Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor

Vipin Jain1, Benjamin Hilton2, Bin Lin3, Satyakam Patnaik1, Fengting Liang1, Eva Darian3, Yue Zou2, Alexander D. MacKerell Jr3 and Bongsup P. Cho1,*

1Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI 02881, 2Department of Biomedical Sciences, East Tennessee State University, Johnson city, TN 37614 and 3Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA

Received September 7, 2012; Revised October 11, 2012; Accepted October 15, 2012

ABSTRACT

The environmental arylamine mutagens are implicated in the etiology of various sporadic human cancers. Arylamine-modified dG lesions were studied in two fully paired 11-mer duplexes with a -G*CN- sequence context, in which G* is a C8-substituted dG adduct derived from fluorinated analogs of 4-aminobiphenyl (FABP), 2-aminofluorene (FAF) or 2-acetylaminofluorene (FAAF), and N is either dA or dT. The FABP and FAF lesions exist in a simple mixture of ‘stacked’ (S) and ‘B-type’ (B) conformers, whereas the N-acetylated FAAF also samples a ‘wedge’ (W) conformer. FAAF is repaired three to four times more efficiently than FABP and FAF. A simple A- to -T polarity swap in the G*C/A/G*CT transition produced a dramatic increase in syn-conformation and resulted in 2- to 3-fold lower nucleotide excision repair (NER) efficiencies in Escherichia coli. These results indicate that lesion-induced DNA bending/thermodynamic destabilization is an important DNA damage recognition factor, more so than the local S/B-conformational heterogeneity that was observed previously for FAF and FAAF in certain sequence contexts. This work represents a novel 3'-next flanking sequence effect as a unique NER factor for bulky arylamine lesions in E. coli.

INTRODUCTION

Structural and conformational damage in specific areas of the genome can trigger tumorigenesis. For example, disruption of a gene that encodes the tumor suppressor p53 protein has been found in the majority of sporadic human cancer (1). Although human cells are equipped with repair pathways to safeguard the genome from various DNA damage, some lesions may go unrepaired, thereby serving as a faulty template to produce a complex array of mutations and genomic instability, ultimately leading to cancer initiation (2).

Arylamines and heterocyclic amines are well-known environmental mutagens/carcinogens, which have been implicated in the etiology of breast, liver and bladder cancers in humans (3). Metabolic activation of these amines in vivo produces C8-substituted dG as the major DNA adducts (4). For example, the human bladder carcinogen 4-aminobiphenyl produces ABP (Figure 1a). Similarly, AF and AAF are the major DNA adducts derived from 2-amino- and 2-nitrofluorene and 2-acetylaminofluorene (Figure 1a). The ABP and AF adducts in fully paired duplex DNA have been shown to adopt an equilibrium of two prototype conformers: ‘B-type’, in which the carcinogen resides in the major groove of a relatively unperturbed double helical DNA, and ‘stacked (S)’, in which the carcinogen is base displaced and the glycosidic linkage to the modified guanine is syn (Figure 1c) (5,6). The aromatic moieties of ABP are not coplanar as in AF, which results in a much lower S-state population than AF. AF-induced S/B-heterogeneity is dependent on the flanking sequence, which modulates mutational and repair outcomes (6,7). AAF is chemically identical to AF except for a single acetyl group on the central nitrogen (Figure 1a), leading to sampling of an additional W-conformation, in which the fluorene moiety is in the minor groove along with a syn glycosidic linkage (Figure 1c) (7,8). The B and S conformers...
exhibited by AAF are similar to those obtained for ABP and AF.

Nucleotide excision repair (NER) is the major cellular pathway for removing bulky DNA lesions in cells. Accumulated evidence suggests that efficiency of NER is governed by various structural, cellular and biological factors (9–11). Sequence context, in particular, plays an important role in NER of bulky DNA lesions (7,12). The most notable sequence effects were observed in the NarI sequence (5'...CG1G2CG3CC...3'), which is well known for inducing higher frequency of /C02 deletion mutations when adducted by AAF at G3 position despite the similar chemical reactivities of three guanines (13,14). Fuchs and coworkers have shown that AAF in duplex is an excellent substrate for *Escherichia coli* UvrABC and human exonuclease repair systems (15–18). They reported that in *E. coli*, the relative repair efficiencies of AAF at G1, G2 and G3 were in a ratio of 100:18:66, respectively, whereas the human exonuclease exhibited 38:100:68 ratio (17,18). Mu *et al*. (19) have recently carried out a human NER study of these lesions in HeLa cell extracts and found similar sequence-dependent NER efficiencies. Their molecular dynamics (MD) simulation data indicated that the greater NER efficiencies are correlated with base sequence–dependent local un-twisting and minor groove opening together with weaker stacking interactions (19). Recently, we conducted *E. coli* UvrABC NER studies on the NarI sequence duplexes (5'-G1G2CG3CC-3'), in which guanines are modified by either AF or AAF (7). Results showed that the bulky AAF adducts repair in a conformation-specific manner, with the highly S/W-conformeric G3 and G1 duplexes incised considerably more efficiently than the highly B-conformeric G2 duplex (G3/C2$>$G2). Conversely, the repair rate of N-deacetylated AF was 2- to 3-fold lower than AAF, and the order of incision efficiencies was opposite of that observed for the AAF case. We have coined the term ‘N-acetyl factor’ to describe the complexity of NER recognition of AF versus AAF (7).

Here, we describe an unusual 3’-next flanking base effect on the conformational properties and *E. coli* NER efficiencies of three prototype arylamine adducts in the G*C*N sequence context (Figure 1b: G* = ABP, AF, or AAF; N = A or T). Results from spectroscopy (19F NMR and induced circular dichroism [ICD]), thermodynamic quantification (differential scanning calorimetry...
The EMSA assay were processed as reported previously (20,27). N-(FABP), m300 Duplex samples (about 20–30 ODS) were dissolved in C6D6 at E. coli repair recognition of bulky arylamine-DNA adducts in E. coli. This work represents a novel 3'- next flanking sequence effect as a unique NER factor for bulky arylamine lesions in E. coli.

MATERIALS AND METHODS

Adduct synthesis

Modified duplexes were prepared following the published procedures (7,8,20–23). The modified oligos were characterized by electrospray time-of-flight mass spectrometry analysis as reported previously (24). An identical set of unmodified duplexes was also prepared as controls.

Differential scanning calorimetry

DSC measurements were performed using a VP-DSC Micro-calorimeter from Microcal Inc. (Northampton, MA) according to the procedures published previously (22). All sample solutions were 0.12 mM concentration. Tm, was the temperature at half the peak area. ΔG and ΔS values were determined by the procedures of Chakrabarti et al. (25). The uncertainties in the values of Tm, ΔH, ΔG and ΔS represent the random errors inherent in the DSC measurements.

UV-Melting (Cary100 Bio, Beckman) and Circular Dichroism (CD) (J-810, Jasco) experiments were performed using the previously reported procedures (7,20,23,26).

Dynamic 19F NMR

Duplex samples (about 20–30 ODS) were dissolved in 300 μl of pH 7.0 buffer (100 mM NaCl, 10 mM Na3PO4 and 100 μM EDTA in 10% D2O/90% H2O) and filtered into a Shigemi tube using a 0.2-μm membrane filter. All 1H and 19F NMR experiments were conducted using a dedicated 5-mm 19F/1H 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. 19F NMR spectra were acquired in the 1H-decoupled mode and referenced to CFCl3 by assigning external C6F6 for 5 min and

Nucleotide excision assay

DNA substrates of 58 bp containing a single FABP, FAF or FAAP, each adducted at either G*CT or G*CA sequences, were constructed as described previously (28,29). UvrA, UvrB and UvrC proteins were overexpressed in E. coli and then purified as described previously (30). The 5'-terminally labeled DNA substrates were incubated and incised by UvrABC as described previously (28,29). Briefly, the DNA substrates (2 nM) were incubated in the UvrABC reaction buffer (50 mM Tris–HCl, pH 7.5, 50 mM KCl, 10 mM MgCl2, 5 mM DTT) at 37°C in the presence of UvrABC (10 nM UvrA, 250 nM UvrB and 100 nM UvrC). The Uvr proteins were diluted and premixed in Uvr storage buffer before addition to the reaction. Aliquots were collected at 0, 5, 10, 15 and 20 min into the reaction. The reaction was terminated by heating at 95°C for 5 min. The products were denatured by addition of formamide loading buffer and heating to 95°C for 5 min, followed by quick chilling on ice. The incision products were then analyzed by electrophoresis on a 12% polyacrylamide sequencing gel under denaturing conditions with Tris/Borate/EDTA (TBE) buffer. To quantify the incision products, radioactivity was measured using a Fuji FLA-5000 Image Scanner with MultiGauge V3.0 software. The DNA incised (in femtomoles) by UvrABC was calculated based on the total molar amount of DNA used in each reaction and the ratio of the radioactivity of incision products to total radioactivity of DNA. At least three independent experiments were performed for determination of the rates of incision.

MD and PMF calculations

PMF calculations were performed on the GCA and GCT 11-mers initiated from the canonical B form of DNA for ‘anti’ simulations where the glycosidic bond is in the anti form. ‘Syn’ simulations were initiated from models based on an NMR structure (PDB: 1C0Y) in which the glycosidic bond is in the syn form. MD simulations were performed with the programs CHARMM and NAMD, using the CHARMM27 additive nucleic acid force field. Modified G* lesions were created based on the CHARMM General Force Field followed by additional optimization of the dihedral parameters linking the G base to the adduct. Determination of the PMFs followed the protocol of Banavali and MacKerell (31) with details of the simulations included in the supporting information.
Results

Model systems

As model systems, 11-mer DNA duplexes \( \{d(5\prime\text{-CCATCG}^*\text{CNACC}-3\prime),d(5\prime\text{-GGTNGCGATGG}-3\prime)\} \) were prepared, in which \( \text{G}^* \) is FABP, FAF or FAAF and \( \text{N} \) is either \( \text{dA} \) or \( \text{dT} \) (designated as G*CA and G*CT duplexes, respectively) (Figure 1b). The two sequences are chemically isomeric, differing only on the polarity of the 3′-next flanking A:T versus T:A. The utility of fluorine-tagged lesions as effective structure probes has been documented (32). Both the G*CA and G*CT sequences have been used previously for the studies of bulky adducts (22,33).

\[19F \text{ NMR spectroscopy} \]

Figure 2a–c compares the \( \text{19F NMR spectra} \) (−114 to −121 ppm) of modified DNA duplexes at 20°C for the G*CT and G*CA sequence contexts (see Supplementary Figure S1 for full temperature ranges). \( \text{19F signal assignments were made based on the H/D solvent effect, exchange spectroscopy, adduct-induced CD} \) assignments were made based on the H/D solvent effect, exchange spectroscopy, adduct-induced CD (ICD_{290-350\,nm}) and chemical shifts as previously described (6,26,32,34).

\( \text{FABP-duplexes} \)

A clear conformational difference exists between the two isomeric FABP-modified G*CA and G*CT duplexes (Figure 2a). The single signal at −116.9 ppm for FABP-G*CA has been previously assigned to the B-conformer (22). In contrast, FABP-G*CT exhibited two signals at −116.9 (B) and −118.0 (S) ppm in a 40:60% ratio and adopted a two-site exchange (EXSY spectra at 5 and 17°C, inset, Supplementary Figure S1a). A large chemical shift gap (−2 ppm) between the two signals suggests differences of their electronic environments. In addition, the −116.9 ppm signal revealed a large H/D effect (+0.24 ppm) compared to the −118.0 ppm signal (+0.14 ppm) (data not shown) on increasing the D_2O content from 10 to 100%. The results indicate the exposed fluorine atom in the B-conformer, as observed in the MD/PMF simulations (Supplementary Figure S2).

\( \text{FAF-duplexes} \)

Although not as dramatic, a similar sequence effect was observed for FAF. The FAF-G*CA duplex showed signals at −117.4 and −118.6 ppm in a 34:66% ratio (Figure 2b), which have been assigned to B- and S-conformers, respectively (20,22,23,26,35). The S conformer population was increased to 90% in the FAF-G*CT duplex. Consistent with this assignment, the downfield −117.4 ppm signal revealed a larger H/D effect (+0.19 ppm) compared with the −118.8 ppm signal (+0.03 ppm) (data not shown), again consistent with the MD/PMF simulations (Supplementary Figure S2).

\( \text{FAAF-duplexes} \)

At least three major signals are present (Figure 2c) for FAAF in the G*CA and G*CT sequences. These signals are shifted considerably (−2 ppm) to the downfield compared with FAF, a phenomenon associated with the N-acetyl factor. We reported previously S/B/W-conformer assignment of FAAF-modified 12- and 16-mer duplexes in several sequence contexts (i.e. TG*A, CG*C, CG*G and GG*C) (7,8). The results revealed that \( \text{19F signals of B-, S- and W-conformers appear consistently going from downfield to upfield in order of −115.0 to −115.5, −115.5 to −117.0 and −116.5 to −118.0 ppm, respectively.} \)

The major \( \text{19F signals in Figure 2c fit that pattern.} \)

---

**Figure 2.** \( \text{19F NMR (a–c) and CD (d–f) spectra of FABP-, FAF- and FAAF-modified 11-mer duplexes, respectively, in the G*CA and G*CT sequence contexts at 20°C.} \) The B, S and W notation used in the \( \text{19F NMR} \) (a–c) signal assignments represent major groove ‘B’, base displaced stacked ‘S’ and minor groove ‘W’ conformers, respectively (see Figure 1c legend). The CD spectra (d–f) of modified duplexes show the typical B-DNA characteristic (positive and negative ellipticity at 275 and 245 nm, respectively) and different ICD_{290-350\,nm} patterns represent the lesion conformation in the duplex (see CD in Results).
G*CA/G*CT duplexes (5'-CCATCG*C/NAC-3') match well with those observed for 16-mer (5'-CTCTCG G,C-G*CCATCAC-3') (7) and 12-mer (5'-CTCTCG G*CCCTC-3') duplexes (8), both of which have the CG*C sequence context (underlined). In line with this observation, their proton spectra displayed a mixture of broad imino signals arising not only from those involved in Watson–Crick hydrogen bonds (12–14 ppm) but also from the lesion site and its vicinity (11–12 ppm) (Supplementary Figure S3). Although the two sequences are similar in the total syn conformation (S+W) (i.e. 83 vs. 78%, respectively, for G*CT and G*CA), the W-conformer population appears to be significantly greater for G*CT (30%) compared with G*CA (14%) (Supplementary Figure S4).

### Induced circular dichroism

Figure 2d-f show the CD of the modified G*CA and G*CT duplexes. We reported that B- and S-conformers are characterized by positive and negative ICD290–350 nm, respectively (26). Accordingly, the B-conformeric FABP-G*CA displayed a strongly negative ICD290–350 nm, whereas an S-shape curve was observed for the S/B-mixture G*CT duplex (Figure 2d). These results are in good agreement with 19F NMR results (Figure 2a). Unlike FABP, FAF on both sequences exhibited a strong positive ICD290–350 nm with the effect much greater for G*CT (Figure 2e), consistent with the greater S-conformer population determined by 19F NMR (Figure 2b). The ICD of FAAF (Figure 2f), which is confined in the narrow 290–320 nm range, has not been defined as clearly as FAF (8).

In addition, the modified duplexes displayed a significant blue shift relative to their respective control duplexes (Supplementary Figure S3a and b and Table 1). All except for FAF-G*CT exhibited significant blue shifts up to 8 nm. The bulky N-acetylated FAAF exhibited greater shifts (ΔG*G-G = 4–8 nm) than FAF and FABP (ΔG*G-G = 0–4 nm). GCA sequences, which are prone to the B-conformer, displayed greater blue shift (ΔG*G-G*CT = 2–4 nm, Table 1) compared with GCT. It is well known that protein-induced DNA bending exhibits significant CD shift at 275 nm of regular B-type DNA (36–38). For instance, the HMG box protein SOX-5 bends DNA by ≈74° on binding, which resulted in a significant blue CD shift (37). These reports suggest that the blue shifts observed in this study result from the distortion of the DNA backbone, particularly bending.

### Gel mobility assay

Two 19-mer sequences (5'-TTACCATCGC/NACCATT C-3', N = T or A) were used to investigate the impact of the A/T polarity swap at the N position on the gel mobility of the modified duplexes. Initially, the abovementioned 11-mer sequences were used but they denatured in the 15% native polyacrylamide gel at 1800 V (data not shown). Figure 3 compares the electrophoretic mobility of the 19-mer GCA and GCT sequences with and without modifications. Differential mobility between the single strand and double strand (ds) oligonucleotides confirmed the integrity of the duplexes (Figure 3). The modified duplexes exhibited retardation in the mobility in a lesion-dependent manner. In both sequences, major retardation effect was observed for FAAF followed by FABP, whereas no retardation was observed for FAF, results consistent with the CD blue shift data above (Table 1). It should be noted that the magnitude of retardation in mobility observed in this study is significantly lower than what was previously observed in benzo[a]pyrene modified duplexes (39–41). We cannot rule out a possibility of C8-substituted dG (this study) versus N2-substituted dG (benzo[a]pyrene) binding. However, the rationale behind this small difference in mobility could be well due to the significantly longer oligonucleotides (19-mer) used in this study as opposed to the aforementioned benzo[a]pyrene cases (11 and 15-mer). The utilization of longer sequences will reduce the number of adducts per helical turn, which might result into diminishing of the lesion-induced bending effect. Tsao et al. reported similar effects of oligonucleotide length on the electrophoretic mobility of benzo[a]pyrene-modified duplexes (41).

### Thermodynamics

Thermodynamics results from UV-optical melting (Supplementary Table S1) and DSC, which is not
dependent on melting patterns and stoichiometry (22) (Table 2), are comparable. Supplementary Figure S6 shows the DSC thermograms of modified duplexes in the G*CA and G*CT sequences relative to the unmodified controls. These curves were transformed into the corresponding thermodynamic charts (Figure 4a and b), and the results are tabulated in Table 2. According to the NMR results (Figure 2), FABP and FAF display a S/B-equilibrium, whereas FAAF produces a complex S/B/W-equilibrium; thus, they will be compared separately.

FABP|FAF-DNA duplexes

Both FABP and FAF resulted in destabilization (Figure 4 and Table 2). FABP reduced $T_m$ for the G*CA and G*CT duplexes by −10.2 and −11.2°C and $\Delta\Delta G_{37}$ by 2.4 and 2.9 kcal/mol, respectively. The G*CA/G*CT transition produced major effect on $\Delta H$ (2.7 vs. 10.8 kcal/mol) and $\Delta S$ (0.8 vs. 25.6 eu), consistent with a significant increase in the S-conformer population from 0 to ~60% (Figure 2a). The B-conformer FABP-G*CA is expected to lead to small entropy compensation, and consequently, the enthalpy reduction dominated the free-energy destabilization (Figure 4a). As expected, the structural disturbance caused by the S/B mixture FABP-G*CT duplex leads to a considerable reduction of melting enthalpy; however, most of it is compensated by entropy (Figure 4b).

FAF modification resulted in a similar destabilization effect: $\Delta T_m$ by −9.1 and −9.6°C and $\Delta\Delta G_{37}$ by 2.7 and 2.9 kcal/mol, respectively, for G*CA and G*CT. However, compared with FABP, FAF in both sequences yielded significantly larger reduction in enthalpy ($\Delta\Delta H = 13.4$ and 16.4 kcal/mol) and entropy ($\Delta\Delta S = 34.2$ and 43.8 eu) compensation (Figure 4 and Table 2). FAF exhibits more S-conformer than FABP in both sequences because of stronger stacking effect. As a result, sequence dependence on the thermodynamics was not as dramatic as in FABP. It is clear from Figure 4 that FAF (over FABP) and G*CT (over G*CA) combinations produce consistently greater enthalpy/entropy compensation, that is, FAF/G*CT > FABP/G*CT > FABP/G*CA from the highest to the lowest. As expected the N-deacetylated FABP-G*CT and FAF-G*CA exhibited two site exchange (Supplementary Table S2).

FAAF-duplexes

FAAF modification resulted in the most significant reduction of $\Delta T_m$ by −21.8 and −23.3°C and $\Delta\Delta H$ by 39.4 and 41.1 kcal/mol, respectively, for G*CA and G*CT sequence contexts (Figure 4a and Table 2). This is due to the bulky acetyl group in FAAF (Supplementary Figure S2). Like FABP and FAF, however, entropy compensation contributed a stabilization, that is,

<table>
<thead>
<tr>
<th>GCA Sequence</th>
<th>GCT Sequence</th>
</tr>
</thead>
</table>

Table 2. Thermal and thermodynamic parameters derived from DSC curves

<table>
<thead>
<tr>
<th></th>
<th>CCATCG*CAACC</th>
<th>CCATCG*CTACC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−ΔH (kcal/mol)</td>
<td>−ΔS (eu)</td>
</tr>
<tr>
<td>Controla</td>
<td>79.1</td>
<td>214.9</td>
</tr>
<tr>
<td>FABPa</td>
<td>76.4</td>
<td>214.1</td>
</tr>
<tr>
<td>FAFa</td>
<td>65.7</td>
<td>180.7</td>
</tr>
<tr>
<td>FAAFa</td>
<td>39.7</td>
<td>106.1</td>
</tr>
</tbody>
</table>

The average standard deviations for $\Delta G$, $\Delta H$, and $T_m$ are ±0.2, ±2.0, and ±0.4, respectively.

aThe results were calculated from integration of the DSC curve directly.

b$\Delta\Delta H = \Delta H$ (modified duplex) − $\Delta H$ (control duplex).

c$\Delta\Delta S = \Delta S$ (modified duplex) − $\Delta S$ (control duplex).

d$\Delta\Delta G = \Delta G$ (modified duplex) − $\Delta G$ (control duplex).

e$\Delta T_m = T_m$ (modified duplex) − $T_m$ (control duplex).

![Figure 4](https://academic.oup.com/nar/article-abstract/41/2/869/1061270/16)
\[ \Delta G = 5.6 \text{ and } 5.4 \text{ kcal/mol, respectively, for the G*CA and G*CT.} \]

**Molecular dynamics/potential mean force calculations**

To further understand the impact of lesion modification on the G*CA and G*CT duplexes, MD simulations were performed in combination with potential mean force (PMF) calculations. PMF calculations yield the free energy as a function of the extent of flipping of the modified G* base (Supplementary Figure S7). Individual PMFs were determined with G* in the either anti or syn orientations, whereas only the anti orientation was studied for the unmodified duplexes.

Figure 5 shows the free energies from the PMFs. In the unmodified and anti-G* PMFs, there is a deep minimum in 330–360° corresponding to the Watson-Crick (WC)-base paired state (31), which in case of lesions corresponds to the B-state. Deep minima are also present in 330–360° in the syn-G* PMFs, which corresponds to the S-state. The conformer assignments were made based on the published experimental NOE data (Supplementary Figures S8–S11 and Supplementary Tables S3–S6) (42–45). The relative energies of the flipped states are highest in the unmodified duplexes in all cases indicating that the lesions lower the relative energies of the flipped state and/or destabilize the low-energy B- or S-states. Notably, the free-energy surfaces show the relative energies of the flipped states to be lower in the syn PMFs, consistent with the conformational thermodynamic data discussed above. As such, the lowest energies of the flipped state occur with FAAF (Table 2) indicating that the syn FAAF may sample a wider range of conformations as compared with FABP and FAF, consistent with the \(^{19}\)F NMR data (Figure 2). Further validation of the PMFs is the energies of the minima being about 15 kcal/mol, which is in good agreement with the experimental \(\Delta G^\#\) of 14.1 kcal/mol required for a B/S conversion. Representative B/S/W structures from the NOE-based PMFs are shown in Supplementary Figures S12–S14. For all three lesions, the presence of WC base pairing in the B-state and the stacking of the adduct into the duplex in the S-state are evident. In contrast, the structure of the W-state, which is only populated by FAAF, is highly distorted.

In the anti-G* PMFs, the average solvent accessible surface areas (SASAs) of both the lesion and fluorine atom are high in the vicinity of the B state, consistent with the location adduct in the major groove (Supplementary Figure S2). For the syn PMFs, the SASA values are low in the regions corresponding to the S-state, consistent with adducts being stacked inside the helices. However, the SASA values are higher with FAAF compared with FABP and FAF, suggesting an altered environment for FAAF. In addition, the syn PMFs of FAAF exhibits SASA minima in the 60–120° region, which encompasses the W-state. These suggest a fundamental difference in conformational properties of FAF/FABP versus FAAF, consistent with the significant difference in the \(^{19}\)F chemical shifts for the FAAF species and thermodynamic data.

Shown in Supplementary Figure S15 are bending probability distributions for the B-, S- and W-states for the three lesions in both the G*CA and the G*CT contexts. In the B- and S-states, the extent of bending is significantly larger with FAAF (cyan) versus FABP (red) and FAF (blue). These results are consistent with the experimental data obtained from the greater blue shift in CD (Supplementary Figure S5 and Table 1), although the changes in G*CA occur only in the S-state (Supplementary Figure S15). In addition, the simulations indicate the extent of bending for FABP to be similar to that of FAF. The significant increase in bending in FAAF is consistent with the greater destabilization of the duplexes caused by

![Figure 5](https://academic.oup.com/nar/article-abstract/41/2/869/1061270) Free-energy profiles as a function of the pseudo-dihedral angle phi (Supplementary Figure S7) from PMF calculations over the sampling range of 0.5–3 ns for G*CT-FABP, G*CA-FABP, G*CT-FAF, G*CA-FAF, G*CT-FAAF and G*CA-FAAF modified sequences (red and blue) and the unmodified GCT and GCA (black).
the FAAF lesion (Table 2). Concerning the bending, calculation of local helicoidal parameters revealed significant differences in twist and tilt for base pair 8, where the A/T switch occurs (Supplementary Table S7). For example, twist values are systematically larger and tilt values are less negative in GCT versus GCA sequences. These differences suggest that the local structural alteration associated with the A/T switch is being propagated to the overall helix.

**Escherichia coli UvrABC incision**

DNA substrates containing lesions in the defined sequences were subjected to incisions by the *E. coli* UvrABC system in a kinetic assay. These substrates were radioactively labeled at the 5'-end of the adducted strand and the major incision products separated on a urea-PAGE gel running under denaturing conditions (Supplementary Figure S16). The incision occurred at the eighth phosphate bond 5'-G*, which is consistent with the currently accepted mechanisms of UvrABC-based NER (28,29). The substrates were incised at differing efficiencies depending on not only the type of DNA adduct but also the sequence context (Figure 6). Specifically, the G*CA sequences were incised at higher rates by ~2-fold than G*CT, while the order of incision rate of adducts is FAAF > FAF & FABP for both sequences, with FAAF being incised with 2- to 3-fold greater efficiency than the other lesions.

**DISCUSSION**

**Conformational and thermodynamic effects on the G*CA/G*CT transition**

The NMR/ICD results indicate that lesion stacking is affected considerably by a polarity swap at the 3'-next flanking base (GCA → GCT). The effect was most significant for FABP, which produced a dramatic increase in S-conformer (0–60%) (Figure 2a). This is an extraordinary DNA sequence effect. A similar trend was observed for FAF, although the S-conformer was only 24% greater in G*CT than in G*CA (Figure 2b). Unlike FABP and FAF, the impact of the A/T swap on FAAF was minimal; specifically, the syn-glycosidic S- and W-conformers remained relatively unchanged (~78 to ~83%) (Figure 2c). Interestingly, the increase of W-conformer (14–30%) appeared to be compensated by a concomitant decrease of S-conformer (64–53%). These data indicate that the N-acetyl group in FAAF can push the low-energy syn-S/W-equilibrium toward W (see ‘N-acetyl factor’). Overall, these results indicate that the A/T swap has the largest impact on the most stable system, whereas the least stable FAAF lesion is the least impacted.

As expected, all modified duplexes were consistently destabilized compared with the controls (Figure 4 and Table 2): FAAF > FAF ≈ FABP. The G*CA/G*CT transition led to further destabilization, which was associated with increases in lesion stacking (greater S/W) for all three lesions. Obviously, a higher population of the syn-S/W-conformer states is expected to disrupt the double helical structure, which would significantly reduce the enthalpy, accompanied by a compensatory increase in entropy (22).

**Lesion-induced DNA bending as a major NER recognition factor**

For each lesion, a greater proportion of B-conformer was observed in G*CA (FABP: 100%, FAF: 34% and...
FAAF: 22%) than G*CT (FABP: 40%, FAF: 10% and FAAF: 17%). Moreover, good correlation between the magnitudes of change in conformer populations and incision efficiencies was found among the lesions. In FABP, the A/T polarity switch caused a 60% change in S-conformer proportion and a 3-fold reduction in repair efficiency. The changes were significantly lower in FAAF and FAAF (24 and 5–15%, respectively), as were the repair efficiencies (2.0- and 1.8-fold, respectively). At a glance, the results seem to suggest B-conformers have greater reparability than S-conformers. This feature is in clear contrast to the trend that has been observed previously for AF and AAF in certain sequence contexts, that is, the S-conformer is more repairable than the B-conformer (7,35). This type of conformation-specific repair is not only restricted to arylamines but also applied to other bulky lesions. For instance, Geacintov et al. have reported that the base-displaced cis-N²-dG adducts of benzo[a]pyrene are incised more efficiently than the minor groove-orientated trans-N²-dG adducts (33).

However, the E. coli repair results in this study seem to match well with events of adduct-induced DNA bending/distortion, as evidenced by blue shifts in CD (Table 1) and retardation of mobility in electrophoretic mobility shift assay (Figure 3). The slowed mobility indicates flexibility at the lesion site as observed by Tsuo et al. for (+)-trans-anti-[BP]-N²-dG lesion in the TG*T sequence context with concomitant thermal destabilization (41,48). Similarly, the bulky N-acetyl FAAF exhibited significantly slower electrophoretic mobility compared with FAF and FABP within the same sequence context. In case of sequence, the G*CA duplex exhibited consistently greater bending than its G*CT counterpart, with the effect being significantly greater for FAAF than FAF and FABP. A similar CD pattern has been reported for AAF-modified NvrI duplexes related to the formation of a B-Z junction (49). Clearly, the A/T swap alters the conformational equilibrium anti (B-) to syn (S- or W-). It should be noted that the G*CA (–TCG*CAA –) sequence contains a stretch of alternating pyrimidine:purine bases, which are predisposed to DNA bending (50–52). In contrast, such a stretch is interrupted in the highly S-conformeric G*CT (–TCG*CTA –) sequence. It is possible that the B-conformer may facilitate DNA bending, due to the exposure of the carcinogen moiety to the major groove’s hydrophilic environment. In both sequences, a major effect was observed with FAAF, followed by FABP and FAF (Figure 3). However, MD/PMF simulations indicate that the major changes in G*CA occur in the S-state. Also, unlike the CD data, there were no significant differences in electrophoretic mobility between the two sequences (Figure 3). The reason for the inconsistency in the mobility, CD and MD data demonstrating the sequence effect is not apparent, but the greater bending and flexibility of FAAF over FABP or FAF is in good agreement with the observed repair efficiencies (FAAF >> FAF ≈ FABP; Figure 6).

The repair results in Figure 6 along with previously reported work on polycyclic aromatic hydrocarbons (41,53) and arylamines (7) indicate that lesion-induced destabilization of DNA is a major determining factor for repair. However, these lesions were consistently repaired two to three times more efficiently in G*CA than in G*CT, which was not consistent with relative thermodynamic stabilities observed for each. The inconsistency is likely due to the second step of damage recognition (54) that becomes much more significant for FAAF versus FABP and FAAF within a given sequence. Unlike the initial step of damage recognition by UvrA2, which depends on DNA conformation and sequence, the second step of recognition is well known to be characterized by the direct interaction of UvrB with adduct itself on DNA strand opening (47,54–56). In other words, the structure and chemistry of the lesions matter more with UvrB than UvrA2. Recently, Liu et al. reported the NER incision efficiencies of the bulky benzo[a]pyrene and equine estrogen substrates using human HeLa cell extracts and bacterial UvrABC proteins (53). They demonstrated that despite having differences in the prokaryotic and eukaryotic NER proteins, XPC-RAD23B and UvrB, respectively, they exhibit common feature of β-hairpin intrusion for damage recognition. In addition, it was found that local thermodynamic destabilization near the lesion site assists the insertion of β-hairpin, thus recognition.

Clearly, this study shows that the thermodynamic destabilization of the DNA duplex along with lesion flexibility promotes strand opening and thus the second step of damage recognition. The presence of the N-acetyl group (see below) may make FAAF more efficiently recognized than FAF and FABP at the second step due to its flexible nature and greater destabilization of the DNA double helix. As for the G*CA/G*CT transition, the initial recognition step conducted by UvrA2 should be a major determinant factor as the same efficiency of recognition at the second step is expected for the same type of lesion. Thus, bending appears to be an important factor for the DNA damage recognition. Indeed, a recent crystal study by Jaciuk et al. (10) found that in the active site of UvrA, the fluorescein-modified duplexes were bent by ~15° and the structure was related to the kinked structure of psoralen and PAH adducts according to NMR (57). They concluded that the UvrA2 protein does not have direct chemical contacts with a lesion per se, but indirectly senses the overall helical distortion (unwinding and bending) (10). Because energy is required for the bending, formation of the pre-bent DNA induced by bulky lesions would likely enhance the UvrA2 binding and thus damage recognition.

Cai et al. have reported a similar repair trend for the 5’-CACACCG*ACAC-3’ sequence versus 5’-CCATC CG*CTACC-3’, in which G* is the major mutagenic lesion derived from the environmental carcinogen benzo[a]pyrene, 10S (+)-trans-anti-B[a]P-N²-dG (39). A greater repair (1.6-fold) of the CG*CA duplex over the CG*CT counterpart was attributed to its higher bending of the distant 5'-end sequences, as evidenced by gel experiments and MD simulations; these findings are consistent with the bending argument made in this study. Although the sequence contexts (underlined above) near the lesion site, including the 3'-next flanking base, are identical to those used in this study, they did not
The A to T polarity swap in the arylamine-modified DNA duplex. A similar observation regarding the differences in the orientation of AF and AAF was reported by Mu et al. (19) who have conducted a NER study of these lesions in human HeLa cells. MD simulations in that work indicate that the greater repair susceptibility of AAF stems from the increased bending/thermodynamic destabilization as well as the steric constraint in the duplex (‘N-acetyl factor’) (7). A number of different damage recognition parameters have been implicated in the molecular mechanisms of NER (9,55,58). However, it is known that thermal/thermodynamic destabilization and DNA distortion/bending are important factors for damage recognition by repair proteins (9,39,53). The results of this study show that lesion-induced DNA bending/thermodynamic destabilization is a more important NER factor than the usual S/B conformational heterogeneity, as has been observed previously for AF and AAF in certain sequence contexts (7,35). This work represents a novel 3’-next flanking sequence effect as a unique NER factor for bulky arylamine lesions in E. coli. Taken together, the results of this study demonstrate the complexity in DNA recognition factors for repair of bulky arylamine lesions in E. coli.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Tables 1–7, Supplementary Figures 1–16, Supplementary Method and Supplementary References [59–79].

FUNDING
National Institutes of Health (NIH) [CA098296 to B.C., GM051501 to A.D.M. and CA86927 to Y.Z.]; the National Science Foundation/RI-EPSCoR [0554548] and the RI-INBRE core facility supported by National Institutes of Health [P20 RR016457]. Funding for open access charge: NIH/NCI.

Conflict of interest statement. None declared.

REFERENCES


Syn-[AF]-C8-dG adduct opposite dC in a DNA duplex. Biochemistry, 37, 81–94.


