Structure of human glutathione S-transferase class Mu genes

John B. TAYLOR,* John OLIVER,* Robin SHERRINGTON† and Sally E. PEMBLE*†
* Cancer Research Campaign Molecular Toxicology Group and † The Molecular Psychiatry Research Group,
University College and Middlesex Hospital Medical School, Windyay Building,
Cleveland Street, London WIP 6DB, U.K.

Nucleotide sequencing of a human cosmid clone shows that the exon–intron structures of a glutathione S-transferase multigene family are conserved between man and rat, that the human gene family is clustered and that gene conversion events have occurred within the cluster. In addition, between man and rat, there is a high degree of nucleotide sequence identity not only in exons but also in some introns. These conserved sequences are coincident with homologous sequences subject to gene conversion in both species, and hence the utilization of gene conversion by this gene family has itself been conserved. By using transient-expression assay the conserved/converted regions are shown to be capable of modulating transcriptional activity. The data suggest that DNA repair by gene conversion may be a chemical immunity mechanism, which could result in acquired resistance to toxins and, in particular, drug resistance due to glutathione S-transferase in tumours.

INTRODUCTION

The soluble glutathione S-transferases (GSTs; EC 2.5.1.18) are a group of dimeric isoenzymes exhibiting catalytic activities that include the conjugation of GSH with genotoxic and cytotoxic electrophiles, the biosynthesis of physiologically active compounds such as leukotriene C4 and prostaglandin E2 and the binding of hydrophobic compounds such as steroid hormones (Ketterer et al., 1986; Listowsky et al., 1988). The subunits of mammalian GSTs fall into three classes, Alpha, Mu and Pi, according to their primary structures (Mannervik et al., 1985; Pickett, 1987), and their gene structure is conserved between rat (Telawski-Hopkins et al., 1986) and mouse (Daniel et al., 1987) for Alpha class genes and between rat (Okuda et al., 1987) and human (Cowell et al., 1988; Morrow et al., 1989) for the Pi class genes. For Mu class genes, three rat sequences have been compared and gene conversion (i.e. non-reciprocal genetic exchange; see Maizels, 1989) has been suggested to occur in this family (Lai et al., 1988; Morton et al., 1990). We report evidence that germ-line gene conversion also occurs in homologous regions of two human Mu class genes. The sequences subject to gene conversion are also conserved between the species and for at least one human gene are capable of modulating promoter activity in transient-expression assays. These sequences may also undergo somatic gene conversion and, since they are capable of modulating transcriptional activity in a synergistic manner, they offer a novel mechanism for acquired resistance to cytotoxic insult.

MATERIALS AND METHODS

Characterization of cosmid sequences

With the use of a cDNA clone encoding rat GST subunit 4 (J. B. Taylor, unpublished work), a human genomic cosmid library of 2 x 10^9 independent recombinants was screened for cross-hybridizing sequences under conditions described in Cowell et al. (1988) except that washing stringency was 0.5 x SSC (75 mm-NaCl/7.5 mm-sodium citrate buffer, pH 7) containing 0.1 % (w/v) SDS at 60 °C. Cross-hybridizing cosmid fragments were determined by restriction-enzyme analysis and Southern blotting before sub-cloning for DNA sequencing as described in Cowell et al. (1988).

Cell culture, transfections and chloramphenicol acetyltransferase assays

The cell lines HepG2 and MCF7 were grown in Dulbecco’s modified Earle’s medium and RPMI 1640 respectively (obtained from Gibco). All media, with one exception, were supplemented with 10 % (v/v) foetal bovine serum and 10 μg of gentamycin sulphate/ml (‘supplements’), and calcium phosphate transfection and chloramphenicol acetyltransferase assay conditions were as described by Dixon et al. (1989). Exceptional culture media were required in order to assess the hormonal response of chloramphenicol acetyltransferase-containing plasmid constructs transfected into MCF7 cells. In this case the routine medium (RPMI 1640 with supplements) was replaced by Dulbecco’ modified Earle’s medium with supplements immediately before transfection for 4 h in order to avoid the precipitation of the calcium phosphate by high concentrations of Ca²⁺ in RPMI 1640. The cells were allowed to recover for 24 h in RPMI 1640 with supplements and then the medium was replaced by RPMI without supplements but containing additions of hormones at 300 ng of testosterone/ml, 20 ng of 3,3’,5-tri-iodo-L-thyronine/ml and 400 ng of dexamethasone/ml (cell-culture grades; all from Sigma Chemical Co.) where appropriate.

RESULTS

Characterization of human class Mu cosmid sequences

With the use of a cDNA clone encoding rat GST subunit 4 (su4), a class Mu polypeptide, cross-hybridizing human genomic sequence, was isolated from a cosmid library provided by D. Kioussis (Kioussis et al., 1987). Some nucleotide sequences from one cosmid, cosH1.10, which yielded several cross-hybridizing...
(a) Nucleotide sequences encompassing exons 3, 4 and 5 of two human Mu class genes (designated GSTmu2 and GSTmu3) in cosmid H1.10 and (b) a sample intron sequence from GSTmu3 compared with the new *a* human repetitive sequence

(a) The nucleotide sequences encompassing exons 3, 4 and 5 of two human Mu class genes (designated GSTmu2 and GSTmu3) in cosmid H1.10. The 3'-end of intron 2, through exons 3, 4 and 5 into intron 5, of the GSTmu3 gene is aligned with corresponding regions of the GSTmu2 gene, the rat su4 gene (Lai et al., 1988) and the human GSTmu1 cDNA (DeJong et al., 1988; Seidégard et al., 1988). Exons, in upper-case letters, are highlighted by the *mu* cDNA sequence; introns are in lower-case letters. Dashes (-) represent the same nucleotide residue as in GSTmu3. *t* represents a space inserted to maximize sequence fit [note that a dash may represent a space (•) when sequences are compared with GSTmu2]. Examples of conserved (rat/human) intronic oligomeric sequences are, relative to the nucleotide position in parentheses, *mu3(302):su4, CTCTGTTGG; mu2:mu3(276):su4, CATCTCTT; mu2:mu3(370):su4, CTGTTGTC; mu2(439):su4, TGAGTGCC; mu3(459):su4, GGGAAGGGAT; mu3(651):su4, GCCTGGTGG; mu3(670):su4, GCTGAGAGT; mu2(692):su4, GCTGAGAGT]. Single nucleotide changes are ignored, numerous additional sequences of similar or greater length are evident. Underlined is a potential glucocorticoid regulatory element CTTCATGATGTCTT [cf. a functional glucocorticoid regulatory element CTTCATGATGTCTT and their optimum palindromic AGAACANNNTGTCTT (see the text)]. The identity of the human gene was determined from both strands. In the text, the position of the decamer deletion in intron 3 of GSTmu3 (residues 432-441) is compared with a similarly placed 33-residue deletion in rat su4 that corresponds to residues 1328-1360 guac...guag of rat su4 [data from Lai et al., 1988] and Morton et al., 1990,]. (b) A sample intron sequence from GSTmu3 compared with the new *a* human repetitive sequence (Yang et al., 1983; Henthorn et al., 1986). This repetitive region actually spans 379 residues (nos. 1252 to 1630), not all of which are shown.

Fig. 1. (a) Nucleotide sequences encompassing exons 3, 4 and 5 of two human Mu class genes (designated GSTmu2 and GSTmu3) in cosmid H1.10 and (b) a sample intron sequence from GSTmu3 compared with the new *a* human repetitive sequence
respective (results not shown). \textit{GSTmu2} and \textit{GSTmu3} also show strong identity within the whole, or portions, of their corresponding intron sequences. Also of special interest is the strong nucleotide sequence identity between all or part of the introns of both \textit{GSTmu2} and \textit{GSTmu3} and those of the rat genes. These conserved regions are coincident with those subject to gene conversion in the rat, i.e. a region encompassing exons 3, 4 and 5 shows conservation between the species and also undergoes gene conversion in the rat. In contrast, and in common with introns for other conserved multigene families, only the first few nucleotide residues of intron 5 are conserved both between and within the species. The possibility that cosH1.10 is contamination of rat origin can be excluded since a human repetitive sequence, 'new alu' (Fig. 1b), is present in intron 5 of \textit{GSTmu3}, and is part of a cluster of four different tandem repeats (not shown). Furthermore, polymerase-chain-reaction analysis of whole DNA preparations, with primers corresponding to the intron 4-exon 5 region, the \textit{CCCA}-(CCCAAATTCCTCTCACTC)-region, the \textit{tetramer} region, the \textit{CCCAATTCCTCTCACTC}, function such as corticoid regulatory element (Tes+T3) contains three \textit{AATTC} repeats (residues 1–575) in \textit{pSS0.2CAT}, \textit{pe4-2CAT} is the inverse orientation; \textit{pe4-i5CAT} is the \textit{Rsu1–PstI} fragment (residues 575–861), \textit{pi5-4CAT} is the inverse orientation. Orientations were checked by double-strand sequencing across the junctions. \textit{pBS1.0CAT}, a size insert control, has 800 bp containing an alu repetitive and part of an L1 repetitive at –99 bp in \textit{pSS0.2CAT} (Dixon et al., 1989). Transfections were in triplicate and the results are the means of at least three independent experiments with different plasmid preparations that were double-purified by CsCl-gradient centrifugation. Promoter fidelity was checked by nuclease protection analysis with the 171bp \textit{HindIII–SstII} transcription start-containing fragment of \textit{pSS0.2CAT}. Inclusion of either dexamethasone (dex) or 3,3',5-tri-ido-L-thyronine \textit{(T)} \textit{3} in the absence of testosterone \textit{(tes)} had no induction effect (not shown).

### Evidence for sites of frequent gene conversion of \textit{GSTmu2} and \textit{GSTmu3}

\textit{GSTmu2} and \textit{GSTmu3} share two regions of 100 % identity, namely 55 nucleotide residues in intron 2 (residues 27–81, \textit{GSTmu3}) and 174 nucleotide residues of exons 3 and the 5'–ends of introns 3 (residues 165–338, \textit{GSTmu3}). Both regions are surrounded by scattered nucleotide substitutions present at approximately every eight bases and they also have structural motifs in common at their boundaries. These motifs are the presence of a tetramer \textit{TGGG}, which is a Chi-related element associated with gene rearrangement in mouse immunoglobulin genes (Kenter & Birshstein, 1981), and of direct repeats, which are also associated with gene conversion (Stachelek & Liskay, 1988). Specifically, in the 55-nucleotide region, a single tetrameric direct repeat (ACTG) occurs at each boundary and adjoins a \textit{TGGG} tetramer at the 5' boundary (TGGGACTG). In the 174-nucleotide region, the association is less clear since the 3' boundary (CCCCATTCCTCTCTCTC)GTC), as well as containing a \textit{TGGG} (CCCA, underlined) motif, contains three sequences (AATTC, CCCAAT, TCCTCTC) for which direct repeats occur elsewhere within the 174-nucleotide-long sequence (residues 191, 216 and 280). It is noteworthy that the first repeat (AATTC) contains a tetramer contiguous with the \textit{CCCA} (i.e. \textit{CCCCATTCCTCTCTC}). Overall, the conserved/converted region is also unusual in that the \textit{TGGG} (CCCA) tetramer occurs 22 times in \textit{GSTmu3}, 21 times in \textit{GSTmu2}, 17 times in \textit{mu4} and 14 times in \textit{as}, whereas the expected frequencies are seven for \textit{GSTmu3} and \textit{GSTmu2} and six and five respectively for the shorter rat sequences.

### The conserved/converted region can modulate promoter activity

Conservation of nucleotide sequences (see Fig. 1 legend) implies a selective pressure resulting from an advantageous function such as cis-acting regulatory sequences. Transcriptionally active fragments from \textit{GSTmu3} have been identified in transient expression assays. In MCF7, a human mammary carcinoma cell line, the region of intron 2 through to intron 4 confers testosterone-dependent inducibility either by dexamethasone or by 3,3',5-tri-ido-L-thyronine (Table 1). In a human hepatoma cell line, HepG2, we have also localized a cis-acting suppressor function to the intron 4–exon 5 region of \textit{GSTmu3} (Table 1). These activities are orientation-independent, although not always with equal efficiency. Several putative regulatory sequences that may account for these effects are present in the conserved/converted region of \textit{GSTmu3}. In intron 3, the latter half (underlined) of the sequence TGGGTGTCACACTGCCCTTGCAATGATGTTCT is consistent with a glucocorticoid regulatory element (consensus TGGTTC, cf. its optimized palindromic \textit{AGAACANNTTGGTTC}; Klock et al., 1987) and is very like a glucocorticoid regulatory element of the rat tryptophan oxidase gene (CTTTCATGATGTCCT; Danesch et al., 1987). The adjacent sequence TGGGTGTCACACTGCC is a perfect palindromic (underlined) with a spacer tetranucleotide ('hyphen'), CTAC, separating the elements. This structure is characteristic of regulatory elements and is consistent with sequences present in genes both repressed and induced by androgen (TGGNTG; Persson et al., 1990). Co-operativity between cis-acting regulatory sequences for steroid hormones is well characterized (reviewed in Beato, 1989), and at least one \textit{GSTmu3} class gene of the rat is repressed by androgen (Chang et al., 1987). It is also of note that in hormonal regulation of promoters it is the context of the regulatory region and not the nucleotide sequence of a site that determines whether it induces or represses transcription (Beato, 1989).

A hyphenated imperfect palindromic (CTTGGGAGGGTGTCACACTGCCCTTGCAATGATGTTCT) is present in intron 2 (residues 85–102, \textit{mu3}), that,

### Table 1. Relative chloramphenicol acetyltransferase activities generated by \textit{GSTmu3–CAT} chimaeric gene constructs

Values were calculated as percentage chloramphenicol acylation, corrected for protein concentration, and are expressed relative to the activity obtained with the \textit{GST} \textit{p} promoter in \textit{pSS0.2CAT} (Cowell et al., 1988; Dixon et al., 1989), which is active in numerous cell types. \textit{GSTmu3} fragments, with \textit{HindIII} linkers, were inserted in \textit{pSS0.2CAT} (Dixon et al., 1989), which has a \textit{HindIII} site at –99 bp, as follows: \textit{pi2-4CAT} is the \textit{EcoRI–RsuI} fragment (residues 1–575) in \textit{pSS0.2CAT}, \textit{pe4-2CAT} is the inverse orientation; \textit{pe4-i5CAT} is the \textit{Rsu1–PstI} fragment (residues 575–861), \textit{pi5-4CAT} is the inverse orientation. Orientations were checked by double-strand sequencing across the junctions. \textit{pBS1.0CAT}, a size insert control, has 800 bp containing an alu repetitive and part of an L1 repetitive at –99 bp in \textit{pSS0.2CAT} (Dixon et al., 1989).

<table>
<thead>
<tr>
<th>Relative chloramphenicol acetyltransferase activity</th>
<th>MCF7 cells</th>
<th>pSS0.2CAT</th>
<th>pi2-4CAT</th>
<th>pe4-2CAT</th>
<th>pBS1.0CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No addition</td>
<td>1.0 ±0.01</td>
<td>1.3 ±0.3</td>
<td>1.0 ±0.2</td>
<td>0.9 ±0.2</td>
<td></td>
</tr>
<tr>
<td>Tes</td>
<td>0.7 ±0.3</td>
<td>1.4 ±0.3</td>
<td>1.1 ±0.1</td>
<td>0.7 ±0.2</td>
<td></td>
</tr>
<tr>
<td>Tes+ dex</td>
<td>0.6 ±0.2</td>
<td>3.3 ±0.1</td>
<td>3.4 ±0.2</td>
<td>0.9 ±0.1</td>
<td></td>
</tr>
<tr>
<td>Tes+ Tg</td>
<td>0.9 ±0.1</td>
<td>3.5 ±0.1</td>
<td>2.8 ±0.4</td>
<td>1.1 ±0.1</td>
<td></td>
</tr>
<tr>
<td>HepG2 cells</td>
<td>pSS0.2CAT</td>
<td>pi2-4CAT</td>
<td>pBS1.0CAT</td>
<td>pi5-4CAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0 ±0.2</td>
<td>0.3 ±0.1</td>
<td>0.6 ±0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pi2-4CAT</td>
<td>pe4-i5CAT</td>
<td>pi5-4CAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1 ±0.2</td>
<td>1.0 ±0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
excepting the length of the hyphen, fits a consensus element (CTTTGGNNNCCCGNG) present in 3',5'-tri-iodo-L-thyronine-responsive genes and reported recently (Sap et al., 1990). The length of the hyphen need not prohibit binding of a factor, since, for example, the 3',5'-tri-iodo-L-thyronine receptor has also been reported to bind to oestrogen-responsive elements irrespective of whether a gap is present, although activation of reporter promoters may not result (Sap et al., 1990).

Another sequence of interest centres on the decameric deletion at residue 433 in intron 3. Relative to GSTmu2, a hyphenated palindrome, GAGCGGGCAGCTGAGTTGCCTGGTGCT, is lost due to the deletion (ACACTGAGTG), although a second hyphenated palindrome, GAGCAGCCCTCTGGTGTC, is generated in GSTmu3. This may reflect the mechanism of deletion or both sequences may be cis-acting regulatory elements. In addition, of several regulatory elements shown to bind transcription factors that act synergistically with steroid receptors (Schüle et al., 1988), at least two consensus sequences are also present in the conserved/converted introns. These are the SP1 binding element (GGGGCGG) in GSTmu3 (residues 238–244, intron 3) and GSTmu2 and a possible NF1 element (TCCTTGG) in GSTmu3 (residues 83–89, intron 2; see also GSTmu3, residue 651; Fig. 1 legend), GSTmu2 and rat su4.

Evidence for restriction of gene conversion affecting intron 4–exon 5 to sub-classes of human GSTmu

There are two oligomers (10- and 11-mers) in intron 4 that are identical in rat su4 and GSTmu3 (residues 651 and 670; Fig. 1 legend). Although these might have a role in the down-regulation in HepG2 cells, a further complexity is that, over the length of intron 4 and the 5'-end of exon 5, GSTmu2 and rat su4 have greater sequence similarity to each other than either has to GSTmu3 (Table 2 and Fig. 2a). A sequence that would correspond to intron 4–exon 5 of GSTmu3 has not yet been reported in the rat (see 5'-end of exon 5 in Fig. 2a). We have used this loss of identity in intron 4–intron 5 to provide evidence that GSTmu2 and GSTmu3 are not merely the products of a recent gene duplication. Using the intron 4–exon 5 regions of GSTmu2 and of GSTmu3 in Southern-blot analyses of human genomic DNA restricted with HincII characterizes the number of sequences similar to these regions in the genome and demonstrates that the two genes are members of two sub-classes. The GSTmu2-derived fragment hybridizes strongly to three fragments in some DNA samples and to two in GSTmu1 homozygous null individuals (Fig. 3a), and is itself present in the largest fragment [5.3 kb; deduced from Southern blots of CosHI.10 (results not shown)]. The latter DNA hybridization profile with two hybridizing fragments is due to a deletion genotype yielding a null allele at this locus (Board, 1981; Seidegård et al., 1988), and the relative intensity of hybridization, in conjunction with the family tree, shows that the deletion undergoes Mendelian segregation (Fig. 3). The probe also hybridizes very weakly to two additional fragments. None of these five fragments contains the homologous GSTmu3 locus, since the GSTmu3-derived intron 4–exon 5 fragment hybridizes to, and is contained within, a 5.9 kb HincII

![Fig. 2](image_url)

(a) Deduced amino acid sequences encoded by exons 5 in cosmid H1.10, GSTmu1 and rat Mu class subunit genes and (b) codon usage in exon 3

(a) Deduced amino acid sequences encoded by exons 5 in cosmid H1.10, GSTmu1 and rat Mu class subunit genes. Only those amino acid residues that differ from those encoded by GSTmu2 are illustrated. (b) Codon usage in exon 3. mu1, mu2, and mu3 are human sequences, su3, su4 and su6 are rat sequences and GT75 and GT55 are mouse sequences [references are given in the text except Pearson et al. (1988) and Towsey et al. (1989) for mouse sequences]. Underlined are high codon degeneracies of which only one is used.

Table 2. Percentage identity of the intron 4-exon 5 regions of rat and human Mu class genes

<table>
<thead>
<tr>
<th>Degree of identity (%)</th>
<th>Overall (210 bases)</th>
<th>Intron 4 (103 bases)</th>
<th>Exon 5 (101 bases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mu3:mu2</td>
<td>mu3:rat</td>
<td>mu3:mu2</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>69</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>mu2:rat</td>
<td>81</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td></td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td></td>
<td>82</td>
</tr>
</tbody>
</table>

The Table shows a comparison of the 'escape from gene conversion' area, to emphasize that human mu2 is more like rat su4 than it is like human mu3. Note the 74 % identity in the introns 5 of mu2:rat is the same percentage as the identity of the exons 5 of mu2:mu3.
Fig. 3. Southern blots of *Hind*II digests of familial DNA samples probed with exons 5

Fragments arrowed are 5.9, 5.3, 4.7, 4.6, 3.8, 3.2, 2.8, 2.6, 2.3 and 1.0 kb. Integers refer to the familial relationships illustrated in (c). (a) Probed with the *PstI*-EcoR I fragment of GSTmu2, residues 631–1058. The relative intensities of fragments f and g (also fragments b and g) show that 9 is homozygous positive (2.8 kb fragment), 8 is homozygous null and progeny are all heterozygotes. All others are homozygous null. (b) Reprobed with the *RsaI*-PvuII fragment of GSTmu3, residues 575–769. The signals from fragments seen with the GSTmu2 probe probably result from residual mu2 probe, since these are not seen in the viral blots. Note the 3.8 kb fragment in 11 and 15, the 2.6 kb fragment in 1 and 15, and the 2.3 kb fragment in 2 and 7. (c) Familial relationships: 21 is half-sister to 4, 11, 14, 18 and 8, and *5 signifies male.

fragment (Fig. 3b). The GSTmu3 fragment also highlights several polymorphisms (Fig. 3b legend) and in particular hybridizes with different intensities to a 1.0 kb fragment. This variation in intensity is not consistent with a simple deletion such as is seen for the null allele. The intensity may reflect the percentage homology to the probe arising from gene conversion events at different times, since varying the copy number would require conservation of the size of the restriction fragment in the multiple copies. It is noteworthy that these distinct hybridization profiles were obtained under a low-stringency wash (60 °C and 0.5 × SSC), which therefore emphasizes the non-identity of the region. Hence, on the basis of this differential hybridization to the two probes, this region allows a sub-division of the class and offers an example of escape (Walsh, 1987) from gene conversion between the sub-classes, presumably due to loss of sequence similarity. In summary, therefore, although GSTmu2 and GSTmu3 have converted sequences covering intron 2 through to exon 4, the evidence is consistent with gene conversion restricted to their sub-class in the intron 4-exon 5 region.

Other unusual sequence characteristics

We have also considered selection pressure at the protein level. In this case no open reading frame that might yield a conserved (rat/human) amino acid sequence is obvious in the intron sequences, especially since any reading frame would be disrupted by the spacers (•) inserted to maximize nucleotide similarity. This is true of reading frames that are either sense or anti-sense relative to the GSTmu exons. However, for the nine reported nucleotide sequences of human, rat and mouse Mu class subunits, the codon usage present in exon 3 is unusual (Fig. 2b). In seven sequences five amino acid residues have codons with six degenerate options and two residues have a codon with four degenerate options, yet only one codon option is used in each case. The other two sequences are also remarkable in that they have, in one case six, and in the other case five, of the same seven codons conserved. We have used an anti-sense probe for exon 3 in Northern-blot analyses of RNA from several tissues taken from 12-day and 18-day mouse foetuses and from adult rats. We have observed no hybridization signal and are unable to explain the codon preference, although we note that a hyphened palindrome (AGCCAGTGGCT) is present in the oligomeric sequence conserved in all nine genes.

DISCUSSION

Gene conversion within the germ-line was proposed as a mechanism to allow the evolution of any gene family that maintains microdiversity in a set of otherwise very similar and functional molecules (Baltimore, 1981), and sequence identity in the introns of gene families is evidence for its occurrence. Although the differential hybridization profiles of the intron 4-exon 5 regions and loss of identity in intron 5 provide evidence that GSTmu2 and GSTmu3 are not a recent gene duplication, the high degree of sequence identity in introns 2 and 3 shows that gene conversion has occurred subsequent to the gene duplication. In addition, the intron 4-exon 5 regions of the two genes appear to have undergone gene conversion restricted to their sub-class. Similar gene conversion events, localized to the same regions of the genes, have also occurred in the homologous rat gene family, although a sub-class has not been characterized. Since intron sequences are usually non-functional, however, there is usually no selection pressure to conserve the intron sequences across the species even if gene conversion were occurring within the family. In the GSTmu gene family, therefore, the sequences subject to gene conversion are unusual, since there is conservation of the intron sequences, presumably due to regulatory elements within them. Hence the conserved superposition of germ-line gene conversion, both in rat and man, and the resulting microdiversity in regulatory elements might result in varying the control of expression of the genes. Such an interpretation is consistent with the restriction of gene conversion to a sub-class, since it is possible that a regulatory element within an intron 4 is linked with the property conferred upon the subunit when the 5'-end of exon 5 is incorporated. This phenomenon could also allow variations in expression of the members of the gene family within the population according to cis-acting sequences without a resulting change in amino acid sequence. For example, a conversion event such as the 174-nucleotide-long region of 100% identity covering exon 3 and the 5'-end of intron 3 of the human genes would not result in a change in amino acid sequence.
The data agree with a model in which the expression of a Mu class gene containing a given exon, and possibly an enzymic function conferred by that exon, is linked to the regulatory element adjacent to the exon. Such a model is also consistent with the gene conversion data in the rat, which show that the high degree of identity in the exons extends for various lengths into the neighbouring introns, for example 40, 60 and 100 nucleotide residues of the 3'-end of intron 2 depending upon which genes are compared (Lai et al., 1988; Morton et al., 1990; Fig. 1). It is also noteworthy that deletion events in intron 3 (see Fig. 1 legend) have occurred at the same position in the rat and human gene families even though the sequences surrounding the deletions are highly homologous within the species. This is suggestive of a mechanism operating at similar boundaries. The model is also in agreement with the mosaicism (Taylor et al., 1988) seen in the primary structures of subunits from other organisms such as Schistosoma mansoni, a parasitic helmith. This mosaicism arises from short regions with a high degree of identity that correlate to exons 2, 4 and 7 of the appropriate subunit gene family of higher eukaryotes and suggests that function is associated with exon domains.

It is unlikely, however, that germ-line gene conversion is responsible for all of the sequence identity in cosH11.10. Here the presence of two sequences of 100% identity implies that the conversion events were very recent, especially since both sequences are surrounded by regions with single scattered nucleotide changes; and, since there are two recent events, this is suggestive of a somatic mechanism. A somatic recombination mechanism is also consistent with the derivation of cosH11.10, since the clone was selected from a cosmid clone bank (Kioussis et al., 1987) derived from a T-cell line that may have, and will have had in its stem cell, an active recombinational mechanism. For example, gene conversion has been suggested as a means of somatic hypermutation in mammalian immunoglobulin genes and is essential in generating diversity in avian B-cells (Maizels, 1989).

The data also suggest that the application of gene conversion to this region is itself subject to selection, since there is a coincidence, in both species, of the occurrence of gene conversion and of the high frequency of sequences (TGGG) reported to facilitate it. Selection for frequent gene conversion within the gene family could also explain the maintenance of the GSTm1 deletion genotype in all human populations. In all races the homozygous null genotype is present at similar frequency (approx. 45%), which results in a similar occurrence of the heterozygote. If a selective advantage accrues from frequent gene conversion, then the heterozygote genotype ought to be beneficial, since it could result in frequent misalignment due to the deletion, thereby both facilitating and adding to potential diversity.

A putative selection pressure that reflects the diverse properties of the class Mu GSTs, such as hormone binding and xenobiotic detoxification, is also evident for the selection of somatic gene conversion. The transposition of regulatory elements with result in co-operativity effects leading to altered expression of the proteins, for example by induction or in cell type. By analogy to the somatic gene conversion in avian B-cells, altered expression of class Mu GSTs provides a means for the adaptation of a cell to its chemical or hormonal environment.

With respect to tumour chemotherapy, however, gene conversion in this gene family may be detrimental. Somatic gene conversion has previously been suggested as a mechanism that would result in the expression of recessive alleles during tumorigenesis (Cavenee et al., 1983). It has also been induced in mouse L cells following exposure to an alkylating cross-linking anti-tumour drug (mitomycin C), to a simple alkylating agent (N-methyl-N'-nitro-N-nitrosoguanidine) and to a reactive metabolite of a polycyclic aromatic hydrocarbon (+)-anti-benzo[a]-pyrene-7,8-diol 9,10-epoxide (Wang et al., 1988), the latter being a good substrate for Mu class GSTs (Robertson et al., 1986). Three bulky carcinogens structurally related to (+)-anti-benzo[a]pyrene-7,8-diol 9,10-epoxide have also been shown to stimulate gene conversion (Bhattacharyya et al., 1989). Hence the GSTmu class introns are likely subjects of DNA repair by gene conversion initiated by, for example, alkylating agents or oxy radicals (Hayes & Wolf, 1988), and such repair could result in altered levels of expression of the proteins, in alteration of their cell-specific regulation and in changes in their structures (or combinations of these), thereby allowing the selection of that cell which successfully detoxifies the source of DNA damage (Carmichael et al., 1986). Thus DNA repair by gene conversion may select GST-class-Mu-dependent genotoxin-resistant clonal events in tumours.

J. B. T. and S. E. P. thank the Cancer Research Campaign. J. O. thanks the Middlesex Hospital Trustees and R. S. thanks the Wellcome Foundation for financial support. We also thank Professor B. Ketterer, Head of the Cancer Research Campaign Molecular Toxicology Group, for help and encouragement, Dr. P. Beverley for the MCF7 line, Dr. D. Edwards (Calgary) for Northern blots of mouse foetal RNA and Dr. J. Cowell for sharing his chloramphenicol acetyltransferase assay expertise.

REFERENCES

Human glutathione S-transferase class Mu genes


Received 5 June 1990/16 August 1990; accepted 23 August 1990


Vol. 274