European Journal of Human Genetics (2000) 8, 540–544

www.nature.com/ejhg

ARTICLE

Extended structural variation of a pentanucleotide repeat in the GSTP1 gene: characterisation in a normal population and in thyroid and gastric tumours

Cíntia Alves[1], Filipe Silva[1], Leonor Gusmão[1], Raquel Seruca[1], Paula Soares[1], Rui M Reis[1] and António Amorim[1,2]

[1]*Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP);* [2]*Faculty of Sciences, University of Porto, Portugal*

The promoter region of the human *GSTP1* gene contains a polymorphic short tandem repeat (STR) locus consisting of pentanucleotide repeat units (ATAAA). In this work we report the existence of a total of 26 alleles in a Caucasian population. While differences in size (ranging from one to five base pairs) were responsible for the major variation, in five size-defined classes, two alternative sequences were found. Automatic fragment sizing and sequencing analysis revealed that this polymorphism is of a highly complex nature in contrast with previous reports. A genetic population study was carried out on a random sample from Portugal showing no deviation from Hardy-Weinberg equilibrium. Somatic instability studies were also performed on gastric and thyroid tumours using this STR: no instability was detected in thyroid tumour tissues when compared with their normal counterpart but in gastric tumour tissues microsatellite instability (MSI) was detected in 9.6% of the cases and loss of heterozygosity (LOH) also in 9.6% of the cases studied. The results obtained with GSTP1 in gastric cancer were compared with previously reported data on MSI using BAT-26 and several dinucleotide repeat markers. *European Journal of Human Genetics* (2000) **8**, 540–544.

Keywords: GSTP1; STR; polymorphism; Portugal; cancer; instability

Introduction

Short tandem repeats (STRs) are microsatellite DNA sequences which are widely spread throughout the human genome, occurring with a frequency of one locus every 6–10 kb.[1] The repeat unit can vary typically from two to six base pairs in length.[2] They are highly polymorphic in respect to the number of repeats[3] and, in some cases, in respect to the base sequence[4] and in the presence of more or less variable interrupting sequences.[5] Hence, STR loci provide a rich source of genetic markers which are useful in the elaboration of linkage maps, in the identification of disease genes and in DNA typing.

The glutathione S-transferase P1 (GSTP1) belongs to a group of enzymes that have been implicated in cellular detoxification.[6] High levels of GTSP1 expression have been associated with many human solid tumours and seem to determine drug resistance.[7] In this work we study a polymorphic STR locus in the promoter region of the human *GSTP1* gene, consisting of pentanucleotide repeats showing length and sequence variation. This repeat region was originally reported to have allele sizes differing in five base pair increments, with 10 alleles observed.[8] Later, Smith *et al*[9] described this region as having an imperfect TAAAA repeat interrupted by a single TA and TAA insertion, with a total of 11 alleles. In this work we report further variation of this STR in a Caucasian population including more complex size and

Correspondence: Cíntia Alves, IPATIMUP, Rua Dr Roberto Frias s/n, 4200 Porto, Portugal. Tel: +351 22 5570700; Fax: +351 22 5570799; E-mail: calves@ipatimup.pt
Received 20 July 1998; revised 22 February 2000; accepted 29 February 2000

Table 1 DNA fragment size and sequence of the alleles in the GSTP1 locus sequenced in this study

Fragment length (bp)[a]	*Repeat sequence*
186	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_6$ AAA $(ATAAA)_6$
196	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_6$ ATA $(ATAAA)_8$
200	-$(ATAAA)_4$ AAA $(ATAAA)_4$ ATAA $(ATAAA)_4$ ATAA $(ATAAA)_2$ A TTAAA $(ATAAA)_3$
201	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_9$ AAA $(ATAAA)_6$
203	-$(ATAAA)_9$ ATA $(ATAAA)_5$ AT $(ATAAA)_6$ or -$(ATAAA)_6$ ATA $(ATAAA)_5$ AT $(ATAAA)_9$
205	-$(ATAAA)_{12}$ AT $(ATAAA)_9$
206	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{10}$ AAA $(ATAAA)_6$
208	-$(ATAAA)_7$ ATA $(ATAAA)_5$ AT $(ATAAA)_9$
210	-$(ATAAA)_{13}$ AT $(ATAAA)_9$
211	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{11}$ AAA $(ATAAA)_6$
213	-$(ATAAA)_8$ ATA $(ATAAA)_5$ AT $(ATAAA)_9$ or -$(ATAAA)_9$ ATA $(ATAAA)_4$ AT $(ATAAA)_9$
214	-$(ATAAA)_4$ AAA $(ATAAA)_4$ ATAA $(ATAAA)_6$ ATAA $(ATAAA)_7$
215	-$(ATAAA)_{14}$ AT $(ATAAA)_9$
216	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{12}$ AAA $(ATAAA)_6$
218	-$(ATAAA)_9$ ATA $(ATAAA)_5$ AT $(ATAAA)_9$
220	-$(ATAAA)_{15}$ AT $(ATAAA)_9$
222	-$(ATAAA)_5$ AT $(ATAAA)_{10}$ AT $(ATAAA)_9$
223	-$(ATAAA)_{10}$ ATA $(ATAAA)_5$ AT $(ATAAA)_9$ or -$(ATAAA)_9$ ATA $(ATAAA)_5$ AT $(ATAAA)_{10}$
226	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{12}$ AAA $(ATAAA)_8$ or -$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{13}$ AAA $(ATAAA)_7$
228	-$(ATAAA)_9$ ATA $(ATAAA)_5$ AT $(ATAAA)_{11}$ or -$(ATAAA)_{10}$ ATA $(ATAAA)_5$ AT $(ATAAA)_{10}$
229	-$(ATAAA)_3$ ATAAC ACAAA $(ATAAA)_{11}$ ATA $(ATAAA)_4$ AAA $(ATAAA)_5$

[a]length, in base pairs, obtained by automatic fragment sizing.

sequence differences, based on automatic fragment sizing and sequencing analysis.

Furthermore, we present our results on population studies performed on a Portuguese sample and also the results obtained from somatic instability (MSI) and loss of heterozygosity (LOH) studies performed on gastric and thyroid tumours using this STR.

Material and methods

Blood samples and DNA extraction

Population control sample was composed of unrelated volunteer donors and individuals under paternity testing. Blood was obtained by venipuncture and genomic DNA was extracted using the 'salting out' method described by Miller *et al*.[10] Tumour tissues and their normal counterpart were collected from Portuguese patients subjected to surgery at S João Hospital (Oporto) and genomic DNA was extracted using the method described by Mülenbach *et al*.[11]

PCR amplification and fragment visualisation

PCR was carried out using the primers described by Harada *et al*:[8]

5'-AGCCTGGGCCACAGCGTGAGACTACGT (forward)

5'-TCCCGGAGCTTGCACACCCGCTTCACA (reverse).

The amplifications were performed in 25 μl reaction mixture containing 200 μM dNTP (Promega, Madison, USA), 0.25 μM each primer (Pharmacia, Uppsala, Sweden) and 0.5 U AmpliTaq DNA polymerase (Perkin Elmer, Foster City, CA, USA) in 1 × PCR buffer (Perkin Elmer, Foster City, CA, USA). The reaction was carried out in 30 cycles at 94°C for 1 min, 65°C for 1.5 min and 72°C for 30 s.

The PCR products were separated in native polyacrylamide gel electrophoresis (T9C5) using a discontinuous buffer system.[12] The DNA fragments were visualised by silver staining according to Budowle *et al*.[13]

All samples were amplified at least twice and analysed by two different observers.

Sequencing

Individual STR alleles were excised after electrophoresis and extracted from the gel with TE buffer by freezing for 30 min

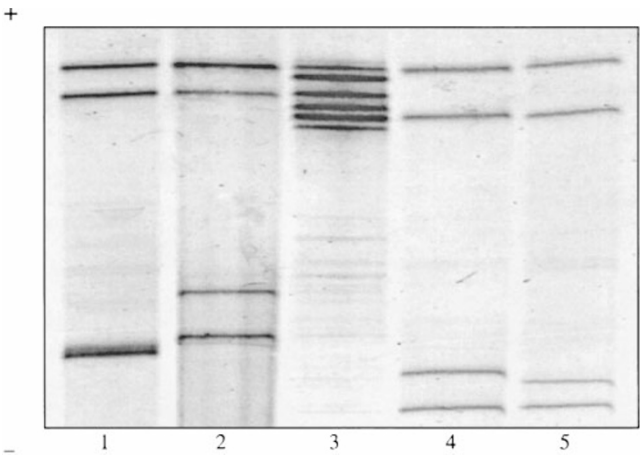

Figure 1 Examples of GSTP1 STR genotypes in heterozygotes with identical homoduplexes but different heteroduplex patterns. 1: 200/213; 2: 200/213'; 3: allelic ladder (200/206/213/218/223/228); 4: 200/223'; 5: 200/223. Alleles with the symbol ' represent sequence variants.

and heating at 65°C for 15 min three times. The extracted STR alleles were then reamplified using the same conditions as described before. The fragments were purified using Microspin S-200 HR columns (Pharmacia, Uppsala, Sweden), according to the manufacturers' instructions.

The sequencing reaction was done with the reverse primer using a Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer, Foster City, CA, USA), in a total reaction volume of 5 µl. The samples were additionally purified with an MgCl2 and ethanol based protocol.

Electrophoresis was carried out on an ABI PRISM 377 DNA Sequencer (Perkin Elmer, Foster City, CA, USA). Samples were loaded on a pre-warmed 6% polyacrylamide gel containing 6 M urea and 1 × TBE buffer and the running conditions were 1680 V, 50 mA, 150 W, 51°C during approximately 3.5 h. The results were analysed using the ABI PRISM 377-18 Data Collection software (version 1.1).

Genotype analysis

Genotyping was done by automatic fragment sizing using the ABI PRISM 377 DNA Sequencer and the GeneScan Analysis 2.1 software (Perkin Elmer, Foster City, CA, USA). All samples were run with an internal size standard (GeneScan TAMRA-500) and compared with a ladder containing the most frequently observed alleles. Allele frequencies were estimated from a sample of 279 unrelated individuals from Portugal. Hardy-Weinberg equilibrium was assayed by Guo and Thompson exact test,[14] using GENEPOP software.[15]

Results and discussion

A total of 21 alleles differing in size were identified in a sample of 279 individuals from a Portuguese Caucasian population. Sequencing analysis revealed their repeat unit structure and the presence of sequence variants in five alleles. The basic repeat structure consists of ATAAA units. Table 1 shows the highly complex structural variability within the GSTP1 alleles.

Apart from the sequencing analysis, structural variants that exist in alleles of the same size can also be distinguished in native polyacrylamide gels by heteroduplex analysis. In Figure 1, two examples are shown in which the heteroduplex patterns are different in heterozygous individuals with identical homoduplexes. Even though only variants in alleles 203, 213, 223, 226 and 228 bp were detected in this work, further variants may exist.

Our population studies were performed on a sample of 279 unrelated Portuguese individuals. Gene frequencies were estimated (Table 2) and the population showed no deviation from Hardy-Weinberg equilibrium (Table 3). The observed heterozygosity is 0.72, not significantly deviating from the expected value of 0.74, and the polymorphic information content (PIC) is 0.70. Analysis of nuclear families demonstrated co-dominant inheritance of the alleles and no germinal instability was detected in 81 mother/child pairs.

Table 2 Allele frequencies (size variants) in our Portuguese sample

Allele (bp)	*No. of alleles observed*	*Frequency (%)*
186	2	0.36
196	1	0.18
200	229	41.03
201	5	0.90
203	8	1.43
205	7	1.25
206	6	1.08
208	1	0.18
210	14	2.51
211	74	13.26
213	6	1.08
214	5	0.90
215	15	2.69
216	1	0.18
218	151	27.05
220	2	0.36
222	1	0.18
223	24	4.30
226	1	0.18
228	3	0.54
229	2	0.36
Total	558	100.00

These facts, together with the high level of variation, demonstrate that this polymorphism can be a useful tool in evolutionary biology studies.

Somatic instability studies using this STR were performed on 52 gastric tumour samples (30 intestinal, 13 atypical and nine diffuse types using Lauren's Histological Classification) and 20 thyroid tumour samples (11 goitres, four follicular adenomas, four papillary carcinomas and one poorly differentiated carcinoma). No genotypic alterations in the GSTP1 STR were detected in thyroid tumours when compared with the normal counterpart. Genotypic alterations in dinucleotide repeat markers have been described before in thyroid tumours,[16] but in a far lower frequency than the observed in colon and gastric cancers. These 20 thyroid tumours were previously tested for MSI[16] using dinucleotide repeat markers and instability was detected in two cases and found only in one out of six loci analysed. Our results confirm stability of the GSTP1 locus and no association with tumour development in this type of tissue. However, the fact that the genotypic distribution in these cases shows significant deviations to the Hardy-Weinberg equilibrium (Table 4), does not exclude some kind of susceptibility genotypes.

In our 52 gastric tumour samples, microsatellite instability was detected in 9.6% (five in 52) of the cases studied (three intestinal, one diffuse and one atypical). LOH was also detected in another five cases (three intestinal and two atypical). Examples of some of the differences observed between genotypes of tumour tissues and their normal counterpart are shown in Figure 2.

From the 52 gastric tumour samples analysed for GSTP1, 50 samples were previously analysed for MSI using other markers, including mono and/or dinucleotide repeats

Table 3 Genotype distributions in our Portuguese sample

	No. of individuals	
Observed genotypes	*Observed*	*Expected*
186–200	1	0.822
186–218	1	0.542
196–200	1	0.411
200–200	46	46.869
200–201	3	2.056
200–205	2	2.878
200–206	3	2.467
200–210	8	5.756
200–211	24	30.424
200–213	5	2.467
200–214	4	2.056
200–215	6	6.167
200–216	1	0.411
200–218	63	62.081
200–220	1	0.822
200–223	12	9.867
200–226	1	0.411
200–228	1	1.233
200–229	1	0.822
201–211	1	0.664
201–218	1	1.355
203–205	1	0.101
203–211	3	1.063
203–218	3	2.169
203–223	1	0.345
205–215	1	0.189
205–218	2	1.898
205–223	1	0.302
206–211	1	0.797
206–218	2	1.627
208–210	1	0.025
210–211	1	1.860
210–218	3	3.795
210–228	1	0.075
211–211	8	4.849
211–214	1	0.664
211–215	4	1.993
211–218	19	20.061
211–223	4	3.189
213–218	1	1.627
215–218	2	4.066
215–223	2	0.646
218–218	23	20.332
218–220	1	0.542
218–222	1	0.271
218–223	4	6.506
218–228	1	0.813
218–229	1	0.542
Other	0	18.072
Total = 279; P = 0.834 (s.d. ± 0.009)		

P = probability; s.d. = standard deviation.

(Table 5). Of the five MSI positive cases for the GSTP1 locus, three showed instability using BAT-26 and/or dinucleotide markers, two of them only with BAT-26 (data not shown).

Of the 50 cases where GSTP1 data (MSI) could be compared with previously reported results on MSI[17] only 8 of the GTSP1 MSI negative cases showed MSI in both mono and/or dinucleotide markers that were studied (classes 4 and 5, Table 5).

Table 4 Genotype distributions in gastric and thyroid tumour samples

	No. of individuals			
	Gastric		*Thyroid*	
Observed genotypes	*Observed*	*Expected*	*Observed*	*Expected*
186–213	0	0.00	1	0.13
196–218	1	0.30	0	0.00
200–200	10	10.05	6	4.39
200–206	1	0.89	0	0.00
200–213	3	4.47	3	2.44
200–214	1	0.45	0	0.00
200–216	2	1.79	0	0.00
200–218	15	13.85	4	3.90
200–223	2	1.34	0	1.46
200–226	1	0.45	0	0.00
200–228	1	0.45	0	0.00
203–218	1	0.30	0	0.00
206–218	1	0.60	0	0.00
208–213	1	0.10	0	0.13
208–223	0	0.03	1	0.08
211–211	0	0.01	1	0.03
211–216	1	0.08	0	0.00
211–218	1	0.60	0	0.41
213–218	6	3.01	1	1.03
216–218	1	1.20	0	0.00
218–218	2	4.52	1	0.72
218–223	1	0.90	1	0.62
223–229	0	0.00	1	0.08
Other	0	4.10	0	4.58
Total	52	52.00	20	20.00
	P = 0.741; s.d. ± 0.009		P = 0.026; s.d. ± 0.007	

P = probability; s.d. = standard deviation.

We can observe (Table 5) that in gastric cancer there are cases (class 2) where MSI is positive for GSTP1 but negative for BAT-26 and/or dinucleotide markers and, conversely, cases where MSI is negative for GSTP1 but positive for BAT-26 and/or dinucleotide markers (classes 4 and 5). We observe further that one tumour case with LOH in GSTP1, is MSI positive in BAT-26 and/or dinucleotide markers. Another

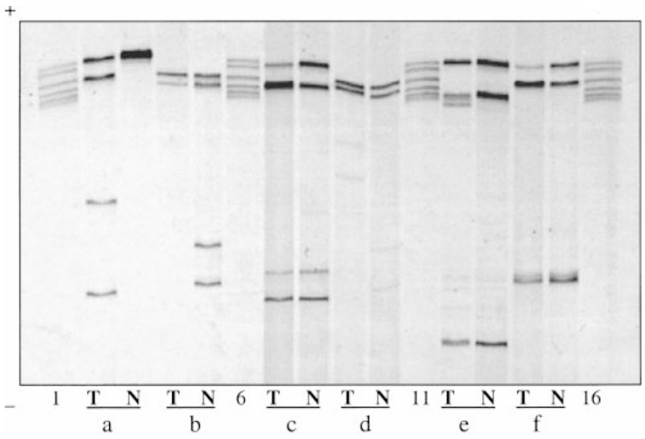

Figure 2 Examples of genotypes observed in gastric tumour tissues and their normal counterpart: 1, 6, 11 and 16 ladder; N: normal tissue; T: tumour tissue. **a**: MSI; **b**: LOH; **c**: LOH; **d**: MSI; **e**: MSI; **f**: LOH.

Table 5 Comparison of MSI and LOH results in gastric tumour cases obtained from the GSTP1 locus with those obtained from BAT-26 and/or dinucleotide markers

		GSTP1		*BAT-26 and/or dinucleotide markers*
Classes	*No. of gastric tumour cases observed*	*MSI*	*LOH*	*MSI*
1	3 (2I + 1A)	+	–	+
2	2 (1I + 1D)	+	–	–
3	3 (1I + 2A)	–	+	–
4	1 (1I)	–	+	+
5	7 (5I + 2A)	–	–	+
6	34 (19I + 8A + 7D)	–	–	–
Total	50 (13A + 29I + 8D)[a]			

[a]in two cases, a comparison study was not possible since instability was not analysed using BAT-26 and/or dinucleotide markers; A: atypical; I: intestinal; D: diffuse.

finding is that a diffuse type of gastric tumour studied (class 2) showed MSI in GSTP1, which is interesting considering that this is usually not observed when using mono- and dinucleotide markers.[17]

These observations suggest that the repair mechanisms which are inactivated in the GSTP1 STR are distinct from those operating on mono- and dinucleotide repeats.

The lack of association between this STR and mono- and dinucleotide MSI diagnostic markers does not exclude the importance of this polymorphism in tumorigenesis. Indeed, it has been shown that GSTP1 is expressed at high levels in solid tumours.[7] It remains to be demonstrated if the polymorphism of this STR, located in the promoter region of the gene, has some involvement on the expression levels.

Acknowledgements

This work was partially supported by JNICT Programa PRAXIS XXI ref. BM/7112/95, BD/3127/94, BPD/11812/97 and 2/2.1/BIA/100/94.

References

1 Beckmann JS, Weber JL: Survey of human and rat microsatellites. *Genomics* 1992; **12**: 627–631.
2 Koreth J, O'Leary JJ, McGee JOD: Microsatellites and PCR genomic analysis. *J Pathol* 1996; **178**: 239–248.
3 Fowler JCS, Burgoyne LA, Scott AC, Harding HW: Repetitive deoxyribonucleic acid (DNA) and human genome variation. A concise review relevant to forensic biology. *J Forensis Sci* 1988; **33**: 1111–1126.
4 Möller A, Meyer E, Brinkmann B: Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. *Int J Leg Med* 1994; **106**: 319–323.
5 Urquhart A, Kimpton CP, Downes TJ, Gill P: Variation in Short Tandem Repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Leg Med* 1994; **107**: 13–20.
6 Moffat GJ, McLaren AW, Wolf CR: Sp1-mediated transcriptional activation of the human Pi class glutathione S-transferase promoter. *J Biol Chem* 1996; **271**: 1054–1060.
7 O'Brien ML, Tew KD: Glutathione and related enzymes in multidrug resistance. *Eur J Cancer* 1996; **32A**: 967–978.
8 Harada S, Nakamura T, Misawa S: Polymorphism of pentanucleotide repeats in the 5' flanking region of glutathione S-transferase (GST) π gene. *Hum Genet* 1994; **93**: 223–224.
9 Smith CM, Bora PS, Bora NS, Jones C, Gerhard DS: Genetic and radiation-reduced somatic cell hybrid sublocalization of the human *GSTP1* gene. *Cytogenet Cell Genet* 1995; **71**: 235–239.
10 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.
11 Müllenbach R, Lagoda PJL, Welter C: An efficient salt-chloroform extraction of DNA from blood and tissues. *Trends Genet* 1989; **5**: 391.
12 Luis JR, Caeiro B: Application of two STRs (VWF and hTPO) to human population profiling. A survey in Galicia. *Hum Biol* 1995; **67**: 789–795.
13 Budowle B, Chakraborty R, Gusti AM, Eisenberg AJ, Allen RC: Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am J Hum Genet* 1991; **48**: 137–144.
14 Guo SW, Thompson EA: Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* 1992; **48**: 361–372.
15 Raymond M, Rousset F: GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 1995; **86**: 91–102.
16 Soares P, Dos Santos NR, Seruca R, Lothe RA, Sobrinho-Simões M: Benign and malignant thyroid lesions show instability at microsatellite loci. *Eur J Cancer* 1997; **2**: 293–296.
17 Oliveira C, Seruca R, Seixas M, Sobrinho-Simões M: The clinicopathological features of gastric carcinomas with microsatellite instability may be mediated by mutations of different 'target genes'. *Am J Pathol* 1998; **153**(4): 1211–1219.