A subgroup of class α glutathione S-transferases

Cloning of cDNA for mouse lung glutathione S-transferase GST 5.7

Piotr Zimniak[a], Michael A. Eckles[a], Manju Saxena[b] and Yogesh C. Awasthi[b]

[a]Departments of Medicine and Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA and [b]Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555, USA

A full-length cDNA clone encoding the previously purified mouse glutathione S-transferase GST 5 7 [(1991), Biochem. J. 278, 793–799] has been isolated from a mouse lung cDNA library in λgt11. Sequencing of the clone revealed the presence of microheterogeneity in GST 5.7. Comparison of the deduced protein sequence with other glutathione S-transferases, together with previous information available on GST 5 7, indicates that the enzyme belongs to a novel subgroup within the α class of glutathione S-transferases. Members of the subgroup, which also include the rat GST 8-8 and perhaps chicken GST CL3, show high sequence homology with each other, but only moderate similarity to other α class enzymes They share a substrate specificity profile that resembles π-class enzymes, and are active in the conjugation of lipid peroxidation products.

Glutathione S-transferase; Cloning; cDNA; Sequence analysis; Mouse lung

1. INTRODUCTION

Glutathione S-transferases (GST; EC 2.5.1.18) of mammalian tissues are classified into three major classes, α, μ, and π [1,2]. Some GST isoenzymes not corresponding to either of these classes in regard to their immunogenicity and kinetic properties have also been described, and the possibility of other minor classes of GSTs being expressed has been suggested [3,4]. In earlier studies from one of our laboratories, a mouse lung GST isoenzyme immunologically and kinetically distinct from the α, μ, and π classes of GSTs was reported [5]. Sequence studies of CNBr fragments of this isozyme later revealed some homology between this mouse form and rat GST 8-8, indicating a structural similarity between these two isoenzymes [6]. This mouse GST isoenzyme, referred to as GST 5.7, has been shown to be expressed in liver [7] and lung; its kinetic characteristics indicate a relationship to both α- and π classes of GSTs [6]. Determination of the primary structure of this isoenzyme may, therefore, provide further insight into the structure–function relationship among GSTs. In the present communication, we report the nucleotide and derived protein primary structure of this mouse GST isoenzyme through the sequence of its cDNA, and discuss its structural and functional interrelationship with other mammalian GSTs.

Correspondence address P. Zimniak, University of Arkansas for Medical Sciences, mail slot 567, 4301 W Markham, Little Rock, AR 72205, USA. Fax: (1) (501) 671-2510.

2. MATERIALS AND METHODS

2.1. Materials

Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer at the Molecular Biology Core Facility of the Department of Biochemistry and Molecular Biology, University of Arkansas. Polymerase chain reactions (PCR) were carried out on a Hybaid thermal cycler. The pGEM-3zf(+) plasmid and restriction endonucleases were from Promega, Madison, WI. The λgt11 cDNA library from adult male BALB/c mouse lung was from Clonetech, Palo Alto, CA. The TA Cloning kit, containing the PCR vector, was from Invitrogen, San Diego, CA Sequencing was carried out using the Sequenase II kit from USB, Cleveland, OH; a kit for random primer labeling of DNA was purchased from the same supplier. [α-^{32}P]ATP was from DuPont-NEN, Boston, MA.

2.2. PCR and library screening

Primers for PCR were designed based on the N-terminal sequence of two CNBr fragments of purified mouse lung and liver GST 5.7 [6] with the help of organism-specific codon usage tables obtained from the file server at the European Molecular Biology Laboratory. Primer P1. based on a peptide sequence located in the C-terminal portion of the protein, was 5′-CTCICC(A/G)AA(A/G)TC(A/G)GA(G/C)AGIAC-3′. Primer P2, derived from a protein sequence located close to the center of the molecule, was 5′-AT(T/C)GC(C/T)GT(G/C)GC(T/C)CC(A/T/C)TT(T/C)AA-3′ The positions of the primers relative to the sequence are shown in Fig. 1. PCR was carried out with the two primers using total DNA isolated from a mouse lung cDNA library in λgt11 as the template The expected inclusive distance of the primers, approximately 220 bp, was estimated from the rat GST 8-8 sequence, to which the known fragments of GST 5.7 are homologous [6]. A 220 bp band, present among the products of the PCR reaction, was isolated from an agarose gel, labeled by the random priming method, and used for hybridization screening of the original library in λgt11. A positive clone was plaque purified, and an insert of approx. 950 bp was released from purified λ DNA by EcoRI digestion The insert was subcloned into EcoRI-cut pGEM-3Zf(+).

2 3 DNA sequencing

Restriction mapping of the cDNA fragment revealed the existence of two *Pst*I and a unique *Cla*I site. These sites, together with appropriate sites in the multiple cloning region of plasmid pGEM-3Zf(+) (*Pst*I and *Acc*I, which is end compatible with *Cla*I), were used to construct several partial clones for sequencing. Sequencing by the dideoxy method was carried out either directly on the plasmids, or on single-stranded DNA obtained after superinfection of transformed cells with a filamentous phage. Plasmids suitable for sequencing were prepared by the Triton/LiCl lysis method [8]. A modified sequencing protocol with combined labeling and termination reactions carried out at 48°C [9] was used. The pUC/M13 forward and reverse sequencing primers were employed except in one case, where the inverse complement to primer P2 was used.

2.4. Sequence analysis

Sequences were analyzed with the help of the computer program PCGene (Intelligenetics, Mountain View, CA); comparisons with published sequences were carried out by the Blast program [10] available as a network service from the National Center for Biotechnology Information, NIH.

3. RESULTS AND DISCUSSION

A mouse GST, designated GST 5.7, has been previously isolated from lung and liver [6]. The protein had a blocked N-terminus, but a partial internal amino acid sequence could be obtained by Edman microsequencing of two CNBr fragments [6]. Oligonucleotides have been designed based on parts of these sequences. The distance between the two oligonucleotides, estimated under the assumption that GST 8-8 can be used as a model for GST 5.7, should be approx. 220 bp. The oligonucleotides were used as primers for PCR, with total DNA from a mouse lung cDNA library as the target. The amplified 220 bp fragment was purified by agarose gel electrophoresis, subcloned into the PCR vector, and sequenced. According to the sequence data, the fragment was 225-bp long (including the two primers, which were 20 and 21 nucleotides) and yielded an uninterrupted protein only in one reading frame; the deduced protein sequence between, and exclusiding, the primers was 85% identical to the GST 8-8 sequence between residues 113 and 166 (data not shown). The primers were excluded since their design (and degeneracy) would influence the resulting protein sequence. The similarity to GST 8-8, as well as the presence of those parts of the peptide sequences that were not used for primer design, confirmed the identity of the partial clone, which was subsequently used for screening of the library. The screening yielded a clone of approx. 950 bp. The insert was liberated from λ DNA with *Eco*RI, subcloned into pGEM-3Z(f+), and restriction mapped. The map is shown in Fig. 1. Shorter fragments were prepared as described in Methods, and were subjected to sequencing as shown by the arrows in Fig. 1. Both strands were read at least once. The resulting sequence, together with its deduced translation, is shown in Fig. 2.

The cDNA sequence contains an open reading frame of 666 nucleotides. The ATG in position 1 has the ex-
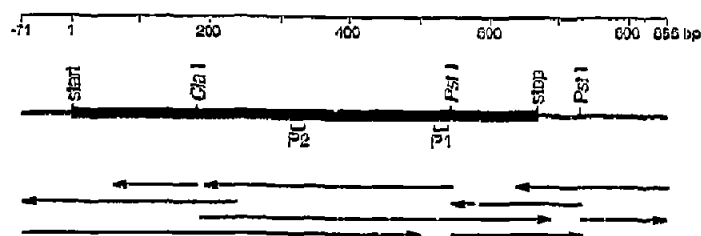

Fig 1. cDNA clone encoding mouse lung GST 5 7. The coding region is highlighted by the thick line. Restriction sites used in preparation of subclones for sequencing, and the location of primers used for initial PCR screening of the library, are shown. Arrows denote the sequencing strategy

pected consensus context for a eukaryotic initiation codon, in this case G in positions −3 and +4 [11]. The polyadenylation site, which begins 152-bp downstream from the stop codon (underlined nucleic acid sequence in Fig. 2), is of a rare type, namely ATTAAA rather than ATAAA [12]. The broken bars over the sequence in Fig. 2 denote the positions of the PCR primers.

The deduced amino acid sequence of 222 residues yields a protein of 25,576 Da. The underlined segments of the protein sequence in Fig. 2 denote the sequenced, N-terminal portions of the 12 and 6 kDa CNBr peptides [6]. The sequences deduced from the cDNA match the peptide sequences obtained previously [6] except for one residue in the central (12 kDa) peptide and three residues in the C-terminally located (6 kDa) peptide; the positions of these residues are marked with asterisks in Fig. 2. There are several likely reasons for this discrepancy. Microheterogeneity of GSTs is a well-known phenomenon. The isolated and the cloned enzyme may thus represent different variants of GST 5.7. A similar situation has been recently reported for chicken liver GST CL3 [13]. The expression of some GSTs is sex dependent [7] and is likely to differ in various strains of animals; in connection with that, it should be noted that GST 5.7 was originally purified from female CD-1 mice [6], whereas the cDNA library used in the present work was prepared from male BALB/c mice. Finally, the 6 kDa fragment of GST 5.7, which yielded the amino acid sequence with the majority of the variations (3 out of 18 residues), was purified from mouse liver [6], whereas the cDNA clone was isolated from a mouse lung library. Tissue-specific expression is one of the characteristic features of mammalian GSTs [14]. It should be noted that only 1 out of the 4 amino acid substitutions can be explained by a point mutation; in the other cases, a change in at least two nucleotides is required.

The calculated isoelectric point of the protein deduced from the cDNA sequence is 6.2 if, in keeping with previously obtained evidence [6], a blocked N-terminus is assumed. This value is somewhat higher than the experimental p*I* of 5.7 [6]. It should be kept in mind, however, that the p*I* calculation assumes that all

```
                    -71   CGAACCGCTTCTTTCTCCAGTGCCTGGAGACAAAATCCC  -32

ACAAGAATAACGAAGCTCTGAAGCACGAGTC ATG GCA GCC AAA CCT AAG CTC   21
                                 M   A   A   K   P   K   L     7

TAC TAC TTT AAT GGC AGG GGA CGC ATG GAG TCG ATC CGC TGG CTG   66
 Y   Y   F   N   G   R   G   R   M   E   S   I   R   W   L    22

CTC GCT GCG GCT GGA GTG GAG TTT GAG GAA GAA TTT CTT GAG ACA  111
 L   A   A   A   G   V   E   F   E   E   E   F   L   E   T    37

AGG GAA CAG TAT GAG AAG ATG CAA AAG GAT GGA CAC CTG CTT TTC  156
 R   E   Q   Y   E   K   M   Q   K   D   G   H   L   L   F    52

GGC GAA GTA CCC TTG GTT GAA ATC GAT GGG ATG ATG CTG ACA CAG  201
 G   E   V   P   L   V   E   I   D   G   M   M   L   T   Q    67

ACC AGG GCC ATC CTC AGC TAC CTC GCT GCC AAG TAC AAC TTG TAT  246
 T   R   A   I   L   S   Y   L   A   A   K   Y   N   L   Y    82

GGG AAG GAC CTG AAG GAG AGA GTC AGG ATT GAC ATG TAT GCA GAT  291
 G   K   D   L   K   E   R   V   R   I   D   M   Y   A   D    97

GGC ACC CAG GAC CTG ATG ATG ATG ATT GCC GTG GCT CCA TTT AAA  336
 G   T   Q   D   L   M   M   M   I   A   V   A   P   F   K   112

ACC CCC AAG GAA AAA GAG GAG AGC TAT GAT TTG ATA CTC TCA GGA  381
 T   P   K   E   K   E   E   S   Y   D   L   I   L   S   R   127

GCT AAA ACC CGT TAC TTC CCA GTG TTT GAA AAG ATT TTA AAA GAC  426
 A   K   T   R   Y   F   P   V   F   E   K   I   L   K   D   142

CAC GGA GAG GCT TTT CTC GTT GGC AAC CAG CTC AGT TGG GCA GAC  471
 H   G   E   A   F   L   V   G   N   Q   L   S   W   A   D   157

ATC CAG CTC CTA GAA GCC ATT TTG ATG GTG GAA CAA CTC AGT GCC  516
 I   Q   L   L   E   A   I   L   M   V   E   E   L   S   A   172

CCT GTA CTG TCC GAC TTC CCT CTG CTG CAG GCA TTT AAG ACA AGA  561
 P   V   L   S   D   F   P   L   L   Q   A   F   K   T   R   187

ATC AGC AAC ATT CCT ACA ATT AAG AAG TTC CTG CAA CCC GGA AGT  606
 I   S   N   I   P   T   I   K   K   F   L   Q   P   G   S   202

CAG AGG AAG CCT CCT CCA GAT GGC CCC TAT GTT GAG GTG GTC AGG  651
 Q   R   K   P   P   P   D   G   P   Y   V   E   V   V   R   217

ATT GTC CTG AAG TTC TAG TGCAGCGTGCTTTAAGGTGGCACCCACAAGuTTCC  704
 I   V   L   K   F   -                                       222

AATTGCAGTGTCACCACAGGCCAGCTGCAGAGCATTCCAGAAGCAAGGTATATAGATCC  763

CAGGAGTCAAGCTCTTCAAACAACAAAACCACTCTCCCCACAATGACAAATGCCAATTA  822

AATACAGTGGAAAACTAAAAAAAAAAAAAAAAA                           855
```

Fig. 2. Nucleotide sequence of the cDNA coding for GST 5.7, and deduced protein sequence. The polyadenylation signal is underlined in the nucleotide sequence. Amino acid sequence obtained previously [6] on CNBr fragments of the isolated protein are underlined in the protein sequence, with asterisks indicating the differences between the original sequences and the sequence deduced from the cDNA. The location of primers used for PCR screening of the library is shown by broken bars above the nucleotide sequence.

charged residues are fully accessible, and that their pKs are not modified by the protein environment. The calculated pI value is therefore only an approximation. In addition, replacing the amino acids deduced from the cDNA by the residues found previously by protein sequencing in the four positions marked in Fig. 2 reduces the calculated pI to 5.4. This is consistent with the existence of variants of the protein, as discussed above.

A comparison of the protein sequence derived from the cDNA with selected other GSTs is presented in Fig. 3. For rat liver chain 8, the sequence derived from cDNA [15] rather than the directly obtained protein sequence [16] was used in the alignment. GST 5.7 shows 91% identity to rat liver chain 8; if conservative substitutions are taken into account, the similarity becomes 94%. Thus, the two proteins resemble each other very

closely, and GST 5.7 is likely to be the mouse ortholog of rat GST 8-8. The latter enzyme has been classified as an α-type GST [15]. GST 5.7 is, however, only 60% homologous to the mouse Ya subunit, a typical example of a α GST (Fig. 3), and to a large group of other α-class enzymes from various sources (not shown). It should be stressed, however, that the classification of GSTs into the α, μ, and π groups, although convenient, is not always clear-cut. The criteria used for the classification include mainly substrate specificity, immunologic crossreactivity, and sequence information. It could be argued that existing GSTs present a continuum of variability in terms of these parameters, and their assignment into classes will by necessity remain arbitrary, especially near class boundaries. The recently cloned chicken GST CL3 [13] may serve as an illustration. GST CL3 and GST 5.7 share 69% identical residues (Fig. 3), a value that is greater than the homology between GST 5.7 and typical α-class enzymes, but smaller than the similarity of GST 5.7 and GST 8-8. By this criterion, GST CL3 could be considered an intermediate form between typical α GSTs and GST 5.7. This conclusion is supported by the analysis of particular regions of the proteins. For example, between residues 50 and 100 many pairs of amino acids are discernible which are identical in Ya and the chicken enzyme, while GSTs 5.7 and 8-8 share a different residue in the same position. On the other hand, regions can be found (and are marked with broken bars above the sequence) in which GSTs 5.7, 8-8, and CL3 are identical, but clearly different from Ya.

The differences in primary structure between the proteins compared in Fig. 3 are not evenly distributed along the chain. The most obvious variable regions are located between residues 100 and 130, and in the C-terminal portion of the proteins. A peptide likely to be part of the hydrophobic binding pocket (H-site) of the active center of the enzyme is located within the variable area at the C-terminus, at least in the rat subunit 1 [17]; the corresponding fragment is underlined in the mouse Ya sequence in Fig. 3. It is noteworthy that, within this area, GST 5.7, 8-8, and CL3 resemble each other more than form Ya, although in a less pronounced way than in the fragments labeled with the broken bars. This homology in sequence of the presumed substrate binding site could form the structural basis of the similarity in kinetic properties (see below).

Substrate specificity is often used for comparisons of GST isoforms. It is conspicuous that the activity toward ethacrynic acid relative to that for chlorodinitrobenzene is high for GST 5.7, 8-8, and CL3 [6,13,15], whereas it is low for typical α-class GSTs. Since high activity towards ethacrynic acid is characteristic for the π class of GSTs, GST 5.7 could be related to the π enzymes. However, GST 5.7 shows only 28% homology to the mouse π class GST P (alignment not shown). This, together with the lack of immunological cross-reactivity between

Fig. 3. Alignment of the deduced protein sequence of GST 5.7, the sequence of rat GST 8-8 [15], chicken GST CL3 [13], and mouse GST subunit Ya [18]. Only those residues that differ from GST 5.7 are shown. Broken bars above the sequence denote areas where GST 5.7, GST 8-8, and GST CL3 are identical but differ from Ya. The solid underline in the Ya sequence marks the area identified as part of the active center of rat chain 1 [17].

GST 5.7 and either α- or π-class enzymes, appears to place GST 5.7 between these major classes of GSTs.

In summary, we conclude that GST 5.7, along with GST 8-8 and CL3, should be classified as an α-type GST. However, we also postulate that GSTs 5.7 and 8-8, and to a lesser extent CL3, form a distinct subgroup within that class. GST 8-8 and probably GST 5.7 are involved in the conjugation of toxic products of lipid peroxidation [15], and thus play an important role in the metabolic response to oxidative stress and in the defense mechanisms against lung tumors. Ironically, the same proteins may render tumor chemotherapy ineffective by accelerating drug conjugation and clearance. The mouse is widely used as a model in both toxicology and oncology, which lends special importance to the study of mouse enzymes. Therefore, a better understanding of

the molecular biology of the function and regulation of GST 5.7, to which the results presented in this paper constitute a first step, should have biochemical as well as clinical implications.

REFERENCES

[1] Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M.K., Warholm, M. and Jornvall, H. (1985) Proc. Natl. Acad. Sci. USA 82, 7202–7206.
[2] Mannervik, B. and Danielson, U.H. (1988) CRC Crit. Rev. Biochem. 23, 283–337.
[3] Singhal, S.S., Ahmad, H., Sharma, R., Gupta, S., Haque, A.K. and Awasthi, Y.C. (1991) Arch. Biochem. Biophys. 285, 64–73.
[4] Meyer, D.J., Coles, B., Pemble, S.E., Gilmore, K.S., Fraser, G.M. and Ketterer, B (1991) Biochem. J. 274, 409–414.
[5] Medh, R.D. and Awasthi, Y.C. (1990) Clin. Chem. Enz Commun 3, 267–280.
[6] Medh, R.D., Saxena, M., Singhal, S.S., Ahmad, H. and Awasthi, Y.C. (1991) Biochem. J. 278, 793–799.
[7] Singhal, S.S., Saxena, M., Ahmad, H. and Awasthi, Y.C. (1992) Biochim. Biophys. Acta 1116, 137–146.
[8] He, M., Wilde, A. and Kaderbhai, M.A. (1990) Nucleic Acids Res. 18, 1660
[9] Schuurman, R. and Keulen, W. (1991) Biotechniques 10, 185
[10] Altschul, S.F, Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol Biol. 215, 403–410.
[11] Kozak, M. (1991) J. Cell Biol. 115, 887–903.
[12] Wickens, M (1990) Trends Biochem. Sci. 15, 277–281.
[13] Chang, L.-H., Fan, J.-Y., Liu, L.-F., Tsai, S.-P and Tam, M.F (1992) Biochem. J. 281, 545–551.
[14] Tu, C.-P, Weiss, M.J., Li, N.Q and Reddy, C.C. (1983) J. Biol. Chem. 258, 4659–4662.
[15] Stenberg, G., Ridderstrom, M., Engstroem, A., Pemble, S.E. and Mannervik, B. (1992) Biochem. J. 284, 313–319.
[16] Alin, P., Jensson, H., Cederlund, E., Jornvall, H. and Mannervik, B. (1989) Biochem J. 261, 531–539.
[17] Hoesch, R.M. and Boyer, T.D. (1989) J. Biol Chem. 264, 17712–17717.
[18] Pearson, W.R., Reinhart, J., Sisk, S.C. and Anderson, K. (1988) J. Biol. Chem. 263, 13324–13332.