Glutathione S-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue

Sally E. PEMBLE*, Anthony F. WARDLE and John B. TAYLOR

Cancer Research Campaign Molecular Toxicology Group, University College London, Windeyer Building, 46 Cleveland Street, London W1P 6DB, U.K.

We have isolated a cDNA clone that encodes rat glutathione S-transferase (GST) subunit 13, a GST originally isolated from rat liver mitochondrial matrix by Harris, Meyer, Coles and Ketterer [(1991) Biochem. J. **278**, 137–141]. The 896 bp cDNA contains an open reading frame of 678 bp encoding a deduced protein sequence of which the first 33 residues (excluding the initiation methionine residue) correspond to the N-terminal sequence reported by Harris *et al*. Hence like many other nuclear-encoded, mitochondrially located proteins, there is no cleavable mitochondrial presequence at the N-terminus. GST subunit 13 was originally placed into the Theta class of GSTs on the basis of sequence identity at the N-terminus; however, this is the only identity with the Theta class and in fact GST subunit 13 shows little sequence similarity to any of the known GST classes. Most importantly it lacks the SNAIL/TRAIL motif that has so far been a characteristic of soluble GSTs, although it does possess a second motif (FGXXXXVXXVDGXXXXXF) reported for GST-related proteins (Koonin, Mushegian, Tatusov, Altschul, Bryant, Bork and Valencia [(1994) Protein Sci. **3**, 2045–2054]. Southern and Northern blot analyses of rat DNA and mRNA are consistent with GST subunit 13's being the product of a single hybridizing gene locus. Searches of EST databases identified numerous similar human DNA sequences and a single pig sequence. We have derived a human cDNA sequence from these EST sequences which shows a high nucleotide similarity (77%) to rat GST subunit 13. The largest open reading frame is identical in length with subunit 13 and yields a deduced protein sequence identity of 70%. Most unusually the 3′ non-coding nucleotide sequence identity is also 77%. We conclude that these cDNAs belong to a novel GST class hereby designated Kappa, with the rat GST subunit 13 gene designated *rGSTK1* and the human gene being called *hGSTK1*.

INTRODUCTION

Glutathione S-transferases (GSTs; EC2.5.1.18) catalyse the conjugation of glutathione (GSH) to a variety of potentially genotoxic electrophiles and have been purified from various cellular fractions (reviewed in [1–3]). The soluble (cytoplasmic) GSTs of rats and humans comprise a supergene family of dimeric proteins whose subunits have been placed into five multigene families, namely Alpha, Mu, Pi, Theta and Sigma, according to their amino-acid sequence identities. These families have short regions of high identity in common and share a common evolutionary pathway [4–7]. In contrast, a membrane-bound GST from the microsomal fraction is structurally unrelated although it has a number of enzymic properties in common with the soluble enzymes, such as activity towards the 'universal' substrate CDNB [8].

A third subcellular localization of GST activity, namely the mitochondrial matrix, has been reported several times [9–11] and the purification and partial characterization of three mitochondrial GSTs have been reported by Kraus [12]. However it has been suggested that these activities may be due to cytosolic contamination [13]. Harris et al. [14] reported the purification of a GST isolated from the mitochondrial matrix of rat liver, GST 13-13, and on the basis of N-terminal analysis of the first 33 residues they placed this GST within the Theta family of cytosolic GSTs. This GST does not bind to GSH–agarose and when tested with a variety of standard GST substrates these authors found GST activity limited to conjugation with CDNB and ethacrynic acid. Since then, Addva et al. [15] have isolated a mitochondrial matrix GST from mouse liver. On the basis of its substrate profile and antibody cross-reactivity they have placed this GST within the Alpha class of GSTs. Here we report the cloning of GST 13-13 and show that it has features that are distinct from the cytosolic class Alpha, Mu, Pi, Sigma and Theta enzymes. Thus GST 13-13 is assigned to class Kappa and the rat and human enzymes are designated rGSTK1-1 and hGSTK1-1 in accordance with the nomenclature guidelines of Mannervik et al. [4] and Hayes and Pulford [2]. Apart from a short region of sequence identity at the N-terminus as noted by Harris et al. [14], there is no further primary sequence relationship to identify Kappa with another GST class.

MATERIALS AND METHODS

Materials

The rat liver cDNA library in λgt10 was prepared as described [16]. Restriction and modifying enzymes used for cloning procedures were obtained from Gibco–BRL and Pharmacia. NA45 paper was supplied by Schleicher and Schuell. Qiagen columns used in the purification of plasmid DNA were supplied by Hybaid (Teddington, Middx., U.K.). The oligo-labelling kit used to radiolabel probes was from Pharmacia. Chromospin columns and a rapid hybridization buffer (Expresshyb) were all from Clontech (Cambridge Bioscience). PCR reactions were performed with *Taq* polymerase; deoxyribonucleotides and reaction buffer IV were all supplied by Advanced Biotechnology. Gel-purified PCR products were cloned directly into the pGEM T-vector (Promega). Oligonucleotides were supplied by Oswel DNA

Abbreviations used: DAB, *N*,*N*-dimethyl-4-aminoazobenzene; GSH, glutathione; GST, glutathione S-transferase.

* To whom correspondence should be addressed at: Lablink Studio Ltd., 121c Cleveland Street, London W1P 5PN, U.K.

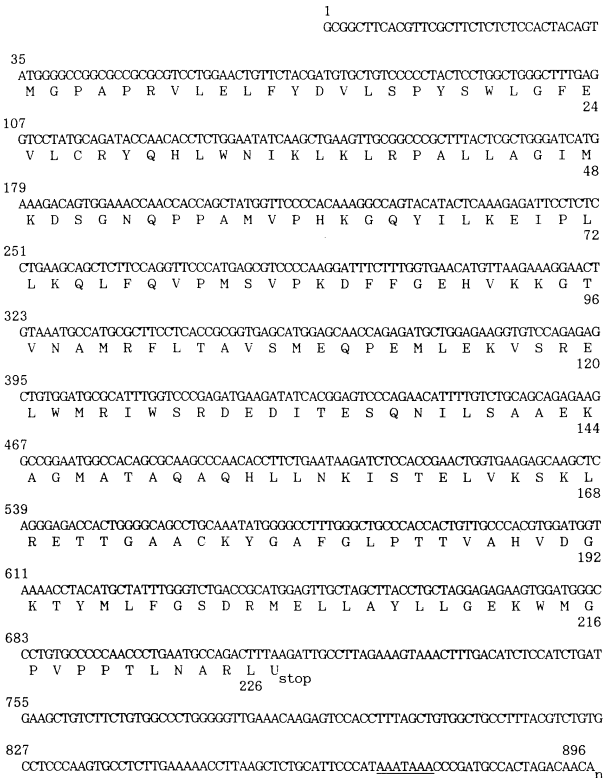

Figure 1 Nucleotide and deduced amino-acid sequence of rGSTK1

Numbers at the left of the figure refer to the DNA sequence; numbers at the right refer to the protein sequence. The polyadenylation site AAATAAA is underlined.

Service (University of Southampton, Southampton, Hants., U.K.). The Sequenase sequencing kit was supplied by USB. Nusieve agarose was from FMC. Radiolabels ([^{32}P]dCTP and [^{35}S]dATP) and nylon membranes (Hybond N plus) used in Northern blotting were from Amersham. All other chemicals were from Sigma or BDH.

Preparation of cDNA fragments by PCR

Rat liver first-strand cDNA was prepared as described [16], diluted 10-fold in sterile water and used directly in PCR. Rat liver λgt10 cDNA library lysate (10^9 plaque-forming units/ml) was prepared for PCR reactions as described [7]. cDNA (1 μl) or lysate (5 μl) was used for each PCR reaction, which was performed in a 50 μl volume of 10 mM Tris/HCl, pH 8.3, containing 200 mM ammonium sulphate, 0.1% Tween-20, 1.5 mM $MgCl_2$ (Advanced Biotechnology buffer IV), 0.2 mM deoxynucleotide triphosphates (all four) and *Taq* polymerase (2.5 units). Degenerate primers were used at a concentration of 100 pmol, and non-degenerate primers at a concentration of 20 pmol, per 50 μl reaction volume.

Analysis of PCR products

Depending on their sizes, PCR products were separated on 3% (w/v) Nusieve/1% (w/v) agarose, 2% (w/v) Nusieve/1% (w/v) agarose and 1.5% agarose gels. Specific bands were isolated by electroelution on to NA45 paper, eluted at 65 °C in 1.5 mM NaCl and recovered by ethanol precipitation. After centrifugation, DNA fragments were ligated into the pGEM T-vector, and transfected into *Escherichia coli* strain JM101. Recombinant plasmid DNA was prepared with Qiagen columns as described in the manufacturer's protocol. All sequencing was carried out with Sequenase and the protocol for double-stranded DNA sequencing described by the manufacturer.

Cloning of rGSTK1-1

Discovery PCR was performed with a 64-degenerate sense primer (A) and a 256-degenerate anti-sense primer (B). Primer A was a 17-mer (GAG/ACTNTTC/TTAC/TGAC/TGT) deduced from residues 8–13 (ELFYDV) of the N-terminal peptide sequence for rat GST subunit 13 [14]. The anti-sense primer B was a 15-mer (NAGG/ATGC/TTGG/ATANCG/T) deduced from residues 27–31 (RYQHL). The template was first-strand rat liver cDNA. The PCR cycling conditions used with these degenerate primers were: 94 °C, 4 min; 94 °C, 1 min/42 °C, 1 min/72 °C, 1 min (35 cycles); 72 °C, 5 min. This gave the predicted 72 bp DNA fragment, which was subcloned and sequenced.

Two more primers, one sense and one anti-sense, were designed to be complementary to segments of this DNA sequence to give the cDNA sequence for rGSTK1 in two overlapping clones. Rat liver poly(A)$^+$ RNA used to synthesize the double-stranded cDNA for the λgt10 library and the single-stranded cDNA for PCR were from the same source. The major part of the *rGSTK1* sequence was obtained by using an 18-mer sense primer N2377 (GATGTGCTGTCCCCCTAC) corresponding to the amino-acid sequence DVLSPY and a non-specific oligo-dT primer X996 ($GCGGCCGCT_{15}$) designed to hybridize to the poly(dA) tail of all the synthesized first-strand cDNAs. This primer pair gave an approx. 900 bp PCR product with rat liver first-strand cDNA as template. With rat liver λgt10 cDNA library as DNA template, an overlapping 130 bp PCR product containing the N-terminal sequence as well as some 5′ leader sequence was obtained with a forward primer for λgt10 (AGCAAGTTCAGCCTGGTTAGG) and an anti-sense primer N2379X (CATAGGACCTCAAAGCCCAG). The anti-sense primer corresponded to LGFEVLC within the rGSTK1 sequence. Both the 900 and 130 bp products were obtained by using the following PCR cycling conditions: 94 °C, 4 min/94 °C, 1 min/55 °C, 1 min/72 °C, 1 min (35 cycles); 72 °C, 5 min. Both PCR products were purified by gel electrophoresis and electroelution, recovered by ethanol precipitation and cloned into the pGEM T vector.

Preparation of radiolabelled probes

The DNA insert used to generate radiolabelled probes was prepared by PCR, with the 100-fold diluted cloned plasmids as DNA template and specific primers. The 900 bp insert from pMTA2 was regenerated by PCR with primer N2377 (sense) and P1958 (TTATGGGAATGCAGAGCTTAA) an anti-sense primer complementary to the end of the 3′ non-coding sequence. The 130 bp insert of pMTFC2 was regenerated by using the same primers used to isolate the fragment in the first instance, i.e. λgt10 forward primer and N2379X. Both DNA fragments were purified by two cycles of gel separation and electroelution. Probes were labelled with [^{32}P]dCTP with the Pharmacia oligo-labelling kit and purified with Chromospin-TE columns.

Northern and Southern blots

The preparation of Northern and Southern blots was as described previously [16] except that Hybond N was used as the blotting membrane. Expresshyb hybridization solution was used in accordance with the manufacturer's instructions for both Northern and Southern blots. Blots were washed several times in

2 × SSC/0.1 % SDS at room temperature, followed by two washes for 20 min each in 0.1 × SSC/0.1 % SDS at 50 °C (Northern) and 60 °C (Southern).

RESULTS

Isolation of cDNA clones for rat GST subunit 13, rGSTK1

With rat liver first-strand cDNA as DNA template, the degenerate primers A and B gave a 72 bp PCR product that was subsequently subcloned into the pGEM T-vector. Of the six recombinant clones selected, one of these, p70.2, contained an insert of 72 bp encoding the N-terminal sequence of rGSTK1, i.e. ELFYDVLSPYSWLGFEVLCRYQHL. The amino acid sequence not defined by the primer pair is underlined and reveals the identity of two residues, tryptophan (W) and cysteine (C) that were undetected in the original sequence, ELFYDVLSPYSXLGFEVLXRYQHL, of Harris et al. [14], presumably owing to low yields. As described in the Materials and methods section, this gave enough unequivocal DNA sequence to design specific primers, corresponding to DVLSPY and LGFEVLC, which generated PCR products of 900 and 130 bp respectively and were presumed to contain overlapping clones containing the C-terminus and N-terminus of rGSTK1 mRNA.

Six recombinants were selected for each terminal fragment and all were partly sequenced to identify positive clones. Three of the recombinants selected for the 900 bp fragment were positive and pMTA2 was selected for further analysis. For the 130 bp fragment, five of the six recombinants were positive and pMTFC2 was fully sequenced. Both pMTA2 and pMTFC2 were sequenced on both strands by using a mixture of oligonucleotide primers and partial subcloning. The sequences of these clones, as well as the original cloned fragment in p70.2, overlapped without any mismatches. The DNA sequence for rGSTK1 and its deduced amino acid sequence from the two clones are shown in Figure 1. The total cDNA is 896 bp long, contains an open reading frame of 678 bp including a putative initiation methionine codon at residues 35–37 and a 3′ non-coding sequence of 184 bp containing a single polyadenylation signal AAATAAA 19 bp from the poly(A) tail.

The putative initiation codon is assigned for several reasons. The first 34 nucleotides contain no other ATG and, in accordance with the Kozak consensus sequence, the DNA surrounding the putative initiation ATG has a purine at positions −3 and +4 (see Figure 1), which would optimize the translational efficiency [17]. Also, this methionine residue would be removed by aminopeptidase *in vivo*, resulting in the N-terminal sequence determined by Harris et al. [14]. The first in-frame stop codon is 678 bp downstream of this initiation codon, whereas the other two reading frames have seven and eleven stop codons in this region.

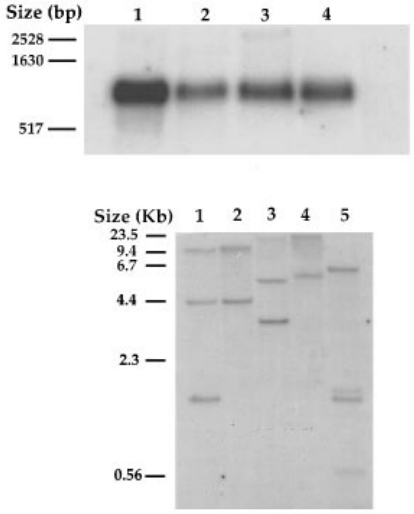

Figure 2 Northern analysis (upper panel) and Southern analysis (lower panel) of rat genomic DNA

Upper panel: poly(A)$^+$ RNA from rat liver, epididymis, DAB-induced hepatoma and heart was analysed under denaturing conditions in a 1.1 % (w/v) agarose gel. The RNA was transferred to Hybond-N membrane hybridized to ^{32}P-labelled DNA inserts derived from pMTA2 and pFC2 and autoradiographed for 4 days. Poly(A)$^+$ RNA (5 μg) was loaded in each lane as follows: lane 1, liver; lane 2, epididymis; lane 3, DAB-induced hepatoma; lane 4, heart. DNA size markers were from a mixed digest of pAT153 and processed in the same way as the RNA samples. Lower panel: genomic DNA isolated from the liver of a single rat was digested to completion with restriction nucleases as indicated below. DNA fragments were separated on a 0.7 % agarose gel, transferred to Hybond N, hybridized to ^{32}P-labelled inserts from pMTA2 and pFC2 and autoradiographed. The cDNA contains single internal sites for *Pst*1, *Bgl*2 and *Eco*RI. Lane 1, *Eco*RI; lane 2, *Hin*dIII; lane 3, *Bam*H1; lane 4, *Pst*1; lane 5, *Bgl*II. Size markers (kb) correspond to a *Hin*dIII digest of λ DNA.

```
rK1                                                    GCGGCTTCACGTTCGCTTCTCTCTCCACTACAGT
huK1                                                   ACT...CT..CGGAGC..G...C

rK1     ATGGGGCCGGCGCCGCGCGTCCTGGAACTGTTCTACGATGTGCTGTCCCCCTACTCCTGGCTGGGCTTTGAG
huK1    ........CCT.......AC.G....G..C.....T..C.............................C...

rK1     GTCCTATGCAGATACCAACACCTCTGGAATATCAAGCTGAAGTTGCGGCCCGCTTTACTCGCTGGGATCATG
huK1    A...G...C.G..T..GA.TA.......C.....C...C...........AGCC.CA.AA.A.........

rK1     AAAGACAGTGGAAACCAACCACCAGCTATGGTTCCCCACAAAGGCCAGTACATACTCAAAGAGATTCCTCTC
huK1    ..............A.G..T....G.C..C......G......A.TA.....GGCA..T..CT.AAAG...

rK1     CTGAAGCAGCTCTTCCAGGTTCCCATGAGCGTCCCCAAGGATTTCTTTGGTGAACATGTTAAGAAAGGAACT
huK1    ....GA..C.ATC.....A.......CCA.T..............GTC..TGATGC..G.A.......G.

rK1     GTAAATGCCATGCGCTTCCTCACCGCGGTGAGCATGGAGCAACCAGAGATGCTGGAGAAGGTGTCCAGAGAG
huK1    ..GTC.........T...........C....A.T.......T.................A.C....C.G...

rK1     CTGTGGATGCGCATTTGGTCCCGAGATGAAGATATCACGGAGTCCCAGAACATTTTGTCTGCAGCAGAGAAG
huK1    ............G.C.....AA.GA.......C.....C...C.G....G...CC..G.G..T.........

rK1     GCCGGAATGGCCACAGCGCAAGCCCAACACCTTCTGAATAAGATCTCCACCGAACTGGTGAAGAGCAAGCTC
huK1    ..T..T...T.TG...AA.........GGGA......G.A......G.A..GCC.AA........A.C.....
scrofa                                                                 1.........A.C.....

rK1     AGGGAGACCACTGGGGCAGCCTGCAAATATGGGGCCTTTGGGCTGCCCACCACTGTTGCCCACGTGGATGGT
huK1    .A............A...........G...C..A..........A.....T...C..G.....T........C
scrofa  .A............A...G.............................T......G..T...C......AC

rK1     AAAACCTACATGCTATTTGGGTCTGACCGCATGGAGTTGCTAGCTTACCTGCTAGGAGAGAAGTGGATGGGC
huK1    C.....C.....T.......C........G......C....G..GC.......G..G..............
scrofa  .....AC.......C.....C........G......C....G..AC.......G..C..............

rK1p     P  V  P  P  T  L  N  A  R  L  U(stop)
rK1     CCTGTGCCCCCAACCCTGAATGCCAGACTTTAAGATTGCCTTAGAAAGTAA----------------------
huK1    ...A.A..T...G..G.....................CGGAGG.AGC.AACTCTTCGTATAAAAAAAGC
scrofa  ........T...G.TG.A..............198

rK1     ----------------------------------------ACTTTGACATCTCCATCTGATGAAGCTGTC
huK1    AGGCCATCTGCTTAACCCTTGGTCCCACCATAAGGCACTGGG...CG..TT....T.......AG..G.A.T
scrofa                                                                285.T..

rK1     TTCTGTGGCCCTGGGGGTTGAAACAAGAG----------------------------TCCACCTTTAGC--
huK1    ..............A.C..TCTGTNTTTCCCCTACCCCCAAGGATGCCAGGAAGACG......A.....CA
scrofa  C..............A..308

rK1     TGTGGCTGCCTTTACGTCTGTGCCTCCCAAGTGCCTCTTGAAAAACCTTAAGCTCTGCATTCCCA-TAAATA
huK1    ......AA.......T...A......A.........T.CAG.G.G..CC..-T.....T......CA.....

rK1     AACCCGATGCCACTAGACAACA(n)
huK1    ...TA......TC..G..(n)
```

Figure 3 Comparison of the DNA sequences of *rGSTK1* with a human sequence derived from ESTs in the databases and a pig EST sequence

Databases screened were EMBL and GenBank with a BLASTN search program. The species source of each sequence is indicated on the left side of the figure where rK1 is *rGSTK1* (rat GST subunit 13), huK1 is *hGSTK1* and scrofa is pig EST. Dots indicate identical nucleotide residues; dashes indicate spaces inserted to maximize alignment; numbers at each end of the pig sequences refer to the numbering of GenBank sequence SSC1A02, which is 510 nucleotides long. The C-terminal 10 residues and the stop codon (TAA) of *rGSTK1* (rK1p) are shown to illustrate the conservation of the stop codon and the end tripeptide ARL. The polyadenylation sites (AAATAAA) are underlined. Note the high nucleotide sequence identity between the 3′ non-coding regions of the rat and human clones. This is also seen in a part of the pig sequence, residues 285–308.

```
rK1p   M G P A P R V L E L F Y D V L S P Y S W L G F E .
huK1p  . . . L . . T V . . . . . . . . . . . . . . . .  24

rK1    V L C R Y Q H L W N I K L K L R P A L L A G I M .
huK1p  I . . . . . N I . . . N . Q . . . S . I T . . .  48

rK1p   K D S G N Q P P A M V P H K G Q Y I L K E I P L .
huK1p  . . . . . K . . G L L . R . . L . M A N D L K .  72

rK1p   L K Q L F Q V P M S V P K D F F G E H V K K G T .
huK1p  . R H H L . I . I H F . . . . L S V M L E . . S  96

rK1p   V N A M R F L T A V S M E Q P E M L E K V S R E .
huK1p  . S . . . . . . . . N L . H . . . . . . A . . . 120

rK1p   L W M R I W S R D E D I T E S Q N I L S A A E K .
huK1p  . . . . V . . . N . . . . . P . S . . A . . . . 144

rK1p   A G M A T A Q A Q H L L N K I S T E L V K S K L .
huK1p  . . . S A E . . . G . . E . . A . P K . . N Q . 168

rK1p   R E T T G A A C K Y G A F G L P T T V A H V D G .
huK1p  K . . . E . . . R . . . . . Q . I . . . . . . . 192

rK1p   K T Y M L F G S D R M E L L A Y L L G E K W M G .
huK1p  Q . H . . . . . . . . . . . . . . . . . . . . . 216

rK1p   P V P P T L N A R L
huK1p  . I . . A V . . . .
                        226
```

Figure 4 Comparison of the deduced amino acid sequences of rGSTK1 and the human sequence hGSTK1

The identity comparison of deduced protein sequences for rGSTK1 (rK1p) and hGSTK1 (huK1p). Numbers at the right of the figure refer to the amino acid residue. Dots indicate that the amino acid is identical.

From the cDNA sequence it is predicted that the mature rGSTK1-1 protein would be 225 amino acids long with a calculated molecular mass of 25496 Da and a pI of 6.2.

Northern and Southern blot analyses

A Northern blot of mRNA isolated from normal rat liver, epididymis, *N*,*N*-dimethyl-4-aminoazobenzene (DAB)-induced hepatoma and heart was hybridized with the combined pMTA2/pMTFC2 probe (Figure 2, upper panel). A single cross-hybridizing band representing an mRNA of 1000–1100 nt was detected in each track. The signal obtained for mRNA isolated from liver was significantly stronger than for other tissues.

Southern blot analysis (Figure 2, lower panel) with the identical probe under high stringency showed only two cross-hybridizing bands for rat DNA digested with *Eco*R1 (11.5 and 5.4 kb) and *Hin*dIII (9.1 and 4.2 kb), three bands for *Bam*H1 (11.5, 5.0 and 3.3 kb) and *Bgl*II (9.4, 4.3 and 1.5 kb) and four for *Pst*I (5.6, 1.5, 1.35 and 0.5 kb). There is no internal restriction site within the cDNA for rGSTK1 for either *Hin*dIII, *Eco*RI or *Bam*H1; however, there are single internal sites for *Bgl*II or *Pst*1.

Identification of highly conserved sequences of other species

Analysis of EST databases revealed numerous short cloned sequences from human sourced mRNA and one from *Sus scrofa* (pig; Figure 3), all of which had a high degree of similarity to the cDNA for rGSTK1. With the exception of the first 24 bases all of the human sequence has been reported several times on both strands. The human sequence is highly similar to the cDNA for rGSTK1 over its full length, and its largest open reading frame has initiation and termination codons that are coincident with those of *rGSTK1*. This results in 70% identity between the deduced amino acid sequences (Figure 4). Interestingly the high degree of similarity between the two coding sequences (77.5%) is maintained over the 3′ non-coding sequences (78%). Similarly the *S. scrofa* sequence is highly similar to the other mammalian sequences, especially to the human.

DISCUSSION

Subunit size relative to other GSTs

The amino acid sequence deduced from the cDNA for rGSTK1 predicts a protein of 225 residues, not including the initiation methionine residue. This length is within the size range of mammalian soluble GST subunits because typically these are 218 or 219 residues for the Alpha class, 220 or 221 residues for the Mu class, 209 residues for the Pi class, and 239–244 residues for the Theta class [7]. This makes the rGSTK1 subunit somewhat shorter than those of class Theta and closer to the Mu and Alpha classes in length.

On the basis of the N-terminal protein sequence, rat GST subunit 13 was allocated to GST class Theta [14], yet with the exception of this short N-terminal identity we show here that the sequence similarity between rGSTK1 and other reported GSTs is low. GST subunits are divided into an N-terminal domain of mixed α-helices and β-sheets (called domain I; approx. the first 100 residues) and the remainder of the protein is called domain II and is essentially α-helical. These domains are connected by a short sequence of residues and each domain contains an amino acid motif found in most GSH-binding proteins [18]. In GSTs motif I is located approx. 60–80 amino acids from the N-terminus and forms part of the GSH-binding site. This motif has a core consensus sequence 'SNAIL/TRAIL' of which the isoleucine residue is nearly always present. However, in rGSTK1, with the exception of an isoleucine residue at position 70, the amino acid sequence in this area shows little similarity to other GSTs although computer prediction of the tertiary structure of rGSTK1 yields a mixture of α-helical and β-sheet structures for the N-terminal half of the protein. However, as for other GSTs, the remainder of rGSTK1, from residue 90 onwards, is solely α-helical and contains a region consistent with motif II, namely FGXXXXVXXVDGXXXXXF, which is centred on residue 190. Therefore motif I is absent from rGSTK1 whereas motif II is present; this is also true of the human sequence.

There are proteins that have been defined as GSTs on the basis of their ability to conjugate GSH but do not contain the SNAIL/TRAIL motif. The microsomal GSTs have activity towards the universal GST substrate CDNB, and rat leukotriene C synthetase has been reported to have GSH-conjugating activity towards DCNB [19]. However, neither of these proteins has any significant sequence identity with the cytosolic GSTs and they differ in the numbers and sizes of their subunits. Rat and human microsomal GSTs are trimeric proteins [20] whose subunits are only 155 residues long whereas human leukotriene C4 synthetase is a monomeric integral membrane protein 150 residues in length [21].

Hence rGSTK1 both shares and lacks characteristics of the cytosolic GSTs. Its size is typical of a soluble GST subunit (see above) and its active enzyme is a dimer [14]. It also has limited identity with class Theta at the N-terminus and it can be cut into two domains with structural characteristics typical of the soluble GSTs. Furthermore there occurs within domain II a peptide motif with characteristics in common with the soluble families. Yet *rGSTK1* and the similar human sequence (Figures 3 and 4) have no other significant identity with these families, which are characteristic of the membrane-bound GSTs. We therefore conclude that *rGSTK1*, the human sequence and the pig sequence, which has only been reported in part (Figure 3), are members of

a new class of soluble intra-organellar GSTs, which we name Kappa. On this basis the human gene would be designated *hGSTK1*, the subunit hGSTK1 and the protein hGSTK1-1; this nomenclature (see [2,3]) is used henceforth.

Relationship of *rGSTK1* to other gene loci

Like GST subunit 7 of the rat GST Pi family, it is probable that *rGSTK1* is the only member of the rat GST Kappa family. The rGSTK1 cDNA sequence contains single restriction sites for *Bgl*II and *Pst*1. Southern blot analysis with a mixture of the 900 and 130 bp fragments as probes under stringent conditions reveals only two or three hybridizing bands for rat DNA restricted with *Eco*RI, *Hin*dIII, *Bam*H1 and *Bgl*II, and four bands in a *Pst*1 digest (Figure 2). Likewise a Northern blot analysis of RNA from several tissues (Figure 2) shows only a single hybridizing band, which is consistent with a single gene in this family. At present it is unknown whether *hGSTK1* is the only member of the human Kappa family; however, use of the *rGSTK1* clone as a cross-hybridization probe yields only a single band in Northern blots of RNA from human tissues (results not shown).

Targeting to the mitochondria

Many proteins that are targeted to the mitochondrion have an N-terminal presequence that is cleaved when the protein enters the organelle. Such presequences are approx. 20 residues long and have some features in common, although there is no consensus amino acid sequence. Presequences are generally rich in positively charged and hydroxylated amino acids and have the capability of forming amphiphilic α-helices [22]. Harris et al. [14] reported that when the N-terminal protein sequence of GST subunit 13 was compared with the N-termini of Theta class GSTs, it contained a five-residue extension (GPAPR). This is very short for a mitochondrial presequence, and the possibility existed that there was an N-terminal peptide that had already been cleaved off and so would not be in the mature protein. However, the cDNA sequence shows that there is an ATG codon that fulfils the Kozak consensus immediately preceding this extension peptide. Because the N-terminal amino acid of the mature rGSTK1 subunit is a glycine residue (G), the preceding methionine residue encoded by the cDNA sequence must be cleaved off and can be considered to meet the requirements of an initiation methionine residue. Therefore because rGSTK1 does not contain a cleavable presequence but was purified from mitochondria, the targeting signal must lie within the mature protein sequence. This is not unique to rGSTK1 because a cleavable presequence is not an essential feature of targeting of proteins to the mitochondrion: there are numerous exceptions. These include all the outer-membrane proteins, some proteins of the intermembrane space and the inner membrane, and a limited number of matrix proteins such as isopropyl malate synthase, rat chaperonin 10 and the β-subunit of human electron-transferring flavoprotein [22–25]. The targeting information in all these cases is located within the protein, as seems to be true for rGSTK1.

Possible peroxisomal targeting sequence

Intriguingly, rGSTK1 terminates in the tripeptide ARL, which could function as a peroxisomal (microbody) targeting signal (reviewed in [26]). The term microbodies includes the peroxisomes in animal cells, glyoxysomes in plants and the glycosomes in certain parasites such as trypanosomes. Like mitochondrial proteins, microbody proteins are synthesized on free polysomes but are generally imported without proteolytic processing. It will therefore be interesting to establish whether GSTK1-1 is present in rat and human peroxisomes and, if so, whether it will yield clues to its endogenous substrates.

Other points of interest

Other observations are worthy of note, especially given the evidence that rGSTK1 is the only member of the rat GST Kappa family. First, we and others have suggested that the mitochondrial GST (i.e. Kappa) is the evolutionary precursor of GST class Theta, which itself is the precursor of classes Alpha, Mu, Pi and Sigma [6,7]. This divergent evolutionary tree was based in part by the assignment of rGSTK1 to class Theta on the basis of its N-terminal sequence [14], a practice that is shown to be doubtful by our cloning of the mitochondrial enzyme. The observation that class Kappa N-terminal sequences show sequence identity solely with class Theta but share motif II with all classes is consistent with this proposed evolutionary pathway. Other observations also favour divergent evolution, namely that the proteins are dimers composed of similar-sized subunits and that the subunits are split into two domains of similar tertiary structures. The absence of the SNAIL/TRAIL motif, because it is a negative, acts neither for nor against this argument. Although it is possible that all these features could result from convergent evolution (an argument best illustrated by their possession of motif II in common with several other GSH-binding proteins with no apparent structural similarities otherwise), it is the fact that the proteins have several features in common that favours a divergent evolutionary pathway. Hence the data are consistent with class Kappa's being the progenitor of class Theta and thus of the soluble GSTs. This point should be resolved when the crystal structures and enzyme mechanisms of several bacterial GSTs and of class Kappa GSTs are characterized and compared with GSTs containing the SNAIL/TRAIL motif.

Secondly, Harris et al. [14] originally reported the isolation of three mitochondrial GSTs of pI values 9.3, 8.9 and 7.5, with the greatest activity being associated with the GST of pI 9.3. It was this protein that was subsequently purified and N-terminally sequenced. However, this pI does not agree with the estimated pI for the deduced protein. The reason for the discrepancy is unknown, although the simplest explanation is that rGSTK1 is post-translationally modified, which would set this GST apart from other soluble families.

Thirdly, Figure 2 shows a Northern blot of mRNA isolated from normal rat liver, epididymis, DAB-induced hepatoma and heart, hybridized with the combined (900 and 130 bp) *rGSTK1* probe. A hybridizing poly(A)$^+$ RNA of 1000–1100 nt was observed in each track, a size consistent with the cDNA when allowance is made for a poly(A) tail. Although a Northern blot alone cannot be taken as directly quantitative, the band intensity obtained for the DAB-induced hepatoma was noticeably less than that for normal liver, presumably reflecting the loss of mitochondria and the utilization of glycolysis by the tumour.

Finally, there is no obvious substrate to be deduced from the sequence by comparison with other proteins in the databases other than the GSH-binding motif II in domain II. By comparison with known functions of cytosolic GSTs it is possible that peroxides produced by oxidative metabolism are substrates, and expression of the cDNA clone will facilitate the search for the role of the enzyme. It would seem that the role is a well-conserved one, given the remarkable conservation of the 3′ non-coding nucleotide sequence between man, rat and pig. This is suggestive of a regulatory role for the non-coding sequence and also itself suggests that the function of the protein is fundamental

and well regulated. Certainly it is known that loss of mitochondrial GSH levels always results in cell death unless the cell uses the glycolytic pathway for its energy production, as happens in many tumour cells. The possession of the highly conserved GST Kappa might be another aspect of this dependence on GSH-maintained redox stability, and it will be of value to determine the relevance of this novel GSH-utilizing enzyme.

We thank Dr. Laki Buluwela and Dr. Charles Pagel of Charing Cross and Westminster Medical School (London, U.K.) for providing and assisting with computer requirements, and the Cancer Research Campaign for financial support.

REFERENCES

1 Ketterer, B. and Christodoulides, L. G. (1994) Adv. Pharmacol. **27**, 37–69
2 Hayes, J. D. and Pulford, D. J. (1995) Crit. Rev. Biochem. Mol. Biol. **30**, 445–600
3 Mannervik, B. and Widersten, M. (1995) in Advances in Drug Metabolism in Man (Pacifici, G. M. and Fracchia, G. N., eds.), pp. 407–459, European Commission DG XII – Science, Research and Development, Eur 15439 EN, Luxembourg
4 Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jornvall, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7202–7206
5 Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991) Biochem. J. **274**, 409–414
6 Buetler, T. M. and Eaton, D. L. (1992) Environ. Carcinogen. Ecotoxicol. Rev. C **10**, 181–203
7 Pemble, S. E. and Taylor, J. B. (1992) Biochem. J. **287**, 957–963
8 Morgenstern, R., Guthenberg, C. and Depierre, J. W. (1982) Eur. J. Biochem. **128**, 243–248
9 Wahlander, A., Soboll, S., Sies, H., Linke, I. and Muller, M. (1979) FEBS Lett. **97**, 139–146
10 Jocelyn, P. C. and Cronshaw, A. (1985) Biochem. Pharmacol. **34**, 1588–1590
11 Botti, B., Ceccarelli, D., Tomasi, A., Muscatello, U. and Masini, A. (1989) Biochim. Biophys. Acta **992**, 327–332
12 Kraus, P. (1980) Hoppe-Seyler's Z. Physiol. Chem. **361**, 9–15
13 Ryle, C. M. and Mantle, T. J. (1984) Biochem. J. **222**, 553–556
14 Harris, J. M., Meyer, D. J., Coles, B. and Ketterer, B. (1991) Biochem. J. **278**, 137–141
15 Addva, S., Mullick, J., Fang, J.-K. and Avadhani, N. G. (1994) Arch. Biochem. Biophys. **310**, 82–88
16 Pemble, S. E., Taylor, J. B. and Ketterer, B. (1986) Biochem. J. **240**, 885–889
17 Kozak, M. (1987) Nucleic Acids Res. **15**, 8125–8148
18 Koonin, E. V., Mushegian, A. R., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Bork, P. and Valencia, A. (1994) Protein Sci. **3**, 2045–2054
19 Bach, M. K., Brashler, J. R., Peck, R. E. and Morton, D. R. (1984) J. Allergy Clin. Immunol. **74**, 353–357
20 McLellan, L. I., Wolf, C. R. and Hayes, J. D. (1989) Biochem. J. **258**, 87–93
21 Lam, B. K., Penrose, J. F., Freeman, G. J. and Austen, K. F. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 7663–7667
22 Glover, L. A. and Lindsay, J. G. (1992) Biochem. J. **284**, 609–620
23 Lill, R. and Neupert, W. (1996) Trends Cell Biol. **6**, 56–60
24 Ryan, M. T., Hoogenraad, N. J. and Hol, P. B. (1994) FEBS Lett. **337**, 152–156
25 Finocchiaro, G., Columbo, I., Garavaglia, B., Gellera, C., Valdameri, G., Garbuglio, N. and Didonato, S. (1993) Eur. J. Biochem. **213**, 1003–1008
26 De Hoop, M. J. and Ab, G. (1992) Biochem. J. **286**, 657–669

Received 18 April 1996/8 July 1996; accepted 9 July 1996