# Influence of C-5 substituted cytosine and related nucleoside analogs on the formation of benzo[a]pyrene diol epoxide-dG adducts at CG base pairs of DNA

Rebecca Guza[1], Delshanee Kotandeniya[1], Kristopher Murphy[1], Thakshila Dissanayake[1,2], Chen Lin[3], George Madalin Giambasu[2], Rahul R. Lad[4], Filip Wojciechowski[5], Shantu Amin[6], Shana J. Sturla[4], Robert H.E. Hudson[5], Darrin M. York[2,7], Ryszard Jankowiak[3], Roger Jones[7] and Natalia Y. Tretyakova[1,*]

[1]Department of Medicinal Chemistry and the Masonic Cancer Center, [2]Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, [3]Department of Chemistry, Kansas State University, Manhattan, KS 66505, USA, [4]Institute of Food, Nutrition and Health, ETH Zurich, 8092 Zurich, Switzerland, [5]Department of Chemistry, The University of Western Ontario, London, Ontario, Canada, [6]Department of Chemistry, Pennsylvania State University and [7]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA

## ABSTRACT

Endogenous 5-methylcytosine ($^{Me}$C) residues are found at all CG dinucleotides of the *p53* tumor suppressor gene, including the mutational 'hotspots' for smoking induced lung cancer. $^{Me}$C enhances the reactivity of its base paired guanine towards carcinogenic diolepoxide metabolites of polycyclic aromatic hydrocarbons (PAH) present in cigarette smoke. In the present study, the structural basis for these effects was investigated using a series of unnatural nucleoside analogs and a representative PAH diolepoxide, benzo[a]pyrene diolepoxide (BPDE). Synthetic DNA duplexes derived from a frequently mutated region of the *p53* gene (5'-CCCGGC ACCC GC[$^{15}$N$_3$,$^{13}$C$_1$-G]TCCGCG-3', + strand) were prepared containing [$^{15}$N$_3$, $^{13}$C$_1$]-guanine opposite unsubstituted cytosine, $^{Me}$C, abasic site, or unnatural nucleobase analogs. Following BPDE treatment and hydrolysis of the modified DNA to 2'-deoxynucleosides, $N^2$-BPDE-dG adducts formed at the [$^{15}$N$_3$, $^{13}$C$_1$]-labeled guanine and elsewhere in the sequence were quantified by mass spectrometry. We found that C-5 alkylcytosines and related structural analogs specifically enhance the reactivity of the base paired guanine towards BPDE and modify the diastereomeric composition of $N^2$-BPDE-dG adducts. Fluorescence and molecular docking studies revealed that 5-alkylcytosines and unnatural nucleobase analogs with extended aromatic systems facilitate the formation of intercalative BPDE–DNA complexes, placing BPDE in a favorable orientation for nucleophilic attack by the $N^2$ position of guanine.

## INTRODUCTION

Benzo[a]pyrene (B[a]P, Scheme 1) is the best known representative of polycyclic aromatic hydrocarbons (PAH) produced from incomplete combustion of organic material. B[a]P and other PAHs are ubiquitously present in cigarette smoke, urban air and cooked food. Studies in animal models have shown that B[a]P is a potent systemic and local carcinogen that induces skin, stomach and lung tumors, and is considered a likely causative agent for smoking-induced cancer (1,2). Metabolic activation of B[a]P by cytochrome P450 monooxygenases produces DNA-reactive 'bay region' diol epoxides, e.g. (+)-*anti*-benzo[a]pyrene-*r*-7,*t*-8-dihydrodiol-*t*-9,10-epoxide [(+)-*anti*-BPDE] and (−)-*anti*-benzo[a]pyrene-*s*-7,*t*-8-dihydrodiol-*t*-9,10-epoxide [(−)-*anti*-BPDE] (Scheme 1) which are considered the ultimate carcinogenic species of B[a]P (3–5).

---

*To whom correspondence should be addressed. Tel: +612 626 3432; Fax: +612 626 5135; Email: trety001@umn.edu
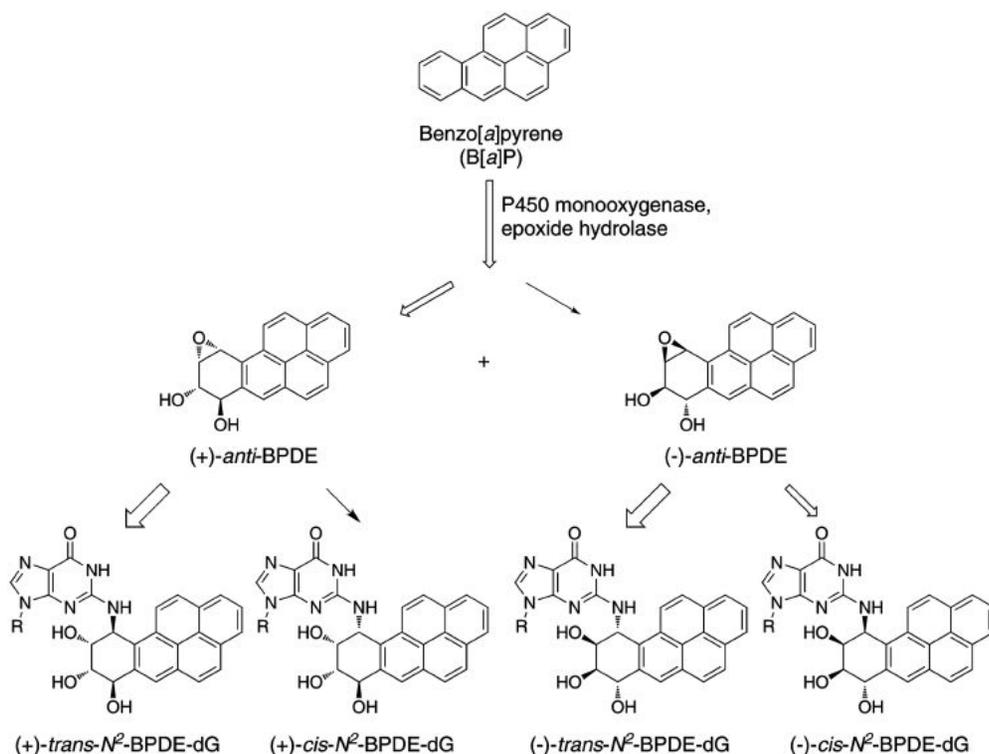
**Scheme 1.** Metabolic activation of benzo[*a*]pyrene to BPDE and the formation of guanine adducts.

Spectroscopic studies have shown that BPDE forms physical complexes with DNA, which precede their chemical reactions with DNA. Both major groove binding and intercalative BPDE-DNA binding motifs have been identified (6). In the intercalative complex, the pyrene ring system of BPDE is 'sandwiched' between two adjacent base pairs of DNA, leading to a red shift of the fluorescence maximum of BPDE from 343 to 353–354 nm (7). Importantly, intercalative BPDE–DNA complexes facilitate the nucleophilic attack by the exocyclic amino group of guanine at the C-10 epoxide position of BPDE and catalyze the 9,10-epoxide ring opening (8,9). *Trans* addition of guanine to the C-10 position of (+) and (−)-*anti*-BPDE produces (+)-*trans*-$N^2$-BPDE-dG and (−)-*trans*-$N^2$-BPDE-dG, while *cis* addition results in (+) and (−)-*cis*-$N^2$-BPDE-dG isomers (Scheme 1). While *trans*-$N^2$-BPDE-dG adducts are the most common diastereomers *in vivo* (10), the *cis* adducts are preferentially generated under high salt conditions (11,12). *Cis* and *trans* $N^2$-BPDE-dG have distinct conformations in DNA, which influences their recognition by DNA repair proteins (13,14). The BPDE moiety of the *trans* adducts is typically found in an external minor groove conformation, with B-DNA helix undergoing minimal distortion (15,16). In contrast, the *cis* adducts assume intercalative conformation by displacing the modified guanine residue and its partner cytosine into the major groove [the (−)-*cis* isomer] or into the minor groove of DNA [(+)-*cis*] (17). Adduct conformations can be further influenced by DNA supercoiling and by the local sequence context (18,19).

5-Methylcytosine ($^{Me}$C) is an important endogenous DNA modification that plays a central role in epigenetic regulation, chromatin structure and DNA repair (20). The majority of $^{Me}$C residues are found at 5′-CG-3′ dinucleotides, including all 5′-CG-3′ sites within the coding sequence of the human *p53* tumor suppressor gene (21). Interestingly, BPDE and other diol epoxides produced upon bioactivation of PAHs present in tobacco smoke exhibit an enhanced reactivity towards the $N^2$-position of guanine in $^{Me}$C:G base pairs (22–25). In particular, $N^2$-BPDE-dG adducts are preferentially formed at guanine bases within $^{Me}$CG dinucleotides found at the major mutational 'hotspots' characteristic for smoking-induced lung cancer, e.g. codons 157, 158, 245, 248 and 273 of the *p53* tumor suppressor gene (22–25). Our studies with partially methylated CG dinucleotides revealed that this reactivity enhancement can be attributed to $^{Me}$C that is base paired with target guanine, rather than the methylated cytosine present in the 5′-flanking position (22).

The origins of 5-methylcytosine mediated effects on $N^2$-BPDE-dG adduct formation are not well understood. Previous experimental and computational studies with mitomycin C, an antitumor agent that forms precovalent intercalative complexes with DNA, suggest that the nucleophilicity of the $N^2$-position of guanine is enhanced as a result of the inductive electronic effects of the C-5 methyl group transmitted to the $N^2$-amino group of G through $^{Me}$C:G hydrogen bonds (26). Consistent with this explanation, the presence of electron withdrawing

fluoro group on the C-5 of cytosine reduces the reactivity of G:C base pairs towards mitomycin C (26,27). These studies suggest that the transmitted electronic effect of the methyl group may be an important factor in determining reactivity of alkylating agents, such as BPDE, towards the $N^2$-position of guanine. However, our experiments with non-intercalating DNA alkylating agents have shown that cytosine methylation decreases the yields of $O^6$-alkyldeoxyguanosine adducts induced by alkyldiazonium ions (28,29) and does not affect the formation of $N^2$-ethyl-deoxyguanosine adducts of acetaldehyde (30). Therefore, C-5 cytosine methylation may enhance the reactivity of PAH diolepoxides towards CG dinucleotides by a mechanism that is distinct from its effects on electron density distribution.

Although cytosine methylation does not interfere with the overall DNA structure (the C-5 methyl substituent is readily accommodated in the major groove of DNA) (31–33), $^{Me}$C decreases the major groove charge density, stabilizes DNA helix and enhances base stacking (32,34). The introduction of a hydrophobic methyl group at the C-5 position of cytosine may encourage the physical binding of BPDE to DNA grooves. Furthermore, since

C-5 methylation increases the molecular polarizability of cytosine (32,35), it is likely to stabilize intercalative complexes of aromatic carcinogens such as BPDE with DNA via increased π–π stacking interactions (36). Indeed, Geacintov *et al*. (8) have shown that $^{Me}$C enhances the intercalative binding of BPDE to poly(dG-$^{Me}$C) duplexes and shifts the orientation of (−)-*trans* $N^2$-BPDE-dG adducts from an external minor groove structure to an intercalative conformation (37,38). The preferential precovalent complex formation at $^{Me}$C:G base pairs is expected to increase the yields of $N^2$-BPDE-dG adducts at methylated CG dinucleotides.

The purpose of this work was to identify the mechanism(s) by which the presence of the C-5 methyl group on cytosine alters the reactivity of endogenously occurring 5′-$^{Me}$CG-3′ steps in DNA towards BPDE and other PAH diolepoxides. A series of C-5 substituted cytosines and related structural analogs, including 5-ethyl-C, 5-propyl-C, N4-ethyl-C, 5-fluoro-C, 5-chloro-C (39), 5-bromo-C, 5-iodo-C, difluorotoluene (40), 5-propynyl-C, pyrrolo-C, 6-phenylpyrrolo-C (41) and perimidin-2-one nucleoside (dPER) (42) (Scheme 2), were placed into synthetic DNA duplexes opposite [$^{15}$N$_3$, $^{13}$C$_1$]-guanine and
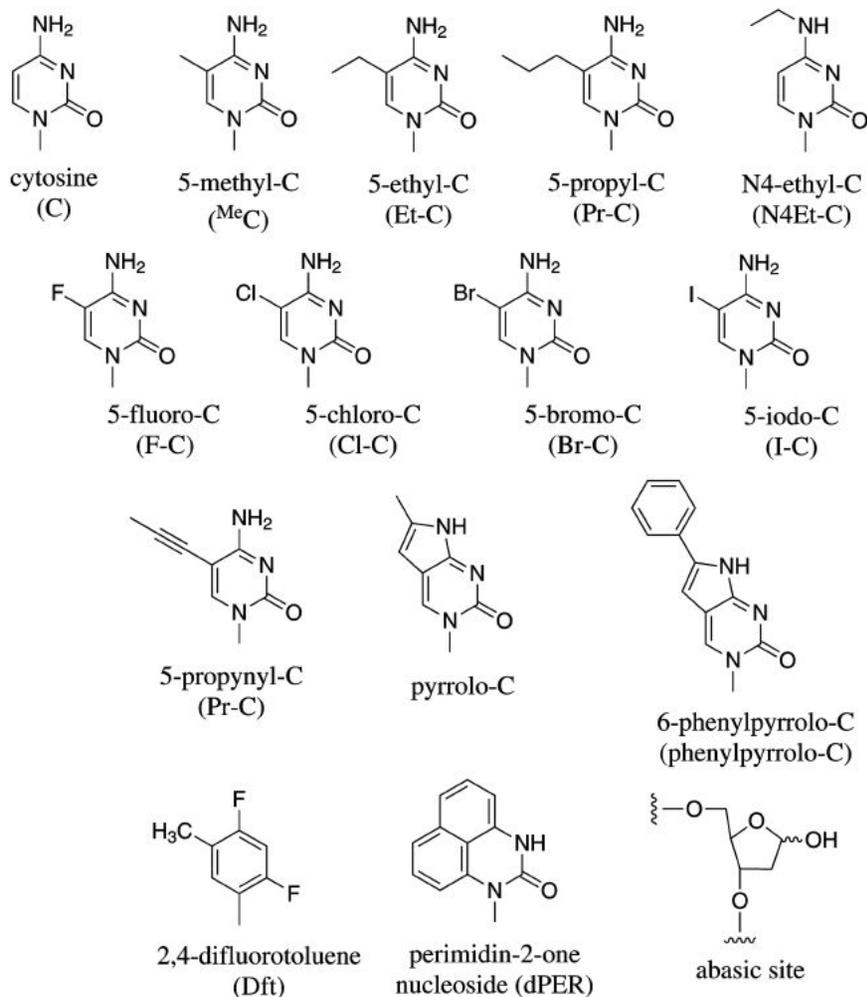


Scheme 2. Nucleoside analogs employed in the present study.

the formation of $N^2$-guanine adducts at the modified C:G base pairs was quantified by stable isotope labeling mass spectrometry. The pre-covalent BPDE:DNA interactions were evaluated by low temperature laser-excited fluorescence spectroscopy, quantum mechanical calculations and molecular docking experiments. Our results demonstrate that the presence of C-5 alkylcytosines and base analogs with extended aromatic systems facilitates the formation of intercalative BPDE-DNA complexes and increases the rates of $N^2$-BPDE-dG adduct formation at structurally modified G:C base pairs. Similar results were observed for other PAH diolepoxides, e.g. benzo[c]phenanthrene-3,4-diol-1,2-epoxide (B[c]PhDE), benzo[g]chrysene-11,12-diol-13,14-epoxide (B[g]CDE), dibenzo[a,l]pyrene-11,12-diol 13,14-epoxide (dB[a,l]PDE) and 5-methyl chrysene-1,2-diol-3,4-epoxide (5-MeCDE) (Supplementary Figure S1). Taken together, these findings offer a likely explanation for the increased re-activity of PAH diolepoxides and other DNA intercalating agents towards endogenously methylated G:C sites.

## MATERIALS AND METHODS

### Caution

Benzo[a]pyrene diol epoxide (BPDE), benzo[c]phenanthrene-3,4-diol 1,2-epoxide (B[c]PhDE), benzo[g]chrysene-11,12-diol 13,14-epoxide (B[g]CDE), dibenzo[a,l]pyrene-11,12-diol 13,14-epoxide (dB[a,l]PDE) and 5-methyl chrysene1,2-diol 3,4-epoxide (5-MeCDE) are carcinogenic and should be handled with extreme caution.

### Materials

(±)-*Anti*-BPDE, (−)-*anti*-BPDE, benzo[a]pyrene-7S-*trans*-7,8-dihydrodiol [(+)-BP78D] and benzo[a]pyrene-r-7,t-8,9,10-tetrahydrotetraol (*trans-anti* BPT or BPT) were obtained from the NCI Chemical Carcinogen Repository (Midwest Research Institute, Kansas City, MO, USA). B[c]PhDE, B[g]CDE, dB[a,l]PDE and 5-MeCDE were synthesized as described in the literature (43–46). DNase I, PDE I and PDE II were bought from Worthington Biochemical Corporation (Lakewood, NJ, USA). Alkaline phosphatase and DOWEX 50WX8-200 beads were purchased from Sigma-Aldrich (St Louis, MO, USA). Solid phase extraction (SPE) C18 SPE cartridges (50 mg) were obtained from Water Associates. Micro BioSpin 6 columns were purchased from Bio-Rad (Hercules, CA, USA) and illustra™ NAP-5 Columns were obtained from GE Healthcare (Piscataway, NJ, USA). All HPLC and liquid chromatography/mass spectrometry (LC/MS) grade solvents were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

### Synthesis of DNA strands containing unnatural base analogs

Synthetic oligodeoxynucleotides representing codons 153–158 of the human *p53* tumor suppressor gene (5′-CCCGG CACCCGC [$^{15}N_3$, $^{13}C_1$-G]TCCGCG-3′) were prepared by solid phase synthesis on an ABI 394 DNA Synthesizer (Applied Biosystems, CA, USA) (47). Structurally modified C:G base pairs were incorporated at the first position of *p53* codon 157, which is a prominent mutational 'hotspot' in smoking induced lung cancer (48).

Nucleoside phosphoramidites (PA) of 5-fluoro-C, 5-bromo-C, 5-iodo-C, difluorotoluene, 5-propynyl-C, protected abasic site and pyrrolo-C were purchased from Glen Research Corporation (Sterling, VA, USA). 1,7,NH₂-$^{15}N_3$-2-$^{13}C_1$-dG-phosphoramidite and perimidin-2-one nucleoside (dPER) phosphoramidite were prepared as reported previously (42,49,50). $O^4$-ethyl-5-chloro-2′-deoxyuridine phosphoramidite was synthesized and incorporated into the oligodeoxynucleotide following the methods of Kang *et al.* (39). 5-Ethyl-2′-deoxycytidine and 5-propyl-2′-deoxycytidine were prepared from 5-iodo-2′-deoxycytidine by the methods described by Robins *et al.* (51) and converted to the corresponding PA derivatives by standard phosphoramidite chemistry (47). 6-Phenylpyrrolo-dC phosphoramidite was prepared as described previously (52). Manual coupling was employed for the incorporation of structurally modified nucleotide units during solid phase DNA synthesis. Synthetic oligodeoxynucleotides were purified by reversed phase HPLC as described elsewhere, (22,28) followed by desalting using illustra™ NAP-5 size exclusion cartridges (GE Healthcare).

Synthetic oligodeoxynucleotides containing protected abasic site (1,2,5-O-tris(*tert*-butyldimethylsilyl)-3-deoxyhexitol) were synthesized using commercial phosphoramidite (Glen Research Corporation, Sterling, VA, USA) and purified using a Supelcosil LC-18-DB column (10 mm × 250 mm, 5 mm, Supelco, Bellefonte, PA, USA) eluted at 40°C and a flow rate of 3 ml/min. HPLC buffers were 100 mM triethylammonium acetate, pH 7.0 (A) and acetonitrile (B). A linear gradient of 15–50% B in 35 min was employed, followed by isocratic elution at 50% B for an additional 5 min. The *tert*-butyldimethylsilyl protecting groups were removed by incubation in 40% acetic acid for 4 h at room temperature, and the deprotected oligodeoxynucleotide containing 1,2,5-trihydroxy-3-deoxyhexitol was purified on a Supelcosil LC-18-DB column (10 mm × 250 mm, 5 mm) maintained at 40°C. The 1,2,5-trihydroxy-3-deoxyhexitol moiety was converted to an abasic site by incubation in 100 mM sodium acetate buffer, pH 6.0, containing 5 mM sodium periodate for 5 min at room temperature.

Following desalting, DNA strands containing structurally modified bases were characterized by HPLC-ESI⁻ MS (Supplementary Table S1) (22), UV melting (Supplementary Table S2) and exonuclease ladder sequencing (Supplementary Tables S3 and S4) (53). To obtain double-stranded DNA, equimolar amounts of the complementary strands (200 μM) were combined in 10 mM Tris–HCl buffer, pH 8.0, containing 50 mM NaCl, followed by heating to 90°C and slowly cooling to room temperature. CD spectra and UV melting temperatures of each structurally modified duplex were obtained

as described in the Supplementary Data and are shown in Supplementary Figures S2 and S3, respectively.

## DNA treatment with PAH diolepoxides and analysis of $N^2$-PAH-dG adducts

Double-stranded DNA 19-mers containing centrally positioned X:[$^{15}N_3$,$^{13}C_1$-dG] base pairs, where X = cytosine, 5-methylcytosine or unnatural nucleobase analog (Supplementary Table S2, Scheme 2), were dissolved in 50 mM Tris–HCl buffer, pH 7.5, to reach a concentration of 40 μM (22). Stock solutions of (±)-*anti*-BPDE, (−)-*anti*-BPDE, B[c]PhDE, B[g]CDE, dB[a,l]PDE and 5-MeCDE were prepared in dry DMSO, and their concentrations were determined by UV spectrophotometery (BPDE: $\varepsilon_{345} = 48\,800\,M^{-1}\,cm^{-1}$; B[c]PhDE: $\varepsilon_{266} = 32\,106\,M^{-1}\,cm^{-1}$, B[g]CDE: $\varepsilon_{269} = 10\,881\,M^{-1}\,cm^{-1}$, dB[a,l]PDE: $\varepsilon_{299} = 21\,717\,M^{-1}\,cm^{-1}$, 5-MeCDE: $\varepsilon_{266} = 6580\,M^{-1}\,cm^{-1}$). An appropriate amount of each PAH diolepoxide was added to the DNA solution to achieve 8 μM concentration, with DMSO making up 10% of the volume. Following 18 h incubation on ice, the reaction mixtures were dried under reduced pressure and re-dissolved in 10 mM Tris–HCl/15 mM MgCl$_2$ buffer, pH 7. Enzymatic digestion of DNA to 2′-deoxyribonucleosides was achieved by incubation with DNAse I (105 U), PDE I (105 mU), PDE II (108 mU) and alkaline phosphatase (22 U), for 18 h at 37°C. To confirm that the enzymatic digestion was complete, a small aliquot of each digest was removed and analyzed by HPLC-UV (22). The resulting PAH diolepoxide–deoxyguanosine adducts were isolated by SPE on C18 SPE cartridges (50 mg, from Water Associates) as described previously (25). SPE fractions containing modified nucleosides were concentrated under vacuum and analyzed by HPLC-ESI-MS/MS.

## HPLC-ESI-MS/MS analysis

Capillary HPLC-ESI$^+$-MS/MS analyses were performed with an Agilent 1100 capillary HPLC system interfaced to a Finnigan Quantum Discovery triple quadrupole (TSQ) mass spectrometer. Chromatographic separation of $N^2$-BPDE-dG diastereomers was achieved with a Waters 3 μm Atlantis C18 column (300 μm × 150 mm) maintained at 15°C and eluted at a flow rate of 3.5 μl/min. The HPLC solvents were 15 mM ammonium acetate (A) and methanol (B), with a linear gradient of 47–50% B in 30 min, followed by isocratic elution at 50% B for 20 min. $N^2$-BPDE-dG stereoisomers eluted as follows: (−)-*trans*-$N^2$-BPDE-dG ($t_R$, 43.0 min), (+)-*cis*-$N^2$-BPDE-dG ($t_R$, 45.5 min), (−)-*cis*-$N^2$-BPDE-dG ($t_R$, 49.2 min) and (+)-*trans*-$N^2$-BPDE-dG ($t_R$, 51.3 min). The mass spectrometer was operated in the ESI$^+$ mode. Selected reaction monitoring (SRM) analysis was performed by following the transitions corresponding to a loss of deoxyribose from protonated molecules of the adducts: $m/z$ 570.1→454.1 ($N^2$-BPDE-dG) and $m/z$ 574.1→459.1 ($^{15}N_3$, $^{13}C_1$-$N^2$-BPDE-dG) (22). Isotopic labeling experiments involving $N^2$-B[c]PhDE-dG, $N^2$-B[g]CDE-dG, $N^2$-dB[a,l]PDE and $N^2$-5-MeCDE-dG

were conducted analogously as described in Supplementary Data.

The extent of $N^2$-BPDE-dG formation at the $^{15}N_3$, $^{13}C_1$-dG was calculated according to the equation:

$$\text{Percentage reaction at X} = \frac{A_{15N,13C-BPDE-dG}}{(A_{15N,13C-BPDE-dG} + A_{BPDE-dG})},$$

where $A_{BPDE-dG}$ and $A_{15N,13C-BPDE-dG}$ are the areas under the HPLC-ESI-MS/MS peaks corresponding to the unlabeled and [$^{15}N$, $^{13}C$]-labeled $N^2$-BPDE-dG, respectively.

Results were expressed as a ratio of relative reactivity of $^{15}N_3$, $^{13}C_1$-G when paired with specific base analog versus the reactivity of the standard G: C base pair. Relative contributions of $N^2$-BPDE-dG isomers to the total adduct number were expressed as percentages in pie chart format. Statistical analyses of the data were conducted as described in the Supplementary Data.

Chromatographic separation of (−)-*trans*-$N^2$-BPDE-dG and (−)-*cis*-$N^2$-BPDE-dG adducts resulting from DNA treatment with (−)-*anti*-BPDE was achieved with a YMC ODS-AQ column (320 μm × 100 mm, 5 μm). The column was maintained at 30°C and eluted at a flow rate of 8 μl/min isocratically with 5 mM ammonium acetate containing 0.02% formic acid and 44.5% B methanol. Under these conditions, the retention times of (−)-*trans*-$N^2$-BPDE-dG and (−)-*cis*-$N^2$-BPDE-dG were 10.0 and 11.5 min, respectively (Supplementary Figure S4). The mass spectrometer was operated in the ESI$^+$-MS/MS mode, and the quantitative analysis of $N^2$-BPDE-dG was performed as described above.

## Kinetic analysis of $N^2$-BPDE-dG adduct formation in G:C and G:$^{Me}$C base pairs

Double-stranded DNA 19-mers containing centrally positioned X:[$^{15}N_3$,$^{13}C_1$-dG] base pairs (5′-CCCGGCACCC GC[$^{15}N_3$, $^{13}C_1$-G]TCCGCG-3′, + strand; GGG CCG TGG GCG XAG GCG C, – strand, where X = cytosine or $^{Me}$C; 25–1000 pmol), were dissolved in 50 mM Tris–HCl buffer (45 μl, pH 7.5). (±)-*Anti*-BPDE (400 pmol, in 5 μl of DMSO) was added to the DNA sample, followed by rapid mixing by vortex. Following 1 min incubation, the reactions were quenched with the addition of 2-mercaptoethanol (20 nmol). Samples were spiked with B[g]CDE-dG (2.7 pmol, internal standard for mass spectrometry), dried under reduced pressure, and re-dissolved in 10 mM Tris–HCl/15 mM MgCl$_2$ buffer, pH 7. Enzymatic digestion of DNA to 2′-deoxyribonucleosides and SPE of $N^2$-BPDE-dG and [$^{15}N_3$,$^{13}C_1$]-$N^2$-BPDE-dG] adducts were conducted as described previously (25). SPE fractions containing $N^2$-BPDE-dG were concentrated under vacuum and analyzed by HPLC-ESI$^+$-MS/MS as described above. $N^2$-BPDE-dG and [$^{15}N_3$,$^{13}C_1$]-$N^2$-BPDE-dG] were quantified from the corresponding HPLC-ESI$^+$-MS/MS peak areas using calibration curves constructed by analyzing solutions containing known amounts of B[g]CDE-dG, $N^2$-BPDE-dG and [$^{15}N_3$,$^{13}C_1$]-$N^2$-BPDE-dG.

**Low-temperature fluorescence spectroscopy**

Fluorescence spectra of benzo[*a*]pyrene-r-7,t-8,9,10-tetrahydrotetraol (*trans-anti* BPT or BPT) and its complexes with DNA were recorded at 77 K using excitation wavelengths of 346 or 355 nm (54). Double-stranded DNA 19-mers containing a centrally positioned X:G base pair (5'-CCCGGCACCC GCGTCCGCG-3', + strand; GGG CCG TGG GCG XAG GCG C, – strand, where X = cytosine, 5-methylcytosine, or dPER) were dissolved in 50 mM Tris–HCl buffer, pH 7.5, to reach a concentration of 40 μM. Stock solutions of BPT were prepared in DMSO and its concentrations were determined by UV spectrophotometry ($\varepsilon_{345}$ = 48 800 $M^{-1}$ $cm^{-1}$). An appropriate aliquot of the BPT solution was added to the DNA solution to achieve 8 μM concentration of BPT, with DMSO making up 10% of the volume. Control samples in the absence of DNA contained 8 μM BPT dissolved in 50 mM Tris–HCl, pH 7.5 buffer, with DMSO making up 10% of the volume. For all fluorescence measurements, samples were placed into quartz tubes and immediately (within ∼5–8 s) immersed in a liquid nitrogen cryostat with quartz optical windows. Fluorescence spectra were taken at 77 K using laser excitation at 346 and 355 nm excitation wavelenghts provided by the excimer pumped dye laser (i.e. a Lambda Physik FL 2002). Scanmate tunable dye laser system was operated at the frequency of 10 Hz. A 1 m McPherson monochromator (model 2601) and a Princeton Instruments photodiode array were used for the dispersion and the detection of fluorescence. A Princeton Instruments FG-100 pulse generator was used for time-resolved spectroscopy with detector delay times from 60 to 100 ns and a gate width of 200 ns. Acquisition time (60 s) was applied to all spectra. The resolution for the fluorescence spectra was ±0.4 nm.

**Computational studies**

All density functional calculations were performed in accord with the standardized protocol used to construct the *QCRNA* database, a recently developed online database of quantum calculations for RNA catalysis (http://theory.chem.umn.edu/Database/QCRNA) (55). All structures were optimized in the gas phase with B3LYP/6-31++G** as implemented in the Gaussian03 suite of programs and described by Giese *et al.* (55). The molecular structure of (−)-*anti* BPDE was built in Gaussian03 (Gaussian 03 revision E.01, Gaussian, Inc., Wallingford, CT, USA) using B3LPY and HF/6-31G (55). The intercalation sites (within GC/CG and AT/CG stacks) were obtained from the PDB data bank with the PDB ID 151D (56) and 1CXO (57), respectively. The structures of the DNA duplexes were obtained by merging the crystal structure of the intercalation sites with the rest of the duplex that were assumed to be standard B-DNA motifs. The duplexes were energy minimized for 20 000 steps, heated to 300 K for 40 000 steps, and cooled to 0 K for 40 000 steps for three cycles using Molecular Dynamics with 1 fs integration steps. The intercalation sites were restrained using a harmonic penalty function.

All calculations were performed using Generalized Born (GBSA model) (58) model in AMBER 10 software (AMBER 10, University of California, San Francisco), using the parm99 force field for nucleic acids with the ff99bsc0 correction (59–61).

The docking calculations were carried out for each of the BPDE enantiomers in both axial and equatorial conformations (Supplementary Figure S6) using a grid centered at the intercalation site using Lamarckian Genetic Algorithm (LGA) as implemented in Autodock 4 (62). Molecular Dynamics was performed on the structures of the complexes for 420 ps using GBSA in Amber 10. Computational results are summarized in Supplementary Table S6.

## RESULTS

**Selection of nucleobase analogs and oligonucleotide synthesis**

In order to investigate the mechanisms by which endogenous $^{Me}$C increases the efficiency of PAH-guanine adduct formation at the base paired guanine, a series of synthetic oligodeoxynucleotide duplexes were prepared representing codons 153–158 of the *p53* tumor suppressor gene (Supplementary Tables S1 and S2). This region of the *p53* gene was selected because it is endogenously methylated in all human tissues (21), and because endogenously methylated codon 157 is specifically targeted for BPDE-dG adduct formation as compared with neighboring guanine bases (24). Furthermore, mutations at the *p53* codon 157 (GTC→TGC) are characteristic for smoking-induced lung cancer (48). In each duplex, $^{15}N_3$, $^{13}C_1$-dG (G) was placed at the first position of *p53* codon 157, while cytosine or its structural analog (X) was introduced in the base paired position. All synthetic oligodeoxynucleotides were purified by HPLC, desalted and characterized by ESI⁻MS (Supplementary Table S1). Exonuclease ladder sequencing of the oligodeoxynucleotides using MALDI-TOF MS confirmed the presence of unnatural nucleobases at specified sites within DNA sequence (Supplementary Tables S3 and S4).

Nucleoside analogs selected for this study span a range of structures, including C-5 substituted cytosine and related aromatic and heteroaromatic nucleobase variants (Scheme 2 above). Analogs with C-5 alkyl substituent on cytosine [5-methylcytosine ($^{Me}$C), 5-ethylcytosine (Et-C), $N^4$-ethylcytosine (N4Et-C) and 5-propylcytosine (Prop-C)] were chosen to investigate the possibility that the presence of alkyl groups at the C-5 of cytosine enhances physical and intercalative binding of BPDE and other PAH diol epoxides to CG dinucleotides, thereby increasing the probability of $N^2$-BPDE-dG adduct formation. In addition, the C-5 alkyl groups on cytosine can donate electron density to the C:G base pair, increasing the nucleophilicity of its partner guanine (26,27).

A series of C-5 halogenated cytosines [5-fluorocytosine (F-C), 5-chlorocytosine (Cl-C), 5-bromocytosine (Br-C)

and 5-iodocytosine (I-C)] were included to probe the effects of electron withdrawing groups at the C-5 of cytosine on reactivity of C:G pairs towards BPDE. The steric impact of these substituents is complementary to that of σ electron-donating alkyl groups. Furthermore, base analogs with extended aromatic systems [5-propynylcytosine (Pr-C), 6-methylpyrrolo[2,3-d] pyrimidine-2(3H)one deoxyribonucleoside (pyrrolo-C), 6-phenylpyrrolo[2,3-d]pyrimidine-2(3H)one deoxyribo-nucleoside (phenylpyrrolo-C) and perimidin-2-one deoxyribonucleoside (dPER)] were introduced in order to enhance the π–π stacking interactions of modified base pairs with BPDE. Two additional base analogs, 2,4-difluorotoluene (Dft) (40) and the abasic site, were included to probe the reactivity of guanine towards BPDE in the absence of complementary hydrogen bonding to cytosine.

### Effects of nucleobase analogs on DNA duplex structure and stability

CD spectra of DNA 19-mers containing various nucleobase mimics (Scheme 2) were similar to the corresponding spectra of duplexes containing unsubstituted dC and $^{Me}$C (Supplementary Figure S2), indicating that the presence these nucleobase modifications does not alter the overall DNA structure. Furthermore, UV variable temperature analysis resulted in hyperbolic thermal denaturation curves indicative of the presence of stable B-DNA duplexes (Supplementary Figure S3). UV melting temperatures of the duplexes containing $^{Me}$C, Et-C and Pr-C were slightly higher than for the duplex containing unsubstituted cytosine (Supplementary Table S2), although these differences were not statistically significant. This increase is consistent with the known ability of $^{Me}$C to enhance base stacking and increase DNA duplex stability (35,63). Small $T_m$ increases were also observed for duplexes containing Br-C, Cl-C, I-C and pyrrolo-C (Supplementary Table S2). In contrast, the thermodynamic stability of DNA containing N4Et-C, Dft, dPER and the abasic site was decreased relative to the duplexes containing standard G:C base pairs, probably a result of the disrupted Watson–Crick hydrogen bonding between guanine and these nucleoside analogs (42,64,65). Indeed, the introduction of N4-ethyl group is expected to disrupt the H-bonding interactions between the $N^4$ of ethylated dC and the $O^6$ of its partner guanine, while no hydrogen bonding with G is possible for Dft, dPER and the abasic site (40,42). Only insignificant changes in $T_m$ were observed upon the introduction of 5-propyl-C, 5-fluoro-C and 6-phenylpyrrolo-C into DNA duplexes (Supplementary Table S2).

### Stable isotope labeling HPLC-ESI-MS/MS approach

In order to evaluate the effects of nucleobase analogs on the $N^2$-BPDE-dG adduct formation at the base paired guanine, isotopically tagged deoxyguanosine ($^{15}N_3$, $^{13}C_1$-dG) was placed in DNA duplexes (5′-CCCG GCACCCGC [$^{15}N_3$, $^{13}C_1$-G] TCCGCG-3′, + strand) opposite cytosine or a cytosine analog (X) (Scheme 3,

Supplementary Table S2). The resulting structurally modified DNA duplexes were treated with (±)-*anti*-BPDE or (−)-*anti*-BPDE and enzymatically digested to 2′-deoxynucleosides. The resulting diastereoisomeric $N^2$-BPDE-dG adducts were quantified by HPLC-ESI$^+$-MS/MS (Figure 1 and Supplementary Figure S4) (22,25). $N^2$-BPDE-dG adducts formed at the $^{15}N_3$, $^{13}C_1$-labeled guanine undergo a +4 mass shift and thus can be readily distinguished from adducts originating from unlabeled deoxyguanosines elsewhere in the sequence (22,25). The extent of $N^2$-BPDE-dG formation at the $^{15}N$, $^{13}C$-labeled guanine was calculated from HPLC-ESI-MS/MS peak areas as described in the 'Materials and Methods' section. Similar methodology was employed to quantify the site-specific formation of $N^2$-guanine adducts induced by other PAH diolepoxides (Supplementary Figure S1).

### Influence of C-5 cytosine substituents within CG base pairs on relative reactivity of (±)-anti-BPDE and (−)-anti-BPDE

Normalized yields of $N^2$-BPDE-dG adducts originating from structurally modified G:X base pair (where X = unnatural cytosine analog) in comparison with that from native G:C pairs are shown as bar graphs in Figures 2 and 4-6, while the relative contributions of individual $N^2$-BPDE-dG stereoisomers to the total adduct number are represented as pie charts in the same figure. We found that the yields of BPDE adducts at guanine placed opposite $^{Me}$C were ~2-fold higher than at guanines within standard G:C base pairs (Figure 2). This observation was confirmed by kinetic experiments in which (±)-*anti*-BPDE was allowed to react with increasing amounts of DNA containing a central $^{Me}$C:G base pair or unmethylated C:G (Figure 3). $N^2$-BPDE-dG adduct formation at the guanine base paired to $^{Me}$C was 1.7-2 fold higher than in the absence of $^{Me}$C (Figure 3A). In contrast, $^{Me}$C had no effect on $N^2$-BPDE-dG adduct formation at guanine bases elsewhere in the sequence (Figure 3B), suggesting that cytosine methylation specifically activates its partner guanine in the opposite strand of DNA towards carcinogen attack. Similar results were obtained for other PAH diol epoxides, e.g. (±)-B[*c*]PhDE, (±)-B[*g*]CDE, (±)-dB[*a,l*]PDE) and (±)-5-MeCDE, with the yields of $N^2$-guanine adducts 1.2- to 2.8-fold higher at $^{Me}$C:G base pairs as compared to unmethylated C:G (Supplementary Figure S1, Supplementary Table S5). An even greater reactivity increase (up to 3-fold) was observed for cytosine analogs containing a larger C-5 alkyl substituent, e.g. Et-C and propyl-C, while the presence of N4-ethyl-dC was less activating (Figure 2).

To investigate the effects of electron withdrawing groups on C-5 of cytosine on reactivity of CG base pairs towards BPDE, isotopically labeled guanine was placed in a DNA duplex opposite F-C, Cl-C, Br-C, or I–C (Figure 4). We found that the presence of halogenated cytosine either increased or decreased the yields of BPDE-dG adducts, depending on halogen identity (Figure 4). A 20–40% decline in reactivity was observed when guanine was paired with 5-fluoro-dC (F-C) as
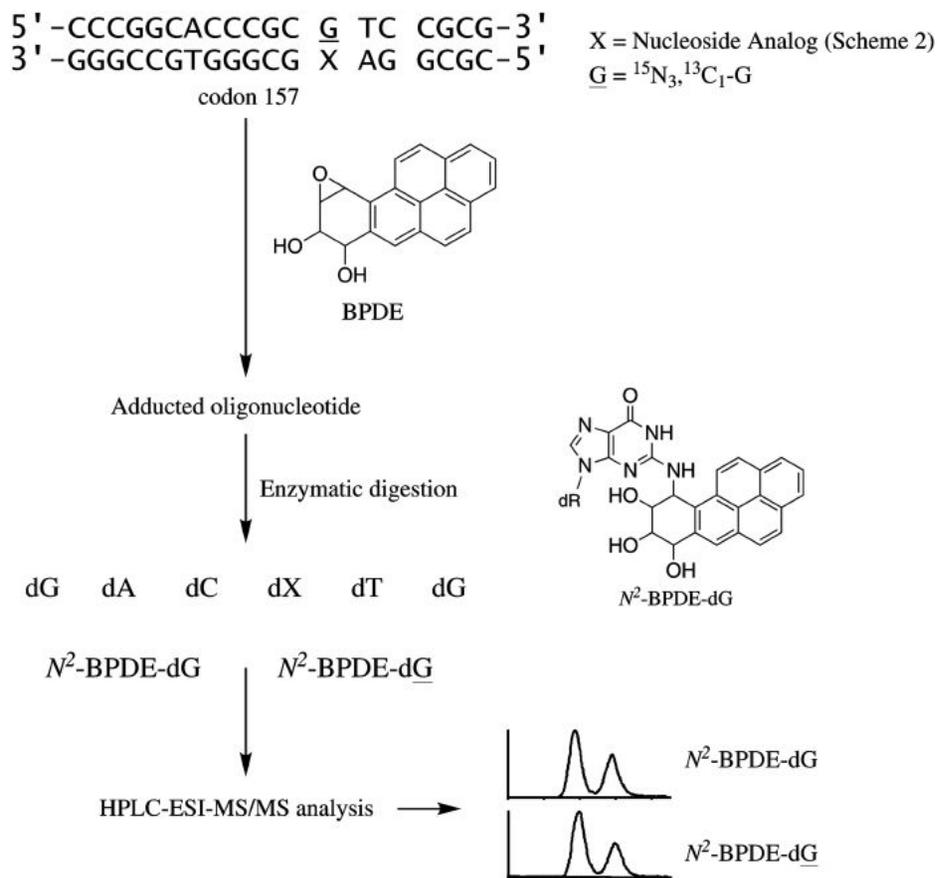
**Scheme 3.** Experimental scheme for probing the reactivity of structurally modified C:G base pairs towards BPDE.

compared with unsubstituted cytosine ($P$-value = 0.0142), suggesting that the negative inductive effect of the C-5 fluoro group decreases the nucleophilicity of the $N^2$-position of guanine (26,27). However, guanine-BPDE adduct formation increased when it was base paired with Cl-C, Br-C, or I-C (Figure 4), indicating that electronic factors alone cannot explain the observed reactivity trends. A similar effect was observed when a structurally related PAH diolepoxide, B[c]PhDE, was reacted with DNA duplexes containing $^{Me}$C, Cl-C, Br-C and I-C (Supplementary Figure S5).

The most significant increase in BPDE adduct formation (4- to 8-fold) was observed for G:X base pairs containing nucleobase analogs designed to enhance carcinogen intercalation into the DNA duplex (Pr-C, Dft, pyrrolo-C, phenylpyrrolo-C and dPER, Figures 5 and 6). When pyrrolo-C or phenylpyrrolo-C was placed opposite target guanine in a DNA duplex, the relative reactivity towards ($\pm$)-anti-BPDE was increased by 60–243% as compared to the native G:C base pair ($P \leq 0.0001$) (Figure 5). An even greater effect (8-fold increase) was observed when dPER containing DNA duplex was treated with ($-$)-anti-BPDE (Figure 6B). This can be attributed to the increased $\pi$–$\pi$ stacking interactions between BPDE and modified G:C base pairs containing nucleobase analogs with extended aromatic systems. Interestingly, a smaller overall effect was

observed for reactions with ($\pm$)-anti-BPDE (Figure 6A), suggesting that dPER specifically stabilizes the precovalent complex with the ($-$)-anti diolepoxide. The introduction of an abasic site opposite the labeled guanine also increased its reactivity toward ($\pm$)-anti-BPDE (Figure 6). In contrast, the presence of 5-propynyl group on cytosine did not increase BPDE-dG adduct formation, possibly because of the unfavorable geometry of the propynyl substituent which extends far out of the DNA base stack and into the major groove of DNA (66).

**Effects of cytosine analogs on the optical composition of $N^2$-BPDE-dG adducts**

Four diastereoisomers of $N^2$-BPDE-dG: ($+$)-trans-$N^2$-BPDE-dG, ($+$)-cis-$N^2$-BPDE-dG, ($-$)-trans-$N^2$-BPDE-dG and ($-$)-cis-$N^2$-BPDE-dG, can be formed when DNA is treated with ($\pm$)-anti-BPDE (Scheme 1). Two of these, ($-$)-trans-$N^2$-BPDE-dG and ($-$)-cis-$N^2$-BPDE-dG, are formed following treatment with ($-$)-anti-BPDE. When guanine is base paired with normal cytosine, the major $N^2$-BPDE-dG adducts generated following treatment with racemic anti-BPDE are ($+$)-trans-$N^2$-BPDE-dG (89%), followed by ($-$)-trans-$N^2$-BPDE-dG (5%), ($-$)-cis-$N^2$-BPDE-dG (4%) and ($+$)-cis-$N^2$-BPDE-dG (2%) (pie charts in Figure 2A). A similar
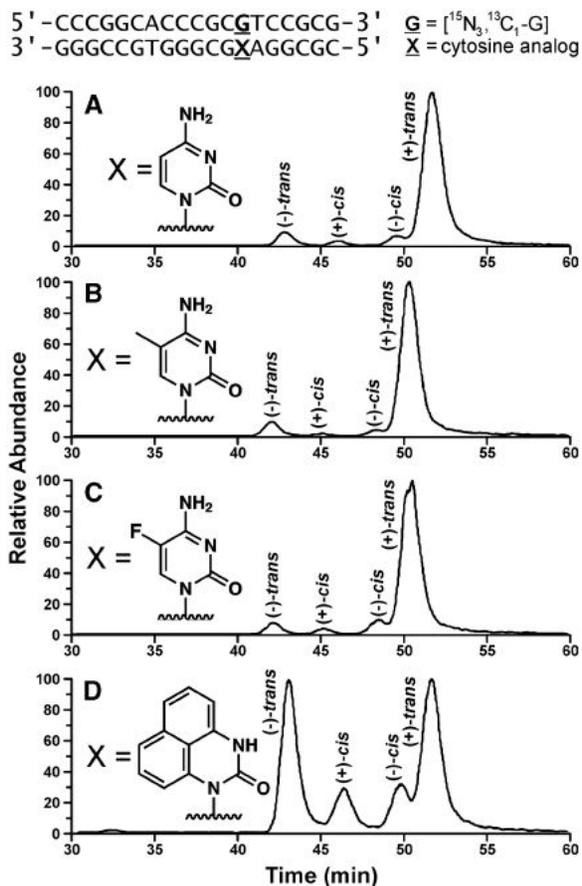
**Figure 1.** HPLC-ESI-MS/MS analysis of $N^2$-BPDE-dG diastereomers formed at the $^{15}N_3$, $^{13}C_1$-labeled guanine when base paired with cytosine (**A**), 5-methylcytosine (**B**), 5-fluorocyrosine (**C**) or perimidin-2-one nucleoside (**D**) following treatment with ($\pm$)-*anti*-BPDE. Selected reaction monitoring of $^{15}N_5$,$^{13}C_1$-$N^2$-BPDE-dG was conducted by following the transitions $m/z$ 574.1 [M+H]$^+$ → 459.1 [M+2H-dR]$^+$.

stereoisomer contribution [85–93% (+)-*trans*-$N^2$-BPDE-dG, 4–6% (−)-*trans*-$N^2$-BPDE-dG, 2–5% (−)-*cis*-$N^2$-BPDE-dG and 1–5% (+)-*cis*-$N^2$-BPDE-dG] is observed for guanine base paired to $^{Me}$C, Et-C, Propyl-C, N4Et-C, Cl-C, Br-C, pyrrolo-C or phenylpyrrolo-C (see the pie chart graphs in Figures 2A, 4A, 5A and 6A). However, a strikingly different stereospecificity was observed when the target guanine was placed opposite Dft, dPER or the abasic site (Figure 6A). For the guanine:Dft pairs, (+)-*trans*-$N^2$-BPDE-dG and (+)-*cis*-$N^2$-BPDE-dG adducts were formed in equal amounts (33–39% each), while the minor isomers were (−)-*trans*-$N^2$-BPDE-dG and (−)-*cis*-$N^2$-BPDE-dG (12–18% each). (+)-*Trans*-$N^2$-BPDE-dG and (−)-*trans*-$N^2$-BPDE-dG were the most abundant isomers formed at the G: dPER base pairs (Figure 6). As discussed above, the presence of dPER appears to increase the reactivity of modified DNA duplexes towards (−)-*anti*-BPDE. When target guanine was placed opposite an abasic site, (+)-*cis*-$N^2$-BPDE-dG became the major adduct (46%), followed by (+)-*trans*-$N^2$-BPDE-dG (25%), (−)-*cis*-$N^2$-BPDE-dG (18%)

and (−)-*trans*-$N^2$-BPDE-dG (11%, see the pie charts in Figure 6A). These differences are not due to salt effects, since all synthetic oligodeoxynucleotides used in these experiments were desalted and annealed under carefully controlled conditions.

A smaller, but still significant, shift in stereochemistry was observed when the same duplexes were treated with optically pure (−)-*anti*-BPDE (pie charts in Figures 2B, 4B, 5B and 6B). While (−)-*trans*-$N^2$-BPDE-dG was the major diastereomer produced at the guanine bases paired to C, $^{Me}$C, Et-C, propyl-C, N4Et-C, Cl-C, Br-C, Dft, pyrrolo-dC, phenylpyrrolo-C and dPER (55–75%), (−)-*cis*-$N^2$-BPDE-dG became the main adduct when $^{15}N_3$, $^{13}C_1$-labeled guanine was placed opposite F-C, I-C, Pr-C, or the abasic site (pie charts in Figures 4B and 6B). Taken together, our results indicate that the presence of unnatural cytosine analogs strongly influences the stereoisomeric composition of $N^2$-BPDE-dG adducts produced at the base paired guanine, probably by changing BPDE conformation in the precovalent BPDE:DNA complexes.

## Spectroscopic studies of trans-anti-BPT complexed with DNA

Fluorescent spectra of molecules with pyrene moiety are known to be sensitive to solvent effect (67,68). Low-temperature fluorescence spectroscopy is frequently used to probe interactions between these chromophores and their environments through spectral shifts and relative peak intensity changes. This technique has been previously applied to study the conformation of BPDE-DNA adducts in DNA and physical complexes of benzo[a]pyrene tetrols with anti-PAH monoclonal antibody (69). In the present study, fluorescence studies were conducted in order to evaluate the pre-covalent interactions between (−)-*anti*-BPDE and DNA duplexes containing standard and structurally modified base pairs. Since the diolepoxide is hydrolytically unstable in the presence of DNA (7), *trans-anti*-BPT was used as a model compound. This molecule has the same stereochemistry at the C-7, C-8, C-9 and C-10 of the hydrocarbon as (+)-*anti*-BPDE and retains a similar molecular shape. Our working hypothesis was that the affinity of BPDE and BPT towards C:G base pairs in DNA depends on the identity of the C-5 cytosine substituent.

Figure 7 shows low-temperature fluorescence spectra of BPT alone and in the presence of DNA duplexes containing a central G:C, $^{Me}$C:G or dPER:G base pair. Two laser wavelengths at 346 and 355 nm were chosen to excite BPT/DNA physical complexes. These two wavelengths have been previously shown to be selective for external (346 nm) and intercalated complexes of PAHs with DNA (355 nm) (69,70). The origin-band emission spectra shown in Figure 7A were obtained with an excitation wavelength $\lambda_{ex}$ of 346 nm and 60 ns delay time of the observation window immediately following mixing of *trans-anti*-BPT with DNA. Curves **a–c** in Figure 7A were obtained in the presence of DNA duplexes containing a central C:G, $^{Me}$C:G or dPER:G base pair, respectively. Spectrum **d** corresponds to BPT in the absence of
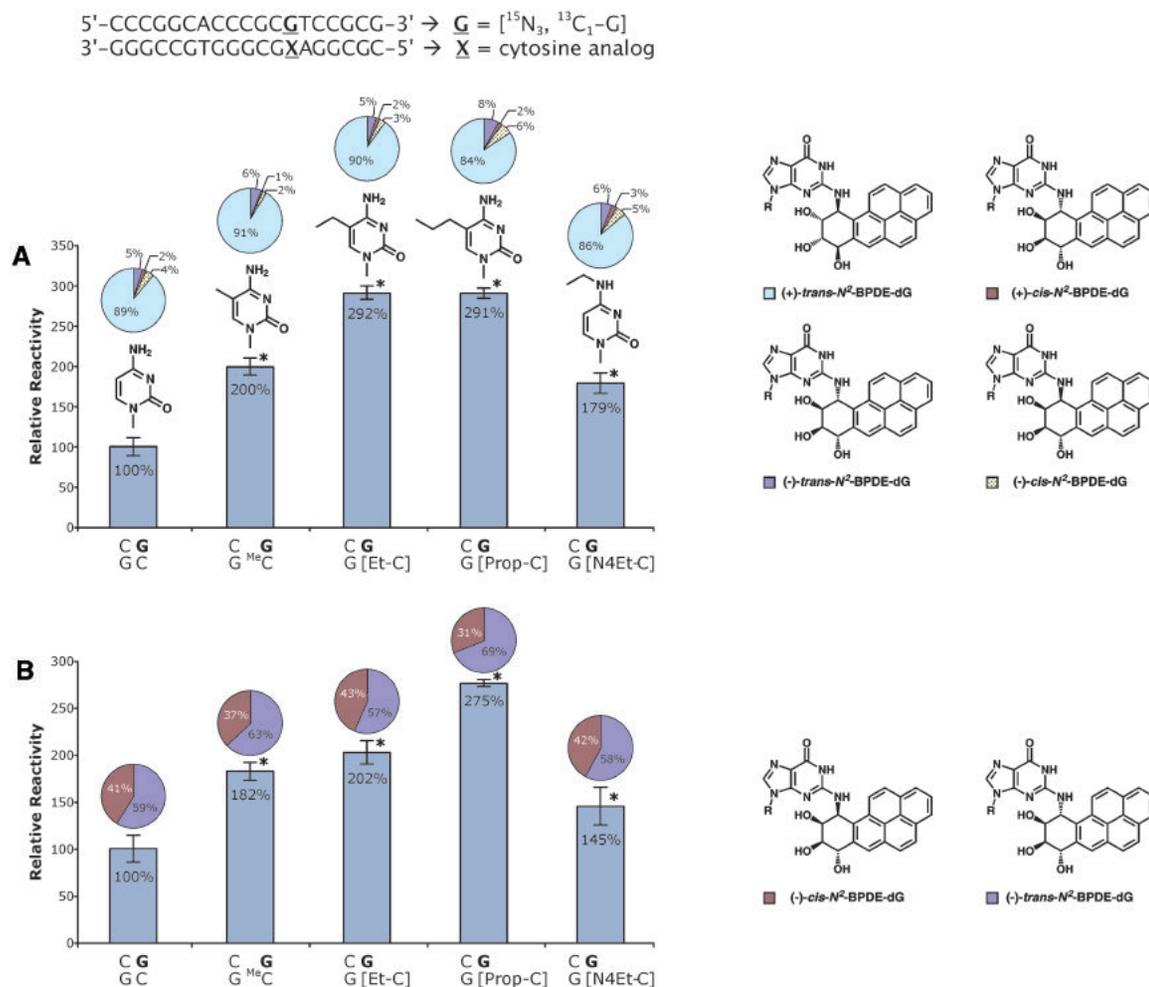
**Figure 2.** Influence of alkyl substituents on cytosine on the yields of $N^2$-BPDE-dG adducts at the base paired guanine. Synthetic DNA duplexes derived from *p53* exon 5 [5′-CCCGGCA CCCGC[$^{15}$N$_3$, $^{13}$C$_1$-G]TCCGCG-3′, (+) strand] containing cytosine, 5-methylcytosine, 5-ethylcytosine, 5-propylcytosine, or N$^4$-ethylcytosine opposite $^{15}$N$_3$, $^{13}$C$_1$-G were treated with ($\pm$)-*anti*-BPDE ($N = 4$) (**A**) or ($-$)-*anti*-BPDE ($N = 6$) (**B**) and the extent of $N^2$-BPDE-dG adduct formation at the isotopically tagged guanine was determined by HPLC-ESI$^+$-MS/MS. Insert: pie charts showing the relative contributions of ($-$)-*trans*-$N^2$-BPDE-dG, (+)-*cis*-$N^2$-BPDE-dG, ($-$)-*cis*-$N^2$- BPDE-dG and (+)-*trans*-$N^2$-BPDE-dG to the total adduct number.

DNA. Note that the fluorescence spectra of BPT alone (curve **d**) and BPT mixed with DNA duplex containing a central G:C base pair (curve **a**) are nearly identical, with the (0,0)-band observed at 376.4 nm (Figure 7A). In contrast, a 1 nm red shift of the (0,0)-band is observed when BPT is combined with $^{Me}$C and dPER containing duplexes (curves **b** and **c** in Figure 7A), suggesting that BPT has an increased affinity towards DNA containing modified base pairs. Identical spectra were obtained for other delay times of the observation window (data not shown), suggestive of a defined association between BPT and structurally modified DNA duplexes.

Fluorescence spectra **a′**, **b′** and **c′** in Figure 7B were recorded with a different excitation wavelength ($\lambda_{ex} = 355$ nm), which is selective for intercalated conformations. As a result, curves **a′**, **b′** and **c′** are characterized by an increased emission at $\lambda = 380$–381 nm (see the dashed arrow and the asterisk in Figure 7B). Note that the spectra obtained in the

presence of $^{Me}$C-containing DNA duplex (curve **b′**) and dPER-containing DNA (curve **c′**) are characterized by a 1–1.2 nm red shift of the (0,0)-band (i.e. 377.4 nm in the $^{Me}$C analog and 377.6 nm in the dPER analog) and the band at 380–381 nm, consistent with increased $\pi$–$\pi$ interactions of BPT with structurally modified duplexes.

Figure 7C shows the difference between emission spectra obtained at $\lambda = 346$ and 355 nm excitation wavelengths. Curves **a\***, **b\*** and **c\*** correspond to BPTs in the presence of DNA duplexes containing unmodified C, $^{Me}$C and dPER, respectively. Spectrum **a\*** = **a′**–**a** reveals that the intercalative complex of BPTs with native DNA has an origin band near 380.1 nm. Curve **b\*** = **b′**–**b** reveals a broader origin band (most likely due to stronger electron–phonon coupling) that is more red-shifted than **a\***, ~381.1 nm, and shows slightly different frequencies of the vibronic bands near 388–394 nm. Spectrum **c\*** obtained for BPT in the presence of dPER-containing DNA is also red shifted with the (0,0)-band near
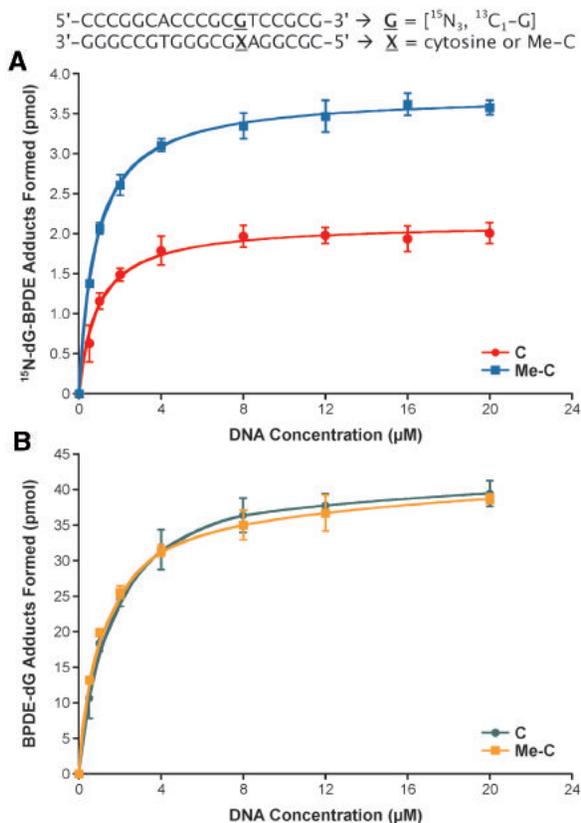
**Figure 3.** Effects of cytosine methylation on the kinetics of formation of $N^2$-BPDE-dG adducts at the base paired guanine (**A**) and at guanine bases elsewhere in the sequence (**B**). Increasing amounts of double-stranded DNA duplexes derived from *p53* exon 5 (5'-CCCGGCACCCGC[$^{15}$N$_3$, $^{13}$C$_1$-G]TCCGCG-3', (+) strand) containing a single 5-methylcytosine opposite $^{15}$N$_3$, $^{13}$C$_1$-G were incubated with ($\pm$)-*anti*-BPDE ($N = 3$) for 1 min and quenched with 2-mercaptoethanol. The extent of $N^2$-BPDE-dG adduct formation at the isotopically tagged guanine ($^{15}$N$_3$, $^{13}$C$_1$-G) and at unlabeled guanines elsewhere in the sequence was quantified by HPLC-ESI$^+$-MS/MS.



**Figure 4.** Influence of C-5 halogen substituents on cytosine on the yields of ($\pm$)-*anti*-BPDE or ($-$)-*anti*-BPDE-induced $N^2$-BPDE-dG adducts at the base paired guanine. Synthetic DNA duplexes derived from *p53* exon 5 (5'-CCCGGCACCCGC[$^{15}$N$_3$, $^{13}$C$_1$-G]TCCGCG-3') containing cytosine, 5-fluorocytosine, 5-chlorocytosine, 5-bromocytosine, or 5-iodocytosine opposite the target G were treated with ($\pm$)-*anti*-BPDE ($N = 4$) (**A**) or ($-$)-*anti*-BPDE ($N = 6$) (**B**) and the extent of $N^2$-BPDE-dG adduct formation at the isotopically tagged guanine was determined by HPLC-ESI$^+$-MS/MS. Insert: pie charts showing relative contributions of ($-$)-*trans*-$N^2$-BPDE-dG, (+)-*cis*-$N^2$-BPDE-dG, ($-$)-*cis*-$N^2$-BPDE-dG and (+)-*trans*-$N^2$-BPDE-dG to the total adduct number. * = significantly different ($P < 0.05$).

381.4 nm. This is suggestive of increased $\pi$–$\pi$ interactions between the pyrenyl residue of BPT and $^{Me}$C:G or dPER:G containing DNA as compared to unmethylated DNA (Figure 7). We conclude that BPT assumes both quasi-external and intercalated geometries in its complexes with DNA. Furthermore, a greater contribution of intercalated structures is observed in the presence of $^{Me}$C and dPER as compared to unmodified duplex. A similar trend is expected for DNA complexes with BPDE and other PAH diolepoxides.

### Computational analysis of BPDE intercalation into duplex DNA containing $^{Me}$C:G base pair

Density functional calculations reveal that BPDE exists in two low-energy conformations (Supplementary Figure S6). In conformer I, the 7,8-dihydroxyl groups of BPDE are in a pseudo-diequatorial orientation, such that the C-10 position of the 9,10-epoxide is open for the *trans* attack by the $N^2$-position of guanine. In conformer II, the 7,8-hydroxyl groups of BPDE are in a pseudo-diaxial conformation, making the C-10 position of BPDE more

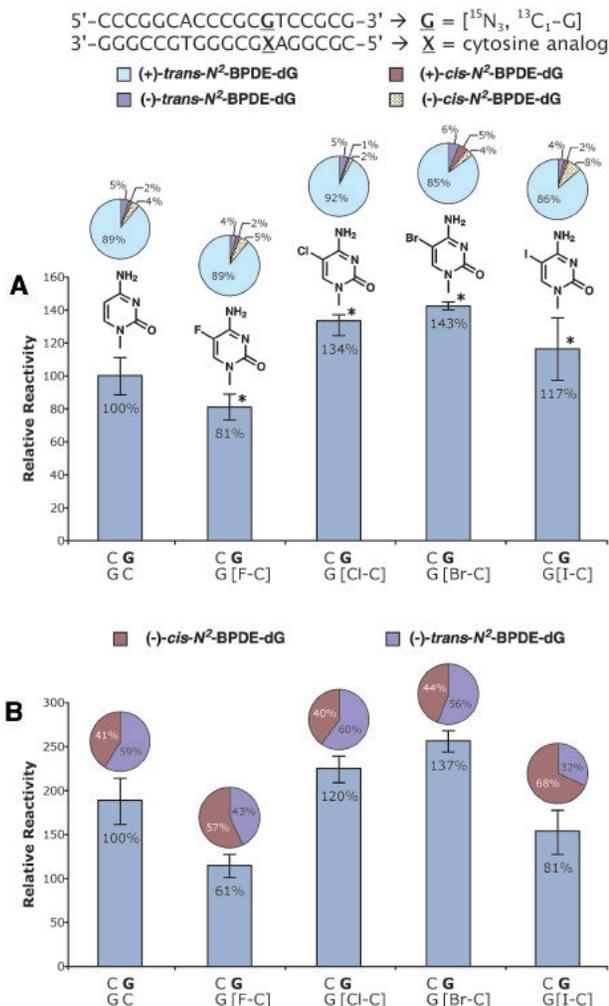sterically hindered and less accessible for nucleophilic attack. Conformer II may instead favor a reaction via an S$_N$1 mechanism, with an intermediate formation of a carbocation triol. Our observation of a 20- to 90-fold excess of (+)-*trans* $N^2$-BPDE-dG adducts as compared to (+)-*cis* products in native DNA (Figure 1, bar graphs in Figure 2) is consistent with a greater stability of conformer I of (+)-*anti*-BPDE as compared to conformer II (2.5 kcal/mol energy difference based on density functional calculations, see panels C and D in Supplementary Figure S6), favoring the formation of the *trans* product. In contrast, the calculated free energies of conformers I and II of ($-$)-*anti*-BPDE are nearly equal (panels A and B in Supplementary Figure S6), consistent with the
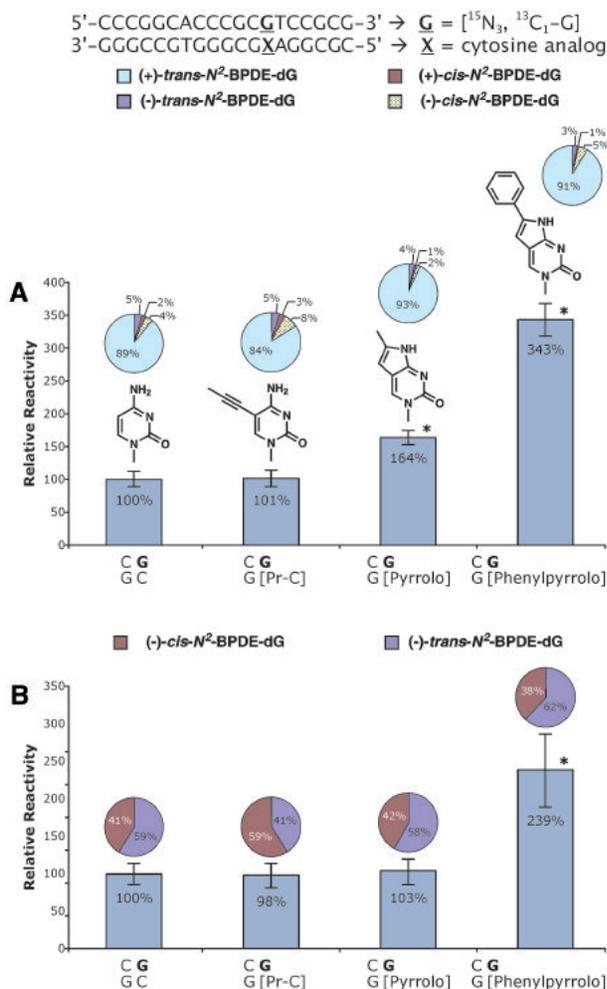
**Figure 5.** Influence of cytosine analogs with extended aromatic systems on the yields of $N^2$-BPDE-dG adducts at the base paired guanine. Double-stranded DNA duplexes derived from *p53* exon 5 (5′-CCCGGCACCCGC[$^{15}$N$_3$, $^{13}$C$_1$-G]TCCGCG-3′) containing cytosine, 5-propynylcytosine (Pr-C), pyrrolocytosine (pyrrolo-C) or 6-phenylpyrrolocytosine (phenylpyrrolo-C) opposite the target G were treated with ($\pm$)-*anti*-BPDE ($N = 4$) (**A**) or ($-$)-*anti*-BPDE ($N = 6$) (**B**) and the extent of $N^2$-BPDE-dG adduct formation at the isotopically tagged guanine was determined by HPLC-ESI$^+$-MS/MS. Insert: pie charts showing relative contributions of ($-$)-*trans*-$N^2$-BPDE-dG, ($+$)-*cis*-$N^2$-BPDE-dG, ($-$)-*cis*-$N^2$-BPDE-dG and ($+$)-*trans*-$N^2$-BPDE-dG isomers. * = significantly different ($P < 0.05$).

**Figure 6.** $N^2$-BPDE-dG adduct formation at guanine bases placed in a DNA duplex opposite unnatural nucleobase analogs unable to form Watson–Crick hydrogen bonds with G. DNA duplexes derived from *p53* exon 5 (5′-CCCGGCACCCGC[$^{15}$N$_3$, $^{13}$C$_1$-G]TCCGCG-3′) containing cytosine (C), difluorotoluene (Dft), abasic site (abasic), or perimidin-2-one nucleoside (dPER) opposite the labeled G were treated with ($\pm$)-*anti*-BPDE ($N = 4$) (**A**) or ($-$)-*anti*-BPDE ($N = 6$) (**B**), and the extent of $N^2$-BPDE-dG adduct formation at the isotopically tagged guanine was determined by HPLC-ESI$^+$-MS/MS. Insert: pie charts showing the relative contributions of ($-$)-*trans*-$N^2$-BPDE-dG, ($+$)-*cis*-$N^2$-BPDE-dG, ($-$)-*cis*-$N^2$- BPDE-dG and ($+$)-*trans*-$N^2$- BPDE-dG isomers. * = significantly different ($P < 0.05$).

observation of similar amounts of ($-$)-*trans* and ($-$)-*cis*-$N^2$-BPDE-dG adducts upon reactions of ($-$)-*anti*-BPDE with DNA (pie chart graphs in Figure 2B).

Our molecular docking experiments were aimed at identifying the structural factors that control the formation and the stereochemical identities of $N^2$-BPDE-dG adducts formed at endogenously methylated C:G base pairs of DNA. Double-stranded oligonucleotides 5′-CCC GCGTCCGC-3′, 3′-GGGCG$^{Me}$CAGGCG-5′ were considered. Open intercalation sites were constructed immediately upstream or downstream from the $^{Me}$C:G base pair using intercalated complexes of a similar structure found in the Protein Databank. A rigid-body docking protocol was performed, followed by molecular dynamics
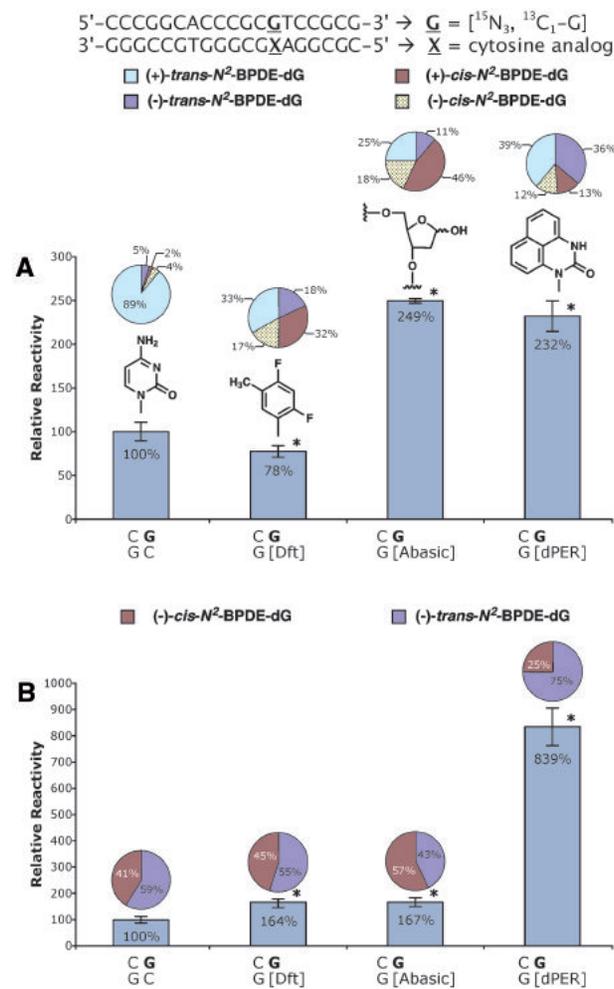
simulated annealing that took into account the full flexibility of the BPDE–DNA complex. All possible combinations of *anti* BPDE stereoisomers [($+$) or ($-$)], diolepoxide conformations (pseudo-diequatorial or pseudo-diaxial) and transition state geometries (*cis* or *trans*) were considered (Supplementary Table S6). BPDE–DNA complexes for which the distance between the exocyclic amino group of G and the C10 of BPDE′s (d1) was $\leq$4.5 Å (which are likely to lead to BPDE adduct formation) were selected for further analysis.

A set of structural indexes was chosen to assess the interactions between BPDE enantiomers and the $^{Me}$C:G site in DNA. First, we considered the formation of hydrogen bonds between the 7-OH and 8-OH of BPDE
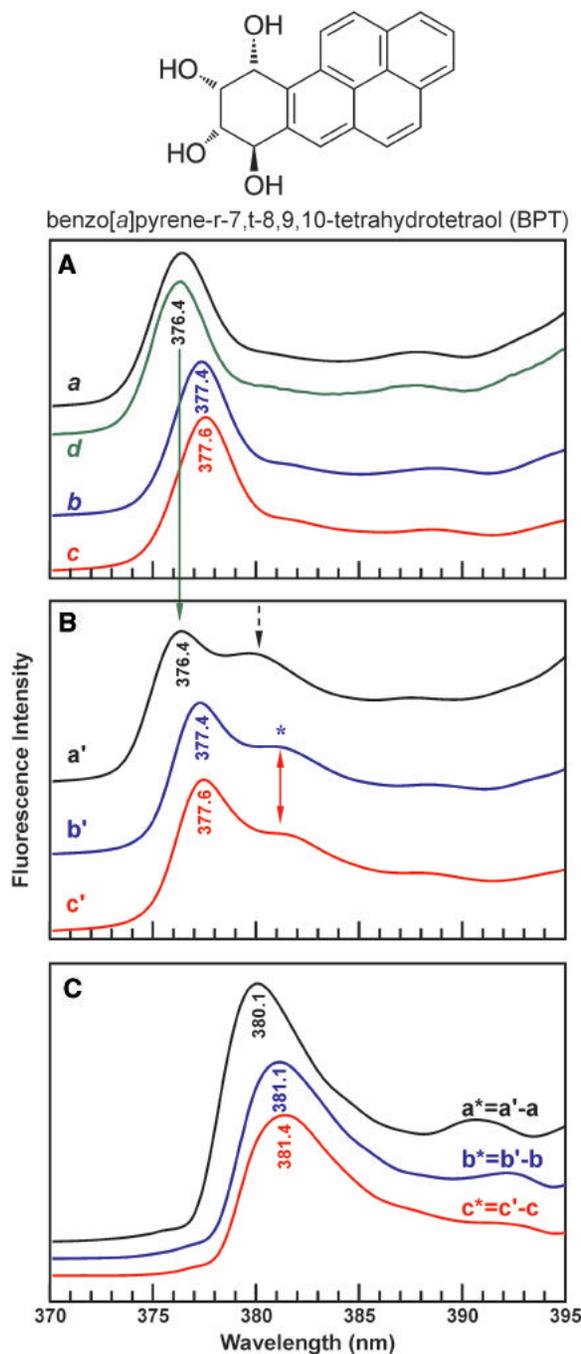
**Figure 7.** Low temperature (77 K) fluorescence origin band spectra of *trans-anti* BPT obtained in the absence and in the presence of DNA duplexes containing a central C:G, $^{Me}$C:G or dPER:G base pair. Spectra in (**A**) were obtained with a $\lambda_{ex}$ of 346 nm and a 60 ns delay time of the observation window immediately after mixing BPT with a 5-fold molar excess of DNA. Curves **a–c** correspond to the fluorescence spectra of BPT mixed with DNA duplexes containing C, $^{Me}$C or dPER, respectively. Spectrum **d** corresponds to BPT alone. Spectra in (**B**) were obtained with $\lambda_{ex}$ of 355 nm and otherwise identical conditions as in (**A**). Curves **a′–c′** correspond to the spectra of BPT mixed with DNA duplexes containing C, $^{Me}$C, or dPER, respectively. (**C**) shows the difference between the emission spectra obtained at 346 and 355 nm (**a\* = a′–a**, **b\* = b′–b**, and **c\* = c′–c**).

and potential hydrogen bond donors/acceptors within the DNA due to their role in defining product stereochemistry. Second, the relative orientations of the 9,10-epoxy group of BPDE and the $N^2$-guanine in DNA were evaluated since the epoxide acts as a leaving group in the nucleophilic substitution reaction leading to BPDE-dG adduct formation. Finally, we took into account the distances between the C10 of BPDE and the exocyclic amino groups of the guanine residues found in the intercalation site (d1 for the guanine base paired with $^{Me}$C and d2 for the 5′-flanking guanine), with d1 value defining the distance between the reactive site of BPDE and its target guanine (Supplementary Table S6).

The proximity of BPDE to the reactive site within the $^{Me}$C:G base pair is dictated by specific hydrogen binding patterns between BPDE and DNA. We found that DNA complexes with the pseudo-diequatorial conformers of BPDE (conformer I in Supplementary Figure S6) are characterized by a closer proximity of the 9,10-epoxide to the exocyclic $N^2$-position of guanine in DNA and more favorable hydrogen bonding interactions of BPDE with the neighboring nucleobases and the phosphodiester backbone of DNA (Supplementary Table S6).

Our docking experiments suggest a possible explanation for the enhanced reactivity of BPDE towards methylated C:G base pairs. In the precovalent BPDE–DNA complex, the pyrenyl group of BPDE is stacked directly above the $^{Me}$C nucleobase, an interaction which is facilitated by the increased hydrophobic and π–π stacking interactions of the methylated cytosine with BPDE (Figure 8A). In this orientation, the epoxide group of BPDE is activated by hydrogen bonding interactions with the 5′-neighboring G (Figure 8A), facilitating *trans* epoxide ring opening by lowering the energy for the transition state for an $S_N2$-type attack. Examination of molecular models predicts that the distance between the reactive centers within guanine and (+)-*anti*-BPDE (d1) in this complex is ∼3.60 Å as compared to 3.84 Å for (−)-*anti*-BPDE, consistent with the predominance of (+)-*trans* adducts observed experimentally (pie charts in Figure 2A).

Alternatively, BPDE can stack with the $^{Me}$C:G base pair in an opposite orientation, with the epoxide ring of BPDE facing the target G and the 7-OH group forming hydrogen bonds with the neighboring C (Figure 8B). In this case, the reaction with the target G can only take place by an $S_N1$-type mechanism, giving rise to *cis* $N^2$-BPDE-dG adducts. The latter reaction is expected to be slower than the $S_N2$ type substitution due to the requirement for carbocation formation and because of a longer distance between the reactive atoms of BPDE and guanine in the reactive complex (d1 ∼3.95 Å for all arrangements that yield *cis* products).

Taken together, our computational results are consistent with the interpretation that the increased reactivity of BPDE towards endogenously methylated CG dinucleotides is due to the facilitated π–π stacking between methylated cytosine and the pyrenyl moiety of BPDE. Furthermore, these molecular models help to explain the observed stereochemistry of $N^2$-BPDE-dG adducts [(+)-*trans* > (−)-*trans* > (+)-*cis* ≈ (−)-*cis*].
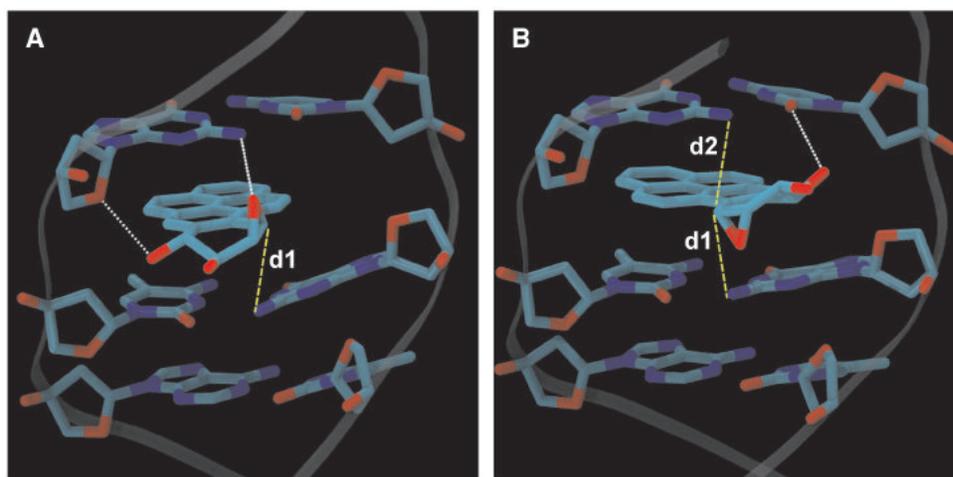
**Figure 8.** Intercalation of the pseudo-diequatorial conformer of (−)-*anti*-BPDE above (**A**) and below the plane of the $^{Me}$C:G base pair (**B**) prior to nucleophilic attack to form a covalent bond with the $N^2$ position of guanine. The sequence that was used in the modeling study was an 11-mer representing the experimental sequence used in the paper: 5′-CCCGC[G]TCCGC-3′/3′-GGGC G[$^{Me}$C]AGGCG-5′ (where the $^{Me}$C:G base pair is indicated by the bracketed residues). Hydrogen bonded heavy atom donor–acceptor distances are indicated by white dotted lines, and the distances between C10 of BPDE epoxide and the $N^2$ exocyclic amine of the reactive guanine (i.e. d1 in Supplementary Table S5) are indicated by a yellow dashed line. Note the formation of a hydrogen bond between the BPDE epoxide oxygen and the exocyclic amino group of the 5′-flanking guanine (A), suggesting that the transition state of the reaction is stabilized by the DNA architecture.

## DISCUSSION

5-Methylcytosine ($^{Me}$C) is among the most important endogenous modifications found in the human genome. The presence of $^{Me}$C in gene promoter regions alters chromatin structure, changes histone acetylation status (34,71) and mediates DNA–protein interactions, allowing for the control of gene expression (71–75). In mammalian cells, $^{Me}$C is produced by enzymatic methylation of the C5 position of cytosine by DNA C-5 cytosine methyltransferases and is found predominantly at CG dinucleotides (76). About 80% of all CG sites are endogenously methylated (72).

Healthy tissues have unique cytosine methylation patterns that are carefully maintained upon DNA replication and cell division (76). However, abnormal cytosine methylation has been observed in aging and in many human diseases, including cancer (77). These epigenetic changes may be caused by endogenous and exogenous electrophiles that target CG dinucleotides in DNA (78–80). For example, endogenously methylated CG dinucleotides found within the *p53* tumor suppressor gene are the sites of the major mutational 'hotposts' in smoking induced lung cancer, suggesting that tobacco carcinogens preferentially modify $^{Me}$CG sequences (81,82). Previous studies revealed that C-5 cytosine methylation increases the reactivity of the $N^2$-position of guanine in C:G base pairs towards mitomycin C, mitoxantrone, esperamicins A1 and C and PAH diol epoxides (22–25,83). In contrast, the presence of $^{Me}$C reduces the yields of $O^6$-alkylguanine adducts induced by alkyldiazonium ions (28) and has little effect on the reactivity of guanine bases towards formaldehyde (30) and aflatoxin B1-*8,9*-epoxide (84).

The main goal of the present investigation was to identify the chemical determinants responsible for $^{Me}$C-mediated effects on guanine adduct formation by carcinogenic PAH diolepoxides. In theory, $^{Me}$C may influence the reactivity of neighboring guanine bases towards electrophiles by modifying DNA structure, mediating pre-covalent binding of carcinogen to DNA, or by affecting the local geometry and electronics of C:G pairs. To distinguish between these possibilities, a series of nucleobase mimics was developed featuring a range of C-5 functionalized cytosines and related structural analogs (Scheme 2), and their effects on the reactivity of guanine introduced in the base paired position were investigated.

We found that the introduction of hydrophobic C-5 alkyl groups on cytosine specifically facilitates the formation of $N^2$-BPDE-dG adducts at its partner guanine (Figures 2 and 3). Furthermore, the relative reactivity of guanine towards BPDE increases as the size of the C-5 alkyl group at the base paired cytosine is increased in the alkyl series (methyl-ethyl-propyl in Figure 2). A smaller reactivity increase was observed for N-4 ethylcytosine:G base pairs which contain an alkyl group at a different position of cytosine.

If C-5 methylation of cytosine activates the guanine within $^{Me}$C:G base pair towards BPDE adduction by an electronic effect transmitted via G:C hydrogen bonds, the introduction of electron withdrawing halogen substituents at the same position should reduce the reactivity of base paired guanine towards BPDE (27). Consistent with previous studies involving mitomycin C (26,27,85), the relative reactivity of guanine base paired to F-C decreases by 20–40% relative to unsubstituted C:G base pair, likely due to the electron withdrawing effects of the fluoro substituent (Figure 4) (26,27,85). In contrast, $N^2$-BPDE-dG adduct formation was enhanced when Cl-C or Br-C were placed opposite guanine in a DNA duplex, while the effect

of I-C was dependent on diolepoxide stereochemistry (Figure 4). This discrepancy may be explained by size differences of the halogen substituents and by the ability of Cl-C, Br-C and I-C, but not F-C, to serve as π-electron donors. Collectively, these results indicate that electronic effects alone cannot explain the effects of C-5 cytosine methylation on reactivity of CG base pairs towards BPDE and other electrophiles.

The most pronounced reactivity enhancement (up to 8-fold) was observed for structural analogs designed to facilitate the intercalation of BPDE into DNA duplexes (pyrrolo-C, phenyl pyrrolo-C and dPER) (Figures 5 and 6). BPDE intercalation between base pairs of the DNA duplex is a required step for covalent adduct formation (86). Intercalative BPDE–DNA interactions also significantly accelerate the hydrolysis of BPDE to the corresponding tetraols by stabilizing the transition state for nucleophilic attack by water (7). Geacintov *et al.* (8) have shown that BPDE prefers to intercalate within oligodeoxynucleotide duplexes of poly(dG-$^{Me}$C). Furthermore, the (−)-*trans* $N^2$-BPDE-dG adducts undergo a shift to an intercalated structure in the presence of neighboring $^{Me}$C (37).

Our non-line narrowing fluorescence studies reveal that trans-*anti* BPT (that was used as a hydrolytically stable analog of BPDE) preferentially forms intercalative complexes with DNA containing $^{Me}$C and dPER as compared to unmodified DNA duplex (Figure 7). These results are supported by fluorescence quenching experiments conducted with benzo[a]pyrene-7S-trans-7,8-dihydrodiol [(+)-BP78D] that was mixed with DNA duplexes containing nucleobase analogs designed to optimize π–π stacking interactions with BPDE (Pr-C, Dft and dPER) (Supplementary Figure S7 and Table S7). Examination of molecular models of BPDE–DNA complexes (Figure 8) suggests that stacking interactions between the pyrene ring system of BPDE and 5-methylcytosine place the epoxy ring of BPDE in a favorable orientation for nucleophilic attack by the amino group of the base paired guanine, leading to increased $N^2$-BPDE-dG adduct yields. One exception is Pr-C:G base pairs which exhibit a strong association with (+)-BP78D (Supplementary Figure S7 and Table S7), but do not display increased reactivity towards BPDE (Figure 5), probably because the presence of C-5 propynyl substituent places BPDE into the major groove of DNA.

In addition to its effects on reactivity of C:G base pairs towards BPDE, $^{Me}$C alters the diastereomeric composition of the resulting DNA adducts. Four lesions are produced upon nucleophilic attack of the exocyclic $N^2$ position of guanine by (±)-*anti* BPDE: (+)-*trans*-$N^2$-BPDE-dG, (−)-*trans*-$N^2$-BPDE-dG, (+)-*cis*-$N^2$-BPDE-dG and (−)-*cis*-$N^2$-BPDE-dG (Scheme 1). While (+)-*trans*-$N^2$-BPDE-dG is typically the major product formed (80–90% of total adducts), minor stereoisomers may play an important role in mutagenesis due to their distinct conformations in double-stranded DNA (87–91). For example, nucleotide excision repair of (+)-*trans* and (−)-*trans* adducts is less efficient than that of (+)-*cis* and (−)-*cis* adducts (87). These differences can be explained by different adduct conformations, with *cis* lesion assuming internal

(base displaced) configurations and *trans* adducts located preferentially in the minor groove of DNA (15,17).

Although fluorescence techniques in combination with PAGE have been previously employed to demonstrate DNA sequence effects on $N^2$-BPDE-dG stereochemistry (92), these methods do not provide quantitative information for $N^2$-BPDE-dG isomers generated at a given guanine. The stable isotope labeling HPLC-ESI$^+$-MS/MS methodology developed in our laboratory (22) makes it possible to quantify the stereoisomeric composition of $N^2$-BPDE-dG adducts originating from specific sites within the DNA duplex (Figure 1, pie charts in Figures 2, 4–6). (±)-*Anti*-BPDE reaction with DNA duplex containing standard G:C base pair yielded 89% (+)-*trans*-$N^2$-BPDE-dG, 5% (−)-*trans*-$N^2$-BPDE-dG, 4% (−)-*cis*-$N^2$-BPDE-dG and 2% (+)-*cis*-$N^2$-BPDE-dG at the target guanine (Figure 1A, pie charts in Figure 2A). When the same duplex was treated with (−)-*anti*-BPDE, 59% (−)-*trans*-$N^2$-BPDE-dG and 41% (−)-*cis*-$N^2$-BPDE-dG was produced at the labeled guanine (pie chart in Figure 2B).

The presence of alkyl or halogen substituents at the C-5 of cytosine did not change the stereochemical composition of $N^2$-BPDE-dG adducts produced at structurally modified G:C base pairs (pie charts in Figures 2 and 4), suggesting that these modifications do not alter the transition state geometry for $N^2$-BPDE-dG adduct formation. The introduction of base analogs with extended aromatic systems further increased the relative yields of the (+)-*trans* product (pie charts in Figure 5), probably by stabilizing the precovalent complex of modified C:G base pairs with the pseudo-diequatorial conformer I of (+)-*anti*-BPDE (see above).

More pronounced differences in the stereochemistry of $N^2$-BPDE-dG adducts were observed for modified base pairs unable to participate in standard Watson–Crick hydrogen bonds, e.g. those containing Dft, the abasic site and dPER (pie charts in Figure 6). Dft is a nonpolar nucleoside mimic that lacks hydrogen bonding functionalities, but preserves DNA structure *via* strong stacking interactions with neighboring bases in the DNA duplex (40). Our fluorescence studies reveal an increased association of (+)-BP78D (a hydrolytically stable analog of (−)-*anti*-BPDE) with Dft-containing DNA (Supplementary Table S7). These results suggest that the presence of Dft leads to the preferential intercalation of (−)-*anti*-BPDE rather than (+)-*anti*-BPDE, thereby decreasing the relative contributions of (+)-*trans* and (+)-*cis* $N^2$-BPDE-dG adducts at modified base pairs. The presence of an abasic site opposite the target guanine increases the contribution of *cis* $N^2$-BPDE-dG adducts, suggesting a preference for reaction through the $S_N1$-type mechanism (pie charts in Figure 6). We hypothesize that the geometry of the pre-covalent BPDE–DNA complex at the G-AP site pair differs from that of the standard G:C base pair due to a spacial void associated with the introduction of an abasic site opposite target guanine. The presence of dPER completely changes the stereochemical outcome of the BPDE reaction with guanine in the opposite strand (Figures 1D and 6),

producing large numbers of (−)-*trans* adducts which are rare at standard G:C base pairs (Figures 2 and 6).

Taken together, our results make it possible to propose a likely model of how BPDE–DNA interactions are modified by 5-methylcytosine, an important endogenous base modification that controls gene expression (75) and governs the formation of smoking-induced lung cancer mutational hotspots in the *p53* tumor suppressor gene (24). The majority of cytosine analogs considered in this study displayed an increased reactivity towards BPDE, which was associated with a facilitated physical complex formation between BPDE and DNA. Density functional calculations reveal that the electronic properties of the modified C:G base pair are affected by the C-5 cytosine substituent, but that these changes alone do not predict reactivity (66). We conclude that C-5 methylation on cytosine increases the yields of $N^2$-BPDE-dG lesions at the base paired guanine mostly by facilitating the formation of pre-covalent intercalative complexes with BPDE. Additionally, the presence of $^{Me}$C and unnatural nucleobase analogs mediates the conformation of the BPDE molecule in its complex with DNA, leading to changes in stereochemistry of the resulting $N^2$-BPDE-dG adducts.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

## FUNDING

*Conflict of interest statement*. None declared.

## REFERENCES

1. Intenational Agency for Research on Cancer (IARC). (1983) Polynuclear aromatic compounds, part I, chemical, environmental, and experimental data. In: IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. IARC, Lyon, France, pp. 33–451.
2. Pfeifer,G.P., Denissenko,M.F., Olivier,M., Tretyakova,N., Hecht,S.S. and Hainaut,P. (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*, **21**, 7435–7451.
3. Yang,S.K., Roller,P.P. and Gelboin,H.V. (1977) Enzymatic mechanism of benzo[a]pyrene conversion to phenols and diols and an improved high-pressure liquid chromatographic separation of benzo[a]pyrene derivatives. *Biochemistry*, **16**, 3680–3687.
4. Shimada,T. and Fujii-Kuriyama,Y. (2004) Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. *Cancer Sci.*, **95**, 1–6.
5. Mazerska,Z. (2006) Metabolism of chemical carcinogens. In Baer-Dubowska,W., Bartoszek,A. and Malejka-Giganti,D. (eds), *Carcinogenic and Anticarcinogenic Food Components*. CRC Press, Boca Raton, FL, pp. 37–67.
6. Geacintov,N.E., Yoshida,H., Ibanez,V., Jacobs,S.A. and Harvey,R.G. (1984) Conformations of adducts and kinetics of binding to DNA of the optically pure enantiomers of anti-benzo(a)pyrene diol epoxide. *Biochem. Biophys. Res. Comm.*, **122**, 33–39.
7. Geacintov,N.E., Yoshida,H., Ibanez,V. and Harvey,R.G. (1982) Noncovalent binding of 7 beta, 8 alpha-dihydroxy-9 alpha, 10 alpha-epoxytetrahydrobenzo[a]pyrene to deoxyribonucleic acid and its catalytic effect on the hydrolysis of the diol epoxide to tetrol. *Biochemistry*, **21**, 1864–1869.
8. Geacintov,N.E., Shahbaz,M., Ibanez,V., Moussaoui,K. and Harvey,R.G. (1988) Base-sequence dependence of noncovalent complex formation and reactivity of benzo[a]pyrene diol epoxide with polynucleotides. *Biochemistry*, **27**, 8380–8387.
9. Meehan,T., Gamper,H. and Becker,J.F. (1982) Characterization of reversible, physical binding of benzo[a]pyrene derivatives to DNA. *J. Bio. Chem.*, **257**, 10479–10485.
10. Jeffrey,A.M., Weinstein,I.B., Jennette,K.W., Grzeskowiak,K., Nakanishi,K., Harvey,R.G., Autrup,H. and Harris,C. (1977) Structures of benzo[a]pyrene-nucleic acid adducts formed in human and bovine bronchial explants. *Nature*, **269**, 348–350.
11. Wolfe,A.R., Yamamoto,J. and Meehan,T. (1994) Chloride ions catalyze the formation of cis adducts in the binding of anti-benzo[a]pyrene diol epoxide to nucleic acids. *Proc. Natl Acad. Sci. USA*, **91**, 1371–1375.
12. Meehan,T., Wolfe,A.R., Negrete,G.R. and Song,Q. (1997) Benzo[a]pyrene diol epoxide-DNA cis adduct formation through a trans chlorohydrin intermediate. *Proc. Natl Acad. Sci. USA*, **94**, 1749–1754.
13. Shukla,R., Jelinsky,S., Liu,T., Geacintov,N.E. and Loechler,E.L. (1997) How stereochemistry affects mutagenesis by N2-deoxyguanosine adducts of 7,8-dihydroxy-9, 10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene: configuration of the adduct bond is more important than those of the hydroxyl groups. *Biochemistry*, **36**, 13263–13269.
14. Suh,M., Ariese,F., Small,G.J., Jankowiak,R., Liu,T.M. and Geacintov,N.E. (1995) Conformational studies of the (+)-*trans*, (−)-*trans*, (+)-*cis*, and (-)-*cis* adducts of *anti*-benzo[a]pyrene diolepoxide to $N^2$-dG in duplex oligonucleotides using polyacrylamide gel electrophoresis and low-temperature fluorescence spectroscopy. *Biophys. Chem.*, **56**, 281–296.
15. Cosman,M., de los Santos,C., Fiala,R., Hingerty,B.E., Singh,S.B., Ibanez,V., Margulis,L.A., Live,D., Geacintov,N.E. and Broyde,S. (1992) Solution conformation of the major adduct between the

carcinogen (+)-*anti*-benzo[a]pyrene diol epoxide and DNA. *Proc. Natl Acad. Sci. USA*, **89**, 1914–1918.

16. de los Santos,C., Cosman,M., Hingerty,B.E., Ibanez,V., Margulis,L.A., Geacintov,N.E., Broyde,S. and Patel,D.J. (1992) Influence of benzo[a]pyrene diol epoxide chirality on solution conformations of DNA covalent adducts: the (-)-*trans-anti*-[BP]G.C adduct structure and comparison with the (+)-*trans-anti*-[BP]G.C enantiomer. *Biochemistry*, **31**, 5245–5252.

17. Cosman,M., Hingerty,B.E., Luneva,N., Amin,S., Geacintov,N.E., Broyde,S. and Patel,D.J. (1996) Solution conformation of the (-)-*cis-anti*-benzo[a]pyrenyl-dG adduct opposite dC in a DNA duplex: intercalation of the covalently attached BP ring into the helix with base displacement of the modified deoxyguanosine into the major groove. *Biochemistry*, **35**, 9850–9863.

18. Jiang,G., Jankowiak,R., Grubor,N., Banasiewicz,M., Small,G.J., Skorvaga,M., Van Houten,B. and States,J.C. (2004) Supercoiled DNA promotes formation of intercalated cis-$N^2$-deoxyguanine adducts and base-stacked *trans*-$N^2$-deoxyguanine adducts by (+)-7R,8S-dihydrodiol-9S,10R-epoxy-7,8,9,10-tetra-hydrobenzo[a]pyrene. *Chem. Res. Toxicol.*, **17**, 330–339.

19. Rodriguez,F.A., Cai,Y., Lin,C., Tang,Y., Kolbanovskiy,A., Amin,S., Patel,D.J., Broyde,S. and Geacintov,N.E. (2007) Exocyclic amino groups of flanking guanines govern sequence-dependent adduct conformations and local structural distortions for minor groove-aligned benzo[a]pyrenyl-guanine lesions in a GG mutation hotspot context. *Nucleic Acids Res.*, **35**, 1555–1568.

20. Riggs,A.D. and Jones,P.A. (1983) 5-methylcytosine, gene regulation, and cancer. *Adv.Cancer Res.*, **40**, 1–30.

21. Tornaletti,S. and Pfeifer,G.P. (1995) Complete and tissue-independent methylation of CpG sites in the *p53* gene: implications for mutations in human cancers. *Oncogene*, **10**, 1493–1499.

22. Matter,B., Wang,G., Jones,R. and Tretyakova,N. (2004) Formation of diastereomeric benzo[a]pyrene diol epoxide-guanine adducts in *p53* gene-derived DNA sequences. *Chem. Res. Toxicol.*, **17**, 731–741.

23. Denissenko,M.F., Pao,A., Tang,M. and Pfeifer,G.P. (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in *p53*. *Science*, **274**, 430–432.

24. Denissenko,M.F., Chen,J.X., Tang,M.S. and Pfeifer,G.P. (1997) Cytosine methylation determines hot spots of DNA damage in the human *p53* gene. *Proc. Natl Acad. Sci. USA*, **94**, 3893–3898.

25. Tretyakova,N., Matter,B., Jones,R. and Shallop,A. (2002) Formation of benzo[a]pyrene diol epoxide-DNA adducts at specific guanines within *K-ras* and *p53* gene sequences: stable isotope-labeling mass spectrometry approach. *Biochemistry*, **41**, 9535–9544.

26. Das,A., Tang,K.S., Gopalakrishnan,S., Waring,M.J. and Tomasz,M. (1999) Reactivity of guanine at m5CpG steps in DNA: evidence for electronic effects transmitted through the base pairs. *Chem. Biol.*, **6**, 461–471.

27. Dannenberg,J.J. and Tomasz,M. (2000) Hydrogen-bond acid/base catalysis: a density functional theory study of protonated guanine-(substituted) cytosine base pairs as models for nucleophilic attack on mitomycin in DNA. *J. Am. Chem. Soc.*, **122**, 2062–2068.

28. Ziegel,R., Shallop,A., Upadhyaya,P., Jones,R. and Tretyakova,N. (2004) Endogenous 5-methylcytosine protects neighboring guanines from N7 and $O^6$-methylation and $O^6$-pyridyloxobutylation by the tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Biochemistry*, **43**, 540–549.

29. Rajesh,M., Wang,G., Jones,R. and Tretyakova,N. (2005) Stable isotope labeling-mass spectrometry analysis of methyl- and pyridyloxobutyl-guanine adducts of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone in *p53*-derived DNA sequences. *Biochemistry*, **44**, 2197–2207.

30. Matter,B., Guza,R., Zhao,J., Li,Z.z., Jones,R. and Tretyakova,N. (2007) Sequence distribution of acetaldehyde-derived N2-Ethyl-dG adducts along duplex DNA. *Chem. Res. Toxicol.*, **20**, 1379–1387.

31. Rauch,C., Trieb,M., Wellenzohn,B., Loferer,M., Voegele,A., Wibowo,F.R. and Liedl,K.R. (2003) C5-methylation of cytosine in B-DNA thermodynamically and kinetically stabilizes BI. *J. Am. Chem. Soc.*, **125**, 14990–14991.

32. Norberg,J. and Vihinen,M. (2001) Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides. *J. Mol. Struct.*, **546**, 51–62.

33. Hodges-Garcia,Y. and Hagerman,P.J. (1992) Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry*, **31**, 7595–7599.

34. Hausheer,F.H., Rao,S.N., Gamcsik,M.P., Kollman,P.A., Colvin,O.M., Saxe,J.D., Nelkin,B.D., McLennan,I.J., Barnett,G. and Baylin,S.B. (1989) Computational analysis of structural and energetic consequences of DNA methylation. *Carcinogenesis*, **10**, 1131–1137.

35. Sowers,L.C., Shaw,B.R. and Sedwick,W.D. (1987) Base stacking and molecular polarizability: effect of a methyl group in the 5-position of pyrimidines. *Biochem. Biophys. Res. Commun.*, **148**, 790–794.

36. Geacintov,N.E., Cosman,M., Hingerty,B.E., Amin,S., Broyde,S. and Patel,D.J. (1997) NMR solution structures of stereoisometric covalent polycyclic aromatic carcinogen-DNA adduct: principles, patterns, and diversity. *Chem. Res. Toxicol.*, **10**, 111–146.

37. Huang,X., Colgate,K.C., Kolbanovskiy,A., Amin,S. and Geacintov,N.E. (2002) Conformational changes of a benzo[a]pyrene diol epoxide-$N^2$-dG adduct induced by a 5'-flanking 5-methyl-substituted cytosine in a $^{Me}$CG double-stranded oligonucleotide sequence context. *Chem. Res. Toxicol.*, **15**, 438–444.

38. Zhang,N., Lin,C., Huang,X., Kolbanovskiy,A., Hingerty,B.E., Amin,S., Broyde,S., Geacintov,N.E. and Patel,D.J. (2005) Methylation of cytosine at C5 in a CpG sequence context causes a conformational switch of a benzo[a]pyrene diol epoxide-$N^2$-guanine adduct in DNA from a minor groove alignment to intercalation with base displacement. *J. Mol. Biol.*, **346**, 951–965.

39. Kang,J.I. Jr, Burdzy,A., Liu,P. and Sowers,L.C. (2004) Synthesis and characterization of oligonucleotides containing 5-chlorocytosine. *Chem. Res. Toxicol.*, **17**, 1236–1244.

40. Kool,E.T. and Sintim,H.O. (2006) The difluorotoluene debate–a decade later. *Chem. Commun.*, **35**, 3665–3675.

41. Wojciechowski,F. and Hudson,R.H. (2008) Fluorescence and hybridization properties of peptide nucleic acid containing a substituted phenylpyrrolocytosine designed to engage Guanine with an additional H-bond. *J. Am. Chem. Soc.*, **130**, 12574–12575.

42. Gong,J. and Sturla,S.J. (2007) A synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA. *J. Am. Chem. Soc.*, **129**, 4882–4883.

43. Misra,B. and Amin,S. (1990) An improved synthesis of anti-benzo[c]phenanthrene-3,4-diol 1,2-epoxide via 4-methoxybenzo[c]phenanthrene. *J. Org. Chem.*, **55**, 4478–4480.

44. Krzeminski,J., Lin,J.M., Amin,S. and Hecht,S.S. (1994) Synthesis of Fjord region diol epoxides as potential ultimate carcinogens of dibenzo[a,l]pyrene. *Chem. Res. Toxicol.*, **7**, 125–129.

45. Sharma,A.K., Kumar,S. and Amin,S. (2004) A highly abbreviated synthesis of dibenzo[def,p]chrysene and its 12-methoxy derivative, a key precursor for the synthesis of the proximate and ultimate carcinogens of dibenzo[def,p]chrysene. *J. Org. Chem.*, **69**, 3979–3982.

46. Amin,S., Camanzo,J., Huie,K. and Hecht,S.S. (1984) Improved photochemical synthesis of 5-methylchrysene derivatives and its application to the preparation of 7,8-dihydro-7,8-dihydroxy-5-methylchrysene. *J. Organ. Chem.*, **49**, 381–384.

47. Gait,M.J. (1984) *Oligonucleotide Synthesis: A Practical Approach.* IRL Press, Washington, DC.

48. Hainaut,P. and Pfeifer,G.P. (2001) Patterns of p53 G–>T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis*, **22**, 367–374.

49. Shallop,A.J., Gaffney,B.L. and Jones,R.A. (2003) Use of $^{13}$C as an indirect tag in $^{15}$N specifically labeled nucleosides. Syntheses of [8-$^{13}$C-1,7,NH$_2$-$^{15}$N$_3$]adenosine, -guanosine, and their deoxy analogues. *J. Org. Chem.*, **68**, 8657–8661.

50. Zhao,H., Pagano,A.R., Wang,W., Shallop,A., Gaffney,B. and Jones,R.A. (1997) Use of a $^{13}$C atom to differentiate two $^{15}$N-labeled nucleosides. Syntheses of [$^{15}$NH$_2$]-Adenosine, [1,7,NH$_2$-$^{15}$N$_2$]- and [2-$^{13}$C,1,NH$_2$-$^{15}$N$_2$]-Guanosine, and [1,7,NH$_2$-$^{15}$N$_3$]- and [2-$^{13}$C,1,7,NH$_2$-$^{15}$N$_3$]-2'-Deoxyguanosine. *J. Org. Chem.*, **62**, 7832–7835.

51. Robins,M.J. and Barr,P.J. (1983) Nucleic acid related compounds. 39. Efficient conversion of 5-iodo to 5-alkynyl and derived 5-substituted uracil bases and nucleosides. *J. Org. Chem.*, **48**, 1854–1862.

52. Hudson,R., Damoiseaux,R. and Viirre,R.D. (2004) Fluorescent 7-deazapurine derivatives from 5-Iodocytosine via a tandem cross-coupling-annulation reaction with terminal alkynes. *Synlett*, **13**, 2400–2402.

53. Tretyakova,N., Matter,B., Ogdie,A., Wishnok,J.S. and Tannenbaum,S.R. (2001) Locating nucleobase lesions within DNA sequences by MALDI-TOF mass spectral analysis of exonuclease ladders. *Chem. Res. Toxicol.*, **14**, 1058–1070.

54. Miksa,B., Chinnappan,R., Dang,N.C., Reppert,M., Matter,B., Tretyakova,N., Grubor,N.M. and Jankowiak,R. (2007) Spectral differentiation and immunoaffinity capillary electrophoresis separation of enantiomeric benzo(a)pyrene diol epoxide-derived DNA adducts. *Chem. Res. Toxicol.*, **20**, 1192–1199.

55. Giese,T.J., Gregersen,B.A., Liu,Y., Nam,K., Mayaan,E., Moser,A., Range,K., Faza,O.N., Lopez,C.S., de Lera,A.R. *et al.* (2006) QCRNA 1.0: a database of quantum calculations for RNA catalysis. *J. Mol. Graphics Modell.*, **25**, 423–433.

56. Lipscomb,L.A., Peek,M.E., Zhou,F.X., Bertrand,J.A., VanDerveer,D. and Williams,L.D. (1994) Water ring structure at DNA interfaces: hydration and dynamics of DNA-anthracycline complexes. *Biochemistry*, **33**, 3649–3659.

57. Gilquin,B., Roumestand,C., Zinn-Justin,S., Ménez,A. and Toma,F. (1993) Refined three-dimensional solution structure of a snake cardiotoxin: analysis of the side-chain organization suggests the existence of a possible phospholipid binding site. *Biopolymers*, **33**, 1659–1675.

58. Roe,D.R., Okur,A., Wickstrom,L., Hornak,V. and Simmerling,C. (2007) Secondary structure bias in generalized born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *J. Phys. Chem. B.*, **111**, 1846–1857.

59. Wang,J., Cieplak,P. and Kollman,P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.

60. Aduri,R., Psciuk,B.T., Saro,P., Taniga,H., Schlegel,H.B. and SantaLucia,J. (2007) AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.*, **3**, 1464–1475.

61. Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. III, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the dscription of $\alpha/\gamma$ conformers. *Biophys. J.*, **92**, 3817–3829.

62. Morris,G.M., Goodsell,D.S., Halliday,R.S., Huey,R., Hart,W.E., Belew,R.K. and Olson,A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **19**, 1639–1662.

63. Frederick,C.A., Saal,D., van der Marel,G.A., van Boom,J.H., Wang,A.H. and Rich,A. (1987) The crystal structure of d(GGm 5CCGGCC): the effect of methylation on A- DNA structure and stability. *Biopolymers*, **26(Suppl.)**, S145–S160.

64. Ahlborn,C., Siegmund,K. and Richert,C. (2007) Isostable DNA. *J. Am. Chem. Soc.*, **129**, 15218–15232.

65. Sagi,J., Guliaev,A.B. and Singer,B. (2001) 15-mer DNA duplexes containing an abasic site are thermodynamically more stable with adjacent purines than with pyrimidines. *Biochemistry*, **40**, 3859–3868.

66. Moser,A., Guza,R., Tretyakova,N. and York,D.M. (2009) Density functional study of the influence of C5 cytosine substitution in base pairs with guanine. *Theor. Chem. Acc.*, **122**, 179–188.

67. Nakajima,A. (1971) Solvent effect on the vibrational structures of the fluorescence and absorption spectra of pyrene. *Bull. Chem. Soc. Japan*, **44**, 3277.

68. Kalyanasundaram,K. and Thomas,J.K. (1977) Environmental effects on vibronic band intensities in pyrene monomer fluorescence and their application in studies of micellar systems. *J. Am. Chem. Soc.*, **99**, 2039–2044.

69. Jankowiak,R., Lu,P.Q., Small,G.J. and Geacintov,N.E. (1990) Laser spectroscopic studies of DNA adduct structure types from enantiomeric diol epoxides of benzo[a]pyrene. *Chem. Res. Toxicol.*, **3**, 39–46.

70. Kim,S.K., Geacintov,N.E., Brenner,H.C. and Harvey,R.G. (1989) Identification of conformationally different binding sites in benzo[a]pyrene diol epoxide-DNA adducts by low-temperature fluorescence spectroscopy. *Carcinogenesis*, **10**, 1333–1335.

71. Newell-Price,J., Clark,A.J. and King,P. (2000) DNA methylation and silencing of gene expression. *Trends Endocrinol. Metab.*, **11**, 142–148.

72. Ehrlich,M., Gama-Sosa,M.A., Huang,L.H., Midgett,R.M., Kuo,K.C., McCune,R.A. and Gehrke,C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.*, **10**, 2709–2721.

73. Millar,D.S., Holliday,R. and Grigg,G.W. (2003) Five not four: history and significance of the fifth base. In Beck,S. and Olek,A. (eds), *The Epigenome, Molecular Hide and Seek*. Wiley, New York, pp. 3–20.

74. Widschwendter,M. (2007) 5-methylcytosine-the fifth base of DNA: the fifth wheel on a car or a highly promising diagnostic and therapeutic target in cancer? *Dis. Markers*, **23**, 1–3.

75. Bird,A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.

76. Cedar,H. and Razin,A. (1990) DNA methylation and development. *Biochim. Biophys. Acta.*, **1049**, 1–8.

77. Millar,D.S., Holliday,R. and Grigg,G.W. (2003) Five not four: history and significance of the fifth base. In Beck,S. and Olek,A. (eds), *The Epigenome: Molecular Hide and Seek*. Wiley-VCH, Weinheim, pp. 3–38.

78. Valinluck,V., Liu,P., Kang,J.I. Jr, Burdzy,A. and Sowers,L.C. (2005) 5-halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2). *Nucleic Acids Res.*, **33**, 3057–3064.

79. Valinluck,V., Tsai,H.H., Rogstad,D.K., Burdzy,A., Bird,A. and Sowers,L.C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.*, **32**, 4100–4108.

80. Baskunov,V.B., Subach,F.V., Kolbanovskiy,A., Kolbanovskiy,M., Eremin,S.A., Johnson,F., Bonala,R., Geacintov,N.E. and Gromova,E.S. (2005) Effects of benzo[a]pyrene-deoxyguanosine lesions on DNA methylation catalyzed by EcoRII DNA methyltransferase and on DNA cleavage effected by EcoRII restriction endonuclease. *Biochemistry*, **44**, 1054–1066.

81. Pfeifer,G.P., Tang,M. and Denissenko,M.F. (2000) Mutation hotspots and DNA methylation. *Curr. Top. Microbiol. Immunol.*, **249**, 1–19.

82. Pfeifer,G.P. (2000) *p53* mutational spectra and the role of methylated CpG sequences. *Mutat. Res.*, **450**, 155–166.

83. Mathur,P., Xu,J. and Dedon,P.C. (1997) Cytosine methylation enhances DNA damage produced by groove binding and intercalating enediynes: studies with esperamicins A1 and C. *Biochemistry*, **36**, 14868–14873.

84. Ross,M.K., Mathison,B.H., Said,B. and Shank,R.C. (1999) 5-Methylcytosine in CpG sites and the reactivity of nearest neighboring guanines toward the carcinogen aflatoxin B1-8,9-epoxide. *Biochem. Biophys. Res. Commun.*, **254**, 114–119.

85. Tomasz,M. and Arunangshu,D. Opposite effects of methylated and fluorinated cytosine on the reactivity of CpG with benzo[a]pyrene diolepoxide (BPDE) in oligonucleotides. *Book of Abstracts, 219th ACS National Meeting*. San Francisco, California, March 26–30, 2000.

86. Becker,J.F., Strunk,S.J., Martinez,K.G., Waltman,T. and Meehan,T. (1986) Decay-associated emission spectra and other spectral evidence for the physical intercalation of 7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene into DNA. *Cancer Biochem. Biophys.*, **9**, 67–73.

87. Hess,M.T., Gunz,D., Luneva,N., Geacintov,N.E. and Naegeli,H. (1997) Base pair conformation-dependent excision of benzo[a]pyrene diol epoxide-guanine adducts by human nucleotide excision repair enzymes. *Mol. Cell. Biol.*, **17**, 7069–7076.

88. Alekseyev,Y.O. and Romano,L.J. (2000) In vitro replication of primer-templates containing benzo[a]pyrene adducts by exonuclease-deficient Escherichia coli DNA polymerase I

(Klenow fragment): effect of sequence context on lesion bypass. *Biochemistry*, **39**, 10431–10438.

89. Zhang,Y., Yuan,F., Wu,X., Rechkoblit,O., Taylor,J.S., Geacintov,N.E. and Wang,Z. (2000) Error-prone lesion bypass by human DNA polymerase η. *Nucleic Acids Res.*, **28**, 4717–4724.

90. Huang,X., Kolbanovskiy,A., Wu,X., Zhang,Y., Wang,Z., Zhuang,P., Amin,S. and Geacintov,N.E. (2003) Effects of base sequence context on translesion synthesis past a bulky (+)-*trans-anti*-B[*a*]P-$N^2$-dG lesion catalyzed by the Y-family polymerase pol κ. *Biochemistry*, **42**, 2456–2466.

91. Zhao,B., Wang,J., Geacintov,N.E. and Wang,Z. (2006) Polη, Polζ and Rev1 together are required for G to T transversion mutations induced by the (+)- and (-)-trans-*anti*-BPDE-$N^2$-dG DNA adducts in yeast cells. *Nucleic Acids Res.*, **34**, 417–425.

92. Marsch,G.A., Jankowiak,R., Suh,M. and Small,G.J. (1994) Sequence dependence of benzo[*a*]pyrene diol epoxide-DNA adduct conformer distribution: a study by laser-induced fluorescence/ polyacrylamide gel electrophoresis. *Chem. Res. Toxicol.*, **7**, 98–109.