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 sites, 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%; \( P < 0.0001, \chi^2 \) 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 formation and G-to-T transversions.
formation 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 mutual 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 *cII* and *lacI* mutation assays as described below.

**supF Mutation Assay.** The pSP189 plasmid, which contains the *supF* 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 SssI 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 *supF* gene was sequenced.

**cII Mutation Assay.** The lambdaLIZ shuttle vector containing the *cII* 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 *cII* 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 *cII* genes undergoes lyogenic growth, but phage with mutant *cII* genes undergo lytic growth and give rise to plaques. When incubated at 37°C, non-*cII* mutants also undergo a lytic cycle. The *cII* 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 Tris-EDTA buffer, and boiled for 5 min. A 433-bp gene and flanking regions was amplified by PCR with two primers: 5'-GTACCCGA-TGTAAAGCGGCGGT-3' (positions −55 to −36) and 5'-GAGTCACGGTTG-1191 to 1208) and +1191 to +1208). We confirmed each mutation by sequencing the same region on the opposite strand and by using additional primers: 5'-TGTGGAATTGTGAGCGGA-3' (position +347 to +363) and 5'-ATTACCGAGTTGCCGCCT-3' (position +797 to +913).

**Methylation Analysis.** The DNA methylation pattern along the *cII* and *lacI* genes in embryonic mouse fibroblasts was determined by Maxam-Gilbert sequencing in combination with LMPCR (30). We compared the cytosinespecific sequencing ladders from genomic DNA of cells with those from methylated and mock-methylated *cII* PCR products. *In vitro* methylation, Maxam-Gilbert cytosine-specific chemical modifications reactions. DNA cleavage, and LMPCR were carried out as described previously, using specific primers for *lacI* (36) and *cII* (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 *cII* 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 *cII* 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 nuclease 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 *cII* gene were amplified by LMPCR as described previously (37).

**RESULTS**

**BPDE Mutagenesis in a Methylated supF 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 assay used the shuttle vector pSP189 (39), which contains the *supF* gene as a mutational target. The vector was methylated *in vitro* at all CpG sequences, using the CpG-specific DNA methyltransferase SssI. 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

**lacI Mutation Assay.** The lambdaLIZ shuttle vector containing the *lacI* 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−, MarA−, MarC-β, Mvr−, HsdR−; Stratagene) on 25-cm NYagar 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 *lacI* 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 *lacI* gene and flanking regions was amplified by PCR with two primers: 5'-GTACCCGACCATCGAATTGTCGACCTGTTA-3' (positions +55 to −36) and 5'-GAGTCACGGTTG-1191 to 1208) and +1191 to +1208). We confirmed each mutation by sequencing the same region on the opposite strand and by using additional primers: 5'-TGTGGAATTGTGAGCGGA-3' (position +347 to +363) and 5'-ATTACCGAGTTGCCGCCT-3' (position +797 to +913).
produce a selectable phenotype (40). Interestingly, benzo[a]pyrene contains relatively high frequency of CpG sequences. This contrasts with the cII gene in the genome of BigBlue mice because both of these genes contain a single CpG site. We chose the HPRT transgene in the endogenous lacI gene, for example, which contains only four CpGs that can introduce one nucleotide. Tandem mutations are underlined. *, +, and # indicate the occurrence of two mutations in the same plasmid. CpG sequences are underlined.

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></th>
<th>supF control</th>
<th>supF + BPDE</th>
<th>p53 lung cancer&lt;sup&gt;a&lt;/sup&gt;</th>
<th>&lt;sup&gt;b&lt;/sup&gt;</th>
<th>&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unmethylated</td>
<td>Methylated</td>
<td>cII control</td>
<td>cII + BPDE</td>
<td>lacI + BPDE</td>
</tr>
<tr>
<td>G to T</td>
<td>5 (71)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5 (55)</td>
<td>34 (65)</td>
<td>48 (68)</td>
<td>6 (9)</td>
</tr>
<tr>
<td>G to T at CpG</td>
<td>1 (14)</td>
<td>1 (11)</td>
<td>8 (15)</td>
<td>20 (28)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>G to A</td>
<td>0 (0)</td>
<td>2 (22)</td>
<td>4 (8)</td>
<td>9 (13)</td>
<td>42 (62)</td>
</tr>
<tr>
<td>G to C</td>
<td>1 (14)</td>
<td>1 (11)</td>
<td>7 (13)</td>
<td>5 (7)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>A to T</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>A to C</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>A to G</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>8 (12)</td>
</tr>
<tr>
<td>Del/Ins</td>
<td>0 (0)</td>
<td>1 (11)</td>
<td>4 (8)</td>
<td>4 (5)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Tandem</td>
<td>2 (4)</td>
<td>3 (4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>7 (100)</td>
<td>9 (100)</td>
<td>52 (100)</td>
<td>71 (100)</td>
<td>68 (100)</td>
</tr>
<tr>
<td>% of all mutations at CpG</td>
<td>17 (14%)</td>
<td>19 (11%)</td>
<td>13/52 (25%)</td>
<td>30/71 (42%)</td>
<td>41/68 (60%)</td>
</tr>
<tr>
<td>% of all G-to-T mutations at CpG</td>
<td>1/5 (20%)</td>
<td>1/5 (20%)</td>
<td>8/34 (23%)</td>
<td>20/48 (42%)</td>
<td>3/6 (50%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Lung cancer mutations in p53 coding sequence, exclusive of nonsmokers and occupational exposures (4).

<sup>b</sup> Percentage of all mutations in parentheses.

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 of a selectable phenotype (40). Interestingly, benzo[a]pyrene contains relatively high frequency of CpG sequences. This contrasts with the cII gene in the genome of BigBlue mice because both of these genes contain a single CpG site. We chose the HPRT transgene in the endogenous lacI gene, for example, which contains only four CpGs that can introduce one nucleotide. Tandem mutations are underlined. *, +, and # indicate the occurrence of two mutations in the same plasmid. CpG sequences are underlined.

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></th>
<th>supF control</th>
<th>supF + BPDE</th>
<th>p53 lung cancer&lt;sup&gt;a&lt;/sup&gt;</th>
<th>&lt;sup&gt;b&lt;/sup&gt;</th>
<th>&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unmethylated</td>
<td>Methylated</td>
<td>cII control</td>
<td>cII + BPDE</td>
<td>lacI + BPDE</td>
</tr>
<tr>
<td>G to T</td>
<td>5 (71)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5 (55)</td>
<td>34 (65)</td>
<td>48 (68)</td>
<td>6 (9)</td>
</tr>
<tr>
<td>G to T at CpG</td>
<td>1 (14)</td>
<td>1 (11)</td>
<td>8 (15)</td>
<td>20 (28)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>G to A</td>
<td>0 (0)</td>
<td>2 (22)</td>
<td>4 (8)</td>
<td>9 (13)</td>
<td>42 (62)</td>
</tr>
<tr>
<td>G to C</td>
<td>1 (14)</td>
<td>1 (11)</td>
<td>7 (13)</td>
<td>5 (7)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>A to T</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>A to C</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>A to G</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>8 (12)</td>
</tr>
<tr>
<td>Del/Ins</td>
<td>0 (0)</td>
<td>1 (11)</td>
<td>4 (8)</td>
<td>4 (5)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Tandem</td>
<td>2 (4)</td>
<td>3 (4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>7 (100)</td>
<td>9 (100)</td>
<td>52 (100)</td>
<td>71 (100)</td>
<td>68 (100)</td>
</tr>
<tr>
<td>% of all mutations at CpG</td>
<td>17 (14%)</td>
<td>19 (11%)</td>
<td>13/52 (25%)</td>
<td>30/71 (42%)</td>
<td>41/68 (60%)</td>
</tr>
<tr>
<td>% of all G-to-T mutations at CpG</td>
<td>1/5 (20%)</td>
<td>1/5 (20%)</td>
<td>8/34 (23%)</td>
<td>20/48 (42%)</td>
<td>3/6 (50%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Lung cancer mutations in p53 coding sequence, exclusive of nonsmokers and occupational exposures (4).

<sup>b</sup> Percentage of all mutations in parentheses.

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

Fig. 2. 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, +A, and C, +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 (+CH<sub>3</sub>); Lane 6, mock-methylated (-CH<sub>3</sub>) cII PCR products. The arrows indicate the positions of cytosines at methylated CpG sequences.
Mutational spectra of the cII transgene for spontaneous mutations (A) and BPDE-induced mutations (B) in embryonic mouse fibroblasts. The sequence positions 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.

![Image](cancerres.aacrjournals.org)
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, 2 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 M) had to be used than that in the mutagenesis experiments (0.2 M). However, in previous experiments we did not see any significant change of adduct patterns over a dose range of 0.2–4 M (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’-GCGGT), 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 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-url)

![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).](image-url)
sequences (codons 157, 158, 245, 248, and 273; see Fig. 6). Such specificity is not observed with other tumors, 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 tumor 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, hII, 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 CpGs (53, 54). Here, we have provided evidence that other mutational processes can occur selectively at methylated CpG sequences, in this case producing G-to-T transversion mutations. This type of mutational mechanism may be most important for tumorigenesis associated with environmental exposure, such as smoking-associated lung cancers.

ACKNOWLEDGMENTS

We thank Steven Bates for fibroblast cell culture and Jeff Longmate for help with the statistical analysis.

REFERENCES


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.


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

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, contact the AACR Publications Department at permissions@aacr.org.