Spectroscopic and Theoretical Insights into Sequence Effects of Aminofluorene-Induced Conformational Heterogeneity and Nucleotide Excision Repair[†,▽]

Srinivasa Rao Meneni,[‡] Steven M. Shell,[§] Lan Gao,[∥] Petr Jurecka,[⊥,#] Wang Lee,[‡] Jiri Sponer,[⊥] Yue Zou,[§] M. Paul Chiarelli,[∥] and Bongsup P. Cho*[,‡]

Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, 41 Lower College Road, Kingston, Rhode Island 02881, Department of Biochemistry and Molecular Biology, Quillen College of Medicine, East Tennessee State University, Johnson City, Tennessee 37614, Department of Chemistry, Loyola University, Chicago, Illinois 60626, Institute of Biophysics, Academy of Sciences of the Czech Republic, v.v.i. Kralovopolska 135, 612 65 Brno, Czech Republic, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Prague, Czech Republic, and Department of Physical Chemistry, Palacky University, Tr. Svobody 26, 77146 Olomouc, Czech Republic

Received May 6, 2007; Revised Manuscript Received July 11, 2007

ABSTRACT: A systematic spectroscopic and computational study was conducted in order to probe the influence of base sequences on stacked (S) versus B-type (B) conformational heterogeneity induced by the major dG adduct derived from the model carcinogen 7-fluoro-2-aminofluorene (FAF). We prepared and characterized eight 12-mer DNA duplexes (-AG*N- series, d[CTTCT**AG*N**CCTC]; -CG*N- series, d[CTTCT**CG*N**CCTC]), in which the central guanines (G*) were site-specifically modified with FAF with varying flanking bases (N = G, A, C, T). S/B heterogeneity was examined by CD, UV, and dynamic ^{19}F NMR spectroscopy. All the modified duplexes studied followed a typical dynamic exchange between the S and B conformers in a sequence dependent manner. Specifically, purine bases at the 3′-flanking site promoted the S conformation (G > A > C > T). Simulation analysis showed that the S/B energy barriers were in the 14−16 kcal/mol range. The correlation times ($\tau = 1/\kappa$) were found to be in the millisecond range at 20 °C. The van der Waals energy force field calculations indicated the importance of the stacking interaction between the carcinogen and neighboring base pairs. Quantum mechanics calculations showed the existence of correlations between the total interaction energies (including electrostatic and solvation effects) and the S/B population ratios. The S/B equilibrium seems to modulate the efficiency of *Escherichia coli* UvrABC-based nucleotide excision repair in a *conformation-specific* manner: *i.e.*, greater repair susceptibility for the S over B conformation and for the -AG*N- over the -CG*N- series. The results indicate a novel structure−function relationship, which provides insights into how bulky DNA adducts are accommodated by UvrABC proteins.

A mutation is defined as a heritable change in genome sequence and may occur as a result of an imbalance between repair and replication (*1*). Most mutagens and carcinogens are metabolized *in vivo* into reactive electrophiles, which subsequently react with cellular DNA to produce DNA adducts (*2*). Relating adduct formation with a specific mutation at the molecular genetic level is a challenging task (*3*−*7*). Earlier efforts to simply connect the types and extent of DNA adduct formation with mutations have largely been unsuccessful. This is due in part to an oversimplified view that damaged DNA adopts a single major structure that is responsible for a specific mutation (*6*, *8*). Certain adducts derived from bulky carcinogens adopt multiple DNA conformations, and the resulting heterogeneities are modulated by the nature of the carcinogen as well as the base sequence contexts surrounding the lesion (*8*−*12*). Small energy differences among conformers could shift the adduct population balance, consequently influencing the choice of incoming dNTP within the active site of a polymerase (*8*). Since mutation is inherently an infrequent biological event, observed mutations could happen due to replication of certain minor adduct conformers that have escaped the repair process (*6*).

Arylamines are an important group of bulky mutagens, which include heterocyclic amines found in overcooked foods and cigarettes (*13*). The prototype arylamine, 2-aminofluo-

[†] We are grateful to the NIH (R01CA098296 to B.P.C. and R01CA86927 to Y.Z.) for their financial support for this work. This research was made possible in part by the use of the RI-INBRE Research Core Facility supported by the NCRR/NIH (P20 RR016457). J.S. was supported by Grants LC06030, MSM0021622413, and AVOZ50040507 by Ministry of Education of the Czech Republic. P.J. was supported by grants (MSM6198959216 and LC512) from the Ministry of Education of the Czech Republic.

[▽] Dedicated to Professor Chang Kiu Lee on the occasion of his 60th birthday.

* To whom correspondence should be addressed. Tel: 401 874 5024. Fax: 401 874 5766. E-mail: bcho@uri.edu.

[‡] University of Rhode Island.

[§] East Tennessee State University.

[∥] Loyola University.

[⊥] Institute of Biophysics and Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

[#] Palacky University.

10.1021/bi700858s CCC: $37.00
Published on Web 09/18/2007

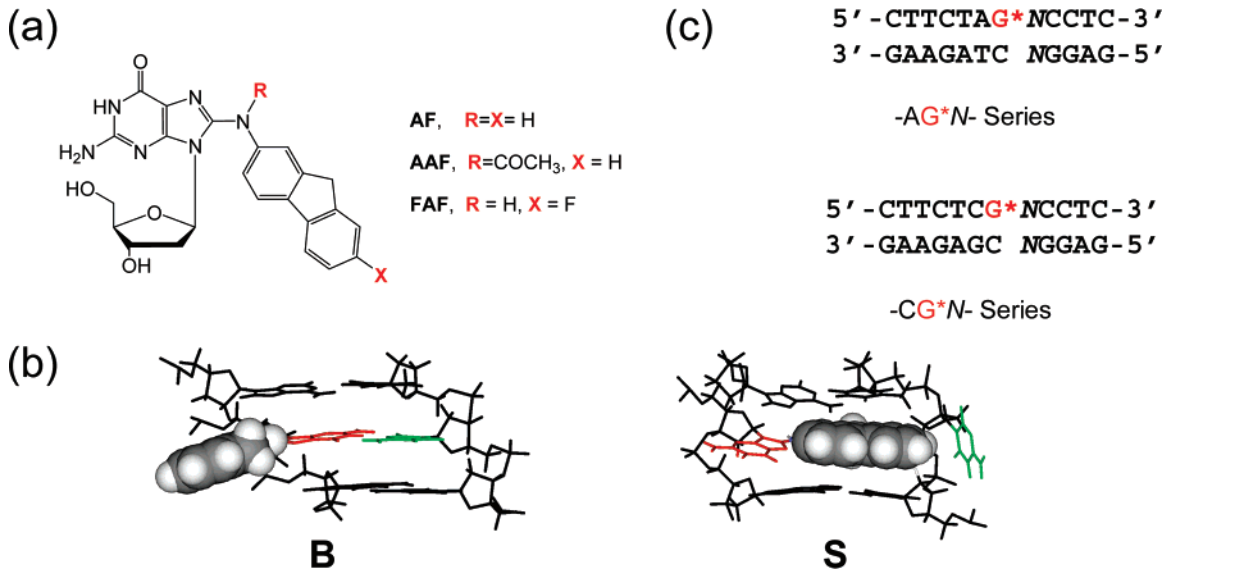

FIGURE 1: (a) Chemical structures of C8-substituted dG-aminofluorene adducts. AF adduct, *N*-(2′-deoxyguanosin-8-yl)-2-aminofluorene; AAF adduct, *N*-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene; FAF adduct, *N*-(2′-deoxyguanosin-8-yl)-7-fluororo-2-aminofluorene. (b) The major groove views of the central trimer segments of the B and S conformers of the AF-modified -AG*A- duplex, as an example. The modified dG and the complementary dC are shown in red and green, respectively, and the AF moiety is highlighted with gray CPK. In the B conformer, *anti*-[AF]dG maintains Watson−Crick hydrogen bonds, thereby placing the AF ring in the major groove. The AF moiety of the S conformer stacks into the helix with the modified dG in the *syn* conformation. (c) Sequence contexts of the -AG**N*- and -CG**N*- duplex series (G* = FAF adduct, *N* = G, A, C, T).

rene, and its derivatives produce two major C8-substituted dG adducts: AF[1] and AAF *in vivo* (Figure 1a) (*13*, *14*). Despite structural similarities, they exhibit unique mutation and repair activities and have been subjected to extensive structure−activity studies (*13*). Translesion synthesis of AF adducts is achieved with high fidelity polymerases, whereas replication of AAF adducts requires specialized bypass polymerases. Accurate nucleotide incorporation has been shown to be feasible, albeit slow, opposite the AF lesion in the active site of a DNA polymerase I *Bacillus* fragment, but not opposite the rigid AAF adduct which blocks nucleotide incorporation (*15*). In a similar study involving T7 DNA polymerase, unlike the AAF adduct, low electron density was observed around the AF moiety in the polymerase active site, suggesting conformational flexibility even in the solid state (*16*). Solution structures of the *N*-deacetylated AF adduct have been studied extensively with various DNA sequence contexts (*17*−*31*). The AF in fully paired DNA duplexes is in sequence dependent equilibrium between the external B-type (B) and stacked (S) conformations, as defined by its location (major groove and base-displaced, respectively) or the glycosidyl (χ) configurations (anti- and syn-, respectively) of the modified dG (Figure 1b). A minor groove binding wedged (W) conformer has also been observed in duplexes in which the lesion is mismatched with purine bases (*23*, *24*, *30*). The dynamics of the AF-induced B/S/W heterogeneity have been shown to be modulated by both the base sequence contexts and the length of primers, and contribute to polymerase activity through a long-range effect (*31*).

Probing the base sequence effects is complicated due to a delicate balance among various contributing chemical forces. Aromatic base stacking is the most likely factor for determining the local sequence dependent conformational variability within the double helical DNA; however, the relationship is complex (*32*). Intrinsic base stacking (the neat, direct base−base forces) is a relatively simple interaction dominated by common van der Waals terms modulated by electrostatics and steric factors (*33*). Nucleobases possess dipolar electrostatic distribution and thus prefer large van der Waals overlaps between coplanar aromatic rings. The nature of intercalator−base stacking is very similar to base−base stacking (*34*). However, capturing selectively the role of DNA stacking is very difficult since the stacking association arises not just from the intrinsic base stacking energy terms, but from their complex interplay with many other forces including solvation effects (*35*, *36*). The overall balance of forces varies with environment changes and nucleic acid structure (*37*, *38*).

Nucleotide excision repair (NER) is the major cellular pathway for removal of a wide variety of bulky lesions, otherwise, mutagenesis occurs (*39*−*41*). Sequence dependence of the efficiencies of DNA repair and replication *may account for* the presence of mutational "hot spots", such as the *Escherichia coli Nar*I exonuclease sequence (5′-CG_1G_2-CG_3C-3′) (*11*). Extensive efforts have been directed toward understanding the detailed molecular mechanisms by which bulky adducts are recognized in both bacterial and mammalian NER systems. Although the human NER proteins show no homology to the prokaryotic proteins, the overall strategy is the same: recognition, helix unwinding, incision, and patch (*41*). The hallmark of NER is its ability to repair a wide variety of lesions. In *E. coli* NER, the DNA damage

[1] Abbreviations: NER, nucleotide excision repair; AF, aminofluorene adduct (*N*-[deoxyguanosin-8-yl]-2-aminofluorene); B conformer, B-type external groove-bound conformer; CD, circular dichroism; DFT, density functional theory; FAF, fluoroaminofluorene adduct (*N*-[deoxyguanosin-8-yl]-7-fluoro-2-aminofluorene); ^{19}F NMR, ^{19}F nuclear magnetic resonance spectroscopy; QM, quantum mechanics; LC-TOF-MS, liquid chromatography time-of-flight mass spectrometry; S conformer, stacked base-displaced conformer; VDW, van der Waals.

recognition has been shown to be modulated by the adduct size and structure, as well as the neighboring base sequences. This has led to several intriguing recognition hypotheses such as "multipartite" and "indirect read" models (*42*, *43*). Similarly dramatic sequence effects have also been observed on mutational outcomes in both bacterial and mammalian cells (*44*). These results clearly reinforce the importance of understanding flanking sequence effects and the delicate balance between repair and mutation. The inherent sequence dependence of the AF-induced S/B/W heterogeneity could be a key driving force to various repair and mutational outcomes.

In the present study, we have prepared a series of 12-mer DNA duplexes in which the fluorine probe 7′-fluoro-2-aminofluorine (FAF) (Figure 1) is site-specifically incorporated at the central guanines and the immediate flanking bases are varied systemically. The utility of FAF as a useful fluorine structure probe has been well documented (*28*−*31*). Incorporation of a fluorine atom at the remote C7 position of AF has been shown to maintain similar carcinogenicity in various rat tissues including the liver (*8*). The profiles of DNA adduct formation and optical behavior of AF and FAF are also similar. The S/B conformational heterogeneities of these modified duplexes were examined by CD and UV melting curve experiments, as well as dynamic ^{19}F NMR spectroscopy. The van der Waals (VDW) energy and quantum mechanics (QM) calculations were carried out in order to gain theoretical insight into the observed sequence effects on S/B heterogeneities. We also measured the NER activities of the modified sequences in the *E. coli* UvrABC system, and the results were examined in terms of the S/B population ratios obtained from the dynamic ^{19}F NMR data. Our results showed that both the base−carcinogen stacking interaction and solvation play a critical role in the sequence effect on the AF heterogeneity. The ^{19}F NMR/NER correlations suggested that the S conformer is more susceptible to repair than the B conformer, and that excision efficiency is preferred for those with dC over dA on the 5′-side of the lesion. Taken together, the results provide novel insights into how bulky carcinogen−DNA adducts are accommodated by NER proteins in a conformationally specific manner. A preliminary account of the present work has been communicated (*45*).

EXPERIMENTAL PROCEDURES

CAUTION! *2-Aminofluorene derivatives are mutagens and suspected human carcinogens and must be handled with caution.*

Crude oligodeoxynucleotides in 10−15 μmol scales in desalted form were obtained from Sigma-Genosys (The Woodlands, TX). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Synthesis and Purification of FAF-Modified Oligodeoxynucleotides. We prepared two sets of 12-mer oligonucleotides (namely, -AG*N*- series, d[CTTCT**AG*****N***CCTC]; -CG*N*- series, d[CTTCT**CG*****N***CCTC]), in which the FAF-modified guanines (G*) are flanked by four natural bases (N = G, A, C, T] (Figure 1c). Additional FAF adduction occurred when *N* was G as in the -AGG*- and -CGG*- sequences.

Preparation of FAF-modified oligodeoxynucleotides was conducted using the procedures described previously (*28*, *29*). Briefly, unmodified oligodeoxynucleotides were treated with *N*-acetoxy-*N*-(trifluoroacetyl)-7-fluoro-2-aminofluorene in a pH 6.0 sodium citrate buffer and placed in a shaker for 18−24 h at 37 °C (Supporting Information (SI) Figure S1). Typical HPLC profiles and their on-line UV spectra are shown for the -CGG- sequence (Supporting Information Figures S2, S3). The early eluting peak at 19.5 min is that of an unreacted oligodeoxynucleotide. Two late eluting ones at 32.8 and 36.7 min exhibited shoulder absorptions in the 300−350 nm range, which are characteristic for modification at the C8 of dG. Purification of the modified oligodeoxynucleotides up to >97% purity was accomplished by repeated injections onto reverse phase HPLC (SI Figure S2). The HPLC system consisted of a Hitachi EZChrom Elite unit with a L2450 diode array as a detector and employed a Waters XTerra MS C18 column (10 × 50 mm, 2.5 μm) with a 60 min gradient system involving 3 to 15% acetonitrile in pH 7.0 ammonium acetate buffer (0.10 M) with a flow rate of 2.0 mL/min. A typical yield for modification was in the range of 20−50% depending on sequences.

Enzyme Digestion/HPLC. The FAF-modified 12-mer oligonucleotides were determined to be C8-substituted guanine adducts by a standard sequential enzyme digestion procedure to 2′-deoxynucleosides and cochromatography with the standard FAF adduct, *N*-[deoxyguanosin-8-yl]-7-fluoro-2-aminofluorene (*28*). Each sample (∼1 OD) was dissolved in 3 mL of Bis-Tris- EDTA (pH 7.0) buffer and 0.5 mg of DNase and incubated at 37 °C with shaking for 3−4 h. Snake venom phosphodiesterase I (100 μL, ∼0.05 unit) and alkaline phosphatase (∼2.5 unit) were then added, and the incubation was continued for 1−2 days. The sample was loaded on a Millipore Centricon YM-3 centrifugal filter (Yellow, MW cutoff = 3000) and centrifuged, and the filtrate was extracted three times with H_2O-saturated 1-butanol. The butanol fractions were combined, back-washed with water, and then evaporated to dryness using a SpeedVac (Thermo Savant). The sample was dissolved in methanol and subjected to reverse phase HPLC. Supporting Information Figure S3 shows a typical HPLC chromatogram (-CG*G- sequence) of the resulting digests and the HPLC conditions.

Sequence Analyses by Enzyme Digest/LC-TOF-MS. The FAF-modified oligonucleotides were sequenced using a Waters ESI-time-of-flight-mass spectrometer (LC-TOF-MS) in the negative ion mode based on the exonuclease (3′-5′ or 5′-3′) strategies described previously (*29*). Briefly, 5−10 μg of FAF-modified oligodeoxynucleotides was combined with ca. 0.01 unit of a 3′-5′- or a 5′-3′-exonuclease in a 1 mM solution of $MgCl_2$ and incubated for different times up to 2 h. Enzyme digests were separated online using a 10 min water/acetonitrile gradient. The aqueous phase was 5 mM in both ammonium acetate and dimethylbutylamine (pH 5.6), and the acetonitrile was made 0.1% in formic acid. The spray voltage was 3.4 kV. Illustrative assignment details are given in Supporting Information Figures S4−S7 for the FAF-modified -CG*G- and -CGG*- sequences. The molecular ion results of all the eight FAF-modified sequences used in the present study are summarized in SI Table S1.

UV-Melting Experiments. UV-melting data were obtained using a Beckman DU 800 UV/vis spectrophotometer equipped with a 6-chamber, 1 cm path length T_m cell, and a Peltier temperature controller, according to the procedure described previously (*29*). Duplex solutions (0.2−14 μM) were pre-

pared in a pH 7.0 buffer containing 0.2 M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA. Thermodynamic parameters were calculated using the program MELTWIN version 3.5.

Circular Dichroism (CD) Spectra. CD measurements were conducted on a Jasco J-810 spectropolarimeter equipped with a variable Peltier temperature controller (*29*). Typically, 2 ODS of each strand was annealed with an equimolar amount of a complementary sequence in 400 μL of a neutral buffer containing 0.2 M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA. Spectra were acquired using a 1 mm path length cell.

NMR Experiments. Approximately 50−80 ODS of pure modified oligonucleotides were annealed with an equivalent amount of complementary sequences to produce the corresponding fully paired 12-mer duplexes. The duplex samples were centrifugated using a Pall Microsep MF centrifugal device (Yellow, MW cutoff = 1000). The centrifuged samples were dissolved in 300 μL of a neutral buffer (10% D_2O/90% H_2O containing 100 mM NaCl, 10 mM sodium phosphate, and 100 μM tetrasodium EDTA, pH 7.0) and filtered into in a Shigemi tube through a 0.2 μm membrane filter.

All ^{1}H and ^{19}F NMR results were recorded using a 5 mm ^{19}F/^{1}H dual probe on a Bruker DPX400 Avance spectrometer operating at 400.0 and 376.5 MHz, respectively. Imino proton spectra were obtained using phase sensitive jump−return sequences at 5 °C and referenced relative to DSS (SI Figure S8). ^{19}F NMR spectra were acquired in the ^{1}H-decoupled mode and referenced to $CFCl_3$ by assigning external C_6F_6 in C_6D_6 at −164.90 ppm. Long-term acquisition with ^{1}H-coupling did not improve signal-to-noise ratios. One-dimensional ^{19}F NMR spectra were measured between 5 °C and 60 °C with increments of 5 and 10 °C. Additional temperatures were used as needed to clarify signal exchange process (see figure legends). Temperatures were maintained by a BRUKER-VT unit using liquid N_2. Spectra were obtained by collecting 65,536 points using a 37,664 Hz sweep width and a recycle delay of 1.0 s. A total of 1600 scans were acquired for each dynamic NMR spectrum. High quality spectra at 20 °C were obtained in separate acquisitions using 4000 scans (Supporting Information Figure S9). All FIDs were processed by zero-filling, exponential multiplication using a 20 Hz line broadening factor and Fourier transformation. The S/B population ratios were obtained by area integration of the base-line corrected spectra (Bruker, Billerica, MA). NOESY/exchange ^{19}F NMR spectra were obtained in the phase-sensitive mode using the following parameters: sweep width 4529 Hz, number of complex data points in t_2 1024, number of complex FIDs in t_1 256, number of scans 96, dummy scans 16, recycle delays 1.0 s, and mixing time 400 ms. The data were apodized with sine function using 2 Hz line broadening in both dimensions and Fourier transformed with the 1024 × 256 data matrix.

Complete Line Shape Analysis. Complete line-shape analysis was carried out using the simulation program WINDNMR-Pro (version 7.1.6, *J. Chem. Educ*. Software Series; Reich, H. J., University of Wisconsin, Madison, WI). Examples of simulation are shown in Supporting Information Figure S10 for the -AG*N- duplex series. The values of frequencies and the S/B-population ratios were determined at the slow exchange limit (5 °C). Subsequently, several spectra were recorded at various temperatures between 5 and 60 °C including at coalescence and into the fast exchange region. The samples were then cooled back to the slow exchange limit to ensure that no irreversible process have occurred at the higher temperatures (*46*).

Modeling. The NMR structures of the S- and B-conformeric AF-modified 11-mer duplexes (*19*, *20*) were adjusted to that of the target 12-mer sequence (Figure 1c). The hydrogen atom on C7 of AF was replaced with a fluorine atom (Supporting Information Figure S11). The additional base pair was added with Insight II (Accelrys Software, Inc.), using classic B-DNA conformation as a guide. Partial charges in S conformer were obtained with Gaussian 03 (Gaussian, Inc.) (*47*). Force field parameters for FAF-dG were obtained by finding the closest analogies in the parm99 (*48*) and GAFF (*49*) force fields, and full details are given in Supporting Information Table S5. HF calculations with the 6-31G* basis set were used to calculate the electrostatic potential using Gaussian 03 (*50*), and the restrained electrostatic potential (RESP) fitting algorithm was employed to fit the charge to each atom center (*47*). Full details of RESP results are given in Supporting Information Tables S6 and S7. Molecular dynamics (MD) simulations were carried out with AMBER 8.0, employing the modified Cornell et al. force field parm99 version (*51*).

Van der Waals (VDW) Energy Analysis. 5 ns MD simulation was carried out for each system. The MD structures stabilized after 2 ns with the 2−5 ns time frame used for analysis. The unstacked complementary dC of the S conformer displaced into the major groove was excluded from the calculation, similar to the FAF moiety in the major groove of the B conformer. The VDW stacking interactions were computed with the program ANAL of AMBER 8. The stacking interaction energies between G-FAF and its adjacent Watson−Crick base pairs were evaluated for both S and B conformers. A total of 3000 snapshots were utilized. MOILview was used to do a cluster analysis of the 2−5 ns trajectory to identify the most representative structures (*52*), representing the lowest average root-mean-square deviations (rmsd) to all of the others in the cluster. The average values and standard deviations are given in Supporting Information Tables S3 and S4.

Quantum Mechanics (QM) Calculations

Geometries. QM calculations were carried out based on MD structures. The G*:C modified pair and its two flanking base pairs were dissected for QM calculations, and all nucleobases were terminated by hydrogen atoms at the glycosidic positions (Figure 2). Each "AMBER" nucleobase geometry was replaced (overlay via rms fit using heavy atoms) with an in vacuo QM-optimized planar monomer. Then, an additional constrained gradient optimization was carried out for the hydrogen-bonded base pairs (each pair separately) with the positions of the heavy atoms frozen and the hydrogen atoms relaxed. Optimizations were carried out in Gaussian 03 (*50*) using B3LYP (*53*) density functional and 6-311G(d,p) basis set.

Interaction Energy Calculations. Interaction energy ΔE^{AB} of a dimer (the individual base pair stack or pair) A-B is the difference between the electronic energy of the dimer E^{AB} and the electronic energies of the infinitely separated monomers, i.e., $\Delta E^{AB} = E^{AB} - E^{A} - E^{B}$ (*33*, *36*).

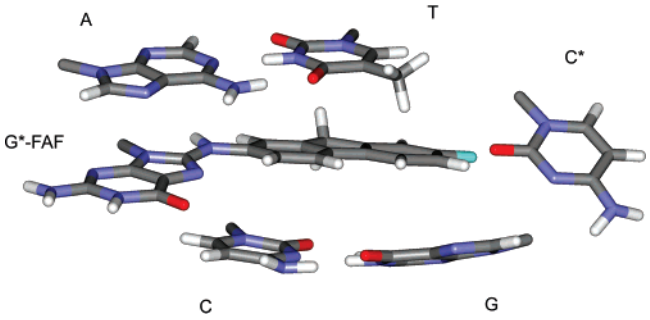

FIGURE 2: A model for quantum mechanic calculations of the S conformer -AG*C- duplex. G*-FAF, FAF-modified G; C*, complementary C opposite the lesion.

The total interaction energy for a given conformer and sequence was calculated as a sum of dimer interaction energies considering all pairs of interacting monomers. Thus, the total interaction energy of the -AG*C- sequence in its S-conformer ΔE(S) (Figure 2) was obtained as a sum of nine interactions: A-T, A-G*[FAF], T-G*[FAF], T-C*, G*[FAF]-C*, G*[FAF]-C, G*[FAF]-G, C*-G, and C-G (C* stands for C opposite the adduct). The in vacuo (intrinsic) interaction energies were calculated using the DFT−D method with empirical dispersion energy correction (*54*) with TPSS exchange-correlation functional (*55*), 6-311++G(3df,3pd) basis set and the B-0.96-27 dispersion parameters using TurboMole 5.7 code (*56*). The gas phase interaction energy calculations were followed by evaluation of the solvation energy of the whole model (three base pairs) with continuum solvation calculation (C-PCM method with the B3LYP/6-31G* solute description) using the Gaussian 03 program (*50*).

Interactions in the corresponding B structures, ΔE(B), were determined in the same way, and the interaction energy difference between S and B conformers was calculated as $\Delta\Delta E$(S−B) = ΔE(S) − ΔE(B). We present only the differences $\Delta\Delta E$(S−B) while all dimer interaction terms and solvation energies are given in SI Table S8. The main sources of possible errors in the QM calculations are the approximations of the solvation model. However, these errors for S and B conformers should largely cancel out to give reasonable S−B differences, provided the MD simulations are sufficiently representative.

Nucleotide Excision Repair Assays

NER assays were conducted using the *E. coli* UvrABC nuclease system. DNA duplex substrates were constructed by ligating an AF- or FAF-12-mer (CTTCTAG*NCCTC), with a flanking 20-mer (GACTACGTACTGTTACGGCT) and a 19-mer (GCAATCAGGCCAGATCTGC) oligonucleotide on the 5′ and 3′ sides, respectively (*57*−*59*). The 20-mer was 5′-terminally labeled with ^{32}P. The 5′-terminally labeled DNA substrates (2 nM) were incised by UvrABC (UvrA, 15 nM, UvrB, 250 nM, and UvrC, 100 nM) in the UvrABC buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 1 mM ATP) at 37 °C in a time-course dependent manner (*60*).

Quantitative data of radioactivity were obtained using Fuji FLA-5000 Image Scanner. The amount of DNA incised (in pmol) by UvrABC was calculated based on the total molar amount of DNA used in each reaction and the percentage of radioactivity in the incision products as compared to the total radioactivity. At least three independent experiments were performed for determination of the rates of incision.

RESULTS

Circular Dichroism

Circular dichroism (CD) results show sequence-dependent differences in the S/B population ratios. Figures 3a and 3b show overlays of CD spectra of the FAF-modified -AG*N- and -CG*N- duplex series, respectively. The profiles of FAF-induced ellipticities in the 290−360 nm range ($ICD_{290-360nm}$) were uniquely sequence dependent, similar to those observed for the non-fluoro AF duplexes reported previously (*29*). Table 1 summarizes the CD patterns in terms of the $CD_{-/+}$ values, which have been defined as the ratio of the negative over the positive areas between 250 and 360 nm (*29*). The results indicated that all the FAF duplexes adopted global B-DNA conformations. However, they differed in the extent of the carcinogen-induced ellipticity between 290 and 360 nm: *i.e.*, the greater the $CD_{-/+}$ values, the greater the population of B-type conformer (*30*). In particular, the sequences with a thymine base 3′ to the lesion exhibited the largest negative induced shifts, 0.85 and 0.39, for -AG*T- and -CG*T- duplexes, respectively. These CD results were in accordance with the ^{19}F NMR and van der Waals energy (VDW) data (see below). The general similarity of the CD patterns between the FAF and AF duplexes indicated that the replacement of the remote C7 hydrogen with fluorine did not change the overall adduct conformation. Thus, FAF was an excellent probe for conformational analysis of AF by ^{19}F spectroscopy.

UV-Thermomelting Data

As in the CD experiments, the thermodynamic profiles (Table 1) of the FAF-modified duplexes were found to be similar to those of the AF-analogues. The effects of the FAF lesions on the thermal (ΔT_m) and thermodynamic ($-\Delta\Delta G°$ and $-\Delta\Delta H°$) stabilities were clearly sensitive to the immediate flanking base sequences (*61*). An excellent correlation (R^2 = 0.90) (not shown) between $-\Delta G°$ and T_m indicated that the modified duplexes followed an adduct-induced destabilization of the DNA secondary structure at or near the adduct site due to the loss of base stacking and base pairing at or near the lesion site. As expected, the duplex stability was typically sensitive to the nature and the polarity of base pairs neighboring the lesion site (*29*, *62*). Therefore, the T_m and $-\Delta G°$ values of duplexes with a 3′-flanking G:C base pair were consistently higher than those with an A:T pair (*i.e.*, -AG*A-, -AG*T-, -CG*A-, and -CG*T-). Consequently, the -CG*N- duplexes exhibited consistently greater T_m and $-\Delta G°$ stabilities than the -AG*N- series (Table 1). These results indicated the importance of the total number of hydrogen bonds available at the flanking region for duplex stabilization. In the case of the AF analogues, we reported a general inverse trend (R^2 = 0.675−0.826) between $T_m/-\Delta G°$ and % B conformer and suggested the importance of the carcinogen stacking interaction in the enhanced stability of the S conformer (*29*). No such correlations were observed for the FAF-modified duplexes in the present study. The reason for this discrepancy is not clear and could be complex. It seems that a major contributing factor is the aforemen-

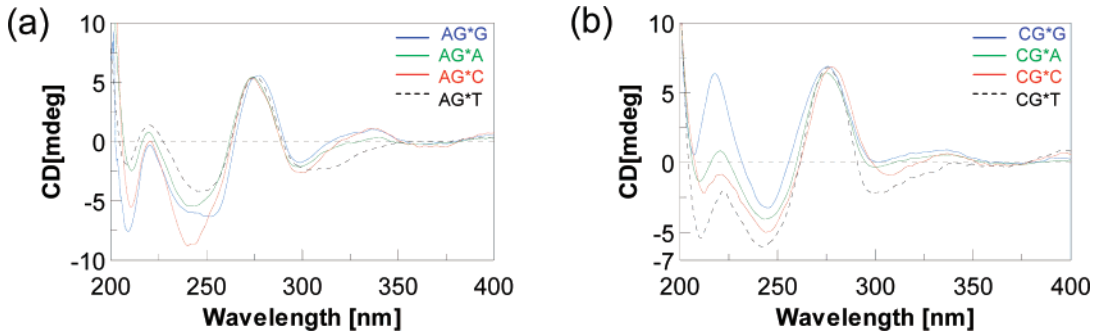

FIGURE 3: CD spectral overlays of the FAF-modified duplexes in the 200−400 nm region at 15 °C for (a) the -AG*N- and (b) -CG*N- duplex series.

Table 1: Effects of FAF-modification on the Thermal and Thermodynamic Stability of DNA Duplexes

sequence[a]	$-\Delta G^b$ (kcal/mol)	$-\Delta H^b$ (kcal/mol)	$T_m{}^c$ (°C)	$\Delta\Delta G^d$ (kcal/mol)	$\Delta\Delta H^e$ (kcal/mol)	$\Delta T_m{}^f$ (°C)	$CD_{-/+}{}^g$ (nm)
-AG*G- -TC C-	9.8 (12.7)	75.8 (101.0)	51.0 (56.2)	2.9	25.2	−5.2	0.20
-AG*A- -TC T-	8.2 (11.1)	63.5 (87.8)	45.6 (54.2)	2.9	24.3	−8.6	0.36
-AG*C- -TC G-	9.1 (11.0)	63.4 (74.4)	50.4 (56.6)	1.9	11.0	−6.2	0.39
-AG*T- -TC A-	8.4 (10.2)	68.6 (76.8)	45.6 (52.8)	1.8	8.2	−7.2	0.85
-CG*G- -GC C-	10.2 (13.5)	73.8 (87.8)	54.1 (63.5)	3.3	14.0	−9.4	0.11
-CG*A- -GC T-	9.4 (11.7)	75.2 (77.7)	49.2 (59.8)	2.3	2.6	−10.6	0.08
-CG*C- -GC G-	10.0 (12.7)	67.2 (74.2)	54.5 (64.4)	2.7	7.0	−9.9	0.06
-CG*T- -GC A-	9.3 (11.7)	72.3 (76.0)	49.9 (59.6)	2.4	3.7	−9.7	0.39

[a] The central trimer portion of the 12-mer duplex (G* = FAF adduct, see Figure 1). [b] The results of curve fit and T_m-lnC_t dependence were within ±15% of each other, and therefore these numbers are averages of the two methods. The average standard deviations for -Δ*G*°, -Δ*H*°, and T_m are ±0.22, ±6.33, and ±0.4, respectively. The values in parentheses are for the unmodified control duplexes. [c] T_m values at 14 μM taken from the $1/T_m - \ln C_t/4$ Meltwin plots . [d] $\Delta\Delta G = \Delta G°$ (FAF-modified duplex) − $\Delta G°$ (control duplex). [e] $\Delta\Delta H = \Delta H°$ (FAF-modified duplex) − $\Delta H°$ (control duplex). [f] $\Delta T_m = T_m$ (FAF-modified duplex) − T_m (control duplex). [g] Ratio of the negative over the positive areas in the 250−360 nm region (*29*) (see Figure 3).

tioned flanking sequence effect, not the AF/FAF swap, which has been demonstrated to be conformationally compatible at various sequence contexts (*8*).

^{19}F NMR Spectroscopy

Signal Assignments. The S/B ^{19}F signal assignments were made using the general strategy described previously (*28*, *30*, *31*). Figure 4 shows dynamic ^{19}F NMR spectra of the -AG*N- and -CG*N- duplexes recorded in the −115 to −121 ppm range between 5 and 60 °C. All eight FAF-modified duplexes exhibited two major well-resolved ^{19}F signals at or below 20 °C, which represented unique ^{19}F environments, namely, the S and B conformation (see below). S/B heterogeneity was evidenced by the varying intensities of the imino proton signals in the ^{1}H NMR spectra (Supporting Information Figure S8). The low intensity imino proton signals at 10−12 ppm usually represent those of the modified dG at the lesion site and the nearby base pairs. The slow interconversion of the two conformers in each duplex was confirmed by temperature dependent NOESY/exchange spectroscopy. For example, Figures 5a and 5b show the NOESY/exchange spectra of the -AG*T- duplex taken at 5 and 20 °C, respectively. Strong temperature dependence of the off-diagonal cross-peaks implied a transfer of magnetization from one peak to another, indicating that the two conformers (S and B) originated from the same chemical structure (*28*). Similar NOESY profiles were observed for all the duplexes studied in the present study.

Specific S/B signal assignments were made based on ring current and H/D isotope effects (*28*, *31*). The S conformer was more susceptible to a ring current effect, which enabled the base-displaced FAF to be better shielded than the B conformer. The rationale for the H/D effect was that the ^{19}F resonance of the exposed FAF residue in a B conformer is more susceptible to solvent-induced shielding (usually >0.2 ppm) than the buried FAF in an S conformer (usually <0.1 ppm) upon increasing the deuterium content from 10% to 100% (*63*). The H/D isotope effects were measured at 20 and 60 °C for all the duplexes studied, and the results are summarized in Supporting Information Table S2. In all cases, the downfield B conformer signals exhibited consistently greater H/D shielding effects (0.13−0.22 ppm) than the upfield S conformer signals (0.01−0.06 ppm) at 20 °C.

Sequence-Dependent S/B Heterogeneity. S/B heterogeneity was markedly sequence-dependent, especially on the 3′-flanking bases. Supporting Information Figure S5 compares the ^{19}F NMR spectra of the -AG*N- and -CG*N- duplex series acquired at 20 °C in a stacked format. Specifically, the population of S conformer for the -AG*N- series decreased in order of 3′-dG (68%) > dA (61%) ≫ dC (40%)

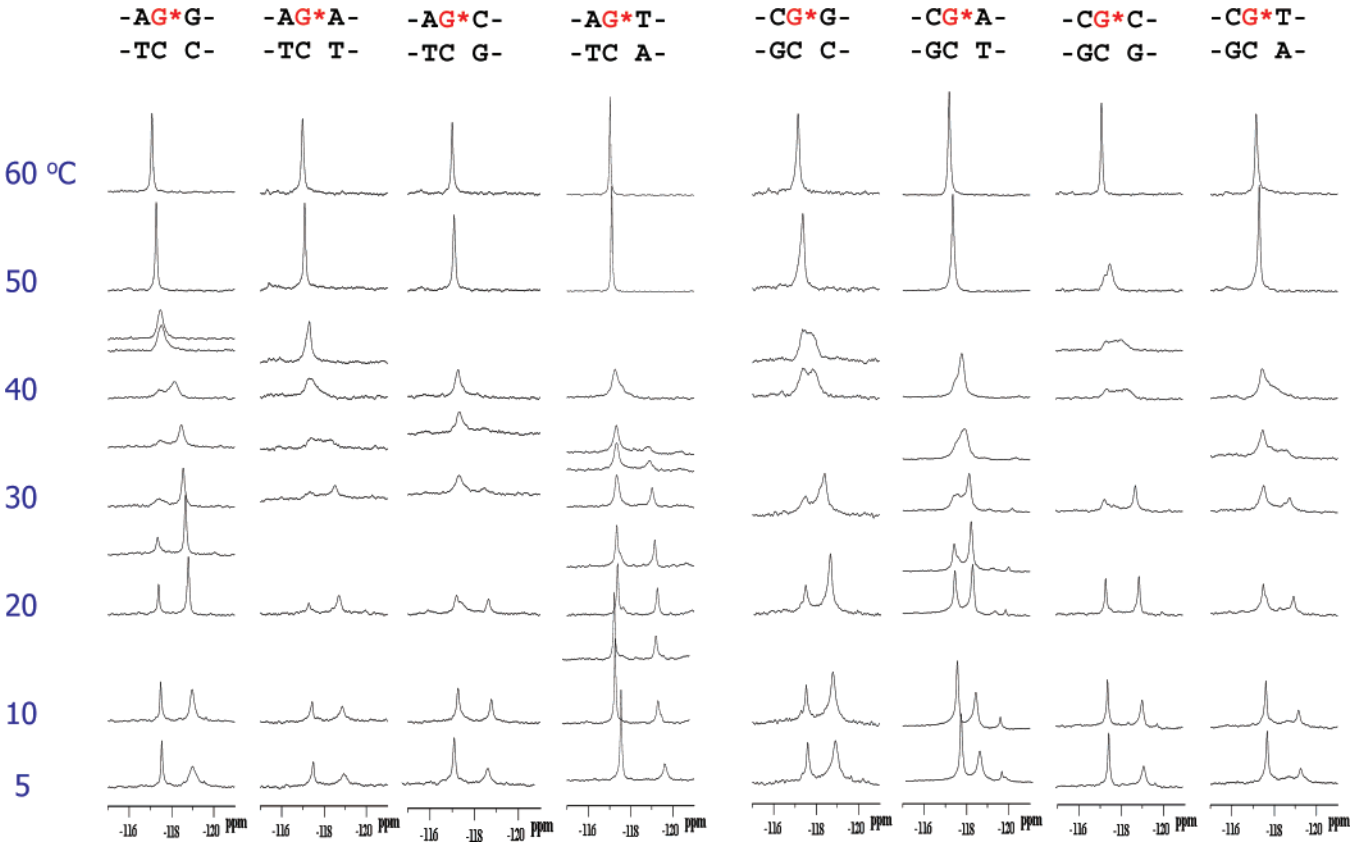

FIGURE 4: Dynamic ^{19}F NMR spectra (−115 to −121 ppm) of the -AG*N- and -CG*N- duplexes (G* = FAF adduct, N = G, A, C, T). Spectra of all sequences were obtained at seven standard temperatures (labeled as 5, 10, 20, 30, 40, 50, and 60 °C). Additional temperatures were recorded for -AG*G (25, 35, 45, 47 °C), -AG*A- (35, 43 °C), -AG*C- (37 °C), -AG*T- (15, 25, 33, 35 °C), -CG*G- (45 °C), -CG*A- (25, 35 °C), -CG*C- (45), and -CG*T- (35 °C).

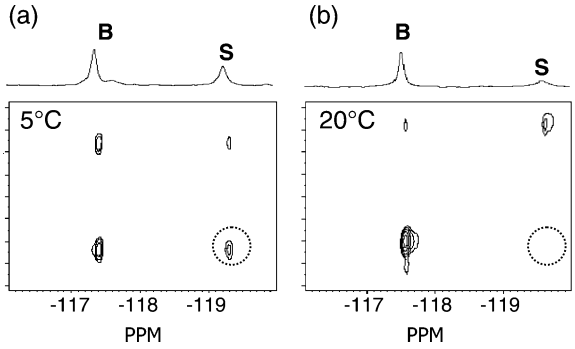

FIGURE 5: Temperature dependent ^{19}F NMR NOESY spectra of the -AG*T- duplex recorded in 10% D_2O/90% H_2O pH 7.0 buffer at (a) 5 °C and (b) 20 °C. B = B-type conformer; S = stacked conformer (see Figure 1). Off diagonal cross-peaks are shown in the dotted circles.

> dT (36%). A similar trend was obtained for the -CG*N- series, 3′-dG (66%) > dA (56%) ∼ dC (53%) ≫ dT (37%).

Sequence-Dependent ^{19}F NMR Chemical Shifts. The ^{19}F signals for the B conformer in both the -AG*N- and -CG*N- series appeared at less than 0.13 ppm (Supporting Information Figure S9 and SI Table S2). This was in contrast to the S conformer signals in the same contexts, which exhibited significant deshielding (0.3−0.7 ppm) going from 3′-T to A or G. The results clearly indicated the sensitivity of the S conformer ^{19}F chemical shifts to the nature of the 3′-flanking base. The stacking efficiencies between the carcinogen and the 3′-flanking base play an important role (see below). The largest chemical shift difference between the two signals was observed for the -AG*T- duplex.

Thermodynamics of the FAF-Modified Duplex System. An overall thermodynamic picture of the FAF-modified duplex system can be characterized by two different equilibriums: first, the equilibrium between the S and B conformers in the duplex, and second, a melting equilibrium between the duplex and single strands. Dynamic ^{19}F NMR data (Figure 4) provided valuable kinetic and thermodynamic information on both equilibria (Table 3, see below). Description of the thermodynamic diagram and assumptions undertaken in the present study are given in Supporting Information Figure S12. The dynamic NMR results showed that the S signal increased with increasing temperatures, i.e., $\Delta\Delta H$(S−B) was positive. The S conformer was enthalpically disfavored, but entropically favored with respect to the B conformer. Using the van't Hoff equation, the values of $\Delta\Delta H$(S−B) were estimated to be about 2−6 kcal/mol, which depended on the sequence contexts. These values may be inaccurate since the temperature dependence of the S signal was obscured by an unknown heat capacity difference between the S and B conformers, and melting of the duplex at higher temperatures.

Dynamic ^{19}F NMR Spectra. All eight FAF duplexes exhibited a typical two-site chemical exchange behavior (Figure 4). The S/B ^{19}F signals were in slow exchange at or below 5 °C, but upon raising the temperature they broadened and moved closer together, and eventually coalesced between 40 and 55 °C. Sharp signals at 60 °C were due to the fast rotating modified G* FAF residues of denaturing duplexes (see below), and substantial H/D isotope effects (0.16−0.20 ppm) were observed (SI Table S2). The B conformer signals of certain duplexes (i.e., -AG*C-, -AG*T-, and

Table 2: Dynamic NMR and Thermodynamic Parameters for the FAF-modified 12-mer Duplexes

sequence[a]	% B[b]	% S[b]	$\kappa_{30°C}$ (s^{-1})[c]	$\tau_{30°C}$ (ms)[d]	k_C (s^{-1})[e]	$\Delta G°_{30°C}$ (cal/mol)[f]	$\Delta G^{‡}_{30°C}$ (kcal/mol)[g]
-AG*G- -T CC-	32	68	200	5	1214	−438	14.6
-AG*A- -TC T-	39	61	191	5.2	1202	−94	14.6
-AG*C- -TC G-	60	40	263	3.8	1240	388	15.4
-AG*T- -TC A-	64	36	200	5.0	1715	334	14.6
-CG*G- -GC C-	36	64	435	2.3	1097	−419	14.1
-CG*A- -GC T-	44	56	450	2.2	739	−140	14.1
-CG*C- -GC G-	47	53	200	5.0	1377	−70	14.6
-CG*T- -GC A-	63	37	446	2.2	1321	309	14.1

[a] The trimer core sequence of the FAF-modified 12-mer duplexes used in this study. See Figure 1c for full sequence contexts. G* = FAF adduct. [b] % population of B and S conformers by integration of ^{19}F NMR signals at 20 °C (SD = ±3%). [c] Rate constant. (SD ± 25%.) The data were obtained by complete line shape analysis of temperature dependent ^{19}F NMR results using the WINDNMR-Pro (see Experimental Procedures). [d] Exchange time ($1/k$) indicates the amount of time the adduct spends in one conformation before jumping into another conformation. [e] Rate constants at a coalescence temperature, a lower limit on the exchange rate between the two conformers. $k_C = 2.22\Delta\upsilon$ (the chemical shift difference between the two signals in Hz at slow exchange, i.e., at 5 °C). [f] The energy difference between the two conformers: $\Delta G° = -RT \ln K$ [K_{eq} = S/B]. [g] The S/B interconversion energy barrier @ 30 °C obtained from Eyring equation (*46*): $\Delta G^{‡}_C = 4.58T_C$ (10.32 + log T_C/k_C) cal/mol. SD ± 0.2 kcal/mol.

-CG*T-) were split between 20 and 30 °C, suggesting the existence of additional B-like conformer intermediates (B*). A similar observation was made in the primer-template sequence contexts (*31*).

It should be noted that for all sequences the intensity of the downfield fluorine signal initially decreased with increasing temperatures and then increased again. A key to understanding this behavior is to realize that the downfield signal belonging to the ^{19}F exposed to solvent is a sum of the signals coming from the B conformer and denatured modified single strands. This is depicted in Figure 6, which captures the S/B (X_S, red curve) and melting equilibria (melting curve, black line) for the -AG*G- duplex as modeled by the van't Hoff equation. The diagram is only an illustration of principle, and the thermodynamic characteristics used to generate it are approximate and were chosen to resemble the steepness of the X_S curve at low temperatures (e. g., $\Delta\Delta H$(S−B) ∼5 kcal/mol); the real behavior in ^{19}F NMR (Figure 4) is somewhat obscured by nonideal DNA melting etc. The blue and orange curves represent signals of ^{19}F exposed to solvent (X_B) and shielded (X_S) by the flanking base pairs, respectively. At low temperatures, the ratio of the two signals corresponded primarily to the S/B equilibrium, whereas at high temperatures it represented exclusively the melting process.

We conducted lineshape analysis by assuming two-site exchange for all the dynamic ^{19}F NMR spectra (Figure 4, see Supporting Information Figure S10) (*46*). The resulting kinetic and thermodynamic parameters are summarized in Table 2. The S/B interconversion energy barriers ($\Delta G^{‡}$) obtained from the Eyring equation, $\Delta G^{‡}_C = 4.58T_C$(10.32 + log T_C/k_C) cal/mol, at 30 °C was found to be 14−15 kcal/mol. The energy differences ($\Delta G° = -RT \ln K_{eq}$) of the two conformers were relatively small (<500 cal/mol). The coalescence temperature (T_C) is defined as the temperature at which the two signals are merged into a single, flat-topped peak (*46*). Determination of exact T_C values for each duplex was difficult since sufficient numbers of dynamic spectra were not obtained. A close inspection of emerging signals, however, indicated that T_C decreased in the order of -CG*C- > -CG*G- > -AG*G-, -CG*A- > -AG*A-, -AG*C-, -AG*T-, -CG*T-. These results were consistent with the T_m trend (Table 1) and appeared to be in line with the number of Watson−Crick hydrogen bonds surrounding the lesion site, i.e., the -CG*C- and -CG*G- sequences have three hydrogen bonds on both sides of the lesion (Table 1). The exchange rate constants at T_C ($k_C = 2.22\Delta\nu$, $\Delta\nu$ = the separation in frequency in Hz between the two conformer signals at 5 °C, when dynamic exchange is minimal) were determined to be in the range of 739−1715 s^{-1}. The correlation times ($\tau = 1/\kappa$) of the modified duplexes were found to be in the narrow range of 2.2−5.2 ms at 20 °C and signify the amount of time the lesion spends in one conformation before changing to another conformation.

Quantum Mechanics (QM) Calculations

Full computational evaluation of the free energy changes is unfeasible, due to major methodological uncertainties, even when resorting to the force field description (*36*, *38*). In addition, the variation of % S conformer corresponded to a few tenths of a kcal/mol of the free energy change between S and B (Table 2). Therefore, instead of evaluating the ΔG directly we focused on some of the key contributions to the free energies, with the aim to obtain theoretical insight into the forces governing the S/B equilibrium. Fairly accurate quantum mechanical calculations can be performed to evaluate interaction (base stacking and base pairing) and solvation energies, which can be correlated with the interaction enthalpy or other measurable quantities.

The NMR S/B Ratios and QM Calculations. First, we determined whether the $\Delta\Delta H$(S−B) calculated using the QM methods correlates with the S/B ratio (more precisely, with ln(S/B)) from the NMR experiment. The S/B ratio was determined by the free energy difference between S and B. The underlying assumption for $\Delta\Delta H$(S−B) to correlate with the S/B ratio is that the entropy change between S and B is similar for all four sequences in a series. If this were the case, the $\Delta\Delta G$(S−B) trend would be given primarily by the enthalpy change $\Delta\Delta H$(S−B). More details about the model are given in Supporting Information Figure S12. The results of our QM calculations and the enthalpy difference between S and B are summarized in Table 3. The correlation coefficients (R^2) of the calculated interaction energy differences with ln(S/B) were 0.86 and 0.44 for the -AG*N- and -CG*N- series, respectively. With inclusion of the continuum solvent correction, R^2 were 0.63 and 0.57 for the -AG*N- and -CG*N- series, respectively. Thus, the S/B ratio is only marginally correlated with the base interaction energies.

The *in vacuo* interaction energy loss ($\Delta\Delta E_{int}$(S−B)) of B to S conversion was clearly due to the loss of hydrogen bonding at the lesion site, which ranged between 17.6 and

Table 3: Calculated Quantum Mechanics (QM) Stabilization Energies (kcal/mol) of the FAF-Modified DNA Duplexes

		$\Delta\Delta E$(S−B)[c]				
sequence[a]	$\Delta\Delta E_{int}$(S−B)[b]	DFT only	disp only	$\Delta\Delta E_{solv}$[d]	$\Delta\Delta E_{tot}$(S−B)[e]	$\Delta\Delta$E(S−B)[f] weighted
-AG*G- -TC C-	17.6	17.8	−0.2	−9.3	8.3	5.6
-AG*A- -TC T-	21.9	18.0	3.9	−13.8	8.1	4.9
-AG*C- -TC G-	24.0	22.6	1.4	−15.0	8.9	3.6
-AG*T- -TC A-	25.2	20.2	5.1	−14.2	11.1	4.0
-CG*G- -GC G-	19.4	20.8	−1.4	−17.2	2.2	1.4
-CG*A- -GC T-	24.7	24.4	0.3	−25.4	−0.6	−0.4
-CG*C- -GC G-	23.3	23.7	−0.4	−18.2	5.1	2.7
-CG*T- -GC A-	24.2	24.1	0.1	−16.7	7.5	2.8

[a] The trimer core sequence of the FAF-modified 12-mer duplexes used in this study. See Figure 1c for full sequence contexts. G* = FAF adduct. [b]Difference between stabilization energies of S and B conformers calculated by TPSS/6-311++G(3df,3pd)/B-0.96−27 DFT + dispersion method. [c]DFT and empirical dispersion components of the total interaction energy difference form the previous column. [d]Difference between solvation energies of S and B conformers (S − B), calculated with COSMO/B3LYP/6-31G*, electrostatic contribution only. [e]$\Delta\Delta E$(S−B) = $\Delta\Delta E_{int}$(S−B) + $\Delta\Delta E_{solv}$. [f]$\Delta\Delta E$(S−B) corrected for the expected population of the S structure at the melting temperature; to be compared with $\Delta\Delta H$, see text.

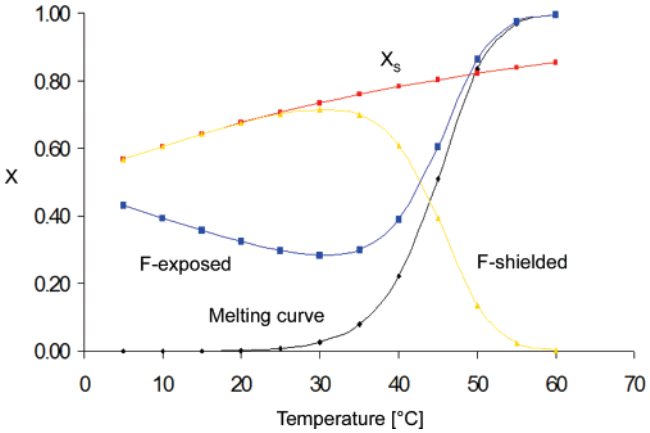

FIGURE 6: An illustration of temperature dependence of the S/B (X_S, orange) and melting (melting curve, black) equilibria for the -AG*G- duplex modeled by van't Hoff equation. The F-shielded (yellow) signal represents S conformers. The F-exposed (blue) signal is a sum of the B conformer and denatured FAF-modified single strand signals. See text.

25.2 kcal/mol for the -AG*N- series and 19.4−24.2 kcal/mol for the -CG*N- series. This roughly corresponded to the interaction energy of the G:C pair *in vacuo*. The DFT−D interaction energy can be decomposed to the DFT part (Table 3, column 3) and the dispersion part (column 4). The former covered the electrostatic interaction, polarization contribution, and short-range repulsion between interacting molecules. The latter was the attractive London dispersion interaction which usually dominates aromatic stacking. Only small differences in dispersion energies (Table 3, column 4) indicated that the dispersion interaction lost in the original B duplex was almost completely compensated by the dispersion stabilization of stacking of the FAF moiety with its flanking pairs in the S conformer. Therefore, the S destabilization was due almost exclusively to the loss of hydrogen bonding.

In an aqueous environment, the energy penalty for S conformer coming from the loss of hydrogen bonding is largely diminished because of the hydration gain corresponding to hydration of the guanine and cytosine Watson−Crick edges exposed to solvent in the S conformation (see Table 3, column 5, and the sum of the energy differences and hydration corrections in column 6). For instance, in the case of -CG*T- sequence (last line) the dispersion contribution to S destabilization *in vacuo* is only 0.1 out of 24.2 kcal/mol. Then, the difference in hydration energies of S and B is −16.7 kcal/mol in favor of S, and the resulting S energy destabilization thus falls to 7.5 kcal/mol only. The final energy destabilization still originates in the lost hydrogen bonding, but is much smaller than *in vacuo*.

Sequence Dependence of the Melting Enthalpy Differences. Next, we estimated the difference between the melting enthalpy of the modified and unmodified duplexes and compared them with van't Hoff thermodynamic results. The melting enthalpy was not calculated directly because our calculations were hampered by our inability to estimate the enthalpy of the melted single strands, among other issues (*38*). However, as explained in Supporting Information Figure S12, the $\Delta\Delta H$ of melting was approximated by the enthalpy difference between the S and B conformers: $\Delta\Delta H \cong \Delta\Delta E$(S−*B*)weighted = $X_S\Delta\Delta E$(S−B), where X_S is the molar fraction of S at the melting temperature. Because we did not know X_S, we used the values at 20 °C. The corresponding errors should be moderate, and the resulting $\Delta\Delta E$(S−B) values are given in Table 3 (last column). The magnitude of the calculated $\Delta\Delta E$(S−B) corresponded well with the measured $\Delta\Delta H$ of melting (Table 1), but no correlations were observed due to significant error with both the experimental and theoretical values.

The standard deviation of the experimentally measured ΔH of melting was estimated to be 6.3 kcal/mol and may be larger for the resulting $\Delta\Delta H$ values. In particular, experimental values above 20 kcal/mol seemed to be overestimated as they were much larger than both the theoretical and NMR estimates. Smaller $\Delta\Delta H$ were to be expected from the nearest neighbor model, in which the ΔH associated with breaking one base pair is about 9 kcal/mol. If we assume that there is only one base pair broken in the S structure (i.e., the G[FAF]:

Table 4: Van der Waals Base Stacking Interaction Energies (kcal/mol)

sequence[a]	S conformer[b]	B conformer[b]	% S[c]
-AG*G-	−30.05 ± 2.41	−29.72 ± 1.92	68
-AG*A-	−32.04 ± 2.46	−30.03 ± 1.93	61
-AG*C-	−30.89 ± 2.29	−29.92 ± 2.02	40
-AG*T-	−28.66 ± 3.39	−30.09 ± 1.90	36
-CG*G-	−29.06 ± 2.28	−29.37 ± 2.15	64
-CG*A-	−29.21 ± 2.23	−29.70 ± 2.18	56
-CG*C-	−29.61 ± 2.13	−29.64 ± 2.20	53
-CG*T-	−26.04 ± 3.88	−29.61 ± 2.27	37

[a] The trimer core sequence of the FAF-modified 12-mer duplexes used in this study. See Figure 1c for full sequence contexts. G* = FAF adduct. [b] The sum of van der Waals interaction energies in kcal/mol between the trimer containing FAF-G and 3′- and 5′-base pairs. [c] Taken from Table 2.

C pair), then the real enthalpy loss should be even smaller, since the stacking enthalpy of the G:C pair was not completely lost but was replaced by stacking of the FAF moiety. Therefore, the $\Delta\Delta H$ values around and below 10 kcal/mol, as those obtained from the QM calculations, seemed to be more realistic.

Van der Waals (VDW) Base Stacking Interactions

We monitored the pure van der Waals (VDW) energies by the Amber Lennard-Jones term as sampled in MD simulations. In all the S conformer duplexes examined, the G[FAF] stacked well with adjacent base pairs. Table 4 lists the sum of total stacking stabilization of G[FAF] with 5′- and 3′-base pairs. The duplexes with 3′-T (-AG*T-, -CG*T-) in each series were noticeably less stabilized by the VDW stacking energies. This contrasts the B-type conformers, in which the VDW energies of the two 3′-T duplexes were comparable to others, but represented a significant increase in stacking stabilization relative to the corresponding S conformers (by +1.43 and +3.57 kcal/mol, respectively). The VDW results supported a general trend, i.e., the greater the VDW stacking, the greater the S conformer population. The full details of the VDW interactions are given in Supporting Information Tables S3 and S4.

E. coli UvrABC Nucleotide Excision Repair

Control Experiments. The biological and spectroscopic utilities of 7′-fluoro-2-aminofluorene (FAF) as a fluorine structure probe have well been documented (*28*−*31*). The CD profiles of FAF-modified DNA duplexes (this study, Figure 3) are similar to those of the AF ones we reported previously (*29*). Control repair experiments were performed to verify that the incision efficiency was not also affected by the hydrogen/fluorine swap (*45*). For that, we employed the -AGG- sequence (5′-CTTCTAGGCCTC-3′), in which each of the two central guanines were modified with either AF or FAF. Incision efficiencies of the resulting four sequences were found to be comparable under identical experimental conditions (*45*). These results confirmed that the fluorine substitution had little or no effect on the incision efficiency of the AF adduct.

*Incision of the -AG*N- and -CG*N- Series.* Figure 7 shows a plot of the *E. coli* UvrABC incision efficiency *versus* % S conformer for the eight FAF-modified sequences. The -AG*N- and -CG*N- series are labeled as red and blue

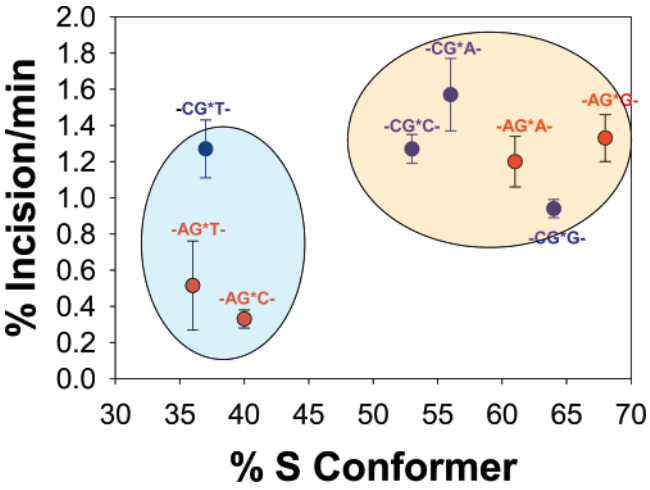

FIGURE 7: Incision efficiencies (% DNA incision/min) of the -AG*N- (red) and -CG*N- (blue) sequences versus % S conformer by *Escherichia coli* UvrABC nuclease.

symbols, respectively. An excellent correlation ($R^2 = 0.93$) was obtained when only the -AG*N- series were considered (*45*). Consistently greater incision efficiencies were observed for duplexes that are predominantly S conformer, the -AG*G- (68%) and -AG*A- (61%) sequences. However, no such correlation was found for the -CG*N- series. Consequently, when eight data points were plotted together, a poor correlation was obtained. Nevertheless, the trend persisted: *i.e.*, low NER was associated with low % S conformation (shaded cyan; -AG*C-, -AG*T-, -CG*T-), and high NER was associated with high % S conformation (shaded orange; -AG*G-, AG*A-, -CG*G-, -CG*A-, -CG*C-).

In other words, the lesions with higher S/B ratio appeared to be more susceptible to NER repair than most of the lesions with lower S/B ratio (except CG*T). Moreover, the -CG*N- series exhibited greater incision efficiencies than the -AG*N- counterparts.

DISCUSSION

The precise mechanisms by which DNA adducts exert mutations are of immense interest from the standpoint of structure−function relationships (*2*−*4*). It is known that certain bulky carcinogen−DNA adducts produce complex sequence-dependent conformational heterogeneities in DNA (*4*, *8*−*10*). A profound sequence effect on mutation has been reported for a variety of lesions on both bacterial and mammalian cells (*64*−*69*). The key question is how base sequence context governs equilibria between adduct conformations. Is one sequence context more prone to repair or miscode during replication than another? Are there unique or a combination of conformers that are more or less repair susceptible, and/or more likely to produce mutations upon replication? Answers to these questions can, in principle, be gleaned from understanding the nature of chemical forces necessary to maintain adduct structures/conformations and their interactions with relevant proteins. The established sequence dependence of AF-induced S/B heterogeneity make it uniquely suited for addressing these questions.

Sequence Effects on S/B Conformational Heterogeneity. We employed eight of the 16 possible flanking sequence contexts that are modified by the fluorine probe FAF (Figure 1). The ^{19}F NMR results (Figure 4) showed that all the fully

complementary 12-mer duplexes existed as a mixture of S and B conformers and their population balances were modulated by the nature of flanking base sequence contexts. When measured at 20 °C, the population of S conformer for both the -AG*N*- and -CG*N*- series decreased in order of 3′-dG > dA > dC > dT. The results clearly indicated the importance of a purine base flanking 3′ to the lesion site for promoting the S conformation. Although data are limited, the 5′-flanking base did not seem to exert a similar effect on S/B heterogeneity. For example, the population of S conformer was similar for -AG*G- and -CG*G-, and for -AG*A- and -CG*A-.

We carried out three distinct computational analyses to determine which energy terms correlate with the NMR-based S/B population ratios. We first determined the pure van der Waals (VDW) energy (Table 4), which neglects the electrostatic portion of stacking. Not surprisingly, correlations between the VDW energy difference and the S/B ratios were found to be poor. Nevertheless, analysis of the individual VDW contributions suggested that purine bases at the 3′-flanking position increased stacking energies, thereby favoring the S conformation. Conversely, pyrimidine bases reduced stacking which promoted the B conformation. The results were in agreement with the experimental observations: *i.e.*, the nature of 3′-flanking bases was important for promoting (*i.e.*, G and A) or reducing (T) stacking energy, thus yielding greater populations of S and B conformers, respectively. These results were also in accordance with the CD data which showed largest negative CD ellipticities for the -AG*T- and -CG*T- duplexes.

We then evaluated the full intrinsic (direct) base stacking (or base−base) forces. For this purpose we utilized complete QM calculations using the latest dispersion-augmented DFT method (Table 3) (*53*), which provide a full description of the stacking (including also electrostatics, induction, and charge-transfer forces) (*33*−*36*). The magnitude of the S/B energy difference was closer to the expected enthalpy difference than the VDW analysis, and the correlations represent improvement over the pure VDW description discussed above. Finally, we added solvent screening effects to the QM stacking data. The solvent was approximated by the QM-based continuum solvent approach because we were unable to account for the exact shape of the whole DNA dodecamer (*i.e.*, only the trimer segment was considered). Upon inclusion of solvation, we achieved an overall agreement between the magnitude of the calculated interaction and the expected values (Table 1), although the correlation coefficients were still relatively poor.

Next, let us comment on the inaccuracies in the QM calculations. Although evaluation of the interaction energy itself was fairly accurate, especially when considering $\Delta\Delta E$ (major error cancellation), there were at least two additional sources of error in our theoretical model. First, the calculated interaction was critically dependent on the geometries taken from MD simulations, which can be biased by force field and limited sampling of the MD, especially for short simulations. Thus, some simulations may give geometries that poorly represent the real ones. Perhaps less important was the error of the solvation model (see for instance the outlaying solvation energy of -CG*A- in Table 3), which stems from having only three dissected base pairs for solvation. However, attempting to include the sugar-phosphate segments into the calculations would create insurmountable methodological problems and reduce the quality of the predictions. In view of all the approximations, the complexity of the system, and relatively small enthalpy differences, the observed nonideal correlation was not surprising. Alternatively, the free energy difference could be acquired by postprocessing the MD trajectories, using the MM-PBSA method. Although this method is extensively used in the literature, it is based on approximations and the uncertainty of the results would be in the low kcal/mol range (*38*).

The QM results described above indicate that both electrostatics and solvation were important sequence effectors as their sequence variation and magnitude were larger than the variability of the dispersion energy contribution. Although inclusion of these factors did not significantly improve the correlation between the interaction energy ($\Delta\Delta E_{tot}$(S−B)) and the S/B population ratios (Table 3), the enthalpy differences calculated were similar to the values estimated from the NMR and thermomelting measurements. Despite the lack of quantitative correlation, we believe that our present model, the QM description augmented by solvent effects, is a good compromise and provides valuable qualitative insights into the sequence effect.

The Dynamics of S/B Conformational Heterogeneity. The dynamics of S/B heterogeneity were complicated by the concomitant presence of the S/B equilibrium and the melting process of the modified duplex. Nevertheless, valuable kinetic and thermodynamic properties were obtained from UV melting (van't Hoff) and ^{19}F NMR (dynamic exchange) experiments. Analysis of the van't Hoff parameters indicated that the thermal and thermodynamic stabilities of the FAF-modified duplexes were dictated primarily by the polarity and the numbers of hydrogen bonds associated with the flanking base pairs. The small energy differences ($\Delta G°$ < 0.5 kcal/mol) observed between the two conformers from the dynamic ^{19}F NMR experiments suggest a possible facile adduct conformation switch when in contact with repair proteins or the active pocket of a polymerase (*3*, *8*). Rapid dynamics of S/B conformational heterogeneity were observed. The exchange rate constants at the coalescent temperature were found to be faster than average rates of spontaneous base pair opening of the Watson−Crick base pairs in B-DNA (*69*). The highly lipophilic carcinogenic moiety in the S conformer was not involved with any hydrogen bonds, but adopted a stable stacked conformation within the helix.

The S/B exchange occurred in milliseconds and was significant because it was in the physiological range of DNA synthesis. Eckel and Krugh (*18*) have previously coined the term "mutagenic switch" to describe S/B conformational heterogeneity. We have shown recently the dynamic presence of the S/B equilibrium at the replication fork, rationalizing either error-free or error-prone replication (*31*). The results also support a model in which the conformational heterogeneities at positions remote from the primer-terminus influence polymerase function through a long-range DNA−protein interaction (*31*). The S/B equilibrium situation is reminiscent of the extensively studied base flipping by the DNA repair enzyme uracil DNA glycosylase. An important distinction is that the inner helix of the S/B model is always occupied by either the hydrophobic carcinogen (*i.e.*, S

conformer) or the modified dG (*i.e.*, B conformer). Jiang et al. (*70*) have studied the role of hydrogen bonding and steric effects in uracil DNA glycosylase using difluorophenyl as a model uracil. They concluded that the hydrophobic moiety is stabilized within the helix despite a rapid exchange between the stacked and unstacked states ($k_C > 1200\ s^{-1}$), which could also explain the S/B equilibrium in our experiments. It has been shown that base-stacking interaction dominates not only in the duplex stability but also contributes into the dependence of the duplex stability on its sequence (*71*). A recent theoretical study indicated that the fluorescence quenching effect of 2-aminopurine is more pronounced when it is stacked with a purine base than with a pyrimidine base, indicating the importance of extended π-oribital conjugation (*72*). The sequence dependent base stacking and unstacking has been shown to be enthalpic in nature (*71*). Adduct formation induced weakened base stacking, and the resulting conformational flexibility could be an important determining factor for many repair proteins, including the *E. coli* UvrABC system (*73*, *74*).

Nucleotide Excision Repair versus S/B Conformational Heterogeneity. The majority of NER studies to date have assumed that a bulky lesion adopts a single conformation in the DNA duplex (*39*, *40*, *57*−*60*). Clearly, this is not the case for the AF adducts which equilibrated dynamically between the S and B conformations. The sequence dependence of NER has been documented in both bacterial and mammalian systems (*39*, *40*) and may account for the presence of mutational "hot spots", such as the *E. coli Nar*I exonuclease sequence (5′-$CG_1G_2CG_3$C-3′) (*11*). Mekhovich et al. (*75*) have shown that the rate of incision of both AF and AAF adducts in the *E. coli* UvrABC system is significantly faster when they are positioned in the *Nar*I sequence (5′-GGCG*CC-3′) than when located in a random sequence (5-GATG*ATA-3′).

The key issue here is which conformer, S or B, is recognized as a defect in the first step of *E. coli* NER. This is an important question since S/B interconversion occurs in physiologically relevant scales (see above). Our present ^{19}F NMR/NER results provided structural/conformational insights into this very intriguing matter. Control experiments showed that the incision efficiencies of two different AF adduct sequences were not affected by the presence of fluorine, thus justifying the use of FAF as repair probes (*45*).

It has been shown previously that the exclusively stacked AAF adduct is repaired more readily than the conformationally flexible AF adduct (*39*, *40*). The S conformation, which lacks Watson−Crick base pairs at the lesion site, is also structurally similar to the base-displaced *cis*-benzo[*a*]-pyrene-N^2-dG (BPDE-dG) adducts. In contrast, the minor-groove *trans*-BPDE adducts maintain Watson−Crick base pairs at the lesion site (*8*, *9*). The *cis* adducts have been shown to be incised by UvrABC more efficiently than the *trans* adducts (*59*). Thus, our NER results support growing evidence that a base-displaced intercalated conformer is more efficiently repaired than adducts with intact Watson−Crick hydrogen bonds. However, it should be pointed out that carcinogen stacking in certain intercalated S-like conformations is responsible for increased thermal and thermodynamic stabilization of a duplex (*9*, *10*), a known negative indicator for NER (*43*). While the modified dG of the B conformer maintains Watson−Crick bonds at the lesion site, the highly hydrophobic carcinogen moiety in the major groove may create a solvation environment favorable for repair. Zou et al. (*59*) have shown that AF and AAF adducts in the -TG*T- sequence context are incised more efficiently by UvrABC than in the -CG*C- context. Similar results have been obtained with the minor groove *trans*-BPDE adduct by UvrABC proteins from both *E. coli* and the thermophilic *Bacillus caldotenax* (*76*). The presence of the T:A flanking base pairs adjacent to the modified dG* allowed for greater local bend, flexibility, and conformational heterogeneity than in the -CG*C- sequence. This is consistent with the lower T_m and $-\Delta G°$ values observed for the -TG*T- sequence. Interestingly, we have shown in the present study that the -CG*N- duplex series were more stable (greater T_m and $-\Delta G°$ values) than the -AG*N- counterpart, yet they were more susceptible to repair (Figure 7). It appears that important recognition factors of UvrABC NER are not necessarily thermal and thermodynamic stabilities, but rather the quality of Watson−Crick base pairs and conformational flexibility at the lesion site. Taken together, these results illustrated the complexity of the way multiconformeric adducts interact with repair proteins, information that is crucial to the first step in damage recognition in NER.

Several mechanisms have been proposed to explain how the NER machinery recognizes lesions situated within an expanse of intact genomic DNA (*39*−*43*). However, there is no single one-fits-all mechanism. The concept of "thermal/thermodynamic probing" has evolved in recent years to involve more complex "bipartite" or "multipartite" mechanisms (*42*, *43*). The latter focuses on the quality of Watson−Crick base pairs at or near the lesion site and the adduct structures. Alternative possibilities include the indirect readout mechanisms which emphasize the local conformational distortion of either the modified or the complementary strand at the modification site (*77*, *78*).

It has been shown that the DNA sequence itself has no intrinsic or direct effect on the incision efficiency, as the damage recognition factor $UvrA_2B$ also binds to undamaged DNA in a sequence-independent manner (*79*). The DNA sequence at a lesion may only affect the structure of the lesion which is recognized and incised. Since all the substrates examined in this study contained the same adduct FAF, they should be recognized equally well by UvrB (*80*). Our repair results reinforced the importance of Watson−Crick base pairing at the lesion site for the initial recognition by the $UvrA_2B$ proteins. Our CD data indicated, however, that neither the B nor the S conformer significantly altered the global DNA conformation. So, what differentiates B and S conformers in terms of NER recognition? Figure 8 shows the surface models of B and S conformers and highlights the polarity and shape differences in their major groove area: *i.e.*, the extrusion of the highly hydrophobic carcinogen moiety in the B conformer *versus* the adduct-displaced highly polar dG and dC in the S conformer. It is plausible that this subtle electrostatic difference may be responsible for the observed conformation-specific repair. Direct identification of the lesion structures at those sequence-dependent hot spots in cells is almost impossible. However, our ^{19}F NMR/NER results provided insights into the three-way relationship of sequence, conformation specificity, and NER, a biological consequence closely associated with mutagenesis (*81*).

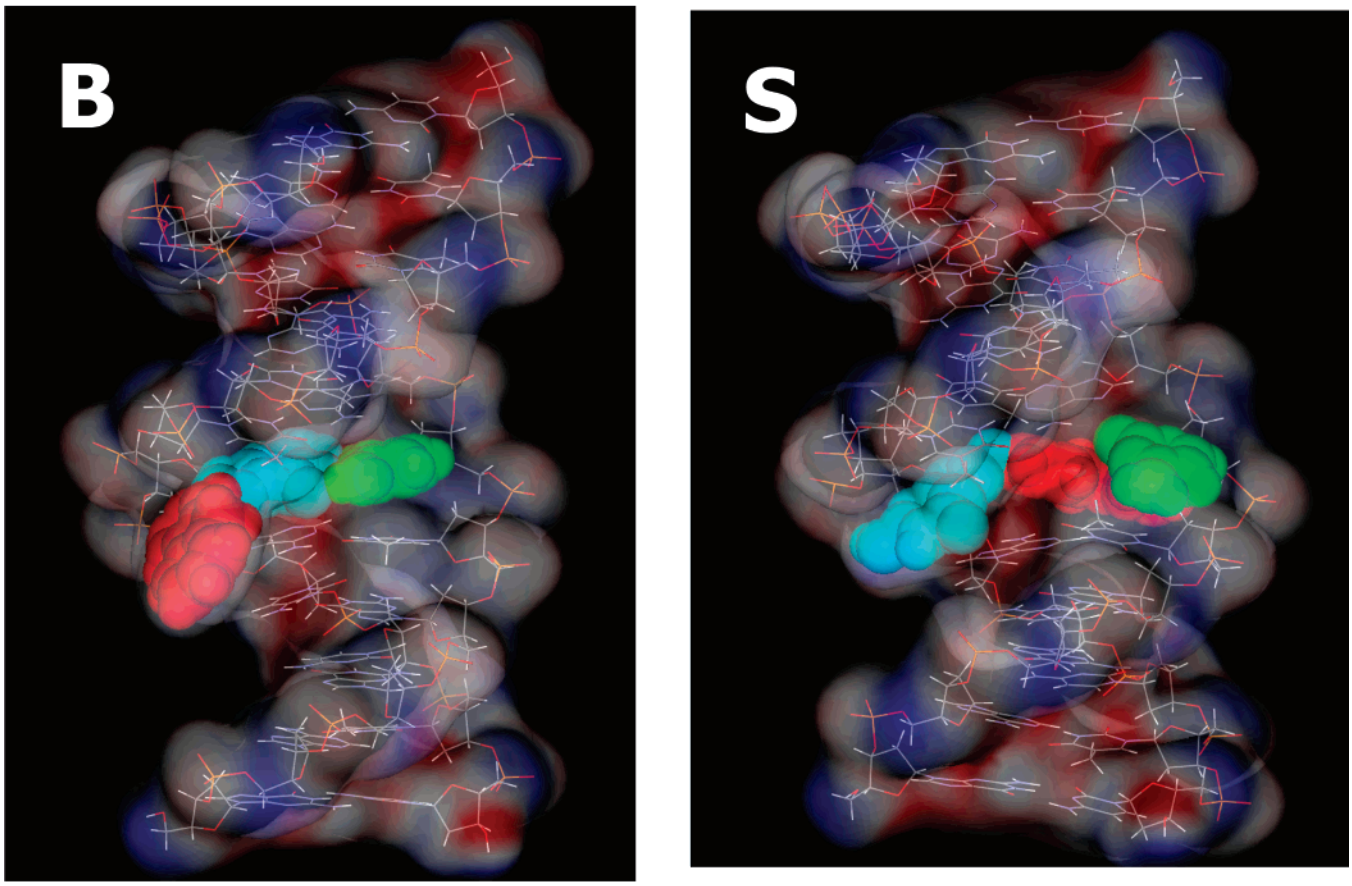

FIGURE 8: Surface model illustrations which highlight the polarity and shape difference in the major groove area of the B and S conformers; *i.e.*, the extrusion of the highly hydrophobic carcinogen moiety (red) in the B conformer *versus* the adduct-displaced highly polar dG (cyan) and dC (green) in the S conformer.

S/B conformational heterogeneity may also play an important role in modulating the mutational outcomes of the AF adduct. Significant sequence effects were observed on both frameshift and substitution mutation frequency of AAF, AF, and PhIP adducts derived both in *E. coli* and mammalian cells (*44*, *64*). Shibutani et al. (*66*) have shown that the mutation frequency and specificity of AF in COS-7 cells vary significantly depending on the nature of the bases flanking the lesion. For example, the base substitution mutation of AF in the 5′-GG*C-3′ context was 50-fold higher than in the 5′-TG*C-3′ context. The presence of a dG 5′ of the lesion increases the overall mutational frequency dramatically. These effects must originate from the sequence dependence of S/B heterogeneity. A more complete set of ^{19}F NMR results will be required to establish meaningful conformation−mutation correlations.

Summary. In the present study we conducted a systematic spectroscopic (UV, CD, ^{19}F NMR) and computational (VDW, QM) investigation in order to gain structural insight into AF-induced S/B conformational heterogeneity. We found that FAF-modified DNA duplexes equilibrated between S and B conformations and interconverted slowly at room temperature. The S/B population ratios were modulated by the 3′-flanking sequence. Specifically, purine bases at the 3′-flanking site promoted the S conformation (G > A > C > T). Van der Waals energy calculations indicated that the stacking interaction between the carcinogen and neighboring base pairs was a key to the observed sequence effect. Quantum mechanics calculations provided a more complete picture and stressed the importance of both electrostatic and solvation forces. The ^{19}F NMR/NER results suggested that the lesions with higher S/B ratio appeared to be more susceptible to NER repair than most of the lesions with lower S/B ratio (except CG*T). The results represented the first quantitative structure−function investigation relating fractions of S conformers with increased NER efficiencies and provided valuable insights into how bulky DNA adducts are accommodated by the NER protein in a conformationally specific manner.

SUPPORTING INFORMATION AVAILABLE

Synthesis (Figure S1), HPLC purification (Figures S2, S3), and ESI-TOF-MS characterization (Figures S4−S7) of FAF-modified 12-mer oligodeoxynucleotides; imino proton (Figure S8) and ^{19}F (Figure S9) NMR spectra of the -AG*N- and CG*N- duplex series; simulated dynamic spectra (Figure S10) of the -CG*N- duplexes; the atom type assignment of FAF-dG (Figure S11); assumed enthalpy ordering diagram (Figure S12); calculated and measured molecular ion mass (Table S1); ^{19}F NMR H/D isotope effects (Table S2); van der Waals parameters for B (Table S3) and S (Table S4) conformers; force field parameters for dG-FAF (Table S5); RESP charge results for B (Table S6) and S (Table S7) conformers; QM calculations, DFT−D dimer interaction energies (Table S8). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hoeijmakers, J. H. J. (2001) Genome maintenance mechanisms for preventing cancer, *Nature 411*, 366−374.
2. Luch, A. (2005) Nature and nature-lessons from chemical carcinogenesis, *Nat. Rev. Cancer 5*, 113−125.
3. Guengerich, F. P. (2006) Interactions of carcinogen-bound DNA with individual DNA polymerases, *Chem. Rev. 106*, 420−452.
4. Lukin, M., and de Los Santos, C. (2006) NMR structures of damaged DNA, *Chem. Rev. 106*, 607−686.
5. Pages, V., and Fuchs, R. P. P. (2002) How DNA lesions are turned into mutations within cells?, *Oncogene 21*, 8957−8966.
6. Seo, K. Y., Jelinsky, S. A., and Loechler, E. L. (2000) Factors that influence the mutagenic patterns of DNA adducts from chemical carcinogens, *Mutat. Res. 463*, 215−246.
7. Seo, K. Y., Nagalingam, A., Tiffany, M., and Loechler, E. L. (2005) Mutagenesis studies on four stereoisomeric N^2-dG benzo[*a*]pyrene adducts in the identical 5′-CGC sequence used in NMR studies: although adduct conformation differs, mutagenesis outcome does not as G→T mutations dominate in each case, *Mutagenesis 20*, 441−448.
8. Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance, *J. Environ. Sci. Health, Part C: Environ. Carcinog. Ecotoxicol. Rev. 22*, 57−90.
9. 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.
10. Patel, D. J., Mao, B., Gu, Z., Hingerty, B. E., Gorin, A., Basu, A. K., and Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance, *Chem. Res. Toxicol. 11*, 391−407.
11. Hoffmann, G. R., and Fuchs, R. P. P. (1997) Mechanisms of frameshift mutations: insight from aromatic amines, *Chem. Res. Toxicol. 10*, 347−459.
12. Elmquist, C. E., Wang, F., Stover, J. S., Stone, M. P., and Rizzo, C. J. (2007) Conformational differences of the C8-Deoxyguanosine adduct of 2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ) within the *Nar*I recognition sequence, *Chem. Res. Toxicol.* 445−454.
13. Heflich, R. H., and Neft, R. E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites, *Mutat. Res. 318*, 73−114.
14. Beije, B., and Moller, L. (1988) 2-Nitrofluorene and related compounds: prevalence and biological effects, *Mutat. Res. 196*, 177−209.
15. Hsu, G. W., Kiefer, J. R., Burnouf, D., Becherel, O. J., Fuchs, R. P., and Beese, L. S. (2004) Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase, *J. Biol. Chem. 279*, 50280−50285.
16. Dutta, S., Li, Y., Johnson, D., Dzantiev, J., Richardson, C. C., Romano, L. J., and Ellenberger, T. (2004) Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis, *Proc. Natl. Acad. Sci. U.S.A. 101*, 16186−16191.
17. Cho, B. P., Beland, F. A., and Marques, M. M. (1994) NMR structural studies of a 15-mer duplex from a *ras* protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity, *Biochemistry 33*, 1373−1384.
18. Eckel, L. M., and Krugh, T. R. (1994) Structural characterization of two interchangeable conformations of a 2-aminofluorene-modified DNA oligomer by NMR and energy minimizations, *Biochemistry 33*, 13611−13624.
19. Mao, B., Gu, Z., Hingerty, B. E., Broyde, S., and Patel, D. J. (1998) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn* [AF]-C^8-dG adduct opposite dC in a DNA duplex, *Biochemistry 37*, 81−94.
20. Mao, B., Gu, Z., Hingerty, B. E., Broyde, S., and Patel, D. J. (1998) Solution structure of the aminofluorene [AF]-external conformer of the *anti* [AF]-C^8-dG adduct opposite dC in a DNA duplex, *Biochemistry 37*, 95−106.
21. Mao, B., Gu, Z., Gorin, A., Hingerty, B. E., Broyde, S., and Patel, D. J. (1997) Solution structure of the aminofluorene [AF] stacked conformer of the *syn* [AF]-C^8-dG adduct positioned at a DNA template-primer junction, *Biochemistry 36*, 14491−14501.
22. Gorin, A., Gu, Z., Hingerty, B. E., Broyde, S., and Patel, D. J. (1999) Solution structure of the aminofluorene [AF]-stacked conformer of the *syn* [AF]-C^8-dG adduct positioned opposite dC at a template-primer junction, *Biochemistry 38*, 10855−10870.
23. Norman, D., Abuaf, P., Hingerty, B. E., Live, D., Grunberger, D., Broyde, S., and Patel, D. J. (1989) NMR and computational characterization of the N-(deoxyguanosin-8-yl)aminofluorene adduct (AF)G opposite adenosine in DNA: (AF)G[*syn*]: A[*anti*] pair formation and its pH dependence, *Biochemistry 28*, 7462−7476.
24. Abuaf, P., Hingerty, B. E., Broyde, S., and Grunberger, D. (1995) Solution conformation of the *N*-(deoxyguanosin-8-yl)aminofluorene adduct opposite deoxyinosine and deoxyguanosine in DNA by NMR and computational characterization, *Chem. Res. Toxicol. 8*, 369−378.
25. Mao, B., Cosman, M., Hingerty, B. E., Broyde, S., and Patel, D. J. (1995) Solution conformation of [AF]dG opposite a -1 deletion site in a DNA duplex: intercalation of the covalently attached aminofluorene ring into the helix with base displacement of the C^8-modified *syn* guanine into the major groove, *Biochemistry 34*, 6226−6238.
26. Mao, B., Hingerty, B. E., Broyde, S., and Patel, D. J. (1995) Solution conformation of [AF]dG opposite a -2 deletion site in a DNA duplex: intercalation of the covalently attached aminofluorene ring into the helix with base displacement of the C8-modified *syn* guanine and adjacent unpaired 3′-adenine into the major groove, *Biochemistry 34*, 16641−16653.
27. Mao, B., Gorin, A., Gu, Z., Hingerty, B. E., Broyde, S., and Patel, D. J. (1997) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn* [AF]-C^8-dG adduct opposite a -2 deletion site in the *Nar*I hotspot sequence context, *Biochemistry 36*, 14479−14490.
28. Zhou, L., Rajabzadeh, G., Traficante, D. D., and Cho, B. P. (1997) Conformational heterogeneity of arylamine-modified DNA: ^{19}F NMR evidence, *J. Am. Chem. Soc. 119*, 5384−5389.
29. Meneni, S. R., D'Mello, R., Norigian, G., Baker, G., Gao, L., Chiarelli, M. P., and Cho, B. P. (2006) Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties, *Nucleic Acids Res. 34*, 755−763.
30. Liang, F., Meneni, S., and Cho, B. P. (2006) Induced circular dichroism characteristics as conformational probes for carcinogenic aminofluorene-DNA adducts, *Chem. Res. Toxicol. 19*, 1040−1043.
31. Meneni, S., Liang, F., and Cho, B. P. (2007) Examination of the long-range effects of aminofluorene-induced conformational heterogeneity and its relevance to the mechanisms of translesion DNA synthesis, *J. Mol. Biol. 366*, 1387−1400.
32. Prive, G. G., Yanagi, K., and Dickerson, R. E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G, *J. Mol. Biol. 217*, 177−199.
33. Sponer, J., Leszczynski, J., and Hobza, P. (1996) Nature of nucleic acid-base stacking: nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs, *J. Phys. Chem. 100*, 5590−5596.
34. Reha, D., Kabelac, M., Ryjacek, F., Sponer, J., Sponer, J. E., Elstner, M., Suhai, S., and Hobza, P. (2002) Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4′,6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study, *J. Am. Chem. Soc. 124*, 3366−3376.
35. Luo, R., Gilson, H. S. R., Potter, M. J., and Gilson, M. K. (2001) The physical basis of nucleic acid base stacking in water, *Biophys. J. 80*, 140−148.
36. Sponer, J., Jurecka, P., Marchan, I., Luque, F. J., Orozco, M., and Hobza, P. (2006) Nature of Base Stacking: Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps, *Chem. Eur. J. 12*, 2854−2865.
37. Spackova, N., Berger, I., Egli, M., and Sponer, J. (1998) Molecular dynamics of hemiprotonated intercalated four-stranded i-DNA: stable trajectories on a nanosecond scale, *J. Am. Chem. Soc. 120*, 6147−6151.
38. McDowell, S. E., Spackova, N., Sponer, J., and Walter, N. G. (2007) Molecular dynamics simulations of RNA: An in silico single molecule approach, *Biopolymers 85*, 169−184.
39. Gillet, L. C. J., and Scharer, O. D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair, *Chem. Rev. 106*, 253−276.

40. Truglio, J. J., Croteau, D. L., Van Houten, B., and Kisker, C. (2006) Prokaryotic nucleotide excision repair: the UvrABC system, *Chem. Rev. 106*, 233−252.
41. Sancar, A., and Reardon, J. T. (2004) Nucleotide excision repair in E. coli and man, *Adv. Protein Chem. 69*, 43−71.
42. Dip, R., Camenisch, U., and Naegeli, H. (2004) Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair, *DNA Repair (Amst) 2*, 1409−23.
43. Geacintov, N. E., Broyde, S., Buterin, T., Naegeli, H., Wu, M., Yan, S., and Patel, D. J. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery, *Biopolymers 65*, 202−210.
44. Shibutani, S. (2004) Requirements for frame-shift deletion during translesion synthesis, *Environ. Mutagen Res.* **26**, 135−141.
45. Meneni, S., Shell, S. M., Zou, Y., and Cho, B. P. (2007) Conformation-specific recognition of carcinogen-DNA adduct in *E. coli*. nucleotide excision repair, *Chem. Res. Toxicol. 20*, 6−10.
46. Friebolin, H. (1998) in *Basic one- and two-dimensional NMR spectroscopy*, 3rd ed., pp 301−329, Wiley-VCH, New York.
47. Wang, J. M., 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.
48. Cheatham, T. E., III, Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat, *J. Biomol. Struct. Dyn. 16*, 845−862.
49. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J. Comput. Chem. 25*, 1157−1174.
50. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. (2004) *Gaussian 03*, revision C.02, Gaussian, Inc., Wallingford CT.
51. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
52. Simmerling, C., Elber, R., and Zhang, J. (1995) in *Modeling of Biomolecular Structure and Mechanisms* (Pullman, A., Ed.) pp 241−265, Kluwer, The Netherlands.
53. Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys. 98*, 5648−5652.
54. Jurečka, P., Černý, J., Hobza, P., and Salahub, D. R. (2007) Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with *ab initio* quantum mechanics calculations, *J. Comput. Chem. 28*, 555−569.
55. Tao, J., Perdew, J. P., Staroverov, V. N., and Scuseria, G. E. (2003) Climbing the density functional ladder: nonempirical meta−generalized gradient approximation designed for molecules and solids, *Phys. Rev. Lett. 91*, 146401−146404.
56. Ahlrichs, R., Bär, M., Häser, M., Horn, H., and Kölmel, C. (1989) Electronic structure calculations on workstation computers: The program system turbomole, *Chem. Phys. Lett. 162*, 165−169.
57. Zou, Y., Liu, T. M., Geacintov, N. E., and Van Houten, B. (1995) Interaction of the UvrABC nuclease system with a DNA duplex containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE, *Biochemistry 34*, 13582−13593.
58. Yang, Z., Colis, L. C., Basu, A. K., and Zou, Y. (2005) Recognition and incision of gamma-radiation-induced cross-linked guanine-thymine tandem lesion G[8,5-Me]T by UvrABC nuclease, *Chem. Res. Toxicol. 18*, 1339−1346.
59. Zou, Y., Shell, S. M., Utzat, C. D., Luo, C., Yang, Z., Geacintov, N. E., and Basu, A. K. (2003) Effects of DNA adduct structure and sequence context on strand opening of repair intermediates and incision by UvrABC nuclease, *Biochemistry 42*, 12654−12661.
60. Luo, C., Krishnasamy, R., Basu, A. K., and Zou, Y. (2000) Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, *N*-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease, *Nucleic Acids Res. 28*, 3719−3724.
61. SantaLucia, J., and Hicks, D. (2004) The thermodynamics of DNA structural motifs, *Annu. Rev. Biophys. Biomol. Struct. 33*, 415−440.
62. Singer, B., and Hang, B. (2000) Nucleic acid sequence and repair: role of adduct, neighbor bases and enzyme specificity, *Carcinogenesis 21*, 1071−1078.
63. Hansen, P. E., Dettman, H. D., and Sykes, B. D. (1985) Solvent-induced deuterium isotope effects on fluorine-19 chemical shifts of some substituted fluorobenzenes. Formation of inclusion complexes, *J. Magn. Reson. 62*, 487−496.
64. Broschard, T. H., Koffel-Schwartz, N., and Fuchs, R. P. P. (1999) Sequence-dependent modulation of frameshift mutagenesis at *Nar*I-derived mutation hot spots, *J. Mol. Biol. 288*, 191−199.
65. Page, J. E., Zajc, B., Oh-hara, T., Lakshman, M. K., Sayer, J. M., Jerina, D. M., and Dipple, A. (1998) Sequence context profoundly influences the mutagenic potency of trans-opened benzo[a]pyrene 7,8-diol 9,10-epoxide-purine nucleoside adducts in site-specific mutation studies, *Biochemistry 37*, 9127−9137.
66. Shibutani, S., Suzuki, N., Tan, X., Johnson, F., and Grollman, A. P. (2001) Influence of flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells, *Biochemistry 40*, 3717−3722.
67. Zhuang, P., Kolbanovskiy, A., Amin, S., and Geacintov, N. E. (2001) Base sequence dependence of in vitro translesional DNA replication past a bulky lesion catalyzed by the exo- Klenow fragment of Pol I, *Biochemistry 40*, 6660−6669.
68. Kolbanovskiy, A., Kuzmin, V., Shastry, A., Kolbanovskaya, M., Chen, D., Chang, M, Bolton, J. L., and Geacintov, N. E. (2005) Base selectivity and effects of sequence and DNA secondary structure on the formation of covalent adducts derived from the equine estrogen metabolite 4-hydroxyequilenin, *Chem. Res. Toxicol. 18*, 1737−1747.
69. Krueger, A., Protozanova, E., and Frank-Kamenetskii, M. D. (2006) Sequence-dependent base pair opening in DNA double helix, *Biophys. J. 90*, 3091−3099.
70. Jiang, Y. L., McDowell, L. M., Poliks, B., Studelska, D. R., Cao, C., Potter, G. S., Schaefer, J., Song, F., and Stivers, J. T. (2004) Recognition of an unnatural difluorophenyl nucleotide by uracil DNA glycosylase, *Biochemistry 43*, 15429 -15438.
71. Yakovchuk, P., Protozanova, E., and Frank-Kamenetskii, M. D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix, *Nucleic Acids Res. 34*, 564−574.
72. Hardman, S. J., and Thompson, K. C. (2006) Influence of base stacking and hydrogen bonding on the fluorescence of 2-aminopurine and pyrrolocytosine in nucleic acids, *Biochemistry 45*, 9145−9155.
73. Yang, W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins, *DNA Repair 5*, 654−666.
74. Malta, E., Moolenaar, G. F., and Goosen, N. (2006) Base flipping in nucleotide excision repair, *J. Biol. Chem.*, *281*, 2184−2194.
75. Mekhovich, O., Tang, M., and Romano, L. J. (1998) Rate of incision of N-acetyl-2-aminofluorene and *N*-2-aminofluorene adducts by UvrABC nuclease is adduct- and sequence-specific: comparison of the rates of UvrABC nuclease incision and protein-DNA complex formation, *Biochemistry 37*, 571−579.
76. Ruan, Q., Liu, T., Kolbanovskiy, A., Liu, Y., Ren, J., Skovraga, M., Zou, Y., Lader, J., Malkani, B., Amin, S., Van Houten, B., and Geacintov, N. E. (2007) Sequence context- and temperature-dependent nucleotide excision repair of a benzo[a]pyrene diol epoxide-guanine DNA adduct catalyzed by thermophilic UvrABC proteins, *Biochemistry 46*, 7006−7015.

77. Isaacs, R. J., and Spielmann, H. P. (2004) A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility, *DNA Repair 4*, 455−464.
78. Buterin, T., Meyer, C., Giese, B., and Naegeli, H. (2005) DNA quality control by conformational readout on the undamaged strand of the double helix, *Chem. Biol. 12*, 913−922.
79. Van Houten, B. (1990) Nucleotide excision repair in Escherichia coli, *Microbiol. Rev. 54*, 18−51.
80. Zou, Y., Luo, C., and Geacintov, N. E. (2001) Hierarchy of DNA damage recognition in *Escherichia coli* nucleotide excision repair, *Biochemistry 40*, 2923−2931.
81. Branzei, D., and Foiani, M. (2007) Interplay of replication checkpoints and repair proteins at stalled replication forks, *DNA Repair (Amsterdam) 6*, 994−1003.

BI700858S