Methylated CpG Dinucleotides Are the Preferential Targets for G-to-T Transversion Mutations Induced by Benzo[a]pyrene Diol Epoxide in Mammalian Cells: Similarities with the p53 Mutation Spectrum in Smoking-associated Lung Cancers

Jung-Hoon Yoon, Leslie E. Smith, Zaohui Feng, Moon-shong Tang, Chong-Soon Lee, and Gerd P. Pfeifer

Department of Biology, Beckman Research Institute of the City of Hope, Duarte, California 91010 [J.-H. Y., L. E. S., G. P. P.]; Department of Biochemistry, College of Natural Sciences, Yeungnam University, Kyongsan, 712-749 Korea [J.-H. Y., C.-S. L.]; and New York University, Department of Environmental Medicine, Tuxedo, New York 10987 [Z. F., M.-S. T.]

ABSTRACT

A large fraction of the p53 mutations in lung cancers from smokers are G-to-T transversions, a type of mutation that is infrequent in lung cancers from nonsmokers and in most other tumors. Previous studies have indicated that there is an association between G-to-T transversion hotspots in lung cancers and sites of preferential formation of polycyclic aromatic hydrocarbon adducts along the p53 gene. p53 codons containing methylated CpG sequences are preferential targets for formation of adducts by (±) anti-7,8,9,10-tetrahydrobenzo[a]pyrene (BPDE). To assess the role of CpG methylation in induction of mutations by BPDE, we analyzed BPDE mutagenesis in three CpG methylated target genes: a supF shuttle vector and the cII and lacI transgenes in embryonic mouse fibroblasts. After methylation of the shuttle vector at all CpG sequences, 42% of all G-to-T transversions were at CpG sites compared with 23% in unmethylated DNA. In the cII transgene, which is methylated at CpG sequences in vivo, 83 of 147 (56%) of the BPDE-induced mutations were G-to-T transversions, and 58% (48 of 83) of all G-to-T transversions occurred at methylated CpG sequences. In the lacI gene, 68% (75 of 111) of the BPDE-induced mutations were G-to-T events, and 58% of 77 (77%) of these occurred at methylated CpG sequences. The occurrence of transversion hotspots at methylated CpGs correlated with high levels of BPDE adducts formed at such sites. This situation mirrors the one in the p53 gene in lung cancers from smokers where 236 of 465 (51%) of the G-to-T transversions occurred at methylated CpG sites. These findings further strengthen a link between polycyclic aromatic hydrocarbons present in cigarette smoke and lung cancer mutations and provide evidence that mutational processes other than C-to-T transition mutations can occur selectively at methylated CpG sequences.

INTRODUCTION

p53 mutations are common in lung cancers from smokers but are less common in nonsmokers. The p53 mutation spectrum in lung cancer is different from mutational spectra in other cancers. An excess of G-to-T transversions is characteristic for these tumors (1–4). We recently reanalyzed the p53 mutation database maintained at the IARC and the primary literature on p53 mutations in lung cancers from smokers and nonsmokers (4). We came to the conclusion that p53 mutation spectra are very different between smokers and nonsmokers and that this difference is statistically highly significant (G-to-T transversions are 30% versus 10%; P < 0.0001, χ² test). A similar difference is seen between lung cancers and nonlung cancers. At a number of mutational hotspots common to all cancers, a large fraction of the mutations are G-to-T transversions in lung cancers but are almost exclusively G-to-A transitions in nonlung tumors (4). This analysis has reinforced the notion that a considerable fraction of p53 mutations in lung cancers may be attributable to direct DNA damage from cigarette smoke carcinogens.

G-to-T transversions are thought of as a “molecular signature” of tobacco smoke mutagens in smoking-associated lung cancers. Among the many carcinogens present in cigarette smoke, PAHs have been implicated strongly in lung cancer. The argumentation is the following: (a) PAHs are prominent carcinogens present in cigarette smoke (5). (b) PAH adducts are found in DNA extracted from human tissues exposed to tobacco smoke (6–10). (c) PAHs can produce predominantly G-to-T mutations in selectable marker genes or in tumor-associated genes (11–17). (d) There is an increased frequency of G-to-T transversions in the p53 gene of lung cancers from smokers compared with lung cancers from nonsmokers and compared with most other cancers (2, 4, 18, 19). (e) More than 85% of the G-to-T transversions can be attributed to guanines on the nontranscribed DNA strand (4). A preferential repair of bulky DNA adducts on the transcribed strand may explain this strand difference of transversion mutations (20–22). (f) Finally, there is a precise correspondence between G-to-T mutational hotspots in the p53 gene and hotspots of adduct formation by PAHs found in tobacco smoke (23–26).

It is of interest that most of the G-to-T transversion hotspots in lung cancers occur at guanines that are part of the dinucleotide sequence 5′-CpG-3′ (CpG). CpG dinucleotides are sequences targeted by DNA methyltransferases in mammalian cells, which methylate the cytosine ring at position 5 (27, 28). In the coding sequence of p53, all CpG dinucleotides between exons 5 and 9 are highly methylated in all tissues analyzed (29, 30). In earlier studies (24, 25), we showed that methylation of CpG sequences determines the preferential formation of DNA adducts of the PAH compound BPDE. The authors of studies with methylated and unmethylated oligonucleotides arrived at a similar conclusion (31).

The formation of DNA adducts at specific sequences is the first step in the mutagenic process. However, a mutational spectrum is shaped by additional factors, such as DNA repair, efficiency and specificity of lesion bypass by DNA polymerases, and selection. It is of note that the types of mutations induced by BPDE have been shown to depend on the particular DNA sequence context in which the lesion is embedded. For example, BPDE induces predominantly G-to-A transitions in the sequence 5′-CGT-3′ (32). Thus, it is conceivable that BPDE adducts at methylated CpG sequences may promote transition mutations rather than transversions. If true, such a finding would question any putative links that have been made between PAH adduct
formulation and lung cancer mutations and would offer alternative explanations for CpG transition mutations in general.

In this study, we investigated the mutational specificity of BPDE, using three mutational target genes that contain methylated CpG sequences.

MATERIALS AND METHODS

Cell Culture and Treatment with BPDE. Embryonic mouse fibroblasts were derived from 13.5-day-old BigBlue mouse embryos, kindly provided by K. Hill, City of Hope (Duarte, CA). Four embryos were minced into small pieces and digested with collagenase overnight. After centrifugation, the pellet (exclusive of cartilage and bones) was resuspended, and fibroblasts were cultured in DMEM containing 10% fetal bovine serum. Racemic BPDE was obtained from the NCI repository (Midwest Research Institute, Kansas City, MO). For the mutagenesis experiments, fibroblasts were grown as monolayers at 20–30% confluence. BPDE was added to the cells to a final concentration of 0.2 μM, and cells were incubated at 37°C for 30 min. The medium was removed, the cells were washed with PBS, and regular growth medium was returned to the cells. They were grown for 5 days to allow mutation fixation. DNA was isolated by standard phenol-chloroform extraction and ethanol precipitation. This DNA was used in the cl II and lac I mutation assays as described below.

**sup F Mutation Assay.** The pSP189 plasmid, which contains the sup F gene as a mutational target, was kindly provided by M. Seidman (NIA, NIH, Baltimore, MD). The plasmid was methylated in vitro using the CpG-specific DNA methylase SsI1 according to the manufacturer’s instructions (New England Biolabs, Beverly, MA). Control DNA was mock-methylated in the absence of S-adenosylmethionine. Completion of methylation was confirmed by incubating an aliquot of the reaction mixture with the methylation-sensitive restriction endonuclease HpaII. Methylated and unmethylated pSP189 was then treated with 2 μM BPDE for 30 min at room temperature. After removal of BPDE by organic solvent extraction, the DNA was ethanol precipitated, redissolved, and transfected into the XP-A cell line XP12BE, using Lipofectamine. Seventy-two h after transfection, the plasmid was rescued by Hirt extraction (33). The DNA was cleaved with DpnI to remove unreplicated plasmid. The plasmid was then electroporated into MB7070 bacteria, which carry a lacZ gene with an amber mutation. Plasmids were isolated from white bacterial colonies, and the sup F gene was sequenced.

**cl II Mutation Assay.** The lambdaLIZ shuttle vector containing the cl II target gene was rescued from total genomic DNA of embryonic mouse fibroblasts by mixing 0.5 μg/μl DNA aliquots with lambda phage packaging extract (Transpack; Stratagene) as described in the Big Blue manual (Stratagene). The cl II mutation assay was performed with the G1250 h7 Escherichia coli host strain (34). To determine the total titer of packaged phage, 200 μl of the G1250 strain were mixed with 1:100 dilutions of phage, plated on TB1 plates in aliquots of 20 and 100 μl, and incubated overnight at 37°C. For mutant selection, 100 μl of the packaged phage were mixed with 200 μl of the G1250 strain, plated on TB1 plates, and incubated at 24°C for 48 h. After incubation at 24°C, lambda phage bearing nonmutant cl II genes undergoes lyogenic growth, but phage with mutant cl II genes undergo lytic growth and give rise to plaques. When incubated at 37°C, non-cl II mutants also undergo a lytic cycle. The cl II mutant frequency was calculated by dividing the number of mutant plaques by the estimated number of total plaques. For sequencing analysis, mutant plaques were selected at random and replated at low density to verify the mutant phenotype and to isolate plaques. Single well-isolated plaques were picked, placed into 25 μl of Tris-EDTA buffer and boiled for 5 min. A 433-bp segment containing the cl II gene and flanking regions was amplified by PCR with two primers (5’-CCACACCTATGGTGATG-3’ (positions –68 to –50) and 5’-CCCTGTCCGAATTGAGTATG-3’ (positions +345 to +365). The PCR products were purified using PCR purification kits (Qiagen, Chatsworth, CA). The PCR products (100 ng) were sequenced with a Big Dye terminator cycle sequencing kit (ABI Prism, PE Applied BioSystems, Foster City, CA) on an ABI DNA sequencer. Each cl II mutant was sequenced in its entirety with PCR primers as mention above. Each mutation was confirmed by sequencing the opposite strand. The resulting DNA sequences were analyzed and compared with that of the wild-type cl II gene with SeqWeb Version 1.2.

**lac I Mutation Assay.** The lambdaLIZ shuttle vector containing the lac I target gene was rescued from total genomic DNA by packaging as described above. Mutations were detected as blue plaques on E. coli K-12 lawns (SCS-8 strain, recA−, MecA−, MecBC−, Mvr−, HsdR−; Stratagene) on 25-cm NZY agar plates containing 1.5 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. The plates were incubated overnight at 37°C, and the mutant blue plaques were counted. The lac I mutant frequency was calculated by dividing the number of mutant blue plaques, excluding sectored and pinpoint plaques, by the estimated number of total clear plaques. Putative mutant plaques were picked and replated at low density in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside to verify the mutant phenotype. Single well-isolated blue plaques were picked, placed into Tris-EDTA buffer, and boiled for 10 min. A 1338-bp segment containing the lac I gene and flanking regions was amplified by PCR with two primers: 5’-GTACCCGATCCACCTGAAGT3’ (positions –55 to –36) and 5’-GAGTCAAGGCGTGTGA-3’ (positions +1270 to +1283) as described previously (35). The PCR products (100 ng) were sequenced on an ABI DNA sequencer. Each lac I mutant was sequenced in its entirety with primer 5’-CCCCGACACCATCGAAATAACGTTTA-3’ (position –53 to –37 relative to the transcription start point) and reverse primer 5’-TGAAGGCGGCGCGCACGTTTTAATC-3’ (position +1191 to +1208). We confirmed each mutation by sequencing the same region on the opposite strand and by using additional primers: 5’-TGTTGAATTGAGTGAATC-3’ (position +347 to +363) and 5’-ATTACCGAGTCCGGCTC-3’ (position +797 to +813).

**Methylation Analysis.** The DNA methylation pattern along the cl II and lac I genes in embryonic mouse fibroblasts was determined by Maxam-Gilbert sequencing in combination with LMPCR (30). We compared the cytosine-specific sequencing ladders from genomic DNA of cells with those from methylated and mock-methylated cl II PCR products. In vitro methylation, Maxam-Gilbert cytosine-specific chemical modification reactions, DNA cleavage, and LMPCR were carried out as described previously, using specific primers for lac I (36) and cl II (37). For quantitation of methylation levels at individual CpG sites, the band signals were measured with a phosphorimager. Signals from the methylated and mock-methylated cl II PCR products were defined as 100 and 0% methylation, respectively, and loading differences between lanes were corrected by use of cytosine signals at adjacent non-CpG sequences.

**Mapping of BPDE Adducts.** BPDE adducts in the cl II gene were mapped after exposure of embryonic fibroblasts to 5 μM BPDE for 30 min. The DNA was purified and cleaved at BPDE adducts, using the UvrABC nucleases complex of E. coli according to published procedures (23). Under these conditions, cleavage by UvrABC is quantitative (23, 38). Sequences of both strands of the cl II gene were amplified by LMPCR as described previously (37).

RESULTS

**BPDE Mutagenesis in a Methylated sup F Shuttle Vector.** To determine the mutational specificity of benzo[a]pyrene diol epoxide in a CpG-methylated target gene, we used three assay systems. The first mutation assay used the shuttle vector pSP189 (39), which contains the sup F gene as a mutational target. The vector was methylated in vitro at all CpG sequences, using the CpG-specific DNA methyltransferase SsI1. This methylated shuttle vector was treated with BPDE and was then transfected into the nucleotide excision repair-deficient XP-A cell line XP12BE. In parallel experiments, pSP189 was mock-methylated in the absence of S-adenosylmethionine, treated with BPDE, and transfected. After 72 h, the plasmid was rescued, and the DNA was cleaved with DpnI to remove unreplicated plasmids. Using HpaII and HhaI digestion, we found that the level of CpG methylation was largely (>80%) maintained in DpnI-treated DNA 72 h after transfection, indicating that there is no active removal of methyl groups from the methylated plasmids and that the methylation pattern is at least partially conserved during DNA replication (data not shown). Conversely, the unmethylated plasmid did not undergo de novo methylation. The rescued plasmids were then electroporated into MB7070 bacteria, which carry a lacZ gene with an amber mutation. Plasmids were isolated from white colonies, and the
produce a selectable phenotype (40). Interestingly, benzo[a]HPRT relatively high frequency of CpG sequences. This contrasts with the in the genome of BigBlue mice because both of these genes contain a two mammalian transgenes. We chose the cII situation, we then studied the mutational specificity of BPDE in the endogenous HPRT gene, for example, which contains only four CpGs that can |

CpG-methylated pSP189

Supplementary Fig. 1. Methylation analysis of the cII transgene in embryonic mouse fibroblasts. The gel shows an analysis of the upper strand of the cII gene from nt +70 to +215. Lanes 1, 2, and 3, Maxam-Gilbert sequencing reactions (G, C, A, and T, respectively); Lane 4, C-specific sequencing reaction of genomic DNA isolated from embryonic mouse fibroblasts; Lane 5, C-specific sequencing reaction of cII PCR products methylated with the CpG-specific methylase SssI (+CH3); Lane 6, mock-methylated (-CH3) cII PCR products. The arrows indicate the positions of cytosines at methylated CpG sequences.

Table 1 Mutations induced by BPDE in CpG-methylated target genes and comparison with G-to-T transversions in lung cancer

<table>
<thead>
<tr>
<th>supF control</th>
<th>supF + BPDE</th>
<th>p53 lung cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unmethylated</td>
<td>Methylated</td>
</tr>
<tr>
<td>G to T</td>
<td>5 (71)^b</td>
<td>5 (55)</td>
</tr>
<tr>
<td>G to T at CpG</td>
<td>1 (14)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>A to G</td>
<td>0 (0)</td>
<td>2 (22)</td>
</tr>
<tr>
<td>G to C</td>
<td>1 (14)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>A to T</td>
<td>1 (14)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>A to T</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Del/Ins</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Tandem</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>7 (100)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>% of all mutations at CpG</td>
<td>1/7 (14%)</td>
<td>1/9 (11%)</td>
</tr>
<tr>
<td>% of all G-to-T mutations at CpG</td>
<td>1/5 (20%)</td>
<td>1/5 (20%)</td>
</tr>
</tbody>
</table>

^a Lung cancer mutations in p53 coding sequence, exclusive of nonsmokers and occupational exposures (4).

^b Percentage of all mutations in parentheses.

supF gene was sequenced (Fig. 1 and Table 1). The mutant frequencies in unmethylated and methylated untreated supF vectors were not substantially different, and 11–14% of the mutations in both plasmids were G-to-T transversions at CpG sites. There was an increase in mutant frequencies between untreated and BPDE-treated shuttle vectors from 1.3 × 10^{-4} to 41.2 × 10^{-4} (32-fold) for the unmethylated DNA and from 1.8 × 10^{-4} to 48.1 × 10^{-4} (27-fold) for the methylated DNA. Thus, background mutations in the treated samples should have only a very minor contribution. Fig. 1 shows the mutational spectra obtained with the unmethylated and methylated shuttle vectors after BPDE treatment. G-to-T transversions were the predominant type of mutation induced into both plasmids (65–68%). However, there was a clear difference in the distribution of mutations along the supF sequence. In the methylated plasmid, 42% of all mutations were at CpGs, whereas only 25% were at these sites in the unmethylated vector (P = 0.05, χ^2 test). After methylation, 42% of all G-to-T transversions were at CpG sites compared with 23% in unmethylated DNA (Table 1). Thus, there was a fairly substantial association of CpG site transversions and methylation.

BPDE Mutagenesis in the cII Transgene. Because transfection experiments using in vitro-methylated DNA may represent an artificial situation, we then studied the mutational specificity of BPDE in two mammalian transgenes. We chose the cII and lacI genes present in the genome of BigBlue mice because both of these genes contain a relatively high frequency of CpG sequences. This contrasts with the HPRT gene, for example, which contains only four CpGs that can produce a selectable phenotype (40). Interestingly, benzo[a]pyrene produces a much higher mutation frequency in the lacI transgene than in the endogenous HPRT locus (41).

We first determined the methylation status of the cII transgene in BigBlue mouse embryonic fibroblasts. The methylation level of individual CpG sites was analyzed by genomic sequencing and LMPCR (30). The results of a methylation pattern analysis along the sequences...
MUTATIONS AT METHYLATED CpG SEQUENCES

**A**

CATATGGTTCCGTGCAAACAAGAATCACCAGACGCGCTCACTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT

**B**

CATATGGGTCCGTGCAAACAAGAATCACCAGACGCGCTCACTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT

Fig. 3. Mutational spectra of the cII transgene for spontaneous mutations (A) and BPDE-induced mutations (B) in embryonic mouse fibroblasts. The sequence position and type of mutation are indicated for all mutations identified by sequencing of the cII transgene. The altered base is shown for mutations found in mock-treated cells (A) and BPDE-induced mutations (B). Single-base deletions are indicated by an X. Single-base additions are indicated by +N. CpG sites are underlined. Nucleotide positions are numbered, and mutational hotspot codons are indicated by a bracket. The positions of mutational hotspots at CpG sites are outlined.

between nucleotides +70 and +215 of the cII gene are shown in Fig. 2. Methylated cytosines appear as a gap in the cytosine sequencing ladders at 5-methylcytosine positions. By phosphorimager analysis we found that all cytosine residues in CpG dinucleotides were >95% methylated, as determined by comparison between the cytosine-specific sequence ladders from mouse fibroblast DNA with those from methylated and mock-methylated cII PCR products. Thus, the cII transgene is highly methylated at CpG sites in mouse fibroblasts and can be used as a target for BPDE mutagenesis experiments. Similarly, the methylation pattern of the cII gene was determined, and the CpG sites were found to be highly methylated (35, 36; data not shown).

Embryonic mouse fibroblasts were treated with 0.2 μM BPDE. To allow mutation fixation, the cells were incubated for 5 days after BPDE treatment. DNA was isolated, packaged into lambda phage particles, and used to determine the mutant frequency and mutational spectra in the cII transgene. The mutant frequency of the untreated cells was 4.83 × 10⁻⁵, and the mutant frequency after BPDE treatment was 34.2 × 10⁻⁵, an increase of 7.1-fold. A mutational spectrum was determined from untreated fibroblasts and from cells treated with BPDE (Fig. 3). In untreated cells, a mutational hotspot was observed as a transition at a CpG site at position +34 downstream of the ATG codon. This “jackpot” mutation was present in 38% of the sequenced mutants. It occurred in all packaging reactions. This mutation may be considered siblings or potentially clonally expanded mutants, present from the early stages of embryo development or cell propagation. In untreated cells, 62% of all mutations were C-to-T transitions, and the majority occurred at CpG sites. This specificity is commonly observed in tissues of untreated cII or lacI transgenic mice and is considered a hallmark of spontaneous mutation (34, 42–46).

We sequenced 147 cII mutants from BPDE-treated cells. The cII mutational spectrum after BPDE treatment was entirely different from the spontaneous one. The frequency of G-to-T transversions was increased from 9% in the control to 56% in the BPDE-treated cells. This is consistent with the known mutational specificity of BPDE (11–17). Most strikingly, however, the majority of these G-to-T transversions (58%) occurred at guanines in methylated CpG sequences (Table 1). This rise in G-to-T mutations was paralleled by the occurrence of several mutational hotspots at positions +35, +103, +193, +196, and +212 (Fig. 3B). It is extremely unlikely that a single mutational event was scored multiple times in this assay system because only 10% of the total genomic DNA obtained after BPDE treatment was packaged into phage. In addition, packaging usually captures ~0.1% of the lambda copies in the DNA (47), and the BPDE-treated cells divided only a maximum of three to four times. Of course, not all CpG sites were mutated, presumably because of phenotypic selection. Only 3 of 68 mutations in untreated cells were G-to-T transversions at CpG sites. After BPDE treatment this number was 48 of 147. The difference was statistically highly significant (P < 0.001, χ² test). The occurrence of G-to-T transversion hotspots in the cII gene was remarkably similar to that observed in the CpG-methylated supF gene (Fig. 1). The percentage of G-to-T transversions at methylated CpG sites was 58% (48 of 83). There are 17 CpG sequences in 240 bp of the cII gene, i.e., 14% of the total sequence carries >50% of the BPDE-induced transversion mutations.

BPDE Mutagenesis in the lacI Transgene. To further substantiate this observation, we analyzed BPDE-induced mutations in the CpG-methylated lacI transgene (Fig. 4). The mutant frequency of lacI was increased from 6.08 × 10⁻⁵ in untreated fibroblasts to 102 × 10⁻⁵ in BPDE-treated cells (a 16.8-fold increase). We sequenced 111 lacI mutants. lacI mutations in untreated cells were not sequenced, but because there was a 16.8-fold increase in mutant
frequency after BPDE treatment, we expect that >94% of all mutations in the treated samples were in fact derived from the mutagenic action of BPDE. Spontaneous mutation spectra in the lacI transgene are usually dominated by C-to-T transitions, many of them occurring at CpG sequences (42, 44–46). However, the BPDE-induced mutation spectrum in lacI showed an abundance of G-to-T transversions (68% of all mutations; Table 1). Strikingly, 77% (58 of 75) of all G-to-T transversions were localized to methylated CpG sequences, scored as either G-to-T or C-to-A events (Fig. 4). Of the 36 non-G-to-T mutations, only 12 (33%) were at CpG sites. This difference was statistically highly significant (P < 0.001, χ² test). The results obtained with the lacI gene were entirely consistent with those obtained with supF and cII. The data point to an important role of methylated CpG sequences in BPDE-induced mutagenesis.

Mapping of BPDE Adducts. The occurrence of G-to-T transversion hotspots may be related to selective formation of BPDE adducts at the guanines of methylated CpG sequences. To test this hypothesis, we treated embryonic fibroblasts with BPDE, cleaved the DNA isolated from these cells with UvrABC, and amplified sequences of the cII gene by LMPCR (Fig. 5). Because of the sensitivity limit with LMPCR and difficulties in amplifying this sequence, a higher dose of BPDE (5 µM) had to be used than that in the mutagenesis experiments (0.2 µM). However, in previous experiments we did not see any significant change of adduct patterns over a dose range of 0.2–4 µM (24). On the lower strand (Fig. 5A), we observed two prominent BPDE adduct hotspots, one at position +192 (the sequence of the lower strand is 5'-TCGTC; the modified guanine is underlined) and one at position +212 (the sequence of the lower strand is 5'-GCGCC; the modified guanine is underlined). If a G-to-T transversion had occurred at position +192, the change would be GAC to GAA (Asp→Glu), a conservative substitution. The adduct hotspot at position +212 correlates with one of the most prominent mutation hotspots (Fig. 3B). The most common change at position +212 after BPDE treatment is GCG to GAG (Ala→Glu). Similarly, on the upper strand (Fig. 5B), we observed a strong adduct hotspot at position +103 (sequence 5'-GCCTT), corresponding to a mutational hotspot (Fig. 3B). Other significantly damaged sites were at position +90, where a G-to-T transversion would produce a silent substitution (GCG→GCT, Ala→Ala), and at position +114 (TCG→TCT, Ser→Ser).

DISCUSSION

PAHs, such as benzo[a]pyrene, have been suspected as etiological agents in mutagenic processes initiated by cigarette smoke in lung tissue (5). These compounds, in particular those that produce bay region diol epoxides upon metabolic activation, are known to produce G-to-T transversions at high frequencies when tested in standard mutagenesis assays. This has been established in test systems such as the his alleles of Salmonella (48), the lacI gene in E. coli (11), and the HPRT gene in mammalian cells (13, 15, 16). However, because these selectable markers do not contain or contain only very few methylated CpG sequences, studies on the sequence specificity of mutagenesis by PAHs have overlooked the role played by methylated CpG sequences. Using three different target genes that contain CpG methylation, we...
showed here that these sequences are in fact strongly targeted by BPDE mutagenesis.

Studies of BPDE mutagenesis in the HPRT gene have revealed a dose dependence in the profile of mutations induced (15, 16). The proportion of G-to-T transversions decreased from 69% at a high dose (300–480 nm) to 52% at an intermediate dose (40–100 nm) to 42% at a low dose (10–20 nm) of BPDE. At the same time the proportion of mutations at AT base pairs increased from 6% to 24% to 36% at the lowest dose (16). Our study, conducted at a dose of 200 nm BPDE, resulted in 56% and 68% G-to-T transversions in the cII and lacI genes, respectively. The proportion of mutations at AT base pairs was 5–10%. These results are consistent with the studies of Wei et al. (16).

However, it should be noted that the transgenes are more GC-rich than the HPRT gene. This may change the mutation profile toward a lesser involvement of mutations at AT base pairs. In addition, the transgenes contain a much higher frequency of CpG dinucleotides. There are only four CpGs in the HPRT coding sequence that can give a selectable phenotype after mutation (40). Mutations at two CpG sites were reported by Wei et al. (16) in both the high and lower dose experiments, i.e., BPDE-induced mutations at CpGs were not unique to a particular dose of the carcinogen.

It is worth considering the role of selection in shaping the mutational spectra in the three reporter genes we have analyzed. In the supF gene, selection is minimal because a change at almost every base position in this tRNA gene leads to a detectable phenotype (39). For cII and lacI, a considerable number of events cannot be scored, including many changes at the third position of a codon. Thus, some primary mutational changes occurring in cells will go undetected.

Although there is a very large number of missense mutations in lacI and cII that have been reported in the literature (35, 37, 42–46), and thus should have a detectable phenotype, we cannot exclude the possibility that some mutants are preferentially selected. An example would be the +34 jackpot mutation in the cII gene. A transition mutation at +34 changes the codon from CGA (Arg) to TGA (Stop), thus completely diminishing cII function and probably resulting in a strong selection advantage. However, a very large number of mutational changes can be selected in both the lacI and cII genes (for examples, see Refs. 35, 37, 42–46).

The specific base changes introduced by DNA polymerases into DNA at the sites of a benzo[a]pyrene adduct are apparently dependent on the particular sequence context of the lesion. For example, it has been shown that the same BPDE adduct can produce a G-to-T transversion mutation in the sequence context 5'-TGC-3', but induces predominantly G-to-A transitions in the sequence 5'-CGT-3' (32). Thus, it is conceivable that BPDE mutagenesis at methylated CpG sequences produces G-to-A transitions. We show here that this is not the case. At 12 of the 13 CpG-specific mutational hotspots observed in all three target genes combined, the vast majority of the mutations were G-to-T transversions. The only exception was one hotspot at position +196 in the cII gene, which showed G-to-A and G-to-C mutations in addition to G to T (Fig. 3B).

The occurrence of transversion hotspots at methylated CpGs correlated with high levels of BPDE adducts formed at such sites (Fig. 5). Hydrophobic effects (49) or increased molecular polarizability and base stacking (50) derived from the methyl group of 5-methylcytosine may facilitate the creation of an intercalation site for BPDE at methylated CpG sequences. The increase in BPDE intercalative binding to methylated CpGs may eventually be reflected in the extent of covalent interactions. In addition, transmission of the electron-donating effect of the methyl group of 5-methylcytosine through the base pairs may increase the nucleophilicity of the 2-amino group of guanine (51). The precise molecular mechanism of methylated CpG targeting by BPDE is a subject of further studies. Because other PAHs similarly form preferentially at methylated CpG sequences (26), it is likely that their mutational specificity is similar to that of BPDE, although this needs to be tested. Nucleotide excision repair apparently does not contribute greatly to the occurrence of mutational hotspots at methylated CpG sites because an XP-A cell line was used in the supF experiments (Fig. 1).

In the p53 gene of human lung tumors linked to tobacco smoking, five of six G-to-T mutational hotspots occurred at methylated CpG

![Fig. 5. Formation of BPDE adducts in the cII gene. Mouse embryonic fibroblasts were treated with BPDE, and DNA was isolated and cleaved at the sites of adducts with UvrABC nuclease. The sequences of both strands of the cII gene were amplified by LMPCR. A, lower strand; B, upper strand. Positions of preferential adduct formation at methylated CpG sequences are indicated by arrows. Note that the cleavage by UvrABC occurs four nucleotides 3' to a BPDE adduct.](image)

![Fig. 6. p53 mutational spectrum of G-to-T transversions in lung cancer. The data are for lung cancers minus nonsmokers and minus occupationally exposed individuals as specified in Ref. 4 (n = 412). *, hotspot codons of G-to-T transversions involving methylated CpG sequences. The mutated guanines are underlined. The spectrum of G-to-T transversions in lung cancers of nonsmokers is not drawn because there were only seven data points (7 of 99 mutations were G to T; 1 G-to-T mutation occurred at codon 148, 1 at codon 158, 1 at 242, and 4 at 249).](image)
sequences (codons 157, 158, 245, 248, and 273; see Fig. 6). Such specificity is not observed with other tumours, in which the frequency of G-to-T transversions is three times lower than in lung cancer (4). In hepatocellular carcinoma, one other type of tumour with a high frequency of G-to-T transversions, the major hotspot is at codon 249, which does not contain a CpG. The base and sequence specificity of mutations induced by BPDE in three CpG methylated target genes is strikingly similar to that observed in the p53 gene of lung cancers (Table 1 and Figs. 3, 4, and 6). The p53 gene is methylated at all CpG sequences in normal human lung tissue (30). Dominant G-to-T transversion hotspots are found at methylated CpG sequences in all four genes, supF, cII, lacI, and p53. These findings suggest that the primary mutagenic action of PAH compounds may be a crucial component of p53 mutagenesis in lung tumors. In summary, we have added an important piece of evidence that further strengthens the link between mutagenic PAHs present in cigarette smoke and the specificity of G-to-T transversion mutations found in the p53 gene of lung cancers from cigarette smokers.

Mutations at CpG sequences are usually recovered as transition mutations, either C to T or G to A. This is the most common single-base substitution mutation found in a variety of genes linked to human genetic diseases and cancer (52, 53). These transition mutations are usually ascribed to deamination of 5-methylcytosines at the RNA/DNA complementary strand of the target DNA in RNA/DNA adducts along the nontranscribed strand of the human HPRT gene. Mutations at CpG in the p53 gene are the predominant mutational signature of DNA damage by tobacco smoke. Carcinogenesis (Lond.), 22: 367–374, 2001.


Mutations at methylated CpG sequences


Methylated CpG Dinucleotides Are the Preferential Targets for G-to-T Transversion Mutations Induced by Benzo[a]pyrene Diol Epoxide in Mammalian Cells: Similarities with the p53 Mutation Spectrum in Smoking-associated Lung Cancers

Jung-Hoon Yoon, Leslie E. Smith, Zaohui Feng, et al.

*Cancer Res* 2001;61:7110-7117.

Access the most recent version of this article at: [http://cancerres.aacrjournals.org/content/61/19/7110](http://cancerres.aacrjournals.org/content/61/19/7110)

This article cites 54 articles, 27 of which you can access for free at: [http://cancerres.aacrjournals.org/content/61/19/7110.full.html#ref-list-1](http://cancerres.aacrjournals.org/content/61/19/7110.full.html#ref-list-1)

This article has been cited by 34 HighWire-hosted articles. Access the articles at: [http://cancerres.aacrjournals.org/content/61/19/7110.full.html#related-urls](http://cancerres.aacrjournals.org/content/61/19/7110.full.html#related-urls)

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

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

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.