

Published in final edited form as:

Chem Res Toxicol. 2013 August 19; 26(8): 1251–1262. doi:10.1021/tx400186v.

Structural and thermodynamic insight into *E. coli* UvrABC mediated incision of cluster di-acetylaminofluorene adducts on the *NarI* sequence

Vipin Jain^{1,‡}, Benjamin Hilton^{2,‡}, Bin Lin^{3,‡}, Anshu Jain¹, Alexander D. MacKerell Jr³, Yue Zou², and Bongsup P. Cho^{1,*}

¹Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI 02881

²Department of Biomedical Sciences, East Tennessee State University, Johnson City, TN 37614

³Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201

Abstract

Cluster DNA damage refers to two or more lesions in a single turn of the DNA helix. Such clustering may occur with bulky DNA lesions, which may be responsible for their sequence dependent repair and mutational outcomes. Here we prepared three 16-mer cluster duplexes in which two fluoroacetylaminofluorene adducts (dG-FAAF) are separated by none, one and two nucleotides in the *E. coli* *NarI* mutational hot spot (5'-CTCTCG₁G₂CG₃CCATCAC-3'): i.e. 5'--CG₁*G₂*CG₃CC--3', 5'--CG₁G₂*CG₃*CC--3', and 5'--CG₁*G₂CG₃*CC--3' [G*=dG-FAAF], respectively. We conducted spectroscopic, thermodynamic, and molecular dynamics studies of these di-FAAF duplexes and the results were compared with those of the corresponding mono-FAAF adducts in the same *NarI* sequence (*Nucleic Acids Res.* 2012, 3939–3951). Our nucleotide excision repair results showed greater reparability of the di-adducts in comparison to the corresponding mono-adducts. Moreover, we observed dramatic flanking base sequence effects on their repair efficiency in the order of *NarI*-G₂G₃ > -G₁G₃ > -G₁G₂. The NMR/CD/UV-melting and MD-simulation results revealed that in contrast to the mono-adducts, di-adducts produced synergistic effect on duplex destabilization. In addition, dG-FAAF at G₂G₃ and G₁G₃ destack the neighboring bases with greater destabilization occurring with the former. Overall, the results indicate the importance of base stacking and related thermal/thermodynamic destabilization in the repair of bulky cluster arylamine DNA adducts.

INTRODUCTION

The human genome is susceptible to chemical assault from a range of internal and external sources including UV radiation and carcinogens. Adduct formation occurs throughout the genome: some sites showed higher adduct formation, yet little or no mutation induction and

*To whom correspondence should be addressed: bcho@uri.edu, Telephone #: 401-874-5024 (Office); 401-874-5766 (FAX).
equally contributed

Supporting Information. The details of preparation, characterization, CD, UV-melting and UvrABC incision experiments of FAAF di-adducts and MD simulation results for torsion angle at the linkage sites. The HPLC chromatogram (Figure S1), MALDI-TOF spectra of Peak 4 (Figure S2), Peak 5 (Figure S3), and Peak 6 (Figure S4), ¹⁹F NMR (Figure S5 and S6), imino proton spectra (Figure S7), RMSD (Figure S8), and total energy from MM/GBSA (Figure S9) of di-adduct duplexes, 12% denaturing PAGE showing incision products (Figure S10). Tabulated values for MM/GBSA energetic analysis (Table S1), clustering analysis (Table S2), percent base pairing (Table S3), base stacking energies (Table S4) and conformer population ratios (Table S5). This material is available free of charge via the Internet at <http://pubs.acs.org>.

vice versa.¹⁻³ The so-called locally multiply damaged sites (LMDS) or clustered lesions are defined as two or more lesions occurring in a short stretch (<24 bp) of DNA.⁴ In general, cluster lesions are considered to be more genotoxic than a single lesion. It also has been hypothesized that cluster lesions are less repairable than individual ones and that mutation of one lesion is synergized by the presence of another lesion in the vicinity.^{5, 6} The repair efficiency of clustered oxidative damage has been shown to depend on the nature of inter-lesion distance, sequence contexts, and their relative orientation.⁷ For instance, Kalam *et al*⁵ have found a significant fraction of tandem mutations when an uracil was adjacent to 8-oxo-G. The presence of an apurinic site either 5' or 3' to 8-oxo-G increased the frequency of 8-oxo-G to T transversions with the effect much greater on the latter.

Evidence suggests that a similar clustering occurs with bulky DNA lesions and may be responsible for their sequence dependent repair and mutational outcomes. Fuchs and coworkers have shown that the binding spectrum of 2-acetylaminofluorene (AAF) is essentially random and all guanine residues exhibited equal reactivity towards *N*-AcO-AAF.^{8, 9} Other factors such as repair and neighboring sequences may also play a role in the formation of mutation hotspots.¹⁰ For example, 5-methyl cytosine in CpG sites can modulate the distribution of guanine adduct formation by benzo[a]pyrene diolepoxides (BPDE), *N*-methyl-*N*-nitroso urea (MNU), and AAF.¹¹ As in oxidative damage, GC-rich regions and runs of contiguous guanines in the genome have a higher probability of producing bulky cluster adducts.¹¹ Cluster adducts also can be produced through a mixture of either an individual or multiple lesions and such adducts may have an important ramification in damage recognition/verification and repair outcomes.⁷

Arylamines are implicated in the etiology of human breast, liver and bladder cancers.¹² The prototype AAF is no longer threat to human exposure, however, its amino and nitro derivatives are environmental pollutants. These compounds produce two major DNA adducts *in vivo* activation: *N*-(2'-deoxyguanosin-8-yl)-2-aminofluorene (dG-AF) and *N*-(2'-deoxyguanosin-8-yl)- 2-acetylaminofluorene (dG-AAF)(Figure 1a, which have been used extensively as models for bulky lesions).¹³ We have shown that dG-AF exists in an equilibrium of external binding *anti*- B-type and *syn* stacked S-conformers. In the bulky *N*-acetylated AAF, however, the *syn*-adduct also could exist in the minor groove binding "wedge (W)-conformer" (Figure 1c).^{14, 15} The population ratios of these different conformations were found to be strongly dependent on the nature of flanking sequence context around the lesion.¹⁴⁻¹⁹

A number of factors have been implicated to be important for the repair of bulky DNA lesions such as arylamines and polycyclic aromatic hydrocarbons.²⁰⁻²³ They include disruption of Watson-Crick hydrogen bonding, DNA bending, thermodynamic destabilization, local conformational flexibility etc.²¹ Striking sequence effects were observed in the *E. coli* UvrABC mediated NER of fluorinated dG-AF and dG-AAF (dG-FAF and dG-FAAF, respectively, Figure 1a) when present at different guanines of the *NarI* sequence (5'--G₁G₂CG₃CC--3'). dG-FAAF exhibited conformation-specific repair, i.e., the highly S/W-conformeric G₃ and G₁ duplexes incise considerably more efficiently than the highly B-conformeric G₂ duplex (G₃ ~ G₁ > G₂). The nucleotide excision repair (NER) efficiency of the *N*-deacetylated dG-FAF was 2- to 3-fold less, but it too displayed strong sequence dependence.¹⁴ Mu *et al* have confirmed the earlier human NER results of dG-AF and dG-AAF by Fuchs group and provided a molecular dynamics-based structural rationale for their sequence and lesion dependence in repair.²⁴ A similar sequence dependence of repair has been reported in the non *NarI* sequences as well.²⁵⁻²⁷

Little is known about the structural and repair consequence of clustered bulky lesions. As elaborated above, previous structure activity relationship studies have focused on singly

modified sequences, presumably due to methodological ease dealing with a single lesion than multiple adducts.^{28, 29} We hypothesize that cluster arylamine lesions could be formed wherever multiple guanines are present in the genome, resulting in major structural/conformational consequences on repair and mutational outcomes. To that end, we prepared duplex NER substrates, in which the bulky dG-FAAF lesions are located on two (G₁G₂, G₂G₃ and G₁G₃) of the three guanines (G₁, G₂, and G₃) of the *NarI* recognition sequence (Figure 1b). We conducted spectroscopic/melting experiments and MD simulations for conformational and thermodynamic analyses and NER measurements using the *E. coli* UvrABC system. The results present the first structure-repair study on the NER of bulky cluster lesions and its sequence dependence.

MATERIALS AND METHODS

Caution: 2-Aminofluorene derivatives are mutagens and suspected human carcinogens and therefore must be handled with caution.

Crude oligodeoxynucleotides (10 μ mol scale) in desalted form were purchased from Eurofins MWG operon (Huntsville, AL). All HPLC solvents were purchased from Fisher Inc. (Pittsburgh, PA).

Preparation and Characterization of FAAF-modified di-adduct oligodeoxynucleotides

We reported previously the preparation of 16-mer *NarI* oligodeoxynucleotides (5'-CTCTCG₁G₂CG₃CCATCAC-3'), in which each of the three guanine were modified by FAAF.¹⁴ We also have documented that the incorporation of fluorine atom at the longest axis does not affect the overall conformational and thermodynamic profiles of aminofluorene-modified duplexes (e.g., dG-FAF, dG-FAAF).^{17, 30} FAAF-modified 16-mer di-adduct oligodeoxynucleotides were prepared similarly with a longer reaction time (Figure S1, Supporting Information).^{14, 15, 31} Experimental details of FAAF modification and HPLC purification are provided in the Supporting Information.

The three di-FAAF-modified 16-mer sequences were each annealed with the corresponding complementary sequence (5'-GTGATGGCGCCGAGAG-3') to form fully paired duplexes, designated as *NarI*-G₁G₂, *NarI*-G₂G₃, and *NarI*-G₁G₃ depending the location of FAAF modification (Figure 1b). It should be noted that only one 16-mer oligodeoxynucleotide sequence was used with FAAF adducts at different G positions.

MALDI-TOF characterization of di-adduct oligodeoxynucleotides

The modified oligodeoxynucleotides were characterized by following the general procedures on 3'-5' and 5'-3' exonuclease digestion/MALDI-TOF mass spectrometry.^{32, 33} In general, 1 μ L of sample containing approximately 100–250 pmol of a modified oligodeoxynucleotide was used for digestion. In the 5'-3' digestion, 0.01 unit of bovine spleen phosphodiesterase (BSP) was added to a solution containing 1 μ L of oligodeoxynucleotide and 7 μ L of water. Similarly, 0.1 unit of snake venom phosphodiesterase (SVP) was added to a solution of 1 μ L oligodeoxynucleotide, 6 μ L ammonium citrate (100mM, pH 9.4), and 6 μ L of water for 3'-5' digestion. In each case, 1 μ L of reaction mixture was withdrawn at regular time intervals until the digestion rate was significantly reduced, the reaction was quenched by mixing the aliquot with 1 μ L of matrix (3-HPA and DHAC in 1:1 ratio). The samples were then immediately spotted on a MALDI plate and the spectra were collected on the Shimadzu Axima Performance mass spectrometer equipped with a 50 Hz nitrogen laser in the positive ion reflectron mode.

UV melting experiments were performed on Cary100 Bio UV/VIS spectrophotometer equipped with a 6 × 6 multi-cell block and 1.0 cm pathlength. **Circular Dichroism (CD)** measurements were conducted on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. The details of UV-melting and CD experiments are presented in the Supporting Information.

Dynamic ^{19}F NMR

Approximately 10 OD_{260nm} (optical density/mL at 260 nm) of a di-FAAF 16-mer oligodeoxynucleotide was annealed with an equimolar amount of the complementary sequence (5'-GTGATGGCGCCGAGAG-3') to produce a fully-paired duplex (Figure 1b). The duplex samples were then dissolved in 300 μL (~0.2 mM) of a pH 7.0 NMR buffer containing 10% D₂O/90% H₂O, 100 mM NaCl, 10 mM sodium phosphate, and 100 μM EDTA, and filtered into a Shigemi tube through a 0.2 μm membrane filter. All ^1H and ^{19}F NMR results were recorded using a HFC probe on a Varian NMR spectrometer operating at 500.0 and 476.5 MHz, respectively, using the usual acquisition parameters described previously.^{17, 30, 34} Imino proton spectra were obtained at 5°C using a WET-1D sequence and referenced relative to that of DSS. ^{19}F NMR spectra were acquired in the ^1H -decoupled mode and referenced relative to that of CFCl₃ by assigning external C₆F₆ in C₆D₆ at -164.9 ppm. ^{19}F NMR spectra were measured between 5 and 65 °C with an increment of 5–10 °C. Temperatures were maintained by a Varian FTS control system.

Molecular Modeling

The di-adducts were modeled in the BB, BS, SB, and SS conformations, where B and S represent the *anti* and *syn* orientations of the glycosidic bond of the modified G base. The different B and S states will be referred to as glycosidic isomers. Their starting conformations were built in the following way. First, d[CTCTCG₁G₂CG₃CCATCAC] was built using the program 3DNA³⁵ as canonical B DNA for the unmodified duplex, where the glycosidic torsion angles are in the *anti* conformation for G₁, G₂ and G₃. Coordinates were then read into CHARMM³⁶ and patched to attach the FAAF adduct to the C8 atoms in the appropriate guanine bases. This yielded the BB conformations of *NarI*-G₁G₃, *NarI*-G₂G₃, and *NarI*-G₁G₂. The S conformations were prepared by rotating the glycosidic torsion angle by 180° from their respective B-type models. This resulted in a total of 12 starting conformations of FAAF-modified di-adducts, plus an unmodified control DNA duplex. Each system was solvated in a cubic box with TIP3P water³⁷ with the distance from DNA duplex to the edge of the box set to 10 Å. 30 sodium ions were added to neutralize each system. All the models were subjected to extensive energy minimization and equilibration using GROMACS³⁸ 4.5.5 with the latest CHARMM36 nucleic acid force field.³⁹ Particle Mesh Ewald⁴⁰ method was used to treat long-range electrostatics using a real-space cutoff of 10 Å, forth-order spline and an approximately 1 Å grid spacing. Lennard-Jones interactions were switched off between 8 Å and 10 Å. Temperature was controlled at 300 K using velocity rescaling^{41, 42} with a stochastic term of 0.1 ps and the pressure was maintained at 1 atm with the Parrinello-Rahman⁴³ barostat with a time constant of 2 ps. The LINCS algorithm^{44, 45} was employed to constrain all covalent bonds involving hydrogens and the equations of motion were integrated with the Leapfrog algorithm with a time step of 2 fs. Each system was minimized with the Steepest Descent method for 1000 steps then equilibrated by a 100 ps NVT simulation, followed by the 100 ns production runs with the Watson-Crick (WC) N1–N3 hydrogen bond restrained between 2.5 and 3.5 Å for the terminal base pairs. Coordinates were saved every 20 ps. The first 20 ns of the trajectories were discarded as equilibration as judged from the RMS differences with respect to the starting structures. The remaining 80 ns were used for analyses. Curves+⁴⁶ and CHARMM were used to perform bending, structure and interaction energy analyses. Clustering analysis was done using the single linkage algorithm.⁴⁷ VMD⁴⁸ and PyMOL⁴⁹ were used for

visualizing trajectories and preparing structural figures. Averages and standard errors were estimated by using 8 block averages of 10 ns length.

Substrate Construction and UvrABC Protein Purification

DNA substrates of 55 bp containing two FAAF adducts simultaneously on two of the three guanine residues were constructed as previously described^{50, 51} and the details are presented in the Supporting Information. UvrA, UvrB, and UvrC proteins were over expressed in *E. coli* and then purified as previously described.⁵² The estimated purity of the three proteins was greater than 95%. A Bio-Rad Protein Assay was used to determine the protein concentration with BSA as the standard based on the manufacturer-recommended procedures.

Nucleotide Excision Assay and Quantification of Incision Products

The 5'-terminally labeled DNA substrates were incised by UvrABC as previously described.^{50, 51} The details of incision assay and quantification are provided in the Supporting Information.

RESULTS

Model Sequences

We previously reported that a simple mixing of the 16-mer *NarI* oligodeoxynucleotide (5'-CTCTCG₁G₂CG₃CCATCAC-3')¹⁴ with the chemically reactive FAAF at room temperature produced approximately equal mixture of mono-, di- and tri-adducts. A longer reaction time increased production of di- and tri-adducts, which were well-separated on a RP-HPLC as shown in Figure S1 of Supporting Information. We grouped them into three peaks, 1–3 (28–35 min), 4–6 (42–60 min), and 7 (84 min) as mono-, di-, tri-adducts on the basis of the relative absorption intensity (1:2:3, respectively) of adduct induced shoulder in the 290–320 nm range.¹⁴ Moreover, we assigned peaks 1, 2 and 3 as G₁, G₃ and G₂, respectively, using the exonuclease-LC/MS methods.¹⁴ In the present study, we have characterized peaks 4, 5 and 6 similarly using MALDI-TOF as described below. The di-adducts were designated as *NarI*-G₁G₂, *NarI*-G₂G₃, or *NarI*-G₁G₃, in which the numbers signify the position of FAAF-modified guanines (Figure 1b).

MALDI-TOF Characterization

The molecular weights of di-FAAF modified 16-mer were measured by MALDI-TOF prior to sequence verification by exonuclease digestion. Figure 2 (see Supporting Information Figure S2 for complete time ranges) represents the 3'→5' (Figure 2a) and 5'→3' (Figure 2b) exonuclease digestion fragments for Peak 4 at different time intervals. In general, the 3'→5' digestion was much faster than 5'→3'. The ions observed at *m/z* 5257.2 at 30 sec (Figure 2a) and *m/z* 5257.5 at 5 mins (Figure 2b) represent (M+H)⁺ ion of the doubly modified *NarI*-16mer oligodeoxynucleotide. Increase in incubation time leads to the digestion of subsequent unmodified bases. However, 3' and 5' digestions were significantly slowed down at *m/z* 3170.2 and 3781.0, respectively. The ion observed at *m/z* 3170.2 at 15 min of 3'-digestion represents (M+H)⁺ ion formed from the doubly modified 5'-CTCTCG₁G₂CG₃-3' fragment (see Figure 2 inset for theoretical MW values). These results indicate a FAAF modification at the G₃ position of *NarI* oligodeoxynucleotide. In 5'-digestion, the ion observed at *m/z* 3781.0 at 80 mins (Figure 2b) represents (M+H)⁺ ion formed from the doubly modified 5'-G₁G₂CG₃CCATCAC-3' fragment, suggesting another FAAF at the G₁ position. The ion at *m/z* 4070.2 of 5'-digestion (Figure 2b) indicates that the digestion slowed down one base prior to the modification. Taken all together, peak 4 was assigned as *NarI*-G₁G₃ (5'-CTCTCG₁[FAAF]G₂CG₃[FAAF]CCATCAC-3'). Similarly, Peak 5 and 6

were characterized as *NarI*-G₂G₃ and *NarI*-G₁G₂ and the results are shown in Supporting Information Figures S3 and S4, respectively. The HPLC elution pattern and mass characterization of the FAAF di-adduct 16-mer oligodeoxynucleotides in the present study is similar to that of the FAF di-adduct 12-mer *NarI* sequence (5'-CTCG₁G₂CG₃CCATC-3') reported by Gao *et al.*³²

Circular Dichroism

Figure 3a shows an overlay of the CD spectra for the three FAAF di-adduct *NarI*-G₁G₃, *NarI*-G₂G₃ and *NarI*-G₁G₂ duplexes relative to the unmodified control (brown) in ~35 μ M concentration. Unlike FAF and FABP, the induced CD in the 290–320 nm range (ICD_{290–320 nm}) for FAAF-induced S/B/W-conformational heterogeneity have not been established. Nonetheless, comparative pattern analysis provided some valuable conformational insight. Unmodified and FAAF di-adduct duplexes both displayed a positive and negative ellipticity at around 270 and 250 nm, respectively, which is characteristic for a regular B-form DNA double helix. The modified duplexes consistently exhibited a reduced intensity around 250 nm, signifying the alteration of the right handed helicity. All modified duplexes displayed slight blue shifts (2–3 nm) relative to the unmodified control duplexes, indicating a lesion impact on the DNA backbone.²⁵ The effect was most prominent in *NarI*-G₁G₃. Also, the modified duplexes exhibited a change in the positive ellipticity at around 270 nm, indicating alterations of base stacking of a B-form DNA. While the *NarI*-G₁G₂ exhibited a slight increase, the *NarI*-G₂G₃ and *NarI*-G₁G₃ duplexes actually displayed a decrease in intensity with the impact greater in the former (Figure 3a).

In addition, sequence-dependent ICD_{290–320 nm} patterns were seen in all three di-adduct duplexes. Previously, we reported that the FAAF-modified *NarI* mono-adduct duplexes displayed negative dip, with the *NarI*-G₂ duplex exhibited a greater dip than G₁ or G₃ duplexes.¹⁴ Similar ICD_{290–320 nm} was observed in the di-adduct duplexes. The *NarI*-G₁G₃ and *NarI*-G₂G₃ duplexes showed a bigger dip than the *NarI*-G₁G₂ duplex.

UV Melting Experiments

Figure 3b shows the UV melting profiles of the three di-FAAF duplexes in comparison to the unmodified control at 4.8 μ M. All duplexes exhibited typical monophasic, sigmoidal, helix to coil transitions with an excellent linear correlation ($R_2 > 0.9$) between T_m^{-1} and $\ln C_t$. It is clear that modified duplexes were thermally and thermodynamically destabilized relative to the controls. Table 1 summarizes the thermal and thermodynamic parameters calculated from the UV-melting experiments. The results were converted into the thermodynamic chart shown in Figure 4. The *NarI*-G₂G₃ duplex was most destabilized with $\Delta\Delta G_{37^\circ\text{C}} = 7.8$ kcal/mol, $\Delta\Delta H = 23.3$ kcal/mol and $\Delta T_m = -17.9$ °C. The ΔT_m value of *NarI*-G₁G₃ was -14.1 °C, approximately 4 °C lower than *NarI*-G₂G₃, but had a slightly higher $\Delta\Delta H = 24.1$ kcal/mol. However, the large entropy ($\Delta\Delta S = 56.0$ eu) compensated for the enthalpy (Figure 4, Table 1), resulting in the overall free energy destabilization of $\Delta\Delta G_{37^\circ\text{C}} = 6.8$ kcal/mol. In comparison, the stability of the *NarI*-G₁G₂ was least affected ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol, $\Delta T_m = -10.0$ °C). The weak entropy compensation ($\Delta\Delta S = 6.0$ eu) for the substantial enthalpy reduction ($\Delta\Delta H = 6.2$ kcal/mol) resulted into the free energy destabilization ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol)(Figure 4, Table 1). Overall, the magnitude of thermal (ΔT_m) and thermodynamic ($\Delta\Delta G$) destabilization was in the order of *NarI*-G₂G₃ > *NarI*-G₁G₃ > *NarI*-G₁G₂.

Dynamic ¹⁹F NMR Spectroscopy

Figure 5 shows the ¹⁹F NMR spectra of di-adduct duplexes as a function of temperature (5–65 °C)(see Supporting Information Figure S5 for complete temperature ranges). As expected, the di-adduct duplexes displayed an array of ¹⁹F signal in the -114 to -118 ppm

range. This is the same chemical shift range in which mono-FAAF adducts display their B- (−115.0 to −115.5 ppm), S- (−115.5 to −117.0 ppm), and W-conformers (−117.0 to −118.0 ppm) conformers.^{14, 15} The complexity observed in the ¹⁹F NMR spectra of di- over mono-FAAF adducts was expected owing to the presence of two FAAF lesions, each of which is capable of exhibiting S/B/W heterogeneity. As a result, it was very difficult to assign signals as we have done for the mono-adducts.¹⁴

Nonetheless, all di-adducts exhibited typical dynamic NMR characteristics, *i.e.*, exchange broadening of ¹⁹F signals with increasing temperatures, reaching an eventual coalescence at 60 °C and above, indicating thermal denaturation of a duplex (Figure 5). In addition, overall signal patterns were found to be flanking sequence dependent. Supporting Information Figure S6a compares the ¹⁹F NMR spectra of *NarI*-G₂G₃ with the mono-FAAF counterparts (*NarI*-G₂ and - G₃) at two different temperatures (5 and 20 °C). Similar comparisons were made for *NarI*-G₁G₃ and *NarI*-G₁G₂ (Figure S6b and S6c, Supporting Information). The first-order approximation was that the signal patterns of the di-adduct duplexes would be the sum of their respective mono-adducts, but that turned out not to be the case. This indicates that the two nearby FAAF adducts interact with each other leading to conformational sampling not occurring with the mono-adducts.

It was noted that signals of *NarI*-G₁G₃ and G₂G₃ are broad even at 5°C (Figure 5), indicating some conformeric dynamic exchanges. This observation suggests a conformational flexibility at the lesion site, consistent with their high entropy values (Figure 4, Table 1). In contrast, the ¹⁹F signals of *NarI*-G₁G₂ were comparatively sharper, in line with their low entropy (Table 1). In all three duplexes, increasing temperatures resulted in coalescence of signals into a single signal (~−115 ppm) that represents a free modified single strand. The *NarI*-G₁G₂ duplex showed a greater duplex melting signal at ~65 °C compared to 60 °C for the other two duplexes. These results indicate the greater thermodynamic stability of G₁G₂, consistent with the UV-melting data above. The ¹⁹F conformational complexity was supported by their imino protons spectra (Figure S7, Supporting Information). The *NarI*-G₁G₃ and *NarI*-G₁G₂ duplexes exhibited multiple signals in the non-Watson-Crick base pairing 11–12 ppm region, which is known occur from the imino protons of modified guanines. In *NarI*-G₂G₃, however, the signals were shifted downfield to 12 ppm.

MD Simulations

To provide molecular explanations of the experimental results described above, we constructed all the FAAF-modified DNA di-adducts as well as an unmodified duplex as control and subjected those models to 100 ns MD simulations in explicit solvent with neutralizing counterions. Each di-adduct lesion site was modeled in both B-type and S-type conformers, resulting in 12 possible di-adducts. Detailed analyses were performed on the MD trajectories between 20 and 100 ns during which the simulations were stable as shown from their RMS differences as a function of time (Figure S8, Supporting Information). However, the RMS differences for all the di-adducts are significantly larger than the unmodified control duplex and in many of the systems brief excursions to larger RMS differences occur during the simulations. The larger RMS differences are associated with increased bending in the di-adducts compared to the control duplex and the excursions suggest that a range of conformations are being sampled during the simulations. The sampling of a range of conformations is consistent with the ¹⁹F-NMR analysis discussed above. Please see Supporting Information for torsion angles at the nucleotide-adduct linkage site.

MM/GBSA Analysis—As each di-adducts has four possible glycosidic isomers (BB, BS, SB and SS) we undertook MM/GBSA analysis⁵³ to determine the most highly populated isomer for each di-adducts (Figure S9 and Table S1, Supporting Information). We note that additional terms may be included in the overall free energy estimate⁵⁴, such as configurational entropy, but those terms have been omitted here and we note the magnitude of the standard deviations. We note that given the changes in chemical structure of the duplexes, the energies between the control duplex and different di-adducts cannot be directly compared. For each of the di-adducts one or two of the isomers have the most favorable free energies (Figure S9, Supporting Information). This include BB and BS for G₁G₃, BB for G₂G₃ and BB for G₁G₂, though with the latter the BS and SB conformations are within 20 kcal/mol of the minimum while the separation is 25 kcal/mol or more for the G₁G₃ and G₂G₃ di-adducts. Notably, none of the SS di-adducts was observed to be stable (see below). Based on this analysis we assume that these lower energy isomers contribute to the experimental results and interpret experimental and computational data based on this assumption.

Some mix of B and S isomers is indicated which could contribute to the conformational heterogeneity observed in the ¹⁹F NMR spectra. With G₂G₃ 3 relatively well-defined peaks are observed in the ¹⁹F spectra at 25 °C (Figure 5), consistent with the BB isomer dominating. Broader distributions of conformations are observed in the G₁G₃ and G₁G₂ NMR spectra, consistent with presence of 1 or 3 low energy geometric isomers, respectively, being accessible based on the MM/GBSA calculations. In addition, the wide distributions in the G₁G₃ and G₁G₂ spectra are consistent with predicted sampling W conformations in some of the S states (see below).

Clustering Analysis—Clustering analysis was undertaken to identify representative conformations of the different di-adducts. Clustering was based on the RMS differences of the central 6 nucleotides with the results shown in Supporting Information Table S2. In G₁G₃ the BB and BS states are dominated by one conformation, with significant sampling of a second conformation in both cases. With G₂G₃ a single conformation dominates the low energy BB state. A similar situation occurs with the BB and BS isomers of G₁G₂, though three highly populated conformations occur with SB. Overall these results are consistent with the ¹⁹F-NMR results (Figure 5) with a range of conformations present with G₁G₃ and G₁G₂ while a more limited number of conformations occur with G₂G₃. It should be noted that additional conformations not sampled in the present simulation may contribute to the heterogeneity of the ¹⁹F-NMR spectra.

Representative structures from the most populous clusters are shown in Figure 6. For G₁G₃ there were two highly populated clusters for both BB and BS, with the 4 structures shown. Consistent with the *anti* vs. *syn* linkages, three of the FAAF adducts are exposed to solvent while in the BS 240 structure the *syn* adduct is stacked in the duplex. Comparison of the BB structures shows 955 structures to have the lower adduct partially stacked against a C base, an interaction that would be expected to impact the ¹⁹F chemical shift. The two G₁G₃ BS structures are similar, with the exception of the C base flipping out of the duplex, thereby allowing for increased hydration of the interior of the duplex, a phenomenon that could also contribute to heterogeneity in the ¹⁹F NMR spectra (Figure 5). With G₂G₃, the BB isomer was indicated to be the dominate species and a single conformation was sampled the majority of the duration of the simulation. The structure in Figure 6 shows the Watson-Crick (WC) pairing to be well maintained with the exception of the base pair between the two modified Gs. These results are consistent with the G₂G₃ ¹⁹F NMR spectra being the most well defined, although the presence of three peaks in that spectra indicate that additional conformations are present that are not being sampled in the present simulations. The ¹⁹F NMR spectra of G₁G₂ (Figure 5) was quite complicated, which is consistent with the BB,

BS and SB glycosidic isomers indicated to be stable based on the MM/GBSA results (Table S2, Supporting Information). In the BB isomer the WC pairing is maintained, but the duplex is distorted due to the FAAF adducts being adjacent to each other and, consequently interacting directly with each other. The overall duplex is also well maintained on the BS isomer, which appears to be associated with the *syn* FAAF adduct sampling a W-type conformation where it is interacting with the minor groove rather than being stacked in the duplex. Similarly, a W-type conformation of the *syn* FAAF adduct is observed in the 821 and 1981 conformation clusters of the SB isomer, while the FAAF adduct is stacked in the 945 conformation cluster. It should be noted that during the SB simulation, the S-conformer was present at the beginning of the simulation, consistent with the initial model, yielding the 945 cluster, with the *syn* FAAF adduct subsequently converting to the W-conformer which was maintained throughout the remainder of the simulation. This suggests that the W-conformers are representative of the SB isomer of G₁G₂. Overall, the simulations indicate that the different di-adducts sample different ranges of conformations, an observation that is consistent with the ¹⁹F-NMR data. However, especially in the case of G₂G₃, it is clear that the simulations are not sampling all the relevant conformations. This may be an outcome with limitations in the estimated stabilities from the GBSM/MM analysis as well as the limited simulation time in this study.

Base Pairing—Additional analysis focused on possible contributions to the observed stabilities of the di-adducts, including base pairing, base stacking, bending and solvation effects for the low energy glycosidic isomers (Table 2). Table 2 also contains the relative T_m, enthalpy and entropy values from the data in Table 1. For the unmodified control duplex, its WC hydrogen bonding is almost fully maintained, while with WC pairing is lost to varying degrees in the di-adducts. This suggests that a loss of base pairing in the di-adducts contributes to their lowered stability. However, as the least stable di-adducts, G₂G₃, maintains a significant amount of WC interactions versus the more stable G₁G₃ di-adducts, it appears that this term does not dominate the relative stabilities of the di-adducts.

Base Stacking—Base stacking typically dominates the thermodynamic stability of DNA and changes in adduct stability in deletion duplexes have been shown to have significant contributions from perturbations of base stacking.⁵⁵ Accordingly, we examined the relationship of base stacking energies to the measured thermodynamic enthalpies of the di-adducts. Compared to the control, all di-adducts have a significant loss of base stacking energies. The loss is the least in the most stable G₁G₂ di-adducts; however, the least stable adduct, G₂G₃, maintains base stacking to a larger degree than the relatively more stable G₁G₃ adduct. With the relative experimental enthalpies of 24.1, 23.3, and 6.2 kcal/mol for G₁G₃, G₂G₃ and G₁G₂, respectively, it is evident that the calculated results are consistent with the relatively small loss in calculated stacking with G₁G₂ while the losses are significantly more with G₁G₃ and G₂G₃. Thus, base stacking is playing an important role in the relative stabilities of the di-adducts.

We note that in all the SS glycosidic isomers the greatest loss of base pairing and base stacking occurred with one exception (Tables S3 and S4, Supporting Information). This is consistent with the unfavorable MM/GBSA energies of these species. This results suggest that the duplex is not able to accommodate the two *syn* adducts simultaneously stacking, leading to SS isomers not being stable. In the presence of two FAAF modified G bases in the *syn* orientations, both of the lesions would be right next to each other in the same strand, leading to a decrease in stacking energy that is much larger than in the others di-adduct geometric isomers. Thus, we consider it unlikely that this conformer has a noticeable population in the studied di-adducts.

Bending and Solvation—DNA bending analyses of the duplexes are presented in Table 2. Compared to the unmodified duplex, the presence of two FAAF adducts induces significant bending in all the di-adducts, leading to the significant RMS differences noted above. The extent of bending is evident in the *NarI*-G₁G₂-BS (3975) and *NarI*-G₁G₂-SB (1981) structures shown in Figure 6 and is consistent with the decrease in the base pairing and stacking in the adducts. In addition, the increased bending is consistent with the more favorable entropy of the di-adducts, which partially compensates for the enthalpy loss especially with G₁G₃ and G₂G₃.

Solvation can also impact both enthalpic and entropic aspects of duplex stability. To analyze the possible contribution of solvation effects on the bulge regions we computed the solvent accessible surface area (SASA) and the hydration energies from the simulations (Table 2). The increase on the SASA values in the di-adducts over the control duplex is significant indicating that solvation effects are making a large contribution to destabilization of the duplexes in the di-adducts. The increased SASA is accompanied by a more favorable hydration energy in the di-adducts, which would lead to a more favorable enthalpy, in contrast to the experimental data. This suggests that the loss of base pairing and stacking in the di-adducts is dominating the change in enthalpy with solvation offsetting those unfavorable contributions. Interestingly, the MM/GBSA contribution to the free energy of solvation is the most favorable in control (Figure S9 and Table S1, Supporting Information) indicating that the entropy component is playing a significant role in the change of solvation. While qualitative in nature these results suggest a model where the di-adducts lead to a more favorable solvation enthalpy, but this is overcome by a larger unfavorable entropy contribution leading to the solvation contribution contributing to destabilization of the di-adduct duplexes. Concerning the different di-adducts, the solvation properties do not appear to make a significant contribution to the relative stabilities.

UvrABC Incisions of *NarI* FAAF di-adducts

The NER kinetic assay results are shown in Figure 7, in which the 55-mer di-FAAF-adduct duplex substrates were incised by UvrABC nuclease. These substrates were radioactively labeled at the 5'-end of the adducted strand. The major incision products can be seen as ~18 and 19-mer fragments separated on a urea-PAGE gel under denaturing conditions (Figure S10, Supporting Information). The incision occurred at the 8th phosphate bond 5' to the modified nucleotide, which is consistent with the previously reported results of UvrABC incision.^{50, 51} Quantitative analysis of the incision products suggest that the di-adduct substrates were incised 1.5–2 times more efficiently than the previously reported mono-adducts (Figure 7).¹⁴ Moreover, the substrates were incised at different efficiencies, depending upon where the two damage sites were located in the sequence (Figure 7). Specifically, di-FAAF-adducts in *NarI*-G₂G₃ (100%) displayed the fastest rate of incision, followed by *NarI*-G₁G₃ (85%) whereas *NarI*-G₁G₂ (44%) had a much lower rate of incision; G₂G₃ (100%) > G₁G₃ (85%) > G₁G₂ (44%) (Figure 7).

DISCUSSION

A combined experimental and computational study of selected di-FAAF-adducts of the *NarI* sequence was undertaken. Results show the di-adduct duplexes to be destabilized to a greater extent than the mono-adducts. In addition, they sample a wider range of conformations as indicated by ¹⁹F NMR spectra with those conformations maintaining significant B-form DNA characteristics as indicated by CD and MD analyses. Notably, the di-adducts lead to enhance incision rates, as discussed below.

Impact of sequence context on di-adduct duplex structure and thermodynamics

The CD results indicate that the overall DNA structure was influenced by how the two bulky FAAF lesions are arranged in the *NarI* sequence (Figure 3a). In particular, base stacking represented by changes in the positive ellipticity at 270 nm was flanking sequence dependent. While a significant decrease was seen in *NarI*-G₂G₃ and *NarI*-G₁G₃, an increase in intensity was noted for *NarI*-G₁G₂. Our MD simulation data indicate the perturbations of base stacking as a major contributor towards the observed sequence effect. In line with CD, the base stacking was maintained to a larger extent in *NarI*-G₁G₂ whereas severe disruption of base stacking in *NarI*-G₁G₃ and -G₂G₃ resulted into significant loss of duplex stability (Table 2 and Table S4, Supporting Information). However, in contrast to CD, the major impact on base stacking was observed in *NarI*-G₁G₃. This difference could be due to the MD simulations not sampling all the relevant conformations contributing to the experimental results, though complications in the interpretation of the CD spectra cannot be excluded. Similar sequence effect was also observed in the thermal and thermodynamic instability (Figure 4, Table 1): *NarI*-G₂G₃ > *NarI*-G₁G₃ > *NarI*-G₁G₂. This trend is not surprising since base stacking interactions are one of the most important factors that contribute towards the thermal and thermodynamic stability of the DNA duplexes.^{17, 18}

Assuming that each lesion can exist in two major conformations about the glycosidic linkage; *anti* (B) or *syn* (S), there are 4 possible combinations of geometric isomers, BB, BS, SB, and SS. All four were modeled and simulated. RMS differences analysis (Figure S8, Supporting Information) indicates that the di-adduct structures are perturbed to a greater extent than the unmodified control duplex, though all are reasonably stable. The larger RMS differences are associated with increased bending in the di-adducts compared to the non-adducted control due to a range of conformations being sampled during the simulations. Consequently, the ¹⁹F NMR spectra of the di-adduct duplexes produced very complex signal patterns (Figure 5). Interestingly, MM/GBSA analysis indicates the geometric isomers containing B conformers, BB and BS for *NarI*-G₁G₃, BB for -G₂G₃ and BB, BS and SB for -G₁G₂ to be favored, with the SS geometric isomer being unfavorable in all cases. Cluster analysis of the conformations from the simulations of the favored conformations indicates each of the isomers to sample a range of conformations, contributing to the heterogeneity in the ¹⁹F NMR spectra.

We reported previously that the single dG-FAAF adduct at G₃ position of the *NarI* sequence exhibits a preference of S-conformation (61%)(Table S5, Supporting Information).¹⁴ Similar S-conformer dominance was found for dG-AF¹⁷ and related dG-C8 adducts derived from the heterocyclic amines 2-amino-3-methylimidazo(4,5-f)quinolone (IQ)⁵⁶ and 2-amino-1-methyl- 6-phenylimid-azo(4,5-b)pyridine (PHIP).⁵⁷ Interestingly in the present study, a greater thermodynamic instability was found for the G₃ containing *NarI*-G₂G₃ and *NarI*-G₁G₃ duplexes which produce a higher combined S-population (~76 and ~95%, respectively) than *NarI*-G₁G₂ (~49%)(Table S5, Supporting Information). It is possible that the closer proximity of two FAAF in *NarI*-G₂G₃ (e.g. just one base apart) compared to *NarI*-G₁G₃ (two bases) could possibly induces a greater DNA distortion. It should be noted, however, that the most thermodynamically stable *NarI*-G₁G₂ has no base in between the lesions. Thus, it appears that assessing the overall thermodynamic stability of di-adducts is complicated by various conformational factors.

Conformation-specific Nucleotide Excision Repair

The *E. coli* UvrABC system exhibited approximately 1.5–2.0 folds better efficiency in the reparability of the di-adducts (G₂G₃, G₁G₃ and G₁G₂) as compared to mono-adducts (G₁, G₂ and G₃) in the same *NarI* sequence (Figure 7). This is not surprising because of the greater thermal/thermodynamic destabilization of the di-FAAF (Table 1) vs. mono-FAAF

duplexes.¹⁴ It is well known that NER efficiency is highly modulated by the damage recognition (UvrA₂) and verification/recognition (UvrB) steps.⁵⁸ Jaciuk et al have shown that the UvrA₂ does not interact with DNA lesion directly, but senses the structural/conformational perturbations induced by a lack of Watson-Crick hydrogen bonding and thermodynamic destabilization.²¹ Moreover, the thermodynamic destabilization of the duplex assists UvrB in the insertion of a β -hairpin into the duplex which is required for the lesion verification/recognition.²² We reported that the repair of mono dG-FAAF adducts in *NarI* sequence occurs in a conformation-specific manner, i.e. the highly S/W-conformeric G₃ and G₁ duplexes produced greater DNA distortion and thermodynamic destabilization and thus incised more efficiently than the B-type G₂ duplex (G₃~G₁ > G₂).¹⁴ The presence of two dG-FAAF in the same *NarI* sequence is expected to synergize destabilization effect because of the significant loss of base pairing and base destacking as observed in the MD-simulations (Table 2). Indeed, the destabilization effect of di-adducts ($\Delta\Delta G_{37^\circ\text{C}} = 4.3$ to 7.8 kcal/mol, $\Delta T_m = -10.0$ to -17.9 °C) is 1.5–2.0 folds greater than that of the mono-adducts ($\Delta\Delta G_{37^\circ\text{C}} = 3.7$ to 4.7 kcal/mol, $\Delta T_m = -5.3$ to -8.3 °C)(Table 1).¹⁴ These results are in contrast to recent reports on oxidative DNA damages where the reparability is compromised due to clustering, leading to enhanced genotoxicity.⁷ The rationale behind this discrepancy could lie in the complexity of processing of oxidative cluster damages that involve base excision repair, non-homologous end joining and homologous recombination proteins.

In addition, we observed different repair efficiencies depending on the location of di-FAAF adducts (G₂G₃, G₁G₃ and G₁G₂). The *NarI*-G₁G₂ duplex showed considerably lower efficiency than *NarI*-G₂G₃ and -G₁G₃ [G₂G₃ (100%) > G₁G₃ (85%) > G₁G₂ (44%)](Figure 7). These NER results are in good agreement with the extent of the loss of base stacking and thermal and thermodynamic destabilization (G₂G₃ > G₁G₃ > G₁G₂). The *NarI*-G₂G₃ duplex displayed the highest repair rate and also underwent the most severe loss of base stacking (Figure 3a, Table 2) and duplex destabilization (*NarI*-G₂G₃: $\Delta\Delta G_{37^\circ\text{C}} = 7.8$ kcal/mol, $\Delta T_m = -17.9$ °C). By contrast, *NarI*-G₁G₂ duplex showed the weakest reparability, consistent with the minimum base stacking (Figure 3a, Table 2) and duplex destabilization (*NarI*-G₁G₂: $\Delta\Delta G_{37^\circ\text{C}} = 4.3$ kcal/mol, $\Delta T_m = -10$ °C). *NarI*-G₁G₃ showed an intermediary reparability and loss of stacking (Figure 3a, Table 2)/destabilization (*NarI*-G₂G₃: $\Delta\Delta G_{37^\circ\text{C}} = 6.8$ kcal/mol, $\Delta T_m = -14.1$ °C). These results confirm the sequence dependence repair of the di-adduct duplexes on duplex destabilization.

It should be noted that repair of AAF involving position G₃ is consistently greater, which suggested the uniqueness of the G₃ position of the *NarI* sequence in both mono-¹⁴ and di-adduct cases. Interestingly, comparison of the incision products indicated that *NarI*-G₁G₂/*NarI*-G₁G₃ and *NarI*-G₂G₃ are incised at different sites in the sequence (Figure S10b, Supporting Information). Specifically, *NarI*-G₁G₂/*NarI*-G₁G₃ are incised to generate a major product of 18-mer while G₂G₃ is incised to produce a 19-mer. These interesting results suggest that the incision position of the di-adducts depends on the position of 5'-adduct, but independent of that of 3'- adduct.

In summary, the nucleotide excision repair of di-dG-FAAF adducts was found to occur approximately 2-fold more efficiently than the corresponding mono-adducts in the *NarI* mutational hot spot sequence. The greater reparability of the di-adducts was primarily due to their greater thermal and thermodynamic destabilization. In addition, the incision efficiency occurred in a sequence-specific manner i.e. the highly destacked and destabilized *NarI*-G₂G₃ and *NarI*-G₁G₃ are repaired more efficiently than the relatively most stable *NarI*-G₁G₂. These results indicate the importance of carcinogen induced destacking and related thermal/thermodynamic destabilization in the repair of cluster arylamine adducts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

This work was supported by the National Institutes of Health (Grant number# CA098296 to BC, CA86927 to YZ and GM051501 to ADM) and is based upon work conducted at a research facility at the University of Rhode Island supported in part by the National Institutes of Health/RI-INBRE (P20 RR016457) and the National Science Foundation EPSCoR Cooperative Agreement #EPS-1004057.

ABBREVIATIONS

dG-AAF-adduct	<i>N</i> -(2'-deoxyguanosin-8-yl)-2-acetylaminofluorene
dG-FAAF-adduct	<i>N</i> -(2'-deoxyguanosin-8-yl)-7-fluoro-2-acetylaminofluorene
ICD	induced circular dichroism
MALDI-TOF	matrix assisted laser desorption/ionization-time-of-flight
SASA	solvent accessible surface area.

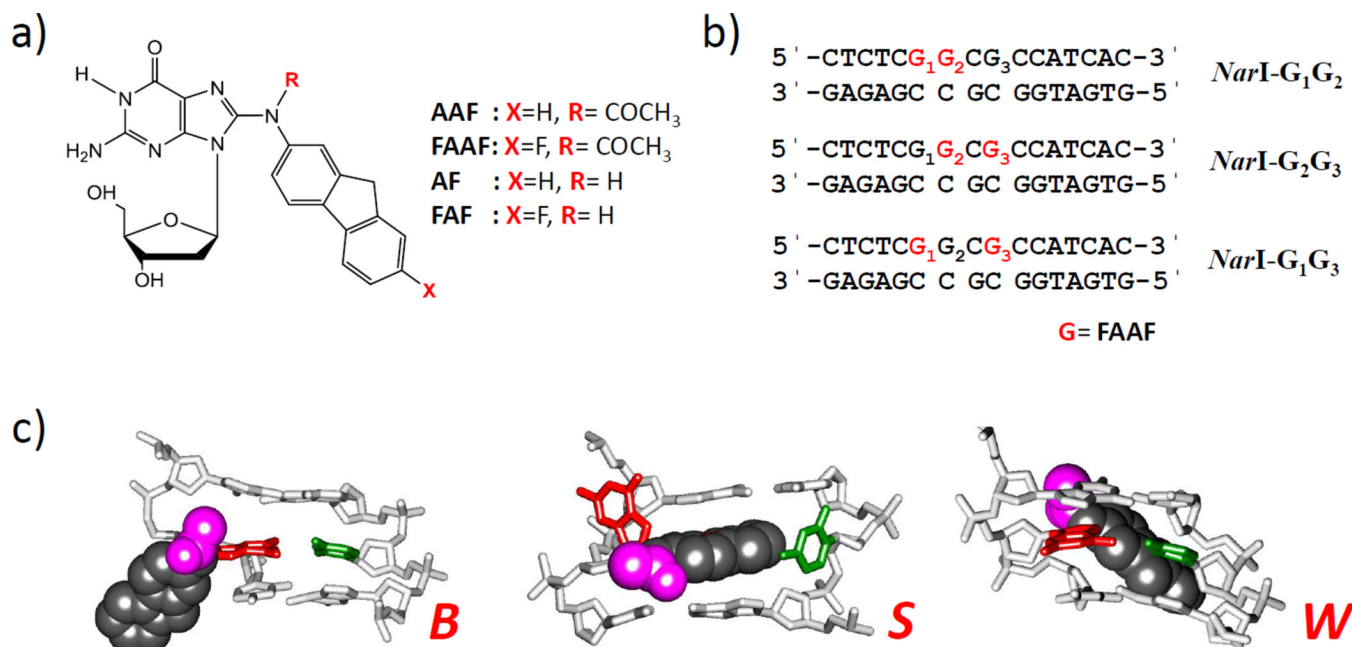
References

1. Boldt J, Mah MC, Wang YC, Smith BA, Beland FA, Maher VM, McCormick JJ. Kinds of mutations found when a shuttle vector containing adducts of 1,6-dinitropyrene replicates in human cells. *Carcinogenesis*. 1991; 12:119–126. [PubMed: 1988171]
2. Burnouf D, Koehl P, Fuchs RP. Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci. USA*. 1989; 86:4147–4151. [PubMed: 2657743]
3. Yang JL, Maher VM, McCormick JJ. Kinds of mutations formed when a shuttle vector containing adducts of (+/-)-7 beta, 8 alpha-dihydroxy-9 alpha, 10 alpha-epoxy-7,8,9, 10-tetrahydrobenzo[a]pyrene replicates in human cells. *Proc. Natl. Acad. Sci. USA*. 1987; 84:3787–3791. [PubMed: 3108878]
4. Ward JF. DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Prog. Nucleic Acid Res. Mol. Biol.* 1988; 35:95–125. [PubMed: 3065826]
5. Kalam MA, Basu AK. Mutagenesis of 8-oxoguanine adjacent to an abasic site in simian kidney cells: tandem mutations and enhancement of G-->T transversions. *Chem. Res. Toxicol.* 2005; 18:1187–1192. [PubMed: 16097791]
6. Shikazono N, Pearson C, O'Neill P, Thacker J. The roles of specific glycosylases in determining the mutagenic consequences of clustered DNA base damage. *Nucleic Acids Res.* 2006; 34:3722–3730. [PubMed: 16893955]
7. Magnander K, Elmroth K. Biological consequences of formation and repair of complex DNA damage. *Cancer Lett.* 2012; 327:90–96. [PubMed: 22353687]
8. Fuchs RP. DNA binding spectrum of the carcinogen N-acetoxy-N-2-acetylaminofluorene significantly differs from the mutation spectrum. *J. Mol. Biol.* 1984; 177:173–180. [PubMed: 6748082]
9. Roy A, Fuchs RP. Mutational spectrum induced in *Saccharomyces cerevisiae* by the carcinogen N-2-acetylaminofluorene. *Mol. Gen. Genet.* 1994; 245:69–77. [PubMed: 7845359]
10. Mah MC, Boldt J, Culp SJ, Maher VM, McCormick JJ. Replication of acetylaminofluorene-adducted plasmids in human cells: spectrum of base substitutions and evidence of excision repair. *Proc. Natl. Acad. Sci. USA*. 1991; 88:10193–10197. [PubMed: 1946440]

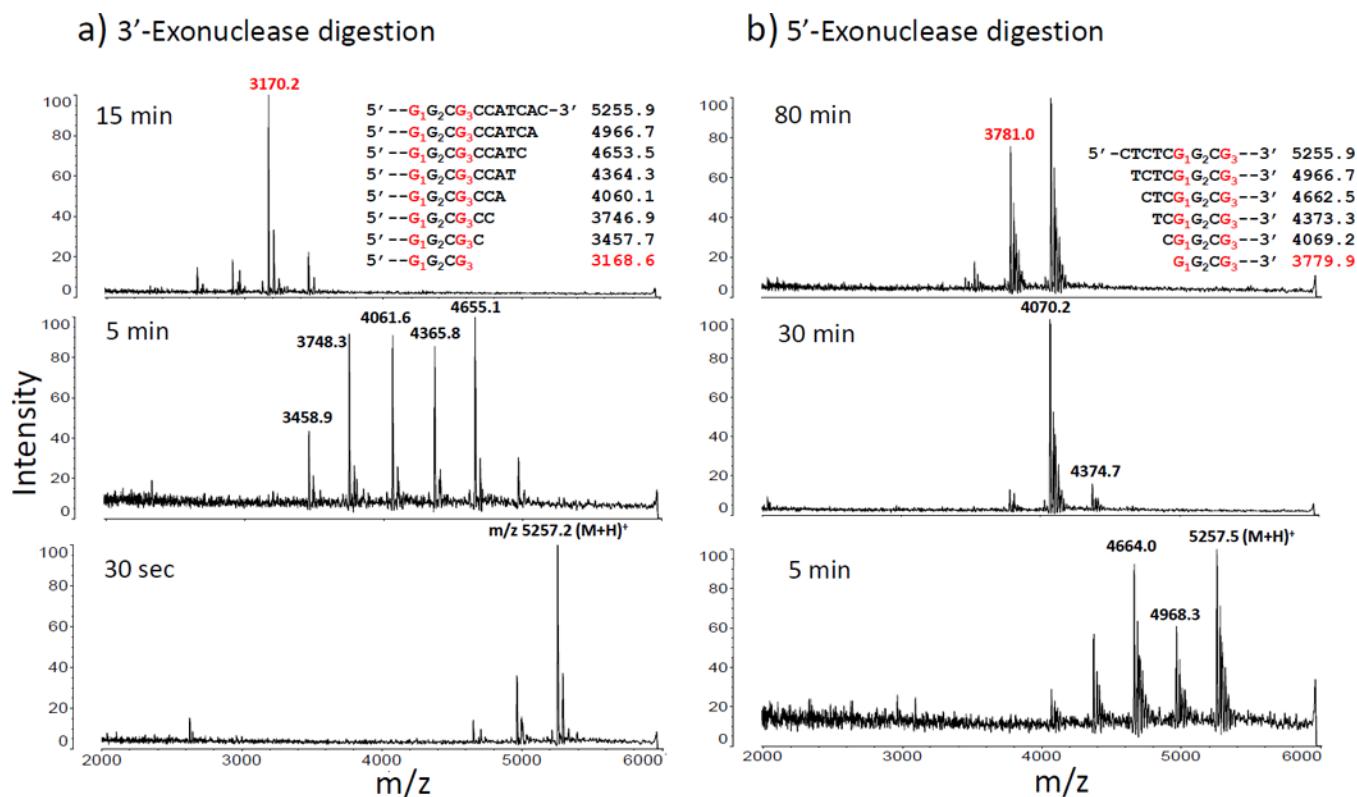
11. Ross MK, Said B, Shank RC. DNA-damaging effects of genotoxins in mixture: modulation of covalent binding to DNA. *Toxicol. Sci.* 2000; 53:224–236. [PubMed: 10696770]
12. Luch A. Nature and nurture - lessons from chemical carcinogenesis. *Nat. Rev. Cancer.* 2005; 5:113–125. [PubMed: 15660110]
13. Heflich RH, Neft RE. Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutation Research/Reviews in Genetic Toxicology.* 1994; 318:73–174.
14. Jain V, Hilton B, Patnaik S, Zou Y, Chiarelli MP, Cho BP. Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the *NarI* sequence. *Nucleic Acids Res.* 2012; 40:3939–3951. [PubMed: 22241773]
15. Patnaik S, Cho BP. Structures of 2-acetylaminofluorene modified DNA revisited: insight into conformational heterogeneity. *Chem. Res. Toxicol.* 2010; 23:1650–1652. [PubMed: 20954689]
16. Cho BP. Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* 2004; 22:57–90. [PubMed: 16291518]
17. Jain N, Li Y, Zhang L, Meneni SR, Cho BP. Probing the sequence effects on *NarI*-induced -2 frameshift mutagenesis by dynamic 19F NMR, UV, CD spectroscopy. *Biochemistry.* 2007; 46:13310–13321. [PubMed: 17960913]
18. Meneni SR, D'Mello R, Norigian G, Baker G, Gao L, Chiarelli MP, Cho BP. Sequence effects of aminofluorene-modified DNA duplexes: thermodynamic and circular dichroism properties. *Nucleic Acids Res.* 2006; 34:755–763. [PubMed: 16449208]
19. Meneni SR, Shell SM, Gao L, Jurecka P, Lee W, Sponer J, Zou Y, Chiarelli MP, Cho BP. Spectroscopic and theoretical insights into sequence effects of aminofluorene-induced conformational heterogeneity and nucleotide excision repair. *Biochemistry.* 2007; 46:11263–11278. [PubMed: 17877372]
20. Geacintov NE, Broyde S, Buterin T, Naegeli H, Wu M, Yan S, Patel D. Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers.* 2002; 65:202–210. [PubMed: 12228925]
21. Jaciuk M, Nowak E, Skowronek K, Tanska A, Nowotny M. Structure of UvrA nucleotide excision repair protein in complex with modified DNA. *Nat. Struct. Mol. Biol.* 2011; 18:191–197. [PubMed: 21240268]
22. Liu Y, Reeves D, Kropachev K, Cai Y, Ding S, Kolbanovskiy M, Kolbanovskiy A, Bolton JL, Broyde S, Van Houten B, Geacintov NE. Probing for DNA damage with beta-hairpins: similarities in incision efficiencies of bulky DNA adducts by prokaryotic and human nucleotide excision repair systems in vitro. *DNA Repair (Amst).* 2011; 10:684–696. [PubMed: 21741328]
23. Min JH, Pavletich NP. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature.* 2007; 449:570–575. [PubMed: 17882165]
24. Mu H, Kropachev K, Wang L, Zhang L, Kolbanovskiy A, Kolbanovskiy MNEG, Broyde S. Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.* 2012; 40:9675–9690. [PubMed: 22904073]
25. Jain V, Hilton B, Lin B, Patnaik S, Liang F, Darian E, Zou Y, Mackerell AD Jr, Cho BP. Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Res.* 2013; 41:869–880. [PubMed: 23180767]
26. Mekhovich O, Tang M, Romano LJ. Rate of incision of N-acetyl-2-aminofluorene and N-2-aminofluorene adducts by UvrABC nuclease is adduct- and sequence-specific: comparison of the rates of UvrABC nuclease incision and protein-DNA complex formation. *Biochemistry.* 1998; 37:571–579. [PubMed: 9425079]
27. Zou Y, Shell SM, Utzat CD, Luo C, Yang Z, Geacintov NE, Basu AK. Effects of DNA adduct structure and sequence context on strand opening of repair intermediates and incision by UvrABC nuclease. *Biochemistry.* 2003; 42:12654–12661. [PubMed: 14580212]
28. Basu AK, Essigmann JM. Site-specifically modified oligodeoxynucleotides as probes for the structural and biological effects of DNA-damaging agents. *Chem. Res. Toxicol.* 1988; 1:1–18. [PubMed: 2979705]

29. Guengerich FP. Interactions of carcinogen-bound DNA with individual DNA polymerases. *Chem. Rev.* 2006; 106:420–452. [PubMed: 16464013]
30. Zhou L, Rajabzadeh M, Traficante DD, Cho BP. Conformational heterogeneity of arylamine-modified DNA: 19F NMR evidence. *J. Am. Chem. Soc.* 1997; 119:5384–5389.
31. Cho BP, Zhou L. Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by 19F NMR spectroscopy. *Biochemistry.* 1999; 38:7572–7583. [PubMed: 10360955]
32. Gao L, Zhang L, Cho BP, Chiarelli MP. Sequence verification of oligonucleotides containing multiple arylamine modifications by enzymatic digestion and liquid chromatography mass spectrometry (LC/MS). *J. Am. Soc. Mass Spectrom.* 2008; 19:1147–1155. [PubMed: 18524623]
33. Zhang LK, Gross ML. Matrix-assisted laser desorption/ionization mass spectrometry methods for oligodeoxynucleotides: improvements in matrix, detection limits, quantification, and sequencing. *J. Am. Soc. Mass Spectrom.* 2000; 11:854–865. [PubMed: 11014447]
34. Jain N, Meneni S, Jain V, Cho BP. Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Res.* 2009; 37:1628–1637. [PubMed: 19151371]
35. Lu XJ, Olson WK. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols.* 2008; 3:1213–1227.
36. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kucsera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
37. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 79:926–935.
38. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 2008; 4:435–447.
39. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, Mackerell AD Jr. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* 2012; 8:348–362. [PubMed: 22368531]
40. Darden T, York D, Pedersen L. Particle Mesh Ewald - An N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993; 98:10089–10092.
41. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 2007; 126
42. Bussi G, Zykova-Timan T, Parrinello M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* 2009; 130
43. Parrinello M, Rahman A. Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters.* 1980; 45:1196–1199.
44. Hess B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 2008; 4:116–122.
45. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997; 18:1463–1472.
46. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ *Nucleic Acids Res.* 2009; 37:5917–5929. [PubMed: 19625494]
47. Sibson R. SLINK: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal.* 1973; 16:30–34.
48. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics Modell.* 1996; 14:33–38.
49. Schrodinger L. The PyMOL Molecular Graphics System. 2010
50. Luo C, Krishnasamy R, Basu AK, Zou Y. Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, N-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease. *Nucleic Acids Res.* 2000; 28:3719–3724. [PubMed: 11000263]

51. Zou Y, Liu TM, Geacintov NE, Van Houten B. Interaction of the UvrABC nuclease system with a DNA duplex containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE. *Biochemistry*. 1995; 34:13582–13593. [PubMed: 7577947]
52. Zou Y, Van Houten B. Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J*. 1999; 18:4889–4901. [PubMed: 10469667]
53. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL 3rd. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem*. 2004; 25:265–284. [PubMed: 14648625]
54. Beveridge DL, Dicapua FM. Free-energy via molecular simulation - applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem*. 1989; 18:431–492. [PubMed: 2660832]
55. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*. 2006; 34:564–574. [PubMed: 16449200]
56. Elmquist CE, Wang F, Stover JS, Stone MP, Rizzo CJ. Conformational differences of the C8-deoxyguanosine adduct of 2-amino-3-methylimidazo[4,5-f]quinoline (IQ) within the *NarI* recognition sequence. *Chem. Res. Toxicol*. 2007; 20:445–454. [PubMed: 17311423]
57. Brown K, Hingerty BE, Guenther EA, Krishnan VV, Broyde S, Turteltaub KW, Cosman M. Solution structure of the 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine adduct in duplex DNA. *Proc. Natl. Acad. Sci. USA*. 2001; 98:8507–8512. [PubMed: 11438709]
58. Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: The UvrABC system. *Chem. Rev*. 2006; 106:233–252. [PubMed: 16464004]

**Figure 1.**

Adduct structures and sequences. (a) Chemical structures of AAF and FAAF adducts; (b) fully paired 16-mer *NarI* duplexes; (c) major groove views of the central trimer segments of the B/S/W-conformer equilibrium of FAAF-modified mono-adduct duplexes. The modified dG and the complementary dC are shown in red and green sticks, respectively, and the aminofluorene moiety is highlighted with shiny grey CPK and the *N*-acetyl with pink CPK. In the B-type conformer, anti-glycosidic FAAF-dG maintains Watson-Crick hydrogen bonds, thereby placing the carcinogen moiety in the major groove. The carcinogens in the S- and W-conformers stack into the helix or wedged into the minor groove, respectively, with the modified dG in the syn conformation.

**Figure 2.**

MALDI-TOF mass spectra of the hydrolysis fragments formed from exonuclease digestion of peak 4 (see Fig. S1) (a) the 3'- exonuclease digestion spectrum at 30 seconds (bottom), 5 (middle) and 15 (top) minutes. (b) the 5'-exonuclease digestion spectrum at 5 (bottom), 30 (middle) and 80 (top) minutes. Insets show the theoretical molecular weights of the fragments that should form after the 3'- and 5'-exonuclease digestion of peak 4 if oligodeoxynucleotide is modified at G₁ and G₃ positions.

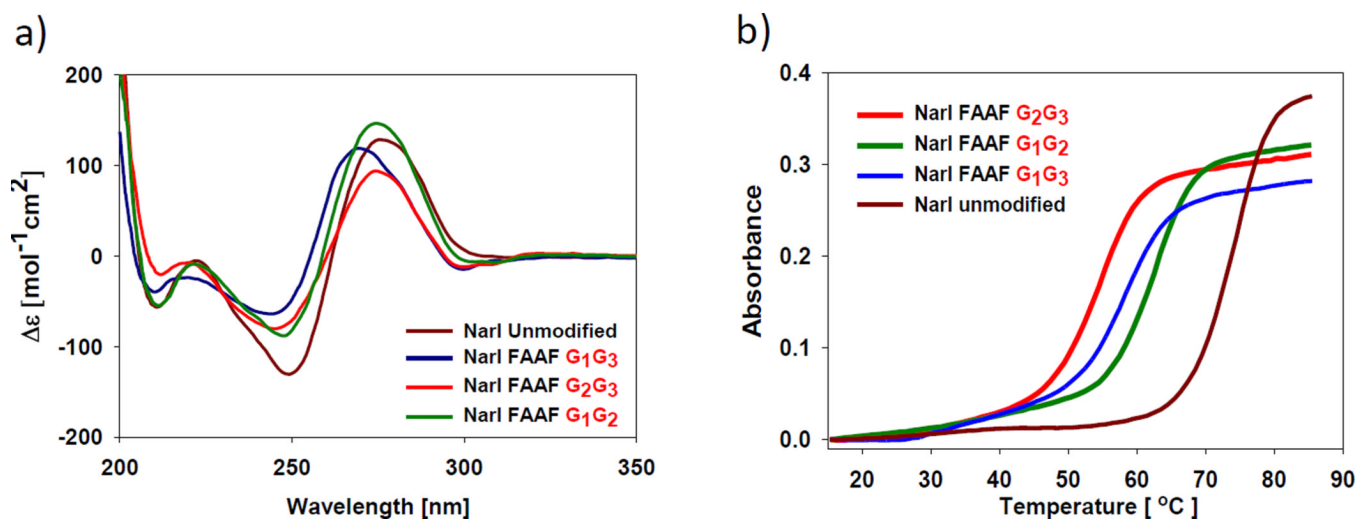


Figure 3. CD and UV-melting profiles. (a) CD spectral overlays recorded at 20 °C and (b) UV-melting curves in 20 mM phosphate buffer containing 0.1 M NaCl at pH 7 of fully paired 16-mer *NarI* di-FAAF-adduct duplexes: G₁G₂ (green), G₁G₃ (blue), and G₂G₃ (red).

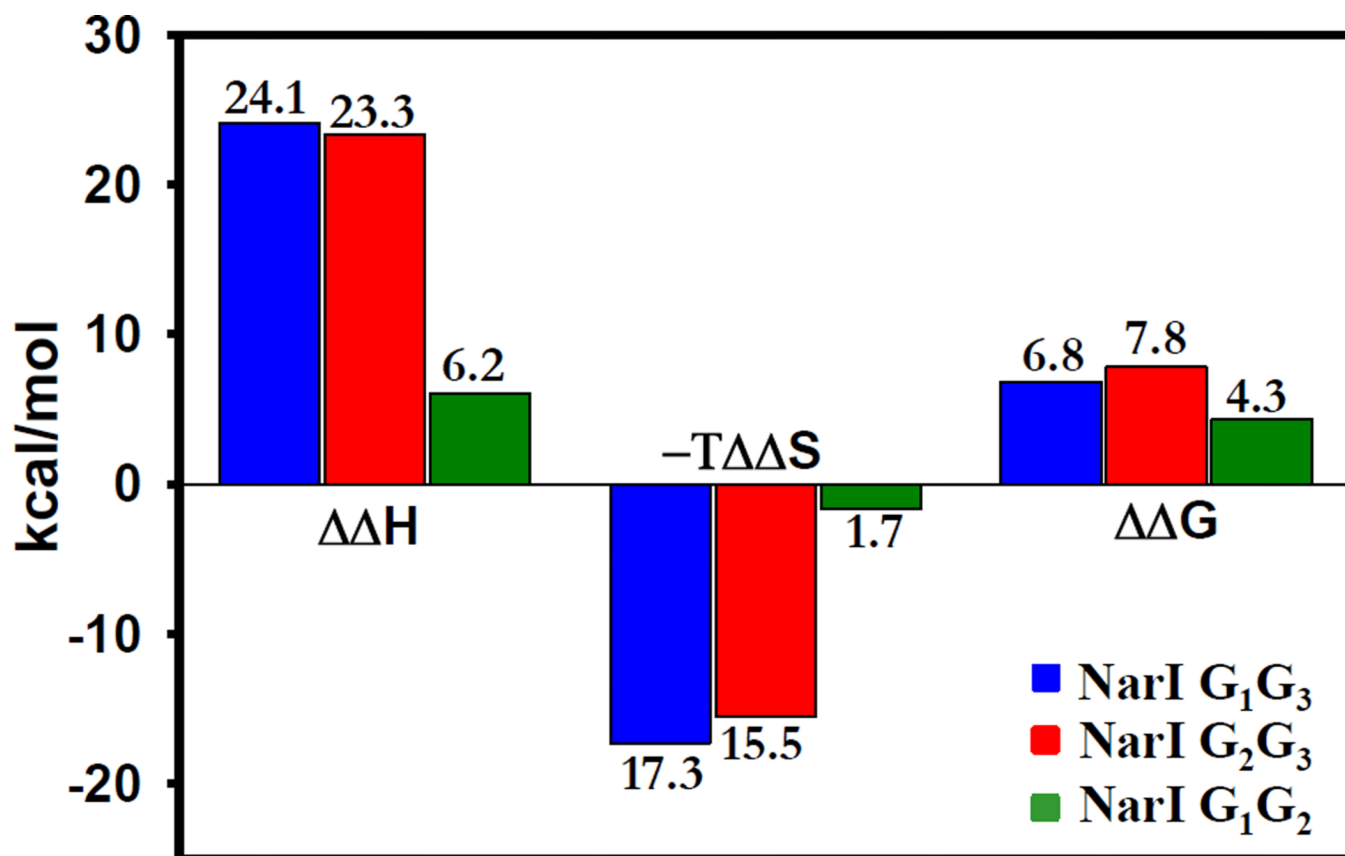


Figure 4.

Comparative thermodynamic parameters histogram of *NarI* di-FAAF adduct duplexes; G₁G₂ (green), G₁G₃ (blue), and G₂G₃ (red). The $\Delta\Delta$ values represent; $\Delta\Delta H = \Delta H$ (modified duplex) - ΔH (control duplex), $T\Delta\Delta S = T\Delta S$ (modified duplex) - $T\Delta S$ (control duplex), $\Delta\Delta G = \Delta G$ (modified duplex) - ΔG (control duplex).

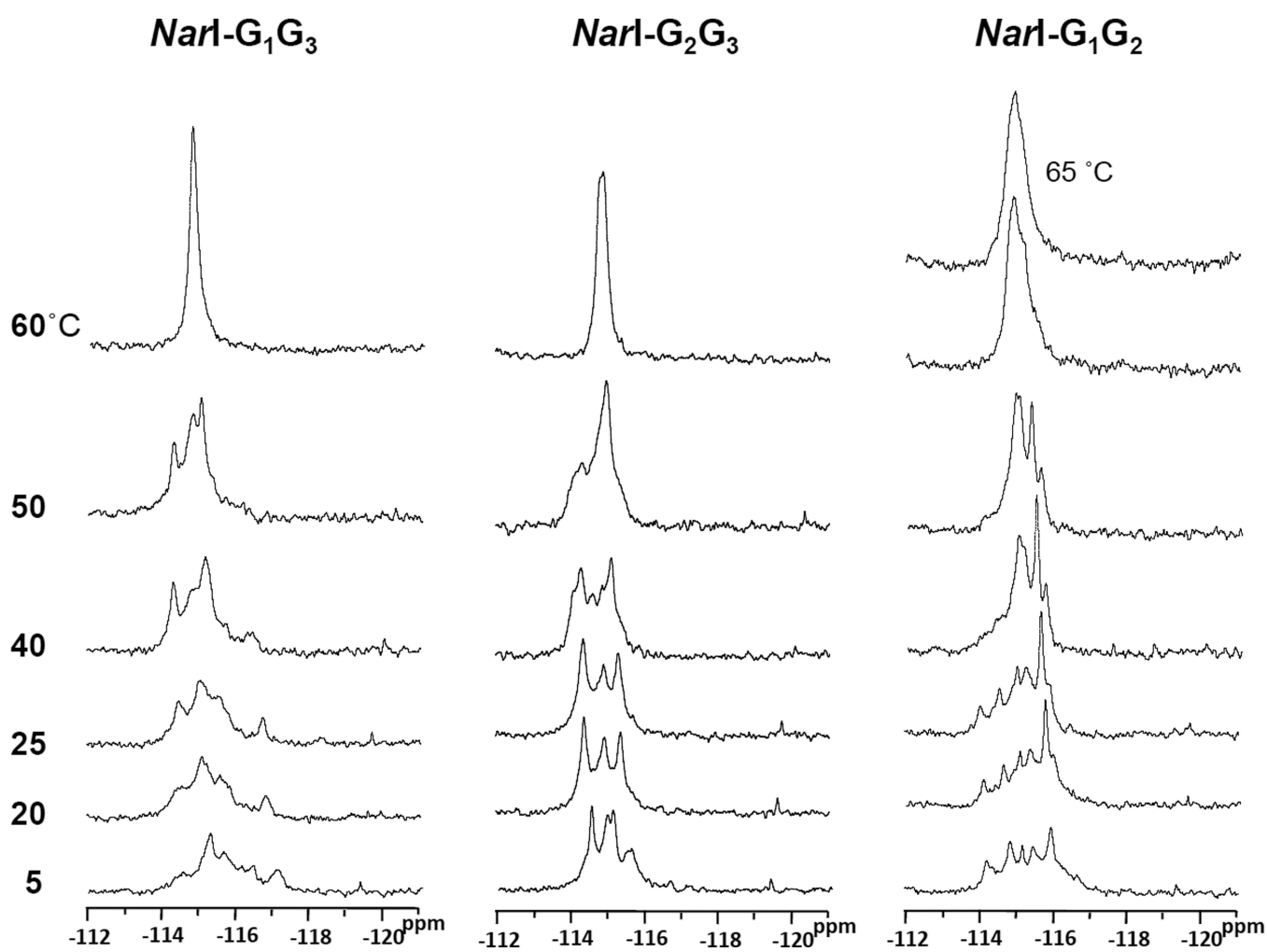
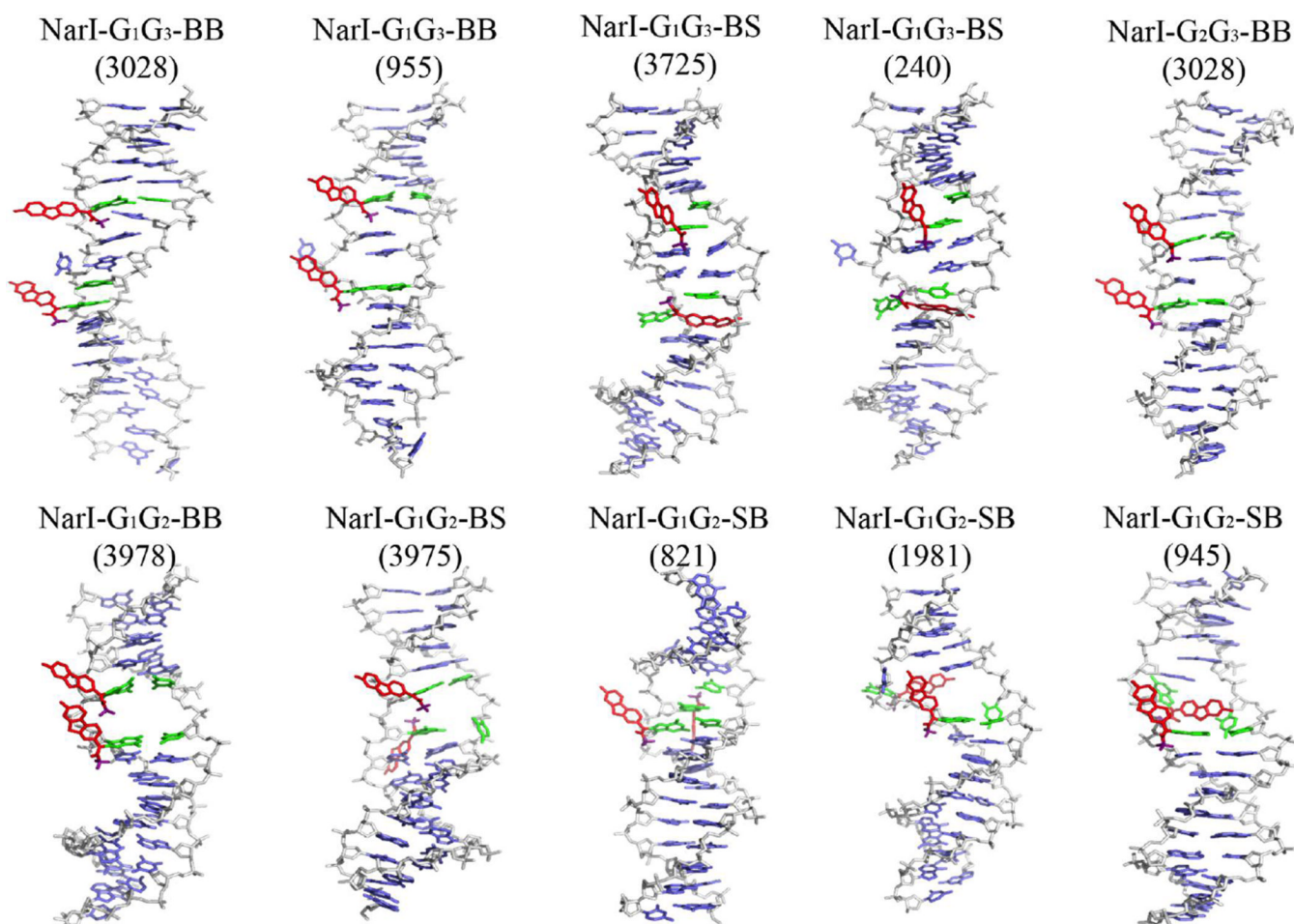


Figure 5.
¹⁹F NMR spectra of di-FAAF-modified duplexes at different temperatures.

**Figure 6.**

Representative conformations of the di-FAAF duplexes for the low energy glycosidic isomers obtained from the large clusters based on RMS Differences of the central 6 nucleotides (B: B-type; S-stacked). FAAF adducts are colored in red, DNA bases are colored in light blue except that G₁, G₂, G₃ and their complimentary cytosine bases are colored in green. Backbone and sugar rings are colored in grey. Number of frames for each cluster are in parentheses as listed in Supporting Information Table S2, which indicates the population of each cluster divided by the total number frames for each isomer (4000 in this study).

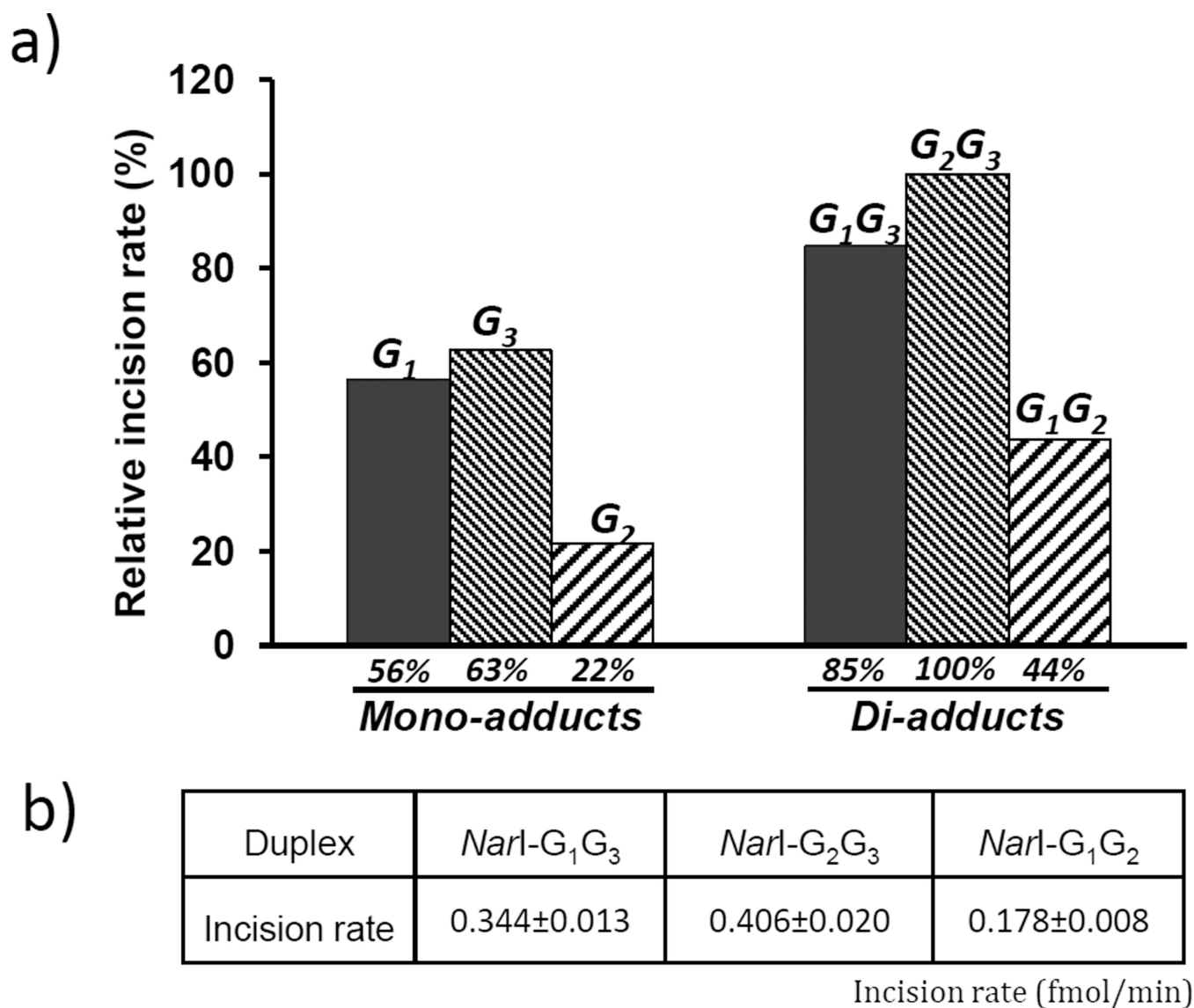


Figure 7. UvrABC nuclease incision kinetic results. (a) comparative relative incision rate histogram of FAAF modified mono- and di-adducts located at different positions calculated by considering *NarI*-G₂G₃ FAAF incision rate as 100%. (b) table showing the absolute incision rate of three di-adduct duplexes.

Thermal and thermodynamic parameters of FAAF modified *NarI* di-adduct duplexes obtained from UV-melting experiments

Table 1

5'-CTCTCG ₁ G ₂ CG ₃ CCATCAC-3' 3'-GAGAGC C GCGGTAGTG-5'									
	$-\Delta H$ kcal/mol	$-\Delta S$ eu	$-\Delta G_{37^\circ C}$ kcal/mol	T_m^b °C	$\Delta\Delta H^c$ kcal/mol	$\Delta\Delta S^d$ eu	$\Delta\Delta G_{37^\circ C}^e$ kcal/mol	ΔT_m^f °C	
Control ^a	121.9	324.7	21.2	70.6	-	-	-	-	-
<i>NarI</i> -G ₁ G ₃ -FAAF ^a	97.8	268.7	14.4	56.5	24.1	56.0	6.8	-14.1	
<i>NarI</i> -G ₂ G ₃ -FAAF ^a	98.6	274.6	13.4	52.7	23.3	50.1	7.8	-17.9	
<i>NarI</i> -G ₁ G ₂ -FAAF ^a	115.7	318.7	16.9	60.6	6.2	6.0	4.3	-10.0	

^a) The average standard deviations for $-\Delta G$, $-\Delta H$, and T_m are ± 0.4 , ± 3.0 , and ± 0.4 , respectively.

^b) T_m values is the temperature at half the peak area.

^c) $\Delta\Delta H = \Delta H$ (modified duplex) $- \Delta H$ (control duplex).

^d) $\Delta\Delta S = \Delta S$ (modified duplex) $- \Delta S$ (control duplex).

^e) $\Delta\Delta G = \Delta G$ (modified duplex) $- \Delta G$ (control duplex).

^f) $\Delta T_m = T_m$ (modified duplex) $- T_m$ (control duplex).

Selected calculated data of the MM/GBSA low energy di-adduct glycosidic isomers. ΔT , $\Delta\Delta H$ and $\Delta\Delta S$ experimental results from Table 1 and Figure 4, with the remaining results from analysis of the MD simulations. The two terminal nucleotides at each end are excluded from the calculations.

Table 2

	ΔT_m (°C)	$\Delta\Delta H$ (kcal/mol)	$-\Delta\Delta S$	Base Pairing ^a (%)	Base stacking (kcal/mol)	Duplex bending (deg.)	SASA (Å ²)	Hydration energy (kcal/mol)
<i>NarI</i> -WT	-	-	-	99.9	-178.2±1.0	20.5±11.8	4176±34	-4286±22
<i>NarI</i> -G ₁ G ₃ -BB	-14.1	24.1	17.3	79.8	-152.4±2.3	37.4±1.9	4718±41	-4605±17
<i>NarI</i> -G ₁ G ₃ -BS				83.3	-155.6±3.6	42.6±4.0	4624±35	-4610±30
<i>NarI</i> -G ₂ G ₃ -BB	-17.9	23.3	15.5	93.5	-160.0±10.0	41.3±8.3	4591±107	-4500±139
<i>NarI</i> -G ₁ G ₂ -BB	-10.0	6.2	1.7	99.7	-170.2±1.5	46.3±5.4	4691±44	-4441±12
<i>NarI</i> -G ₁ G ₂ -BS				85.8	-169.1±9.3	38.1±15.1	4699±56	-4573±17
<i>NarI</i> -G ₁ G ₂ -SB				83.9	-155.4±6.8	34.3±5.5	4661±91	-4397±49

^aThe percent WC base pairing was calculated by applying a cutoff of 3.5 Å for the N1-N3 distance for all base pairs excluding the terminal 2 nucleotides at each end of the duplexes.