Identification of polymorphism at the glutathione S-transferase, GSTM3 locus: evidence for linkage with GSTM1*A

Avis INSKIP,* Jose ELEXPERU-CAMIRUAGA,† Neil BUXTON,† Pilatha S. DIAS,† James MACINTOSH,† Donald CAMPBELL,† Peter W. JONES,‡ Lilian YENGL,* J. Alan TALBOT,* Richard C. STRANGE** and Anthony A. FRYER**§

**Centre for Pathology and Molecular Medicine, School of Postgraduate Medicine, Keele University, Central Pathology Laboratory, North Staffordshire Hospital, Stoke-on-Trent, Staffordshire, U.K., †Department of Neurosurgery, North Staffordshire Hospital, Stoke-on-Trent, Staffordshire, U.K., and ‡Department of Mathematics, University of Keele, Keele, Staffordshire, U.K.

Allelism in the glutathione S-transferase, GSTM3 gene has been identified using PCR with specific primers to exon 6/exon 7. Sequencing showed the mutant GSTM3*B allele to have a three-base deletion in intron 6 with a frequency of 0.158. The mutation

generates a recognition sequence, 5'-AAGATA-3', for the negative transcription factor YY1. GSTM3*B was significantly associated with GSTM1*A.

INTRODUCTION

Polymorphism in carcinogen-metabolizing enzymes has been associated with susceptibility to various pathologies [1,2]. Thus, much attention has centred on the glutathione S-transferase, GSTM1 locus as homozygosity for the null GSTM1*0 allele confers increased risk of several cancers [3–6]. However, while GSTM1 is an attractive candidate, data from some studies are conflicting [3]. The reason for these discrepancies is unclear, although studies showing a protective effect of GSTM1 A,B (but not GSTM1 A or GSTM1 B) in cutaneous basal cell carcinoma [5,6], and differing effects of GSTM1 A and GSTM1 B on susceptibility to several cancers, suggest that a model based only on the hypothesis that homozygosity for GSTM1*0 confers increased risk may be too simplistic [3].

GSTM1 is one of five mu-class genes (M1–M5) situated in tandem on chromosome 1 [7–9]. While the products of these genes demonstrate differences in catalytic efficiencies [10], there is an overlap in their substrates, implying that deficiencies in one gene may be compensated for by expression of other genes in the family. This possibility would be complicated by polymorphism in other mu genes. While this possibility has not been studied in detail, restriction-fragment-length polymorphisms within the gene cluster have been described [11], though it is unclear which genes, if any, are involved. Phenotypic variation in expression of mu-class genes such as GSTM3 has also been reported in human lung and brain [12–15] though the molecular basis for this phenomenon is unknown. Interestingly, Nakajima et al. [14] have shown an influence of GSTM1 genotype on the level of GSTM3 expression in lung: GSTM1*0 homozygotes expressed less GSTM3 than individuals with other GSTM1 genotypes. These findings indicate that the effect of GSTM1 on susceptibility may be influenced by expression of GSTM3. We now describe studies that characterize, for the first time, allelic variation at GSTM3 and show linkage between a GSTM3 allele containing a motif for the YY1 negative transcription factor and GSTM1*A. Our data suggest a mechanism for the differing effects of GSTM1 A and GSTM1 B and indicate that interactions between mu genes may be a significant factor in determining susceptibility.

MATERIALS AND METHODS

Identification of polymorphism in GSTM3

DNA was extracted, using a phenol/chloroform method, from leucocytes from 244 patients admitted to the North Staffordshire Hospital with various pathologies. All samples were collected with Ethical Committee approval and informed consent. As the sequences of exons 6–7 of GSTM1, M2, M3, M4 and M5 are known [9], specific primers to this region of GSTM3 were synthesized (5'-CCTCAGTACTTGAAGAGCT-3' and 5'-CACATGAAAGCCTTCAGGTT-3'). PCR was performed for 35 cycles of denaturation (94 °C, 45 s), annealing (58 °C, 30 s) and primer extension (72 °C, 45 s). Electrophoresis in 4% (w/v) agarose gels revealed two distinct electrophoretic patterns: first, in 181 subjects a single band of about 273 bp (wild-type pattern) and secondly, in 63 subjects a two-band pattern comprising fragments of about 273 bp and 285–290 bp (atypical pattern). Atypical pattern PCR products were further resolved in 20% (w/v) polyacrylamide which revealed, in each case, that the additional band was a doublet of 365 and 387 bp.

To characterize the additional bands, the following strategy was used. First, the specificity of PCR was confirmed by digesting the product with *EcoRV* and *HpaII* (Sigma) using the manufacturer's conditions. These enzymes were chosen because inspection of the GSTM1–M5 sequences [9] shows M3 has restriction sites (5'-GAT/ATC-3' *EcoRV*, 5'-C/CGG-3' *HpaII*) not present in the other mu genes. Secondly, 'hot start' PCR was used to reduce non-specific primer annealing [16]. PCR conditions were as above, except Taq polymerase was added at 94 °C. Thirdly, to eliminate matrix effects between primers [16], an alternative reverse primer in intron 6 (5'-CCTCAGACCTAAGCATCTTA-3') was used.

Sequencing of PCR products

PCR-amplified DNA was sequenced by the Sequenase method (United States Biologicals) according to the manufacturer's instructions using a biotinylated forward primer and a second

§ To whom correspondence should be addressed.

internal primer in exon 7 (5'-ATATACGGTTCTGATCCAA-GAT-3'). Results were confirmed using the fmol cycle sequencing method (Promega, Southampton, U.K.). DNA was sequenced from the sense and antisense strands. Labelled fragments were separated in 6% (w/v) polyacrylamide and autoradiographed.

Each of the three fragments from four subjects with the atypical pattern were isolated from polyacrylamide gels by a 'crush and soak' method [17] and sequenced. All three sequences were identical, indicating that the bands of molecular size 365 and 387 bp comprised heteroduplex fragments. Thus, we predicted that the 273 bp fragment in subjects with the atypical pattern would comprise a mixture of two similar but not identical fragments. Since it was not possible to separate the two homoduplex forms giving rise to the heteroduplexes, the total PCR product from these individuals was sequenced and the sequence of the wild-type pattern subtracted to reveal the additional DNA sequence. The two sequences were compared using the SeqAid software package (SeqAid II version 3.81, Kansas State University, Molecular Genetics Laboratory, KS, U.S.A.). The data obtained indicated the presence of two *GSTM3* alleles and we therefore hypothesized that the wild-type pattern comprised individuals homozygous for either the *GSTM3*A* or *GSTM3*B* alleles while the atypical pattern comprised heterozygotes for these alleles.

Identification of *GSTM3* genotypes

Inspection of the sequences showed an *MnII* restriction site (CCTC(N)₇/(N)₇GAGG) in *GSTM3*A* but not in *GSTM3*B*. Wild-type and atypical pattern PCR products were digested with *MnII* (37 °C, 18 h) and resolved in 20% (w/v) polyacrylamide (200 V, 2 h) stained with ethidium bromide (0.5 µg/ml). Sequencing of DNA from a *GSTM3*B* homozygote confirmed the sequence obtained by subtractive sequencing.

Frequency of *GSTM3* polymorphism

GSTM3 genotype frequencies were determined by *MnII* digestion of DNA from the 244 subjects. Allele frequencies were calculated assuming Hardy-Weinberg equilibrium. Chi-squared tests were used to investigate the dependence of *GSTM3* genotype on *GSTM1* genotype.

Identification of *GSTM1* genotypes

GSTM1 genotype was identified using a modified amplification refractory mutation system-based PCR. Amplification was carried out using primer sets to intron 6/exon 7 and exon 4/exon 5 of *GSTM1* and, as a positive control, β -globin [6].

RESULTS

Identification of *GSTM3* polymorphism

As expected from the published sequence [9], PCR using the *GSTM3*-specific primers yielded an approx. 273 bp fragment (wild-type pattern) on electrophoresis in agarose gels. No sample failed to amplify the *GSTM3* fragment. In 63 of the 244 subjects, however, an additional band (285–290 bp) was detected which appeared as a doublet of 365 and 387 bp on further resolution in polyacrylamide (atypical pattern; Figure 1). These additional bands persisted despite 'hot start'. The use of the intron 6 primer also resulted in a three-band pattern with 157, 250 and 272 bp fragments observed. Restriction endonuclease digestion resulted in the expected *GSTM3*-specific pattern; wild-type pattern gave

Figure 1 PAGE of *GSTM3*

Lanes: 1, *GSTM3*A/GSTM3*B* subject; 2, as lane 1 post-*MnII* digestion; 3, *GSTM3*A/GSTM3*A* subject; 4, as lane 3 post-*MnII* digestion; 5, *GSTM3*B/GSTM3*B* subject; 6, as lane 5 post-*MnII* digestion; 7, *GSTM3*A/GSTM3*B* subject; 8, as lane 7 post-*MnII* digestion; 9, negative control (no DNA); 10, molecular-size markers (pBR322 *HaeIII* digest).

MnII

```
CCTCAGTACT TGGAAGAGCT ACCTGGACAA CTGAAACAAT TCTCCATGTT
: : : : :
CCTCAGTACT TGGAAGAGCT ACCTGGACAA CTGAAACAAT TCTCCATGTT 50
: : : : :
CCTCAGTACT TGGAAGAGCT ACCTGGACAA CTGAAACAAT TCTCCATGTT
```

HpaII

```
TCTGGGAAAA TTCTCATGGT TTGCCGGGGA AAAGgtagga agaagggaaa
: : : : :
TCTGGGAAAA TTCTCATGGT TTGCCGGGGA AAAGgtagga agaagggaaa 100
: : : : :
TCTGGGAAAA TTCTCATGGT TTGCCGGGGA AAAGgtagga agaagggaaa
```

MnII

```
agaagagat acttctctat ctctgcaggc tactactcct cagacctaag
: : : : :
agagaag_at acttctctat ctctgcaggc tactactcct cagacctaag 150
: : : : :
agaag__at acttctctat ctctgcaggc tactactcct cagacctaag
```

MnII

```
catcttattg ctttctctct agCTCACCTT TGTGGATTTT CTCACCTATG
: : : : :
catcttattg ctttctctct agCTCACCTT TGTGGATTTT CTCACCTATG 200
: : : : :
catcttattg ctttctctct agCTCACCTT TGTGGATTTT CTCACCTATG
```

EcoRV

```
ATATCTTTGGA TCAGAACCCT ATATTGACC CCAACTGCCT GGATGAGTTC
: : : : :
ATATCTTTGGA TCAGAACCCT ATATTGACC CCAACTGCCT GGATGAGTTC 250
: : : : :
ATATCTTTGGA TCAGAACCCT ATATTGACC CCAACTGCCT GGATGAGTTC
```

```
CCAAACCTGA AGGCTTTCAT GTG - This report (GSTM3*A)
: : : : :
CCAAACCTGA AGGCTTTCAT GTG - Pearson et al. [9]
: : : : :
CCAAACCTGA AGGCTTTCAT GTG - This report (GSTM3*B)
```

Figure 2 Sequence of *GSTM3* exon 6/exon 7

Comparison of exon 6, intron 6 and exon 7 of *GSTM3* in the published sequence [9] with the alleles described in this report. Differences are underlined. Restriction enzyme recognition sites are in **bold** and cut sites are shown by arrows.

fragment of 202 and 71 bp (*EcoRV*) and 199 and 74 bp (*HpaII*), while the atypical pattern gave fragments of 316, 294, 202 and 71 bp (*EcoRV*) and 313, 291, 199 and 74 bp (*HpaII*) (see Figure

Figure 3 Autoradiograph of sequencing gel from patients with wild-type and atypical patterns

Lanes 1–4, sequence of wild-type pattern. Lanes 5–8, sequence of atypical pattern, showing multiple signals in some tracks suggestive of amplification of two similar sequences which differ at the point indicated (*).

2). These data show that the additional bands contain restriction sites consistent with allelism in GSTM3 and were not artifacts.

The difference in mobility in agarose and polyacrylamide gels suggested the formation of heteroduplexes as a result of amplification of two similar GSTM3-like sequences in some samples. Consequently, DNA from eight subjects with the wild-type pattern and four with the atypical pattern was sequenced. The sequence of the PCR product from the eight wild-type samples (Figure 2) differed from that reported [9] in two places in intron 6 (shown underlined): first, substitution of two bases and a single base deletion, and secondly, a single base substitution. The latter substitution, found using both sequencing methods in both strands, was also found in the four atypical samples.

Each of the three fragments from the four atypical subjects was isolated from the gel and sequenced. Identical results were obtained whichever fragment was analysed, further suggesting

that the additional bands were heteroduplexes arising from allelism in GSTM3. Sequencing both the sense and antisense strand showed that the base sequence was identical with that of the eight wild-type samples sequenced until the base deletion; thereafter radioactivity was detected in more than one lane on the autoradiograph, suggesting simultaneous amplification of two fragments with identical ends (Figure 3). The sequence obtained from the eight wild-type samples was therefore subtracted to reveal the additional sequence. As the eight wild-type DNAs gave identical sequences we termed this allele, *GSTM3*A*. The additional sequence obtained using the subtractive approach had a 3 bp deletion compared with *GSTM3*A* (Figure 2), and was termed *GSTM3*B*.

Digestion of DNA from *GSTM3*A* homozygotes with *MnII* gave the expected 11, 51, 86 and 125 bp fragments. PCR products from individuals with the atypical pattern (i.e. *GSTM3*A/GSTM3*B*) demonstrated the additional 134 bp fragment (Figure 1). Digestion of PCR products from individuals with a single band revealed some subjects to be homozygous for *GSTM3*B*. *MnII* digestion of these samples gave the expected fragments (11, 125 and 134 bp), and one *GSTM3*B/GSTM3*B* sample was sequenced for confirmation.

GSTM3 genotype frequencies

Determination of genotype frequencies in 244 subjects identified 173 (70.9%) *GSTM3*A* homozygotes, 63 (25.8%) *GSTM3*A/GSTM3*B* heterozygotes and 8 (3.4%) *GSTM3*B* homozygotes. Allele frequencies were: *GSTM3*A*, 0.842; *GSTM3*B*, 0.158.

The possibility that GSTM3 was linked to GSTM1 was examined by comparing GSTM3 genotype frequencies in subjects with different GSTM1 genotypes (Table 1). Chi-squared analysis showed that GSTM3 genotype was dependent on GSTM1 genotype ($X^2_6 = 23.84$, $p = 0.0026$). Inspection of the data showed that the largest chi-squared contribution (and greatest departure from expected values) was for GSTM1 A (16.2) compared with 4.9, 1.7 and 0.9 for GSTM1 null, GSTM1 B and GSTM1 A,B, respectively (Table 1). Thus, *GSTM1*A* is associated with an increased frequency of *GSTM3*B*, indicating that the polymorphism is linked to GSTM1 on chromosome 1.

DISCUSSION

This is the first reported characterization of allelic variation in a mu-class gene other than GSTM1. All samples analysed showed the presence of amplified DNA, indicating that GSTM3 does not demonstrate a null allele as found in GSTM1 and GSTT1 [1–3]. It was necessary to exclude other explanations for the additional

Table 1 Evidence for linkage between *GSTM1*A* and *GSTM3*B*

Frequencies of GSTM3 genotypes in subjects with GSTM1 null, GSTM1 A, GSTM1 B and GSTM1 A,B polymorphisms with expected values calculated using Chi-squared tests in parentheses.

	GSTM1 null	GSTM1 A	GSTM1 B	GSTM1 A,B	Total
GSTM3 genotype					
<i>GSTM3*A/GSTM3*A</i>	115 (102.81)	29 (43.25)	21 (18.43)	8 (8.51)	173
<i>GSTM3*A/GSTM3*B</i>	27 (37.44)	28 (15.75)	5 (6.71)	3 (3.10)	63
<i>GSTM3*B/GSTM3*B</i>	3 (4.75)*	4 (2.00)	0 (0.85)	1 (0.39)	8
Total	145	61	26	12	244

PCR products found in some subjects [16]. PCR artifacts have been shown to result from degenerate primer annealing, amplification of non-specific sequences and multiplex primer interactions. Each of these possibilities was excluded, indicating that the amplified bands originated from a GSTM3-like sequence.

GSTM1–GSTM5 are arranged in tandem on chromosome 1, though it is possible that a further mu-like gene or pseudogene is present elsewhere in the genome [9]. It is unlikely that this putative gene gave rise to the additional sequence reported here because of the linkage with *GSTM1**A. The band may result from an unidentified mu-like sequence on chromosome 1, though allelism in GSTM3 is most likely.

Although the differences between *GSTM3**A and *GSTM3**B occur in intron 6, intronic and extragenic sequences appear to be important in mediating transcriptional activity at other detoxifying enzyme loci [18,19]. Interestingly, the base-pair deletion occurs in a region containing a palindromic sequence (Figure 1) in *GSTM3**B, a consensus sequence common to control elements. The region also contains a GA-rich, followed by a CT-rich, sequence. It is known that deletions appear more frequently in regions rich in direct or inverted repeats which have a potential for stable self-complementarity, such as loop and stem structures [20]. Deletions may therefore be generated by slippage of the DNA template at these points during replication. Taylor et al. [7] reported the presence of conserved sequences within introns of mu genes and that suggested conservation implies a selective pressure resulting from an advantageous function of these regions, such as a *cis*-acting regulatory sequence. Indeed, they claimed that perfect hyphenated palindromes, as seen around the site of the mutation described here, are characteristic of regulatory elements. Interestingly, the sequence 5'-AAGATA-3' present in *GSTM3**B, but not *GSTM3**A, is a recognition motif for the negative transcription factor, YY1 [21].

The data presented in this study, and recently by Nakajima et al. [14], identify two distinct interactions between GSTM1 and GSTM3. First, Nakajima et al. showed that individuals with GSTM1 null express lower levels of GSTM3, and possibly GSTM2, though they did not propose a mechanism for this effect. Secondly, our data indicate that in individuals that express GSTM1, GSTM1 A and GSTM1 B may confer different susceptibilities because of linkage with alleles in GSTM3. This suggestion would explain data from our group showing, in several cancers, that the frequency of *GSTM1**B is reduced in the case group [3,6]. The mechanism for this effect was unclear as *GSTM1**A and *GSTM1**B have similar catalytic and structural properties. The identification of the GSTM3 polymorphism with linkage to *GSTM1**A suggests that GSTM1 A confers increased susceptibility because of its association with *GSTM3**B. Thus, individuals with GSTM1 A should demonstrate reduced ex-

pression of GSTM3 because of the YY1 negative transcription factor. This motif has been shown to regulate gene expression from intragenic sequences in several genes including *c-myc* [22]. The data suggest that determining a role for GSTM1 in mediating cancer susceptibility may need to take into account complex interactions between various members of this gene family, since other members may be polymorphic.

We gratefully acknowledge the support of the North Staffordshire Hospital Neurosurgery Trust Fund. We are also grateful to Dr. B. Ollier and Dr. A. Hajeer for helpful advice.

REFERENCES

- 1 Wolf, C. R., Smith, C. A. D. and Forman, D. (1994) *Br. Med. Bull.* **50**, 718–731
- 2 Nebert, D. W., McKinnon, R. A. and Puga, A. (1995) *Proc. Am. Assoc. Cancer Res.* **36**, 664–665.
- 3 Strange, R. C. (1995) in *Proceedings 1995 International ISSX – Workshop on Glutathione S-Transferases*, Taylor and Francis, London, in the press
- 4 Bell, D. A., Taylor, J. A., Paulson, D. F., Robertson, C. N., Mohler, J. L. and Lucier, G. W. (1993) *J. Natl. Cancer Inst.* **85**, 1159–1164
- 5 Fryer, A., Heagerty, A., Bowers, B. et al. (1995) *Proc. Am. Assoc. Cancer Res.*, **36**, 119
- 6 Heagerty, A. H. M., Fitzgerald, D., Smith, A. et al. (1994) *Lancet* **343**, 266–268
- 7 Taylor, J. B., Oliver, J., Sherrington, R. and Pemble, S. E. (1991) *Biochem. J.* **274**, 587–593
- 8 Zhong, S., Spurr, N. K., Hayes, J. D. and Wolf, C. R. (1993) *Biochem. J.* **291**, 41–50
- 9 Pearson, W. R., Vorachek, W. R., Xu, S. J. et al. (1993) *Am. J. Hum. Genet.* **53**, 220–233
- 10 Comstock, K. E., Widersten, M., Hao, X.-Y., Henner, W. D. and Mannervik, B. (1994) *Arch. Biochem. Biophys.* **311**, 487–495
- 11 DeJong, J. L. and Tu, C.-P. D. (1989) in *Glutathione S-Transferase and Drug Resistance* (Hayes, J. D. H., Pickett, C. B. and Mantle, T. J., eds.), pp. 250–261, Taylor and Francis, London
- 12 Suzuki, T., Coggan, M., Shaw, D. C. and Board, P. (1987) *Ann. Hum. Genet.* **51**, 95–106
- 13 Campbell, E., Takahashi, Y., Abramovitz, M., Peretz, M. and Listowsky, I. (1990) *J. Biol. Chem.* **265**, 9188–9193
- 14 Nakajima, T., Elovaara, E., Anttila, S. et al. (1995) *Carcinogenesis* **16**, 707–711
- 15 Strange, R. C., Fryer, A. A., Matheroo, B. et al. (1992) *Biochim. Biophys. Acta* **1139**, 222–228
- 16 Newton, C. R. and Graham, A., eds. *PCR – Polymerase Chain Reaction* (1994), pp. 27–38, Bios Scientific Publishers, Oxford
- 17 Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., *Molecular Cloning: A Laboratory Manual* (1989), 2nd edn., pp. 6.46–6.48, Cold Spring Harbour Laboratory Press, New York
- 18 Watanabe, J., Hayashi, S. and Kawajiri, K. (1994) *J. Biochem.* **116**, 321–326
- 19 Nakachi, K., Imai, K., Hayashi, S., Watanabe, J. and Kawajiri, K. (1991) *Cancer Res.* **51**, 5177–5180
- 20 Albetin, A. M., Hoffer, M., Carlos, M. P. and Miller, J. H. (1982) *Cell* **29**, 319–328
- 21 Shi, Y., Seto, E., Chang, L.-S. and Shenk, T. (1995) *Cell* **81**, 377–388
- 22 Hariharan, N., Kelley, D. E. and Perry, R. P. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9799–9803