The structure of the human glutathione *S*-transferase π gene

Ian G. COWELL,* Kathy H. DIXON, Sally E. PEMBLE, Brian KETTERER and John B. TAYLOR

Cancer Research Campaign Molecular Toxicology Research Group, University College and Middlesex School of Medicine, Windeyer Building, Cleveland Street, London W1P 6DB, U.K.

The human glutathione *S*-transferase π gene has been isolated from a cosmid library. The gene spans approximately 3 kb, is interrupted by six introns and the region around its 5' end has the high G + C and CpG content typical of an HTF (*HpaII* tiny fragment) island. In addition to a TATA box at position –28 relative to the transcription initiation site and two G + C boxes (GGGCGG), the promoter contains a sequence motif matching the phorbol ester- and *ras*-responsive element from the polyoma virus enhancer.

INTRODUCTION

The glutathione *S*-transferases (GSTs, EC 2.5.1.18) are a group of dimeric isoenzymes exhibiting a number of catalytic activities, including the conjugation of glutathione with hydrophobic electrophiles and the reduction of fatty acid and DNA hydroperoxides [1, 2]. In addition, the GSTs bind a number of hydrophobic compounds, such as haem and bilirubin. The GSTs of man and rat may be divided into three classes [3] on the basis of the primary structure of their component subunits [4]. The human enzyme GST- π [5] and the rat GST 7-7 (GSP-P) [6] fall into the π class and are sufficiently closely related at the nucleotide level to allow their cross-hybridization in Northern-blot analysis [7].

GST- π was first reported as an anionic GST from human placenta [8], and has subsequently been shown to be present in a wide range of tissues, including lung [9, 10], kidney [11] and prostate [12], although its expression is low or absent in liver [13]. A number of reports have shown tumour-specific increases in the levels of expression of the GST π family, for example, in the rat GST 7-7 is barely detectable in normal liver, but is expressed constitutively in hyperplastic nodules and hepatocellular carcinoma induced by chemical carcinogens [14, 15]. A number of human tumours [10, 16–19], including cervical carcinoma and some lung carcinomas, have been shown to express raised levels of GST- π compared with the corresponding normal tissues, although unlike its analogue in the rat, GST- π is not expressed in hepatocellular carcinoma. In addition, it has been shown that in some human cell lines selected for their multi-drug resistant phenotype, GST- π is expressed at elevated levels [20–22].

To study the molecular mechanisms involved in the regulation of the GST- π gene in relation to its increased expression in tumours and drug-resistant cell lines, we have isolated the GST- π gene to analyse the promoter and upstream elements for regulatory sequences. We report here the structure and sequence of the GST- π gene and discuss some potentially important features of its promoter.

MATERIALS AND METHODS

Cell culture

The cosmid library was constructed with genomic DNA prepared from the acute lymphocytic leukaemia cell line HPB-ALL [23]. RNA for primer extension experiments was prepared from the lung non-small-cell carcinoma cell line BEN [24] and from HeLa cells.

Screening of the human genomic library

A human genomic cosmid library of approximately 2×10^5 independent recombinants was a gift from D. Kioussis [25]. The library was screened with a ^{32}P -labelled 0.7 kb *EcoRI* fragment from plasmid pGSTr7 [7], which contained the entire coding region of rat GST subunit 7 cDNA. Hybridization was carried out in $6 \times \text{SSC}$ (0.9 M-NaCl, 0.09 M-sodium citrate), 0.02% (w/v) bovine serum albumin (BSA), 0.02% (w/v) polyvinylpyrrolidone, 0.02% (w/v) Ficoll, 0.1% (w/v) SDS, 100 μg of herring sperm DNA/ml, 20 μg of poly (A)/ml and ^{32}P -labelled cDNA probe (2×10^8 c.p.m./ μg) at approximately 10^6 c.p.m./ml, for 12 h at 60 °C. Filters were washed twice for a total of 2 h with $0.2 \times \text{SSC}$, 0.1% (w/v) SDS at 60 °C.

DNA sequencing

Three overlapping restriction fragments, a 5.2 kb *HindIII* fragment, a 2.3 kb *BamHI* fragment and a 3.5 kb *BglII* fragment, which each hybridized to the cDNA probe, were isolated from cosGSTrP5 (see Fig. 1.) and digested further with several restriction enzymes to generate sets of smaller overlapping fragments, which were cloned into M13mp10 and M13mp11. Sequence data were determined by the method of Sanger *et al.* [26]. All of the sequence data presented were determined from both strands.

Primer extension

Total cellular RNA was isolated from cultured BEN and HeLa cells by the method of Chirgwin *et al.* [27].

Abbreviations used: GST, glutathione *S*-transferase; HTF, *HpaII* tiny fragment; BSA, bovine serum albumin; TPA, 12-*O*-tetradecanoylphorbol 13-acetate; TRE, TPA-responsive element.

* To whom correspondence should be addressed.

The sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers: GST- π gene exons 1–5, X08094; exon 6, X08095; exon 7, X08096.

Primer extension was carried out by using an end-labelled oligonucleotide primer which was hybridized with 10 μ g of poly (A) or 50 μ g of total RNA in 500 mM-NaCl/10 mM-Pipes, pH 6.4, at 55 °C for 6 h. The primer was extended with AMV reverse transcriptase at 42 °C for 1 h. The products of the reaction were run on a 6 % acrylamide, 8 M-urea sequencing gel, together with a sequencing ladder as size markers.

Southern-blot analysis

Human leukocyte DNA was digested with *Hind*III and 20 μ g of the digest was run on a 0.8 % agarose gel and blotted onto a Hybond-N membrane (Amersham). Hybridization was carried out in 6 $\times$ SSC, 0.1 % (w/v) SDS, 0.02 % (w/v) BSA, 0.02 % (w/v) polyvinylpyrrolidone, 0.02 % (w/v) Ficoll, 100 μ g of herring sperm DNA/ml, 20 μ g of poly(A) ml, and ^{32} P-labelled GST- π cDNA (2×10^8 c.p.m./ μ g) (K. Dixon, unpublished results) for 18 h at 65 °C. The filter was washed in $0.1 \times$ SSC, 0.1 % (w/v) SDS at 65 °C.

RESULTS AND DISCUSSION

Isolation of genomic clones

Using rat GST subunit 7 cDNA [7] as a probe, eight cross-hybridizing human genomic clones were isolated from a cosmid library. Restriction analysis revealed that three different inserts were present in the eight hybridizing cosmids. A number of common restriction fragments could be generated from each of the cosmids and mapping of the restriction sites showed that the three different genomic clones formed an overlapping set, spanning approximately 60 kb (see Fig. 1). In addition, restriction mapping and Southern-blot analysis of the cosmid DNAs indicated that each of the cosmids contained the whole of the sequence hybridizing to GST subunit 7 cDNA. However, Southern-blot analysis of genomic DNA revealed, on long exposure, fragments

Fig. 1. Organization of the GST- π locus

(a) Restriction maps of the three GST- π gene containing cosmids, cosGSTrP5, cosGSTrP6 and cosGSTrP7, with *Hind*III sites represented by vertical lines; the solid box corresponds to the region of complementarity to the rat GST subunit 7 cDNA. (b) Restriction map and intron/exon arrangement of the GST- π gene, exons are shown as solid boxes numbered 1–7, restriction sites for the enzymes *Bgl*II (Bg), *Sac*II (S), *Pst*I (P), *Hind*III (H) and *Bam*HI (B) are shown above.

Fig. 2. Southern-blot analysis of human leukocyte DNA

Genomic DNA (20 μ g) was digested with *Hind*III and electrophoresed on an 0.8 % agarose gel. Following transfer to a Hybond-N filter (Amersham), hybridization was carried out with ^{32} P-labelled GST- π cDNA under the conditions described in the Materials and methods section.

hybridizing to GST- π cDNA which could not be accounted for by the cosmid DNA. For example, *Hind*III digestion of human leukocyte DNA (see Fig. 2), resulted in two strongly hybridizing fragments of 5.2 kb and 6.8 kb, which could be accommodated by the restriction map of the cosmids (see Fig. 1), but a more weakly hybridizing 3.3 kb fragment was also detected and this was not present in the genomic clones that were isolated. It is thought that the third band probably represents a GST- π pseudogene, which was not detected in the cosmid library due to its lower degree of homology to the rat GST subunit 7 cDNA used in the selection procedure. In comparison, the rat has four processed pseudogenes in addition to the single expressed GST subunit 7 gene [28].

Sequence analysis

The coding sequence (see Fig. 3) of the gene displays a high degree of homology with that of rat subunit 7 [6, 28] and encodes a peptide identical to that deduced from the GST- π cDNA sequence described by Kano *et al.* [5]. Like its homologue in the rat, the GST- π gene is interrupted by six introns (see Fig. 1b), each possessing the canonical GT and AG splicing signals at their 5' and 3' ends respectively. The first exon has a length of 33 bp and terminates at the 3' end, within the codon of the initiation methionine.

Vol. 255

Fig. 4. Determination of the site of transcription initiation

RNA prepared from either BEN or HeLa cells was used in a primer extension experiment using an anti-coding oligonucleotide primer derived from the last base of exon-1 and the first 19 bases of exon-2. Lane 1, size markers; lane 2, BEN poly(A)⁺ RNA; lane 3, HeLa total cellular RNA; lane 4, BEN poly(A)⁺ RNA. The numbers correspond to the size markers, the largest primer extension product is indicated by an arrow.

Determination of RNA start site

The transcription initiation site of the GST- π gene was determined for RNA prepared from BEN and HeLa cells. By using an oligonucleotide primer (5'-ACCACGGTGTAGGGCGGCAT-3') derived from the 5'-portion of exon-2 and the last base of exon-1 of the GST- π gene, poly(A)⁺ or total RNA produced bands corresponding to an initiation site between 28 and 33 bp downstream of the TATA box (see Fig. 4). The largest extension product was taken to represent the RNA start site, while CAP and secondary-structure effects may have accounted for the smaller bands. However, the possibility of multiple start sites cannot be excluded. At 28 bp downstream of a TATAA sequence there is a G residue surrounded by pyrimidine bases, which fits the consensus found by Breathnach & Chambon [29] for the transcription initiation site of a number of eukaryotic genes, although, in this consensus, transcription was found to start with an A residue more frequently than G.

Presence of an HTF island covering the 5' ends of the gene

There is a strong bias towards G+C residues in the region surrounding the 5' end of the gene; from positions -100 to +300 the G+C content reaches 72% and the

proportion of CpG dinucleotides is 9.2%, which is much higher than that found in the mammalian genome as a whole [30]. This nucleotide and dinucleotide composition is diagnostic of an HTF (*HpaII* tiny fragment) island. HTF islands, which are regions of constitutive hypomethylation of CpG dinucleotides, are associated with many genes, particularly those with a so-called house-keeping function [30]. Another feature of interest is a repetitive region close to the 5' end of the gene, this is approximately 110 bp in length, and consists entirely of A and T residues.

Analysis of the upstream region of the GST- π gene and the presence of a putative phorbol ester- and *ras*-responsive element

As well as the TATAA sequence positioned at -28 to -24 from the transcription initiation site there are two GC boxes [31] at positions -43 to -38 and -53 to -48 with the sequence GGGCGG. This motif has been identified at similar positions in the promoters of a number of cellular genes as well as in the 21 bp repeats of the SV40 genome (see ref. [32] for review), and both of these boxes precisely match the consensus sequence 5'(G/T)GGGCGG(G/A)(G/A)(C/T)-3' for the binding site of the transcription factor SP1. This sequence is also present in the promoter of the GST subunit 7 gene, but only as a single copy.

Just upstream of the GC boxes, at position -59 to -66, is the sequence TGA CT CAG which corresponds to the consensus sequence [33] found in promoters of various genes that are responsive to phorbol esters [such as 12-*O*-tetradecanoylphorbol 13-acetate (TPA); this motif has been called a TPA responsive element (TRE)]. An identical sequence is present in the rat GST subunit 7 gene ([28], M. K. Tahir, C. Guthenberg & B. Mannervik, unpublished results) and in a similar position relative to the transcription initiation site. It has recently been shown [34] that c-Ha-*ras* and TPA, both of which can activate the polyoma virus enhancer, act through an enhancer element closely related in sequence to the TREs discussed above. The possibility that *ras* acts through TREs of cellular genes is highly relevant to the tumour-specific induction of GST- π expression, as amplified or activated *ras* genes are frequently found in tumours. The GST subunit 7 gene appears to be *ras*-responsive as GST 7-7 can be induced *de novo* in rat liver epithelial cells following transformation with an N-*ras* gene [35]. In addition, Li *et al.* [36] have shown an increase in GST subunit 7 RNA following expression of an inducible Ha-*ras* gene, introduced into cultured rat liver epithelial cells.

In conclusion, the results demonstrate the isolation of the GST- π gene and the determination of its structure and sequence. Analysis of the sequence of the promoter region has revealed several features which are expected to be important in its function, and of particular interest is a motif closely related to a *ras*-responsive element and TRE from viral and cellular genes, which may be implicated in the tumour-specific expression of GST- π .

I.G.C. is supported by the Frances and Augustus Newman Foundation. K.D.H., S.E.P. and J.B.T. are supported by the Cancer Research Campaign of which B.K. is a Fellow.

REFERENCES

1. Ketterer, B., Meyer, D. J., Coles, B., Taylor, J. B. & Pemble, S. E. (1986) in *Antimutagenesis and Anticarcinogenesis Mechanisms* (Shankel, D. M., Hartman, P. E., Kada, T. & Hollaender, A., eds.), pp. 103–126, Plenum Press, New York
2. Tan, K. H., Meyer, D. J., Coles, B. & Ketterer, B. (1986) *FEBS Lett.* **207**, 231–233
3. Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. & Jornvall, H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7202–7206
4. Pickett, C. B. (1987) *Essays Biochem.* **23**, 116–143
5. Kano, T., Sakai, M. & Muramatsu, M. (1987) *Cancer Res.* **47**, 5626–5630
6. Suguoka, Y., Kano, T., Okuda, A., Sakai, M., Kitagawa, T. & Muramatsu, M. (1985) *Nucleic Acids Res.* **13**, 6049–6057
7. Pemble, S. E., Taylor, J. B. & Ketterer, B. (1986) *Biochem. J.* **240**, 885–889
8. Guthenberg, C., Akerfeldt, K. & Mannervik, B. (1979) *Acta Chem. Scand. Ser. B* **33**, 595–596
9. Dao, D. D., Partridge, C. A., Kuroski, A. & Awasthi, Y. C. (1984) *Biochem. J.* **221**, 33–41
10. Di Ilio, C., Del Boccio, G., Aceto, A., Casaccia, R., Mucilli, F. & Federici, G. (1988) *Carcinogenesis* **9**, 335–340
11. Singh, S. V., Leal, T., Ansari, G. A. S. & Awasthi, Y. C. (1987) *Biochem. J.* **246**, 179–186
12. Tew, K. D., Clapper, M. L., Greenberg, R. E., Weese, J. L., Hoffman, S. A. & Smith, T. M. (1987) *Biochim. Biophys. Acta* **926**, 8–15
13. Hussey, A. J., Stockman, P. K., Beckett, G. J. & Hayes, J. D. (1986) *Biochim. Biophys. Acta* **874**, 1–12
14. Meyer, D. J., Beale, D., Tan, K. H., Coles, B. & Ketterer, B. (1985) *FEBS Lett.* **184**, 139–143
15. Satoh, K., Kitahara, A., Soma, Y., Inaba, Y., Hatayama, Y. & Sato, K. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3964–3968
16. Shea, T. C. & Henner, W. D. (1986) in *Glutathione S-Transferases and Carcinogenesis* (Mantle, T. J., Pickett, C. B. & Hayes, J. D., eds.), pp. 227–231, Taylor and Francis, London
17. Mannervik, B., Castro, V. M., Danielson, U. H., Tahir, M. K., Hansson, J. & Ringborg, U. (1987) *Carcinogenesis* **8**, 1929–1932
18. Shiratori, Y., Soma, Y., Maruyama, H., Sato, S., Takano, A. & Sato, K. (1987) *Cancer Res.* **47**, 6808–6809
19. Soma, Y., Satoh, K. & Sato, K. (1986) *Biochim. Biophys. Acta* **869**, 247–258
20. Batist, G., Tulpule, A., Sinha, B. K., Katki, A. G., Meyers, C. E. & Cowan, K. H. (1986) *J. Biol. Chem.* **261**, 15544–15549
21. Moscow, J. A., Townsend, A. J., Fojo, A., Meyers, C. E. & Cowan, K. H. (1987) *Proc. Am. Assoc. Cancer Res.* **293**, 1160
22. Robson, C. N., Lewis, A. D., Wolf, C. R., Hayes, J. D., Hall, A., Proctor, S. J., Harris, A. L. & Hickson, I. D. (1987) *Cancer Res.* **47**, 6022–6027
23. Minowada, J., Janossy, G., Greaves, M. F., Tsuboda, T., Sahai Srivastava, B. I., Morikawa, S. & Tatsumi, E. (1978) *J. Natl. Cancer Inst.* **60**, 1269–1277
24. Ellison, M. L., Woodhouse, D., Hillyard, C., Dowsett, M., Coombes, R. C., Gillby, E. D., Greenberg, P. B. & Neville, A. M. (1975) *Br. J. Cancer* **32**, 373–379
25. Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, T. & Playfair, J. H. L. (1987) *EMBO J.* **6**, 355–361
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
28. Okuda, A., Sakai, M. & Muramatsu, M. (1987) *J. Biol. Chem.* **262**, 3858–3863
29. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383
30. Bird, A. P. (1986) *Nature (London)* **321**, 209–213
31. McNight, S. L. & Kingsbury, R. (1982) *Science* **217**, 316–324
32. Kadonga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23
33. Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P. & Karin, M. (1987) *Cell (Cambridge, Mass.)* **49**, 729–739
34. Imler, J. L., Schatz, C., Wasylyk, C., Chatton, B. & Wasylyk, B. (1988) *Nature (London)* **332**, 275–278
35. Power, C., Sinha, S., Webber, C., Manson, M. M. & Neal, G. F. (1987) *Carcinogenesis* **8**, 797–801
36. Li, Y., Seyama, T., Godwin, A. K., Winokur, T. S., Lebovitz, R. M. & Lieberman, M. W. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 344–348

Received 9 June 1988/11 July 1988; accepted 12 July 1988