Human Thymine-DNA Glycosylase Maps at Chromosome 12q22-q24.1: A Region of High Loss of Heterozygosity in Gastric Cancer

Christoph Schmutte, Raffaele Baffa, Luisa M. Veronese, Yoshiki Murakumo, and Richard Fishel

Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

ABSTRACT

Spontaneous hydrolytic deamination of 5-methylcytosine leads to T:G mismatches in double-stranded DNA and comprises a major threat for the integrity of both the DNA primary sequence as well as the epigenetic information stored in the DNA methylation pattern. Failure of the cellular DNA repair machinery to recognize and repair such mismatched nucleotides can lead to a mutator phenotype and subsequent carcinogenesis. A thymine-DNA glycosylase (TDG) has been described that initiates T:G mismatch repair by specifically excising the mismatched T. We have studied the TDG genomic locus and the expression of this enzyme to evaluate its role in cancer development. TDG is highly expressed in thymus and is expressed at lower levels in all human tissues analyzed. The TDG gene has 10 exons covering a region of >25 kb and is located on chromosome 12q22-q24.1. Because gastric tumors have been shown to contain a high percentage of C→T mutations at CpG sites, we used a microsatellite found in intron 8 of the TDG locus to screen gastric tumor samples for loss of heterozygosity. Although our analysis showed loss of heterozygosity in 10 of 24 samples (42%), none of those tumor samples revealed a mutation in the coding sequence of the remaining TDG allele as analyzed by single-strand conformational polymorphism. Expression of the TDG was not determined because of the limited availability of RNA in these primary tumor samples. At present, we have found no evidence that TDG is central to the development of gastric cancer, limiting the importance of TDG in T:G mismatch repair and subsequent carcinogenesis.

INTRODUCTION

Human tumors are characterized by a high proportion of C→T transition mutations at CpG sites (1). For example, five of six mutational hot spots in the p53 tumor suppressor gene are CpG sites, and 24% of all mutations found in this gene in human tumors are C→T transitions. Up to 47% of the CpG sites in the p53 gene are mutated to TpG or CpA, respectively, in some tumor types such as colorectal and gastrointestinal tumors (2). This high percentage is even more remarkable because CpG sites are underrepresented in the human genome by a factor of five (3-5). It is believed that all of the CpG mutational hot spots in the p53 gene are caused by spontaneous hydrolytic deamination of 5mC to thymine (6-8), because they seem to be completely methylated (9). The deamination process creates a T:G mismatch that results in a C→T transition mutation if it remains unrepaired during the cell cycle.

Spontaneous hydrolytic deamination of 5mC has been studied in great detail. In double-stranded DNA, 5mC decays to T at a rate of 5.8 x 10^-13/s at 37°C (10, 11). That is approximately 2.5 times the rate of C deaminating to uracil (U) under the same conditions. Assuming 3.8 x 10^9 CpG sites in the human genome, two T:G mismatches are created/cell/day independently of DNA replication. This number of deamination reactions is more than sufficient to account for all mutations at CpG sites found in human tumors (11). How many of these mismatches result in C→T mutations depends on the efficiency of the cellular mismatch repair machinery. Uracil is not a normal Watson-Crick base and is easily recognized and removed by ubiquitous cellular UDGs. Escherichia coli with nonfunctional UDG show an approximately 30-fold higher mutation rate (12). T:G mismatch repair has been shown to be initiated in bacteria, which methylate the 3' C to 5mC at dcm sites (CCWGG), by the very short patch repair (vsr) gene product. Inactivation of this repair pathway leads to mutational hot spots at methylation sites (13, 14). A similar enzyme has not yet been described in human cells and may not exist. However, T:G mismatches are also repaired very efficiently in mammals (15). The formation of C→T mutations at CpG sites has been calculated to be approximately 2 x 10^-8 mutations/year in the primate genome (16). This rate is more than 1000-fold slower than the deamination rate of 5mC to T in vitro, showing that most of the T:G mismatches created by 5mC deamination are successfully repaired.

A human TDG has been identified (17) that initiates T:G mismatch repair by specifically excising T from those mismatches (18, 19). A glycosylase mechanism is especially attractive as a model for T:G mismatch repair at CpG sites, because it avoids the problem of strand discrimination. Although T:G mismatches in several sequence contexts are repaired, the best substrate for this enzyme is a T:G mismatch within a CpG site in which the remaining C is methylated to 5mC (20-22). Thus, TDG could be a major defense mechanism against spontaneous deamination of 5mC and the formation of transition mutations at CpG sites in human cells. Mismatched nucleotides are also recognized and removed by the long patch repair mismatch repair (23-25). Mutational inactivation of the human mismatch repair genes hMSH2, hMLH1, and hPMS2, which are homologues of the bacterial MutS and MutL mismatch repair proteins, has been shown to result in a mutator phenotype (26, 27) and has been linked to hereditary cancer in various human tissues (28-32).

Here we further characterize the TDG locus and evaluate whether inactivation of this gene may be involved in the formation of human gastric cancers that generally show high frequencies of C→T transitions at CpG sites (2).
HUMAN TDG GENE

Mapping of the TDG Gene by Radiation Hybrid Method. PCR reactions were performed using primers x97 (5'-AACGCTAAGGCTTAGTGG-3') and x98 (5'-ACGAATGTGTAGACAGTCTCT-3'), respectively, to screen the Genebridge-4 radiation hybrid panel (Version RHO2.02; Research Genetics, Inc.; Ref. 34). Thirty-five cycles were used with an annealing temperature of 55°C for 30 s and then 72°C for 1 min. Fragments were visualized by agarose gel electrophoresis and sequenced using primer x84 (5'-TACTCACTGCCGGTTAGCAACC-3').

Table I Primer sequences used for the determination of intron sizes and for SSCP analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>5' primer</th>
<th>3' primer</th>
<th>Intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Determination of intron sizes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x20</td>
<td>GTGGGCGGCGGCGCATCA</td>
<td>x28</td>
<td>GGCGCTGAACCTCTTCTTGG</td>
</tr>
<tr>
<td>x27</td>
<td>GGCGGCGATCTCCCTCTTCA</td>
<td>x31</td>
<td>GGGTCGAGATCGTGCAGAT</td>
</tr>
<tr>
<td>x29</td>
<td>ACCAGAAACACAGAACCAA</td>
<td>x32</td>
<td>GGATCAAAATGATTTCTGAATAAAGCTAC</td>
</tr>
<tr>
<td>x72</td>
<td>GCCGATACCAACCCGGCTAAATGT</td>
<td>x35</td>
<td>ACTGAGCCCTGCAATACAC</td>
</tr>
<tr>
<td>x99</td>
<td>CAGGGAGGACCGCTCTTCTAA</td>
<td>x37</td>
<td>CACATCCCTATTCCCTTCT</td>
</tr>
<tr>
<td>x36</td>
<td>CACGAAATCACGATGTTTAAATG</td>
<td>x38</td>
<td>CTCTTTGGGCAAGCTGTA</td>
</tr>
<tr>
<td>x70</td>
<td>CATCTGTTGATGAGGATGACT</td>
<td>x73</td>
<td>GTTTTTTCCTCGTAAGACACC</td>
</tr>
<tr>
<td>x74</td>
<td>GTATCCGTGGTTAGCCGACCA</td>
<td>x26</td>
<td>CAACACTTTAGTCTCTACCGTCTT</td>
</tr>
</tbody>
</table>

B. SSCP analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>5' primer</th>
<th>3' primer</th>
<th>Exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>x78</td>
<td>GGGTGCTCGGCAAGCTGATGAGTA</td>
<td>x154</td>
<td>AGGGCTGCGCAGAGCTGAGGG</td>
</tr>
<tr>
<td>x92</td>
<td>GGCTAATTTTTACATTACATCTTGGTCTTCT</td>
<td>x93</td>
<td>AGAGGTGGTGACATGATGCTGAATTT</td>
</tr>
<tr>
<td>x29</td>
<td>CTTGGGAGAATCTGAATTCTCTAGT</td>
<td>x31</td>
<td>GGGTTCGAGATCCTGGGACAG</td>
</tr>
<tr>
<td>x97</td>
<td>ACCGCAAATGCGCTCTTCTACTG</td>
<td>x98</td>
<td>GAGGATGTTGATTTCTTCT</td>
</tr>
<tr>
<td>x99</td>
<td>CAGACGGACGGCTCTTCTAA</td>
<td>x100</td>
<td>GCATCTCAGTTGTTCAACCCCTAAG</td>
</tr>
<tr>
<td>x102</td>
<td>ACCGAGCTCAACAAATAGT</td>
<td>x103</td>
<td>TCTACGAGATCGTGGAGT</td>
</tr>
<tr>
<td>x101</td>
<td>TGCTTCTCAAGATTTCAGCTT</td>
<td>x104</td>
<td>AATTTAAAGGGGAGAATGAGCAGT</td>
</tr>
<tr>
<td>x108</td>
<td>GAGCTGAGCAAGAGAGCACAAA</td>
<td>x109</td>
<td>TCTTAAAGGGGAAGGGAGAAGAG</td>
</tr>
<tr>
<td>x106</td>
<td>GTTCTGATTTCTGACATAGACA</td>
<td>x107</td>
<td>TCTGAGGAATACACATCTG</td>
</tr>
<tr>
<td>x144</td>
<td>CGATAGAGTGAACAGCTCGT</td>
<td>x145</td>
<td>TGCAACATTTAGGACAGTCG</td>
</tr>
</tbody>
</table>

Table I Analysis of the TDG cDNA. The RACE anchor primer x65 (5'-GATCCTGGAGGATGCGCTGTACACAT-3') was phosphorylated using T4 polynucleotide kinase (New England Biolabs, Beverly, MA), purified by phenol extraction and ethanol precipitation, and subsequently ligated to single-stranded cDNA as described (33). Semi-nested PCR was performed using x113 (5'-ATTCCCCGACTCGCTACCTCTC-3'), x71 (5'-GTCGAGGGTGATATGAGTACCT-3'), x28 (5'-GGGCGTGAACCTCTTCTTGG-3'), and AmpliTaq DNA polymerase (Perkin-Elmer, Foster City, CA) in a total volume of 10 or 25 μl. Fragments were purified by agarose gel electrophoresis and sequenced using primer x84 (5'-TACTCACTGCCGGTTAGCAACC-3').

Analysis of the TDG Gene Expression in Human Tissues by Northern Analysis. A 1.0-kb fragment was PCR-amplified using primers x27 and x73 (5'-GATTTTTCCTCCGTAAGACACC-3'). The fragment was agarose gel purified and isolated using the agarose gel extraction kit (Boehringer Mannheim). The fragment was agarose gel purified and isolated using the agarose gel extraction kit (Boehringer Mannheim). The fragment was gel electrophoresed, and data were submitted to the Whitehead Institute/MIT Center for Genome Research for final analysis.

Analysis of the TDG Gene by FISH. FISH was performed as described previously (35). Metaphase chromosomes from normal peripheral blood lymphocytes were prepared according to standard techniques. Slides were pre-treated with RNase A (Boehringer Mannheim, Mannheim, Germany; 100 μg/ml) in 2X SSC (pH 7.0) for 60 min at 37°C, followed by two washes in 2X SSC and sequential dehydration in 70%, 85%, and 100% ethanol. Subsequently, samples on the slides were denatured in 70% formamide and 2X SSC at 70°C for 2 min and dehydrated. DNA from a BAC clone (339F22) was labeled by nick translation with biotin-14 dATP using the BioNick labeling kit (Life Technologies, Inc.), coprecipitated with 10—15 µg of unlabeled human Cot-I DNA (Life Technologies, Inc.), and resuspended in 50 ml of hybridization solution (50% formamide and 10% dextran sulfate in 12.5× saline-sodium phosphate-EDTA). The probe was denatured and preannealed at 37°C for 1 h before adding the mixture to the slides. After overnight hybridization and washes in 50% formamide, 2X SSC at 45°C, and 2X SSC at 37°C, the biotin-labeled probe was detected by immunologically using FITC conjugated to avidin (FITC-avidin DCS; Vector Laboratories). The chromosones were counterstained with 4,6-diamidino-2-phenylindole dihydrochloride. The evaluation was performed on a Leitz Wetzlar fluorescence microscope. Images were taken using a cooled charge-coupled device camera (Princeton Instruments, Trenton, NJ). Image processing was done with IPLab Spectrum software (Signal Analytics Corp., Vienna, VA).

Mapping of the TDG Gene. A 1.0-kb fragment was PCR-amplified using primers x27 and x73 (5'-GATTTTTCCTCCGTAAGACACC-3'). The fragment was agarose gel purified and isolated using the agarose gel extraction kit (Boehringer Mannheim). Subsequently, the fragment was 5'-labeled by incubation with 50 μCi of [γ-32P]ATP (3000 μCi/mmol; DuPont, Boston, MA), 1 μl of T4 polynucleotide kinase (New England Biolabs), 70 mm Tris-HCl, 10 mm MgCl2, and 5 mm DTT at 37°C for 1 h in a total volume of 10 μl; purified and used to probe a multiple-tissue Northern blot (obtained from Clontech Laboratories, Inc., Palo Alto, CA). β-actin was used as a control. Signals were scanned and analyzed using a PhosphorImager system (Molecular Dynamics, Sunnyvale, CA). TDG mRNA levels were normalized to β-actin signals in the same lane and expressed as the ratio of TDG:β-actin.

LOH Analysis. Forty primary gastric carcinomas and corresponding normal stomach mucosa were obtained from surgically resected stomach cancers at Istituto Regina Elena (Rome, Italy; all RO cases) and at the National Cancer Center Research Institute (Tokyo, Japan; all AJ cases). The samples were taken immediately after resection and snap-frozen in liquid nitrogen. The remaining tissue was routinely processed for histopathological analysis.

DNA from both normal mucosa and tumor was purified by standard techniques (36). Primers x75 (5'-ATTATCAGAGCCTCTCTGCT-3') and x76 (5'-GGTATGCTCTGCTTCTGTAC-3'), which flank a microsatellite in intron 8 of the TDG gene, were used in PCR reactions to test for LOH at this locus. PCR reactions were performed using 100 ng of template and 1 unit of Taq polymerase (Takara; PanVera Corp.) with 50 μM each of dATP, dGTP, and dTFP; 5 μM dCTP; and 10 μCi of (α-35P) dCTP in a total volume of 25 μl. After an initial denaturation step at 94°C for 3 min, DNA was amplified through 25 cycles consisting of a 20-s denaturation step at 94°C, a 30-s annealing at 57°C, a 30-s extension at 72°C, and a final extension at 72°C for 5 min. Assessment of LOH was performed as described previously (37).

Analysis of the TDG Gene by SSCP in Human Gastric Tumor Samples. Segments of the TDG gene were amplified from the genomic DNA of 14 cases at the TDG locus, using sets of oligonucleotide primer pairs flanking all coding exons. The sequences of the primers are shown in Table 1. The conditions for the PCR-SSCP reactions were similar to those for LOH analysis. PCR products were electrophoresed on a 0.5× mutation detection enhancement gel (FMC BioProduct, Rockland, ME) at 5 W for 18 h. The annealing temperature and number of amplification cycles were optimized for each primer pair.

RESULTS

Analysis of the TDG Genomic Locus. To obtain a bacterial clone containing the full-length human TDG gene, we screened a human BAC DNA pool (Research Genetics, Inc.) by PCR according to the manufacturer's protocol using primers derived from the TDG cDNA (38). BAC clone 339F22 was grown, and DNA was isolated and sequenced. Intron sizes were determined by long-range PCR using BAC and HeLa cell DNA, respectively. We found that the human TDG gene has 10 exons covering a region of more than 301 kilobases.
than 25 kb (Fig. 1A). However, we have been unable to find a region containing the sequence of the first 248 bp of the published cDNA sequence (38). Moreover, all primers that anneal to that sequence failed to give a PCR product using BAC DNA, human WBC DNA, or HeLa cDNA as PCR template, respectively. Therefore, we analyzed the 5' end of the TDG cDNA by RACE (33) using HeLa cDNA and also using cDNA from human testis. We determined that the TDG cDNA 5' untranslated region is 254 bp shorter than the published sequence (data not shown). We also found a TCA instead of CCA in positions 416-419 of the cDNA sequence that changes the amino acid code from Pro to Ser. Total length of the cDNA is 3.16 kb (Fig. 1A). Intron-exon boundaries are presented in Fig. 1B. Interestingly, two regions at the 5' end of the TDG gene fulfill the criteria of being a CpG island (Ref. 39; Fig. 2); exon 1 is part of an island that extends into the first intron. A second CpG island of approximately 300 bp is located in the putative promoter region. This pattern is typical for housekeeping genes and is similar to that of the human UNG gene that codes for UDG (40).

We probed a multiple-tissue Northern blot with a 1.03-kb fragment and found expression of a 3.2 kb TDG mRNA in all tissues tested (Fig. 3). The highest expression was found in thymus (Lane 2), whereas all other lanes showed approximately the same signal:control ratio.

The human TDG gene was mapped at 5.5 cR telomeric of marker WI-7420 (lod > 3.0) that is located on chromosome 12 at position 12q22-q24.1 using the Genebridge-4 radiation hybrid panel. The mapping to the long arm of chromosome 12 was confirmed by FISH analysis (Fig. 4).

LOH and SSCP Analysis of the TDG Gene. Sequence analysis of the introns in the TDG gene locus identified a (TA)\(n\) and (CA)\(n\) microsatellite in intron 8 that spans a region of approximately 300 bp.

---

**Fig. 1. A.** Organization of the human TDG genomic locus. Underlined intron sizes are sequenced completely. Dark boxes containing numbers 1-10 indicate the individual exons. B. Exon-intron borders of the TDG gene. Numbers in parentheses between the intron sequences are the nucleotide coordinates of the cDNA.
DISCUSSION

DNA mismatch repair is crucial to maintain the integrity of cellular DNA. TDG is a potentially important enzyme that is proposed to counteract spontaneous deamination of 5mC. This idea is supported by the fact that TDG is ubiquitously expressed in all human tissues analyzed (Fig. 3), and the 5' region of the gene is consistent with a typical housekeeping gene structure. Because C→T transition mutations at methylation sites are frequently found in many human tumor types, it is reasonable to hypothesize that TDG inactivation may contribute to the formation of muta-

This microsatellite was used as a marker for LOH analysis of primary gastric carcinomas. Gastric tumors were chosen, because this cancer type shows a high rate of C→T mutations in the p53 gene (2). Moreover, 31% of their gastric tumor samples display LOH in the region of chromosome 12q (41), where TDG was found to be located. We designed primers that flank this repeat and analyzed 40 human gastric tumor samples for LOH. Ten of 24 informative cases (42%) showed LOH at the TDG locus (Table 2; Fig. 5). Fifteen cases were not informative, and one case showed microsatellite instability (42, 43). Ten cases with LOH at this locus were screened for mutations in the remaining TDG allele using the PCR-SSCP technique. No alteration in the migration pattern of the bands in the SSCP gel were found compared to the pattern produced by PCR products from the corresponding normal mucosa. In addition, four cases (NJ17, RO2, RO26, and RO36) without abnormalities at the TDG locus showed similar SSCP patterns. We also screened 20 cases of breast cancer carcinomas and 17 breast cancer cell lines by SSCP analysis and found no identifiable alterations (data not shown).

Fig. 2. Distribution of CpG sites and the percentage of G + C bp between −1000 and +500 of the human TDG gene calculated by the GCG software package (0, transcription start site).

Fig. 3. Expression of TDG in human tissues by Northern analysis. Lanes contain mRNA from different human tissues as indicated or from peripheral blood leukocytes (PBL). A 1.0-kb PCR product amplified from HtLa cDNA was used as a probe.

Fig. 4. Mapping of the human TDG gene by FISH analysis. The signal is confined to chromosome 12q22-q24.
tional hot spots at sites of mammalian DNA methylation and thus contribute to human tumor development. To test this hypothesis, we first characterized the structure of the TDG gene and mapped it to chromosome 12q22-2q41 (Fig. 4). Interestingly, this is a region of high LOH in gastric cancers (Fig. 5; Ref. 41). Although gastric tumors also contain a high percentage of C→T transition mutations, we found no mutations within the coding sequence of the remaining TDG allele in the gastric tumor samples that displayed LOH. We have been unable to test whether the expression of TDG is altered in these primary gastric tumors because of limited availability of RNA. With this caveat in mind, it is unlikely that the mutational inactivation of TDG is a central step in gastric tumorigenesis, and another gene in this region on chromosome 12q may contribute to human tumor development. To test this hypothesis, we used a T:G base excision repair activity assay (44). Although colorectal tumors show the highest frequency of C→T transitions at CpG sites in the p53 gene of all human tissues analyzed (2), extracts from colon carcinoma tissue and extracts from adjacent mucosa showed approximately the same T:G base excision repair activity. We cannot exclude reduced TDG expression, due to the loss of one allele of the TDG gene, as a contributor to a lower T:G repair activity and thus to a mutator phenotype in gastric cancer. It is also possible that a mutation in the TDG promoter region or DNA hypermethylation of the CpG islands silences TDG gene expression, similar to the reduced expression of hMLH1 in sporadic colon tumors that was correlated with DNA methylation of the promoter region (45).

It is not clear which enzyme system carries the main burden of repairing T:G mismatches created by spontaneous hydrolytic deamination of 5mC. Inactivation of TDG may not lead to a mutator phenotype, because several redundant T:G repair mechanisms exist in the cell. hMSH2, hMLH1, and hPMS2, which are central to the long patch mismatch repair process, are expressed throughout the cell cycle (46), suggesting that mismatch repair is not exclusively active during S-phase of the cell cycle. In addition, TDG also excises U from U:G mismatches, and it has not been excluded that the main function of this enzyme is to backup UDG in its repair of U:G mismatches (19). A family of UDGs with high homology to the central part of the TDG cDNA has been described (47). This homology codes for its U recognition ability, suggesting that the human TDG evolved from bacterial UDGs. Thus, TDG may play only a supporting role in the repair of T:G mismatches, and other TDGs (and UDGs) perform redundant functions. However, it should be noted that the cell has to discriminate between T:G mismatches formed by 5mC decay and the same mispairs created by misincorporation during DNA synthesis. Excision of the wrong base would create a transition mutation in either case. As discussed before, a glycosylase mechanism to counteract spontaneous deamination is especially attractive, because it would avoid these problems of strand discrimination. The finding that TDG is expressed in all human tissues examined (Fig. 4) with highest expression in thymus may further support a significant role of TDG, because thymus DNA also contains the highest percentage of 5mC (48). Interestingly, a TDG has also been found in a thermophilic archaeon Methanobacterium thermophilum, which has an optimal growth temperature of 72°C and therefore has higher rates of hydrolytic 5mC deamination (49). However, this enzyme shows no sequence homology with the human TDG. Perhaps there is a human homologue of this M. thermophilum TDG that plays a larger role in T:G discrimination in the context of a CpG site. Additional studies will clarify this issue.
ACKNOWLEDGMENTS

We thank Bernadette Mandes for technical assistance and Eugenio Santoro and Setsuo Hiroth for the tumor samples. We are also grateful to Hansjürg Alder and the Kimmel Nucleic Acids Facility for oligonucleotide synthesis, sequencing, and technical assistance and to Florencia Bullrich for helpful discussions.

REFERENCES


Human Thymine-DNA Glycosylase Maps at Chromosome 12q22-q24.1: A Region of High Loss of Heterozygosity in Gastric Cancer

Christoph Schmutte, Raffaele Baffa, Luisa M. Veronese, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/57/14/3010

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/57/14/3010.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.