Zeta, a novel class of glutathione transferases in a range of species from plants to humans

Philip G. BOARD[1], Rohan T. BAKER, Gareth CHELVANAYAGAM and Lars S. JERMIIN

Molecular Genetics Group, John Curtin School of Medical Research, Australian National University, GPO Box 34, Canberra, ACT 2601, Australia

Sequence alignment and phylogenetic analysis has identified a new subgroup of glutathione S-transferase (GST)-like proteins from a range of species extending from plants to humans. This group has been termed the Zeta class. An atomic model of the N-terminal domain suggests that the members of the Zeta class have a similar structure to that of other GSTs, binding glutathione in a similar orientation in the G site. Recombinant human GSTZ1-1 has been expressed in *Escherichia coli* and characterized. The protein is a dimer composed of 24.2 kDa subunits and has minimal glutathione-conjugating activity with ethacrynic acid and 7-chloro-4-nitrobenz-2-oxa-1,3-diazole. Although low in comparison with other GSTs, GSTZ1-1 has glutathione peroxidase activity with t-butyl and cumene hydroperoxides. The members of the Zeta class have been conserved over a long evolutionary period, suggesting that they might have a role in the metabolism of a compound that is common in many living cells.

INTRODUCTION

The glutathione S-transferases (GSTs) are found ubiquitously in aerobic organisms and catalyse the conjugation of glutathione to a wide variety of electrophilic substrates [1]. Some members of the GST superfamily also manifest glutathione peroxidase activity and contribute to the metabolism of organic hydroperoxides generated in cells. Characteristically, most species express multiple forms of GST, each with both specific and overlapping substrate preferences [2,3]. The GSTs are known to metabolize a number of environmentally derived carcinogens, and inherited deficiencies have been associated with an increased risk of a variety of cancers. In contrast, overexpression of GSTs in tumour cells can contribute to drug resistance (reviewed in [1]).

The first genetic studies of the GSTs in humans [4] demonstrated the existence of three genetically distinct groups of cytosolic GST isoenzymes. Further studies by Mannervik et al. [5] demonstrated that these groups were found in other mammals and represented three evolutionary classes, termed Alpha, Mu and Pi. Subsequent studies have identified an additional class of cytosolic GSTs termed Theta [6,7]. The Theta class GSTs are characterized by the presence of an active-site serine residue, which contrasts with an active-site tyrosine residue that is conserved throughout the Alpha, Mu and Pi class GSTs [8–11]. More recently, a soluble mitochondrial specific GST termed Kappa has been described [12], although little is known of the structure, function or evolution of this class. A membrane-bound microsomal GST has been well characterized and seems to be structurally and genetically distinct from the cytosolic enzymes [13].

Alignment of mammalian cytosolic GST amino acid sequences shows that members of the Theta class seem to be the closest to the GSTs found in many less advanced species, leading to the suggestion that this class might represent the ancestral class [14]. Alignments of the non-mammalian GST sequences suggest that there might be additional classes that are as distinct from each other as they are from the well-established mammalian Alpha, Mu, Pi and Theta classes. Clearly, the Sigma class, which is well represented in cephalopods, the proposed Phi class in plants and the proposed D (Delta) class in insects are examples of the additional diversity of the GSTs [15–17].

While undertaking sequence alignments to study the evolution of the GST superfamily we detected an unusual group of GST-like proteins that consisted of representatives from a highly diverged range of species extending from plants to humans. To confirm that this group of proteins are GSTs, we have sequenced a human cDNA clone and expressed the recombinant protein in *Escherichia coli*. Because the protein exhibits glutathione-conjugating activity and has several other features in common with the GST superfamily, it seems to define a new evolutionary class of GSTs that, in contrast with the previously characterized classes, occurs in a range of phylogenetic groups. We have termed the new class Zeta, in accordance with the established human GST nomenclature convention [18].

EXPERIMENTAL

Plasmids and DNA sequencing

A human expressed sequence tag (EST) that had sequence similarity to a putative carnation (*Dianthus caryophyllus*) GST [19] was obtained from the I.M.A.G.E. Consortium and the A.T.C.C. This clone (I.M.A.G.E. consortium ID 265310, GenBank accession number N31040) was originally isolated from a melanocyte cDNA library. The complete DNA sequence was obtained from each strand by subcloning restriction fragments and by the use of a Thermo Sequenase Kit (Amersham).

Heterologous expression and purification of GSTZ1

GSTZ1-1 was expressed in *E. coli* as a ubiquitin fusion protein as previously described for other GSTs [20,21]. The ubiquitin moiety

Abbreviations used: EST, expressed sequence tag; GST, glutathione S-transferase.

[1] To whom correspondence should be addressed.

The nucleotide sequence results reported here will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number U86529.

was co-translationally cleaved by the yeast ubiquitin-specific protease Ubp1 expressed on a separate plasmid. The expressed GSTZ1-1 was then purified by affinity chromatography on glutathione–agarose (Sigma) [22].

Enzyme assays

GST activity was measured spectrophotometrically with a range of electrophilic substrates under the following previously described conditions: 1-chloro-2,4-dinitrobenzene [23], ethacrynic acid [23], *p*-nitrophenyl acetate [23], *p*-nitrobenzyl chloride [23], *trans*,*trans*-alka-2,4-dienals and *trans*-alk-2-enols [24], 7-chloro-4-nitrobenz-2-oxa-1,3-diazole [25], dichloromethane [21,26] and menaphthyl sulphate [27]. Glutathione peroxidase activity was determined with cumene hydroperoxide and t-butyl hydroperoxide as substrates [28].

Sequence alignment and tree building

Selected GST nucleotide sequences, chosen to represent the different GST classes, were retrieved from GenBank or from the authors, and translated into protein sequences. The *Arabidopsis* and *Petunia* sequences were classified as members of the previously proposed plant Phi class [16]. The insect sequences, which have been previously referred to as Theta, Theta-like, or D-class, are here termed Delta in keeping with the established nomenclature conventions. This subdivision is clearly supported by the separation shown below in Figure 3. The published protein sequence of human Kappa [12] did not correspond to DNA sequences published in GenBank's EST database, so we chose to generate a consensus of sequences retrieved from the EST database: the consensus sequence used in this study is based on positions 1–372 of R83924, positions 103–202 of R78029, positions 158–379 of AA132173 and positions 381–442 of AA132173.

A multiple sequence alignment (see Figure 2) was produced by using CLUSTAL W [29] and the result was subsequently adjusted manually, making use of known structural alignments [30–33]. A phylogenetic study of the data was then conducted by maximum-likelihood analysis. The alignment in Figure 2 was analysed with the JTT-F substitution model [34], and the local bootstrap values were determined for internal edges, both by means of ProtML [35]. An underlying assumption of this phylogenetic analysis is that the sequences are not biased in terms of base-compositional heterogeneity. This assumption was not assessed, because tools for this analysis are still under development (L. S. Jermiin, unpublished work). Therefore the tree in Figure 3, like many similar trees published so far, might be the result of both compositional and historical components. The tree in Figure 3 is an unrooted tree with the branch lengths drawn to scale: the local bootstrap values are superimposed above the internal edges. The definition of the nine GST classes is well supported, as indicated by the high local bootstrap values of class-defining edges (97.4–100 %).

Molecular modelling

A computer model of the N-terminal domain of GSTZ1 was constructed on the basis of the co-ordinates of the *Lucilia cuprina* GST structure [30], aligning Pro-6 of GSTZ1 with Met-1 of the insect enzyme and allowing Arg-42 and Gly-43 to form an insertion. This alignment yields approx. 20 % sequence identity over the first 75 amino acid positions, effectively the N-terminal domain. The model was built by using the HOMOLOGY module of the Insight Package (Biosym Technologies, San Diego, CA, U.S.A.), following the alignment given in Figure 2. No refinement of the model was attempted apart from a few manual adjustments to the side chains of some residues. The relative insertion in the loop between β strand 2 and helix 2 is only roughly built and was formed by generating several loop structures and selecting one that had a low root-mean-square deviation at the splice sites as well as complementing the rest of the model.

RESULTS

DNA and protein sequence

The cDNA contained in the EST clone N31040, which we have termed pGSTZ1, has been completely sequenced on both strands (Figure 1) and has 103 bp of 5′ non-coding sequence, 648 bp of coding sequence and 404 bp of 3′ non-coding sequence. The cDNA encodes a peptide of 216 residues that has a predicted molecular mass of 24166 Da. An alignment of the deduced amino acid sequence of GSTZ1 with representative sequences from previously described GST classes is shown in Figure 2 and indicates the significant differences between the GSTZ1 sequence and the previously defined classes in mammals. Among the human GSTs, the Theta class enzyme seems to be the most similar to GSTZ1, yet GSTT2 has less than 20 % identity.

We originally identified the GSTZ1 EST because of its similarity to the putative carnation (*D. caryophyllus*) petal GST [19]. Database searches have subsequently identified additional GST genes with a high degree of similarity from *Caenorhabditis elegans* (Z66560) and mouse (D21579). Although the complete sequence for the mouse cDNA is not available, the complete protein-coding sequences from carnation and *C. elegans* have been extracted from databases and are compared with the deduced amino acid sequence of GSTZ1 in Figure 2. We also used the carnation sequence to search the yeast *Saccharomyces cerevisiae* genome (yeast database; http://genome−www.stanford.edu/saccharomyces/) and identified a putative open reading frame on chromosome XII (YLL060C; accession nos. Z73165, Y13138) as a possibly related GST. The alignment

```
                    AAGACACGGGCCTGATTCGTCGAGTCTCACTGAGCCTTAGTCG -103
TCGGCAGGTCCCAGGCGCGAAGTTTCTCGGCCTGGAGGAGGGGGTCGCGCGAAGTGCCAG  -60
ATGCAGGCGGGGAAGCCCATCCTCTATTCCTATTTCCGAAGCTCCTGCTCATGGAGAGTT   60
M  Q  A  G  K  P  I  L  Y  S  Y  F  R  S  S  C  S  W  R  V     20

CGAATTGCTCTGGCCTTGAAAGGCATCGACTACAAGACGGTGCCCATCAATCTCATAAAG  120
R  I  A  L  A  L  K  G  I  D  Y  K  T  V  P  I  N  L  I  K     40

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

ACCCTGAAGATTGATGGAATCACCATTCACCAGTCACTGGCCATCATTGAGTATCTAGAG  240
T  L  K  I  D  G  I  T  I  H  Q  S  L  A  I  I  E  Y  L  E     80

GAGACGCGTCCCACTCCGCGACTTCTGCCTCAGGACCCAAAGAAGAGGGCCAGCGTGCGT  300
E  T  R  P  T  P  R  L  L  P  Q  D  P  K  K  R  A  S  V  R    100

ATGATTTCTGACCTCATCGCTGGTGGCATCCAGCCCCTGCAGAACCTGTCTGTCCTGAAG  360
M  I  S  D  L  I  A  G  G  I  Q  P  L  Q  N  L  S  V  L  K    120

CAAGTGGGAGAGGAGATGCAGCTGACCTGGGCCCAGAACGCCATCACTTGTGGCTTTAAC  420
Q  V  G  E  E  M  Q  L  T  W  A  Q  N  A  I  T  C  G  F  N    140

GCCCTGGAGCAGATCCTACAGAGCACAGCGGGCATATACTGTGTAGGAGACGAGGTGACC  480
A  L  E  Q  I  L  Q  S  T  A  G  I  Y  C  V  G  D  E  V  T    160

ATGGCTGATCTGTGCTTGGTGCCTCAGGTGGCAAATGCTGAAAGATTCAAGGTGGATCTC  540
M  A  D  L  C  L  V  P  Q  V  A  N  A  E  R  F  K  V  D  L    180

ACCCCCTACCCTACCATCAGCTCCATCAACAAGAGGCTGCTGGTCTTGGAGGCCTTCCAG  600
T  P  Y  P  T  I  S  S  I  N  K  R  L  L  V  L  E  A  F  Q    200

GTGTCTCACCCCTGCCGGCAGCCAGATACACCCACTGAGCTGAGGGCCTAGCTCCCAAAT  660
V  S  H  P  C  R  Q  P  D  T  P  T  E  L  R  A               216

CCTGCCCCGTTGGCACAGGGCCACAGGAGCAGAAGCTGGGTGGGCTGAAGAGGCCTGGAA  720
ACGAGAGTCTTAATTGAGGAGATGGGAGACTCGAACTCTAGCCCTGGATCTGCCTTCCTG  780
CTGAAACTTGTTCCACCTCAGTCCCCTCATCTGTCACACGCATGTGGGGTGGAGTAGGGA  840
GATGCGGGGAGCAGGGTGGGCAGGAATACTGTTATCTATGTGACGGGGCAGTCGTGAGGC  900
TGAGATGAGAATGCGGATTAAAATGCCTGGCGTGCTCACCGTAACACCACGGGGAAGGCT  960
GTGTGCCTTTTCTCATCCGCTTTTGTTGTGTGACTCCAAAGAATGCCCGCGCTGAAAT 1020
TTGGCGTGAATTAAACTGAAGCCCAGGCCTCT                            1052
```

Figure 1 Nucleotide and deduced amino acid sequences of GSTZ1

The poly(A) addition signal is underlined.

Figure 2 Structurally based multiple sequence alignment of representative GSTs from the major evolutionary classes

The secondary-structural elements (H, helix; E, β strand) are based on the *Lucilia cuprina* GST structure [30]. The sequences have the following Genbank accession numbers: Human_Zeta (*Homo sapiens*, U86529); Caenorhabditis_Zeta (*Caenorhabditis elegans*, Z66560); Carnation_Zeta (*Dianthus caryophyllus*, M64268); Arabidopsis_Phi (*Arabidopsis thaliana*, D17672); Petunia_Phi (*Petunia hybrida*, Y07721); Human_Theta (*Homo sapiens*, L38503); Mouse_Theta (*Mus musculus*, U48420); Musca_Delta (*Musca domestica*, M83249); Drosophila_Delta (*Drosophila melanogaster*, X14233); Human_Alpha (*Homo sapiens*, M15872); Mouse_Alpha (*Mus musculus*, M73483); Chicken_Alpha (*Gallus gallus*, L15386); Human_Pi (*Homo sapiens*, X15480); Rat_Pi (*Rattus norvegicus*, X02904); Human_Mu (*Homo sapiens*, M63509); Mouse_Mu (*Mus musculus*, J04632); Chicken_Mu (*Gallus gallus*, X58248); Squid1_Sigma (*Ommastrephes sloanei*, M36937; Squid2_Sigma (*Ommastrephes sloanei*, M36938). The yeast *Saccharomyces cerevisiae* sequence YLL060C was obtained from the yeast database (accession nos. Z73165, X13138), the rat Kappa class sequences were obtained from the authors [12] and the human Kappa was determined from EST sequences (see the text).

shown in Figure 2 was used to generate a maximum-likelihood tree (Figure 3). This tree clearly groups GSTZ1 with the sequences from carnation and *C. elegans*. Interestingly, the yeast sequence is grouped separately, with the *Arabidopsis* and petunia GSTs, rather than with the group containing GSTZ1; the alternative scenario with the yeast sequence grouping together with the GSTZ1, *C. elegans* and carnation sequences was supported by only 38.1 % in local bootstrap value. It is clear from Figure 3 that the carnation, *C. elegans* and human GSTZ1 sequences form a distinct cluster, and consequently we have termed this group the Zeta class.

Enzyme activity

GSTZ1 was initially identified as a GST because of its similarity to a carnation petal GST. However, examination of the data

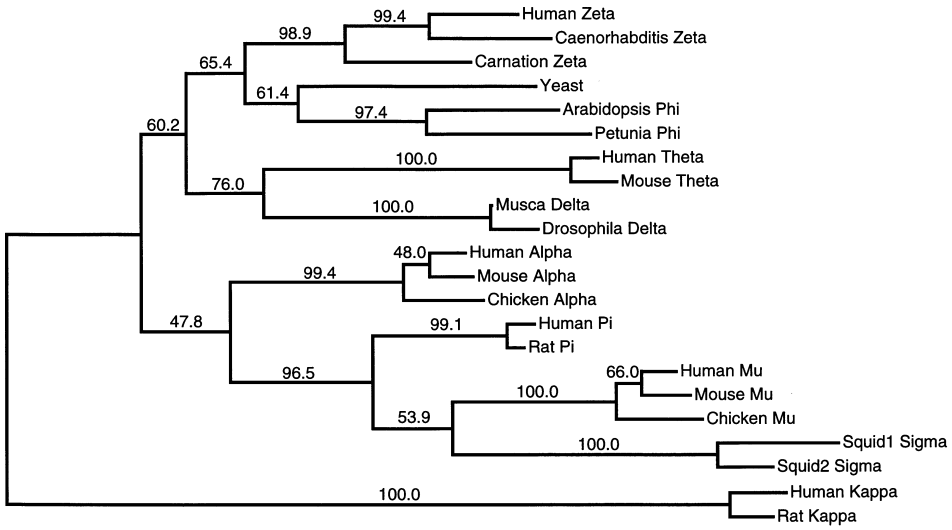

Figure 3 An unrooted maximum-likelihood tree based on the protein alignment in Figure 2

The tree, with horizontal branch lengths drawn to scale and local bootstrap values, shows the phylogenetic relationships between the selected GST sequences.

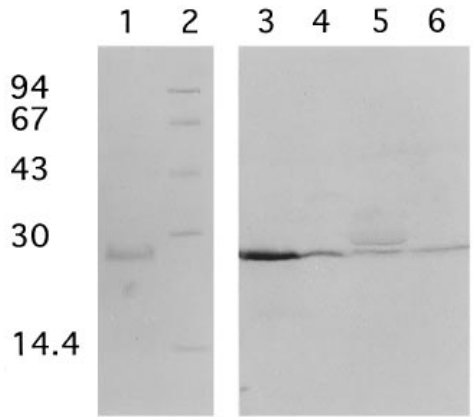

Figure 4 SDS/PAGE of GSTZ1

Lanes 1 and 2 were stained with Coomassie Blue: lane 1, purified recombinant GSTZ1 expressed in *E. coli*; lane 2, molecular size markers (shown in kDa at the left). Lanes 3–6 were blotted on nitrocellulose and the GSTZ1 was detected with a rabbit antiserum to the recombinant GSTZ1-1: lane 3, liver; lane 4, kidney; lane 5, skeletal muscle; lane 6, brain. Equivalent amounts of total protein were added to lanes 3–6.

Table 1 Activity of recombinant human GSTZ1-1 with various substrates

Results are means ± S.D. for at least three determinations. Abbreviation: n.d., not detectable.

Substrate	Specific activity (nmol/min per mg of protein)
Cumene hydroperoxide	160 ± 11
t-Butyl hydroperoxide	170 ± 4
1-Chloro-2,4-dinitrobenzene	n.d.
1,2-Dichloronitrobenzene	n.d.
4-Phenylbut-3-en-2-one	n.d.
Ethacrynic acid	45 ± 15
7-Chloro-4-nitrobenz-2-oxa-1,3-diazole	32 ± 5
1-Menaphthyl sulphate	n.d.
Dichloromethane	n.d.
p-Nitrobenzyl chloride	n.d.
p-Nitrophenyl acetate	14 ± 6
trans-Non-2-enal	n.d.
trans-Oct-2-enal	n.d.
Hexa-2,4-dienal	n.d.
trans,trans-Deca-2,4-dienal	n.d.

relating to the carnation GST indicated that its identification as a GST was only because of sequence similarity with a rat Alpha class isoenzyme [19,36]. The *C. elegans* sequence was also identified by similarity and has not been studied in detail. As it has not previously been demonstrated that any of the proteins now assigned to the Zeta class have GST activity, we therefore expressed recombinant GSTZ1-1 protein in *E. coli* and characterized its properties. The expressed enzyme was purified from the *E. coli* lysate to apparent homogeneity (Figure 4) by affinity chromatography on glutathione–agarose. Compared with enzymes from the Alpha, Mu and Pi classes, GSTZ1-1 had a relatively weak affinity for this matrix and needed to be incubated with the beads for several hours to ensure adequate binding. The enzyme could be eluted either by raising the pH to 9.6 or by the addition of 5 mM glutathione. Purified GSTZ1-1 had an electrophoretic mobility on SDS/PAGE that was compatible with its predicted 24.2 kDa size (Figure 4). Analysis of the protein by sedimentation equilibrium gave an estimated maximum size of 48 kDa, which clearly indicated that the GSTZ1 protein assembles as a dimer, as is observed in the other cytosolic GSTs [1,2].

The purified GSTZ1-1 was tested for enzymic activity with a range of compounds that have been found to be substrates for GSTs from the Alpha, Mu, Pi and Theta classes (Table 1). Although GST activity was demonstrated with a number of compounds, the levels of activity were low and well below those

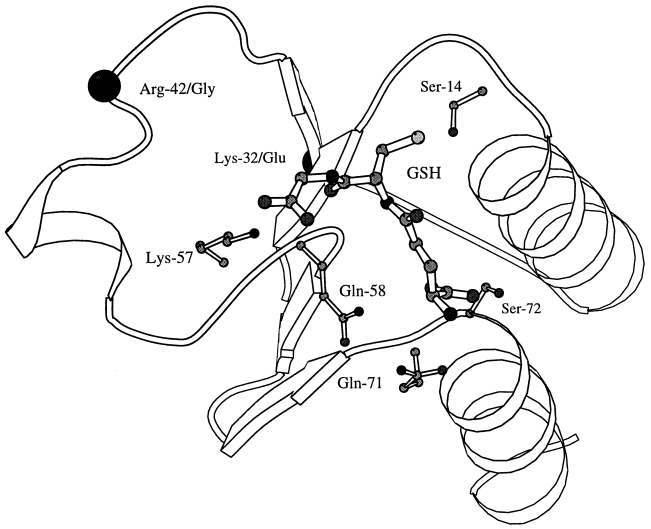

Figure 5 MOLSCRIPT [44] schematic of the modelled N-terminal domain of GSTZ1

Residues that interact directly with glutathione (GSH) are drawn except for Val-58, of which only the backbone atoms that hydrogen-bond with the glutathione are shown. In the diagram, carbon atoms are shown in black, nitrogen in grey and oxygen in white. The thiol sulphur is shown in grey directly adjacent to the putative active-site residue Ser-14. The position of two residues (Arg-42/Gly and Lys-32/Glu) potentially exhibiting genetic polymorphism are shown as black spheres in the carbon backbone.

Table 2 Summary of tissues in which GSTZ1 transcripts have been detected, and the nucleotides at positions 94 and 124

Tissue	EST accession number	nt 94	nt 124
Melanocyte*	N31040	A	A
Melanocyte	H83727	A	G
Synovium	T35781	A	G
Breast	R67650	G	G
Skeletal muscle	Z24885	G	G
Infant brain	H30434	G	G
Placenta	R36472	G	N†
Foetal liver	H90978	G	G
Foetal heart	W79121	A	A
Pancreas adenocarcinoma	AA076211	A	A
Senescent fibroblasts	W23979	A	G

* The clone sequenced in this study.
† There was an ambiguous sequence at this point.

obtained with other GSTs. The highest activity measured was the glutathione peroxidase activity with t-butyl and cumene hydroperoxides. This activity is also present in some Alpha and Theta class isoenzymes [2,21,37].

Structural model

To investigate the structure of the G site of the human Zeta class GST, a tentative atomic model of the N-terminal domain has been constructed with the knowledge that this domain is generally well conserved across the different classes of cytosolic GSTs (Figure 5). The model shows that many of the residues that bind glutathione in the G site are conserved either in the *L. cuprina* structure or in Alpha, Mu, Pi or Sigma class GST structures [30,33,38–40]. In particular, the model shows that Ser-14, equivalent to Ser-9 in the *L. cuprina* enzyme, is appropriately placed to be able to stabilize the glutathione thiolate. The side chain of Gln-58 forms hydrogen bonds with the γ-Glu side-chain oxygen and to the glutathione Gly terminal oxygen, analogous to Gln-54 in the Alpha class GSTs. Likewise, the backbone of Val-59 (Alpha class Val-55) hydrogen-bonds to the backbone of the glutathione Cys. Although Val-59 seems conserved in the Zeta class, the residue type at position 58 can vary. Each of these residue types, however, can still provide a donor proton to support the hydrogen-bonding. The hydroxide and backbone nitrogen of Ser-72 form hydrogen bonds with the γ-Glu carboxylate, as found with Ser-65 in Pi and *L. cuprina* Delta class, Ser-72 in Mu class, Ser-64 in Sigma class and Thr-68 in Alpha class GSTs, whereas Gln-71 [Glu-64 in *L. cuprina* (Delta); Gln-67 in Alpha; Gln-64 in Pi; Gln-71 in Mu and Gln-63 in Sigma] hydrogen-bonds to both the N-terminus and C-terminus of the γ-Glu. Lys-57 interacts electrostatically with the glutathione Gly terminal oxygen in the glutathione in a similar manner to that of His-50 in the *L. cuprina* enzyme. An equivalent positive charge in the Alpha, Mu, Pi and Sigma class GSTs tends to be contributed by residues located prior to or in the loop before helix 2. In other Zeta class members, the positive charge is likely to be contributed by the residue at position 45 (carnation) or 51 (*C. elegans*). In cytosolic GSTs, a negative charge is often contributed to the G site by a residue in the C-terminal domain in the second subunit of the intact dimer. A possible candidate for this interaction in GSTZ1 would be Asp-104. No negatively charged residue in this region is conserved in the Zeta class.

Tissue distribution of GSTZ1-1

The tissues in which GSTZ1-1 is expressed were evaluated by Western blotting with a rabbit polyclonal antiserum raised against purified recombinant GSTZ1-1 (Figure 4). Among the adult tissues studied, the most pronounced expression appeared in the liver, with lower levels in skeletal muscle and brain. It was notable that the antiserum detected two cross-reacting proteins in the muscle. Further studies are required to resolve the origin of this size variation.

To evaluate the expression of GSTZ1-1 in a wider range of tissues we searched the Genbank EST database with the GSTZ1 cDNA sequence. This search revealed the expression of GSTZ1 in the tissues and cell types listed in Table 2. It is notable that the ESTs were found in cDNA libraries from foetal tissues and were not found in libraries from major adult tissues, such as liver and brain and kidney. Although not shown in Table 2, multiple GSTZ1 clones were identified from melanocytes, breast, infant brain and foetal liver.

Variation in GSTZ

Because multiple genes or alleles have been found within each of the human Alpha, Mu, Pi and Theta classes [3], we examined the available sequence data from 13 ESTs encoding GSTZ1 for evidence of genetically determined sequence heterogeneity. Because of the automated sequencing approach utilized in the generation of many EST sequences, there are uncorrected sequencing errors that cloud the identification of genetically determined sequence variation. However, this source of variation can be largely eliminated by ignoring individual sequence variations and by focusing on variations that occur in multiple ESTs. Alignment of the available GSTZ1 sequences revealed

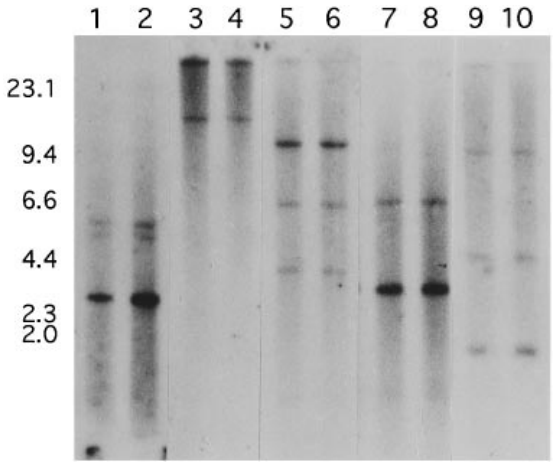

Figure 6 Southern blot of human DNA digested with different restriction enzymes and hybridized with the GSTZ1 cDNA

Lanes 1 and 2, *Eco*RI; lanes 3 and 4, *Bam*HI; lanes 5 and 6, *Hin*dIII; lanes 7 and 8, *Pst*I; lanes 9 and 10, *Sac*I. The molecular sizes shown at the left are in kbp.

consistent variability at two sites. As shown in Table 2, A or G was variably found at nt 94 and A or G was variably found at nt 124. These nucleotide changes are numbered from the initiating ATG codon and result in Lys-32 → Glu and Arg-42 → Gly substitutions respectively in the GSTZ1 sequence. With the exception of the two melanocyte ESTs, each clone represented in Table 2 was isolated from a different library that was prepared from a different individual. Thus these substitutions suggest that there might be genetically determined variation in GSTZ1. However, further studies are required to determine whether this variation is the result of multiple Zeta class gene loci or whether it is allelic. In this regard, the two melanocyte ESTs were cloned from the same library/individual and clearly differ at nt 124. This suggests, but does not prove, that the variation is allelic. As the available EST sequences do not extend over the complete length of the cDNA species, there might be additional variation that has not yet been detected.

Southern-blot analysis was undertaken to gain a greater understanding of the potential complexity of the GSTZ1 locus. Blots of genomic DNA digested with several restriction enzymes revealed multiple DNA fragments that hybridized to the GSTZ1 cDNA (Figure 6). The size and number of these fragments suggest that either the gene is considerably larger than the genes encoding the other GST classes or there is more than one cross-hybridizing gene/pseudogene within this class.

DISCUSSION

The Zeta class of GSTs defined in this study has a number of distinctive features. In particular contrast with the other classes, the Zeta class has been very well conserved over a considerable evolutionary period, such that 38 % identity remains between the carnation and human sequences and 49 % remains between the *C. elegans* and human sequences. Because most previously described classes seem to have evolved within major phylogenetic groups, it was striking to identify closely related GSTs from such divergent species. There is a particularly high percentage identity in the N-terminal region that contains the active site residue in all previously characterized soluble GSTs. This region contains a motif, SSCXWRVIAL, that is completely conserved in all three Zeta class sequences. The first serine residue of this motif aligns with the active-site serine found in the mammalian Theta and insect Delta class sequences. This suggests that Ser-14 in GSTZ1 might be catalytically active; however, this remains to be confirmed by mutagenesis and structural studies. If Ser-14 is involved in catalysis in GSTZ1-1, it suggests that there has been a major evolutionary separation of the cytosolic GSTs into two subfamilies that are characterized by either an active-site serine residue (Theta, Delta and Zeta classes) or an active-site tyrosine residue (Alpha, Mu, Pi and Sigma). The separation of the two subfamilies is clearly evident in the phylogenetic tree shown in Figure 3. A possible exception to this is the group comprising the yeast, *Arabidopsis* and petunia sequences. This group does not seem to have a conserved serine or tyrosine residue in this region; further structural studies of the these GSTs are required given that other plant sequences not shown in this alignment do in fact contain a serine residue at that position [41]. The sequence alignment and the phylogenetic tree (Figures 2 and 3) show that the recently described mitochondrial Kappa class is widely diverged from all the previously defined classes. The mechanism of catalysis and the residues involved are yet to be defined in the Kappa class.

The overall appearance of the alignment in Figure 2 indicates that it is likely that the structure of GSTZ1 strongly resembles that of other cytosolic members of the GST family, and the conservation of residues Gly-156 and Asp-163 support this. However, details of the relative orientation and the occurrence of bulges in helix 4 and/or 5 of the C-terminal domains in the structures of Alpha, Mu, Pi, Delta (*L. cuprina*) and Sigma GSTs hamper the construction of an accurate model for this domain, and in particular of the H site [30,33,38–40]. The N-terminal domain is much more conserved: a model of this domain (Figure 5) shows that there are no obvious residues that would prevent the binding of glutathione. Further, the structural conservation of residues that facilitate the binding of glutathione suggests that the orientation of the substrate in the GSTZ1 G site should be very similar to that found in the other cytosolic GSTs.

The role of the Zeta class GSTs is not clear. Their high degree of sequence conservation over a long evolutionary period and the wide tissue distribution of GSTZ1-1 suggests that they might be involved in the metabolism of a compound that is a common metabolic product or a significant component in the environment of most species. In carnation petals, the Zeta class GSTs are transcribed during senescence and the genes contain ethylene response elements [42]. However, nothing is known about Zeta class gene regulation in other species. The presence of GSTZ1 cDNA species in many libraries derived from foetal tissues suggests that there might be a level of developmental regulation. However, this point is not clear, in view of our observation that GSTZ1-1 is also expressed in adult tissues. Our study of a range of compounds previously shown to be substrates for cytosolic GSTs from the previously described classes revealed a small number of compounds that elicited relatively weak activity. The highest activity was recorded with organic hydroperoxides, and they represent a class of compounds that are likely to be generated in most cells and tissues. Further investigation of the capacity of GSTZ1-1 to utilize other organic peroxides or products of oxygen metabolism is clearly warranted.

It is notable that, although we clearly detected GSTZ1-1 in liver, kidney and brain by immunological methods, we failed to identify ESTs derived from these adult tissues. A search of the EST database with a GSTA1 sequence revealed a number of human liver-derived GSTA1 clones, confirming the presence of adult liver cDNA species in the database. From the EST database search we identified multiple GSTZ1 clones in melanocytes,

breast, infant brain and foetal liver, suggesting that this isoenzyme is relatively abundant in those tissues. The failure to detect GSTZ1 ESTs from adult tissues might reflect their relatively poor abundance, but might also be an artifact of the normalization of cDNA libraries used in EST sequencing programs [43].

In humans there are multiple alleles or gene loci within each of the Alpha, Mu and Pi classes [43]. This heterogeneity might also be found in the Zeta class, as two possible allelic variants have been identified by comparing the available EST sequences. The variation at nt 94 would result in a Lys → Glu substitution at residue 32. On the basis of the *L. cuprina* structure [30] and the model developed in this study, this residue seems to fall at the beginning of $\beta 2$ and would not be expected to interact directly with the active site (Figure 5). It is notable that a glutamic residue is present in this position in both the carnation and *C. elegans* sequences. The second variant at nt 124 encodes an Arg-42 → Gly substitution. This codon seems to fall in the $\beta_2 \rightarrow \alpha_2$ loop region (Figure 5) and, because this region varies significantly in all the GST structures solved so far, it is hard to predict whether this substitution has any functional role.

Many previous studies have focused on the GSTs that utilize the model substrate 1-chloro-2,4-dinitrobenzene and bind to glutathione affinity matrices. Because GSTZ1 has limited activity with known GST substrates and binds relatively weakly to glutathione–agarose, it is evident that this GST has remained undiscovered during conventional protein-purification studies. The discovery of the Zeta class of GSTs supports the view that the full extent of this gene family has yet to be realized.

We thank Ms. Marj Coggan for technical assistance, Dr Jun Adachi for assistance with the phylogenetic analysis, and Dr Peter Jeffrey for assistance with the sedimentation equilbrium analysis. R.T.B. and G.C. are recipients of Australian Research Council Fellowships.

REFERENCES

1 Hayes, J. D. and Pulford, D. J. (1995) CRC Crit. Rev. Biochem. **30**, 445–600
2 Mannervik, B. and Danielson, H. (1988) CRC Crit. Rev. Biochem. **23**, 283–337
3 Board, P. G., Coggan, M., Johnston, P., Ross, V., Suzuki, T. and Webb, G. (1990) Pharmacol. Ther. **48**, 357–369
4 Board, P. G. (1981) Am. J. Hum. Genet. **33**, 36–43
5 Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K. and Jornvall, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7202–7206
6 Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991) Biochem. J. **274**, 409–414
7 Hiratsuka, A., Sebata, N., Kawashima, K., Okuda, H., Ogura, K., Watabe, T., Satoh, K., Hatayama, I., Tsuchida, S. and Ishikawa, T. (1990) J. Biol. Chem. **265**, 11973–11981
8 Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O. and Huber, R. (1991) EMBO J. **10**, 1997–2005
9 Stenberg, G., Board, P. G. and Mannervik, B. (1991) FEBS Lett. **293**, 153–155
10 Board, P. G., Coggan, M. and Parker, M. W. (1996) Biochem. J. **311**, 247–250
11 Tan, K.-L., Chelvanayagam, G., Parker, M. W. and Board, P. G. (1996) Biochem. J. **319**, 315–321
12 Pemble, S. E., Wardle, A. F. and Taylor, J. B. (1996) Biochem. J. **319**, 749–754
13 Anderson, C., Mosialou, E., Wernander, R., Herbert, H. and Morgenstern, R. (1993) Structure and Function of Glutathione Transferases, pp. 109–116, CRC Press, Boca Raton
14 Pemble, S. E. and Taylor, J. B. (1992) Biochem. J. **287**, 957–963
15 Buetler, T. M. and Eaton, D. I. (1992) Environ. Carcinogen. Ecotoxicol. Rev. **C10**, 181–203
16 Blocki, F. A., Ellis, L. B. M. and Wackett, L. P. (1993) Protein Sci. **2**, 2095–2102
17 Toung, Y.-P. S., Hsieh, T.-S. and Tu, C.-P. D. (1993) J. Biol. Chem. **268**, 9737–9746
18 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
19 Meyer, R. C., Goldsbrough, P. B. and Woodson, W. R. (1991) Plant Mol. Biol. **17**, 277–281
20 Baker, R. T., Smith, S. A., Marano, R., Mckee, J. and Board, P. G. (1994) J. Biol. Chem. **269**, 25381–25386
21 Tan, K.-L. and Board, P. G. (1996) Biochem. J. **311**, 247–250
22 Simmons, P. C. and Vander Jagt, D. L. (1977) Anal. Biochem. **82**, 334–341
23 Mannervik, B. and Widersten, M. (1995) in Advances in Drug Metabolism in Man (Pacifici, G. M. and Fracchia, G. N., eds.), pp. 407–459, Office for Official Publications of the European Communities, Luxemburg
24 Brophy, P. M., Southan, C. and Barrett, J. (1989) Biochem. J. **262**, 939–946
25 Ricci, G., Caccuri, A. M., Lo Bello, M. A. P., Pumonte, F. and Federici, G. (1994) Anal. Biochem. **218**, 463–465
26 Nash, T. (1953) Biochem. J. **55**, 416–421
27 Gilham, B. (1971) Biochem. J. **121**, 667–672
28 Beutler, E. (1975) Red Blood Metabolism: A Manual of Biochemical Methods, 2nd edn., pp. 71–73, Grune and Stratton, New York
29 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Nucleic Acids Res. **22**, 4673–4680
30 Wilce, M. C. J., Board, P. G., Feil, S. C. and Parker, M. W. (1995) EMBO J. **14**, 2133–2143
31 Lim, K., Ho, J. H., Keeling, K., Gilliland, G. L., Ji, X., Rüker, F. and Carter, D. C. (1994) Protein Sci. **3**, 2233–2244
32 Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tormarev, S. I., Piatigorsky, J., Armstrong, R. N. and Gilliland, G. L. (1995) Biochemistry **34**, 5317–5328
33 Reinemer, P., Prade, L., Hof, P., Neuefeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Bartunik, H. D. and Bieseler, B. (1996) J. Mol. Biol. **255**, 289–309
34 Jones, D. T., Taylor, W. R. and Thornton, G. M. (1992) Comput. Appl. Biosci. **8**, 275–282
35 Adachi, J. and Hasegawa, M. (1996) Computer Science Monographs, no. 28: MOLPHY Version 2.3: Programs for Molecular Phylogenetics based on Maximum Likelihood, The Institute of Statistical Mathematics, Tokyo
36 Lai, H.-C. J., Li, N.-C. L., Weiss, M. J., Reddy, C. C. and Tu, C.-P. D. (1984) J. Biol. Chem. **259**, 5536–5542
37 Hussey, A. J. and Hayes, J. D. (1992) Biochem. J. **286**, 929–935
38 Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G. et al. (1993) J. Mol. Biol. **232**, 192–212
39 Ji, X., Zhang, P., Armstrong, R. N. and Gilliland, G. L. (1992) Biochemistry **31**, 10169–10184
40 Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O. and Huber, R. (1991) EMBO J. **10**, 1997–2005
41 Chelvanayagam, G., Wilce, M. C. J., Parker, M. W., Tan, K. L. and Board, P. G. (1997) Proteins **27**, 118–130
42 Itzhaki, H. and Woodson, W. R. (1993) Plant Mol. Biol. **22**, 43–58
43 Soares, M. B., Bonaldo, M. De F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 9228–9232
44 Kraulis, J. P. (1991) J. Appl. Crystallogr. **24**, 946–950

Received 23 June 1997/1 September 1997; accepted 3 September 1997