Base Sequence Effects in Bending Induced by Bulky Carcinogen−DNA Adducts: Experimental and Computational Analysis†

Qian Ruan,‡ Ping Zhuang,‡ Sheng Li,‡ Rebecca Perlow,§ A. R. Srinivasan,|| Xiang-Jun Lu,|| Suse Broyde,§ Wilma K. Olson,|| and Nicholas E. Geacintov*,‡

Chemistry and Biology Departments, New York University, New York, New York 10003, and Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903

Received November 16, 2000; Revised Manuscript Received May 29, 2001

ABSTRACT: The covalent binding of bulky mutagenic or carcinogenic compounds to DNA can lead to bending, which could significantly alter the interactions of DNA with critical replication and transcription proteins. The impact of adducts derived from the highly reactive bay region enantiomeric (+)- and (−)-*anti*-7,8-diol-9,10-epoxide derivatives of benzo[*a*]pyrene (BPDE) are of interest because the (+)-7*R*,8*S*,9*S*,-10*R*-*anti*-BPDE enantiomer is highly tumorigenic in rodents, while the (−)-7*S*,8*R*,9*R*,10*S*-*anti*-BPDE enantiomer is not. Both (+)- and (−)-*anti*-BPDE bind covalently with DNA predominantly by *trans* addition at the exocyclic amino group of guanine to yield 10*S* (+)- and 10*R* (−)-*trans*-*anti*-[BP]-N^2-dG adducts. We have synthesized a number of different oligonucleotides with single (+)- and (−)-*trans*-*anti*-[BP]-N^2-dG adducts (G*) in the base sequence context XG*Y, where X and Y are different DNA bases. The G* residues were positioned at or close to the center of 11 base pair (∼1 helical turn) or 16 base pair (∼1.5 turns) duplexes. All bases, except for X and Y and their partners, were identical. These sequences were self-ligated with T4 ligase to form multimers that yield a ladder of bands upon electrophoresis in native polyacrylamide gels. The extent of bending in each oligonucleotide was assessed by monitoring the decrease in gel mobilities of these linear, self-ligated oligomers, relative to unmodified oligonucleotides of the same base sequence. The extent of global bending was then estimated using a sequence-specific three-dimensional model from which the values of the base-pair step parameter roll adjacent to the lesion site could be extracted. We find that (+)-*trans*-*anti*-[BP]-N^2-dG adducts are considerably more bent than the (−) isomers regardless of sequence and that A-T base pairs flanking the [BP]-N^2-dG lesion site allow for local flexibility consistent with adduct conformational heterogeneity. Interestingly, the fit of computed versus observed gel mobilities using classical reptation treatments requires enhancement of unmodified DNA flexibility in gels, compared to aqueous salt solution. The differences in bending between the two stereoisomeric adduct duplexes and the observed base sequence context effects may play a significant role in the differential processing of these lesions by cellular replication, transcription, and repair enzymes.

The binding of proteins to DNA is of essential importance in gene regulation and in nucleosome packaging and can be accompanied by marked deviations of the normal double helical DNA structure (*1*−*5*). The covalent binding of bulky mutagenic or carcinogenic compounds to DNA can also lead to DNA distortions, including bending, and numerous examples of such phenomena have been reported (*6*−*9*). If not excised by repair enzymes, these DNA adducts could significantly alter the interactions of DNA with critical replication and transcription proteins.

We have been interested in the impact of adducts derived from the binding of polycyclic aromatic carcinogen metabolites on double-stranded DNA. The highly reactive bay region enantiomeric (+)- and (−)-*anti*-7,8-diol-9,10-epoxide derivatives of benzo[*a*]pyrene (BPDE) bind predominantly to DNA at the exocyclic amino group of guanine by *cis* or *trans* addition of N^2-dG to the C10 position of BPDE (*10*, *11*). These systems are of particular importance because the (+)-7*R*,8*S*,9*S*,10*R* *anti*-BPDE enantiomer is highly tumorigenic in rodents, while the (−)-7*S*,8*R*,9*R*,10*S* enantiomer is not (*12*, *13*). Furthermore, in gel electrophoresis mobility assays, site-specific oligonucleotide duplexes with single, site-specific 10*S* (+)-*trans*-*anti*-[BP]-N^2-dG adducts[1] are significantly bent, while those with the stereoisomeric 10*R* (−)-*trans*-*anti*-[BP]-N^2-dG adducts (Figure 1) are not (*14*, *15*). Also, the 11-mer oligonucleotides with (+)-*trans* adducts

† This research is supported by NIH Grants CA-28038 and CA-75449 and DOE Grant DE-FG60290ER60931 to S.B., NIH Grant CA-76660 to N.E.G., and NIH Grant GM-20861 to W.K.O. Computational resources were provided to S.B. from the NSF San Diego Supercomputer Center and the DOE National Energy Research Supercomputer Center.

* To whom correspondence should be addressed. Phone: (212) 998-8407. Fax: (212) 998-8421. E-mail: nicholas.geacintov@nyu.edu.

‡ Chemistry Department, New York University.

§ Biology Department, New York University.

|| Department of Chemistry, Rutgers University.

[1] Abbreviations: *anti*-BPDE, 7*r*,8*t*-dihydroxy-*t*9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (+)-*anti*-BPDE, (+)-7*R*,8*S*-dihydroxy-9*S*,10*R*-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (−)-*anti*-BPDE, (−)-7*S*,8*R*-dihydroxy-9*R*,10*S*-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (+)- or (−)-*trans*-*anti*-[BP]-N^2-dG adducts, covalent products derived from the binding of (+)- or (−)-*anti*-BPDE at its C10 position to the N^2-2′-deoxyguanosine residues in DNA by *trans* addition; bp, base pair.

10.1021/bi002643x CCC: $20.00
Published on Web 08/11/2001

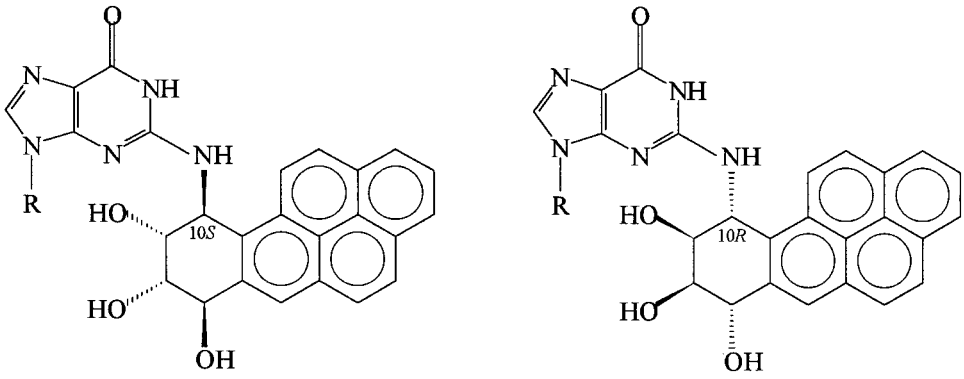

FIGURE 1: Chemical structures of stereoisomeric *anti*-[BP]-N^2-dG adducts.

Table 1: BPDE-Modified DNA Duplex Sequences Investigated

Designation	Sequence
TG*C-11	**5′-CACATG*CACAC**
	TGTA C GTGTGG-5′
TG*C-16	**5′-CTCACATG*CACACTCT**
	GTGTA C GTGTGAGAGA-5′
CG*T-11	**5′-CACACG*TACAC**
	TGTG C ATGTGG-5′
CG*T-16	**5′-CTCACACG*TACACTCT**
	GTGTG C ATGTGAGAGA-5′
AG*C-11	**5′-CACAAG*CACAC**
	TGTT C GTGTGG-5′
AG*C-16	**5′-CTCACAAG*CACACTCT**
	GTGTT C GTGTGAGAGA-5′
CG*A-11	**5′-CACACG*AACAC**
	TGTG C TTGTGG-5′
CG*A-16	**5′-CTCACACG*AACACTCT**
	GTGTG C TTGTGAGAGA-5′

readily form small minicircles upon ligation, while those with (−)-*trans* adducts form circles with much lower efficiencies and also of greater size (*16*). From high-resolution NMR studies, it is known that, in both adducts, the bulky pyrenyl ring systems are positioned predominantly in the minor groove in normal, full duplexes (*17*−*19*). However, in the 10*S* (+)-*trans* adduct, this BPDE residue is oriented toward the 5′-end of the modified strand (*17*, *18*), while in the case of the 10*R* (−)-*trans* adduct, it points in the opposite 3′-direction (*19*). The induced bends are sequence-dependent, and the differences in DNA bending exhibited by the (+)-*trans* and the (−)-*trans* adducts have been rationalized in terms of the differences in adduct conformation (*20*).

It is useful to relate the observed carcinogen-induced bending effects to changes in the base-pair step parameter, roll, since roll predominantly governs DNA bending (*2*, *21*, *22*). Since it is known that this parameter depends on base sequence (*23*), it is important to investigate experimentally how the bases flanking both sides of the BPDE-modified guanines affect DNA bending and roll. In this work, we have synthesized a number of different oligonucleotides with single (+)-*trans-anti*- and (−)-*trans-anti*-[BP]-N^2-dG adducts (G*) in the base sequence context XG*Y, where X and Y are different DNA bases, as shown in Table 1. We focus on the effects of different pyrimidines flanking the modified G*. Flanking pyrimidines are of special importance because pyrimidine/purine steps, especially TG·CA dinucleotide steps, are particularly susceptible to bending via roll (*2*, *3*, *24*, *25*). The modified guanine residues G* were positioned at or close to the center of two types of oligonucleotides, 11 base pairs (∼1 helical turn) or 16 base pairs (∼1.5 turns) long. All bases except for X and Y, and their partners, were identical in each set of oligonucleotides (Table 1). The extent of bending of each oligonucleotide was assessed by monitoring the decrease in the mobilities of these linear, self-ligated DNA oligomers of different lengths, relative to the unmodified oligonucleotides of the same base sequence (*20*). The extent of global bending was then computed using a sequence-specific model from which the values of roll at the carcinogen-modified base-pair step could be extracted.

METHODS

Preparation of Modified Oligonucleotides and Gel Electrophoresis. The oligodeoxyribonucleotide sequences studied in this work are shown in Table 1. The helical repeat of the 11-mer sequences is estimated from the sequence-dependent twist angles of Kabsch et al. (*26*) to be ∼10.5 base pairs per helical turn; thus the sequences 11 and 16 base pairs long comprise approximately 1 and 1.5 helical turns, respectively.

All oligonucleotides were synthesized using the Biosearch Cyclone DNA synthesizer and were purified by standard HPLC protocols. The BPDE-modified oligonucleotides were generated by a direct synthesis method using racemic *anti*-BPDE obtained from the National Cancer Institute Carcinogen Reference Repository. The procedures used and the methods of characterization and verification of adduct stereochemistry were the same as those described earlier (*27*). Forward reaction 5′-end labeling of both unmodified and modified oligonucleotides with [γ-^{32}P]ATP (New England Nuclear) and cold ATP (Pharmacia) was employed as described previously (*20*). About 2 μg of each unmodified and BPDE-modified oligonucleotide was 5′-end labeled with γ-^{32}P (New England Nuclear) using 10 units of T4 polynucleotide kinase (Gibco BRL). After labeling with γ-^{32}P, 2 μL of 0.1 M cold ATP (Pharmacia) and 10 units of T4 polynucleotide kinase were added, and the reaction was continued at 37 °C for an additional hour. The complementary oligonucleotides were cold labeled at the 5′-end. The labeled unmodified and BPDE-modified single-stranded oligonucleotides were repurified using denaturing 20% polyacrylamide gel electrophoresis (7 M urea). The *Bam*H1

Table 2: Average Values and Standard Deviations of Sequence-Dependent Base Pair Step Parameters[a]

step	tilt $\tau_0 \pm \Delta\tau$ (deg)	roll $\rho_0 \pm \Delta\rho$ (deg)	twist $\Omega_0 \pm \Delta\Omega$ (deg)	shift $Dx_0 \pm \Delta Dx$ (Å)	slide $Dy_0 \pm \Delta Dy$ (Å)	rise $Dz_0 \pm \Delta Dz$ (Å)
CG	0.0 ± 4.2	5.4 ± 5.2	36.1 ± 5.5	0.00 ± 0.87	0.41 ± 0.56	3.39 ± 0.27
CA·TG[b]	0.5 ± 3.7	4.7 ± 5.1	37.3 ± 6.5	0.09 ± 0.55	0.53 ± 0.89	3.33 ± 0.26
TA	0.0 ± 2.7	3.3 ± 6.6	37.8 ± 5.5	0.00 ± 0.52	0.05 ± 0.71	3.42 ± 0.24
AG·CT[b]	-1.7 ± 3.3	4.5 ± 3.4	31.9 ± 4.5	0.09 ± 0.69	-0.25 ± 0.41	3.34 ± 0.23
GG·CC[b]	-0.1 ± 3.7	3.6 ± 4.5	32.9 ± 5.2	0.05 ± 0.76	-0.22 ± 0.64	3.42 ± 0.24
AA·TT[b]	-1.4 ± 3.3	0.7 ± 5.4	35.1 ± 3.9	-0.03 ± 0.57	-0.08 ± 0.45	3.27 ± 0.22
GA·TC[b]	-1.5 ± 3.8	1.9 ± 5.3	36.3 ± 4.4	-0.28 ± 0.46	0.09 ± 0.70	3.37 ± 0.26
AT	0.0 ± 2.5	1.1 ± 4.9	29.3 ± 4.5	0.00 ± 0.57	-0.59 ± 0.31	3.31 ± 0.21
AC·GT[b]	-0.1 ± 3.1	0.7 ± 3.9	31.5 ± 4.2	0.13 ± 0.59	-0.58 ± 0.41	3.36 ± 0.23
GC	0.0 ± 3.9	0.3 ± 4.6	33.6 ± 4.7	0.00 ± 0.61	-0.38 ± 0.56	3.40 ± 0.24

[a] Data for the unperturbed protein-bound DNA dimers are given in ref *3*. The protein-bound DNAs were judged to be more representative of the conformations found in carcinogen-liganded DNA than DNA that is entirely unperturbed. Fluctuations of individual dimer steps are measured in terms of the elastic energy, $E_{elastic} = (kT/2)[((\tau - \tau_0)/\xi\Delta\tau)^2 + ((\rho - \rho_0)/\xi\Delta\rho)^2 + ((\Omega - \Omega_0)/\xi\Delta\Omega)^2 + ((Dx - Dx_0)/\xi\Delta Dx)^2 + ((Dy - Dy_0)/\xi\Delta Dy)^2 + ((Dz - Dz_0)/\xi\Delta Dz)^2]$, where ξ is a deformability parameter used in fitting the computed mean-square end-to-end distance to the observed gel mobilities of unmodified sequences, ρ is a specific value of roll generated in the Monte Carlo sampling procedure, ρ_0 is the intrinsic (equilibrium rest state) value, and $\Delta\rho$ is the standard deviation. Analogous definitions apply to the other base-pair step parameters. [b] Mean values of tilt and shift are negated in the designated complementary dimers.

10-mer linker 5′-d(CGGGATCCCG) (New England Biolabs), used as a size standard, was 5′-end labeled by T4 polynucleotide kinase in a phosphate-exchange reaction. All samples were desalted using chromatography cartridges (Bio-Rad). The end-labeled oligonucleotides were annealed with 2.5-fold excess complementary strands designed to produce cohesive single-stranded ends (Table 1), which ensures that the oligonucleotides are in duplex forms during the ligation reaction. About 0.5 μg of the annealed duplexes was incubated in 45 μL of 1 × ligation buffer (Gibco BRL) and 9 units of T4 ligase (Gibco BRL) at 12 °C for about 24 h. The ligated multimers were subjected to electrophoresis on nondenaturing 8% polyacrylamide gels at 4 °C. The relative intensities of each of the bands were analyzed using a Bio-Rad GS-525 imaging system (Bio-Rad).

Computations. (a) Modeling of Carcinogen-Induced DNA Bending. The experimental studies yield relative distances of electrophoretic migration of different DNA oligomers in a fixed time interval. Our goal is to produce structural models of specific carcinogen-modified duplexes of different lengths whose computed mobilities match the observed values. Theories of the motion of DNA in electrophoretic gels (*28*, *29*) permit computation of the mean-square end-to-end distance of DNA from the measured mobility and, conversely, estimation of mobility if the mean-square end-to-end distance is available. We have implemented an algorithm developed by Babcock et al. (*30*) which generates coordinates of a double helical structure from the six local dimer "step" parameters—tilt, roll, twist, shift, slide, and rise—that specify the orientation and displacement of adjacent base pairs (*31*). The mean-square end-to-end distance of a given DNA sequence is computed from sets of step parameters generated in Monte Carlo sampling of the allowed distortions of the constituent dimers. The sampling procedure employs the sequence-dependent step parameters extracted by Olson et al. (*3*) from high-resolution protein-bound DNA structures stored in the Nucleic Acid Database (*32*). The fluctuations of unmodified dimers are constrained by a simple elastic energy model based on the observed structural data (Table 2), and the step parameters of accepted states are accumulated in matrices developed by Marky and Olson (*33*) for the computation of DNA end-to-end dimensions (see Appendix for further details).

Step parameters at the modification site are varied to fit the observed gel mobilities. Knowing that ligand-induced DNA bending occurs almost exclusively via roll (*2*), we focus our analysis on the changes in roll at the carcinogen-bound dimer steps that can account for the experimental migration distances (see, however, the Results and Discussion sections for the influence of other step parameters and deformation sites on computed gel mobilities). We generate ensembles of dimer structures for both (+)- and (−)-*trans-anti*-[BP]-N^2-dG adducts in different sequence contexts and incorporate their dynamic variation, using the methods of Marky and Olson (*33*), in multimers of differing lengths.

(b) Relation between End-to-End Distance and DNA Gel Mobility. A number of factors influence the mobility of a DNA duplex in an electrophoretic gel. However, the motions of a duplex through the interstices of the gel in the presence of an electric field are not well understood. The gel mobility models of Lerman and Frisch (*28*) and Lumpkin and Zimm (*29*), in which the DNA molecule moves by reptation in the fashion of a wormlike chain, make a number of simplifying assumptions: (1) the mobility of the DNA molecule is independent of its length; (2) the radii of the pores of the gel are much smaller than the root-mean-square radius of the DNA strand; and (3) the electric field strength, E, which determines the DNA mobility, is sufficiently small so that the orientations of the DNA molecules about the electric field vector remain random. According to these models, the mobility μ, defined as the velocity of reptation (dx/dt) divided by E, is proportional to the quotient of $\langle R^2 \rangle$, the mean-square end-to-end distance, and the inverse square of N, the number of base pairs:

$$\mu = \frac{1}{E}\frac{dx}{dt} \propto \frac{\langle R^2 \rangle}{N^2} \quad (1)$$

The mobilities of our flexible chain models are estimated with eq 1, neglecting the proportionality coefficient and assuming that the average velocity does not change as a function of time. The distance migrated, d, of the DNA in

the gel is then given by

$$d = a\mu + b \quad (2)$$

where d is the distance migrated, μ is the mobility calculated from eq 1, and a and b are constants for any particular gel, obtained by regression analysis (see below).

The goodness of fit of the distances migrated by a particular multimer series, calculated from eq 2 versus the experimental data, is evaluated by their linear correlation coefficient r. The slower the mobility, the smaller the value of the end-to-end distance. In turn, the average value of the end-to-end distance decreases as the degree of bending or flexibility of duplexes is increased. Different intrinsic flexibilities (see below) are considered, and the best fit for each electrophoretic gel is obtained, i.e., the deformability parameter where r is closest to unity, the value indicative of a perfect match between experiment and theory.

(c) Calculations for Unmodified DNA. The deformability parameter ξ is a factor by which the standard deviations in base-pair step parameters (Table 2) are multiplied. This product gives an effective range of flexibility consistent with the gel measurements and the simplifying assumptions of our model. To treat carcinogen-modified DNA, we first find the value of ξ that best reproduces the experimental mobilities of the unmodified B-DNA duplexes of the same multimer series. We compare the set $\{d\}$ of experimental distances migrated by chains with a given repeating sequence and the computed mobilties $\{\mu\}$ of DNA models of the corresponding sequence and chain length. Typical values for the deformability parameter lie in the range of 2.5−3.6, with optimum r values of 0.991 or better (see below). In practice, ξ is varied at increments of 0.01 over the above range, with individual dimer steps sampling 100 000 low-energy configurations per evaluation of $\langle R^2 \rangle$. The value of ξ yielding the highest correlation coefficient is used to deduce the values of a and b required for the analysis of gel mobilities of the corresponding series of carcinogen-modified DNAs.

Notably, the enhancement of dimer flexibility required to match the distances migrated by the DNA through a nondenaturing polyacrylamide gel significantly underestimates the known DNA persistence length in aqueous salt solution. Apparent thermal fluctuations in the gel are roughly 3-fold larger than those in solution. A comparable enhancement of chain bending was also required by Levene and Zimm (*34*) to fit their tube model to the anomalous mobilities of naturally curved kinetoplast DNA on polyacrylamide gels. The "tight" spacing of such gels, where the average pore size is roughly an order of magnitude smaller than the normal DNA persistence length (*35*, *36*), may contribute to the apparent flexibility. Alternatively, interpretation of the DNA movement through such gels might require a more realistic model of the ease of threading individual molecules through a complex maze (*37*).

(d) Calculations for BPDE-Modified DNA. For simplicity, we assume that the covalent binding of a carcinogen perturbs the intrinsic base-pair step geometry but has no effect on dimer flexibility. Indeed, base-pair step parameters show comparable variations at modified and unmodified steps in our all-atom molecular dynamics simulations of a neutralized and solvated (+)-*trans-anti*-[BP]-N^2-dG modified duplex (see Figure 10 and Supplementary Information, Table S1, for numerical values and discussion below for computational details). We thus use the value of ξ deduced for unmodified DNA strands of the same sequence to sample the configurations of modified chains. We consider a set of perturbed step parameters for the dinucleotide step(s) containing the modified base. In most cases, we limit the modification to a change in intrinsic roll at a single dimer step: the XG* step that is 5′ to G* in the (+)-*trans* adducts and the G*Y step that is 3′ to G* in the (−)-*trans* adducts, based respectively on the observed 5′- and 3′-orientations of the BP moiety in the minor groove of model oligomers (*17*−*19*). [Note, however, that CG* twist, a parameter known to be tightly coupled to roll in protein-bound DNA (*3*), is the most distorted step variable in the short, atomic-level simulation of the carcinogen-modified oligonucleotide duplex in Figure 10.] For each value of roll considered in the polymer-level simulations of BPDE-damaged DNA, we generate a sample of 100 000 low-energy dimer configurations. This set of perturbed dimer structures is then incorporated, as outlined above, into the computation of the mean-square end-to-end distance, with all other dimer steps of the DNA adopting the sets of configurations used to account for the gel patterns of the corresponding unmodified strands. The distance migrated by the modified DNA, which carries the modeled dimer lesion in each chain repeating fragment, is obtained from eq 2, with μ computed according to eq 1 and a and b taken from the regression analysis of gel profiles of the corresponding unmodified sequences. The sampling of intrinsic roll is repeated until the correlation coefficient of computed and experimental migration distances is maximized.

(e) Molecular Dynamics Simulation. Molecular dynamics simulations of the 9 bp (+)-*trans-anti*-[BP]-N^2-dG modified duplex, 5′-d(CCGCG*TCCG)-3′•5′-d(CGGACGCGG)-3′, were carried out using the SANDER module of AMBER 5.0 (*38*) and the Cornell et al. (*39*) force field, including the updated Parm98.dat parameter set (*40*). The particle mesh Ewald approximation (*41*) was used to assess long-range Coulombic interactions. Periodic boundary conditions were applied, and all dynamics runs were performed at constant pressure (NPT).

Parametrization. (a) Partial Charges. Partial charges for the (+)-*trans-anti*-[BP]-N^2-dG residue were determined for 12 different minimum energy conformations (Perlow and Broyde, unpublished data) in order to prevent conformational bias. The partial charges of atoms in the modified guanine were obtained from Hartree−Fock calculations using the 6-31G* basis set in GAUSSIAN 94 (*42*), fitted to atomic centers using the RESP module of AMBER 5.0, and then averaged and normalized to avoid charge imbalance (*43*). The final partial charges with AMBER atom type assignments are listed in Table S1.

(b) Bond Angle and Improper Torsional Parameters. The bond angle and improper torsional parameters of the modified guanine residue that were missing from the AMBER force field (Table S2) were assigned by analogy to the parameters of chemically related fragments in the Parm98.dat parameter set (*40*).

Preparation of Starting Structure. The starting structure of the modified duplex was based on the NMR solution structure of the 11 bp (+)-*trans-anti*-[BP]-N^2-dG duplex, d(CCATCG*CTACC)•d(GTAGCGATGG) (*17*). The sequence was truncated and changed to that of the requisite

9-mer, and the conformational energy was minimized using the DUPLEX molecular mechanics program (*44*) with the distance-dependent dielectric function of Hingerty et al. (*45*) in the Coulombic energy term. This minimized structure was then neutralized with 16 Na^+ ions using the LEaP module of AMBER 5.0 (*38*) and solvated in 3168 TIP3P (*46*) water molecules. The positions of the waters and ions were optimized using the steepest descent method (5000 steps) with a dielectric constant of unity.

Molecular Dynamics Protocol. To relax the positions of waters and ions and bring the system to a realistic density (i.e., ~1 g/cm^3), 25 ps of molecular dynamics was conducted at 10 K with the atoms of the solute fixed by harmonic restraints of 25 kcal/mol. The system was next heated to 300 K uniformly over 15 ps and then held at 300 K for 10 ps, again holding the solute fixed, to equilibrate the waters and ions. The remainder of the dynamics simulation was carried out at 300 K. The harmonic restraints on the solute were released slowly over 50 ps by running successively 5 ps of dynamics with 10.0 kcal/mol restraints, 20 ps with 1.0 kcal/mol restraints, and 25 ps with 0.1 kcal/mol restraints on the solute atoms. Once the restraints on the solute were removed, unrestrained molecular dynamics production runs were carried out for a total of 2 ns. Structures from the full dynamics run were used to assess the relative structures and deformability of modified and unmodified dimer steps and the likely positions of the benzo[*a*]pyrenyl moiety with respect to the modified G*•C base pair.

Conformational Analysis. Base-pair step parameters of the simulated duplex were computed with 3DNA (Lu and Olson, in preparation) at 0.1 ns increments over the full dynamics run. Mean values and standard deviations for each dimer were compared against the corresponding, sequence-independent parameters from high-resolution B-DNA crystal structures, the latter values obtained by equal weighting of the averages and dispersions of step parameters in the 16 common dimers (*3*). The comparable variation of parameters at modified and unmodified CG steps (Figure 10) underlies the assumption of equivalent dimer flexibility in the simulations of normal and carcinogen-damaged multimers.

RESULTS

Experimental Mobilities. A typical autoradiograph of a native electrophoresis gel of the ligation products of the BPDE-modified CG*T and unmodified CGT 11-mers is shown in Figure 2. The bands in lanes 1, 2, and 3 correspond respectively to unligated standard duplexes of the 10-mer *Bam*H1 linker, the unmodified CGT 11-mer, and the (+)-*trans* CG*T 11-mer adduct. Lanes 4 and 7 show the ladders of the *Bam*H1 ligation products. The ligation ladders of unmodified CGT and (+)-*trans*-CG*T are shown in lanes 5 and 6, respectively. Analogous results were obtained with all of the other sequences shown in Table 1 (Supporting Information). Several ligation experiments were also performed with (−)-*trans* adducts in selected sequences (Supporting Information).

While all of the BPDE-modified oligonucleotide fragments studied here migrate more slowly than the unmodified oligonucleotides of the same sequence lengths, the DNA molecules with (+)-*trans*-*anti*-[BP]-N^2-dG lesions exhibit the slowest electrophoretic mobilities of all of the sequences

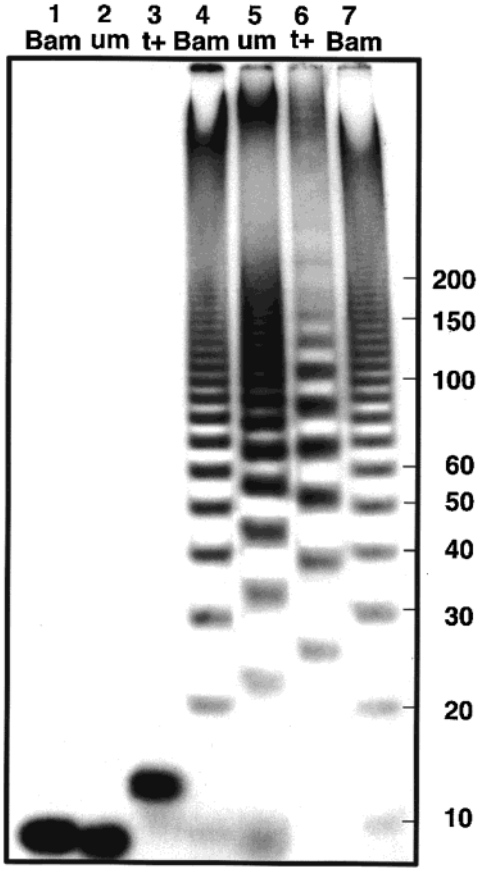

FIGURE 2: Ligation ladders of unmodified CGT-11 duplexes (lane 5), BPDE-modified CG*T-11 duplexes with 10*S* (+)-*trans* adduct stereochemistry (lane 6), and 10-mer *Bam*H1 linkers (lanes 4 and 7). The single bands in lane 1, 2, and 3 are unligated standard duplexes of the linker, unmodified 11-mer, and (+)-*trans* adduct, respectively. The ladder (numbers on the right margin of the figure) refers to the positions of the linker multimers (lane 7) of the designated sizes expressed in numbers of base pairs.

studied. All of the oligonucleotide multimers containing BPDE residues migrate more slowly than the unmodified oligonucleotide multimers with the same number of base pairs and thus are characterized by longer apparent sequence lengths. In all cases, the electrophoretic mobilities of sequences with (−)-*trans* adducts were slower than those of the identical unmodified oligomers. However, consistent with previous observations (*20*), the impact of the *anti*-[BP]-N^2-dG adducts on the lowering of the electrophoretic mobilities was much smaller in the case of the (−)-*trans* than the (+)-*trans* lesions.

The unusual relatively slow mobilities of the (+)-*trans*-oligonucleotide adducts are revealed in plots of R_L versus the number of base pairs in each multimer, where R_L is defined (*47*, *48*) as

$$R_L = \frac{\text{(apparent length of multimer)}}{\text{(sequence length of multimer)}} \qquad (3)$$

The apparent length of a BPDE-modified multimer is obtained from the migration distances of this molecule and the sequence length of an unmodified multimer with the same migration distance as described in, for example, Tsao et al. (*20*). The BPDE-modified multimers are bent and migrate more slowly than the unmodified multimers of the same

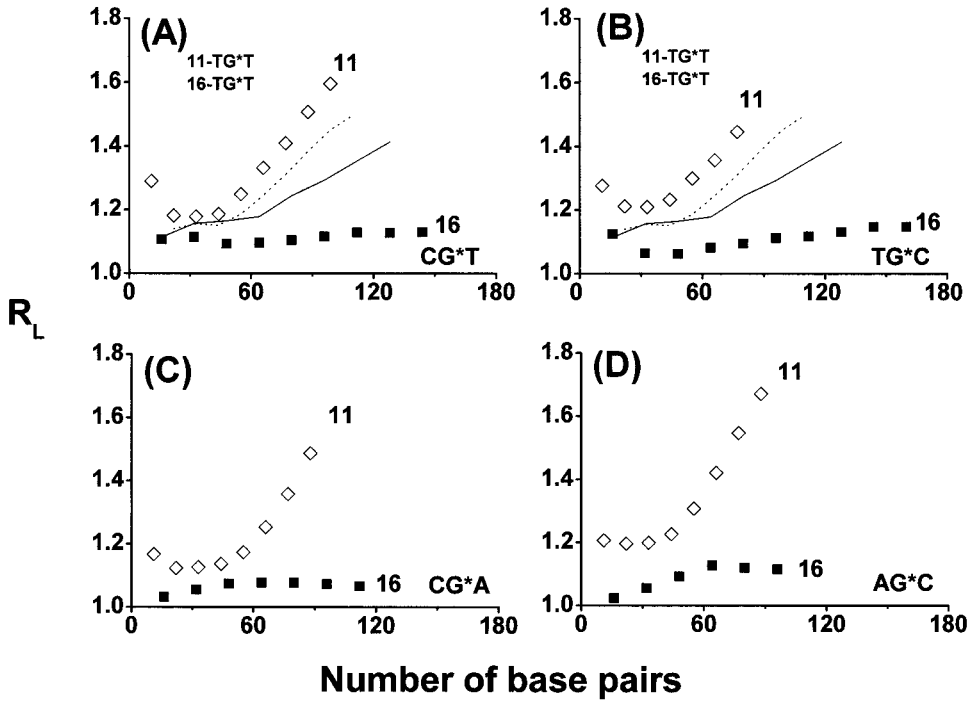

FIGURE 3: Plots of R_L for duplexes with 10*S* (+)-*trans-anti*-[BP]-N^2-dG lesions as a function of the number of base pairs per multimer. (A) -CGT- sequences (◇, CG*T-11, and ■, CG*T-16). (B) -TGC- sequences (◇, TG*C-11, and ■, TG*C-16). In both (A) and (B), the -TGT- sequences are inserted as lines (dotted line for TG*T-11 and solid line for TG*T-16). (C) -CGA- sequences (◇, CG*A-11, and ■, CG*A-16). (D) -AGC- sequences (◇, AG*C-11, and ■, AG*C-16).

sequence length, and thus $R_L > 1.0$. Plots of R_L vs the number of base pairs per molecule for CG*T-11 and CG*T-16, TG*C-11 and TG*C-16, CG*A-11 and CG*A-16, and AG*C-11 and AG*C-16 oligonucleotides with (+)-*trans-anti*-[BP]-N^2-dG lesions are shown in panels A−D of Figure 3, respectively. Previously published data for the TG*T-11 and TG*T-16 sequences (dotted and solid lines), obtained under similar conditions and published earlier (*20*), are also shown for comparison. In all of the cases of 11-mer (+)-*trans* adducts, the R_L values increase to 1.5−1.7 as the size of the oligomers increases. This increase in R_L with increasing number of base pairs, *N*, results from the cumulative effect of the bends in each 11-mer segment of the oligomer (*49*). In the case of the ligated 16-mers, with ∼1.5 helical turns per repeating sequence, a marked phasing effect is observed since R_L does not exceed values of ∼1.1 and does not show the same marked increase with increasing *N* as in the case of the ligated 11-mers. The only exception occurs in the case of the TG*T sequences, in which significant cumulative bending effects are observed, even in the case of the out-of-phase 16-mer ligation products (Figure 3, panels A and B). The results of Tsao et al. (*20*) suggest that there is significant flexibility associated with (+)-*trans* lesions in the ...TG*T... sequence context. The pronounced phasing effects on R_L in the CG*T, TG*C, CG*A, and AG*C sequences containing the identical (+)-*trans-anti*-[BP]-N^2-dG lesions indicate that the bends induced by the [BP] residues are more rigid. A similar phasing effect on R_L was observed by Tsao et al. (*20*) in relation to bends in CG*C and interpreted to indicate greater rigidity than for TG*T. The present results indicate that the flexibility of the TG*T sequences requires flanking Ts on both sides of the (+)-*trans-anti*-[BP]-N^2-dG lesions, since a pronounced phasing effect is observed in the case of the TG*C and CG*T sequences (Figure 3).

We note that the ligation of 11-mers with (+)-*trans-anti*-[BP]-N^2-dG lesions gives rise not only to bent linear molecules but also to small covalently closed circles. Circular ligation products are formed upon ligation of the 11-mer TG*T (16) and CG*C (20) sequences. The circular molecules migrate more slowly than linear ligation products of the same sequence length and can be easily distinguished from linear ones by two-dimensional gel electrophoresis and the appearance of intense bands in the upper portions of the gel lanes (see, for example, ref *20*). Such bands were not included in the data presented here. Furthermore, significant amounts of circular ligation products were not observed in any of the other sequences nor in any of the 16-mers studied in this work.

Computed Mobilities. The distances *d* migrated by the oligomers of different lengths were calculated as described in the Methods section. The calculated values of *d* that give rise to the best fits to the experimental data are shown in Figures 4 and 5 for all unmodified sequences studied (Table 1) and for the same sequences with single (+)-*trans-anti*-[BP]-N^2-dG lesions in Figures 6 and 7. A number of the sequences in Table 1 with (−)-*trans-anti*-[BP]-N^2-dG were also studied (Figure 8).

For the unmodified 11- and 16-mers with different ligation sequences, the calculated mobilities are almost identical to the experimental results. In matching the migration distances for the BPDE-modified DNA, the computed data in Figures 6−8 only consider the perturbation of intrinsic (equilibrium) roll from the unmodified step parameters. However, we also carried out trial calculations in which equilibrium values of tilt and twist were varied, together with roll, in various

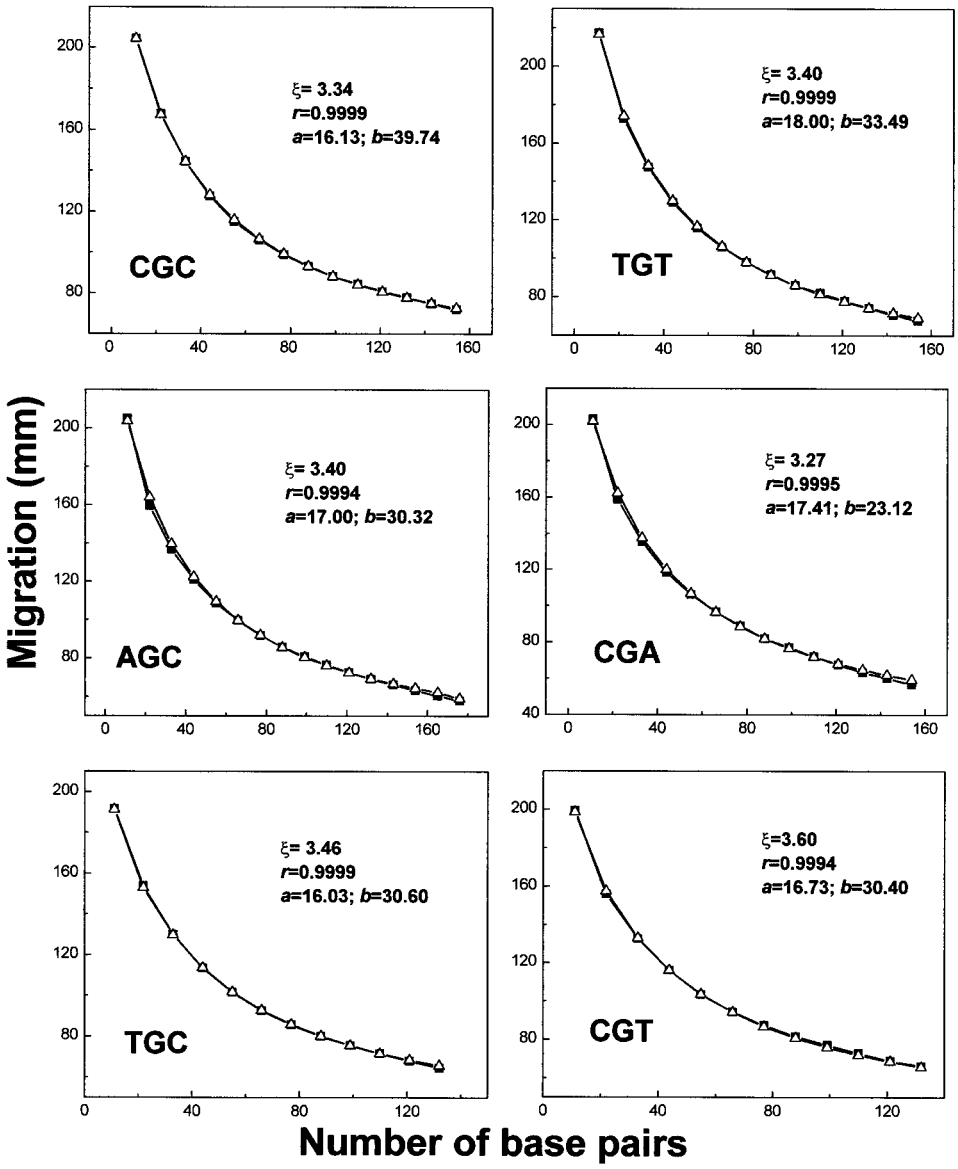

FIGURE 4: Comparison of experimental and computed migration distances of unmodified XGY-11 duplexes (■, experimental, and △, computed). The deformability parameter (ξ) and correlation (r) coefficients of the best-fit models and the slopes a and intercepts b of their linear correlation are given in the figure. Intrinsic base-pair step geometry is taken from Table 2.

realistic combinations (*2*, *3*). We found that roll changes alone sufficed to reproduce the data. Changes in twist and tilt within physically meaningful ranges do not significantly affect the values of $\langle R^2 \rangle$ and thus d in eq 2. By adjusting the roll, the calculated values of d from eq 2 can be made to approach the experimentally observed values in almost all cases. In addition, some trials were carried out in which the deformability parameter ξ at the modification site (see Methods section) was varied from the value deduced from unmodified DNA, using 2−3-fold lower and higher quantities, in combination with changes in roll. The effect on the agreement between experimental and computed migration distances was slight.

The molecular structures in Figure 9 illustrate the distortions of the DNA double helix deduced from this analysis in the CGT sequence context. The observed variations of base-pair step parameters found in high-resolution DNA crystal structures introduce only minor perturbations of global structure in a short unmodified (CGT-16) duplex (Figure 9A). The limited movements of terminal base pairs in this example (4−5 Å root-mean-square deviations of the ends, corresponding to a global bend of 14 ± 7°, in the intrinsic 16 bp structure shown at the atomic level), nevertheless, accumulate sufficiently with increase in chain length to account for the observed (*48*) persistence length of DNA in aqueous salt solution (Olson and Srinivasan, unpublished data). The ends of the same 16-mer show significantly larger deviations from the ends of the equilibrium reference structure (14−15 Å base pair displacements or a 44 ± 22° global bend angle) when the flexibility of each dimer is increased to the (ξ = 3.3) level needed to account for the gel mobilities of carcinogen-free CGT-16 multimers (Figure 9B). The increase in dimer flexibility under gel vs solution conditions corresponds at the level of the 16-mer to a 12-fold increase in conformational entropy as measured by the volumes of the spatial density distributions (small dots in Figure 9) accumulated in the sampling of 50 000 configurations. The 32° intrinsic roll implicated by the binding of BP (Table 3), in

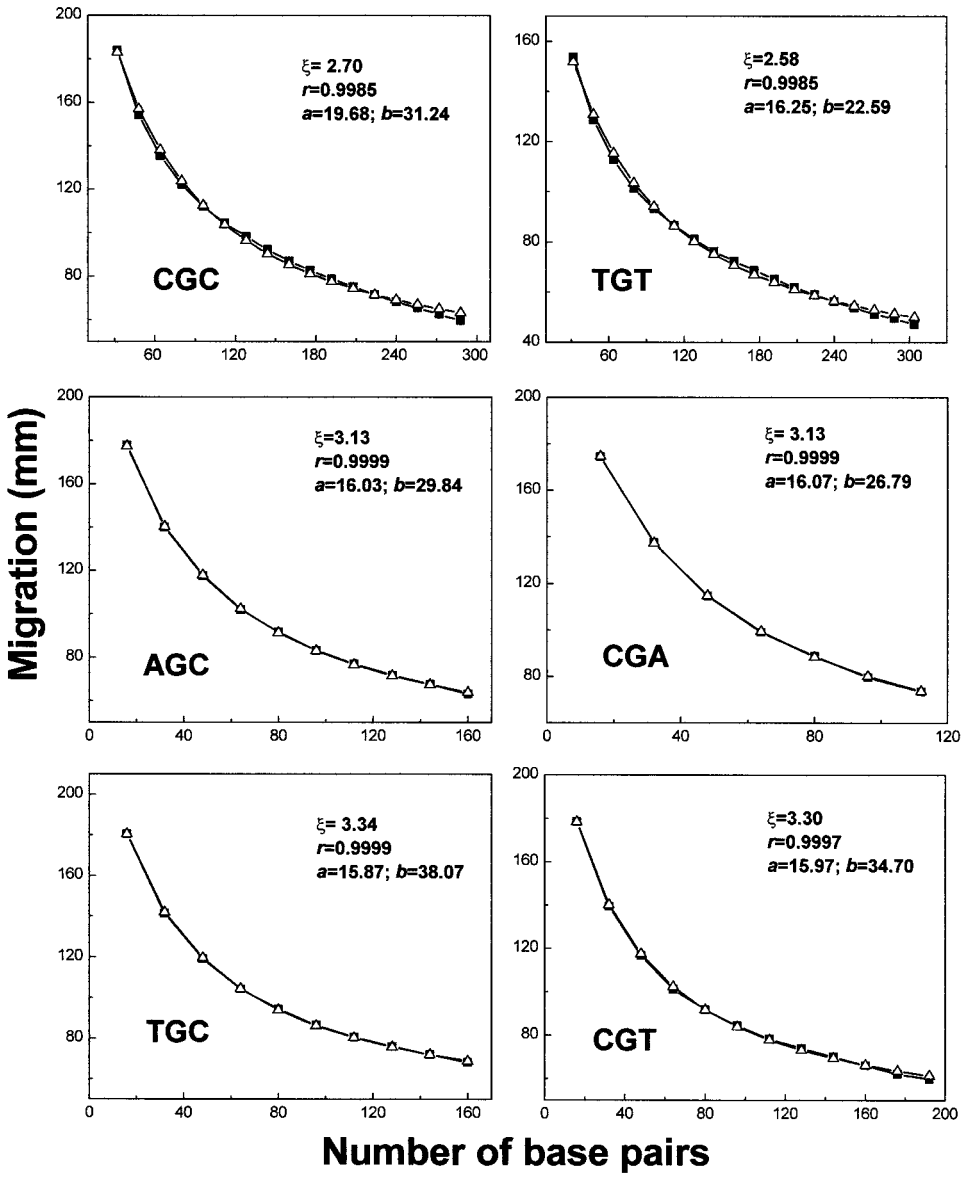

FIGURE 5: Comparison of experimental and computed migration distances of unmodified XGY-16 duplexes (■, experimental, and △, computed). See legend to Figure 4.

Table 3: Inferred Roll Angles (in Deg) of BPDE-Modified DNA[a]

sequence	(+)-*trans*	(−)-*trans*	sequence	(+)-*trans*	(−)-*trans*
CG*C	38		CG*A	32	19
TG*T	33		TG*C	30	20
AG*C	32	21	CG*T	32	

[a] Distortions for (+)-*trans* lesions located at the X*G dimer step and for (−)-*trans* lesions at the G*Y step.

combination with the enhancement of dimer flexibility, brings the two ends of the (CG*T-16) oligomer closer to one another (Figure 9C) with a root-mean-square end-to-end distance of 44.3 Å versus values of 45.7 Å in the unmodified duplex subject to the same local motions and 49.6 Å in the carcinogen-free DNA under "solution" conditions. The change in intrinsic roll around the lesion site similarly adds to the degree and range of global bending (49 ± 25°). The presence of carcinogen, however, has no effect on the entropy of the modified 16-mer relative to the unmodified DNA in the gel as expected from the assumed equivalence of CG and CG* flexibility (Figure 10).

DISCUSSION

We have assumed, for simplicity, that all of the structural changes associated with the [BP]-N^2-dG adducts occur at one dinucleotide base pair step, i.e., the site of the G* lesion and its 5′- or 3′-neighbor for the (+)- and (−)-*trans* adducts, respectively (see Methods). However, some trials were carried out in which the changes in roll were distributed over the two base pairs adjacent to the G*•C pair, with very similar values of $\langle R^2 \rangle$ and d obtained as when the changes are concentrated solely at a single dimer step. In addition, some trials were carried out in which the 3′-neighboring pair adjacent to the G*•C pair provided the dinucleotide step for the (+)-*trans-anti* adduct and the 5′-neighboring pair was used for the (−) adduct. Again, no significant differences in $\langle R^2 \rangle$ and d were obtained. Thus, carcinogen-induced changes in roll cannot be said to be concentrated on just one step or another, or both. Rather, the apparent roll angles deduced from these simulations represent the overall bend angles induced by the BP moiety.

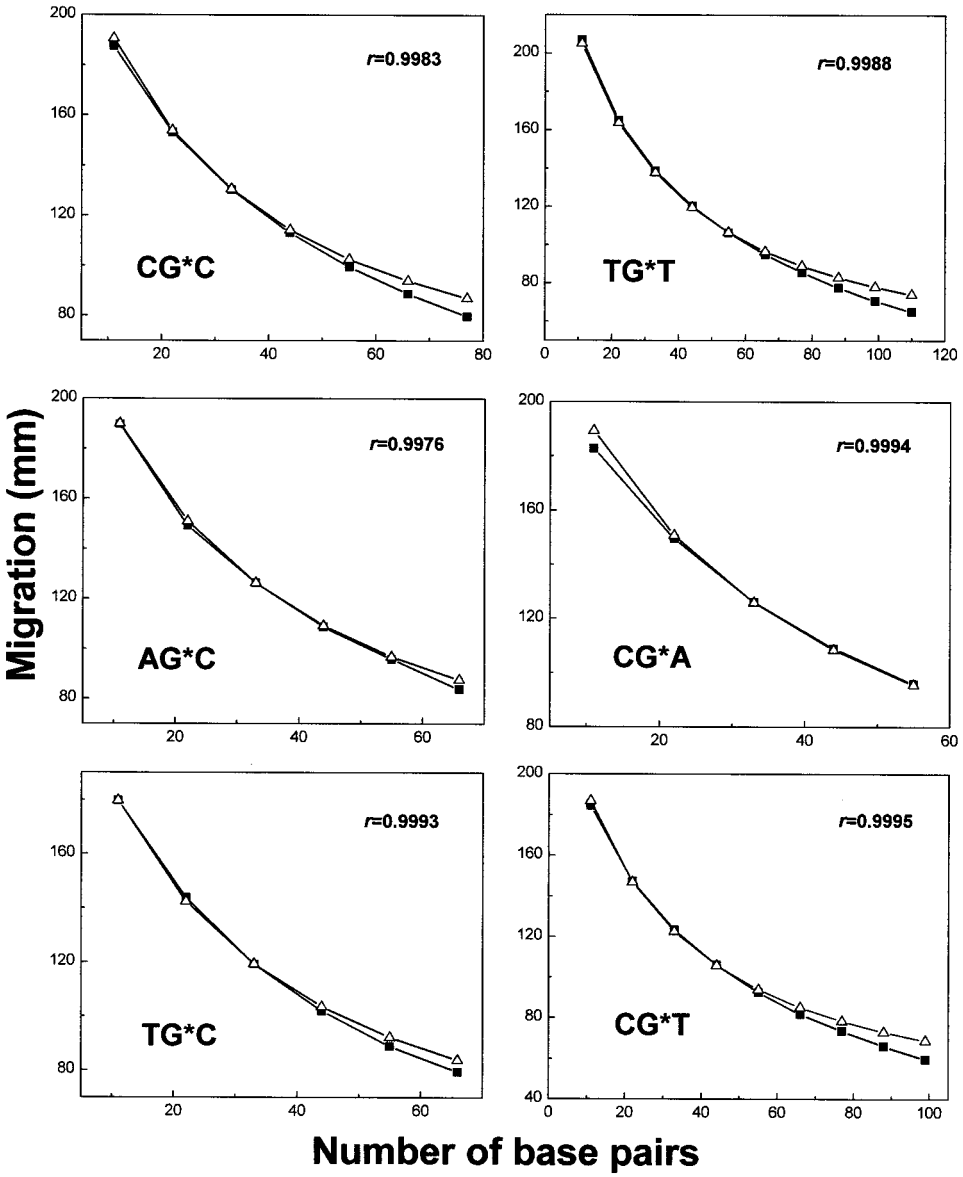

FIGURE 6: Comparison of experimental and computed migration distances of BPDE-modified (+)-*trans* XG*Y-11 duplexes (■, experimental, and △, computed). Best-fit modeling of end-to-end distances of chains with the base-pair step geometries in Tables 2 and 3. See Figure 4 for correlation coefficients *r* of each multimer series and the slopes *a* and intercepts *b* of their linear correlation.

For all six unmodified sequences studied, the fits of the calculated values of *d* to the experimental data are excellent using the sets of base-pair step parameters listed in Table 2 for the 11- and 16-mer ligation products (Figures 4 and 5 and Table 3). Reasonable fits are also obtained for the ligated 11-mers (Figure 6) with single (+)-*trans-anti* lesions. The fits with the identical step parameters are not uniformly as good in the calculation of *d* for the (+)-*trans*-16-mer ligation products (Figure 7).

The differences between computation and experiment are most pronounced in the case of the highly flexible TG*T sequence context. It was shown by Xu et al. (*16*) that, in the TG*T sequence context, multiple adduct conformers are observed by NMR techniques, while in our computational model the existence of a single conformation is assumed. The different conformers noted by NMR may be characterized by different values of intrinsic roll, thus accounting for the failure to identify a single value of roll that fits all of the 11- and 16-mer ligation product data. The flexibilities of the oligomers with the TG*T sequence context is confirmed by the lack of a clear-cut phasing effect on R_L (Figure 3, top panels) (*20*).

In the case of the CG*C sequence context, a single NMR structure was observed, indicating a single adduct conformation (*17*). The relative rigidity of the bends induced by the (+)-*trans* lesions in the CG*C sequence context is confirmed by the strong phasing effect on the R_L values (Figure 3) (*20*). However, Fountain and Krugh (*18*) found some adduct heterogeneity in an oligonucleotide with a TG*C sequence context but different beyond the central trimer from the one used in this work (Table 1). The phasing experiments (Figure 3) suggest that, in the TG*C-11 and TG*C-16 sequences, the bends induced by the (+)-*trans* lesions are rigid rather than flexible. However, except for the TG*T-11 case (*16*), NMR data are not available for either the TG*C or any of the other exact sequences employed in this work. All of these results indicate that T·A or A·T base pairs flanking the modified guanine residue G* on both sides,

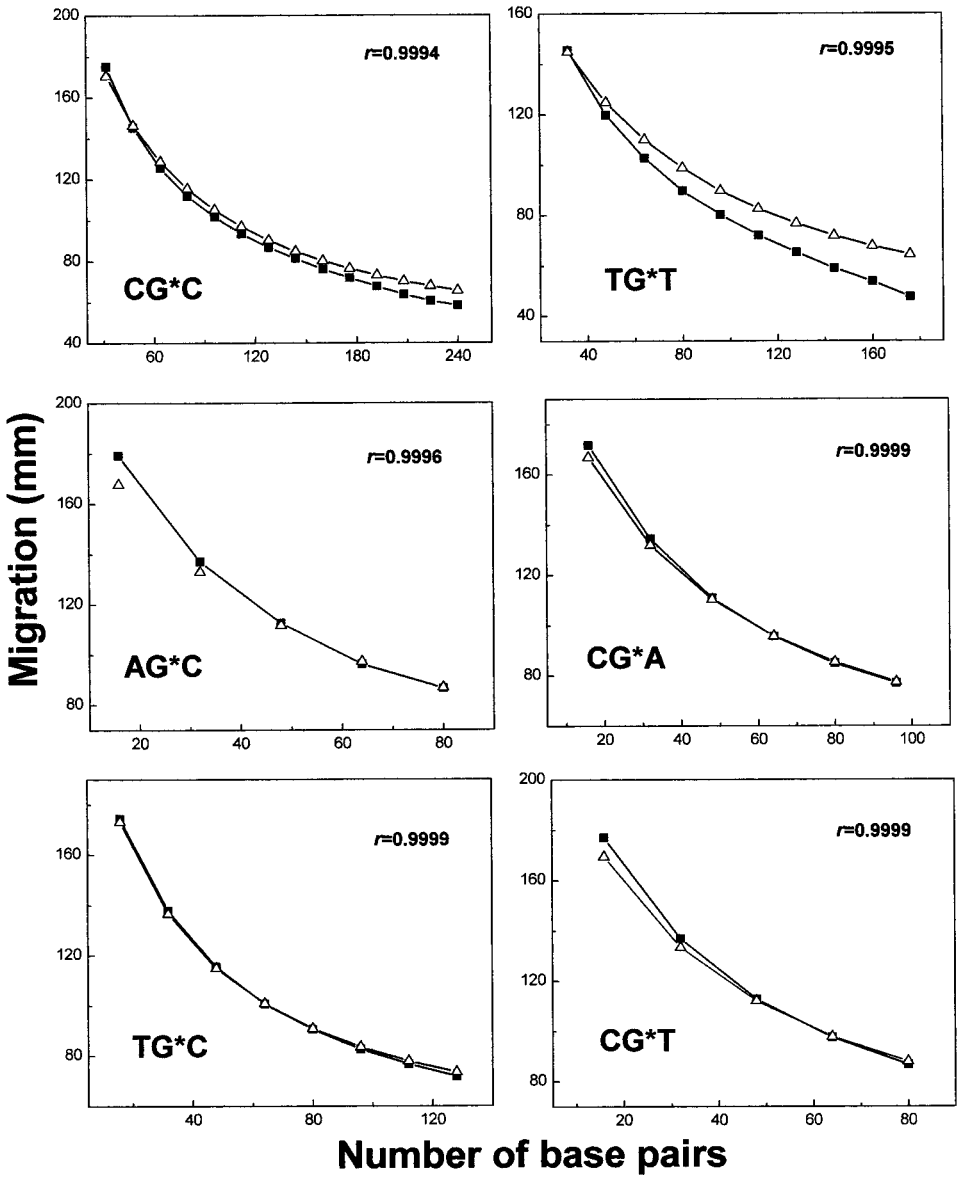

FIGURE 7: Comparison of the experimental and computed migration distances of BPDE-modified (+)-*trans* XG*Y-16 duplexes (■, experimental, and △, computed). Best-fit modeling of end-to-end distances of chains with the base-pair step geometries in Tables 2 and 3. See Figure 5 for correlation coefficients *r* of each multimer series and the slopes *a* and intercepts *b* of their linear correlation.

having only two hydrogen bonds, may allow for considerable local flexibility and thus conformational adduct heterogeneity. It has been suggested that the (+)-*trans*-*anti*-[BP]-N^2-dG adducts, which are mostly positioned in the minor groove (*17*), are to varying extents in equilibrium with base-displaced intercalated conformers in TG*C (*18*) and TG*T sequence contexts (*16*, *50*). The apparent mean roll for the TG*T sequence (33°), which gives the best fit of calculated and observed migration distances, is smaller than that obtained for the CG*C sequence (38°). The lower value for TG*T could represent a mean between two conformers, one of which is less bent. Evidence that base-displaced intercalated conformers are less bent than minor groove conformers has been obtained (*20*).

The high chain flexibility, presumably anisotropic in nature, in the TG*T sequence context is consistent with the high efficiency of circularization of the (+)-*trans*-TG*T oligonucleotides. The inferred roll angles are not significantly different for the AG*C, CG*A, CG*T, and TG*C sequences (Table 3). As compared to TG*T, there is a strong phasing effect on the values of R_L in the latter chains. Thus, the flexibility is significantly reduced when only one flanking A·T or T·A base pair in TG*T, either on the 3′- or the 5′-side of G*, is replaced by a single G·C or C·G base pair. On the basis of the dinucleotide step model, one might have expected a high flexibility associated with the TG*C sequences (*20*), since the TG step is often associated with higher degrees of roll (*2*, *3*, *24*, *25*). The results depicted in Figure 6 indicate that T·A base pairs must flank the modified guanine residues G* on both sides to give rise to the unusual lack of a phasing effect and the high flexibilities of oligonucleotides with (+)-*trans* adducts in the TG*T sequence context.

The apparent intrinsic roll deduced for the BPDE-modified sequences is positive in all cases. This means that the bulky [BP] residues, positioned primarily either on the 5′-side or the 3′-side of the modified guanine residue in the minor groove (*17*, *19*), compress the major groove of DNA.

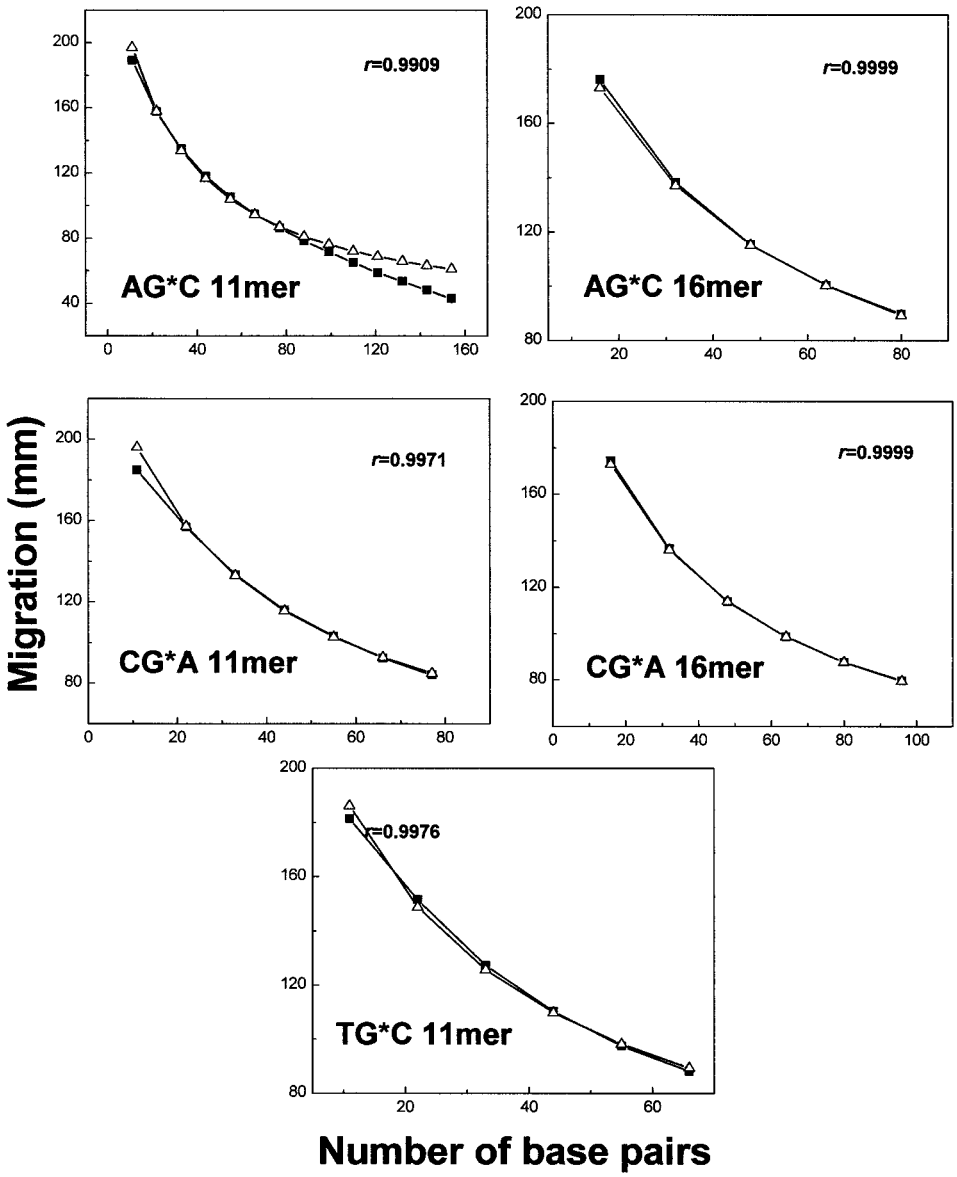

FIGURE 8: Comparison of experimental and computed migration distances of BPDE-modified (−)-*trans* XG*Y-11 and XG*Y-16 duplexes (■, experimental, and △, computed). Best-fit modeling of end-to-end distances of chains with the base-pair step geometries in Tables 2 and 3. See Figures 4 and 5 for correlation coefficients *r* of each multimer series and the slopes *a* and intercepts *b* of their linear correlation.

Negative roll is prevented by the presence of the carcinogen in the minor groove (Figure 9C).

In the case of the sequences studied with (−)-*trans* adducts, excellent agreement between calculated and experimental values of *d* is obtained in all cases, except the AG*C-11 and AG*C-16 sequences (Figure 8). The estimated values of intrinsic roll are much smaller (∼10°) than for the corresponding sequences with (+)-*trans* adducts (Table 3). The sequences with (−)-*trans* lesions are significantly less bent than the same DNA sequences with the stereoisomeric (+)-*trans* adducts. The origins of this effect in relation to the dinucleotide step model (*37*) were previously discussed by Tsao et al. (*20*). It was suggested that, in the case of the (−)-*trans*-*anti*-[BP]-N^2-dG adducts, the pyrenyl residues positioned on the 3′-side of G* in the minor groove (*19*) in the CG*C sequence context may hinder the natural bending into the minor groove associated with the G*C dinucleotide step. In the case of the (+)-*trans* adducts in the CG*C sequence, the bulky pyrenyl residue is on the 5′-side of the G* in the minor groove (*17*). This may reinforce the natural tendency for bending into the major groove at CG dinucleotide steps. A similar argument can be applied to the sequences investigated in the present work in both stereoisomeric adducts: AG*C, CG*A, and TG*C (Table 3). The AG and TG dimers, like CG, tend to bend into the major groove (average positive roll). While GC dimers have an average negative roll in B-DNA crystal structures (corresponding to bending into the minor groove), the GA dimers have average roll near 0° (0.5°) corresponding to little preference in bending direction (*3*). Consequently, the (+)-*trans* adduct induced bending into the major groove would at least meet little natural sequence-dependent resistance and could also account for the inferred anisotropic flexibility.

In some cases, the observed values of *d* for the residual unligated oligomer cannot be fitted well theoretically. All of the ligated strands have unpaired terminal bases, and unpaired strands are well-known to have different mobilities from paired ones (*28*). The effect of unpaired ends in double

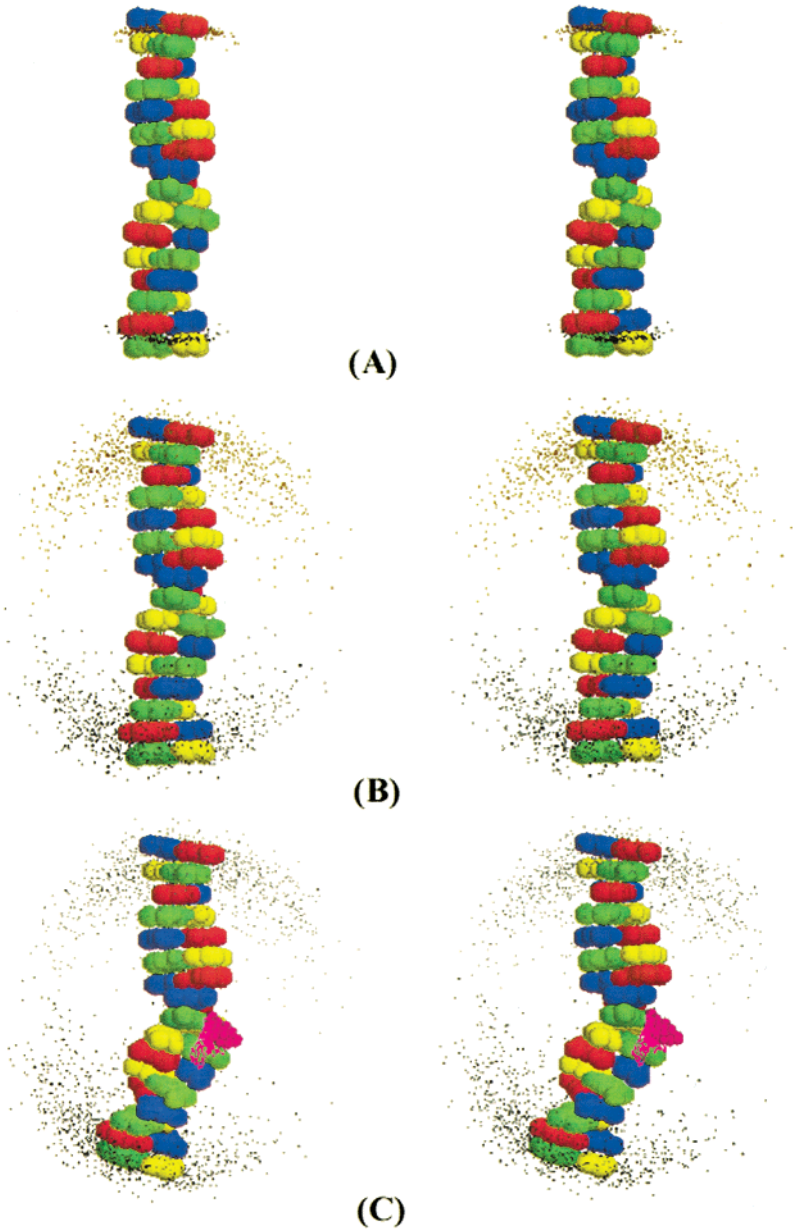

FIGURE 9: Stereopairs illustrating the range of three-dimensional structures accessible to 16 bp d(CTCACACG*TACACTCT)·d(AGAGTGTACGTGTGAG) duplexes (A, B) without carcinogen and (C) carcinogen-bound with a (+)-*trans-anti*-[BP]-N^2-G adduct at the central guanine (G*). Each duplex is described in terms of its intrinsic equilibrium structure (static, color-coded, space-filling representations, where A = red, C = yellow, G = green, and T = blue) and overall flexibility, i.e., the locations of the centers of terminal base pairs (shown as small dots) in 1000 Monte Carlo snapshots with respect to a coordinate frame embedded in the plane of the central G·C or G*·C pair. The spatial density distributions in (A) are accumulated by incorporating, at each base pair step, the ranges of sequence-dependent step parameters observed in high-resolution DNA crystal structures (*3*). The broadened distributions in (B) reflect the enhancement in dimer flexibility (see ξ values in text) needed to account for the gel mobilities of carcinogen-free multimeric sequences with the models of Lerman and Frisch (*28*) and Lumpkin and Zimm (*29*). The model in (C) incorporates both the enhanced flexibility and the 32° intrinsic roll at the CG*·CG dimer step consistent with gel measurements. MD-simulated fluctuations in the CGT sequence context (0.4 ns intervals, over 2 ns) of the BP moiety in the minor groove with respect to the G*·C base pair (magenta images) are superimposed on the equilibrium structure of the bent DNA.

strands is particularly significant if the strands are as short as one of our 11 or 16 bp ligation fragments. Moreover, the experimental error in the measured migration distances of short and fast-moving strands is greater than for longer DNA. In addition, the assumptions of the reptation model cannot be satisfied very well in short oligomers, and it is not clear how these strands move in the network of an electrophoretic gel. The recent work of Mohanty et al. (*51*) offers explanations for the anomalous gel migration observed in very short DNA sequences by explicitly taking into account both hydrodynamic interactions and polyelectrolyte effects.

As noted in Results, while we have employed only intrinsic roll to match migration data, trial calculations in which other step parameters and the deformability parameter (ξ) are varied show that roll alone suffices to reproduce the data. Neither changes in twist/tilt within physically realistic ranges nor alterations in ξ at the modification site significantly affect the values of $\langle R^2 \rangle$ and thus d in eq 2.

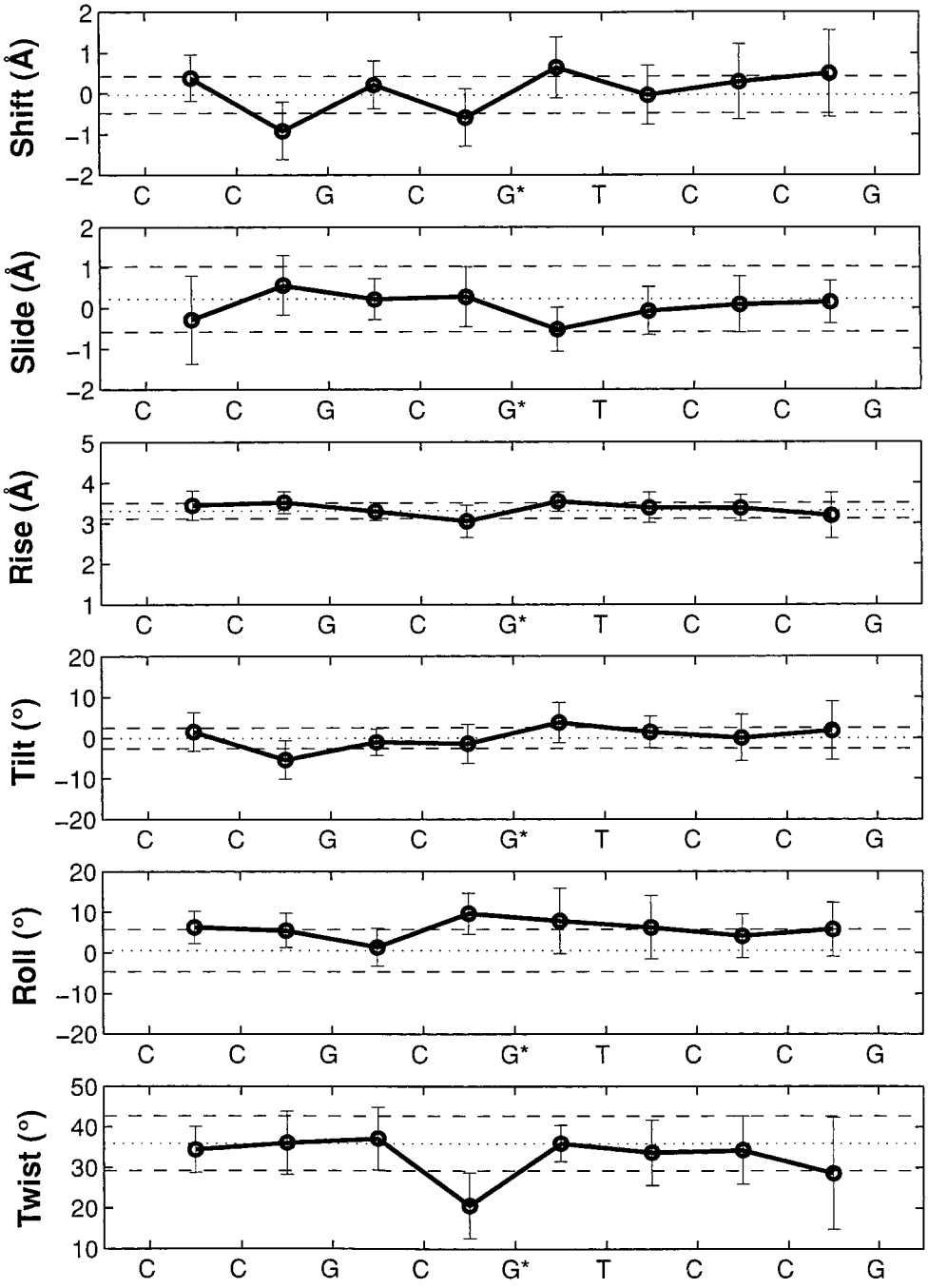

FIGURE 10: Comparison of averages (circles) and standard deviations (error bars) of local base pair step parameters obtained during the course of a 2 ns molecular dynamics simulation of the 9 bp (+)-*trans-anti* [BP]-N^2-dG modified duplex, 5′-d(CCGCG*TCCG)-3′•5′-(CGGACGCGG)-3′, with values extracted from high-resolution B-DNA crystal structures (dotted lines corresponding to sequence-independent, average values and dashed lines to root-mean-square dispersion as described in the text and references therein).

Our theoretical calculation of the mobilities of BPDE-modified, double-stranded DNA sequences in polyacrylamide gels is based on simplifying assumptions (detailed in Methods), grounded in the reptation models of deGennes (*52*) and Doi and Edwards (*53*), in treating the electrophoretic motion of small DNA molecules in gels. Because all of the complexities of the motion of small DNA molecules in gels are still not well understood (*54*), the semiempirical equations on which our computations are based should be regarded as approximate and cannot be used to interpret the experimental results with high accuracy. Furthermore, as Lerman and Frisch (*28*) have noted, understanding of field strength effects is also inadequate, and detailed comparison of experiment and theory is most appropriate for mobility measurements extrapolated to zero field. The apparent enhancement of DNA dimer flexibility in gels compared to aqueous salt solution (Figure 9) might be an artifact of the reptation model, although Levene and Zimm (*34*) inferred a similar chain bending enhancement in gels from their tube model. A more detailed and accurate theory describing the relationships between intrinsic DNA structure and flexibility, and the relationships between these factors and the electrophoretic mobilities of small, rodlike DNA molecules, is needed to provide deeper insights into the results reported here. Finally,

the bending differences deduced for the (+)- and (−)-*trans-anti*-[BP]-N^2-dG adducts and the sequence context effects we observe may play a significant role in the processing of these lesions by cellular replication, transcription, and repair enzymes.

APPENDIX

Local step parameters generated by Monte Carlo sampling are incorporated directly in the transformation matrix $\mathbf{T}_{i,i+1}$ that relates the coordinate frames of adjacent base pairs and the displacement vector $\mathbf{v}_i$ between base-pair centers. We construct $\mathbf{T}_{i,i+1}$ and $\mathbf{v}_i$ so that the parameters are independent of chain direction. Specifically, we adopt the definitions of Babcock et al. (*30*):

$$\mathbf{T}_{i,i+1}(\tau,\rho,\Omega) = \begin{bmatrix} \cos\varphi + (1-\cos\varphi)u_x^2 & (1-\cos\varphi)u_xu_y - u_z\sin\varphi & (1-\cos\varphi)u_xu_z + u_y\sin\varphi \\ (1-\cos\varphi)u_xu_y + u_z\sin\varphi & \cos\varphi + (1-\cos\varphi)u_y^2 & (1-\cos\varphi)u_yu_z - u_x\sin\varphi \\ (1-\cos\varphi)u_xu_z - u_y\sin\varphi & (1-\cos\varphi)u_yu_z + u_x\sin\varphi & \cos\varphi + (1-\cos\varphi)u_z^2 \end{bmatrix}$$

$$\mathbf{v}_i(\tau,\rho,\Omega,Dx,Dy,Dz) = \mathbf{T}_{i,i+1}\left(\frac{\tau}{2},\frac{\rho}{2},\frac{\Omega}{2}\right)\begin{bmatrix} Dx \\ Dy \\ Dz \end{bmatrix}$$

where $\varphi = (\tau^2 + \rho^2 + \Omega^2)^{1/2}$ is the net angle of rotation and ($u_x = \tau/\varphi$, $u_y = \rho/\varphi$, $u_z = \Omega/\varphi$) are the components of the axis of rotation which bring the base-pair frames into coincidence, with τ = tilt, ρ = roll, Ω = twist, Dx = shift, Dy = slide, and Dz = rise.

We find the mean-square end-to-end distance $\langle R^2\rangle$ by substituting the average transformation matrix $\langle\mathbf{T}_{i,i+1}\rangle$, the average vectorial displacement between base pair centers $\langle\mathbf{v}_i\rangle$, and the average products $\langle\mathbf{v}_i^{\mathrm{T}}\mathbf{T}_{i,i+1}\rangle$ and $\langle\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i\rangle$ at dimer step i, $i+1$ into the matrix expression (33):

$$\langle R^2\rangle = 2[1\ \mathbf{0}\ 0]\prod_{i=1}^{N}\begin{bmatrix} 1 & \langle\mathbf{v}_i^{\mathrm{T}}\mathbf{T}_{i,i+1}\rangle & {}^1/_2\langle\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i\rangle \\ \mathbf{0} & \langle\mathbf{T}_{i,i+1}\rangle & \langle\mathbf{v}_i\rangle \\ 0 & \mathbf{0} & 1 \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{0} \\ 1 \end{bmatrix}$$

The above formulation, corresponding to a duplex with $N+1$ base pairs, assumes that the deformations of dimer step $i+1$ are independent of those at all other such steps in the chain. The properties of such a DNA are thus determined by the average configuration of a single dimer unit. The superscript T in the preceding equation is used to denote the row form of the vector $\mathbf{v}_i$, while the **0**'s are null vectors of orders necessary to fill the various arrays.

We find the average components of the generator matrices from a Gaussian distribution of angular variables governed by the harmonic energy function reported in Table 2. We set the equilibrium rest angles and base-pair displacement (τ_0, ρ_0, Ω_0, Dx_0, Dy_0, Dz_0) and the root-mean-square angular fluctuations ($\Delta\tau$, $\Delta\rho$, $\Delta\Omega$, ΔDx, ΔDy, ΔDz) to values listed in the table. In practice, we find stable computed averages in samples of 10^5 dimer configurations.

SUPPORTING INFORMATION AVAILABLE

Seven figures showing ligation ladders of unmodified and BPDE-modified duplexes, one table giving average step parameters in the representative MD-simulated conformational ensemble of the (+)-*trans-anti*-[BP]-N^2-dG adduct in a DNA duplex, and two tables giving partial charges, atom type assignments, and bond angle and improper torsional parameter assignments for the molecular dynamics calculations with AMBER 5.0. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dickerson, R. E., and Chiu, T. K. (1997) *Biopolymers 44*, 361−403.
2. Dickerson, R. E. (1998) *Nucleic Acids Res 26*, 1906−1926.
3. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11163−11168.
4. Maher, L. J., III (1998) *Curr. Opin. Chem. Biol. 2*, 688−694.
5. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) *J. Mol. Biol. 287*, 877−896.
6. Rice, J. A., Crothers, D. M., Pinto, A. L., and Lippard, S. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4158−4161.
7. Schwartz, A., Marrot, L., and Leng, M. (1989) *J. Mol. Biol. 207*, 445−450.
8. Moe, J. G., Reddy, G. R., Marnett, L. J., and Stone, M. P. (1994) *Chem. Res. Toxico.l 7*, 319−328.
9. Zewail-Foote, M., and Hurley, L. H. (1999) *J. Med. Chem. 42*, 2493−2497.
10. Meehan, T., and Straub, K. (1979) *Nature 277*, 410−412.
11. Cheng, S. C., Hilton, B. D., Roman, J. M., and Dipple, A. (1989) *Chem. Res. Toxicol. 2*, 334−340.
12. Buening, M. K., Wislocki, P. G., Levin, W., Yagi, H., Thakker, D. R., Akagi, H., Koreeda, M., Jerina, D. M., and Conney, A. H. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5358−5361.
13. Slaga, T. J., Bracken, W. J., Gleason, G., Levin, W., Yagi, H., Jerina, D. M., and Conney, A. H. (1979) *Cancer Res. 39*, 67−71.
14. Xu, R., Mao, B., Xu, J., Li, B., Birke, S., Swenberg, C. E., and Geacintov, N. E. (1995) *Nucleic Acids Res. 23*, 2314−2319.
15. Liu, T., Xu, J., Tsao, H., Li, B., Xu, R., Yang, C., Amin, S., Moriya, M., and Geacintov, N. E. (1996) *Chem. Res. Toxicol. 9*, 255−261.
16. Xu, R., Mao, B., Amin, S., and Geacintov, N. E. (1998) *Biochemistry 37*, 769−778.
17. Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Singh, S. B., Ibanez, V., Margulis, L. A., Live, D., Geacintov, N. E., Broyde, S., et al. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1914−1918.
18. Fountain, M. A., and Krugh, T. R. (1995) *Biochemistry 34*, 3152−3161.
19. de los Santos, C., Cosman, M., Hingerty, B. E., Ibanez, V., Margulis, L. A., Geacintov, N. E., Broyde, S., and Patel, D. J. (1992) *Biochemistry 31*, 5245−5252.
20. Tsao, H., Mao, B., Zhuang, P., Xu, R., Amin, S., and Geacintov, N. E. (1998) *Biochemistry 37*, 4993−5000.
21. Olson, W. K., and Zhurkin, V. B. (1996) in *Biological Structure and Dynamics* (Sarma, R. H., and Sarma, M. H., Eds.) Vol. 2, pp 341−370, Adenine Press, Schenectady, NY.
22. Olson, W. K., Babcock, M. S., Gorin, A., Liu, G., Marky, N. L., Martino, J. A., Pedersen, S. C., Srinivasan, A. R., Tobias, I., and Westcott, T. P. (1995) *Biophys. Chem. 55*, 7−29.
23. Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995) *J. Mol. Biol. 247*, 34−48.
24. Lyubchenko, Y. L., Shlyakhtenko, L. S., Appella, E., and Harrington, R. E. (1993) *Biochemistry 32*, 4121−4127.
25. Nagaich, A. K., Bhattacharyya, D., Brahmachari, S. K., and Bansal, M. (1994) *J. Biol. Chem. 269*, 7824−7833.
26. Kabsch, W., Sander, C., and Trifonov, E. N. (1982) *Nucleic Acids Res. 10*, 1097−1104.
27. Geacintov, N. E., Cosman, M., Mao, B., Alfano, A., Ibanez, V., and Harvey, R. G. (1991) *Carcinogenesis 12*, 2099−2108.
28. Lerman, L. S., and Frisch, H. L. (1982) *Biopolymers 21*, 995−997.
29. Lumpkin, O. J. (1982) *Biopolymers 21*, 2315−2316.
30. Babcock, M. S., Pednault, E. P., and Olson, W. K. (1994) *J. Mol. Biol. 237*, 125−156.
31. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R.,

Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., and Zhurkin, V. B. (1989) *Nucleic Acids Res. 17*, 1797−1803.
32. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R., and Schneider, B. (1992) *Biophys. J. 63*, 751−759.
33. Marky, N. L., and Olson, W. K. (1994) *Biopolymers 34*, 109−120.
34. Levene, S. D., and Zimm, B. H. (1989) *Science 245*, 396−399.
35. Chrambach, A., and Rodbard, D. (1971) *Science 172*, 440−451.
36. Fawcett, J. S. M., and Morris, C. J. O. R. (1966) *Sep. Sci. 1*, 9−28.
37. Olson, W. K., Marky, N. L., Jernigan, R. L., and Zhurkin, V. B. (1993) *J. Mol. Biol. 232*, 530−554.
38. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C., Darden, T., Merz, K. M., Stanton, R. V., Cheng, A., Vincent, J. J., Crowley, M., Ferguson, D. M., Radner, R., Siebel, G. L., Singh, U. C., Weiner, P., and Kollman, P. A. (1997) *AMBER 5.0 Documentation*, University of California, San Francisco, CA.
39. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
40. Cheatham, T. E., III, Srinivasan, J., Case, D. A., and Kollman, P. A. (1998) *J. Biomol. Struct. Dyn. 16*, 265−280.
41. Darden, T., Perera, L., Li, L., and Pedersen, L. (1999) *Structure (London)*, R55−R60.
42. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P., Johnson, B. G., Rob, M. A., Cheeseman, J. R., Keith, T. A., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Reploge, E. S., Gomperts, R., Martin, R., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzales, C., and Pople, J. A. (1995) *Gaussian 94 (Revision A.1)* Gaussian, Inc., Pittsburgh, PA.
43. Wu, X., Shapiro, R., and Broyde, S. (1999) *Chem. Res. Toxicol. 12*, 895−905.
44. Hingerty, B. E., Figueroa, S., Hayden, T. L., and Broyde, S. (1989) *Biopolymers 28*, 1195−1222.
45. Hingerty, B. E., Ritchie, R. H., Ferrell, T. L., and Turner, J. E. (1985) *Biopolymers 24*, 427−439.
46. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys. 79*, 926−935.
47. Koo, H. S., Wu, H. M., and Crothers, D. M. (1986) *Nature 320*, 501−506.
48. Hagerman, P. J. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 265−286.
49. Hagerman, P. J. (1985) *Biochemistry 24*, 7033−7037.
50. Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S., and Patel, D. J. (1997) *Chem. Res. Toxicol. 10*, 111−146.
51. Mohanty, U., Searls, T., and McLaughlin, L. W. (2000) *Proc. Conversation Biomol. Stereodyn., 11th*, 371−375.
52. De Gennes, P. G. (1979) *Scaling Concepts in Polymer Physics*, p 324, Cornell University Press, Ithaca and London.
53. Doi, M., and Edwards, S. F. (1978) *J. Chem. Soc., Faraday Trans. 2 74*, 560−570, 918−930.
54. Zimm, B. H. (1993) *Curr. Opin. Struct. Biol. 3*, 373−376.

BI002643X