Modelling and bioinformatics studies of the human Kappa-class glutathione transferase predict a novel third glutathione transferase family with similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases

Anna ROBINSON*[1], Gavin A. HUTTLEY†, Hilary S. BOOTH† and Philip G. BOARD*

*John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra City, ACT 2601, Australia, and †Centre for Bioinformation Science, JCSMR and Mathematical Sciences Institute, Australian National University, P.O. Box 334, Canberra City, ACT 2601, Australia

The Kappa class of GSTs (glutathione transferases) comprises soluble enzymes originally isolated from the mitochondrial matrix of rats. We have characterized a Kappa class cDNA from human breast. The cDNA is derived from a single gene comprising eight exons and seven introns located on chromosome 7q34–35. Recombinant hGSTK1-1 was expressed in *Escherichia coli* as a homodimer (subunit molecular mass $\sim$25.5 kDa). Significant glutathione-conjugating activity was found only with the model substrate CDNB (1-chloro-2,4-ditnitrobenzene). Hyperbolic kinetics were obtained for GSH (parameters: K_m^{app}, 3.3 ± 0.95 mM; V_{max}^{app}, 21.4 ± 1.8 μmol/min per mg of enzyme), while sigmoidal kinetics were obtained for CDNB (parameters: $S_{0.5}^{app}$, 1.5 ± 1.0 mM; V_{max}^{app}, 40.3 ± 0.3 μmol/min per mg of enzyme; Hill coefficient, 1.3), reflecting low affinities for both substrates. Sequence analyses, homology modelling and secondary structure predictions show that hGSTK1 has (a) most similarity to bacterial HCCA (2-hydroxychromene-2-carboxylate) isomerases and (b) a predicted C-terminal domain structure that is almost identical to that of bacterial disulphide-bond-forming DsbA oxidoreductase (root mean square deviation 0.5–0.6 Å). The structures of hGSTK1 and HCCA isomerase are predicted to possess a thioredoxin fold with a polyhelical domain (α_x) embedded between the β-strands ($\underline{\beta\alpha\beta}\alpha_x\underline{\beta\beta\alpha}$, where the underlined elements represent the N and C motifs of the thioredoxin fold), as occurs in the bacterial disulphide-bond-forming oxidoreductases. This is in contrast with the cytosolic GSTs, where the helical domain occurs exclusively at the C-terminus ($\underline{\beta\alpha\beta\alpha\beta\beta\alpha}\alpha_x$). Although hGSTK1-1 catalyses some typical GST reactions, we propose that it is structurally distinct from other classes of cytosolic GSTs. The present study suggests that the Kappa class may have arisen in prokaryotes well before the divergence of the cytosolic GSTs.

Key words: disulphide-bond-forming (Dsb) oxidoreductase, evolution, glutathione transferase (GST), human, 2-hydroxychromene-2-carboxylate (HCCA) isomerase, Kappa.

INTRODUCTION

The superfamily of GSTs (glutathione transferases) comprises a complex group of enzymes with a wide biological distribution. They are found in organisms ranging from bacteria to mammals [1–3]. While involved primarily in detoxification mechanisms and redox biochemistry, GSTs have other distinct roles. Some are found in metabolic pathways associated with tyrosine degradation [4,5], steroid hormone biosynthesis [6] and prostaglandin D_2 synthesis [7,8]. Others have non-enzymic roles and can bind hydrophobic compounds such as steroids, bilirubin, haem and fatty acids as ligands rather than substrates [9]. Additional studies show that GSTs have roles in the modulation and formation of ion channels [10], as well as modulating signal transduction pathways [11,12]. GSTs have also been found to play a highly specialized role as lens crystallins in squid and other cephalopods [13].

Enzymes are defined as GSTs if they have GST activity and/or sequence and structural similarity. New GSTs and GST-like enzymes are continually being discovered, and over 1000 sequences containing GST domains have been identified [14]. At present, GSTs are divided into two major, evolutionarily distinct, families: the MAPEGs (microsomal membrane-associated proteins) and cytosolic, soluble enzymes. The cytosolic GSTs are further subdivided into several divergent classes. Eight classes, termed Alpha, Mu, Pi, Theta, Sigma, Omega, Zeta and Kappa, have been described in mammals; additional classes are Phi, Tau and Lambda in plants, Delta in insects and Beta in bacteria [3,15].

Representatives from most classes of soluble GSTs have been well studied, but little research has been carried out on the Kappa-class GSTs. The first representative of this class was isolated from the rat liver mitochondrial matrix in 1991 [16]. It was characterized as a GST due to its typical size, dimerization in solution and activity on selected model GST co-substrates. Due to low N-terminal sequence similarity with other GSTs, it was initially placed in the Theta class, at the time thought to be the progenitor class of the GST superfamily. A cDNA clone encoding a rat Kappa-class GST, and sequences for the human and pig isoenzymes, were reported subsequently [17]. Sequence alignments and phylogenetic studies [5] indicated that the Kappa-class GSTs have little sequence similarity with the cytosolic GSTs, raising doubts about their relatedness. Jowsey et al. [18] recently demonstrated that the mouse and rat Kappa-class GSTs are very similar in terms of gene structure, primary sequence, catalytic properties, tissue specific expression and subcellular localization.

Abbreviations used: AEBSF, 4-(2-aminoethyl)-benzenesulphonyl fluoride hydrochloride; CDNB, 1-chloro-2,4-dinitrobenzene; CLIC, chloride intracellular channel; Dsb, disulphide-bond-forming; EST, expressed sequence tag; GST, glutathione transferase; HCCA, 2-hydroxychromene-2-carboxylate; MAPEG, microsomal membrane-associated protein; rmsd, root mean square deviation; Ub–hGSTK1-1, His–ubiquitin–hGSTK1-1.

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

The nucleotide sequence data reported will appear in the EMBL, GenBank®, DDBJ and GSDB Nucleotide Sequence Databases under the accession number AY520571.

In the present study, we have characterized hGSTK1-1, a recombinant human Kappa-class GST, and confirmed that it has some biochemical characteristics that are typical of soluble cytosolic GSTs. However, our studies also predict that the Kappa-class enzymes may have a novel structure and evolutionary history that is more closely related to bacterial HCCA (2-hydroxychromene-2-carboxylate) isomerases and bacterial disulphide-bond-forming oxidoreductases than to other cytosolic GSTs.

EXPERIMENTAL

Materials

Reagents for solution and kinetic studies were purchased from Sigma-Aldrich Chemical Corp. (St. Louis, MO, U.S.A.) and Merck Pty Ltd (Damstadt, Germany). AEBSF [4-(2-aminoethyl)-benzenesulphonyl fluoride hydrochloride] was purchased from ICN Biomedicals Australasia Pty Ltd (Seven Hills, NSW, Australia). PCR kits were purchased from Advanced Biotechnologies (Epsom, Surrey, U.K.), and restriction enzymes were obtained from Roche Diagnostics Australia Pty Ltd (Castle Hill, NSW, Australia), New England Biolabs Inc. (Beverly, MA, U.S.A.) or Promega Corp. (Annandale, NSW, Australia). Oligonucleotides were purchased from GENSET Pacific Pty Ltd (Lismore, NSW, Australia). The His-tagged ubiquitin-fusion expression plasmid was kindly provided by Dr Rohan Baker (JCSMR, Australian National University). DNA and protein N-terminal sequencing were carried out at the Biomolecular Resource Facility (BRF, JCSMR, Australian National University).

Cloning of hGSTK1

Clones containing the *hGSTK1* coding sequence were identified in the human EST (expressed sequence tag) database (http://ncbi.nlm.nih.gov) by searching with the putative hGSTK1 cDNA sequence [5]. A cDNA clone (R83924) from the Soares breast 3NbHBst *Homo sapiens* library was obtained from the IMAGE Consortium through Incyte Corp. *(*Palo Alto, CA, U.S.A.).

To construct a plasmid for the heterologous expression of recombinant hGSTK1 in *Escherichia coli*, the cDNA coding sequence was amplified by PCR using primers AKEX-F (5′-ataccgcggtggaatggggcccctgccgcg-3′) and HKEX-R (5′-aataagcttaaagtctggcattca-3′). These primers were designed to introduce *Sac*II and *Hin*dIII restriction enzyme sites at the 5′ and 3′ ends respectively to simplify subsequent cloning procedures. PCR reactions were carried out in 20 μl volumes containing $\sim$100–200 ng of template DNA, 10 pmol each of forward and reverse primers, 1× Advanced Biotechnologies (now ABGene) buffer IV [200 mM $(NH_4)_2SO_4$, 75 mM Tris/HCl, pH 8.8, 0.01 % (v/v) Tween 20®], 1.5–2.0 mM $MgCl_2$, 0.2 μM dNTPs, 0.5 unit of *Taq* DNA polymerase and sterile double-distilled water. Samples were initially pre-heated to 95 °C for 2 min, then amplified using a two-step process to optimize the temperature annealing conditions for each primer. Step 1 comprised ten cycles of denaturation at 95 °C for 20 s, primer annealing at 63 °C for 25 s and extension at 72 °C for 1 min. Step 2 involved 24 cycles of denaturation at 95 °C for 20 s, primer annealing at 59 °C for 20 s and extension at 72 °C for 1 min. The reaction was completed with 5 min at 72 °C.

The amplified product was digested with *Sac*II and *Hin*dIII restriction enzymes and ligated between *Sac*II and *Hin*dIII restriction sites downstream of a His_6-tagged human ubiquitin cDNA sequence in a modified pET15 vector termed pHUE (A.-M. Catanzariti, T. A. Soboleva, P. G. Board and R. T. Baker, unpublished work). This created a new plasmid, phGSTK1-R4.

Genes expressed in a pHUE construct are synthesized as ubiquitin fusion proteins and can be readily purified by immobilized metal ion affinity chromatography on nickel–agarose. Pure proteins are cleaved precisely between the C-terminal glycine of the His-tagged ubiquitin and the first residue of the fused protein by digestion with a specific His-tagged ubiquitin protease (A.-M. Catanzariti, T. A. Soboleva, P. G. Board and R. T. Baker, unpublished work). This novel procedure provides high yields of pure proteins that have no additional residues (see below), unlike other His-tagged constructs that are modified at either the N- or C-terminus.

DNA sequencing

Sequencing reactions were carried out in both 5′–3′ and 3′–5′ directions using the ABI PRISM® Big Dye™ Terminator Cycle Sequencing method on an ABI PRISM 377 DNA Sequencer. DNA sequences from both the cDNA clone and the new phGSTK1-R4 vector were verified against each other and also against the genomic sequence.

Protein sequencing

N-terminal sequencing was carried out on proteins isolated from Coomassie Blue-stained SDS/polyacrylamide gel fragments using an Applied Biosystems Procise-HT 494 protein sequencer.

Expression and purification of recombinant hGSTK1

Plasmid phGSTK1-R4 was transformed into *E. coli* BL21(DE3) cells. Fresh overnight cultures were inoculated into Luria broth containing ampicillin (100 μg/ml). Cultures were incubated at 37 °C in a shaking temperature-controlled cabinet to a density of approx. $D_{595} = 0.5–0.8$. Expression of hGSTK1-1 was induced with 0.25 mM isopropyl β-D-thiogalactoside for 2–3 h. Cultures were harvested by centrifuging at 10 000 ***g*** and the pellet was stored at − 20 °C until use.

Thawed cells were resuspended in 10 % of the original volume of lysis buffer (25 mM Hepes/HCl or Tris/HCl, pH 7.5–8, 150 mM NaCl, 10 mM imidazole, 5 mM 2-mercaptoethanol) with 1 mg/ml lyzozyme and 0.1–0.2 mM AEBSF protease inhibitor added. The solution was stirred at 4 °C for 1 h and then sonicated three times for 10–15 s on ice using a Branson Sonifer 250 with the output control set at 6. The resulting homogenate was centrifuged at 12 000 ***g*** for 30–60 min at 4 °C to pellet cell debris.

Recombinant hGSTK1-1 enzyme was isolated and purified by immobilized metal ion affinity chromatography [19]. Tris or Hepes buffers were used at the lowest effective concentration (25 mM) to minimize reduction of nickel ions by secondary or tertiary amines.

Supernatant containing soluble Ub–hGSTK1-1 (His–ubiquitin–hGSTK1-1) was applied to a column of nickel–agarose ($\sim$5 ml bed volume) pre-equilibrated with buffer A (25 mM Hepes/HCl or Tris/HCl at pH 7.5–8, 150 mM NaCl, 10 mM imidazole and 5 mM β-mercaptoethanol). The column was washed with buffer B (25 mM Hepes/HCl or Tris/HCl at pH 7.5–8, 150 mM NaCl, 20 mM imidazole and 5 mM 2-mercaptoethanol) to remove unbound proteins, and Ub–hGSTK1-1 was eluted in buffer C (25 mM Hepes/HCl or Tris/HCl, pH 7.5–8, 150 mM NaCl, 250 mM imidazole and 5 mM 2-mercaptoethanol). The eluted fraction was dialysed against 25 mM Hepes/HCl or Tris/HCl, pH 7.5–8, 150 mM NaCl and 5 mM 2-mercaptoethanol through a Spectra-por membrane (molecular mass cut-off 10 000 Da) to eliminate excess imidazole (final concentration of imidazole < 0.1 mM). The dialysed solution was concentrated by centrifugation through 10 000 Da molecular mass cut-off filters

(Amicon®; Millipore Corp.), to reduce the volume, and incubated with a His-tagged, ubiquitin-cleaving protease for 1 h at 37 °C.

The reaction solution was cooled and particulate matter was removed by filtering or centrifugation. The protein solution was then re-applied to the nickle–agarose chromatography column that had been pre-equilibrated with buffer A. hGSTK1-1 was initially eluted directly in the flow-through. Residual hGSTK1-1 was eluted with ~ 50 ml of buffer B. His-tagged ubiquitin and His-tagged protease remain immobilized on the Ni^{2+}-nitrilotriacetate resin. Imidazole was again dialysed out of the enzyme preparation against two changes of fresh buffer (25 mM Hepes/HCl or Tris/HCl at pH 7.5–8, 150 mM NaCl and 5 mM 2-mercaptoethanol) to a concentration of < 0.1 mM. Minor amounts of contaminant proteins were removed progressively by ultrafiltration though 100 kDa and 10 kDa molecular mass cut-off filters while concentrating proteins.

Throughout the isolation and purification procedure, protein samples were analysed by SDS/PAGE, N-terminal sequencing and/or kinetic assay. Protein concentration was determined spectrophotometrically at A_{280} using the theoretically determined molar absorption coefficient [20] $\varepsilon_{280} = 34\,970\ M^{-1} \cdot cm^{-1}$.

Substrate specificity, pH dependence and kinetic constants

Enzyme activities with a range of substrates were measured spectrophotometrically. Potential substrates and inhibitors were examined in both standard assays as described previously [5,21] and optimized reaction mixtures.

The effect of pH on the conjugating activity of hGSTK1-1 [GSH:CDNB (1-chloro-2,4-dinitrobenzene)] was measured in selected Good's buffers [22] in the pH range 4–11 and at a concentration of 0.1 M. Buffers used were citrate (pH 4.2 and 5.0), Mes (pH 6.0) Hepes (pH 7.4 and 8.0) Bicine (pH 9.0) and Caps [3-(cyclohexylamino)propane-1-sulphonic acid; pH 10.0 and 11.0]. Assays were carried out using a constant amount of recombinant enzyme with the standard reaction mixture of 1 mM GSH and 1 mM CDNB.

Kinetic measurements were carried out in 0.1 M Hepes buffer, pH 8.0, near the enzyme's pH optimum. To determine the apparent K_m for GSH, initial velocities were measured by varying GSH from 0.1 to 30 mM, with CDNB fixed at 1 mM. To measure the apparent K_m for CDNB, GSH was used at a saturating concentration of 10 mM and the CDNB concentration was varied from 0.1 to 2.0 mM.

Reactions were all carried out in triplicate at a constant temperature of 30 °C using final enzyme concentrations of ~ 20–50 $\mu g \cdot ml^{-1}$. Background rates in the absence of enzyme were deducted from enzymic rates prior to calculating specific activities. In pH and kinetic studies, initial rates were calculated over a period of 0.1 min. Kinetic data were subjected to non-linear regression analysis. Calculations were carried out using software packages PRISM (GraphPad Software Inc) or Excel (Microsoft).

Analytical ultracentrifugation

Samples were analysed by analytical ultracentrifugation (XL-A analytical ultracentrifuge; Beckman/Coulter) using a double-sector filled Epon centrepiece. Samples of 100 μl of protein solution and 110 μl of buffer were placed in the respective solvent sectors of the cuvette. Samples were centrifuged at 25 000 rev./min for 24 h at 20 °C. Radial absorbance was measured at 280 nm. Scans taken at 2 h intervals converged after 20 h, indicating that sedimentation equilibrium had been reached. The molecular mass of the protein was estimated from plots of the natural log of the protein concentration against radius^2. The partial specific volume was 0.73 ml/g (average value for proteins in dilute aqueous media) and the density of the buffer was 1.00 g/ml.

Sequence analyses

Publicly available bioinformatics software tools were used for sequence analyses. ProtParam (http://kr.expasy.org/tools/protparam.html) was used to calculate molecular mass, molar absorption coefficient, pI, amino acid composition and ligand binding motifs. PSORT [23] was used to identify cell trafficking signals.

Searches for sequences similar to that of hGSTK1 were carried out using the standard BLAST program (http://ncbi.nlm.nih.gov/blast) [24] and the PSI BLAST program on the non-redundant (nr) database [25]. The PSI BLAST search was carried out using a BLOSUM62 matrix set at default parameters. Sequences identified in similarity searches were analysed by pairwise comparisons and global alignment. Sequence alignments were adapted from those obtained using the NCBI Conserved Domain Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

Modelling studies

To generate structural models, amino acid sequences were submitted to SwissModel [26] using the First Approach Method set at default parameters. Homology models were generated using PDB co-ordinates of DsbA (*E. coli*) [27,28] and TcpG (*Vibrio cholerae*) [29] oxidoreductase structures as search models [PDB Id: 1FVK (DsbA crystal structure), 1A24 (DsbA NMR structure) and 1BED (TcpG crystal structure)]. Manual adjustments were carried out on the models using the DeePVieW SwissPdbViewer software program that is part of the SwissModel package.

Secondary structure prediction

Sequences were submitted to PredictProtein [30] to determine potential secondary structural elements. The sequences of DsbA and TcpG oxidoreductases, whose structures have been solved, were also submitted to obtain a measure of the correctness of the predictions.

RESULTS

cDNA cloning and gene structure

The cDNA sequence for hGSTK1 is shown in Figure 1. The cDNA is 1034 nucleotides long and contains an open reading frame of 678 nucleotides. The sequence contains a short 5′ non-coding region of 55 nucleotides and a 3′ non-coding region of 297 nucleotides, including the poly(A) tail. The 3′ non-coding region contains only one canonical polyadenylation signal, AATAAA, which occurs 21 bases from the terminus. The open reading frame encodes a protein of 226 residues (including the initiating methionine), with an estimated molecular mass of ~25.5 kDa.

A BLAST search of the human genome sequence with the hGSTK cDNA sequence showed exact correspondence to only one locus (51 064; Contig accession number NT007914.12; gi29797157), located on human chromosome 7, banding position q34–35. This locus spans a region of ~5.7 kb and is divided into eight exons and seven introns (Figure 2). Most introns are relatively small (~150–600 bp), except for intron 5, which is 2320 bp. The initial methionine is encoded in exon 1 and the terminal codon is in exon 8. Splice junctions conform to the GT/AG rule.

We compared translation of the cDNA sequence obtained here with that deduced previously by Pemble and co-workers from

```
5′gccgggttccgggaaaaggagctcctgctgccactgctcttccggagcctgcagc

1/1                                         31/11
ATG GGG CCC CTG CCG CGC ACC GTG GAG CTC TTC TAT GAC GTG CTG TCC CCC TAC TCC TGG
Met gly pro leu pro arg thr val glu leu phe tyr asp val leu ser pro tyr ser trp

61/21                                       91/31
CTG GGC TTC GAG ATC CTG TGC CGG TAT CAG AAT ATC TGG AAC ATC AAC CTG CAG TTG CGG
leu gly phe glu ile leu cys arg tyr gln asn ile trp asn ile asn leu gln leu arg

121/41                                      151/51
CCC AGC CTC ATA ACA GGG ATC ATG AAA GAC AGT GGA AAC AAG CCT CCA GGT CTG CTT CCC
pro ser leu ile thr gly ile met lys asp ser gly asn lys pro pro gly leu leu pro

181/61                                      211/71
CGC AAA GGA CTA TAC ATG GCA AAT GAC TTA AAG CTC CTG AGA CAC CAT CTC CAG ATT CCC
arg lys gly leu tyr met ala asn asp leu lys leu leu arg his his leu gln ile pro

241/81                                      271/91
ATC CAC TTC CCC AAG GAT TTC TTG TCT GTG ATG CTT GAA AAA GGA AGT TTG TCT GCC ATG
ile his phe pro lys asp phe leu ser val met leu glu lys gly ser leu ser ala met

301/101                                     331/111
CGT TTC CTC ACC GCC GTG AAC TTG GAG CAT CCA GAG ATG CTG GAG AAA GCG TCC CGG GAG
arg phe leu thr ala val asn leu glu his pro glu met leu glu lys ala ser arg glu

361/121                                     391/131
CTG TGG ATG CGC GTC TGG TCA AGG AAT GAA GAC ATC ACC GAG CCG CAG AGC ATC CTG GCG
leu trp met arg val trp ser arg asn glu asp ile thr glu pro gln ser ile leu ala

421/141                                     451/151
GCT GCA GAG AAG GCT GGT ATG TCT GCA GAA CAA GCC CAG GGA CTT CTG GAA AAG ATC GCA
ala ala glu lys ala gly met ser ala glu gln ala gln gly leu leu glu lys ile ala

481/161                                     511/171
ACG CCA AAG GTG AAG AAC CAG CTC AAG GAG ACC ACT GAG GCA GCC TGC AGA TAC GGA GCC
thr pro lys val lys asn gln leu lys glu thr thr glu ala ala cys arg tyr gly ala

541/181                                     571/191
TTT GGG CTG CCC ATC ACC GTG GCC CAT GTG GAT GGC CAA ACC CAC ATG TTA TTT GGC TCT
phe gly leu pro ile thr val ala his val asp gly gln thr his met leu phe gly ser

601/201                                     631/211
GAC CGG ATG GAG CTG CTG GCG CAC CTG CTG GGA GAG AAG TGG ATG GGC CCT ATA CCT CCA
asp arg met glu leu leu ala his leu leu gly glu lys trp met gly pro ile pro pro

661/221
GCC GTG AAT GCC AGA CTT TAA
ala val asn ala arg leu  *

gattgcccggaggaagcaaactcttcgtataaaaaaagcaggccatctgcttaacccttggctccaccataaggcact
gggactcggatttctctatctgatagaggtattttctgtggccctgggagctgtctgtctttcccctacccccaagga
tgccaggaagacgtccaccattagccatgtggcaacctttacttctatgcctcacaagtgcctttcagagagccccaa
ttctgctttcccacaaaataaacctaatgccatcaggcaaaacaaaaaaaaaaaaaaaaaaaa3′
```

Figure 1 cDNA and deduced amino acid sequence of human Kappa-class GST

The translation termination signal is shown with an asterisk. The consensus polyadenylation signal is boxed.

EST database searches [17]. Three anomalies were identified. The sequence presented here has leucine instead of valine at position 97, leucine instead of glutamine at position 183, and histidine instead of tyrosine at position 208. To see if sequence variations could be attributed to polymorphisms, we searched the EST database with our cDNA sequence. The only putative polymorphisms so far identified in the coding region were not consistent with the previously published sequence. Pemble and co-workers did not find any encoded mitochondrial targeting signals [17]. In case the differences between the two sequences altered contextual effects, or additional signals had been identified in the interim, we carried out a search on the sequence reported here using the latest version of PSORT [23]. No mitochondrial targeting signals were found.

Expression and purification of recombinant enzymes

High yields of Ub–hGSTK1 fusion product and hGSTK1-1 enzyme were readily obtained from the expression and purification procedure described in the Experimental section. Approx. 20 mg of pure hGSTK1-1 protein was routinely recovered from 1 litre cultures.

```
gccgggttccgggaaaaggagctcctgctgccactgctcttccggagcct
Ex1
gcagcATGGGGCCCCTGCCGCGCACCGTGGAGCTCTTCTATGACGTGCTG
     M  G  P  L  P  R  T  V  E  L  F  Y  D  V  L
TCCCCCTACTCCTGGCTGGGCTTCGAGgtgacgctgggaggg...500bp
S  P  Y  S  W  L  G  F  E
..gacctctgcccgcagATCCTGTGCCGGTATCAGAATATCTGGAACATC
             Ex2 I  L  C  R  Y  Q  N  I  W  N  I
AACCTGCAGTTGCGGCCCAGCCTCATAACAGGGATCATGAAAGACAGTGg
M  L  Q  L  R  P  S  L  I  T  G  I  M  K  D  S  G
taggaagggagggt..380bp..accttctcctggcagGAAACAAGCCTC
                                   Ex3  N  K  P  P
CAGGTCTGCTTCCCCGCAAAGGACTATACATGGCAAATGACTTAAAGCTC
  G  L  L  P  R  K  G  L  Y  M  A  N  D  L  K  L
CTGAGACACCATCTCCAGATTCCCATCCACTTCCCCAAGGATTTCTTGTC
L  R  H  H  L  Q  I  P  I  H  F  P  K  D  F  L  S
TGTGATGCTTGAAAAAGgtgaagagagtggga..315bp..gcctctgcc
 V  M  L  E  K  G
ccacagGAAGTTTGTCTGCCATGCGTTTCCTCACCGCCGTGAACTTGGAG
  Ex4   S  L  S  A  M  R  F  L  T  A  V  N  L  E
CATCCAGAGATGCTGGAGAAAGCGTCCCGGGAGCTGTGGATGCGCGTCTG
H  P  E  M  L  E  K  A  S  R  E  L  W  M  R  V  W
GTCAAGGgtgagtgtggggctc..168bp..ttctcttcttcccagAATG
 S  R                                  Ex5  N  E
AAGACATCACCGAGCCGCAGAGCATCCTGGCGgtgagtgtcctggct...
  D  I  T  E  P  Q  S  I  L  A
2320bp..actatattcccttagGCTGCAGAGAAGGCTGGTATGTCTGCA
                  Ex6  A  A  E  K  A  G  M  S  A
GAACAAGCCCAGGGACTTCTGGAAAAGATCGCAACGCCAAAGGTGAAGAA
E  Q  A  Q  G  L  L  E  K  I  A  T  P  K  V  K  N
CCAGCTCAAGGAGACCACTGAGGCAGCCTGCAGATACGGAgtgagcagct
 Q  L  K  E  T  T  E  A  A  C  R  Y  G
ctttat..357bp..aactgcctttctccagGCCTTTGGGCTGCCCATCA
                            Ex7  A  F  G  L  P  I  T
CCGTGGCCCATGTGGATGGCCAAACCCACATGTTATTTGGCTCTGACCGG
  V  A  H  V  D  G  Q  T  H  M  L  F  G  S  D  R
ATGGAGCTGCTGGCGCACCTGCTGGgtaagtaagttaaag..603bp..g
M  E  L  L  A  H  L  L  G
ttttcttcatccagGAGAGAAGTGGATGGGCCCTATACCTCCAGCCGTGA
          Ex8  E  K  W  M  G  P  I  P  P  A  V  N
ATGCCAGACTTTAAgattgcccggaggaagcaaactcttcgtataaaaaa
  A  R  L  *
agcaggccatctgcttaacccttggctccaccataaggcactgggactcg
gatttctctatctgatagaggtattttctgtggccctgggagctgtctgt
ctttcccctacccccaaggatgccaggaagacgtccaccattagccatgt
ggcaacctttacttctatgcctcacaagtgcctttcagagagccccaatt
ctgctttcccacaaaataaacctaatgccatcaggcaaaaca
```

Figure 2 Nucleotide sequence and intron/exon structure of *hGSTK1*

The GSTK gene sequence shown here is derived from the Human Genome (build 33 LOC 51064, gi29797157), and is approx. 5.7 kb in length. The bases encoding exons 1–8 (Ex1–Ex8) are indicated in upper case, and the amino acid sequence is shown below the nucleotide sequence. The translation termination codon is indicated by an asterisk, and the poly(A)$^+$ signal is underlined.

Most extraneous proteins were effectively removed by affinity chromatography washes or 10 kDa molecular mass cut-off filtration steps during protein concentration. One larger protein that appeared in SDS gels to have a molecular mass of ~60 kDa could not be separated by these methods and constantly co-purified with hGSTK1-1 (Figure 3). N-terminal sequencing identified this protein as the *E. coli* chaperonin GroEL or hsp60. hsp60 forms large multisubunit oligomers (> 120 kDa) in its native state [31], which allowed for its ready separation from hGSTK1-1 by filtration through 100 kDa molecular mass cut-off filters during protein concentration procedures.

Sedimentation equilibrium analysis

Duplicate samples of hGSTK1-1 were examined by sedimentation equilibrium analysis. Data taken from three scans obtained at

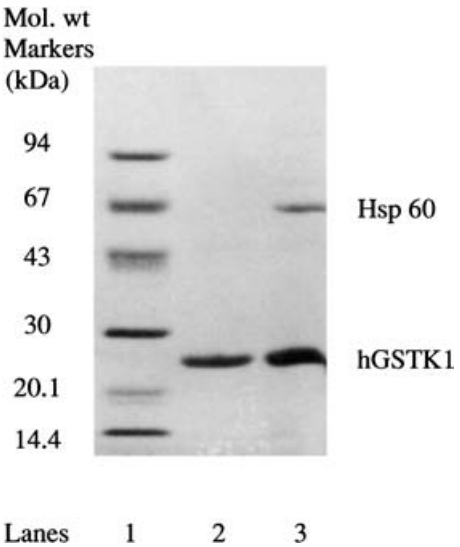

Figure 3 SDS/PAGE analysis of the final stage of hGSTK1-1 isolation and purification

The SDS/12.5 %-polyacrylamide gel shows: lane 1, molecular mass markers; lane 2, pure hGSTK1-1 (~6.0 mg/ml) after filtration of the sample through 100 kDa molecular mass cut-off filters; lane 3, sample of the hGSTK1-1 preparation after nickel–agarose chromatography. The upper band (~60 kDa) on the gel was shown by N-terminal sequencing to be a subunit of hsp60. Hsp60 appeared to consistently co-elute with hGSTK1-1 (~25.5 kDa).

12 min intervals showed that the majority of hGSTK1-1 occurred as a single sedimenting species with a molecular mass of ~46.5 kDa. Once corrections were made for the presence of some monomer (5–7 %), the revised estimate of protein size was ~49–51 kDa. These values indicate that hGSTK1-1 occurs predominantly as a homodimer in solution.

Substrate specificity

A comparison of the specific activities of the human, mouse and rat Kappa-class GSTs with a selection of substrates is shown in Table 1. hGSTK1-1 had significant GSH-conjugating activity with the halogenated aromatics CDNB and 4-nitrobenzylchloride, and some peroxidase activity with cumene and t-butyl hydroperoxide. We also investigated the possibility that hGSTK1-1 could use other thiols as co-substrates. Neither 2-mercaptoethanol nor L-cysteine in the 5–10 mM concentration range showed conjugating activity with CDNB (results not shown). We also tested the hypothesis that pre-incubation with thiol co-substrates may enhance activity, but no increase in catalytic rates was observed.

Kinetic properties

Data from GSH:CDNB conjugation assays in which the GSH concentration was varied and that of CDNB was fixed were readily fitted to the Michaelis–Menton equation. The Michaelis constant and maximal velocity for GSH were calculated to be $K_m^{app} = 3.3 \pm 0.95$ mM and $V_{max}^{app} = 21.4 \pm 1.8$ μmol/min per mg of protein respectively. When CDNB was used as a variable substrate with GSH at a saturating concentration (10 mM), values obtained did not fit the Michaelis–Menton equation, but could be fitted to a Hill plot ($S_{0.5}^{app} = 1.5 \pm 1.0$ mM, $V_{max}^{app} = 40.3 \pm 0.26$ μmol/min per mg of protein and $h = 1.3$). The apparent sigmoidal character suggested possible activation of the enzyme by alkylation of reactive cysteines, as occurs in the microsomal GSTs (MAPEGs), but we found no activation of

Table 1 Activities of hGSTK1-1, mGSTK1-1 and rGSTK1-1 with various substrates

Data for mGSTK1-1 and rGSTK1-1 are taken from Jowsey et al. [18]. For hGSTK1-1, the concentration of GSH was 10 mM (saturating conditions). All values are means ± S.D. of at least three determinations. ND, not detectable; NS, not studied.

	Specific activity (μmol/min per mg)		
Substrate	hGSTK1-1	mGSTK1-1	rGSTK1-1
CDNB	7.4 ± 0.49	154.8 ± 25.0	54.4 ± 17.0
1,2-Dichloronitrobenzene	ND	NS	NS
7-Chloro-4-nitrobenzo-2-oxa-1,3-diazole	0.2 ± 0.03	NS	NS
1,2-Epoxy-3-(4-nitrophenoxy)propane	ND	ND	ND
4-Phenylbut-3-en-2-one	ND	ND	ND
Ethacrynic acid	1.1 ± 0.28	0.29 ± 0.08	0.44 ± 0.12
p-Nitrobenzyl chloride	2.26 ± 0.13	NS	NS
p-Nitrophenylacetate	0.13 ± 0.07	NS	NS
Menaphthyl sulphate	ND	NS	NS
t-Butyl hydroperoxide	0.1 ± 0.02	0.03 ± 0.01	< 0.01
Cumene hydroperoxide	0.2 ± 0.03	1.05 ± 0.04	0.09 ± 0.02
Hydrogen peroxide	ND	NS	NS
trans,trans-Nona-2,4-dienal	0.09 ± 0.01	NS	NS
trans-Nonenal	ND	NS	NS
(±)-2-Bromo-3-(nitrophenyl)propionic acid	ND	NS	NS
Δ^5-Androsten-3,17-dione	ND	NS	NS
Chlorofluroacetic acid	ND	NS	NS

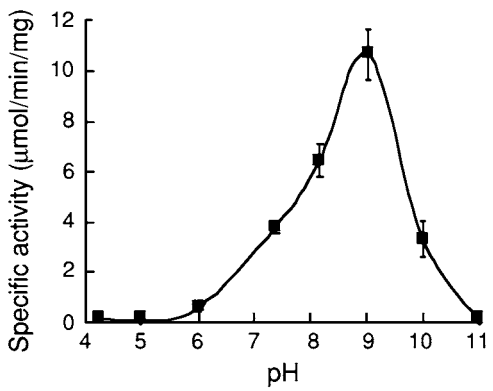

Figure 4 pH dependence of hGSTK1-1

The pH dependence of hGSTK1-1 activity with co-substrates GSH (1 mM) and CDNB (1 mM) is shown. Data are expressed as means ± S.D. ($n = 3$).

hGSTK1-1 by treatment with iodoacetate (results not shown). The sigmoidicity also suggests co-operativity, but we could not fully investigate the enzyme's catalytic mechanism due the limitations of the assay at higher concentrations of CDNB. However, the results we have obtained clearly show that hGSTK1-1 has a low affinity for both CDNB and GSH, indicating that a conjugation reaction may not be the natural role for the Kappa-class GSTs. We anticipate a more comprehensive analysis when appropriate substrates such as HCCA become available.

pH dependence

hGSTK1-1 was shown to have GSH:CDNB conjugating activity over the pH range 6.0–10.5, with optimum activity at approx. pH 9 (Figure 4). The highest activity was observed in the pH range 8.0–9.5. Because of the rapid increase in non-enzymic compared with enzymic rates at higher pH, the pH optima of many cytosolic GSTs have not been rigorously investigated. In the present study, we found a consistent and measurable difference between non-enzymic and enzymic rates up to pH 10.

Sequence similarity searching

From matches generated by a PSI BLAST search of the hGSTK1 sequence against the non-redundant database (as at 4 April, 2003), the 100 highest scoring alignments were selected for further analyses. All matches identified in our results had E-values of less than 2×10^{-29}, showing that they were extremely unlikely to have occurred by chance. High E-values (> 1) indicate random matches. Of the 10 best matches ($> 60\%$ identity), nine represented human, mouse and rat GSTK1 sequences, and one (with 30% identity and 51% similarity) was with a hypothetical Kappa-class GST from the soil-borne eukaryote *Caenorhabditis elegans*.

Of the remaining 90% of matches, the PSI BLAST search identified significant similarity with sequences of HCCA isomerases, dithiol–disulphide isomerases (oxidoreductases), Dsb proteins and polyketide synthases, predominantly from different genera of soil-borne organisms such as *Pseudomonas*, *Ralstonia*, *Alcaligenes*, *Bradyrhizobium* and *Burkholderia*. The largest number of matches ($> 50\%$) was with HCCA isomerase sequences. HCCA isomerases specifically catalyse the isomerization of 2-hydroxychromene-2-carboxylate to *trans-o*-hydroxybenzylidenepyruvate in the naphthalene or analogous degradation pathways in micro-organisms. They exist as part of operon-encoded pathways induced in response to secondary metabolite biosynthesis, transport and polyhydrocarbon catabolism [3,32,33]. The hGSTK1 sequence shared 13–23% identity and 29–44% similarity with sequences in these groups.

Despite the use of this comprehensive search method, there were no matches to GST sequences from any of the other established classes.

A search of the Conserved Domain Database (CDD) (as at 2 April, 2003) with the hGSTK1 sequence showed that it possesses a potential C-terminal domain like that found in DsbA oxidoreductases. An alignment of proteins that contain this domain as well as a consensus sequence was automatically generated as part of the CDD search. The rGSTK1 and selected HCCA isomerase sequences were included in the alignment. Overall, little similarity existed between hGSTK1 and other non-GSTK1 sequences in this alignment (averages of 13–18% similarity), but hGSTK1 shared 38% identity with the consensus sequence.

Investigation of sequence-based relationships between the GST, HCCA isomerase and Dsb oxidoreductase enzyme families

To investigate the possibility that different permutations of relatedness might exist between cytosolic GSTs, HCCA isomerases and Dsb oxidoreductases, a BLAST pairwise comparison set at default parameters was carried out on a selection of representative sequences. They included five different Kappa-class GSTs, 24 cytosolic GSTs (two from each class), two soluble CLIC (chloride intracellular channel) proteins and two representatives of the MAPEG family, as well as six HCCA isomerase and five Dsb oxidoreductase sequences (detailed in Figure 5). The minimum BLAST E-values for each sequence pair were used in the comparison. A maximum E-score of 1000 was assigned if no result was returned by BLAST.

A box-plot of the results is presented in Figure 5. The distribution of BLAST E-scores strongly supports a homologous

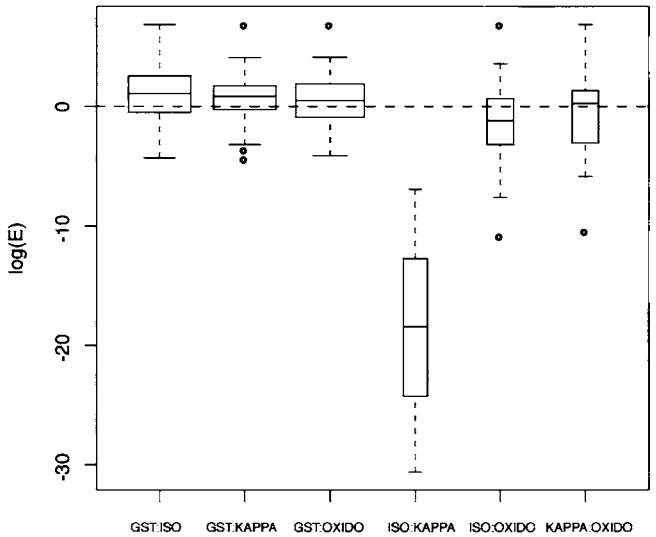

Figure 5 Analysis of sequence similarity between hGSTK1 and related enzymes

Box-plots of log-transformed BLAST E-scores are shown. The boxes represent the first and third quartiles of the distribution plot of comparisons within a group, and the line dissecting each box represents the median. Extensions external to the boxes represent the 5th and 95th percentiles of each distribution plot. Outliers are indicated as open circles. Box width is proportional to the square root of the number of E-values, or the number of pairwise comparisons between the functional classes. Abbreviations: ISO, HCCA isomerases; KAPPA, Kappa-class GSTs; OXIDO, Dsb oxidoreductases. Sequences selected for this study were as follows: gi3046817 GST Sigma *H. sapiens*, gi159854 GST Sigma *Ommastrephens sloanei*, gi183301 GST Mu2 *H. sapiens*, gi309278 GST Mu *Mus musculus*, gi31948 GST Pi *H. sapiens*, gi459939 GST Pi *Rattus norvegicus*, gi306809 GST Alpha *H. sapiens*, gi193703 GST Alpha *M. musculus*, gi6016173 GST Omega 1 (GSTO 1-1) *H. sapiens*, gi27819109 GST Omega 2 (GSTO 2-2) *H. sapiens*, gi21264427 GST Theta 1 *H. sapiens*, gi601918 GST Theta 2 *H. sapiens*, gi409146 GST Delta *Lucilia cuprina*, gi8034 GST Delta *Drosophila melanogaster*, gi2228731 GST Zeta *H. sapiens*, gi167968GST Zeta *Dianthus caryophyllus*, gi12007372 GST Epsilon *Anopheles gambiae*, gi1053076 GST *Proteus mirabilis*, gi1213565 GST *E. coli*, gi3650345 GST *Ochrobactum anthropi*, gi497788 GST Phi *Arabidopsis thaliana*, gi1524316GST Phi *Petunia x hybrida*, gi940381 GST Tau *A. thaliana*, gi11095992 GST Tau *A. thaliana*, gi12643390 CLIC1 *H. sapiens*, gi18490162 CLIC2 *H. sapiens*, gi12643338 GST Kappa cDNA from cord blood cell HDCMD47P *H. sapiens*, gi7643782 GST Kappa cDNA from a dendritic cell *H. sapiens*, gi3041680 GST Kappa *R. norvegicus*, gi3025325 putative GST Kappa *C. elegans*, gi3025313 putative GST Kappa *C. elegans*, gi3820518 HCCA isomerase PhnD *Burkholderia* sp. RP007, gi780657 HCCA isomerase *Rhizobium leguminosarum* bv.viciae, gi4220431 HCCA isomerase *Ralstonia* sp. U2, gi6225744 HCCA isomerase *Pseudomonas putida/aeruginosa* PAO1, gi5578710 HCCA isomerase *Sphingomonas* sp., gi2624890 disulphide oxidoreductase *V. cholerae* (virulence factor synthesis 1BED), gi1706523 thiol:disulphide interchange protein DsbS precursor *Erwinia chrysanthemi*, gi2392344 DsbA *E. coli* (1FVK – chain A), gi762927 Dsb protein *E. coli*, gi1263316 possible Dsb protein *Azotobacter vinelandii*.

relationship between the Kappa-class GSTs and HCCA isomerases. The minimum BLAST scores for all pairwise comparisons between these two groups lie well below a value of 1 (or $\log E < 0$). Comparison between the isomerases and oxidoreductases showed that they were the only other grouping with a majority of pairs exhibiting an E value of < 1, although the median for the Kappa-class GST–oxidoreductase comparison was close to $E = 1$. On the basis of sequence comparisons, the minimal similarity detected (E-values > 1) provides little support for a homologous relationship between any of the other subgroups. It is particularly interesting that the lowest similarity is between the Kappa-class sequences and those of all other soluble GSTs.

Structure prediction and homology modelling

To gain a greater understanding of the structure of hGSTK1, we undertook structure prediction and molecular modelling studies. The DsbA and TcpG oxidoreductases were chosen as representatives of the disulphide-bond-forming proteins, as they are well studied and their structures have been solved. No structures are available for the HCCA isomerases, so we selected a HCCA isomerase from *Pseudomonas* sp. (gi6225745) because it is of similar length (212 amino acids) to hGSTK1 and its potential role in bacteria had been investigated [33].

Due to low sequence identity ($< 25\%$), it was not possible to directly model hGSTK1 and HCCA isomerase sequences using DsbA and TcpG oxidoreductase structures as search models. However, the consensus sequence derived from the CDD alignment shared 33.8% identity with DsbA, 25.3% identity with TcpG, 38.3% identity with hGSTK1 and 40.2% identity with HCCA isomerase. This made it possible to first model the consensus sequence on the DsbA and TcpG structures to generate an intermediate model; the hGSTK1 and HCCA isomerase sequences were then modelled on the consensus intermediate.

C-terminal domain

An alignment of secondary structural elements of the DsbA and TcpG structures and predicted secondary structural elements of the consensus sequence, HCCA isomerase and hGSTK1 is presented in Figure 6. All C-terminal domains are of similar size (between 98 and 110 residues in length) and they possess the same $\alpha\alpha\alpha\alpha\beta\beta\alpha$ structural motif. Invariant glycines and prolines that are essential for configuring reverse turns in the DsbA and TcpG structures are readily aligned in the consensus, HCCA isomerase and hGSTK1 sequences; so are residues whose side chains form a hydrophobic patch adjacent to the active site in the DsbA and TcpG structures [27,29].

The complete sequences of hGSTK1 (226 residues), HCCA isomerase (212 residues) and the consensus (194 residues) were modelled using Swiss-Model, but co-ordinates were generated for only the C-terminal regions (110, 105 and 98 residues respectively). After iteratively fitting the models on to the DsbA and TcpG structures, the C-terminal domains of all proteins appeared to superimpose with good agreement. Relative similarity of fit was determined by pairwise measurements of the rmsd (root mean square deviation) for Cα atoms of selected residues (Table 2). All pairs of structures were very similar, having rmsd values in the range 0.3–1.1 Å and overall rmsd values of 0.5–0.6 Å. Models with an overall rmsd of $\sim$0.5 Å are considered to be essentially identical [34]. On closer examination, we noticed that the highly conserved Gly-Xaa-Pro turn containing the *cis* conformer of proline in DsbA and TcpG structures had been misaligned in the hGSTK1 model. This was due to continuous threading of hGSTK1 residues Ser^{149}, Ala^{150} and Glu^{151}, which do not have counterparts in the DsbA, TcpG, HCCA isomerase or consensus sequences (see Figure 6). Attempts to model these three residues as a small secondary loop were unsuccessful, so they were deleted from the hGSTK1 sequence and hGSTK1(Δsae) was used to generate an alternative model instead.

While the *cis*-proline turn in the hGSTK1(Δsae) model aligned with good agreement (rmsd 0.36 Å), the deletion changed the structural alignment of residues around the terminal β-strands (rmsd 3.93 and 1.02 Å; Table 2). The regional differences translate to greater overall displacement from the DsbA structure (rmsd 3.61 Å), but the configuration remains virtually identical over residues 110–185 (rmsd 0.58 Å), which represents the major portion of the C-terminal domain.

Superimposition of the hGSTK1(Δsae), HCCA isomerase and consensus models showing virtually identical domains is shown in Figure 7. The predicted similarity between the C-terminal domains of hGSTK1 and DsbA is shown in Figure 8. The

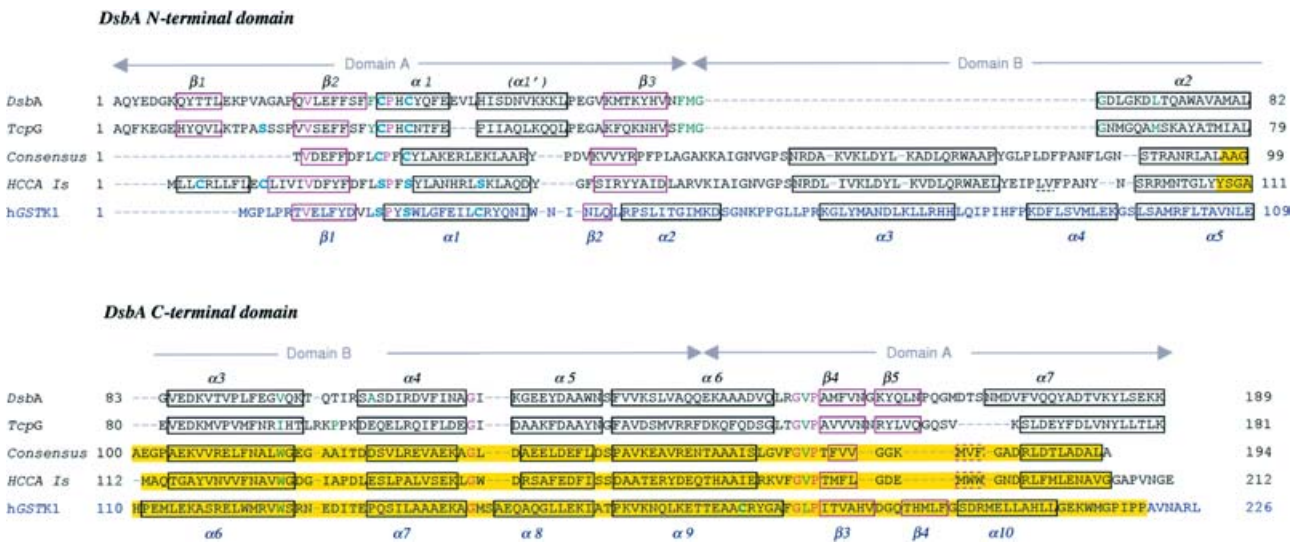

Figure 6 Alignment of sequences and secondary structural elements of *E. coli* DsbA and *V. cholerae* TcpG oxidoreductases with sequences and predicted secondary structural elements of the consensus sequence, a HCCA isomerase from *Pseudomonas* sp. (gi 6225745) and hGSTK1

Sequences and secondary structural elements of *E. coli* DsbA and *V. cholerae* TcpG oxidoreductases are from data in [29]. Helices (black) and strands (purple) are boxed, and secondary structural units of DsbA and hGSTK1 are labelled; strands in broken boxes were modelled but not predicted. A predicted two-residue strand (dotted underline) has not been included, as a minimum of three residues is required for strand formation. Modelled C-terminal domains of the consensus, HCCA isomerase and hGSTK1 sequences are highlighted in yellow. Cysteines (or serines in equivalent positions) are depicted in cyan. Invariant residues are pink. Amino acids forming a hydrophobic patch near the active site in the oxidoreductases and identifiable analogous residues in the consensus, HCCA isomerase and hGSTK1 models are depicted in green. DsbA domains A (thioredoxin fold) and B (helical domain) are indicated by arrows.

Table 2 Rmsd between Cα atoms of aligned residues in hGSTK1, HCCA isomerase and consensus models, and in DsbA as the reference structure

Secondary structural (SS) elements are numbered relative to the C-terminal domain. Numbers of residues compared are shown in parentheses. Rmsd calculations were not carried out for terminal helices, as differences in size and aa composition were too large. wt, wild-type enzyme; Δsae, hGSTK1 with residues Ser[149], Ala[150] and Glu[151] deleted from the sequence (see text for details).

						hGSTK1		
	DsbA	Consensus sequence		HCCA isomerase			Rmsd (Å)	
SS element	Residues	Residues	Rmsd (Å)	Residues	Rmsd (Å)	Residues	wt	Δsae
α-Helix 1 (14)	84–98	104–117	0.36	115–128	0.35	114–127	0.38	0.39
α-Helix 2 (10)	105–114	125–134	0.91	136–145	0.86	135–144	0.91	0.81
Gly reverse turn (1)	116	136	1.01	147	1.07	146	0.99	0.85
α-Helix 3 (8)	119–126	139–146	0.40	150–157	0.42	149–156	0.40	0.47
α-Helix 4 (16)	129–144	149–164	0.45	160–175	0.53	159–174	0.55	0.57
GXX reverse turn (3)	149, 150, 151	169, 170, 171	0.32	180, 181, 182	0.33	G[179], A[180], F[181]	0.32	–
						G[182], L[183], P[184]	–	0.36
β-Strand 1 (5)	152–156	172–176	0.37	183–186	0.38	182–186/185–189	0.40	3.93
β-Strand 2 (5)	158–162	178–182	0.55	189–193	0.50	188–192/195–199	0.48	1.05
Overall rmsd (90)	80–170	100–190	0.56	111–201	0.54	110–200	0.56	3.61
Overall rmsd (75)						110–185	0.58	0.57

largest displacement between the predicted models and the DsbA structure is $\sim$1.0 Å. This occurs at the glycine-mediated sharp reverse turn located at the C-terminal end of helix 4 of DsbA (Table 2 and Figure 8), which is not unexpected. Apart from these minor deviations, the overall rmsd value for all C-terminal domains (residues 110–185 in hGSTK1) relative to the DsbA structure is $<$0.6 Å, indicating that the models are essentially identical.

N-terminal domain

The N-terminal domains of DsbA, TcpG, hGSTK1, HCCA isomerase and the consensus sequence exhibit much greater complexity. Parts of the N-terminal domain are predicted to be structurally similar, while others are very different (see Figure 6). This inconsistency possibly accounts for the difficulty in modelling this region. The first section of the N-terminal region shows some structural similarity between the sequences in the form of a $\beta\alpha\beta$ motif with invariant valines and prolines around the analogous cysteine/serine residues. The N-terminal domain also ends with a predicted α-helix in all sequences, although the residue composition is dissimilar. The remaining features are highly variable. hGSTK1 and HCCA isomerase have serines where DsbA, TcpG and the consensus sequence have cysteines, precluding analogous disulphide bond formation. hGSTK1, HCCA isomerase and the consensus have extended

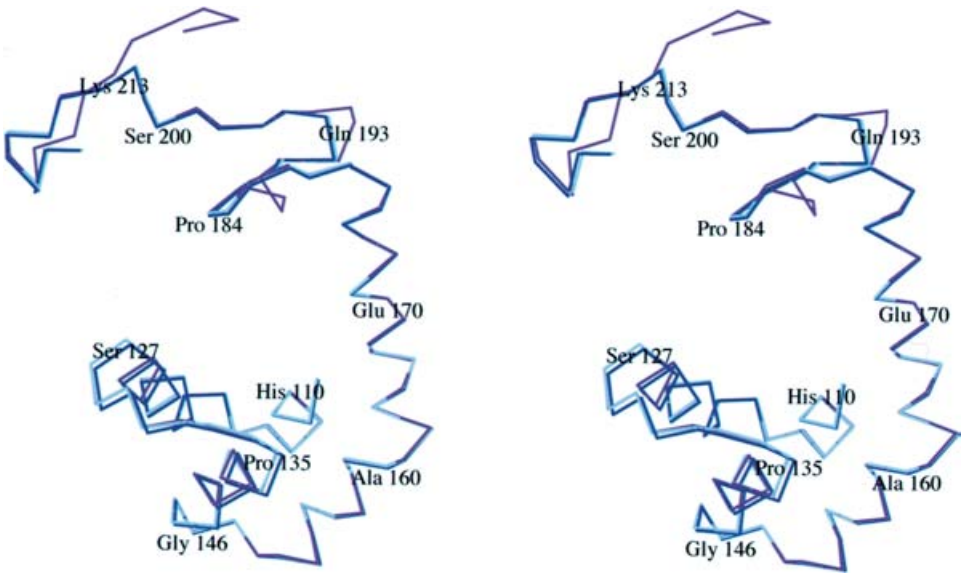

Figure 7 Comparison of modelled C-terminal domains

Shown is a MOLSCRIPT [43] stereoview of superimposed models of C-terminal domains of hGSTK1 (pink), HCCA isomerase (cyan) and the consensus sequence (dark blue). Selected residues in the hGSTK1 C-terminal domain have been labelled to show connectivity. hGSTK1 residues Gly^{146} and *cis*-Pro^{184} have also been labelled to indicate positions where sharp reverse turns occur in all models.

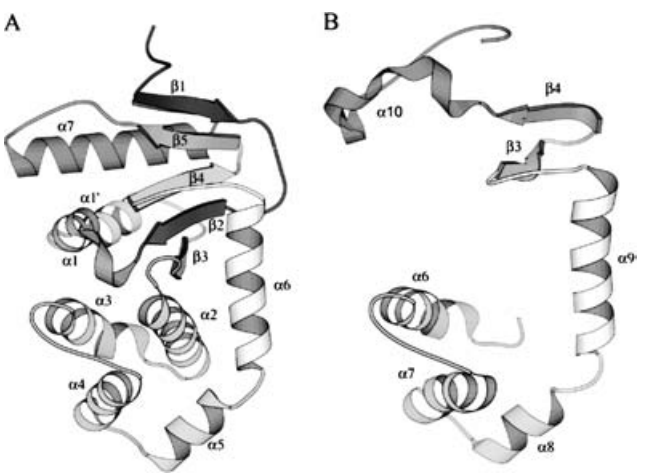

Figure 8 MOLSCRIPT [43] diagrams of DsbA oxidoreductase (A) and the C-terminal domain of hGSTK1 (B)

These schematic drawings show crystallographically determined secondary structural elements of DsbA oxidoreductase and those predicted to form in the C-terminal domain of hGSTK1. Secondary structures have been annotated to correspond to the sequences shown in Figure 6. Strands β4 and β5 of DsbA correspond to strands β3 and β4 in the hGSTK1 sequence; helices α3–α7 of DsbA correspond to helices α6–α10 in the hGSTK1 sequence.

polypeptides, with additional peptide sequences of 58, 47 and 52 residues respectively inserted between the $\beta\alpha\beta$ motif and the helix positioned at the end of the N-terminal region. Finally, HCCA isomerase and the oxidoreductases have N-terminal extensions, but these are differently configured into a helix and a strand respectively. A particularly interesting feature of these variations is that, of the 58-residue insert in hGSTK1, the entire 43 amino acids that have no equivalent in DsbA and TcpG sequences (see Figure 6) are encoded on a single exon (exon 3) in the *hGSTK1* gene (Figure 2).

Comparing topologies

Topology diagrams showing the connectivity of predicted secondary structural elements are shown in Figure 9. Diagrams of glutaredoxin, GST and DsbA [35] have been added to allow comparison. Conspicuous by its presence in all proteins is the prototypical thioredoxin fold within a glutaredoxin framework. Apart from this, and excepting glutaredoxin, all proteins possess an additional but variable helical domain. In the archetypical structure of soluble cytosolic GSTs, the helical domain is external to the thioredoxin fold and connected to the C-terminus, forming the motif $\underline{\beta\alpha\beta\alpha\beta\beta\alpha}\alpha_x$, where the underlined elements represent the N and C motifs of the thioredoxin fold (see Figure 9). In hGSTK1, HCCA isomerase and DsbA, the helical domain bisects the thioredoxin fold (renamed domain A in DsbA to accommodate the bipartite structure; Figure 6 and [27]), which rearranges the secondary structural elements into a $\underline{\beta\alpha\beta\alpha_x\beta\beta\alpha}$ motif. This helical domain, accordingly renamed domain B [27], is different in each protein. In hGSTK1 the helical domain is predicted to comprise 142 residues configured into eight helices, in HCCA isomerase it is 124 residues configured into six helices and in DsbA it has been shown to be 85 residues configured into five helices.

DISCUSSION

In this investigation, we present a detailed study of a human GST of the Kappa class. While the enzyme has some biochemical characteristics that are typical of GSTs, sequence analyses, secondary structure predictions and homology modelling suggest that hGSTK1-1 has a novel structure that differs significantly from that of other classes of soluble cytosolic GSTs.

Genetic characterization

The *hGSTK1* gene is very similar in size and organization to the Kappa-class GST genes described recently in mouse and

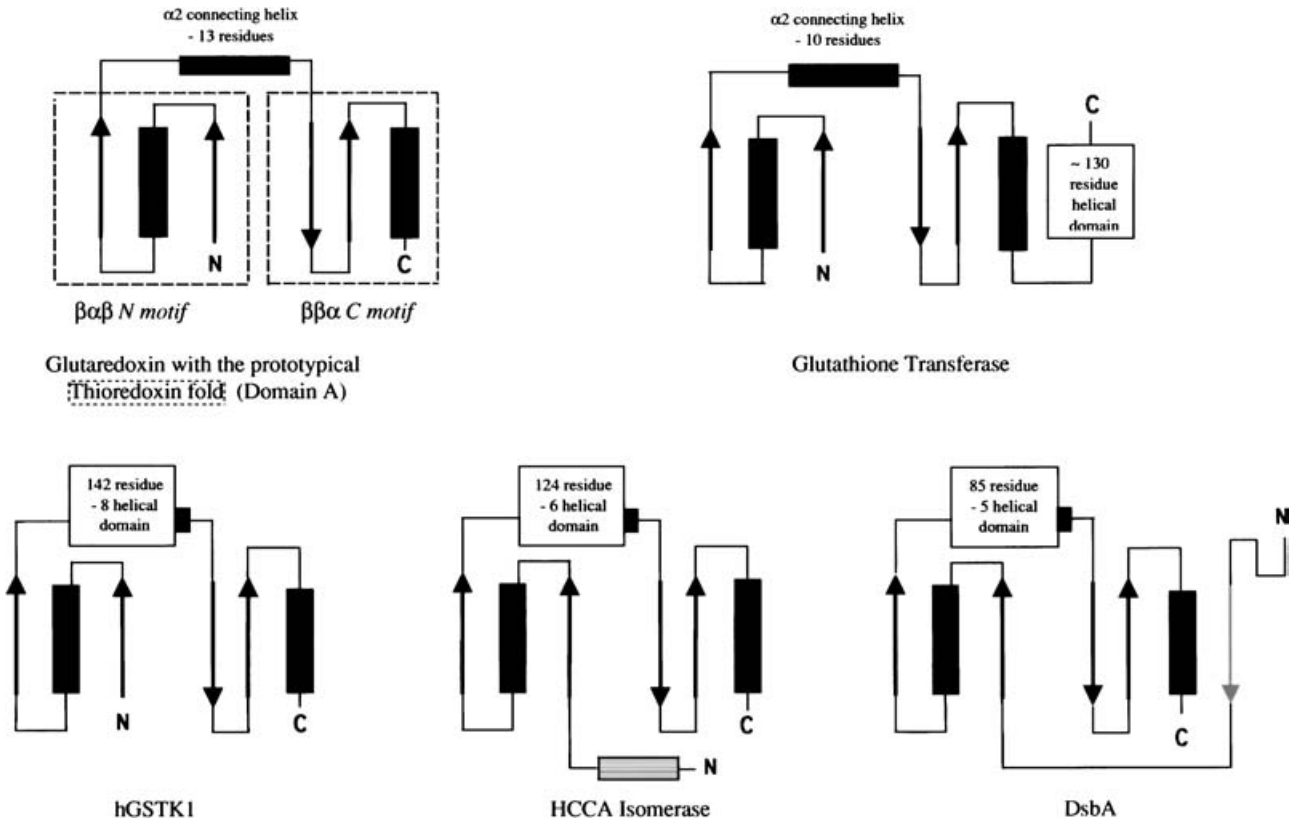

Figure 9 Topology diagrams comparing the predicted secondary structural arrangements of hGSTK1 and HCCA isomerase with those determined by crystal structure analysis of glutaredoxin, GST and DsbA oxidoreductase

β-Strands are drawn as arrows and α-helices as rectangles. The archetypical thioredoxin fold within the glutaredoxin framework is shown top left. The dashed lines indicate the separation of N-terminal ($\beta\alpha\beta$) and C-terminal ($\beta\beta\alpha$) motifs of the thioredoxin fold connected by the $\alpha 2$ helix. Equivalent features are shown in the GST and DsbA structures, and the predicted models of hGSTK1 and HCCA isomerase. Secondary structural elements that are not part of the thioredoxin fold are identified as separate domains or are coloured grey. The number of residues in the inserted helical domains of hGSTK1, HCCA isomerase and DsbA incorporate those resides that are analogous to the $\alpha 2$ connecting helix of glutaredoxin and thioredoxin. Adapted from Structure, vol. **3**, by J. L. Martin, "Thioredoxin – a fold for all reasons", pp. 245–250, © (1995), with permission from Elsevier.

rat [18]. Genes from all three organisms span coding regions of similar length, and are composed of eight exons and seven introns. Initiating and terminating codons are located in exons 1 and 8 respectively. The introns of *hGSTK1*, *mGSTK1* and *rGSTK1* are all mostly small (~100–600 bp), except for intron 5, which is significantly larger, ranging from 1.6 kb in rodents to 2.3 kb in humans. A search of the human genome database indicates that the *hGSTK1* gene is located on the long arm of chromosome 7 (7q.34–35), which is syntenic with the localization of *mGSTK1* on chromosome 6. We found no evidence supporting the existence of other functional Kappa-class genes or pseudo-genes in the human genome database.

Kappa-class GSTs show similar characteristics

Human, mouse and rat Kappa-class enzymes are all composed of 226 amino acids, with a subunit molecular mass of ~25.5–26.0 kDa. The protein sequences show a high degree of similarity (85%), but while mouse and rat sequences share 86% identity, these sequences share only 71% and 69% identity respectively with hGSTK1-1. The human, mouse and rat Kappa-class enzymes also have a similar substrate profile, being active towards predominantly halogenated aromatics. However, the levels of activity vary (Table 1; [18]).

Rat and mouse Kappa-class GSTs have been located primarily in mitochondria [16,18]. Our results using human white blood cells (A. Robinson, unpublished work) support these findings. However, it is not clear how the Kappa-class GSTs are translocated to the mitochondrial matrix. Some nuclear-encoded mitochondrial proteins have targeting signals within their sequences [36,37], but none have been identified in the Kappa-class GSTs so far. It was therefore potentially significant to find the *E. coli* hsp60 protein co-purifying with hGSTK1-1 during isolation and purification procedures. In *E. coli*, hsp60 functions to prevent misfolding and to promote refolding and proper assembly of unfolded polypeptides generated under stress conditions [38]. In eukaryotes, hsp60 is a multi-subunit protein that is associated with the transport and correct folding of proteins in the mitochondrial matrix [39]. As the two hsp60 proteins share 49% identity and 70% similarity, it is plausible to suggest that hGSTK1-1 may be translocated from the cytoplasmic ribosomes by the hsp60 chaperonin. This may account for the lack of any identifiable targeting signals in the primary sequence.

Kappa-class GSTs have characteristics of other enzyme families

Although hGSTK1-1 has typical GST activity, sequence alignments, C-terminal domain models and secondary structure prediction studies persuasively show that it shares greater

similarity with the HCCA isomerases and DsbA oxidoreductases than with GSTs. Whereas all known soluble GSTs possess a helical domain at the C-terminus which is external to the thioredoxin fold ($\beta\alpha\beta\alpha\beta\beta\alpha\alpha_x$), the helical domain in hGSTK1, HCCA isomerase and DsbA is inserted within the thioredoxin fold ($\beta\alpha\beta\alpha_x\beta\beta\alpha$). The different location of the helical domain alters the spatial arrangement of the β-strands in the polypeptide chain. In the cytosolic GSTs the four β-strands are all located in the N-terminal domain, separated only by a short connecting helix (Figure 9). In hGSTK1, HCCA isomerase and DsbA, the β-strands occur as pairs located at opposite ends of the polypeptide chain (see Figures 6 and 9). This placement would require a different folding pattern to recreate the β-sheet, and represents a significant departure from all other known GST structures.

A particularly interesting discovery in the present study was that the insert in the hGSTK1-1 N-terminal helical region is encoded entirely within a single exon, exon 3, of the *hGSTK1* gene (see Figures 2 and 6). This may be evidence of exon shuffling, which has been proposed to play a role in the creation of novel proteins.

Kappa-class GSTs and HCCA isomerases appear to form a third GST family

Overall, hGSTK1-1 exhibits strongest resemblance to HCCA isomerase. First, both enzymes are predicted to form a mixed protein of four strands and ten (hGSTK1) or nine (HCCA isomerase) helices in a $\beta\alpha\beta\alpha_x\beta\beta\alpha$ motif with essentially identical C-terminal domains. Secondly, both enzymes require much higher concentrations of glutathione for maximal activity, indicated by the relatively high K_m of 3.3 mM determined for hGSTK1-1 and the 2.5–5 mM GSH range determined for HCCA isomerases [32]. This is in direct contrast with other mammalian GSTs, whose K_m values are in the 100–200 μM range [40,41]. Thirdly, both enzymes show a pH optimum of $\sim$9. This value is not unexpected for enzymes occurring in soil bacteria, which have intracellular pH values of > 8.0, and is quite consistent with hGSTK1-1 being found in the mitochondrial matrix, where the pH has been shown experimentally to be $\sim$8.0 [42].

In view of such strong similarity with HCCA isomerases, it will be interesting to determine whether hGSTK1-1 can catalyse similar reactions. Investigation of HCCA and its analogues as potential substrates may provide insight into the role of hGSTK1-1 in humans. For example, the chromene nucleus exists as an intermediate in bacterial pathways that dissimilate complex polycyclic aromatic hydrocarbons and heterocyclic aromatic compounds which, by extension, may occur in humans. Chromenes are also found in many naturally occurring plant compounds, such as coumarins, and also in therapeutic agents, such as warfarin.

Conclusion

Although the Kappa-class GSTs are able to catalyse glutathione conjugation reactions, they appear to be structurally distinct from the soluble, cytosolic GSTs and the membrane-bound MAPEG enzymes. The presence of a thioredoxin fold in both the mitochondrial Kappa class and the cytosolic GST classes suggests that they derive from a common ancestor [35], although the relationship must be ancient if undetected at the level of sequence similarity. Alternatively, it is possible that independent evolutionary convergence to the same fold has occurred over time, providing only a functional relationship. If the solution of the crystal structure of a Kappa-class GST confirms the structure predicted here, it will be appropriate to reclassify the Kappa-class enzymes as a third family of GSTs.

We are grateful to Dr Peter Jeffrey and Dr Jun-Yao Fan for carrying out the sedimentation ultracentrifugation experiments. Our thanks also to Dr Peter Milburn and Cameron McCrae at the BRF for sequencing reactions, and to Majorie Coggan for technical support. We also extend our gratitude to work experience students Stephanie Day, Kate Dawson, Jenny Hartley, Alex Boling and Lorrae Webster for their assistance. This project is supported by a grant from the Australian Research Council.

REFERENCES

1 Sheehan, D., Meade, G., Foley, V. M. and Dowd, C. A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem. J. **360**, 1–16
2 Eaton, D. L. and Bammler, T. K. (1999) Concise review of the glutathione S-transferases and their significance to toxicology. Toxicol. Sci. **49**, 156–164
3 Vuilleumier, S. and Pagni, M. (2002) The elusive roles of bacterial glutathione S-transferases: new lessons from genomes. Appl. Microbiol. Biotechnol. **58**, 138–146
4 Fernandez-Canon, J. M., Baetscher, M. W., Finegold, M., Burlingame, T., Gibson, K. M. and Grompe, M. (2002) Maleylacetoacetate isomerase (MAAI/GSTZ)-deficient mice reveal a glutathione-dependent nonenzymatic bypass in tyrosine catabolism. Mol. Cell. Biol. **22**, 4943–4951
5 Board, P. G., Baker, R. T., Chelvanayagam, G. and Jermiin, L. S. (1997) Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochem. J. **328**, 929–935
6 Johansson, A. S. and Mannervik, B. (2001) Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. J. Biol. Chem. **276**, 33061–33065
7 Kanaoka, Y., Ago, H., Inagaki, E., Nanayama, T., Miyano, M., Kikuno, R., Fujii, Y., Eguchi, N., Toh, H., Urade, Y. and Hayaishi, O. (1997) Cloning and crystal structure of hematopoietic prostaglandin D synthase. Cell **90**, 1085–1095
8 Thomson, A. M., Meyer, D. J. and Hayes, J. D. (1998) Sequence, catalytic properties and expresssion of chicken glutathione-dependent prostaglandin D2 synthase, a novel class Sigma glutathione S-transferase. Biochem. J. **333**, 317–325
9 Listowsky, I. (1993) High capacity binding by glutathione S-transferases and glucocorticoid resistance. In Structure and Function of Glutathione Transferases (Tew, K. D., Pickett, C. B., Mantle, T. J., Mannervik, B. and Hayes J. D., eds.), pp. 199–209, CRC Press, Boca Raton
10 Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G. and Board, P. (2001) The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator. J. Biol. Chem. **276**, 3319–3323
11 Adler, V., Yin, Z., Fuchs, S. Y., Benezra, M., Rosario, L., Tew, K. D., Pincus, M. R., Sardana, M., Henderson, C. J., Wolf, C. R. et al. (1999) Regulation of JNK signaling by GSTp. EMBO J. **18**, 1321–1334
12 Cho, S. G., Lee, Y. H., Park, H. S., Ryoo, K., Kang, K. W., Park, J., Eom, S. J., Kim, M. J., Chang, T. S., Choi, S. Y. et al. (2001) Glutathione S-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. J. Biol. Chem. **276**, 12749–12755
13 Tomarev, S. I., Zinovieva, R. D. and Piatigorsky, J. (1992) Characterization of squid crystallin genes. Comparison with mammalian glutathione S-transferase genes. J. Biol. Chem. **267**, 8604–8612
14 Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. (2002) The Pfam protein families database. Nucleic Acids Res. **30**, 276–280
15 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) Nomenclature for human glutathione transferases. Biochem. J. **282**, 305–306
16 Harris, J. M., Meyer, D. J., Coles, B. and Ketterer, B. (1991) A novel glutathione transferase (13-13) isolated from the matrix of rat liver mitochondria having structural similarity to class Theta enzymes. Biochem. J. **278**, 137–141
17 Pemble, S. E., Wardle, A. F. and Taylor, J. B. (1996) Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. Biochem. J. **319**, 749–754
18 Jowsey, I. R., Thomson, R. E., Orton, T. C., Elcombe, C. R. and Hayes, J. D. (2003) Biochemical and genetic characterization of a murine class Kappa glutathione S-transferase. Biochem. J. **373**, 559–569
19 Porath, J. and Olin, B. (1983) Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron and nickel ions. Biochemistry **22**, 1621–1630
20 Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. **182**, 319–326
21 Tong, Z., Board, P. G. and Anders, M. W. (1998) Glutathione transferase zeta-catalyzed biotransformation of dichloroacetic acid and other alpha-haloacids. Chem. Res. Toxicol. **11**, 1332–1338

22 Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S. and Singh, R. M. (1966) Hydrogen ion buffers for biological research. Biochemistry **5**, 467–477
23 Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18**, 298–305
24 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402
25 Altschul, S. F. and Koonin, E. V. (1998) Iterated profile searches with PSI-BLAST – a tool for discovery in protein databases. Trends Biochem. Sci. **23**, 444–447
26 Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. **31**, 3381–3385
27 Martin, J. L., Bardwell, J. C. and Kuriyan, J. (1993) Crystal structure of the DsbA protein required for disulphide bond formation *in vivo*. Nature (London) **365**, 464–468
28 Bushweller, J. H., Billeter, M., Holmgren, A. and Wuthrich, K. (1994) The nuclear magnetic resonance solution structure of the mixed disulfide between *Escherichia coli* glutaredoxin(C14S) and glutathione. J. Mol. Biol. **235**, 1585–1597
29 Hu, S. H., Peek, J. A., Rattigan, E., Taylor, R. K. and Martin, J. L. (1997) Structure of TcpG, the DsbA protein folding catalyst from Vibrio cholerae. J. Mol. Biol. **268**, 137–146
30 Rost, B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. **266**, 525–539
31 Archibald, J. M., Blouin, C. and Doolittle, W. F. (2001) Gene duplication and the evolution of group II chaperonins: implications for structure and function. J. Struct. Biol. **135**, 157–169
32 Kuhm, A. E., Knackmuss, H.-J. and Stolz, A. (1993) 2-Hydroxychromene-2-carboxylate isomerase from bacteria that degrade naphthalenesulfonates. Biodegradation **4**, 155–162
33 Denome, S. A., Stanley, D. C., Olson, E. S. and Young, K. D. (1993) Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway. J. Bacteriol. **175**, 6890–6901
34 Chothia, C. and Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. EMBO J. **5**, 823–826
35 Martin, J. L. (1995) Thioredoxin – a fold for all reasons. Structure **3**, 245–250
36 Raza, H., Robin, M. A., Fang, J. K. and Avadhani, N. G. (2002) Multiple isoforms of mitochondrial glutathione S-transferases and their differential induction under oxidative stress. Biochem. J. **366**, 45–55
37 Rassow, J. and Pfanner, N. (2000) The protein import machinery of the mitochondrial membranes. Traffic **1**, 457–464
38 Chen, L. and Sigler, P. B. (1999) The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. Cell **99**, 757–768
39 Jindal, S., Dudani, A. K., Singh, B., Harley, C. B. and Gupta, R. S. (1989) Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. Mol. Cell. Biol. **9**, 2279–2283
40 Meyer, D. J. (1993) Significance of an unusually low K_m for glutathione in glutathione transferases of the alpha, mu and pi classes. Xenobiotica **23**, 823–834
41 Ketterer, B. and Christodoulides, L. G. (1994) Enzymology of cytosolic glutathione S-transferases. Adv. Pharmacol. **27**, 37–69
42 Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. and Tsien, R. Y. (1998) Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc. Natl. Acad. Sci. U.S.A. **95**, 6803–6808
43 Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **224**, 946–950

Received 30 October 2003/18 December 2003; accepted 7 January 2004
Published as BJ Immediate Publication 7 January 2004, DOI 10.1042/BJ20031656