HUMAN MUTATION 13:294–300 (1999)

RESEARCH ARTICLE

Low Yield of Polymorphisms From EST Blast Searching: Analysis of Genes Related to Oxidative Stress and Verification of the P197L Polymorphism in GPX1

Lena Forsberg,[1]* Ulf de Faire,[2] and Ralf Morgenstern[1]

[1]*Division of Biochemical Toxicology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden*
[2]*Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden*

Communicated by R.G.H. Cotton

To determine new polymorphisms in the antioxidant enzymes superoxide dismutase, glutathione peroxidases, catalase, and microsomal glutathione transferase 1, a search of the human expressed sequence tags (EST) database was performed (with BLAST 2.0). When any mutation, indicated by the BLAST search, gave rise to a nonconservative amino acid change we performed polymerase chain reaction (PCR) restriction analysis and/or sequence analysis of genomic DNA from human subjects in order to verify these potential polymorphisms. Of nine indicated polymorphisms from the EST analysis found in four different antioxidant enzymes, we could verify one, an amino acid substitution Pro-Leu at amino acid position 197 (P197L), in the glutathione peroxidase 1 gene. The corresponding allele frequencies were ≈70/30%. In addition, a silent mutation (1167T/C) in the catalase gene indicated by the BLAST search could also be verified. Six to nine individuals were analyzed per indicated polymorphism, so that only common polymorphisms would be found. The indicated mutations not verified by direct analysis thus cannot be excluded as allelic variation in the human population. These results show that the EST database can be used to search for polymorphisms in genes with high abundance in the human EST database. In addition to the EST analysis, PCR/single-strand conformation polymorphism (SSCP) was employed for the analysis of the microsomal glutathione transferase 1 gene. No polymorphism in the coding sequence could be detected in the gene by either method. The high degree of conservation of the microsomal glutathione transferase 1 gene indicates an important physiological function for this enzyme. Hum Mutat 13:294–300, 1999.

KEY WORDS: **EST; polymorphisms; PCR-SSCP**

INTRODUCTION

Under normal conditions, living organisms maintain a steady-state balance between the production of reactive oxygen metabolites ($^{\cdot}OH$, $O_2^{\cdot-}$ and H_2O_2) and their destruction by antioxidant molecules (e.g., ascorbic acid, glutathione, α-tocopherol), as well as by antioxidant enzymes—superoxide dismutases (SODs), catalase, and glutathione peroxidases (GPXs) [Michiels et al., 1993]. This steady-state balance can be perturbed, however, by increasing the free radical production or by decreasing the defense systems leading to a situation of "oxidative stress" [Sies, 1991]. The reactive oxygen metabolites (ROM) can cause various types of damage to both mitochondrial and genomic DNA, oxidation of proteins and nonprotein thiols and lipid peroxidation [Halliwell et al., 1989].

Three distinct SODs have been described in eukaryotes, one containing copper and zink (SOD1; MIM# 147450) found in cytoplasm, and one containing manganese in its active site found in mitochondria (SOD2; MIM# 147460) [Fridovich, 1985]. The third variant is an extracellular SOD (SOD3; MIM# 185490) [Marklund, 1990].

Received 21 July 1998; accepted revised manuscript 4 December 1998.

*Correspondence to: Lena Forsberg, Institute of Environmental Medicine, Division of Biochemical Toxicology, Karolinska Institute, Box 210, S-171 77 Stockholm, Sweden; Fax: 46-8-334467; E-mail: lena.ekstrom@imm.ki.se

Catalase (CAT; MIM# 115500) appears to be coded by a single-copy gene and is concentrated in peroxisomes in eukaryotes [Masters et al., 1986]. In acatalasemia red blood cell catalase is lowered [Ogata, 1991]. Glutathione peroxidases are selenoproteins (Se-GPXs) containing one atom of selenium as a selenocysteine involved in their catalytic activity [Michiels et al., 1993]. GPX reduces lipidic or nonlipidic hydroperoxides while oxidizing two molecules of glutathione (GSH). Four distinct GPXs exist in eukaryotes: GPX1 (MIM# 138320) is found in cytoplasm and mitochondria [Ursini et al., 1985], GPX2 (MIM# 138319) is a cytosolic form found in liver and colon in human [Chu et al., 1993], GPX3 (MIM# 138321) is an extracellular form found in plasma [Takahashi et al., 1987], and GPX4 (MIM# 138322) is expressed at high levels in testis and is associated to various degrees with intracellular membranes [Ursini et al., 1986]. The latter, uniquely, is able to reduce also phospholipid hydroperoxides. Microsomal glutathione transferase (MGST1; MIM# 138330) belongs to an unrelated enzyme superfamily and is also capable of reducing phospholipid hydroperoxides [Mosialou et al., 1995].

Together, these enzymes constitute vital defenses against oxidative stress; hence, any impairment of function could translate into increased disease susceptibility when, and if, oxidative stress is a significant component of a particular disease.

A major issue in oxidative stress research is indeed its impact on various disorders such as cancer, cardiovascular and neurodegenerative diseases [Halliwell et al., 1989]. Attempts to shed light on a possible link between oxidative stress and disease have often involved the measurement of antioxidant levels and or their supplementation [Gey, 1993]. The results of these studies have often been difficult to interpret based in part on a considerable variation in dietary levels of various antioxidants [Omenn, 1996]. In an effort to study the importance of oxidative stress in disease, we have undertaken to determine the genetic and consequent phenotypic variation of protective enzyme systems in the human population. As a first approach, we have searched for candidate gene variation using the EST database and by single-strand conformation polymorphism (SSCP) analysis. We were able to detect several nucleotide changes in various enzymes and an amino acid substitution in GPX1.

MATERIALS AND METHODS

Isolation of Genomic DNA

Genomic DNA was prepared from whole blood using Genomic DNA purification kit (Qiagen, Germany) according to the manufactures description. Caucasian subjects were blood donors throughout.

BLAST Search

A gapped BLAST 2.0 search (http://www.ncbi.nlm.nih.gov/BLAST/) [Altschul et al., 1997] of the human expressed sequence tags (EST) database was performed against the cDNA sequence of the genes of interest. The analysis was limited to 100 EST database entries and any low-quality sequence was automatically removed by using the gapped option of BLAST. There is an indication that if sequence variants appeared in two or more of the ESTs, an allelic variations exists in the human population. When the indicated mutation gave rise to a nonconservative amino acid change we performed polymerase chain reaction (PCR)-restriction analysis and/or sequence analysis of genomic DNA from human subjects to verify these potential polymorphisms.

SSCP Analysis of the MGST1 Gene

Genomic DNA (0.1 μg) was used as a template for PCR reactions using the primers listed in Table 1. PCR was carried out in 30 cycles, each involving denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1.5 min. One of the PCR products (fragment A, according to Fig. 1) was digested with *Hha*I (New England BioLabs, Beverly, MA) in order to obtain two fragments of

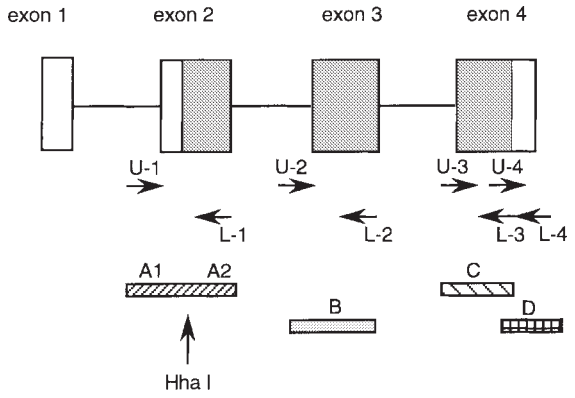

FIGURE 1. **Map of DNA and restriction fragments used for single-strand conformation polymorphism (SSCP) analysis of the MGST1 gene. *Arrows*, polymerase chain reaction (PCR) primers (see Table 1).**

TABLE 1. Sequences of Primers Used for PCR and Sequence Analysis

Gene	Primer	Sequence
CAT	Upper-1	5′-GCCGCCTTTTTGCCTATCCT-3′
	Lower-1	5′-TCCCGCCCATCTGCTCCACG-3′
GPX1	Upper-1	5′-GCCTGGTGGTGGGTTCGAGCC-3′
	Lower-1	5′-GACAGCAGCACTGCAACTGCC-3′
	Upper-2[a]	5′-CAACTTCATGCTCTTCGAGAA-3′
GPX3	Upper-1	5′-AGGTGGAGGCTTTGTCCCTAATT-3′
	Lower-1	5′-CCAAGAAATTCCCCAAGAGTCC-3′
SOD1	Upper-1	5′-GGTGCTGGTTTGCGTCGTAG-3′
	Lower-1	5′-TTCTGCTCGAAATTGATGAT-3′
MGST1	Upper-1	5′-TTTATTGCTTATACTGCAC-3′
	Lower-1	5′-TTGTCAATCTATAGAATGCAGTTGC-3′
	Upper-2	5′-TATGGGGACTCCCTATGTTGC-3′
	Lower-2	5′-CTGCGTACACGTTCTACTCTGTC-3′
	Upper-3	5′-CAGAGCCCACCTGAATGACCTT-3′
	Lower-3	5′-GAGCAAGCTTTTACAGGTACAATTTACTTTTC-3′
	Upper-4	5′-CCTCTACAGCCATCCTGCACTT-3′
	Lower-4	5′-ATTCCTCTGCTCCCCTCCTAC-3′

[a]Nested primer used for sequence analysis.

208 bp and 492 bp, respectively. For the analysis, 3 μl of the PCR products was mixed with 27 μl of loading buffer, containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol. Before loading, the samples were denatured at 100°C for 5 min and chilled on ice. Electrophoretic separation of single-stranded DNA fragments was then analyzed in 8.7% polyacrylamide (BioRad) gels with 10% glycerol. The gels were run in 2× Tris-borate-EDTA buffer [Maniatis et al., 1989] at 4°C for 4–7 hr at 270 V, and the DNA fragments were visualized by silver staining.

Silver Staining

The gels were fixed in 40% methanol/10% acetic acid overnight and washed twice in 10% ethanol, 5% acetic acid. The gels were then oxidized in 3.4 mM $K_2Cr_2O_7$, 3.2 mM HNO_3 for 5 min, washed twice in H_2O, and then stained in 12 mM $AgNO_3$ for 20 min. Bands were developed in 0.28 M Na_2CO_3, 0.02% formaldehyde.

DNA Sequence Analysis

The PCR-SSCP products of the MGST1 gene were sequenced by the PCR product sequencing kit (U.S. Biochemicals, Cleveland, OH) using a ^{32}P-labeled ddNTP (Amersham, England). The PCR products amplified by the other primer listed in Table 1 were purified using the High Pure PCR Product Purification Kit (Boehringer Mannheim, Germany) and the sequence analyzed using the PRISM ready reaction Dyedeoxy™ terminator cycle sequencing kit (Perkin-Elmer) and an ABI model 373 DNA Sequencer. The primers used for amplification were also used for sequencing on both strands of the amplified DNA, except for the GPX1 gene, where a nested primer was used as shown in Table 1.

Polymerase Chain Reaction

DNA was amplified with the primers shown in Table 1. All amplifications were carried out for 30 cycles with 0.15 mM dNTP, 20 pmol of respective primer, and 0.5 U *Taq* Polymerase (Sigma) in the supplied buffer, using the following annealing temperatures: GPX1, 62°C; catalase + GPX3, 56°C; SOD1, 52°C; and microsomal GST 1 products A–D, 55°C. All PCRs were carried out in 30 cycles, each involving denaturation at 94°C for 1 min, annealing for 1 min, and extension at 72°C for 1.5 min. The amplifications of SOD1 and GPX1 also included 10% dimethylsulfoxide (DMSO).

Restriction Analyses

The amplified PCR product of catalase and GPX1 were cleaved with the restriction enzymes *Bst*XI and *Dde*I, respectively (New England BioLabs, Beverly, MA) according to the manufacturers description.

RESULTS

BLAST Search

Table 2 shows possible indications of new polymorphisms resulting in nonconservative amino acid substitutions in oxidative stress-related genes according to BLAST searches. These were chosen for further analysis (several additional polymorphisms that result either in conservative codon changes or silent mutations are also indicated). In addition, one silent polymorphism in the catalase

TABLE 2. Results of BLAST Searches*

Gene	cDNA sequence[a]	Indicated polymorphisms		No. of ESTs[b]/ no. of total ESTs	No. of polymorphic alleles/ no. of alleles sequenced
		Base-pair change[b]	Amino acid change[b]		
CAT	X04076	1167T/C	Silent D388D	6/12	5/16
GPX1	X13709	349G/C	G116R	2/50	0/12
		365C/T	P121L	4/50	0/12
		593C/T	P197L	4/100	5/12
GPX3	X71973	452T/C	F150S	3/15	0/16
SOD1	X01780	41-42CA/AG	P13Q	5/30	0/16
		43G/T	V14L	14/30	0/16
MGST1	J03746	209A/G	E69G	3/26	0/40[c]
		357T/G	I118M	2/29	0/40[c]
		365T/G	L121W	2/29	0/40[c]

*Silent, 3′-noncoding or conservative mutations indicated by EST search results not included in Table 2: SOD1 silent 600T/C (2/17); SOD2 silent 18G/T (3/69); MGST1 3′-mutation 657G/T (2/50); GPX1 silent 186A/T-G (3/17), conservative 464A/G (2/67); GPX3 silent 639T/C (3/15) conservative 31T/C (6/13). Nucleotides are numbered from the initiator ATG and the parenthesis show the no. of ESTs indicating the polymorphism in relation to the ESTs analyzed that cover the area.
[a]The Genebank accession number of the cDNA sequence used as query in the gapped BLAST search. The numbering of nucleotides is from the initiator ATG. Numbering of amino acids is made excluding the initiator.
[b]Number of human ESTs that included the new indicated polymorphic site.
[c]In addition, PCR-SSCP was used for 60–75 individuals. See Figure 1.

gene is included. Genbank accession numbers of the cDNA sequences used as query in the BLAST search and the number of EST clones that include the indicated polymorphism are also given. SOD2, GPX2, and GPX4, which were also analyzed for sequence variations performing a BLAST search, showed no nonconservative mutations.

Detection of Indicated Polymorphisms

Since the present search was aimed at finding common polymorphisms DNA samples from 6–20 individuals were analyzed.

CAT

DNA from eight individuals was amplified by PCR, using the primers listed in Table 1. The indicated polymorphism in Table 2 could be detected in three of these individuals by direct sequence analysis. The T-C point mutation at position 1167 (1167T/C) causes a silent mutation at amino acid position D388. Restriction analyses with the enzyme *Bst*XI (Fig. 2a) of 58 individuals showed 44 homozygotes for C, 2 homozygotes for T, and 12 heterozygotes.

SOD1

DNA from nine individuals was amplified by PCR using the primers listed in Table 1. None of the indicated polymorphisms could be detected.

GPX1

DNA from six individuals was amplified by PCR and the sequence determined by a nested primer, as listed in Table 1. A point mutation C-T 593 (593C/T) causing the amino acid substitution

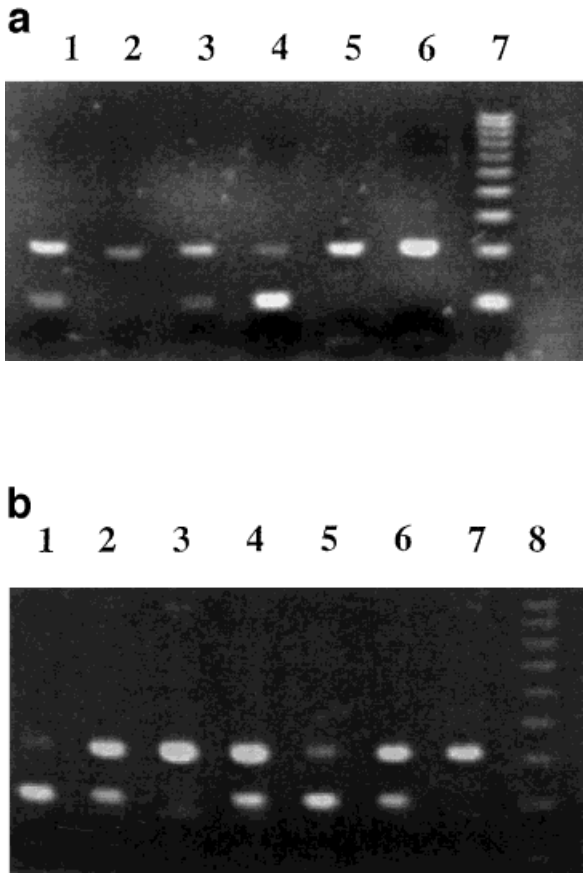

FIGURE 2. **Restriction analyses. a:** Catalase: six samples after digestion with *Bst*XI; *lanes 2,5,6*, homozygous C 1167; *lane 4*, homozygous T 1167; *lanes 1,3*, heterozygous C/T 1167; *lane 7*, 100 bp ladder. **b:** GPX1: 7 samples after digestion with *Dde*I; *lanes 3,7*, homozygous Pro 197; *lanes 1,5*, homozygous Leu 197; *lanes 2,4,6*, heterozygous Pro/Leu 197; *lane 8*, 100-bp ladder.

P197L was identified in three individuals, two homozygotes, and one heterozygote. Restriction analyses with the enzyme *Dde*I (Fig. 2b) of 25 individuals showed 13 homozygotes for Pro, 3 homozygotes for Leu, and 9 heterozygotes. The other two suspected polymorphisms were not detected in six individuals.

GPX3

DNA from eight individuals was amplified by PCR, using the primers listed in Table1, and the indicated polymorphism 452T/C was not detected by direct sequence analysis.

MGST1

DNA from 20 individuals was amplified by PCR, and the sequence or restriction pattern was analyzed. None of the indicated polymorphisms could be detected.

SSCP Analysis of MGST1

In order to investigate any common polymorphic alleles in the MGST1 gene, SSCP analysis was also carried out on all exons and most exon–intron junctions (according to Fig. 1), using DNA from 60–75 individuals. The primer pairs indicated yielded PCR products A–D (Fig 1). No polymorphisms were detected in fragment A1, A2, or C. SSCP analysis of product B showed an altered electrophoretic pattern in 1 of 60 subjects. Sequence analysis revealed two base pair changes in intron 2, –34, and –59 nt from the exon 3-intron 2 boundary, respectively. SSCP analysis of product D revealed altered electrophoretic mobility in two of 70 individuals investigated (Fig. 3a). Direct sequence analysis revealed a T-G substitution at nucleotide position 598 in the noncoding 3′-end. Both donors were heterozygotes (Fig. 3b). In summary, no nucleotide changes resulting in amino acid changes were detected. Furthermore, the mutations indicated by BLAST searches of the EST database could not be verified by this approach, nor by the direct analysis (vide supra).

DISCUSSION

Genetic analysis of polymorphisms in antioxidant enzyme genes may be of interest in populations with different disease states whose pathogenesis involves a relative or absolute excess of free radicals relative to antioxidants. In order to detect new polymorphisms in antioxidant enzymes such as GPXs, SODs catalase, and MGST1, we performed a computer analysis with a BLAST 2.0 search of the human EST database. We showed earlier that already known polymorphisms could be identified in this way (Technical Tips Online (http://www.elsevier.com/locate/tto) T01440). Since ESTs are highly error prone, there is a risk that a sequence variation is a sequence error, rather than a polymorphism. By using only alignments with high-quality sequence, (e.g., identical to wild-type sequence around the indicated sequence variation), we attempted to diminish the risk of sequence mistakes. If the same sequence variation is found in more than one EST (or preferably in several) originating from different libraries, the chances that allelic variations actually exists in the human population increases further. Several nucleotide substitutions were indicated and, as an initial approach, we therefore limited our direct analysis to sequence variations that caused the most non conservative amino acid changes (which are more likely to influence enzyme activity).

The epidemiological screening of candidate genes is greatly facilitated when common polymorphisms are analyzed, therefore only 6–20 individuals were analyzed per indicated polymorphism. For this reason, even though most of the indicated polymorphisms were not verified by direct sequence analysis, they cannot yet be excluded as a polymorphism in the human population.

As shown in Table 2, we attempted to verify 10 of the indicated polymorphisms from the EST analysis found in five different antioxidant enzymes. Two of these polymorphisms could be verified. One is a silent polymorphism predicted in the EST database, a nucleotide substitution 1167T/C in catalase. The other is a common polymorphism resulting in an amino acid substitution in GPX1. A sequence variant not supported by the original cDNA also indicates that at position 272 of GPX1, a T is present instead of A (272A/T). Thus, applying this approach in the search for allelic variation in genes with high abundance in the human EST database will save time and resources compared with established methods such as SSCP. For instance, this method could be one of choice when laying out a net of gene specific polymorphic markers.

When an exhaustive search for polymorphic sites is attempted, SSCP analysis (where the sensitivity for detection of mutation has been found to be 70% [Vidal-Puig et al., 1994] using polyacrylamide gels in the presence of glycerol) or direct sequence analysis are methods of choice. PCR-SSCP was employed for the analysis of the MGST1 gene in addition to the EST analysis.

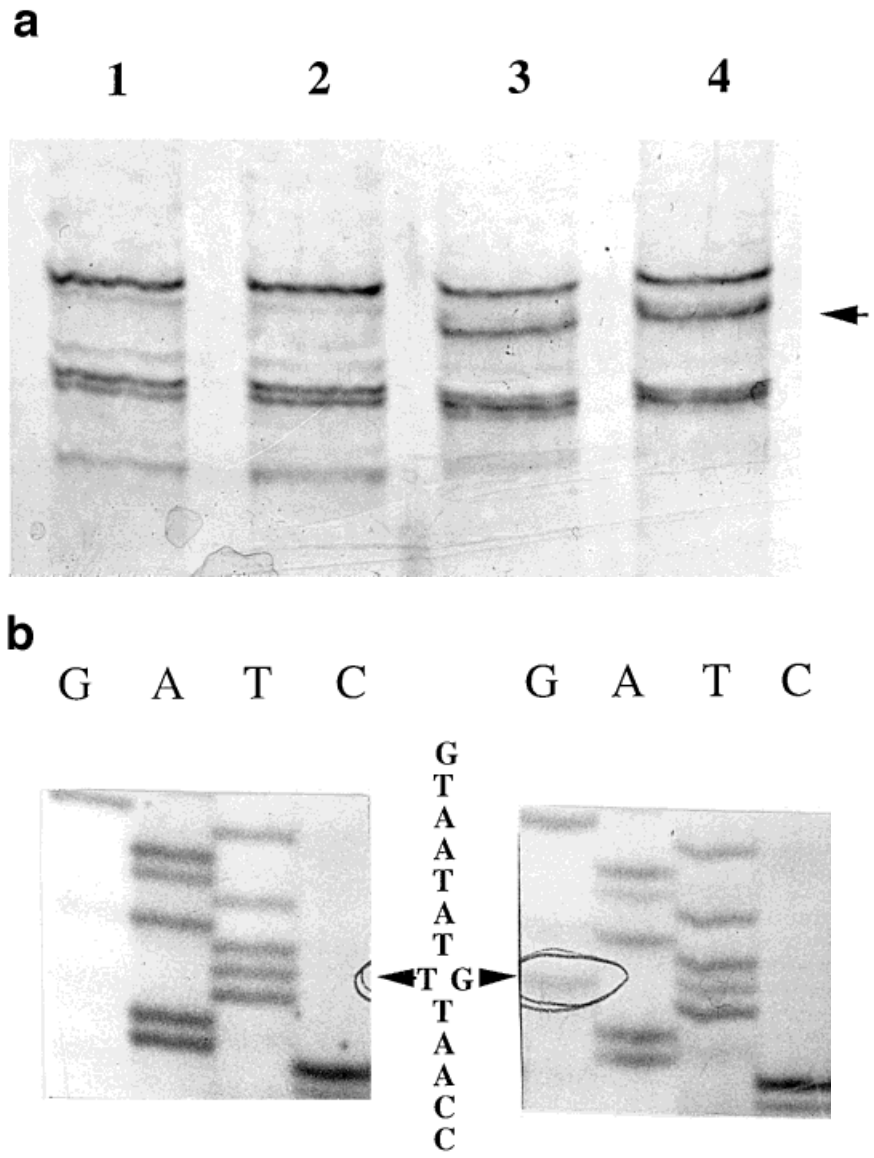

FIGURE 3. **Identification of a mutation in fragment D in the MGST1 gene. a: Single-strand conformation polymorphism (SSCP) analysis of fragment D in the MGST1 gene. *Arrow* (side), the extra band seen in samples 3 and 4. Samples 1 and 2 are wild type. b: Direct sequencing of polymerase chain reaction (PCR) fragment D from wild type and a mutant individual. *Arrow*, T-G point mutation.**

No polymorphism in the coding sequence could be detected in the MGST1 by either method (considering that three indicated polymorphic sites from the EST database could not be verified experimentally). The high degree of conservation of the gene thus indicates an important physiological function for this enzyme.

The above results show that the EST database can be used to search for polymorphisms. One should expect that perhaps 2 of 10 indicated polymorphisms are indeed common and can be verified. Clearly, when aiming at non-conservative changes, a lot of apparent false-positives can be expected. In addition, possible ethnic variation could contribute (e.g., EST libraries are from the United States, whereas a Swedish population was sampled in this survey). Unfortunately, the ethnic origin of the DNA libraries in the human EST database is not given in most cases, making it difficult to discern between these possibilities.

In summary, a common Pro-Leu polymorphism has been identified in the GPX1 gene. After having verified any phenotypic conse-

quences of this variation, screening of this mutation could be used as a tool to study oxidative stress in disease states.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman. 1997. Gapped BLAST and PSI-BLAST a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Chu FF, Doroshow JH, Esworthy RS. 1993. Expression, characterization, and tissue distribution of a new cellular selenium-dependent glutathione peroxidase, GSHPx-GI. J Biol Chem 268:2571–2576.

Fridovich I. 1985. Superoxide dismutases: regularities and irregularities. Harvey Lect 79:51–75.

Gey K. 1993. Prospects for the prevention of free radical disease, regarding cancer and cardiovascular disease. Br Med Bull 49:679–699.

Halliwell B, Gutteridge JMC. 1989. Free radicals in biology and medicine. Oxford: Clarendon Press.

Maniatis T, Fritsch EF, Sambrook J. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor: NY:Cold Spring Harbor Laboratory Press.

Marklund S. 1990. Expression of extracellular superoxide dismutase by human cell lines. Biochem J 299:213–219.

Masters C, Pegg M, Crane D. 1986. On the multiplicity of the enzyme catalase in mammalian liver. Mol Cell Biochem 70:113–120.

Michiels C, Raes M, Toussaint O, Remacle J. 1993. Importance of Se-glutathione peroxidase, catalase, and Cu/Zn-SOD for cell survival against oxidative stress. Free Radical Biol Med 17:235–248.

Mosialou E, Piemonte F, Andersson C, Vos R, Van Bladeren PJ, Morgenstern R. 1995. Microsomal glutathione transferase-lipid-derived substrates and lipid dependence. Arch Biochem Biophys 320:210–216.

Ogata M. 1991. Acatalasemia. Hum Genet 86:331–340.

Omenn G. 1996. Micronutrients (vitamins and minerals) as cancer-preventive agents. IARC Sci Publ 139:33–45.

Takahashi K, Avissar N, Whitin J, Cohen H. 1987. Purification and characterization of human plasma glutathione peroxidase: a selenoglycoprotein distinct from the known cellular enzyme. Arch Biochem Biophys 256:677–686.

Ursini F, Maiorino M, Gregolin C. 1985. The selenoenzyme phospholipid hydroperoxide glutathione peroxidase. Biochim. Biophys. Acta 839:62–70.

Ursini F, Maiorino M, Gregolin C. 1986. Phospholipid hydroperoxide glutathione peroxidase. Int J Tissue React 8:99–103.

Vidal-Puig A, Moller DE. 1994. Comparative sensitivity of alternative single-strand conformation polymorphism (SSCP) methods. Biotechniques 17:490–496.