DNA base sequence effects on bulky lesion-induced conformational heterogeneity during DNA replication

Ang Cai[1], Katie A. Wilson[2], Satyakam Patnaik[1], Stacey D. Wetmore[2] and Bongsup P. Cho[1,*]

[1]Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI 02881, USA and
[2]Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta T1K 3M4, Canada

ABSTRACT

4-Aminobiphenyl (ABP) and its structure analog 2-aminofluorene (AF) are well-known carcinogens. In the present work, an unusual sequence effect in the 5′-CTTCTG$_1$G$_2$TCCTCATTC-3′ DNA duplex is reported for ABP- and AF-modified G. Specifically, the ABP modification at G$_1$ resulted in a mixture of 67% major groove B-type (B) and 33% stacked (S) conformers, while at the ABP modification at G$_2$ exclusively resulted in the B-conformer. The AF modification at G$_1$ and G$_2$ lead to 25%:75% and 83%:17% B:S population ratios, respectively. These differences in preferred conformation are due to an interplay between stabilizing (hydrogen bonding and stacking that is enhanced by lesion planarity) and destabilizing (solvent exposure) forces at the lesion site. Furthermore, while the B-conformer is a thermodynamic stabilizer and the S-conformer is a destabilizer in duplex settings, the situation is reversed at the single strands/double strands (ss/ds) junction. Specifically, the twisted biphenyl is a better stacker at the ss/ds junction than the coplanar AF. Therefore, the ABP modification leads to a stronger strand binding affinity of the ss/ds junction than the AF modification. Overall, the current work provides conformational insights into the role of sequence and lesion effects in modulating DNA replication.

INTRODUCTION

ABP: *N*-(2′-deoxyguanosin-8-yl)-4-aminobiphenyl; AF: *N*-(2′-deoxyguanosin-8-yl)-2-aminofluorene; AAF: *N*-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene

DNA is damaged by many agents in our environment. For example, arylamines are carcinogens derived from cigarette smoke, incomplete combustion of diesel exhaust, and overcooked meat, among others (1,2). These aromatic chemicals do not directly interact with DNA; however, they are metabolized *in vivo* by cytochrome P450 and subsequently activated by cellular enzymes, such as *N*-acetyltrasferase or sulfotransferase to form DNA adducts (3,4). For example, 4-aminobiphenyl (ABP) is a well-known human bladder carcinogen and a potent urinary-bladder carcinogen in experimental animals and is activated by cellular *N*-acetyltransferase to produce C8-substituted dG-ABP as a major DNA lesion (5,6). The mutations of the *p53* tumor suppressor gene in bladder carcinomas of patients who are occupationally exposed to ABP match with the ABP-mutational hotspots of the *p53* gene in cells treated to ABP (7). The structurally similar liver carcinogens 2-aminofluorene (AF) and *N*-2-acetylaminofluorene (AAF) produce similar C8-dG adducts (Figure 1A). AF could be error-free by correctly base pairing with an incoming dC (8), whereas AAF blocks the replication process and requires recruitment of lesion bypass polymerases for translesion synthesis (TLS) (9). In *Escherichia coli*, the TLS of AF produces point and frameshift mutations, whereas bypass of the AAF lesion is frequently accompanied by a frameshift mutation (10).

These bulky DNA adducts have been extensively studied and shown to adopt at least three major conformations in the duplex setting: major groove B-type (B), base-displaced stacked (S), or minor groove wedge (W) (Figure 1B) (11–13). In the S-conformation, the carcinogen stacks between the bases and disrupts base pairing in the double helix, whereas the B-type conformer places the carcinogen in the major groove of the DNA helix that maintains the traditional Watson–Crick base pairings. The W-conformer was originally found in dA mismatch duplexes, and the hydrophobic lesion is placed in the narrow minor groove (14–17). The S/B/W-population ratios have been shown to be influenced by neighboring sequence contexts for the ABP-dG, AAF-dG and AF-dG lesions, which consequently have different repair and mutational efficiencies (11,12,15,17–20). Similarly, bulky DNA lesions at the single-strand (ss)/double-strand (ds) replication fork can

---

*To whom correspondence should be addressed. Tel: +1 401 874 5024; Email: bcho@uri.edu
Present address: Bongsup P. Cho, Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, 7 Greenhouse Road, Kingston, RI 02881, USA.
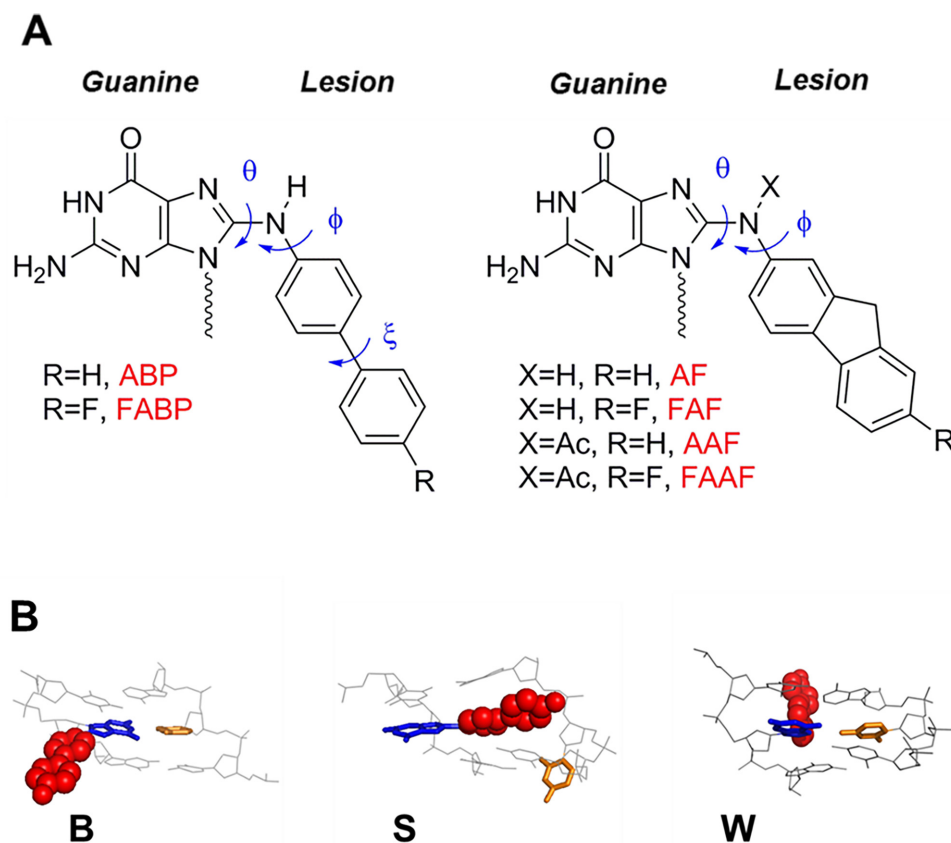
Figure 1. (**A**) Structures of ABP [*N*-(2′-deoxyguanosin-8-yl)-4-aminobiphenyl], AF [*N*-(2′-deoxyguanosin-8-yl)-2-aminofluorene], and AAF [*N*-(2′-deoxyguanosin-8-yl)-2-acetylaminofluorene] and their fluoro models, FABP [*N*-(2′-deoxyguanosin-8-yl)-4′-fluoro-4-aminobiphenyl], FAF [*N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-aminofluorene], and FAAF [*N*-(2′-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene], as well as the definitions of key dihedral angles (θ, φ and ξ). (**B**) Major groove views of arylamine-conformational motifs: B-, S- and W-conformers. (color code: arylamine- red CPK, modified dG- blue, complementary dC- orange).

potentially exist in multiple conformations, which may affect mutational outcomes.

The mutagenicity of DNA adducts is related to how the polymerase processes each of the damaged duplex conformations. When a replicative polymerase encounters a bulky DNA lesion, DNA synthesis pauses or stops to allow the recruitment of bypass polymerases. This in turn affects whether the synthesis is accurate or error prone. For example, AF- and AAF-modified duplexes favor the S-conformer due to the 3′-flanking purine base (15,17,19,21–23). AAF at G_3 in the *Nar*I sequence (5′-$G_1G_2CG_3CN$-3′) induces a greater frequency of −2 frame shift mutations, with the G_3-mutational frequency regulated by the nature of the N base, N = C or T (23–25). Miller *et al.* reported a significant reduction in nucleotide incorporation rates by Klenow fragment (Kf) exo⁻ at positions neighboring and downstream of the lesion site (26). We investigated the conformational effect of AF on nucleotide insertion efficiencies catalyzed by Kf exo⁻ on TG*A and CG*A systems. The results showed that AF in the CG*A duplex sequence adopts a greater population of S-conformer than the TG*A sequence. The potentially S-conformeric CG*A at the ss/ds junction thermodynamically favors A over C opposite the lesion site (22).

Additionally, we previously reported an unusual 3′-flanking T effect on a duplex sequence (5′-CCATCG*CNACC-3′; N = T or A), i.e., the FABP modification at G* exhibited a 40%:60% B:S in the G*CT context whereas 100% B in the G*CA sequence context. Though not as dramatic, a similar trend was observed for the AF and AAF lesions (11).

In the present paper, another unusual fluorine-labeled FABP (FABP)-induced sequence effect on a 16mer duplex sequence (5′-CTTCTG₁G₂TCCTCATTC-3′) was investigated (Figure 1). FABP-modification at G_1 exhibited a 67%:33% ratio of B:S conformations, whereas the B-conformation was adopted exclusively at the G_2 modification. We conducted a series of conformational [NMR, induced circular dichroism (ICD), molecular dynamics (MD) simulations], thermodynamic [UV-melting, differential scanning calorimetry (DSC)], and binding [surface plasmon resonance (SPR), isothermal titration calorimetry (ITC)] experiments for two 16mer DNA templates [(5′-CTTCTG₁*G₂TCCTCATTC-3′) and (5′-CTTCTG₁G₂*TCCTCATTC-3′); G_1*/G_2* = FABP or FAF (fluorine-labeled AF)] (Figures 1 and 2). The data showed that the B-conformer is a thermodynamic stabilizer and the S-conformer is a destabilizer in full duplex settings.
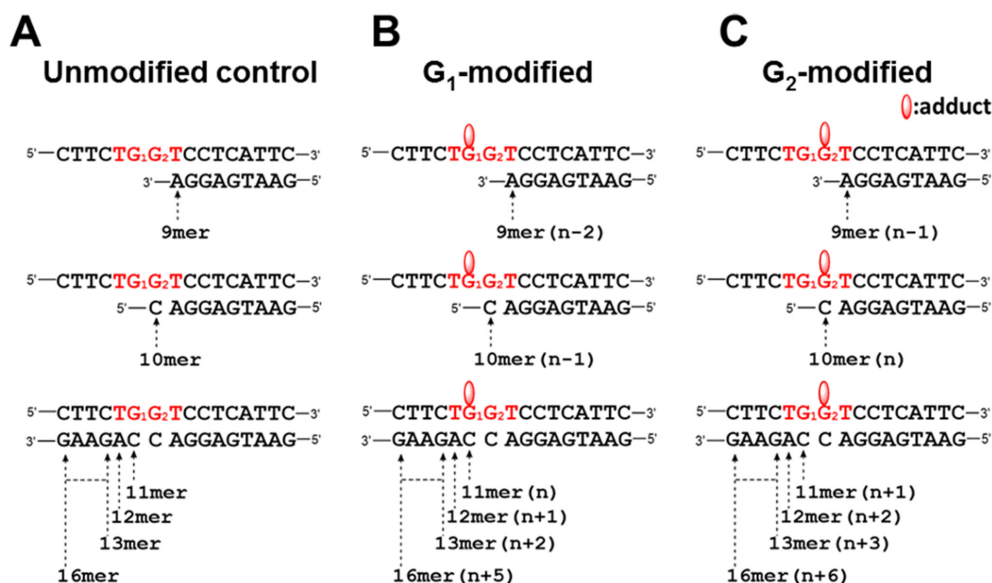
Figure 2. DNA replication models for (**A**) unmodified control, (**B**) FABP/FAF-modified G_1 (TG*GT) templates and (**C**) FABP/FAF-modified G_2 (TGG*T) templates. n: lesion site (solid oval lobes).

However, the opposite was true for the ss/ds sequence contexts. Lesion stacking increased the binding affinity of the complementary strands in the order of FABP > FAF and TG*GT > TGG*T. The twisted nonplanar FABP was a better stacker than the planar FAF in ss/ds systems, which contrasts the situation in the duplexes. Overall, the present study provides valuable insights into the important distinction between ss/ds and full ds sequence contexts, and the flanking base sequences and lesion stacking in modulating bulky lesion-induced conformational heterogeneity.

MATERIALS AND METHODS

Caution : 4-Aminobiphenyl derivatives are known human bladder carcinogens. Aminofluorene analogues are animal carcinogens; therefore, must be handled with caution.

Oligodeoxynucleotides (oligo, 10 μmol scale) in desalted form were obtained from Integrated DNA Technologies (Iowa) and were purified by using a reverse phase HPLC. The HPLC system comprised a Hitachi EZChrom Elite HPLC with a L-2450 diode array detector and a Phenomenex Luna C18 column (150 × 10 mm, 5.0 μm) (Phenomenex, Torrance, CA, USA). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA, USA).

Preparation of lesion-modified oligodeoxynucleotides

FABP-modified 16mer templates (5′-CTTCTG$_1$G$_2$TC CTCATTC-3′) were prepared according to the published procedures (25). For each lesion, we obtained two mono-adducts [d(5′-CTTCTG$_1$*G$_2$TCCTCATTC-3′)] and [d(5′-CTTCTG$_1$G$_2$*TCCTCATTC-3′)] (G* = FABP) and one di-adduct [d(5′-CTTCTG$_1$*G$_2$*TCCTCATTC-3′)]. Briefly, 5–10 mg of *N*-acetoxy-*N*-trifluoroacetyl-7-fluoro-4-aminobiphenyl dissolved in 500 μl absolute ethanol was added to 10 mL sodium citrate buffer (10 mM, pH 6.0) containing ∼200 ods of unmodified 16mer oligo and incubated in a 37°C water bath shaker for 1 h. Unreacted carcinogens were removed by extraction with anhydrous diethyl ether. An acetonitrile gradient of 14–25% in triethylammonium acetate (TEAA) buffer (100 mM, pH 7.0) for 30 min was used to separate the two mono-adduct strands from di-adduct and unreacted 16mer oligos. Supplementary Figure S1A shows the HPLC profile and photodiode array UV spectra of FABP-modified 16mers. The combined collection of the two mono-adduct oligos was lyophilized and stored in −80°C. FAF-modified 16mer oligos were prepared similarly according to published procedures (27) and the resulting HPLC profile is shown in Supplementary Figure S2A.

MALDI-TOF mass spectrometry characterization of mono-adducts

The FABP- and FAF-modified mono-adducted 16mer oligos above were digested by 3′–5′ exonuclease enzyme digestion and characterized by MALDI-TOF mass spectrometry following published procedures (28). Briefly, 1 μL (0.1 unit) of snake venom phosphodiesterase (SVP) was added to a solution of 1 μL of modified oligo (∼100–200 pmol), 6 μl of ammonium citrate (100 mM, pH 9.4), and 6 μl of water for 3′–5′digestion. In each case, 1 μl of the reaction mixture was withdrawn at regular time intervals and was quenched by mixing with 1 μl of matrix (3-hydroxypicolinic acid and diammonium hydrogen citrate in a 1:1 ratio). All MALDI spectra were obtained using Shimadzu Axima Performance MALDI-TOF mass spectrometer equipped with a 70 Hz nitrogen laser in positive ion reflectron mode (15).

UV melting

UV thermal melting experiments were performed on a Cary100 Bio UV-Vis spectrophotometer (Santa Clara, CA) equipped with a 6 × 6 multi-cell chamber (1.0 cm path

length) and connected with a built-in Peltier temperature controller. Samples were prepared in the buffer (containing 0.2 M NaCl, 10 mM sodium phosphate and 0.2 mM EDTA at pH 7.0) in the range of 0.8–5.6 μM. Samples were scanned forward/reverse five times during melting experiments. Thermodynamic parameters were calculated using the program MELTWIN version 3.5 according to previous reports (25).

Differential scanning calorimetry (DSC)

DSC calorimetry measurements were performed on a Nano-DSC from TA Instrument (Lindon, UT, USA) according to previous reports (29). Solutions at 0.1 mM containing unmodified or modified templates with various primers were dissolved in a pH 7.0 buffer (0.1 M NaCl, 20 mM sodium phosphate), and the solutions were incubated at 95°C for 5 min and cooled at room temperature for 3 h to ensure duplex formation. After 15 min degassing, samples were loaded into the instrument cells. A buffer scan was used as control and then subtracted from the sample scan and normalized for heating rate. All samples were scanned against the control buffer from 15 to 85°C at a rate of 1°C/min by measuring five repetitions. Raw data were collected as microwatts versus temperature. ΔH is an integration of ΔCp^{ex} over temperature T. T_m was the temperature at half the peak area. Thermodynamic parameters (ΔG and ΔS) were determined according to previously described procedures (30,31).

Circular dichroism (CD) spectra

Circular dichroism (CD) experiments were conducted on a Jasco J-810 spectropolarimeter (Easton, MD) equipped with a Peltier temperature controller. Typically, a modified template (20 μM) was annealed with an equimolar amount of a primer in 400 μl buffer (containing 0.2 M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA, pH 7.0). The samples were hybridized by heating to 95°C for 5 min and slowly cooled to room temperature for 3 h to ensure duplex formation. Spectra were acquired every 0.2 nm from 200 to 400 nm at a rate of 50 nm/min for five repetitions and smoothed with 25-point adaptive algorithms provided by Jasco (17).

Dynamic ^{19}F-NMR

Approximately 0.1 mM of full and ss/ds duplex samples (Figure 2) were dissolved in 250 μl of NMR buffer (pH 7.0) containing 10% D_2O/90% H_2O, 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA. Solutions were filtered into a Shigemi tube through a 0.22 μm membrane filter. All ^{1}H- and ^{19}F-NMR experiments were performed on a Varian NMR spectrometer with a HFC probe operating at 500.0 and 476.5 MHz, respectively (13,18,32). Imino proton spectra was obtained at different temperature points (5–50°C) using phase-sensitive jump-return sequences and referenced to DSS. ^{19}F-NMR spectra were recorded with increment of 5–10°C (from 5 to 60°C) in the [1]H-decoupled mode and referenced to C6F6 in C6D6 at

−164.9 ppm. Computer line shape simulations were processed as reported previously using MestReNova 10.0.2 (Mestrelab Research, Spain) (13,33).

Surface plasmon resonance (SPR)

SPR binding measurements were conducted using a Biacore T200 instrument (GE Healthcare, Marlborough, MA, USA). Two isomeric 16mer biotinylated DNA template strands (5′-biotin-CTTCTG$_1$*G$_2$TCCTCATTC-3′ and 5′-biotin-CTTCTG$_1$G$_2$*TCCTCATTC-3′; G* = FABP or FAF) were each modified with FABP or FAF (21,27,29,34). HPLC gradient conditions for biotinylated FABP-modified oligos were 16%–25% gradient on acetonitrile for 55 min in TEAA buffer (100 mM, pH 7.0) with a flow rate of 3.0 ml/min (Supplementary Figure S3A). The biotin-FABP-modified sequences were characterized by MALDI-TOF-MS using 3′–5′exonuclease enzyme digestion (Supplementary Figure S3B and C).

Carboxymethylated CM 5 chip was activated by amine coupling kit according to published procedures (21,29,34). Streptavidin (SA, 50 μg/ml dissolved in sodium acetate buffer, pH 4.5) was coated on flow cells 2, 3 and 4 after the injection of EDC/NHS mixture for 7 min. Flow cell 1 was left as blank. Ethanolamine (1 M) was injected over the surface to block unreacted esters. The unbound SA was removed by five pulses of 50 mM NaOH and three pulses of running buffer. The surface was stabilized for 20–30 min in running buffer before DNA coating. Flow cells 2, 3 and 4 were coated around 300 RU of unmodified DNA sequence (TG$_1$G$_2$T), FABP modified (TG$_1$*G$_2$T) and (TG$_1$G$_2$*T) DNA sequences by the manual control mode, respectively. Samples of 9mer to 16mer primer sequence were prepared separately in HBS-P+ buffer (10 mM HEPES, 150 mM NaCl, 0.05% surfactant P20 at pH 7.4) and injected over the chip surface with 100 s contact time and 360 s dissociation time at a flow rate of 15 μl/min at 25°C.

Due to distinct molecule weights, the primers (9mer to 16mer) were used at different concentrations to gain steady-state associations: 300 nM (9mer), 500 nM (10mer), 500 nM (11mer), and 750 nM (12, 13 and 16mer). After the system reached a steady state, excess primers were washed off by the HBS-P+ buffer, and the system was regenerated by adding NaOH (50 nM, 30 s). We were unable to measure the usual K_D (k_d/k_a) kinetics of duplex formation for each elongation. Instead, we used dissociation rates (k_d), which are concentration-independent, to estimate binding strengths for complementary strands.

Dissociation constants (k_d) were analyzed in k_d-alone fitting mode by Scrubber software (version 2.0, Myszka and collaborators, BioLogic Software). The goodness of the fit was determined from the residual standard deviation.

Isothermal titration calorimetry (ITC)

VP-ITC microcalorimeter from MicroCal Inc. (Northampton, MA, USA) was performed to measure full duplex formation of unmodified and modified templates (2.5 μM) with their complementary (25 μM) sequences prepared in 10 mM phosphate buffer (pH 7.0), containing 0.2 M NaCl and 0.2 mM EDTA. The samples were degassed thoroughly

under vacuum using ThermoVac (MicroCal, Northampton, MA, USA) for 20 min. ITC experiments consisted of 25 consecutive injections and each injection involves 10 μl of complementary sequence with injection time 10 and 240 s intervals between each injection with 307 rpm stirring at 25°C. The integrated and normalized areas of raw data were plotted as ΔH of injectant against the molar ratio. The titration data were fitted using a single-site binding model provided in ORIGIN7.0 software supplied by the manufacturer. ΔH, ΔS, K and N (reaction stoichiometry) of 16mer full duplex formation were obtained (35,36).

Molecular modeling

Nucleoside model. The rotational barrier about the nucleoside dihedral angles (θ, φ and ξ; Figure 1A) for both the *anti* and *syn* orientations of FAF and FABP-dG was determined by constraining each dihedral angle in 10° increments from 0 to 360° and performing B3LYP/6–31G(d) optimizations. The interdependence of the conformational energy about θ and φ was investigated, while the conformation about ξ for FABP-dG was only investigated using the lowest energy conformation about θ and φ. All quantum mechanical calculations were performed using Gaussian 09 (revision D.01) (37).

ds DNA model. The initial undamaged helix (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) was generated using the NAB module of AMBER 14 (38). The helix was modified to position the FABP-dG or FAF-dG adduct at G$_1$ or G$_2$ such that the lesion was paired opposite dC. At each position, three orientations of the lesion site were considered to generate the B, S and W adducted DNA conformations. The resulting duplexes were neutralized with Na$^+$ ions and solvated in a TIP3P octahedral water box with sides 8.0 Å from the solute. All natural nucleotides were modeled with the AMBER ff14SB force field (39,40). FABP-dG and FAF-dG parameters were assigned according to the GAFF (41) and AMBER ff14SB force fields using ANTECHAMBER 1.4 (Supplementary Table S1) (42), and partial charges were developed using the RESP charge fitting procedure in the R.E.D.v.III.4 program based on a HF/6–31G(d) calculation (Supplementary Table S1) (43).

Each system was initially minimized with a 500 kcal/mol Å^2 constraint on the DNA for 500 steps of steepest descent and 500 steps of conjugate gradient minimization. Subsequently, 1000 steps of unrestrained steepest descent and 1500 steps of unrestrained conjugate gradient minimization were performed on the entire system. Each system was then heated over 20 ps at constant volume to 300 K. Heating was performed using a Langevin thermostat (γ = 1.0) and a 10 kcal/mol Å^2 constraint on the DNA. Subsequently, each system was simulated for 100 ns at 300 K using a 2 fs time step. To prevent unnatural unraveling of the DNA helices throughout the MD simulations, a 25 kcal/mol Å^2 harmonic distance restraint was used to maintain the heavy atom–heavy atom distance in the Watson–Crick hydrogen bonds for the terminal G:C base pairs between 2.0 and 4.0 Å. During simulations, the SHAKE algorithm was used to constrain covalent bonds involving hydrogen, and the periodic boundary condition was applied. The particle mesh Ewald algorithm was used for long-range electrostatic interactions, with a nonbonded cutoff of 8.0 Å. All simulations and analysis were completed with Amber 14 using the pmemd and cpptraj modules, respectively (38).

RESULTS

Experimental DNA sequences

Chemical structures of ABP and AF, and their fluorine analogues, FABP and FAF, are shown in Figure 1A. In the present study, 16mer (5′-CTTCTG$_1$G$_2$TCCTCATTC-3′) was applied for the $n-1$ and n TLS duplexes. NMR results indicate that the fully paired FABP modified 16mer duplexes exhibited 67%:33% of B:S and 100% B population ratios for G$_1$ and G$_2$ modifications.

As shown in Figure 2, the two 16mer template strands (TG$_1$*G$_2$T or TG$_1$G$_2$*T: G* = FABP or FAF) were each annealed with various lengths of complementary strands (9 to 16mer) to generate two discrete model TLS systems ($n-2$, $n-1$, n, $n+1$, $n+2$, $n+5$ for TG$_1$*G$_2$T and $n-1$, n, $n+1$, $n+2$, $n+3$, $n+6$ for TG$_1$G$_2$*T, respectively; n is the lesion site). In the present study, we used these TLS model systems consistently for ^{19}F-NMR, CD, DSC, UV melting, and SPR experiments. Our strategy is to investigate the effects of the B-/S-conformational heterogeneity on thermodynamic and duplex binding parameters. The results will help us understand the uniqueness of flanking base sequence in the TGGT context in promoting lesion-induced conformational heterogeneity.

MALDI-TOF characterization

FABP- and FAF-modified 16mer oligos were characterized by exonuclease enzyme digestion followed by MALDI-TOF mass spectrometry, according to published procedures (15,28,29). Supplementary Figure S1B shows the MALDI-TOF MS spectra of the 3′-5′ SVP exonuclease digestions of peaks 1 and 2 from FABP-HPLC profile (Supplementary Figure S1A) at different time points (0, 0.5, 2 and 5 min). Briefly, the m/z of 4,960.6 at 0 min represents the control mass to charge ratio of the modified template without SVP. Within 0.5 min of 3′-5′ exonuclease digestion (Supplementary Figure S1B), the lower masses showed up corresponding to the 15mer to 6mer fragments. The fragment at $m/z =$ 1944.6 (theoretical 1945.3), which was the most detectable mass between 2 and 5 min digestion, was assigned to a 6mer containing the FABP lesion. These results confirmed that peak 1 from the HPLC profile was the FABP-modified-G$_1$ sequence (5′-CTTCTG$_1$[FABP] G$_2$TCCTCATTC-3′, TG$_1$*G$_2$T). Peak 2 from HPLC profile was characterized similarly. Supplementary Figure S1C shows the control molecular ion m/z of 4960.2 at 0 min. The m/z 2273.6 ion peak in the 0.5–5 min of digestion corresponded to the 7mer fragment containing FABP-modified guanine. These results confirm peak 2 as the FABP-modified-G$_2$ sequence (5′-CTTCTG$_1$G$_2$[FABP] TCCTCATTC-3′, TG$_1$G$_2$*T).

The FAF-modified 16mer templates have been characterized similarly. Supplementary Figure S2B–C shows the MALDI-TOF MS spectra of the 3′–5′ SVP exonuclease digestion of the FAF-modified 16mer oligo (peaks 1 and

2) at different time intervals. The m/z of 4971.5 (theoretical 4971.0) in peak 1 at 0 min (before digestion, as a control) indicated the 16mer DNA sequence containing a FAF-modified DNA base. From 0 min to 3 min, with the disappearance of the original oligonucleotide (4971.5) the digestion process has been observed when new lower masses showed up corresponding to the smaller fragments from 3′ cleavage (e.g. 4681.7, 4377.9, 3760.3, 3167.0, 2878.1, 2588.9, 2284.6 and 1955.5). After digestion for 3 min, the digestion stopped at $m/z = 1955.4$ (theoretical 1955.1), which represented the FAF-modified 6mer fragment and confirmed peak 1 as FAF-modified-G_1 sequence (5′-CTTCTG$_1$[FAF] G$_2$TCCTCATTC-3′, TG$_1$*G$_2$T). Following the same procedure, we found the SVP digestion of peak 2 stalled at $m/z = 2284.0$ (theoretical 2284.3) after 3 min and confirmed peak 2 as FAF-modified-G_2 sequence (5′-CTTCTG$_1$G$_2$[FAF]TCCTCATTC-3′, TG$_1$G$_2$*T). Taken together, the SVP enzymatic digestion from the 3′-end is slowed down significantly at lesion sites of 6mer and 7mer representing TG$_1$*G$_2$T and TG$_1$G$_2$*T, respectively.

Dynamic ^{19}F-NMR spectroscopy

Dynamic ^{19}F-NMR experiments were performed to measure the extent of conformational heterogeneity during primer elongation. Figure 3 and Supplementary Figure S4 respectively show the ^{19}F-NMR spectra of the FABP- and FAF-modified duplex series as function of temperature (5–60°C). The same sequence systems were used for CD, DSC UV-melting, and SPR experiments for comparison. Briefly, FABP-G_1*-16mer duplex showed a mixture of S- (33%) and B- (67%) conformations at 25°C (highlighted in Figure 3A). By contrast, the corresponding FABP-G_2*-16mer duplex exclusively revealed B-conformer (100%) with no sign of conformational heterogeneity.

*FABP-TG$_1$*G$_2$T.* In Figure 3A, no indication of conformational heterogeneity at $n − 2$ (9mer) and $n − 1$ (10mer) was found, whereas the 3′-end of the primer did not reach the lesion site. A slight heterogeneity was observed at n (11mer). Interestingly, however, it disappeared at $n + 1$. At $n + 2$ (16/13mer), we observed complex ^{19}F signals indicating the presence of multiple conformations. These ^{19}F signals were coalesced into a sharp single peak around −116.5 ppm upon around and >50°C; this result indicates duplex melting. The conformational pattern was steadied at fully paired 16mer duplex ($n + 5$), which showed a ratio of 33% S and 67% B populations at 25°C (highlighted in Figure 3A).

The ^{19}F-signals were assigned according to the general protocols based on relative chemical shifts, chemical exchange patterns, and H/D isotope effects (27). We have investigated a number of ^{19}F-NMR spectra of arylamine-modified duplexes in various sequence settings and found that the fluorine of the S-type conformer resonates always upfield relative to that of the external binding B-type conformer and they are in slow exchange and eventually coalesce into a denatured ss. In addition, the ^{19}F tag in the B-conformation is fully exposed to the solvent in the major groove area, whereas that of the S-conformer is shielded from van der Waals interactions with neighboring base pairs. In addition, the fully exposed ^{19}F tag of the ma-

jor groove binding B-conformer always exhibits significant deuterium isotope effect (0.18 ppm) going from 10% to 100% D$_2$O (Supplementary Figure S5A–B). This result is in contrast to the buried ^{19}F of the stacked conformer, which shows no isotope effect (0.08 ppm) (Supplementary Figure S5A).

*FABP-TG$_1$G$_2$*T.* G$_2$-modfied series shown in Figure 3B is quite different from the G$_1$-modified above. Some minor heterogeneities are observed on $n − 1$ (16/9mer) and $n + 3$ (16/12mer); in general, all duplex sequences exhibited one major signal around −116.5 ppm. This signal can be assigned as the B-like conformer because of its close proximity to the coalescent signal (single strand) and also by comparison with the G$_1$ modified duplex above.

*FAF-TG$_1$*G$_2$T and FAF-TG$_1$G$_2$*T.* Supplementary Figure S4 shows the dynamic ^{19}F-NMR spectra for the FAF-modified series. The results indicate both FAF-TG$_1$*G$_2$T and FAF-TG$_1$G$_2$*T duplexes adopted more S-conformation than the FABP series. FAF-TG$_1$*G$_2$T had approximately 42% more S than the FABP counterpart TG$_1$*G$_2$T (75% S and 25% B) in the full duplex, whereas the S increased from 0% to 17% in the FAF/TG$_1$G$_2$*T. These results reflected the planar nature of FAF relative to FABP preferring to intercalate/stack. Moreover, the B-conformer exhibits a significant deuterium isotope effect (~ 0.18 ppm), going from 10% to 100% D$_2$O, whereas the stacked conformer shows little isotope effect (0.04 ppm) (Supplementary Figure S6A and B).

Induced circular dichroism (ICD)

Supplementary Figure S7 shows CD overlays of unmodified and FABP-modified 16mer templates with increasing primer length. The S curve consisted of +@270 nm, and the B-form DNA duplex consisted of −@250 nm. Given the conformational heterogeneity, FABP-modified TLS duplexes exhibited lesion ICD between 290 and 360 nm (19,25). This result is in contrast to the unmodified control series, which showed no such ICD. The ICD gradually increases along with the primer elongation from 9mer to 16mer. In the fully paired 16mer duplex, the CD data agrees with the above ^{19}F-NMR data, which showed a B/S-conformeric TG$_1$*G$_2$T and exclusively B-conformeric TG$_1$G$_2$*T on fully paired 16mer duplexes. The predominant B-conformer of TG$_1$G$_2$*T highly affects the bending of DNA helix around the adduct site. G$_1$ and G$_2$ FABP-modified duplexes displayed slight blue shifts ($\sim$3 nm) relative to the unmodified controls, indicating adduct-induced DNA bending (28). Similar results were obtained for FAF-modified series (Supplementary Figure S8).

UV melting experiments

The UV-melting profiles of all the ss/ds duplex models were measured at the $0.8−4.0\,\mu$M range for FABP and FAF samples. The UV monophasic and sigmoidal melting curves between lnCt and T_m^{-1} ($R^2 > 0.9$) confirmed that the duplexes adopted a typical helix-coil transitions (13). Supplementary Figure S9 shows the plots of melting (T_m) for FABP and
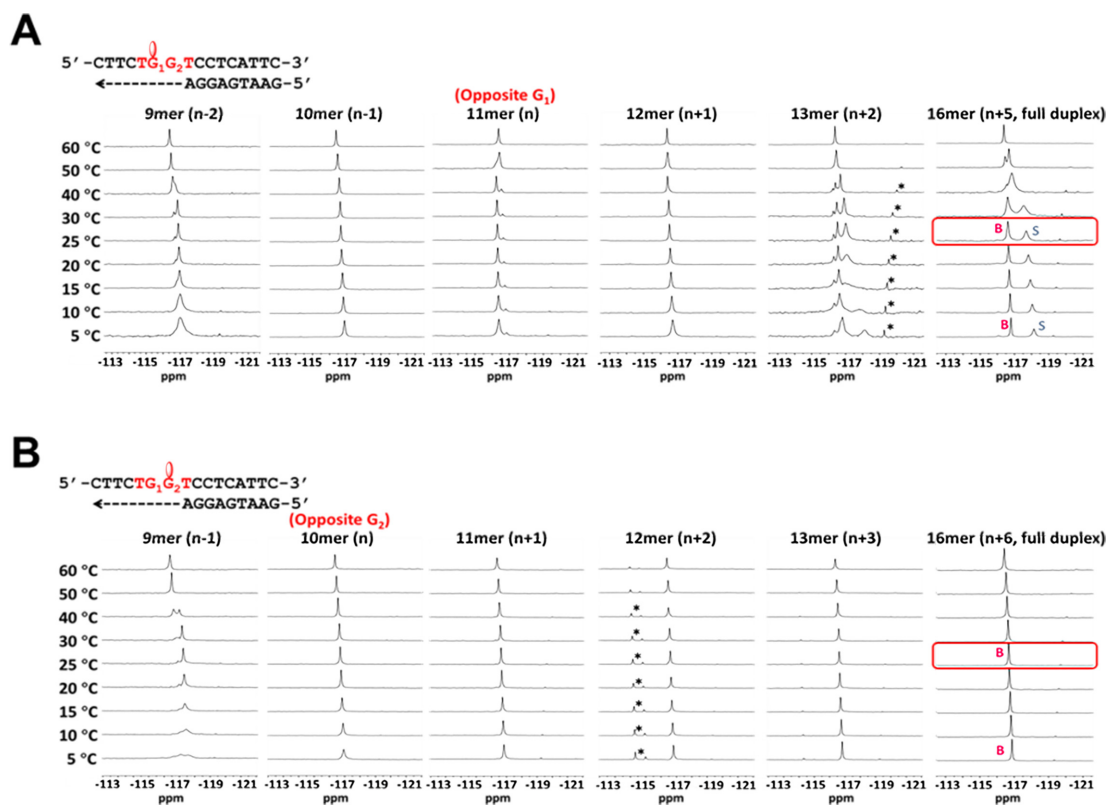
Figure 3. Dynamic ^{19}F-NMR spectra (5–60°C) of FABP-modified duplexes with primer elongation. (**A**) FABP-modified G_1 (TG*GT) showing a 67% B- and 33% S-conformational mixture in full duplex. (**B**) FABP-modified G_2 (TGG*T) exhibiting 100% B-conformer in full duplex. For all the duplexes, the ^{19}F signals around 60°C coalesced into a sharp single peak around -116.5 ppm indicating a denatured duplex. * indicates unassigned minor conformers.

FAF lesions at the TG_1*G_2T and TG_1G_2*T systems (red dash lines and blue dash lines, respectively) relative to the unmodified controls (dark gray solid lines) as a function of primer elongation for TLS systems. Thermal and thermodynamic parameters calculated from UV melting curves are summarized in Supplementary Tables S2 and S3. For all the modified and unmodified duplexes, T_m increases with the elongation of primers. However, modified systems showed stalls around the lesion sites.

The lesion effect in the TG_1*G_2T system was minimal at one and two bases before the lesion sites (ΔT_m of 0.35 and 0.67°C for 9mer and 10mer, respectively). This finding is in contrast to that of the TG_1G_2*T system, which showed minimal lesion effect at those sites (9mer ΔT_m of 1.45°C). However, the effect was significant when the primer elongation was opposite the lesion [11mer and 10mer for TG_1*G_2T ($\Delta T_m = -2.54$°C) and TG_1G_2*T ($\Delta T_m = -5.21$°C), respectively]. In full duplex (16mer), TG_1*G_2T showed greater negative thermodynamic effect to TG_1G_2*T (ΔT_m of -7.68°C to -6.29°C), which can be due to the S-conformer (see ^{19}F-NMR results above).

DSC

DSC experiments were performed for the FABP-modified TLS models (TLS/TG_1*G_2T versus TLS/TG_1G_2*T). Figure 4 shows DSC thermogram overlays of excess heat capacity C_p^{ex} versus temperature. T_m is the maximum point of the

curves, and transition enthalpy (ΔH) is the area under the curves. DSC results are summarized in Table 1 (29,44).

Control TLS. Figure 4A shows the DSC thermograms for the control TLS system. For the natural DNA, T_m and ΔH values increased progressively with increasing primer length from 9mer ($T_m = 42.9$°C, $\Delta H = -71.4$ kcal/mol) to 13mer ($T_m = 59.8$°C, $\Delta H = -87.2$ kcal/mol) and increased significantly at full duplex (16mer, $T_m = 63.1$°C, $\Delta H = -98.5$ kcal/mol).

$TG_1[FABP]G_2T$ Duplex. In FABP-G_1 profiles (Figure 4B), the 9mer primer, which is two bases ($n - 2$) before the lesion (n), showed a minimal lesion effect, i.e. $T_m = 43.3$°C compared with the T_m of the control duplex of 42.9°C. However, 10mer ($n - 1$) to 16mer ($n + 5$) primers were stalled, i.e. their T_m values were clustered between 51.0 and 53.0°C. The 16mer full duplex showed a significant increase in enthalpy ΔH, indicating a full duplex formation.

$TG_1G_2[FABP]T$ Duplex. DSC patterns for the TG_1G_2*T system (Figure 4C) were similar to those of the TG_1*G_2T. Curves for 10mer (n) to 13mer ($n + 3$) were clustered between T_m 45.0°C and 53.0°C, which is lower than that of the TG_1*G_2T counterpart. The increase in ΔH was not as incremental as that of T_m. In general, the exclusively B-conformeric 16mer TG_1G_2*T full duplex showed greater T_m and ΔH than the G_1 counterpart that contains a mixture of B- and S-conformers (see ^{19}F-NMR above). This
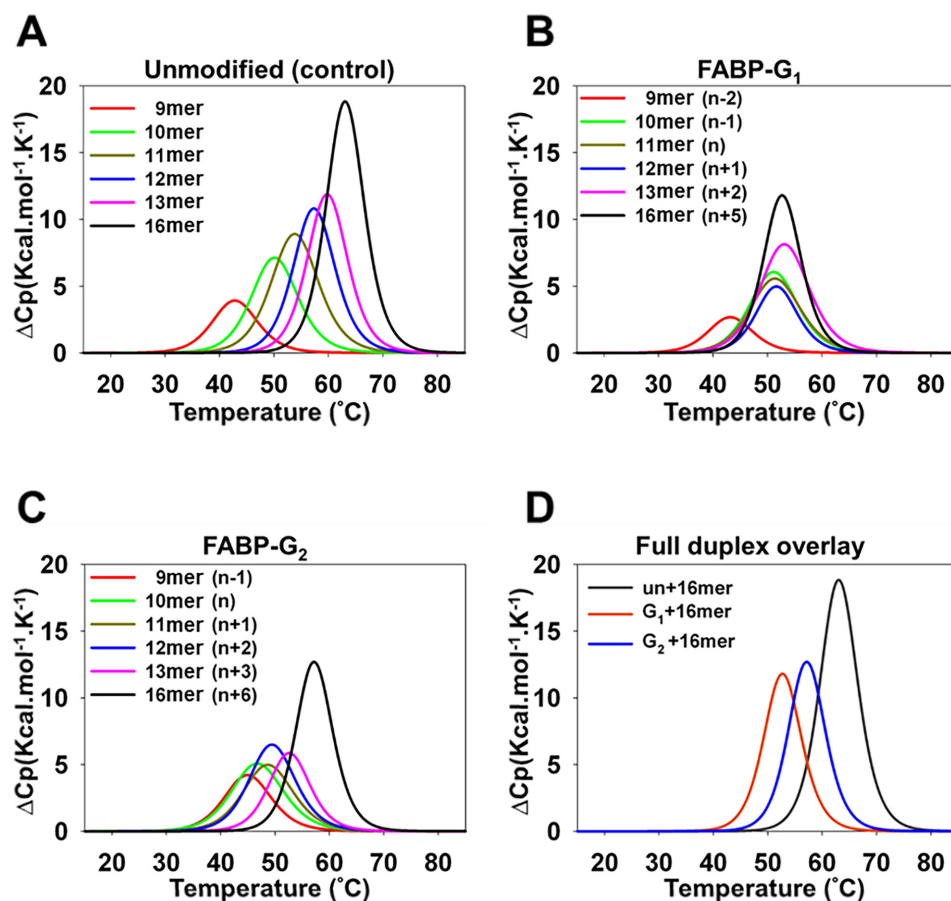
Figure 4. DSC heat capacity ΔC_p versus temperature (15–85°C) for primer elongation (9mer to 16mer) of the (**A**) control, (**B**) FABP-modified G_1 (TG*GT) and (**C**) FABP-modified G_2 (TGG*T). (**D**) Full duplex overlays of unmodified and FABP G_1/G_2-modified sequences. FABP-G_1 and FABP-G_2 show smaller T_m and ΔH than the control/unmodified in full duplex exhibiting that S-conformer (FABP-G_1) might produce thermodynamic destabilization in full duplex.

Table 1. Thermal and thermodynamic parameters derived from DSC

Sequence	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G_{37°C}$ (kcal/mol)	$T_m{}^a$ (°C)	$\Delta\Delta H°{}^b$ (kcal/mol)	$\Delta\Delta G_{37°C}{}^c$ (kcal/mol)	$\Delta T_m{}^d$ (°C)
FABP-G_1							
9mer	71.8 (71.4)	207.3 (206.1)	7.6 (7.5)	43.3 (42.9)	-0.4	-0.1	0.4
10mer	72.1 (72.1)	202.5 (203.4)	9.3 (9.0)	51.3 (50.0)	0.0	-0.3	1.3
11mer	69.6 (74.4)	194.8 (207.9)	9.2 (9.9)	51.5 (53.8)	4.8	0.7	-2.3
12mer	70.0 (82.0)	195.8 (228.7)	9.3 (11.1)	51.6 (56.9)	12.0	1.8	-5.3
13mer	71.6 (87.2)	199.5 (242.2)	9.7 (12.1)	53.4 (59.8)	15.6	2.4	-6.4
16mer	90.6 (98.5)	258.3 (273.4)	10.5 (13.8)	52.8 (63.1)	7.9	3.3	-10.3
FABP-G_2							
9mer	66.5 (71.4)	189.5 (206.1)	7.8 (7.5)	45.0 (42.9)	4.9	-0.3	2.1
10mer	64.6 (72.1)	182.1 (203.4)	8.1 (9.0)	46.9 (50.0)	7.5	0.9	-3.1
11mer	67.8 (74.4)	191.1 (207.9)	8.6 (9.9)	48.5 (53.8)	6.6	1.3	-5.3
12mer	74.8 (82.0)	212.1(228.7)	9.1 (11.1)	49.7 (56.9)	7.2	2.0	-7.2
13mer	84.4 (87.2)	239.2 (242.2)	10.2 (12.1)	52.8 (59.8)	2.8	1.9	-7.0
16mer	96.7 (98.5)	273.0 (273.4)	12.1 (13.8)	57.3 (63.1)	1.8	1.7	-5.8
Standard deviation	±3.0	±8.0	±0.3	±0.3			

[a] T_m value is the maximum point of the DSC curve.
[b] $\Delta\Delta G = \Delta G$ (modified duplex) − ΔG (control duplex).
[c] $\Delta\Delta H = \Delta H$ (modified duplex) − ΔH (control duplex).
[d] $\Delta T_m = T_m$ (modified duplex) − T_m (control duplex).

finding suggests that the B-conformation can be a thermo-dynamic destabilizer during primer elongation and yet the S-conformer might produce thermodynamic destabilization in the full duplex. TG_1*G_2T and TG_1G_2*T duplexes in TLS settings show consistently smaller T_m and ΔH compared with the unmodified controls (Figure 4D).

ITC for duplex formation

ITC provides the nature of duplex formation through a sigmoidal curve and the respective thermodynamic parameters (ΔH, ΔG, ΔS) through plots of heat capacity (C_p) versus molar ratios. The results indicate that the lesion destabilizes the duplex formation.

Figure 5 shows ITC thermograms of duplex formation for FABP- and FAF-modified TGGT sequences at 25°C. In all cases, unmodified controls (black solid line) showed well-defined sigmoidal curves with sharp changes in heat capacity at starting and ending points. The results indicate the typical cooperative 1:1 binding process (e.g. molar ratio of $N = \sim 1$) expected for formation of Watson–Crick-based DNA duplexes. By contrast, the ITC curves for modified full duplexes were not well defined in terms of stoichiometry and overall shape, probably due to heterogeneity. Supplementary Table S4 summarizes the ITC-derived thermodynamic parameters ΔH, ΔS, K and N for unmodified and modified sequences, for example, the binding constant order is TGGT (49.9×10^7 M^{-1}) > TG_1*G_2T (18.7×10^7 M^{-1}) > TG_1G_2*T (5.9×10^7 M^{-1}) in FABP case. In general, the ΔH and the binding constant of the modified sequences are consistently lower than that of the controls, and the FABP shows more variances in shapes than the FAF counterpart.

SPR on DNA duplex formation

SPR experiments were conducted to probe the effect of conformational heterogeneity on duplex formation and dissociation. The results were obtained in a buffer-only environment without a polymerase. With different concentrations used for each complementary strand, we measured dissociation rates (k_d) to evaluate their binding strengths (see Materials and Methods).

Supplementary Figure S10 shows the entirety of SPR results, and Table 2 lists the respective dissociation rate constants. Supplementary Figure S11 shows the overlay of normalized SPR sensograms for FABP and FAF lesions on the unmodified (green) and the TG_1*G_2T (blue)/TG_1G_2*T (purple) sequences with increasing length of complementary strands. As expected, in each complementary sequence, SPR data show gradual reduction in k_d values with elongation in unmodified and modified sequences.

Rapid duplex formation took place for all duplexes; however, different dissociation rates were observed for modified duplexes. In general, slower dissociation indicates greater duplex binding with thermodynamic stability. This appears to be generally true with post-lesion duplexes (16/11 to 16/16mer). The 16/9mer duplexes, which represents the prelesion $n - 2$ and $n - 1$ site for G_1 and G_2, respectively, showed generally faster dissociation (Figure 6A, Table 2): $TG_1G_2[FABP]T$ (7.1×10^{-2} s^{-1}) < $TG_1[FABP]G_2T$ (12.0×10^{-2} s^{-1}) < unmodified TG_1G_2T (16.5×10^{-2} s^{-1}). As

for the 16/10mer (Figure 6B), however, dissociation was dramatically slower for $TG_1[FABP]G_2T$ (0.343 s^{-1}) than the unmodified (0.515 s^{-1}) and $TG_1G_2[FABP]T$ (2.370 s^{-1}). These results are consistent with duplex stability in the order of TG_1*G_2T > unmodified > TG_1G_2*T in FABP case. The 16/9mer is one base before ($n - 1$) the B/S-conformeric TG_1G_2*T, and 16/10mer is one base ($n - 1$) before the TG_1*G_2T duplex. The results suggest that the modified $n - 1$ duplexes may form a stable S-conformer at the ss/ds junction, resulting in slow dissociation. Similar SPR patterns were observed for the FAF lesion (Supplementary Figure S11B); however, general dissociation rates were slightly faster compared to the FABP lesion. This finding indicates that FABP is a better stacker in the ss/ds duplex setting.

Molecular modeling

Molecular modeling was used to understand the conformation of the lesion, and the strength of discrete interactions between the lesion and the surrounding environment. Through using various model sizes, we can better understand the effects of the different compositions of the FABP and FAF bulky moieties on the conformational preferences in the DNA environment.

Nucleoside model. To investigate the difference in the preferred orientation of the bulky moiety with respect to the nucleobase for FAF-dG and FABP-dG, the conformational energy was mapped as a function of the dihedral angles that dictate rotation about the C–N–C linker (i.e. θ and φ, Figure 1) for both the *anti* and *syn* orientations of the lesions (Supplementary Figure S12). Due to the symmetry within the FABP bulky moiety, the calculations revealed one minima for *anti*-FABP that positions the first ring of the bulky moiety planar to the adducted dG (θ and φ = 0°, Supplementary Figure S13). *syn*-FABP has a greater degree of conformational flexibility, with three minima about θ. In the global minimum, the first ring is planar to the adducted G, while the bulky moiety is twisted by 70° with respect to the adduct G in the two additional local minima. The rotational barrier between these minimum energy orientations is low (< 2.5 kcal/mol). Based on the lowest energy orientation of *anti* and *syn*-FABP-dG, the two rings in the FABP moiety were determined to preferentially adopt a twisted orientation (i.e., ξ, the angle between the two rings is ~40°, Supplementary Figure S14). Nevertheless, the rotational barrier within the bulky moiety is small, being only ~2.5 kcal/mol.

FAF-dG displays the same conformational preference about the θ and φ dihedral angles as FABP-dG (Supplementary Figure S12), although the rotational barrier with respect to θ is slightly (~ 1 kcal/mol) higher for *syn*-FAF-dG than *syn*-FABP-dG. Since the FAF bulky moiety is not symmetrical like FABP, the two minima with respect to the φ dihedral angle represent distinct conformations (Supplementary Figure S13). Specifically, φ = 0° positions the methylene linker towards N7, while φ = 180° positions the linker towards N9. The rotational barrier about φ (up to ~7 kcal/mol) is substantially higher than about θ (up to ~3.5 kcal/mol). Due to the nature of the methylene linker between the two rings, the bulky moiety of FAF-dG is planar in contrast to the twisted bulky moiety of FABP-dG. This
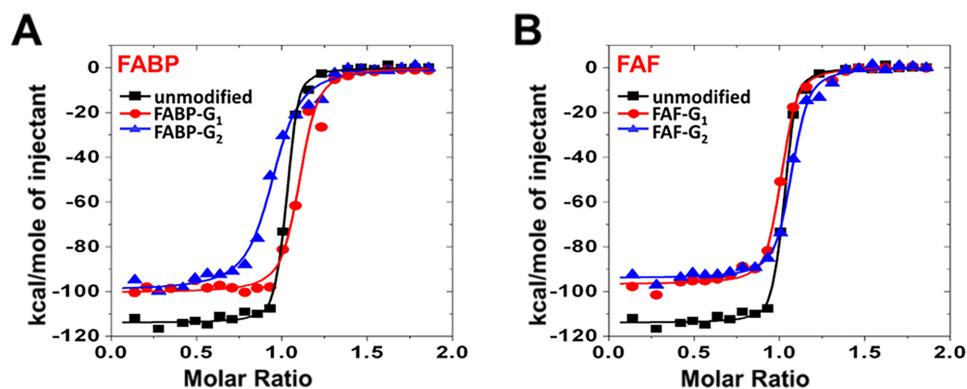
Figure 5. ITC binding thermograms of duplex formation for (**A**) FABP- and (**B**) FAF- modified 16mer TGGT full duplexes at 25°C.

Table 2. Dissociation off-rate constants (k_d, $\times 10^{-2}$, s^{-1}) from SPR

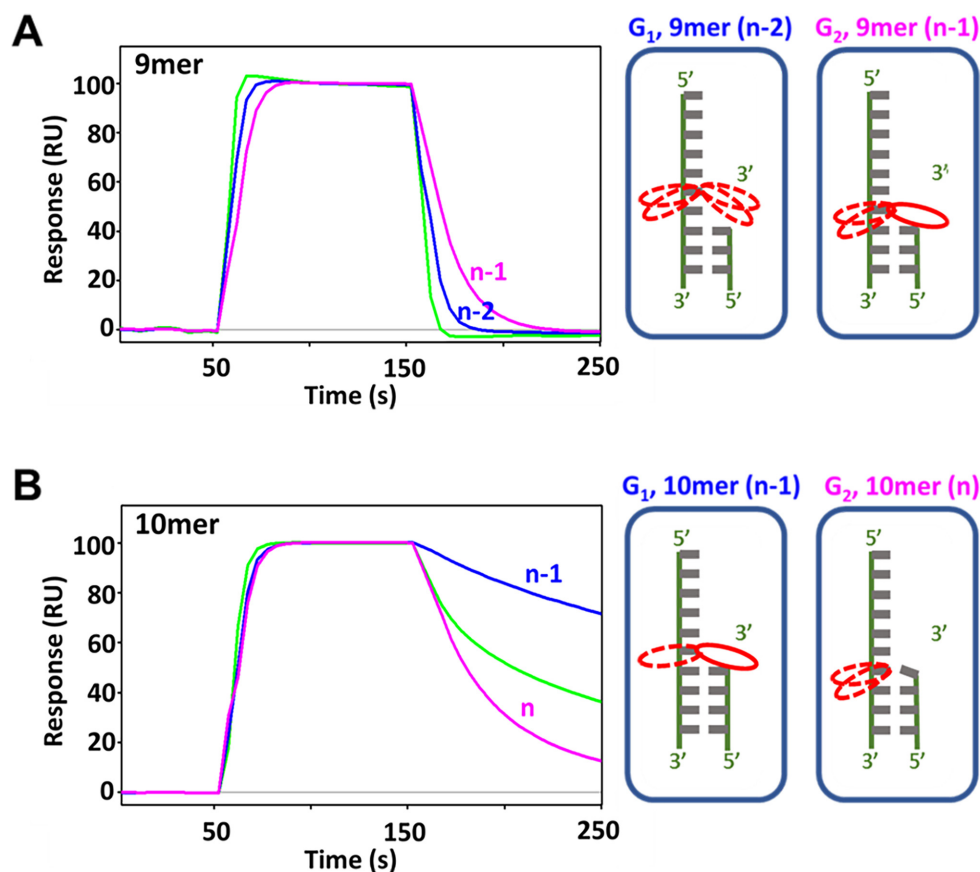|		FABP			FAF		
Sequence	[Conc.] (nM)	Unmodified	FABP-G_1	FABP-G_2	Unmodified	FAF-G_1	FAF-G_2
9mer	300	16.500	12.000	7.100	28.000	13.600	9.000
10mer	500	0.515	0.343	2.370	0.883	0.384	2.810
11mer	500	0.072	0.216	0.769	0.185	0.321	1.960
12mer	750	0.028	0.139	0.361	0.083	0.261	1.290
13mer	750	0.020	0.073	0.122	0.041	0.169	0.382
16mer	750	0.013	0.049	0.041	0.021	0.082	0.086

*Errors are within the limits of 5%.

Figure 6. Normalized SPR sensorgrams of FABP-modified 16mer TG$_1$G$_2$T sequences with 9mer and 10mer primers, and progression of lesion conformational flexibility (hatched lines) and conformational rigidity (solid lines). (**A**) 9mer and (**B**) 10mer. Slower dissociation shows better stability of the primer−template duplexes. Unmodified (green); G$_1$-modified (blue); G$_2$-modified (purple). HBS-P+ buffer is the dissociation solution. The green unmodified trace cannot be labeled as either *n* or *n* − 1 because there's no lesion present.

difference will likely influence the adducted DNA conformational preference.

DNA model. To gain more information about the preferred conformation of FABP- and FAF-dG at the G_1 and G_2 positions in the 16mer duplexes, MD simulations were performed with the lesion opposite dC for duplexes adopting the B- or S- conformation. All simulations showed a stable rmsd throughout the 100 ns simulation (Supplementary Table S5). Regardless of the lesion, duplex conformation or modification position, minimal distortions were observed to the overall helical structure and the flanking base pairs formed canonical Watson-Crick hydrogen bonds throughout the simulation (Supplementary Table S6 and Supplementary Figures S15 and S16). Furthermore, a consistent conformational preference occurred at the lesion site in terms of the glycosidic orientation or the orientation of the bulky moiety with respect to the adducted dG (as defined by the χ, θ and φ dihedral angles) regardless of the lesion or position (Supplementary Table S5).

In the B-conformer of the FABP-modified 16mer duplexes, the adduct maintains Watson–Crick hydrogen bonding with the pairing dC (Supplementary Table S7), and the bulky moiety is positioned in major groove parallel to the adducted G regardless of the lesion position (θ and $\varphi \approx 0°$, Supplementary Table S5 and Figure 7A). There is free rotation about the linker between the two bulky moiety rings (i.e. standard deviation in ξ of up to $\sim$115°, Supplementary Table S5). The preference about θ, φ and ξ for FABP-dG in the helix is consistent with the inherent conformational preference predicted for the nucleoside model. Regardless of whether the lesion is at G_1 or G_2, the hydrogen-bonding and stacking interactions between the adducted and flanking base pairs are similar (average interaction strength of approx. –98 kcal/mol, Supplementary Table S8), and the duplexes are destabilized by the solvation of the non-polar bulky moiety by a similar extent (SASA $\approx$ 155 Å^2, Supplementary Table S5). Due to the similar stabilization (lesion site interactions) and destabilization (solvent accessibility of the bulky moiety) effects, the MMPBSA binding energies of the modified 16mer duplexes are within 2 kcal/mol, with the G_1 modification being less stable (Figure 7A). Overall, there is little difference in the structure and stability of the B-conformeric duplex modified at G_1 and G_2.

In the FABP-modified S-conformeric duplex, the pairing dC is positioned in the major groove and forms a hydrogen bond with the DNA backbone of the 3′ flanking base for up to 33% of the simulation (Supplementary Table S7). In the S-conformer, the bulky moiety intercalates into the helix; however, the bulky moiety position is sequence dependent (Figure 7B). Specifically, at G_1, the adducted G is positioned in the major groove to afford stacking between both rings of the bulky moiety and the flanking bases (approx. –24 kcal/mol, Supplementary Table S8). Furthermore, the lesion position is stabilized by a hydrogen bond between the linker N–H and O4′ of the adducted nucleotide (for 63% of the simulation, Supplementary Table S7). Consequently, this positioning of the lesion reduces the SASA of the hydrophobic bulky moiety compared to the B-conformeric duplex (by $\sim$120 Å^2, Supplementary Table S5). To maintain the lesion site stacking interactions, the two rings in the

FABP bulky moiety maintain a planar relative orientation ($\xi = 174.4 \pm 22.7°$, Supplementary Table S5). In contrast, at G_2, the adducted dG stacks with the flanking bases and the bulky moiety is positioned towards the minor groove, which results in minimal intrahelical stacking with the terminal ring of the bulky moiety. This permits free rotation about the linker between the two bulky moiety rings ($\xi = 53.2 \pm 138.8°$, Supplementary Table S5). Although significant stacking interactions occur between the lesion and the flanking bases (approx. –24 kcal/mol, Supplementary Table S8), the bulky moiety position at G_2 does not shield FABP from solvent to the same extent as at G_1 (difference of $\sim$30 Å^2). Furthermore, at G_2, due to the position of the lesion, a unique hydrogen bond occurs between the linker N–H of FABP-dG and N2 of G_1 (24% occupancy). Since this interaction is not as persistent and has a longer distance (by $\sim$0.4 Å) than the hydrogen bond with the linker at G_1 (Supplementary Table S7), this may contribute to the destabilization of the helix for G_2* relative to G_1*. Due to the decreased stabilization (reduced hydrogen bonding) and greater destabilization (increased solvent exposure of FABP) at G_2, the binding energy of the G_2 modified 16mer is 6.5 kcal/mol less than that of the G_1 modified duplex (Figure 7B).

For FAF-dG adducted DNA in the B-conformation, the lesion maintains Watson-Crick hydrogen bonding with the pairing dC (Supplementary Table S7), and the bulky moiety is positioned in the major groove and parallel to the adducted dG (θ and $\varphi \approx 0°$, Supplementary Table S5 and Figure 7C). At G_2, FAF-dG is stabilized by stacking and hydrogen-bonding interactions to a similar extent as seen for FABP-dG (approx. –98 kcal/mol, Supplementary Table S8). However, at G_1, weaker stacking interactions occur between the adducted and flanking base pairs (by $\sim$ 3.5 kcal/mol). Furthermore, although the SASA of the FAF-moiety is inherently larger than that of FABP (Supplementary Table S5), the SASA of FAF-dG at G_1 is greater than the SASA of the lesion at G_2 (by $\sim$10 Å^2). Therefore, due to decreased stabilization of the lesion site through hydrogen bonding and stacking interactions, and increased destabilization due to greater solvent exposure of the hydrophobic moiety, the helix binding energy of the FAF-modified duplex in the B-conformer with the lesion at G_1 is 6.8 kcal/mol less than that at G_2 (Figure 7C).

When the FAF-modified 16mer adopts the S-conformer, the pairing dC is positioned in the major groove and forms a hydrogen bond with the backbone of the 3′ flanking base (up to 71% of the simulation, Supplementary Table S7 and Figure 7D), regardless of the modification position. In the S-conformer, at both G_1 and G_2, the bulky moiety adopts a planar geometry and FAF-dG is intercalated into the helix such that both the adducted G and bulky moiety to stack with the flanking bases (θ and $\varphi \approx 0°$, Supplementary Table S5 and Figure 7D). As a result, the SASA of the hydrophobic moiety ($\sim$69 Å^2, Supplementary Table S5) and the magnitude of the lesion site interactions (approx. –68 kcal/mol, Supplementary Table S8) are independent of the damage site. As a result, the binding energies for the duplexes are similar, with the G_1 modification being 2 kcal/mol less stable than the G_2 counterpart. It is also noteworthy that the stacking interactions at the lesion site for the FAF-modified
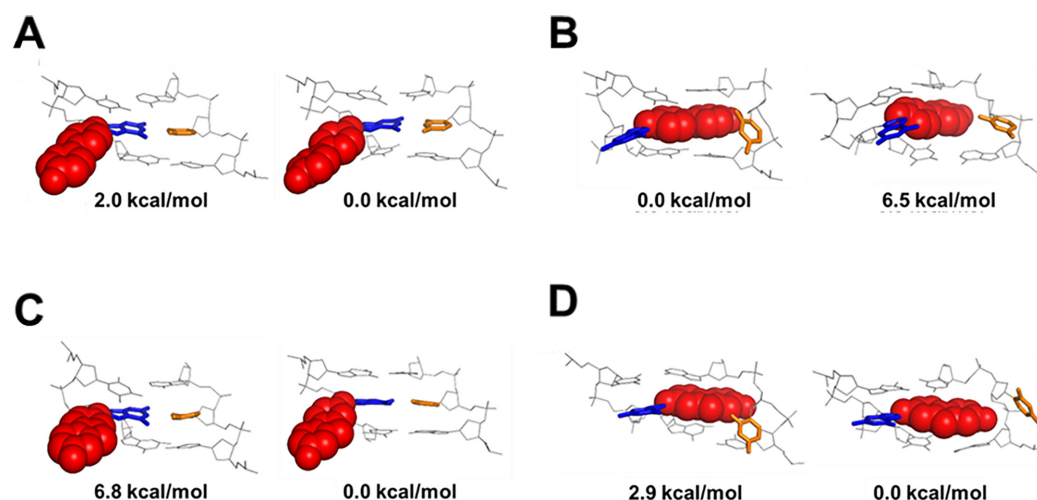
Figure 7. Representative MD structures and associated relative helix binding energies of with the lesion at G_1 (left) or G_2 (right) of (**A**) FABP-modified duplexes, (**B**) FABP-modified duplexes in the S-conformer, (**C**) FAF-modified duplexes in the B-conformer and (**D**) FAF-modified duplexes in the S-conformer.

duplex in the S-conformer are up to 6 kcal/mol stronger than those in the FABP-modified helices and the flanking bases in the S-conformer (Supplementary Table S5).

DISCUSSION

Unusual flanking sequence effects on the TG_1G_2T sequence

In this study, we present a striking flanking sequence effect on a 16mer fully paired DNA duplex 5′-CTTCTG$_1$G$_2$TCCTCATTC-3′, in which lesion modification at G_1 (TG$_1$*G$_2$T sequence) and G_2 (TG$_1$G$_2$*T sequence) yielded a dramatic difference in conformational heterogeneity. Dynamic ^{19}F-NMR results indicate that the FABP-modification at G_1 and G_2 resulted in 67%:33% of B:S and 100% B population ratios, respectively. Modification with the planar FAF in the same sequence contexts exhibited greater S-conformer population: 25%:75% B:S and 83:17% B:S population ratios for the G_1 and G_2 positions, respectively. This represents a significant increase in the S-conformation from 33% to 75% and from 0% to 17% going from FABP to FAF in the TG$_1$*G$_2$T and TG$_1$G$_2$*T sequence context, respectively. The greater S-conformer population of FAF over FABP at duplex DNA can be attributed to their aromatic coplanarity, i.e. the single-carbon atom methylene bridging between two aromatic rings in FAF restricts the twisting of the biphenyl. Indeed, nucleoside modeling showed that the FABP-dG lesion is twisted by ~ 40°, while FAF-dG is inherently planar.

This sequence effect is analogous to our recent reporting on a 11mer DNA duplex sequence 5′-CCATCG*CNACC-3′ (N = T or A), in which the FABP modification exhibited a 40%:60% ratio of B:S and 100% B population ratios, and the FAF modification resulted in 10%:90% and 34%:66% B:S in the G*CT and G*CA sequence context, respectively (27). These 3′-next flanking sequence effects influence primer kinetics, DNA replication, and repair (11). Lesion stacking was primarily responsible for the greater population of the stacked S-conformer compared with that

of the G*CA counterpart; the S-conformation decreased the binding affinity of the complementary strands. These adduct-induced conformational heterogeneities are caused by 3′-next flanking sequence effect, which is different from the TG$_1$G$_2$T context considered in the present study, which corresponds to flanking sequence effects.

Our calorimetric results showed a greater thermodynamic stability for the B-conformeric G_2 over the S/B-conformeric G_1 in the full duplex, which indicates that the B-conformer is a thermodynamic stabilizer in fully paired duplex DNA. As shown in ITC curves (Figure 5), lesion modification has reduced enthalpic energies (~13% and ~17% average for FABP- and FAF-G$_1$/G$_2$, respectively) during duplex formation, as expected. The lesion destabilizes the duplex formation. Correlating the ITC data with the conformation heterogeneity is difficult; hence, the difference in ring planarity had a major effect on the shape of binding isotherms (C), i.e., FABP has greater sequence differences compared with FAF.

Modeling data indicates that the conformational preference of FABP-dG is due to a destabilization of the S-conformer at G_2 compare to the complex at G_1. Specifically, the S-conformer of FABP-dG modification is positioned with the adducted G in the major groove at G_1, while the bulky moiety was positioned in the minor groove at G_2. This differential positioning of FABP-dG in the S-conformer lead to reduced stabilization (hydrogen bonding) and greater destabilization (greater SASA of the hydrophobic bulky moiety) at G_2 and provides a structural explanation for the reduced S-conformer at G_2. Conversely, the conformational preference of FAF-dG is due to a destabilization of the B-conformer at G_1 relative to G_2. As seen for the S-conformer of FABP-dG modified DNA, the destabilization of the B-conformer with modification at G_1 is due to the reduced stabilization (stacking) and greater destabilization (greater SASA of the hydrophobic bulky moiety) at G_2, which provides a structural explanation for the observed reduced B-conformer contribution at G_1. Nevertheless, re-

gardless of the position or modification, MD predicts the adducted DNA duplex to be more stable in the B-conformer at the G_1 and G_2 positions than the S-conformer, which explains the thermodynamic data. Finally, comparison of the lesion site interactions for FABP-dG and FAF-dG modified DNA reveals that planar FAF affords stronger stacking between the lesion and flanking base pairs (by up to 6 kcal/mol) than twisted FABP. This rationalizes the greater occurrence of the S-conformation for FAF-dG adducted DNA than FABP-dG adducted DNA.

Thermodynamic effect on lesion heterogeneity during primer elongation

We studied six discrete DNA replication model duplexes prepared by annealing the 16mer TG_1*G_2T and TG_1G_2*T ($G* =$ FABP or FAF; Figure 2) modified template with appropriate primer strands. These experiments were performed in buffer solutions without a polymerase. The main objective of these experiments is to study the effect of the lesion-induced S/B-heterogeneity on their thermal and thermodynamic stabilities during the primer elongation that mimics a TLS.

The thermodynamic results show a systemic increase in T_m and ΔH with primer elongation from 9mer to 16mer in unmodified control duplexes. However, the FABP modification generally as a lower T_m and ΔH, as well as changed their calorimetric curve patterns for the TG_1*G_2T and TG_1G_2*T systems. As mentioned above, for full duplexes, the exclusively B-conformeric G_2* was more thermally stable than G_1*. However, this finding is not observed during the primer elongation from 10mer to 13mer, in which average T_ms of TG_1G_2*T are lower those of TG_1*G_2T. These results imply that the B-conformer is a thermodynamic stabilizer in the duplex setting but is a destabilizer during primer elongation. In other words, the S-conformer may be a thermodynamic stabilizer in the TLS process (as presented below).

The impact of lesion heterogeneity on ss/ds duplex formation

We performed SPR experiments to investigate the effect of lesion heterogeneity on strand binding. SPR is an excellent real-time technique that allows monitoring the formation and dissociation of DNA duplexes, DNA–RNA hybridization, and triplex formation. Supplementary Figure S11 shows an overlay of SPR sensorgrams as a function of various complementary strands (9mer to 16mer). Sensorgrams from SPR for each primer elongation and dissociation rate k_d (Table 2) was used to evaluate lesion and sequence effects. This finding is due to the different concentrations of each primer elongation; thus, conducting a global binding analysis is difficult for K_D affinity values (see Results). Instead, we utilize k_d dissociation rates (Table 2) to examine lesion effects.

The binding strength for each primer elongation was increased with increasing extent of Watson–Crick hydrogen bonding. As a result, dissociation rates are steadily reduced with primer elongation. Not surprisingly, the binding strength for the unmodified templates was generally greater than that of the modified ones throughout the DNA replication process. Initially, the effect was relatively minimal at

prelesion sites ($n-1$ and $n-2$ for G_1* and G_2*, respectively) (Figure 6A, 9mer). According to NMR results, the FABP-G_1* lesion with the 9mer (16/9mer) is likely to exist in a B/S-mixture that can be readily accommodated in the flexible $n-2$ environment. As for G_2*, this represents $n-1$, in which the mostly B-conformer can exist at the 16/9mer ss/ds junction. The 16/9mer sequences appeared to be slightly more stabilized than the controls. However, the effect of lesion modification on dissociation was clearly seen around the lesion sites. The 16/10mer corresponds to $n-1$ and n for G_1* and G_2*, respectively. The effect on dissociation was in the order of $n-1$ $G_1* >$ control $> n$ G_2*. The most significant lesion effect of G_1* at $n-1$ is probably due to the presence of S-conformer population that can stabilize the ss/ds replication fork (see above). This situation is illustrated in Figure 6B. For G_2*, the primer has advanced up to the lesion site (n); hence, such lesion-induced stacking cannot attain the same stability. As a result, n G_2* (pink) was less stable than that of the unmodified control (green). The opposite was true for the 16/11mer, in which the binding stability at G_1* was reduced at the lesion site presumably due to the clash with the new primer. This is not the case for the B-conformeric G_2* at $n+1$, which did not change much other than the usual gain of binding energy through elongation. The trend remained unchanged for the rest of the elongation ($n+2$, n+3) and for the full ($n+16$) duplex.

As shown in Supplementary Figure S11, the nature of the lesion (FABP versus FAF) had no noticeable effect on dissociation rates, except that FAF dissociates generally faster than FABP (Table 2). A similar situation was noted in our previous G*CA/G*CT examples (27). Although no polymerase was involved, these SPR results are in line with the NMR/thermodynamic/modeling results on the corresponding duplex given that the greater S-conformation of the G_1* sequence posed a major hindrance toward the binding affinity. In other words, the lesion effect was propagated primarily upstream of the template strand and gradually diminished.

We conducted systematic ^{19}F-NMR, CD, DSC, SPR and molecular modeling studies for the elongation of the TG_1*G_2T and TG_1G_2*T sequences modified by FABP and FAF. Model systems mimic a hypothetical TLS process involving the bulky structurally-related arylamine lesions. DNA adducts in the TG_1*G_2T duplex exhibited greater populations of the S-conformer compared with that in the TG_1G_2*T duplex. Bulky lesions exhibited unique sequence-dependent conformational heterogeneities at various elongation positions, including the replication fork that differentially contributes to template-primer bindings and thermodynamics. We found that the B-conformer is a major thermodynamic stabilizer in the duplex settings and the S-conformer is a destabilizer. The opposite was observed during the primer elongation process involving ss/ds duplex sequences, i.e. the S-conformation promotes lesion stacking with nascent base pairs at the ss/ds replication fork. The results are supported by SPR sensograms, which revealed that the bulky adduct in the stacked conformation increases the binding affinity of the complementary strands in the order of FABP > FAF and $TG_1*G_2T > TG_1G_2*T$. Interestingly, unlike the coplanar FAF, the twisted FABP is a better stacker at the ss/ds junction than in fully paired

duplexes. The extent of primer dissociation rates across the lesion were influenced by the flanking DNA sequences and the stacking ability of the lesion. Future studies include examination of the effects of the conformational heterogeneity in DNA replication and repair.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Skipper,P.L., Kim,M.Y., Sun,H.L., Wogan,G.N. and Tannenbaum,S.R. (2010) Monocyclic aromatic amines as potential human carcinogens: old is new again. *Carcinogenesis*, **31**, 50–58.
2. Turesky,R.J. and Le Marchand,L. (2011) Metabolism and biomarkers of heterocyclic aromatic amines in molecular epidemiology studies: lessons learned from aromatic amines. *Chem. Res. Toxicol.*, **24**, 1169–1214.
3. Hein,D.W. (1988) Acetylator genotype and arylamine-induced carcinogenesis. *Biochim. Biophys. Acta*, **948**, 37–66.
4. Kim,D. and Guengerich,F.P. (2005) Cytochrome P450 activation of arylamines and heterocyclic amines. *Annu. Rev. Pharmacol. Toxicol.*, **45**, 27–49.
5. Zhang,Y. (2013) Understanding the gender disparity in bladder cancer risk: the impact of sex hormones and liver on bladder susceptibility to carcinogens. *J. Environ. Sci. Health C*, **31**, 287–304.
6. Liang,F. and Cho,B.P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry*, **49**, 259–266.
7. Feng,Z., Hu,W., Rom,W.N., Beland,F.A. and Tang,M.-S. (2002) 4-Aminobiphenyl is a major etiological agent of human bladder cancer: evidence from its DNA binding spectrum in human p53 gene. *Carcinogenesis*, **23**, 1721–1727.
8. 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.
9. Belguise-Valladier,P. and Fuchs,R.P. (1991) Strong sequence-dependent polymorphism in adduct-induced DNA structure: analysis of single N-2-acetylaminofluorene residues bound within the NarI mutation hot spot. *Biochemistry*, **30**, 10091–10100.
10. Shibutani,S., Suzuki,N. and Grollman,A.P. (1998) Mutagenic specificity of (acetylamino)fluorene-derived DNA adducts in mammalian cells. *Biochemistry*, **37**, 12034–12041.
11. Jain,V., Hilton,B., Lin,B., Patnaik,S., Liang,F., Darian,E., Zou,Y., MacKerell,A.D. and Cho,B.P. (2013) Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Res.*, **41**, 869–880.
12. Patnaik,S. and Cho,B.P. (2010) Structures of 2-Acetylaminofluorene modified DNA Revisited: Insight into conformational heterogeneity. *Chem. Res. Toxicol.*, **23**, 1650–1652.
13. Meneni,S.R., Shell,S.M., Gao,L., Jurecka,P., Lee,W., Sponer,J., Zou,Y., Chiarelli,M.P. and Cho,B.P. (2007) Spectroscopic and theoretical insights into sequence effects of Aminofluorene-Induced conformational heterogeneity and nucleotide excision repair†,∇. *Biochemistry*, **46**, 11263–11278.
14. Jain,N., Meneni,S., Jain,V. and Cho,B.P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.*, **37**, 1628–1637.
15. Jain,V., Hilton,B., Lin,B., Jain,A., MacKerell,A.D. Jr, Zou,Y. and Cho,B.P. (2013) Structural and thermodynamic insight into Escherichia coli UvrABC-mediated incision of cluster diacetylaminofluorene adducts on the NarI sequence. *Chem. Res. Toxicol.*, **26**, 1251–1262.
16. Patnaik,S. and Cho,B.P. (2010) Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol.*, **23**, 1650–1652.
17. 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.
18. Jain,N., Meneni,S., Jain,V. and Cho,B.P. (2009) Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.*, **37**, 1628–1637.
19. 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.
20. 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.
21. Xu,L., Vaidyanathan,V.G. and Cho,B.P. (2014) Real-time surface plasmon resonance study of biomolecular interactions between polymerase and bulky mutagenic DNA lesions. *Chem. Res. Toxicol.*, **27**, 1796–1807.
22. Vaidyanathan,V.G. and Cho,B.P. (2012) Sequence effects on translesion synthesis of an aminofluorene-DNA adduct: conformational, thermodynamic, and primer extension kinetic studies. *Biochemistry*, **51**, 1983–1995.
23. Broschard,T.H., Koffel-Schwartz,N. and Fuchs,R.P.P. (1999) Sequence-dependent modulation of frameshift mutagenesis at NarI-derived mutation hot spots. *J. Mol. Biol.*, **288**, 191–199.
24. Koffel-Schwartz,N. and Fuchs,R.P. (1995) Sequence determinants for -2 frameshift mutagenesis at NarI-derived hot spots. *J. Mol. Biol.*, **252**, 507–513.
25. Jain,N., Li,Y., Zhang,L., Meneni,S.R. and Cho,B.P. (2007) Probing the sequence effects on NarI-induced -2 frameshift mutagenesis by dynamic 19F NMR, UV, and CD spectroscopy. *Biochemistry*, **46**, 13310–13321.
26. Miller,H. and Grollman,A.P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry*, **36**, 15336–15342.
27. Jain,V., Vaidyanathan,V.G., Patnaik,S., Gopal,S. and Cho,B.P. (2014) Conformational insights into the lesion and sequence effects for arylamine-induced translesion DNA synthesis: 19F NMR, surface plasmon resonance, and primer kinetic studies. *Biochemistry*, **53**, 4059–4071.
28. Sandineni,A., Lin,B., MacKerell,A.D. Jr and Cho,B.P. (2013) Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis. *Chem. Res. Toxicol.*, **26**, 937–951.
29. Xu,L. and Cho,B.P. (2016) Conformational insights into the mechanism of Acetylaminofluorene-dG-Induced frameshift

mutations in the NarI mutational hotspot. *Chem. Res. Toxicol.*, **29**, 213–226.

30. Liang,F. and Cho,B.P. (2011) Conformational and thermodynamic impact of bulky aminofluorene adduction on simulated translesion DNA synthesis. *Chem. Res. Toxicol.*, **24**, 597–605.

31. Chakrabarti,M.C. and Schwarz,F.P. (1999) Thermal stability of PNA/DNA and DNA/DNA duplexes by differential scanning calorimetry. *Nucleic Acids Res.*, **27**, 4801–4806.

32. Jain,V., Hilton,B., Patnaik,S., Zou,Y., Chiarelli,M.P. and Cho,B.P. (2012) Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.*, **40**, 3939–3951.

33. Busto,N., Valladolid,J., Martínez-Alonso,M., Lozano,H.J., Jalón,F.A., Manzano,B.R., Rodríguez,A.M., Carrión,M.C., Biver,T., Leal,J.M. *et al.* (2013) Anticancer activity and DNA binding of a bifunctional Ru(II) Arene Aqua-Complex with the 2,4-Diamino-6-(2-pyridyl)-1,3,5-triazine ligand. *Inorg. Chem.*, **52**, 9962–9974.

34. Vaidyanathan,V.G., Xu,L. and Cho,B.P. (2012) Binary and ternary binding affinities between exonuclease-deficient Klenow fragment (Kf-exo(-)) and various arylamine DNA lesions characterized by surface plasmon resonance. *Chem. Res. Toxicol.*, **25**, 1568–1570.

35. Ma,H., Wang,L., Niesen,D.B., Cai,A., Cho,B.P., Tan,W., Gu,Q., Xu,J. and Seeram,N.P. (2015) Structure activity related, mechanistic, and modeling studies of gallotannins containing a Glucitol-Core and alpha-Glucosidase. *RSC Adv.*, **5**, 107904–107915.

36. Takach,J.C., Mikulecky,P.J. and Feig,A.L. (2004) Salt-dependent heat capacity changes for RNA duplex formation. *J. Am. Chem. Soc.*, **126**, 6530–6531.

37. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G.A. *et al.* (2013) *Gaussian09, Revision D.01 ed*. Gaussian, Inc., Wallingford.

38. Case,D.A., Babin,V., Berryman,J.T., Betz,R.M., Cai,Q., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Gohlke,H. *et al.* (2014) University of California, San Francisco.

39. Li,P. and Merz,K.M. (2014) Taking into account the Ion-Induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.*, **10**, 289–297.

40. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.

41. 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 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

42. Wang,J., Wang,W., Kollman,P.A. and Case,D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **25**, 247–260.

43. Cornell,W.D., Cieplak,P., Bayly,C.I. and Kollman,P.A. (1993) Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.*, **115**, 9620–9631.

44. Liang,F. and Cho,B.P. (2007) Probing the thermodynamics of aminofluorene-induced translesion DNA synthesis by differential scanning calorimetry. *J. Am. Chem. Soc.*, **129**, 12108–12109.