoding the remaining amino acid residues 177–240 were unsuccessful. From the available data, it appears that *λMT1* contains a partial clone of *mGSTT1* which ends after exon 4. It is unlikely that this clone represents a pseudogene, since the available coding sequence from exons 1–4 is identical with the corresponding mGSTT1 cDNA, and all confirmed splice sites were conserved when compared with the *mGSTT2* gene structure [14] and with the newly characterized *hGSTT1* gene structure [40]. These features suggest that the gene is functional. Since a 3′ acceptor splice site was identified after exon 4, and since deduced amino acid residues 177–240 of the mGSTT1 cDNA correspond to exon 5 in *hGSTT1* [40], it is expected that *mGSTT1* also comprises five exons. A restriction map of *mGSTT1*, as far as it has been characterized here, is given in Figure 6, and its sequence obtained from the various hybridizing fragments and PCR products is given in Figure 7. Detailed sizes of the exons and introns of *mGSTT1* are presented in Table 3.

Exons (numbered E1–E5) and introns (numbered I1–I4) are indicated by upper- and lower-case letters respectively. All splice sites obey the GT/AG rule and are underlined. Amino acid residues encoded by the exons are given in single letter code above the nucleotide sequence. Those nucleotides and amino acid residues presented in italics are derived from the mGSTT1 cDNA sequence (Figure 1). Nucleotides are numbered from the initiating methionine residue (designated +1). Since approx. 1750 bp (n_{1750}) of the intron 2 sequence was estimated by PCR analysis, nucleotide numbers following this region are estimates. The total lengths of introns 3 and 4 are unknown (n), thus the number of nucleotides following this region are underestimates. The ▽ symbol represents the most 5′ nucleotide of the mGSTT1 cDNA sequence. The stop codon is indicated by an asterisk and the poly(A) addition signal is underlined. Significant restriction endonuclease sites used for subcloning and interpreting Southern blot data are underlined and designated.

Table 3 Exon/intron sizes and encoded amino acid residues of *mGSTT1*

The sizes of all exons were determined by comparing the gene sequence with the cDNA. Intron sizes were estimated by sequence and PCR analyses.

Exon	Size (bp)	Encoded amino acids	Intron	Size (bp)
1	112*	1–37	1	1252
2	88	38–67	2	~ 2596
3	151	68–117	3	> 257
4	176	118–176	4	?
5	398†	177–240		

* Only the coding region of exon 1 is reported since the extent of the 5′ non-coding region of *mGSTT1* is unknown.

† Includes 192 bp of coding sequence plus 206 bp of 3′ non-coding sequence up to and including the poly(A) addition signal.

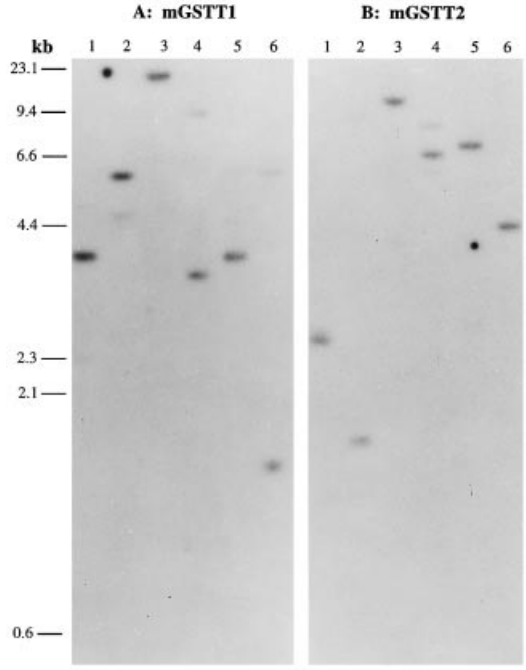

Figure 8 Southern blot of mouse genomic DNA

DNA samples digested with various restriction endonucleases were probed with the coding regions of mGSTT1 (**A**) and mGSTT2 (**B**). Lanes corrrespond to DNA digested with *Acc*I (1), *Hin*dIII (2), *Bam*HI (3), *Eco*RI (4), *Xba*I (5) and *Pst*I (6). Hybridizing fragment sizes were determined by the use of lambda *Hin*dIII DNA markers.

Judging from the available sequence data, the size of *mGSTT1* is expected to be at least 5 kb.

Southern blot analysis

Southern blot analysis of mouse genomic DNA digested with a variety of restriction enzymes and probed with the 700 bp *Bst*YI–*Aat*II fragment from pMT1a showed strongly hybridizing single fragments in all six lanes (Figure 8). Less strongly hybridizing fragments of approx. 2.4 and 4.8 kb (*Acc*I), 4.7 kb (*Hin*dIII), 9.4 kb (*Eco*RI), 4.5 and 6.5 kb (*Xba*I) and 6.5 kb (*Pst*I) were also observed. These blot patterns are consistent with

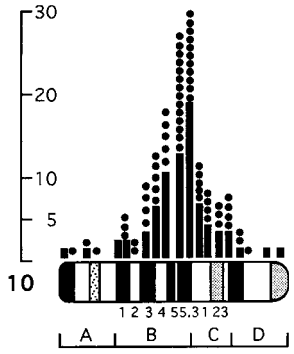

Figure 9 Silver grains over mouse chromosome 10 probed with an mGSTT1 cDNA

Plot of grains over approx. 130 chromosomes 10 showing the probable localization to bands B5–C1, with two very tall peaks suggesting a more precise location to sub-band B5.3 or the proximal half of C1. The solid lines indicate grains scored from Balb/c mice and the dots indicate grains scored from C57BL mice

the mGSTT1 cDNA sequence and the available *mGSTT1* sequence. An additional *Acc*I and *Xba*I site may be located in intron 3 or 4 of *mGSTT1* since three hybridizing fragments were observed in these lanes. Since these data are consistent with the cDNA and genomic sequences of mGSTT1, and since the blot pattern is relatively simple, it seems likely that there is only one gene encoding mGSTT1 in the mouse genome. The hybridizing patterns from this blot suggest that the gene may be as large as 8–10 kb. The obvious difference in the hybridization patterns of the mGSTT1 and mGSTT2 blots indicates that no cross-hybridization occurs between murine Theta class GST sub-families.

In situ hybridization

In situ hybridization with the mGSTT1 probe used for the Southern blot experiment revealed a significant accumulation of silver grains over bands B4–5 and C1 of mouse chromosome 10. A total of 196 grains from approx. 50 labelled chromosome spreads were plotted to all chromosomes and 35.2% of the grains were over chromsosome 10; the background was uniformly distributed (results not shown). In the overall plot, tall peaks of 11, 15 and 18 grains were present over bands B4, B5 and C1 respectively on chromosome 10. Scores from Balb/c and C57BL mice were not apparently different. To confirm the localization of *mGSTT1*, 143 grains over 130 high-quality chromosomes 10 were plotted on to the very accurate idiogram prepared by Evans [26] (Figure 9). This plot shows the three tallest peaks containing 50.3% of the grains over bands B5–C1, confirming that these bands are the probable location of *mGSTT1*. The position of the two very tall peaks of grains strongly suggest a narrower localization of *mGSTT1* to sub-band B5.3 or to the proximal half of band C1.

DISCUSSION

cDNA cloning

The characterization of cDNAs encoding mGSTT1 (present study) and mGSTT2 [14] has confirmed that murine Theta class GSTs may be divided into two subfamilies that have orthologues in the rat and human. Furthermore, comparison of the deduced amino acid sequences of mammalian Theta class GSTs (Figure 2)

confirms that the catalytically important N-terminal serine residue [37,41–43] is conserved among all members of the mammalian Theta class, along with additional N-terminal residues that appear to be involved in the GSH binding site of these GSTs [43,44].

Protein expression

Since Theta class GSTs do not bind to immobilized GSH affinity matrices, lengthy multi-step procedures have been employed to purify these isoenzymes from tissues, usually resulting in low yields and problems with instability. Recently, the expression of recombinant Theta class GSTs using cDNAs has led to a more detailed characterization of the subfamily 2 Theta class GSTs hGSTT2-2 [8,37] and rGSTT2-2 [13]. Comparatively less, however, is known about the substrate specificities of subfamily 1 Theta class GSTs. These latter isoenzymes are of particular interest because their activity towards DCM leads to the production of a reactive metabolite thought to induce liver and lung tumours in mice [15,45] and the risk to humans remains to be fully elucidated [6]. To investigate further the substrate specificity of subfamily 1 Theta class GSTs in general, we have expressed mGSTT1-1 with an N-terminal 6 × His tag which allowed its rapid single-step purification by immobilized-metal affinity chromatography at a yield of approx. 14 mg/l *E. coli* culture.

Recombinant mGSTT1-1 was found to be most active towards EPNP, followed by 4-NBC, although these activities were 2 and 10 times lower respectively than the activity of rGSTT1-1 [2] towards these substrates. The activity of recombinant mGSTT1-1 towards DCM is in good agreement with that reported for mGSTT1-1 purified from liver cytosol [15] and also for rGSTT1-1 [2]. Both rodent isoenzymes are much more active towards these substrates than their human orthologue hGSTT1-1, which was also recently expressed in another laboratory [38]. The specific activities of subfamily 1 Theta class GSTs toward EPNP, 4-NBC and DCM obviously vary between mammalian species; however, they remain indicative of this subfamily, in contrast to subfamily 2 isoenzymes which are inactive towards these substrates [5,8,10,13,17]. Similarly, mGSTT1-1 exhibited only a low activity towards MS and no activity towards EA, typical substrates of subfamily 2 isoenzymes [5,8,10,13,17]. Crystallographic [43] and homology modelling studies [46] suggest that differences in the volume of the second substrate-binding site (H-site) and differences in C-terminal residue interactions between subfamily 1 and 2 isoenzymes may explain the varied substrate specificities of Theta class GSTs.

The close comparison of recombinant and native mGSTT1-1 activity towards DCM suggests that the N-terminal 6 × His tag on the recombinant isoenzyme does not inhibit its activity. Sherratt et al. [38] recently attempted to express recombinant hGSTT1-1 with an N-terminal 6 × His tag, but found that the product was predominantly insoluble, whereas the small amount of soluble material had a low affinity for the Ni–NTA resin. They attributed these features to the 6 × His tag possibly being hidden within the α/β region of the N-terminal domain of hGSTT1-1. In our laboratory, the successful expression of both recombinant mGSTT1-1 (present study) and hGSTT2-2 [37] with an N-terminal 6 × His tag suggests that the N-terminal domain does not occlude the histidine tag from binding the Ni–NTA resin. In fact, crystallographic studies of hGSTT2 [43] and molecular modelling of hGSTT1 [46] suggest that if a C-terminal histidine tag displaced the C-terminal helix, it might be expected to have a greater influence on the active site than an N-terminal tag.

One common substrate of both Theta class subfamilies is cumene hydroperoxide. Human, rat and mouse Theta class GSTs show similar specific activities towards this substrate, with the exception of rGSTT2-2 which exhibits comparatively higher specific activity (Table 2). The potent activity of Theta class GSTs towards cumene hydroperoxide points to an important role of these isoenzymes in protection against oxidative damage. This role is supported by the observations that rGSTT2-2 is active towards a variety of polyunsaturated fatty acid hydroperoxides [10] and recombinant hGSTT2-2 utilizes a range of secondary lipid-peroxidation products [8]. In the present study, recombinant mGSTT1-1 showed activity towards cumene hydroperoxide and t-butyl hydroperoxide, but no activity was detected towards secondary lipid-peroxidation products.

Genomic organization

The characterization of a partial *mGSTT1* gene suggests that, like *mGSTT2* [14] and *rGSTT2* [18], it also comprises five exons interrupted by four introns. All of the intron–exon boundaries of *mGSTT1* were found to be identical with those of *mGSTT2* [14], *rGSTT2* [18] and the newly characterized *hGSTT1* and *hGSTT2* genes [40]. The strong conservation of the mouse, rat and human Theta class GST genes suggests that the gene structure pre-dates the human/rodent divergence. Both murine Theta class GST genes are structurally distinct from reported mouse genes for the other GST classes, and they contain the smallest number of exons even though they encode the longest mammalian GST subunits known to date.

The *mGSTT1* gene is expected to be at least 5 kb long and is larger than *mGSTT2* ($\sim$ 3 kb, [14]). This difference in size may be attributed to the varying intron sizes between the two genes. Although introns 3 and 4 of *mGSTT1* remain to be sized, introns 1 and 2 of *mGSTT1* are clearly three times longer than those in *mGSTT2* and already account for most of the size difference between the two genes. The sizes of *mGSTT1* and *mGSTT2* are comparable with their orthologues in rat (*rGSTT2*, $\sim$ 4 kb; [18]) and human (*hGSTT1*, $\sim$ 8 kb; *hGSTT2*, $\sim$ 3.7 kb; [40]).

Southern blots of mouse genomic DNA digested with restriction endonucleases and probed with mGSTT1 and mGSTT2 suggest that only one gene encodes each of the murine Theta class GSTs in the mouse genome. Similar conclusions using Southern blot analysis have been reached for orthologous Theta class GSTs in rats and humans [6,7,11]. Thus it appears that the Theta class GST family is small compared with the complexity of the Alpha and Mu classes, which comprise multiple genes [47]. Searches of the mouse EST database (ftp://ncbi.nlm.nih.gov/blast/db/est mouse, May 2, 1998) with the mGSTT1 protein sequence, using the TFASTY program [23], found 11 mGSTT1 orthologues with greater than 95% protein sequence identity (including gi|1875927, gi|1528096, gi|1355490, gi|1861683 and gi|1918260). In this database search, mGSTT2 orthologues shared less than 65% identity with the mGSTT1 protein sequence (EST sequence identities can be higher than the 51% identity for the full-length sequence because EST sequences may overlap only a portion of the coding sequence). However, there were several additional EST sequences that shared 86% (gi|3031467) to 91% (gi|1865609) identity with mGSTT1, and less than 60% identity with mGSTT2; whereas another group of EST sequences shared 70% (gi|1701712) to 73% (gi|2403175) with mGSTT1 while sharing less than 55% identity with mGSTT2. Thus there may be additional murine Theta class GST genes, or pseudogenes, which did not cross-hybridize on the Southern blot shown in Figure 8. It is unclear whether these additional genes would cross-hybridize under the *in situ* hybridization conditions used for gene mapping.

Since the Southern blot membranes probed with either

mGSTT1 or mGSTT2 were derived from the one gel, and hence exposed to identical electrophoretic conditions, it is possible to compare directly the hybridization patterns between the two. The hybridization patterns in each Southern blot are notably distinct from each other, indicating that no cross-hybridization between mGSTT1 and mGSTT2 occurs under the stringencies of the experiment, as expected from the 51.1 % deduced amino acid sequence identity between the two subunits (Table 1).

Gene expression

There are few data available concerning the factors regulating the expression of the Theta class GSTs. The results presented here show clearly that there is tissue-specific expression of both mGSTT1 and mGSTT2. Furthermore, it is evident that compounds such as BHA, that are known to induce transcription of mouse Alpha and Mu class GSTs [39], have little effect on the expression of mGSTT1 and mGSTT2. However, phenobarbital, another compound known to induce many detoxication enzymes, had a modest effect on *mGSTT1* and *mGSTT2*. Further work is required to gain a better understanding of the regulation of these enzymes, as their response to potential inducers differs significantly from that of other GSTs.

Gene mapping

The location of *mGSTT1* has been mapped to chromosome 10, probably at bands B5–C1, and more precisely to sub-band B5.3 or the proximal half of C1; the same position as *mGSTT2* [14]. The absence of any significant hybridization over any other chromosome indicates that it is unlikely that there are well-conserved reverse-transcribed pseudogenes derived from *mGSTT1* dispersed throughout the mouse genome. However, a *hGSTT2* pseudogene has been found in tandem with *hGSTT2* on chromosome 22q11.2 [40]. This pseudogene escaped detection by *in situ* hybridization experiments [7] because of its close proximity to *hGSTT2*. The strong signal over chromosome 10 suggests that any additional murine Theta class GST genes or pseudogenes would be located near *mGSTT1* and *mGSTT2*.

Previous work on the human and rodent Alpha, Mu and Pi class GST genes has shown that they are grouped in class-specific clusters on distinct chromosomes [47]. Furthermore, both human Theta class GSTs are clustered on chromosome 22q11.2 [7,48]. The results of this study and our previous study [14] are consistent with this clustering pattern, as both the murine Theta class GST genes are located on chromosome 10 at bands B5–C1.

The significant difference in deduced amino acid sequence between mGSTT1 and mGSTT2, as well as the existence of orthologous gene pairs in rats and humans, suggests that the duplication event occurred before the divergence of primates and rodents. Although the loci of *mGSTT1* and *mGSTT2* are closely associated, the absence of cross-hybridization in addition to distinct amino acid sequences between the two suggest that, like the human Theta class GSTs genes [40], they have not been homogenized by gene-conversion events. In contrast, gene conversion appears to have been a major evolutionary factor within the human Mu class gene cluster at 1p13.3 [49,50].

Synteny between mammalian Theta class GST loci

Mapping of the human, rat and mouse Alpha, Mu and Pi class GST genes has shown that the chromosomal locations of these genes are conserved within mammals [47]. Along with *hGSTT1* [48] and *hGSTT2* [7], there are two other human genes that map to 22q11.2, i.e. breakpoint cluster region homologue (*Bcr*; [51]) and guanine nucleotide binding protein, alpha z subunit (*Gnaz*; [52]). Both of these genes have orthologues that map by linkage to mouse chromosome 10 at 34.5 cM [53]. The *mGSTT1* and *mGSTT2* genes may thus be in the same syntenic segment as these two genes. The mouse zinc-finger protein, autosomal gene (*Zfa*), which maps to 24.5 cM, has been localized by *in situ* hybridization to band B [54], and the transformed mouse 3T3 cell double minute-1 and 2 genes (*Mdm1* and *Mdm2*), which map to 64 cM, have been physically localized to band C [55]. It may be concluded that *mGSTT1* and *mGSTT2* almost certainly fall between 24.5 and 64 cM along with *Bcr* and *Gnaz*. Mapping of the murine Theta class GST genes thus demonstrates that the syntenic group which involves the enzyme-encoding loci *mGSTT1*, *mGSTT2*, *Bcr* and *Gnaz* is conserved between mice and humans.

We thank Ms. M. Coggan and Mr. R. Marano for expert technical assistance. A. T. W. is the recipient of an Australian Postgraduate Research Award. The award of a part-time Senior Research Fellowship by the Research Foundation of The Queen Elizabeth Hospital is gratefully acknowledged by G.C.W. A portion of this work was supported by a grant from the American Cancer Society and the National Institutes of Health, National Library of Medicine (W.R.P.).

REFERENCES

1 Mannervik, B., Ålin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jörnvall, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7202–7206
2 Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991) Biochem. J. **274**, 409–414
3 Meyer, D. J. and Thomas, M. (1995) Biochem. J. **311**, 739–742
4 Board, P. G., Baker, R. T., Chelvanayagam, G. and Jermiin, L. S. (1997) Biochem. J. **328**, 929–935
5 Hussey, A. J. and Hayes, J. D. (1992) Biochem. J. **286**, 929–935
6 Pemble, S., Schroeder, K. R., Spencer, S. R., Meyer, D. J., Hallier, E., Bolt, H. M., Ketterer, B. and Taylor, J. B. (1994) Biochem. J. **300**, 271–276
7 Tan, K.-L., Webb, G. C., Baker, R. T. and Board, P. G. (1995) Genomics **25**, 381–387
8 Tan, K.-L. and Board, P. G. (1996) Biochem. J. **315**, 727–732
9 Hiratsuka, A., Sebata, N., Kawashima, K., Okuda, H., Ogura, K., Watabe, T., Satoh, K., Hatayama, I., Tsuchida, S., Ishikawa, T. and Sato, K. (1990) J. Biol. Chem. **265**, 11973–11981
10 Hiratsuka, A., Okada, T., Nishiyama, T., Fujikawa, M., Ogura, K., Okuda, H., Watabe, T. and Watabe, T. (1994) Biochem. Biophys. Res. Commun. **202**, 278–284
11 Pemble, S. E. and Taylor, J. B. (1992) Biochem. J. **287**, 957–963
12 Ogura, K., Nishiyama, T., Okada, T., Kajita, J., Narihata, H., Watabe, T., Hiratsuka, A. and Watabe, T. (1991) Biochem. Biophys. Res. Commun. **181**, 1294–1300
13 Jemth, P., Stenberg, G., Chaga, G. and Mannervik, B. (1996) Biochem. J. **316**, 131–136
14 Whittington, A. T., Webb, G. C., Baker, R. T. and Board, P. G. (1996) Genomics **33**, 105–111
15 Mainwaring, G. W., Nash, J., Davidson, M. and Green, T. (1996) Biochem. J. **314**, 445–448
16 Mainwaring, G. W., Williams, S. M., Foster, J. R., Tugwood, J. and Green, T. (1996) Biochem. J. **318**, 297–303
17 Hiratsuka, A., Kanazawa, M., Nishiyama, T., Okuda, H., Ogura, K. and Watabe, T. (1995) Biochem. Biophys. Res. Commun. **212**, 743–750
18 Ogura, K., Nishiyama, T., Hiratsuka, A., Watabe, T. and Watabe, T. (1994) Biochem. Biophys. Res. Commun. **205**, 1250–1256
19 Clapp, J. J. and Young, L. (1970) Biochem. J. **118**, 765–771
20 Benton, W. D. and Davis, P. W. (1977) Science **196**, 180–182
21 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
22 Southern, E. M. (1975) J. Mol. Biol. **98**, 503–517
23 Pearson, W. R., Wood, T., Zhang, Z. and Miller, W. (1997) Genomics **46**, 24–36
24 Sisk, S. C. and Pearson, W. R. (1993) Pharmacogenetics **3**, 167–181
25 Webb, G. C., Lee, J. S., Campbell, H. D. and Young, I. D. (1989) Cytogenet. Cell Genet. **50**, 107–110
26 Evans, E. P. (1989) Standard Idiogram. In: Genetic Variants and Strains of the Laboratory Mouse, 2nd Edn (Lyon, M. F. and Searle, A. G., eds), pp. 576–577, Oxford University Press, Oxford
27 Laemmli, U.K. (1970) Nature (London) **227**, 680–685
28 Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) J. Biol. Chem. **193**, 265–275
29 Habig, W. H., Pabst, M. J. and Jakoby, W. B. (1974) J. Biol. Chem. **249**, 7130–7139

30 Mannervik, B. and Widersten, M. (1995) In: Advances in Drug Metabolism in Man (Pacifici, G. M. and Fracchia, G. N., eds.), pp. 409–459. European Commission, Luxembourg
31 Gillham, B. (1973) Biochem. J. **135**, 797–804
32 Lawrence, R. A. and Burk, R. F. (1976) Biochem. Biophys. Res. Commun. **71**, 952–958
33 Brophy, P. M., Southan, C. and Barrett, J. (1989) Biochem. J. **262**, 939–946
34 Ricci, G., Caccuri, A. M., Lo Bello, M., Pastore, A., Piemonte, F. and Federici, G. (1994) Anal. Biochem. **218**, 463–465
35 Nash, T. (1953) Biochem. J. **55**, 416–421
36 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. et al., (1992) Biochem. J. **282**, 305–306
37 Tan, K.-L., Chelvanayagam, G., Parker, M. W. and Board, P. G. (1996) Biochem. J. **319**, 315–321
38 Sherratt, P. J., Pulford, D. J., Harrison, D. J., Green, T. and Hayes, J. D. (1997) Biochem. J. **326**, 837–846
39 Pearson, W. R., Reinhart, J., Sisk, S. C., Anderson, K. S. and Adler, P. N. (1988) J. Biol. Chem. **263**, 13324–13332
40 Coggan, M., Whitbread, L. A., Whittington, A. T. and Board, P. G. (1998) Biochem. J. **334**, 617–623
41 Wilce, M. C. J., Board, P. G., Feil, S. C. and Parker, M. W. (1995) EMBO J. **14**, 2133–2143
42 Board, P. G., Coggan, M., Wilce, M. C. J. and Parker, M. W. (1995) Biochem. J. **311**, 247–250
43 Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G. and Parker, M. W. (1998) Structure **6**, 309–322
44 Chelvanayagam, G., Wilce, M. C. J., Parker, M. W., Tan, K. L. and Board, P. G. (1997) Proteins: Struct. Funct. Genet. **27**, 118–130
45 Ahmed, A. E. and Anders, M. W. (1976) Drug Metab. Dispos. **4**, 357–361
46 Flanagan, J. U., Rossjohn, J. Parker, M. W., Board, P. G.and Chelvanayagam, G. (1998) Proteins: Struct. Funct. Genet., in the press
47 Hayes, J. D. and Pulford, D. J. (1995) CRC Crit. Rev. Biochem. Mol. Biol. **30**, 445–600
48 Webb, G., Vaska, V., Coggan, M. and Board, P. (1996) Genomics **33**, 121–123
49 Taylor, J. B., Oliver, J., Sherrington, R. and Pemble, S. E. (1991) Biochem. J. **274**, 587–593
50 Pearson, W. R., Vorachek, W. R., Xu, S.-J., Berger, R., Hart, I., Vannais, D. and Patterson, D. (1993) Am. J. Hum. Genet. **53**, 220–233
51 Justice, M. J., Siracuso, L. D., Gilbert, D. J., Heisterkamp, N., Groffen, J., Chada, K., Silan, C. M., Copeland, N. G. and Jenkins, N. A. (1990) Genetics **125**, 855–866
52 Wilkie, T. M., Gilbert, D. J., Olsen, A. S., Chen, X. N., Amatruda, T. T., Korenberg, J. R., Trask, B. J., deJong, P., Reed, R. R., Simon, M. I. et al. (1992) Nature Genet. **1**, 85–91
53 Taylor, B. A., Burmeister, M. and Bryda, E. C. (1994) Mamm. Genome **5**, S154–S163
54 Page, D. C., Disteche, C. M., Simpson, E. M., de la Chapelle, A., Andersson, M., Alitalo, T., Brown, L. G., Green, P. and Akots, G. (1990) Genomics **7**, 37–46
55 Cahilly-Snyder, L., Yang-Feng, T., Francke, U. and George, D. L. (1987) Somatic Cell Mol. Genet. **13**, 235–244

Received 6 July 1998/22 September 1998; accepted 13 October 1998