Supplementary Data

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.

Hong Mu[1], Konstantin Kropachev[2], Lihua Wang[1], Lu Zhang[2], Alexander Kolbanovskiy[2], Marina Kolbanovskiy[2], Nicholas E. Geacintov[2,*] and Suse Broyde[1,*]

[1]Department of Biology and [2]Department of Chemistry, New York University, 100 Washington Square East, New York, NY 10003, USA.

*To whom correspondence should be addressed. Tel: +1 212 998 8231; Fax: +1 212 995 4015; Email: broyde@nyu.edu.

Correspondence may also be addressed to Tel: +1 212 998 8407; Fax: +1 212 998 8421; Email: ng1@nyu.edu.

INDEX

EXPERIMENTAL SUPPLEMENTARY DATA

SUPPLEMENTARY FIGURES

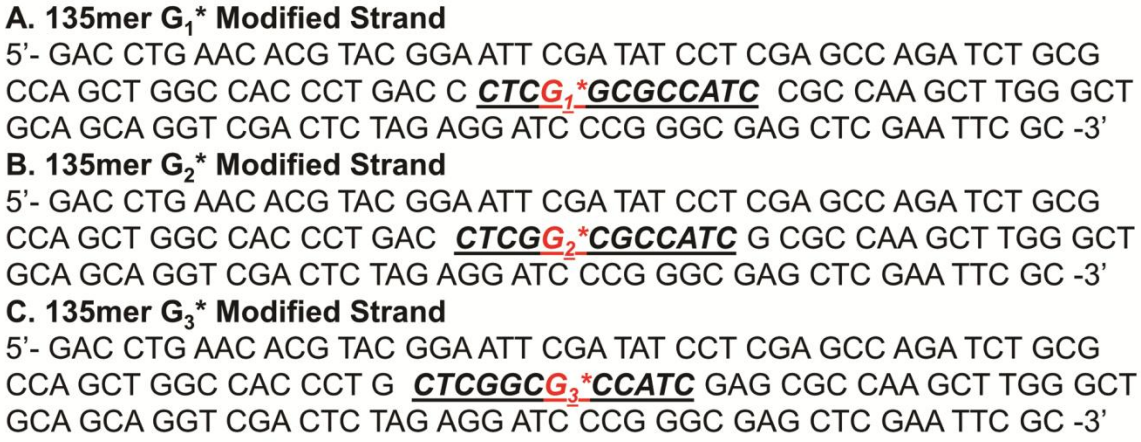


Figure S1. Sequences of 135-mer duplexes for NER experiments. (A) Sequence of the modified 135-mer G_1*strand. (B) Sequence of the modified 135-mer G_2*strand. (C) Sequence of the modified 135-mer G_3*strand. The underlined sequence is the 12-mer *Nar*I sequence containing strand. The complementary strands are normal Watson-Crick partners to each modified strand.

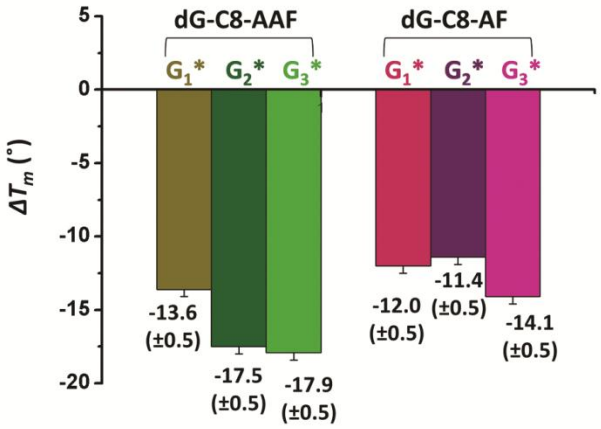


Figure S2. Thermal destabilization of the dG-C8-AAF and dG-C8-AF modified duplexes. ΔT_m = T_m (modified) – T_m (unmodified values). The melting points of the duplexes, T_m, are averages of 3 experiments. The T_m of the unmodified 12-mer duplex is 65.0 (±0.5) ˚C. G_1*, G_2* and G_3* denote 12-mer duplexes 5'-CTC$G_1G_2CG_3$CCATC-3' (Figure 1B in the main text) with adducts at the G_1, G_2, and G_3 positions.

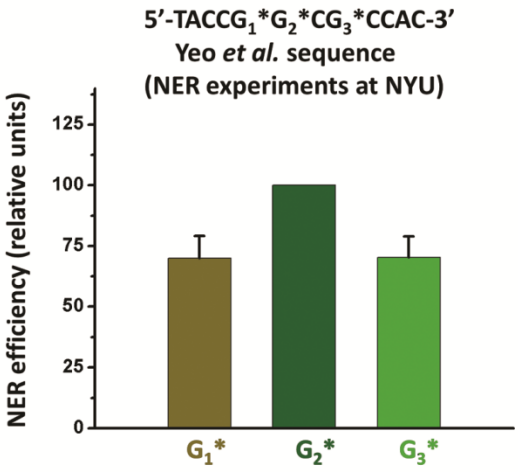


Figure S3. Comparisons of dual incision NER efficiencies of dG-C8-AAF adducts in the *Nar*I sequence used by Yeo *et al.* (2012). This *Nar*I sequence, with the dG-C8-AAF lesion positioned at either G_1*, G_2*, or G_3* in the 12-mer sequence 5'-TACCG_1*G_2*CG_3*CCAC-3' (kindly supplied to us by Prof. O.D. Schärer) was substituted for our own 12-mer sequence 5'-CTCG_1G_2CG_3CCATC-3' into otherwise the same 135-mer double-stranded sequences (Supplementary Figure S1) that were used as templates of NER in HeLa cell extracts in our experiments (Figure 3 in the main text). The results shown in the above Figure are averages of four independent experiments obtained with three different cell extracts (30 min incubation times). In each individual experiment, the NER efficiencies obtained with 135-mer sequences G_1* and G_3* were normalized to the efficiencies obtained with the G_2* sequence (assigned an arbitrary value of 100). The results obtained with the Yeo *et al.* sequence (Yeo, J.E., Khoo, A., Fagbemi, A.F., Schärer, O.D., submitted for publication) is in reasonable agreement with the results obtained with our own 12-mer sequence embedded in the 135-mer linear DNA templates.

MOLECULAR MODELING AND MD SIMULATION SUPPLEMENTARY DATA

SUPPLEMENTARY MATERIALS AND METHODS

Initial structures

All initial structures were modeled using Discovery Studio v2.5 (Accelrys Inc., San Diego, CA, USA). The dG-C8-AF and dG-C8-AAF adducts were incorporated in the centers of the *Nar*I sequence-containing 9-mer duplexes: G_1*, 5'-ACTCG*GCGC-3'; G_2*, 5'-CTCGG*CGCC-3'; and G_3*, 5'-CGGCG*CCAT-3'. For the dG-C8-AF-modified duplexes, the base-displaced intercalated and major groove conformers were constructed based on their respective NMR solution structures (40,41). The terminal base pairs of the NMR structure of the base-displaced dG-C8-AF 11-mer (41) and three base pairs at the 3' end of the lesion-containing strand of the NMR structure of the major groove dG-C8-AF 12-mer (40) were removed to obtain the 9-mer duplexes with G* in the center. The base identities were remodeled to create the G_1*, G_2*, and G_3* duplexes. The 3' and 5' ends of the duplexes were capped with hydrogen atoms. The α' and β' torsion angles were adjusted as shown in Supplementary Table S1 to obtain the rotamers about the long axis of the fluorenyl moiety, governed mainly by rotation of ~ 180° about β'. For the dG-C8-AAF-modified 9-mer duplexes, the base-displaced intercalated conformers were constructed based on the NMR solution structure (35); the major groove conformers were constructed based on the model of Wang and Broyde (49); the minor groove Wedge conformers were constructed based on the model of Shapiro *et al.* (50); the latter two were derived from the major groove and Wedge dG-C8-AF NMR solution structures (37). The α', β' and γ' torsion angles were adjusted as shown in Supplementary Table S1. All acetyl groups were in the less crowded *cis* conformation. The base identities were remodeled to create the G_1*, G_2*, and G_3* duplexes. The initial structure of the unmodified DNA duplex control was a *Nar*I sequence containing the 10-mer B-DNA structure (5'-CTC$G_1G_2CG_3$CCA-3') built using Discovery Studio v2.5 (Accelrys Inc., San Diego, CA, USA). This 10-mer allowed all G_1, G_2 and G_3 positions to be at least three base pairs from the ends.

Force Field

MD simulations were carried out using PMEMD in the AMBER 9 suite of programs (59) with the Cornell *et al.* force field (83), the PARM99 parameter set (84) modified by parmbsc0 (85) and GAFF (86). Partial charges for the dG-C8-AAF adduct were obtained from Wang and Broyde (49). Partial charges for the dG-C8-AF adduct were obtained as described by Cieplak *et al*. (87). Hartree Fock quantum mechanical calculations with 6-31G* basis set were performed without geometry optimization, using the Gaussian 03 package from Gaussian Inc. (88). The RESP algorithm was utilized to fit the charges to each atomic center (87,89). These charges, together with atom types and topologies, are given in Supplementary Table S2. Bond length and bond angle parameters for dG-C8-AAF not present in the parameter set were assigned as in Wang and Broyde (49). All parameters for dG-C8-AF are given in Supplementary Table S2.

MD Computation

Each system was neutralized with Na^+ counterions and solvated with explicit water using the LEAP module of the AMBER 9 suite of programs (59). A periodic rectangular box of TIP3P

water with 10.0 Å buffer was created around the DNA for each sequence. The Particle-Mesh Ewald method with 9.0 Å cutoff for the non-bounded interactions was used in the energy minimizations and MD simulations.

Minimizations were carried out in two stages. First, 5000 steps of steepest descent minimization followed by 5000 cycles of conjugate gradient minimization were conducted for the water molecules and counterions with 500 kcal·mol^{-1}·Å^{-2} restraint on the DNA. Then, 10000 steps of steepest descent minimization followed by 15000 cycles of conjugate gradient minimization were carried out on the whole system without restraints.

A 2.0 fs time step and the SHAKE algorithm were applied in the MD simulations. Each system was heated from 0 K to 300 K over 20 ps with the DNA fixed with a weak restraint of 10 kcal·mol^{-1}·Å^{-2} at constant volume, using the Berendsen coupling algorithm (90) with a 1.0 ps coupling parameter. Each system was equilibrated by 20 ps of constant pressure dynamics at 300K and 1 Atm. Production MD simulations for each system were carried out for 50 ns following equlibration, at 300K and constant pressure of 1 Atm. Temperature and pressure coupling constants were both 0.2 ps.

RMSD

The stabilities of the MD simulations were evaluated for each model. The root mean square deviation (RMSD) of each snapshot in the trajectory relative to its first frame, excluding two base pairs at each end, was plotted as a function of time and is shown in Supplementary Figure S4. For all cases, the MD simulations achieved good stability, fluctuating around the mean, after 20 ns (Supplementary Figure S4). Structural ensembles from 20 ns to 50 ns were employed for analyses of best representative structures, hydrogen bond occupancy, van der Waals interaction energy, duplex helical twist and minor groove width.

Conformational dynamics of β' rotamers for dG-C8-AF-modified duplexes

To investigate the conformational dynamics of the β' rotamer for the base-displaced and major groove conformers of the dG-C8-AF-modified duplexes, we analyzed the time dependence of the β' torsion angles for the 0 ns – 50 ns trajectory of each simulated dG-C8-AF duplex (Supplementary Table S1). In this case we retained the first 20 ns to preserve the initial state following equilibration. The β' torsion angles were defined as in Figure 1A in the main text and calculated using the dihedral command in the Ptraj module of the AMBER 9 suite of programs (59). For the two β' rotamers of the base-displaced conformer, the β' torsion angle values achieved good stability, fluctuating around the mean (Supplementary Figure S5). However, β' rotamer interchanges were observed for the major groove conformer: for the simulations starting with the β' rotamer 1 (Supplementary Table S1), this β' torsion angle fluctuated stably in its own domain with occasional interchange to the β' rotamer 2 (Supplementary Table S1) and back to rotamer 1, for the G_1* and G_2* duplexes; for the simulations starting with the β' rotamer 2 (Supplementary Table S1), the β' torsion angles showed several interchanges to the β' rotamer 1 (Supplementary Table S1) and stabilized in that domain with fluctuations in each rotamer domain (Supplementary Figure S5). In all three modified duplexes (G_1*, G_2*, and G_3* 9-mers), the β' rotamer 1 of the major groove conformer had the largest population of the ensemble following the 20 ns period: rotamer 1 was ~ 90% or more of the population. The less favored

rotamer 2 appeared to have crowding between the fluorenyl ring hydrogen on C1 (Figure 1 in the main text) and the DNA backbone. Hence, structural ensembles for the β' rotamer 1 of the major groove conformer were employed to represent the major groove conformer of the dG-C8-AF for further structural and energetic analyses.

Most Representative Structures

The most representative structure is the frame from an ensemble, which has a conformation that is the closest to all other frames in the ensemble. The most representative structures of all simulated duplexes were obtained using the cluster command in the Ptraj module of the AMBER 10 suite of programs (60). The rms distance metric and K-means clustering method were employed to assign all structures from one ensemble into one cluster and obtain the most representative structure.

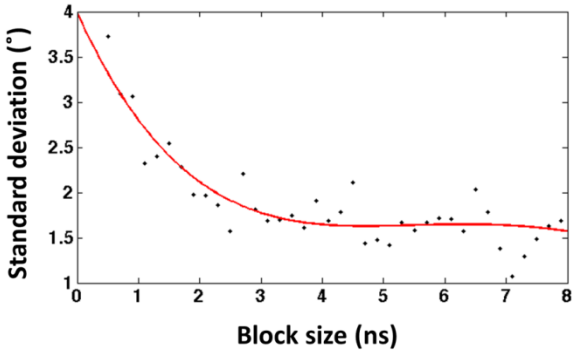
Hydrogen Bond Occupancy

The hydrogen bond occupancy for a given hydrogen bond is the fraction of structures in the ensemble, in which the hydrogen bond exists. The hydrogen bond occupancy for Watson-Crick hydrogen bonds reflects the stability of Watson-Crick base pairing, which contributes to the stability of the DNA duplex. We analyzed the hydrogen bond occupancies of the central trimers of each simulated duplex, using the CARNAL module of AMBER 7 (61). The cut off values were set at 3.4 Å for instantaneous donor (D) – acceptor (A) distance (d_{DA} <= 3.4Å) and 135˚ for instantaneous D−H…A bond angle ($\Theta_{D\text{-}H\ldots A}$ <= 135˚). All hydrogen bonds of the central trimers with occupancies more than 25 % are listed in Supplementary Table S3. The unmodified controls are the analogous base pairs to the trimers of each modified duplexes.

Block averaging

The raw time series data of van der Waals interaction energies and duplex structural properties, such as twist angle and minor groove width, produced by computational MD simulations are a series of correlated data. In order to analyze averages and the variances of averages for these data, we applied the block averaging method (91). In brief, the time series data were divided into "blocks" with a block size that exceeds the longest correlation time, 3 ns in our case (see below). The average for each block was computed and termed "block average". The mean values and the standard deviations of the block averages were used to represent the average and the variance of averages.

The block size was decided upon using the convergence of the standard deviation of the block averages (92). For all raw time series data, we computed the standard deviation of the block averages using a block size from 0.5 ns to 15 ns with an increment of 0.2 ns. The standard deviation values were plotted against the block sizes. The optimal block size was chosen at the point where the standard deviation values converge, reaching a plateau. For example, the convergence of the standard deviations of the block averages for the twist angles of the unmodified G_1 trimer (5'- …CG_1G_2… -3') is shown on the right. The optimal block size is at 3 ns. For all our time series data



from 20 ns to 50 ns, the block size was chosen at 3 ns, giving 10 blocks for each time series data.

van der Waals Interaction Energy

For the base-displaced intercalated conformers, the van der Waals interaction energies between the intercalated lesion aromatic rings and its 3' and 5' neighboring base pairs (lesion – base) were used to evaluate the stacking interaction energies. Due to the difference in lesion intercalation (detailed in the main text), the lesion aromatic rings used for the van der Waals interaction energy calculations were the fluorenyl rings for the dG-C8-AAF-modified duplexes, and both the modified guanine and the fluorenyl rings for the dG-C8-AF-modified duplexes. Since in the major groove and Wedge conformers the lesion-containing base pairs were stacked into the helix between the 3' and 5' neighboring base pairs, with the adduct fluorenyl rings in the major or minor groove (Supplementary Figure S6), the van der Waals interaction energies between the lesion-containing G*:C base pair and its 3' and 5' neighboring base pairs were used to evaluate the stacking interactions. For the Wedge conformers of the dG-C8-AAF, we include van der Waals interactions between the fluorenyl ring system and the DNA atoms in contact with the walls of the minor groove (backbone atoms of nucleotide 6, 7, 14 and 15 of the 9-mer duplexes, Supplementary Table S1). This van der Waals interaction provides considerable stabilization to the fluorenyl rings in the minor groove. For the major groove conformers, the van der Waals interactions between the fluorenyl rings and the DNA atoms in contact in the walls of the major groove were calculated as well. Since these interaction energies were close to zero, they were not included in the stabilizing van der Waals interaction energies for the major groove conformers.

The van der Waals interaction energies for each structure of each simulated conformer were calculated using the ANAL module of the AMBER 9 suite of programs (59). For each structure, the van der Waals interaction energy (E_{vdW}) value was obtained by summing up the van der Waals interaction energies between all interacting residues: lesion aromatic rings and all neighboring bases for the base-displaced conformers, lesion-containing Watson-Crick G*:C base pair and all neighboring base pairs for the major groove conformers, and lesion-containing non-Watson-Crick G*:C base pair (Supplementary Figure S7) and all neighboring base pairs plus fluorenyl rings and minor groove DNA residues in contact for the Wedge conformers. The mean and standard deviation of the block averages for E_{vdW} are given in Supplementary Table S4 for the major groove and Wedge conformers, and in Supplementary Table S6 for base-displaced conformers. The population distributions of E_{vdW} are shown in Figure 4 in the main text for the major groove and Wedge conformers, and in Figure 6 in the main text for base-displaced conformers. Because the van der Waals interaction energy between the central base pair and its neighboring base pairs of the analogous trimers in the unmodified duplex are very similar (Supplementary Table S5), for clarity we used the mean values over all three trimers as the unmodified control for the major groove and Wedge conformers. We used Matlab 7.10.0 to analyze the population distribution and calculate the mean and standard deviation of the block averages for E_{vdW}.

Duplex Helical Twist

The twist of the DNA duplex backbone around the lesion site was represented using the twist angle between the base pairs 3' and 5' to the lesion site (T_{mod}). For all modified duplexes, the

T_{mod} for the structures was calculated using the Curves+ software (62) with the lesion-containing base pair excluded. The twist angle between the 3' and 5' base pairs of each trimer sequence, CGG, GGC and CGC trimers, (two nucleotide steps) of the unmodified duplex was calculated using the Curves+ software (62). Since the twist angles of the three unmodified trimers were very close in value (Supplementary Table S5), we used the average values of these twist angles as the unmodified control (T_{unmod}). The mean and standard deviation of the block averages for T_{unmod} and T_{mod} are given in Supplementary Table S4 and S6. The population distribution of T_{unmod} and T_{mod} are shown in Supplementary Figure S8 and Figure 8A in the main text. DNA duplex unwinding is defined as $T_{unmod} - T_{mod}$ (Figure 5A in the main text). The greater the difference between T_{mod} and T_{unmod}, the more untwisted the modified duplex. The mean and standard deviation of the block averages for the untwisting angles for all conformers are shown as bar plots in Figures 5B and 8A in the main text.

Minor Groove Width

The minor groove widths of the DNA duplexes at the lesion site (W_{mod}) for structures of each modified duplex were calculated using the Curves+ software (62) with the lesion containing base pair excluded. The minor groove width of each trimer sequence, CGG, GGC and CGC trimers, of the unmodified duplex was calculated using the Curves+ software (62). We used the minor groove widths at analogous positions in the unmodified duplexes as the unmodified controls (W_{unmod}). DNA duplex minor groove opening is defined as $\Delta W = W_{mod} - W_{unmod}$. The larger the W_{mod} than W_{unmod}, the more enlarged the minor groove of the modified duplex. The mean values and standard deviation of the block averages for ΔW of all conformers are given in Supplementary Table S4 and S6 and shown as bar plots in Figures 5C and 8B in the main text. Supplementary Figure S9 shows pairs of phosphate atoms for the calculations of minor groove widths at the lesion site of modified duplexes and those at the analogous positions of the unmodified control.

Structural, Dynamic and Energetic Properties of Base-displaced Conformers

Sugar Pucker Phase Angle. The dynamics of the sugar pucker were revealed using the time dependence of the sugar pucker pseudorotation phase angle, ***P***, which indicates the sugar pucker conformation (93). For the base-displaced conformers, the ***P*** values of the modified guanines were calculated with the Altona and Sundaralingam algorithm (94) using the pucker command in the Ptraj module of the AMBER 9 suite of programs (59). The time dependences of the ***P*** values are shown in Supplementary Figure S10.

Conformational dynamics of the γ' cis/trans rotamers for the base-displaced dG-C8-AAF-modified duplexes. To investigate the conformational dynamics of the γ' *cis*/*trans* rotamers for the base-displaced conformers of dG-C8-AAF-modified duplexes, we analyzed the time dependence of the γ' torsion angles for each base-displaced dG-C8-AAF duplex (Supplementary Table S1). The γ' torsion angles are defined in Figure 1A in the main text and calculated using the dihedral command in the Ptraj module of the AMBER 9 suite of programs (59). The time dependence and population distribution of the γ' torsion angle values are shown in Supplementary Figure S13. The γ' torsion angle of the base-displaced dG-C8-AAF in the G_3* duplex alternates between the *cis* and *trans* domains (Supplementary Movie S3), and fluctuates

stably in each domain, indicating the free interchange between *cis* and *trans* conformations of the acetyl group.

Structural origin for cis/trans γ' rotamer preferences of the dG-C8-AAF in the base-displaced intercalated conformation. The time dependence and population distributions of the γ' torsion angle values (Supplementary Figure S13) show that there are sequence dependent preferences for the *cis/trans* rotamers of the acetyl group. We noted that the acetyl group is 100% *trans* in G_1* and 100% *cis* in G_2* (Supplementary Figure S13), in order to study the structural origin of these preference we modeled the *cis* rotamer for the G_1* duplex by rotating the γ' torsion angle of its most representative structure (*trans* rotamer) (Supplementary Figure S14A) to the *cis* domain (0˚), and the *trans* rotamer for the G_2* duplex by rotating the γ' torsion angle of its most representative structure (*cis* rotamer) (Supplementary Figure S14B) to the *trans* domain (180˚). Both *cis* and *trans* rotamers were significantly populated in the G_3* duplex (Supplementary Figure S13); the most representative structure (Supplementary Figure S14C) for each rotamer domain was used to study its structural origin. The structures with $100^\circ \leq \gamma' \leq 260^\circ$ were assigned into the *trans* rotamer cluster, while the rest were assigned into the *cis* rotamer cluster (Supplementary Figure S13). The most representative structure for each cluster was obtained using the methods described above. All structures, modeled and most representative ones, are shown in Supplementary Figure S15, which explains the structural origins of the rotamer preferences, in the figure caption.

van der Waals interaction between the displaced partner C and DNA residues. For the base-displaced conformers, the displaced partner C bases are in dynamic contact with the DNA residues in close vicinity. Their van der Waals interaction energies with nearby DNA residues reflect the position of the displaced partner C and can stabilize this base. For each structure of the base-displaced conformers, we calculated the van der Waals interaction energies between the displaced partner C base and each DNA residue in contact (namely its own sugar ring, the neighboring nucleotides on its own strand and the fluorenyl rings) using the ANAL module of the AMBER 9 suite of programs (59). These energies were summed up to obtain the van der Waals interaction energy value between the displaced partner C base and interacting DNA residues (Partner C – DNA). The mean values and standard deviations of the block averages are given in Supplementary Table S7.

Analyses of Twist Angle Populations

Population Deconvolution of the Twist Angle Distribution. The population distribution of the twist angle of the base-displaced conformer for the dG-C8-AF-modified G_2* duplex (Supplementary Figure S16) reveal two distinct populations. To investigate the proportions and the distributions of these two populations, we assumed the two populations adopt a normal distribution and deconvoluted them using Matlab 7.10.0. The twist angles were separated into two populations with a threshold of 48.65°: population 1 (P1) with twist angles < 48.65°, and population 2 (P2) with twist angle ≥ 48.65°. The normfit function was used to fit normal distribution models to P1 and P2, and obtain mean values (μ1 and μ2) and standard deviations (σ1 and σ2) of the populations. The sum of the deconvoluted models is shown as the following function:

$$\int x = \frac{p}{\sigma 1\sqrt{2\pi}} e^{\frac{-(x-\mu 1)^2}{2\sigma 1^2}} + \frac{1-p}{\sigma 2\sqrt{2\pi}} e^{\frac{-(x-\mu 2)^2}{2\sigma 2^2}}$$

x is the twist angle, and p is the proportion of P1. The p value was obtained by fitting the summed model ($\int x$) to the actual distribution of the twist angles of the base-displaced conformer for the dG-C8-AF-modified G_2* duplex. In brief, p was assigned values from 0.00 to 1.00 with a step of 0.01. The p value, with which the function has the least square fit, was obtained as the best estimation of the P1 proportion. The proportion of P2 is 1 – p.

Most Representative Structure for each Population. The most representative structures for P1 and P2 of the base-displaced dG-C8-AF in the G_2* duplex were obtained using the methods detailed above. In order to focus on the heterogeneity reflected by the twist angles, the clustering was based on the structures of the central trimers. The most representative structures of each cluster are shown in Figure 10C and 10D in the main text.

SUPPLEMENTARY RESULTS

The major groove and Wedge conformational families are little distorted compared to the unmodified control.

For both dG-C8-AAF and dG-C8-AF-modified duplexes, the major groove conformers retain normal B-DNA Watson-Crick base pairing (Supplementary Table S3) and strong stacking interactions that are comparable to the unmodified control (Figure 4 in the main text and Supplementary Table S4). Furthermore, for the dG-C8-AAF-modified duplexes, the Wedge conformers are also close to normal, with one strong non-Watson-Crick hydrogen bond between the G* and partner C as well as a second weak one in two of the three sequences (Supplementary Table S3, Figures S6 and S7), and strong van der Waals stacking interactions comparable to unmodified DNA (Figure 4 in the main text and Supplementary Table S4). The helical twist and minor groove dimensions of these major groove and Wedge conformers are much closer to normal (Supplementary Table S4, Figure S8 and Figure 5 in the main text) than for the base-displaced ones (Supplementary Table S6 and Figure 8 in the main text). Hence we concluded that the base-displaced conformers are the best NER substrates.

Cis/trans rotamers of the acetyl group further modulate the stacking of fluorenyl rings and neighboring bases for the dG-C8-AAF-modified duplexes.

It has been known for many decades from NMR studies that the acetyl group can adopt *cis* and *trans* rotamers about the amide bond (defined by the torsion angle γ', Figure 1), with γ' around 0˚ and 180˚ respectively (95,96). However, the steric restraints imposed by the acetyl group differ in the *cis* and *trans* rotamers. For the base-displaced conformation, our simulations showed a striking sequence-dependent *cis*/*trans* rotamer distribution, being 100% *trans* in G_1*, 100% *cis* in G_2* and 32% *trans*/68% *cis* in the G_3* sequence (Supplementary Figure S13 and Movie S3), although all simulations began in the less crowded *cis* domain (Figure 2 and Table S1). As can be seen from Figure 7A and stereo views in Supplementary Figure S14, in the *cis* rotamer the bulky methyl group protrudes out into the major groove, while the smaller carbonyl oxygen is directed towards the base pair 3' to the modified guanine G*. However, in the *trans* rotamer, the methyl group is directed towards the 3' base pair (Figure 7A and Supplementary Figure S14). Thus, the *trans* rotamer is sterically more restricting: it inhibits how far into the duplex the fluorenyl ring system can intercalate since the methyl group obstructs further insertion (Figure 7A and Supplementary Figure S14). Thus, the sequence dependence of fluorenyl ring stacking is modulated by the *cis/trans* rotamer preference. Explanations of the sequence dependence of this *cis/trans* preference are provided in Supplementary Figure S15.

Populations in the MD simulations with extrusion of the fluorenyl rings into the minor groove suggest a pathway for rotation about the fluorenyl long axis

NMR studies have shown two populations of rotamers representing a flip about the fluorenyl long axis and governed by the β' torsion angle for base-displaced dG-C8-AF (37). We wished to analyze our MD simulations for possible insights on this rotational reorientation, which happens on the NMR time scale of milli to micro seconds, and so could not be observed in our 50 ns simulations. However, we gained some clues from analyses of the distributions of twist angles between the base pairs 5' and 3' to the lesion for all six duplexes (Figure 8A in the main text).

We observed two distinct populations in the G_2* duplex with the base-displaced β' rotamer 1 of dG-C8-AF adduct: one population is characterized by a mean value of the twist angle of ~45˚ that represents ~80% of the ensemble, and a second one with a mean value of twist angle at ~60˚ represents ~20% of the ensemble (Supplementary Figure S16). These populations have different extents of fluorenyl ring protrusion into the minor groove. Figures 10C and 10D show representative structures of each twist population, which illustrate the different positions of the fluorenyl rings in the two populations. The more untwisted (twist angle of ~45˚) population exhibits some stacking interactions between the modified guanine and its 5'-G and a well stacked fluorenyl aromatic ring system, but the displaced partner C and to a large extent its 3'-G, are unstacked; in the less untwisted population (twist angle of ~60˚), the modified G* and its 5' G are well stacked, while the partner C and 3'-G are stacked with each other; these stacking interactions are at the expense of poorer stacking for the fluorenyl rings, which protrude into the minor groove. In addition, while G_1* and G_3* did not have distinct sub-populations with the large twist angle, in each case there were ensemble members with larger twists, associated with greater minor groove protrusion of the fluorenyl moiety. In the population with more minor groove protrusion of the fluorenyl rings, we observed that the fluorenyl moiety can rotate substantially about its long axis so that in extreme cases it is half flipped (Supplementary Figure S18). The rotations of the fluorenyl rings in the minor groove may suggest a pathway for β' rotamer interchange in the base-displaced dG-C8-AF adduct (Supplementary Movie S4).

SUPPLEMENTARY TABLES

Table S1. Structures investigated with MD simulation studies for dG-C8-AAF and dG-C8-AF-modified *Nar*I sequence containing duplexes.

Adduct	Conformers	α' (°)	β' (°)	γ' (°)	Sequences [a]
dG-C8-AAF	Base-displaced	44	0	44	G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
	Major groove	56	16	32	G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
	Wedge	208	317	18	G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
dG-C8-AF	Base-displaced (rotamer 1)	212	138		G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
	Base-displaced (rotamer 2)	197	322		G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
	Major groove (rotamer 1)	200	38		G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')
	Major groove (rotamer 2)	191	213		G_1* (5'-ACTCG*GCGC-3') G_2* (5'-CTCGG*CGCC-3') G_3* (5'-CGGCG*CCAT-3')

a. For all 9-mer duplexes, the nucleotides in the lesion-containing strand were assigned a residue number 1 – 9 from the 5' to the 3' side, and the nucleotides in the partner strand were assigned a number 10 – 18 from the 5' to the 3' side. Structures of these models in the G_2* 9-mer duplexes are given in Figure 2 of the main text.

Table S2. Atom names, atom types, topologies, partial charges and added force field parameters for dG-C8-AF.

Atom Name	Atom Type	Topology	Partial Charge
P	P	M	1.220169
O1P	O2	E	-0.78989
O2P	O2	E	-0.78989
O5'	OS	M	-0.53621
C5'	CT	M	-0.02271
H5'1	H1	E	0.114141
H5'2	H1	E	0.114141
C4'	CT	M	0.052865
H4'	H1	E	0.140361
O4'	OS	S	-0.34429
C1'	CT	B	-0.02634
H1'	H2	E	0.17463
N9	N*	S	0.051403
C4	CB	S	0.059526
N3	NC	S	-0.31313
C2	CA	B	0.439351
N2	N2	B	-0.85497
H21	H	E	0.40509
H22	H	E	0.40509
N1	NA	B	-0.33983
H1	H	E	0.319146
C6	C	B	0.495863
O6	O	E	-0.58194
C5	CB	S	0.074858
N7	NB	S	-0.4638
C8	CK	S	0.368851
N	N2	B	-0.56432
HN	H	E	0.351594
C12	CA	S	0.139257
C13	CA	B	-0.05009
H13	HA	E	0.102591
C14	CA	B	-0.25895
H14	HA	E	0.19615
C15	CB	S	-0.00696
C16	CB	S	0.028318
C17	CA	B	-0.15471
H17	HA	E	0.155965
C18	CA	B	-0.19866

Added Parameters

Bond Length Parameters:

Bond	K_r [a]	r_{eq} [b]
CK-N2	449	1.364
CB-CT	317	1.510

Bond Angle Parameters:

Angle	K_r [c]	θ_{eq} [d]
N*-CK-N2	72.4	123.05
N2-CK-NB	72.4	123.05
CK-N2-H	48.9	117.16
CK-N2-CA	64.9	125.56
CA-CA-N2	69.3	120.13
CA-CB-CT	70	128.78
CB-CB-CT	70	110.85

Torsion Angle Parameters

Torsion	# of paths	$V_n/2$ [e]	γ [f]	n
X-CK-N2-X	4	4.2	180	2
X-CB-CT-X	6	0.0	0	2

a. kcal/(mol Å^2), b. Å,
c. kcal/(mol radian^2), d. degrees,
e. kcal/mol, f. degrees.

H18	HA	E	0.158289	
C19	CA	B	-0.15403	
H19	HA	E	0.153807	
C20	CA	B	-0.19668	
H20	HA	E	0.14786	
C21	CB	S	-0.00277	
C22	CT	3	0.033749	
H221	HC	E	0.052435	
H222	HC	E	0.052435	
C23	CB	S	-0.01508	
C24	CA	S	-0.21204	
H24	HA	E	0.136282	
C3'	CT	M	0.068849	
H3'	H1	E	0.13804	
C2'	CT	B	0.022649	
H2'1	HC	E	0.032795	
H2'2	HC	E	0.032795	
O3'	OS	M	-0.56208	

Table S3A. Hydrogen bond occupancies of the central trimers of the base-displaced (BD), major groove (MG) and Wedge conformers for dG-C8-AAF-modified duplexes and of the unmodified control (Unmod).

Sequence	Base pair	Hydrogen bond	BD (%)	MG (%)	Wedge (%)	Unmod (%)
dG-C8-AAF CG_1*G	**5' C ≡ G**	**O6(G) – N4(C)**	95.5	98.9	98.2	98.7
		N1(G) – N3(C)	99.5	100.0	99.7	100.0
		N2 (G) – O2(C)	99.8	99.9	98.7	99.8
	G* ≡ C	**O6(G*) – N4(C)**	0	94.3	93.9 [a]	98.9
		N1(G*) – N3(C)	0	99.8	0	99.9
		N2 (G*) – O2(C)	0	99.9	0	99.9
		N7(G*) – N4(C)	0	0	30.8	0
	3' G ≡ C	**O6(G) – N4(C)**	99.0	98.4	91.8	98.8
		N1(G) – N3(C)	100.0	100.0	99.6	100.0
		N2 (G) – O2(C)	99.9	99.9	99.7	99.8
	G*:backbone	**N1(G*) – O1P(dC4)**	92.0	0	0	0
		N2(G*) – O1P(dC4)	92.0	0	0	0
dG-C8-AAF GG_2*C	**5' G ≡ C**	**O6(G) – N4(C)**	98.9	98.2	96.8	98.9
		N1(G) – N3(C)	99.9	99.8	99.8	99.9
		N2 (G) – O2(C)	99.5	99.9	99.7	99.9
	G* ≡ C	**O6(G*) – N4(C)**	0	95.9	75.8	98.8
		N1(G*) – N3(C)	0	99.8	0	100.0
		N2 (G*) – O2(C)	0	98.9	0	99.8
		N7(G*) – N4(C)	0	0	0	0
	3' C ≡ G	**O6(G) – N4(C)**	99.5	99.2	97.2	99.1
		N1(G) – N3(C)	100.0	100.0	99.8	100.0
		N2 (G) – O2(C)	99.9	99.9	99.1	99.9
	G*:backbone	**N1(G*) – O1P(dC4)**	91.7	0	0	0
		N2(G*) – O1P(dC4)	90.2	0	0	0
dG-C8-AAF	**5' C ≡ G**	**O6(G) – N4(C)**	97.4	98.5	98.8	99.1
		N1(G) – N3(C)	100.0	99.9	99.9	100.0

CG_3*C		**N2 (G) – O2(C)**	99.8	99.8	99.4	99.9
	G* ≡ C	**O6(G*) – N4(C)**	0	96.3	99.1	99.0
		N1(G*) – N3(C)	0	99.9	0	100.0
		N2 (G*) – O2(C)	0	99.8	0	99.9
		N7(G*) – N4(C)	0	0	30.8	0
	3' C ≡ G	**O6(G) – N4(C)**	99.4	99.0	97.7	99.4
		N1(G) – N3(C)	100.0	100.0	99.8	100.0
		N2 (G) – O2(C)	99.7	99.9	99.9	99.7
	G*:backbone [b]	**N1(G*) – O1P(dC4)**	91.7	0	0	0
		N2(G*) – O1P(dC4)	84.5	0	0	0

a. The hydrogen bond occupancies of Non-Watson-Crick hydrogen bonds are shown in gray highlights.

b. In the base-displaced conformers, the modified guanines form hydrogen bonds with the phosphate oxygen atoms of the backbone (Figure 7 in the main text).

Table S3B. Hydrogen bond occupancies of the central trimers of the base-displaced β' rotamer 1 (BD1), rotamer 2 (BD2) and major groove rotamer 1 (MG1) for dG-C8-AF-modified duplexes and of the unmodified control (Unmod).

Sequence	**Base pair**	**Hydrogen bond**	**BD1 (%)**	**BD2 (%)**	**MG1 (%)**	**Unmod (%)**
dG-C8-AF CG_1*G	**5' C ≡ G**	**O6(G) – N4(C)**	84.0	96.3	97.3	98.7
		N1(G) – N3(C)	98.7	99.9	99.9	100.0
		N2 (G) – O2(C)	96.4	98.8	99.6	99.8
	G* ≡ C	**O6(G) – N4(C)**	0	0	93.7	98.9
		N1(G) – N3(C)	0	0	99.0	99.9
		N2 (G) – O2(C)	0	0	99.9	99.9
		N4(C14)-O1P(dC13) [a]	67.9	35.6	0	0
	3' G ≡ C	**O6(G) – N4(C)**	97.8	95.9	97.3	98.8
		N1(G) – N3(C)	99.8	97.8	99.7	100.0
		N2 (G) – O2(C)	99.5	99.8	99.8	99.8
dG-C8-AF CG_2*G	**5' C ≡ G**	**O6(G) – N4(C)**	98.0	91.4	96.9	98.7
		N1(G) – N3(C)	99.7	94.8	99.9	100.0

		N2 (G) – O2(C)	98.9	94.6	99.7	99.8
	G* ≡ C	**O6(G) – N4(C)**	0	0	97.3	98.9
		N1(G) – N3(C)	0	0	99.8	99.9
		N2 (G) – O2(C)	0	0	99.4	99.9
		N4(C14)-O1P(dC13) [a]	58.2	25.0	0	0
	3' G ≡ C	**O6(G) – N4(C)**	98.5	98.0	97.8	98.8
		N1(G) – N3(C)	100.0	99.8	100.0	100.0
		N2 (G) – O2(C)	99.7	99.6	99.9	99.8
dG-C8-AF CG_3*G	**5' C ≡ G**	**O6(G) – N4(C)**	98.7	98.6	97.0	98.7
		N1(G) – N3(C)	100.0	99.8	100.0	100.0
		N2 (G) – O2(C)	99.8	99.5	99.9	99.8
	G* ≡ C	**O6(G) – N4(C)**	0	0	94.3	98.9
		N1(G) – N3(C)	0	0	99.7	99.9
		N2 (G) – O2(C)	0	0	99.8	99.9
		N4(C14)-O1P(dC13) [a]	76.6	58.9	0	0
	3' G ≡ C	**O6(G) – N4(C)**	98.5	98.3	96.5	98.8
		N1(G) – N3(C)	99.9	99.8	99.8	100.0
		N2 (G) – O2(C)	99.2	99.5	99.8	99.8

a. In the base-displaced conformers, the displaced partner C bases form hydrogen bonds with the phosphate oxygen atoms of the backbone.

Table S4. Structural and energetic properties of major groove conformers of dG-C8-AF duplexes, major groove and Wedge conformers of dG-C8-AAF duplexes, and unmodified control.

Adduct	Sequence	Stacking interaction [a] (kcal/mol)	Twist angle [a] (°)	Minor groove differences [a] (Å)
dG-C8-AAF (major groove)	CG_1*G	-29.8 (± 0.3)	52.1 (± 1.4)	0.3 (± 0.6)
	GG_2*C	-28.5 (± 0.3)	56.9 (± 2.0)	-0.5 (± 0.5)
	CG_3*C	-29.1 (± 0.4)	52.3 (± 3.8)	1.5 (± 1.4)
dG-C8-AAF (Wedge)	CG_1*G	-31.7 (± 0.2)	60.6 (± 2.3)	1.3 (± 0.5)
	GG_2*C	-31.6 (± 1.2)	58.7 (± 2.0)	1.7 (± 0.2)
	CG_3*C	-32.2 (± 0.2)	58.3 (± 0.6)	2.3 (± 0.6)
dG-C8-AF (major groove [b])	CG_1*G	-29.0 (± 0.8)	57.7 (± 2.1)	-2.1 (± 1.5)
	GG_2*C	-30.6 (± 0.1)	61.8 (± 1.5)	-1.9 (± 0.4)
	CG_3*C	-29.9 (± 0.4)	61.0 (± 1.5)	-1.1 (± 0.9)
Unmodified control		-30.8 (± 0.1)	64.5 (± 1.0)	0.0

a. The mean values and standard deviations of the block averages (see Supplementary Methods) for the time series data between 20 ns and 50 ns.

b. For the major groove conformers of dG-C8-AF, the values for β' rotamer 1 (Supplementary Table S1) are shown in the table.

Table S5. Stacking energies and twist angles of unmodified CG_1G, GG_2C and CG_3C trimers.

Sequence	Stacking interaction [a] (kcal/mol)	Twist angle [a] (°)
CG_1G	-30.3 (± 0.2)	59.8 (± 2.1)
GG_2C	-30.9 (± 0.1)	69.2 (± 2.7)
CG_3C	-31.2 (± 0.1)	64.3 (± 3.0)
Unmodified Control [b]	-30.8 (± 0.1)	64.5 (± 1.0)

a. The mean values and standard deviations of the block averages (see Supplementary Methods) for the time series data between 20 ns and 50 ns.

b. Since the values for all three sequences are close, we used the average of the values for all three sequences as the unmodified control.

Table S6. Structural distortions of the base-displaced dG-C8-AAF and dG-C8-AF duplexes.

Adduct	Sequence	Stacking interaction [a] (kcal/mol)	Twist angle [a] (˚)	Minor groove enlargement [a] (Å)
dG-C8-AAF	CG_1*G	-19.3 (± 0.2)	30.6 (± 2.5)	5.0 (± 0.6)
	GG_2*C	-19.2 (± 0.4)	24.8 (± 1.9)	6.6 (± 0.5)
	CG_3*C	-19.5 (± 0.6)	32.2 (± 2.3)	5.4 (± 1.1)
dG-C8-AF (rotamer 1) [b]	CG_1*G	-32.1 (± 0.3)	32.8 (± 4.8)	2.6 (± 0.3)
	GG_2*C	-29.6 (± 1.0)	44.7 (± 7.5) 60.4 (± 3.0)[c]	2.0 (± 1.5) 0.9 (± 1.2)[c]
	CG_3*C	-30.5 (± 0.5)	41.2 (± 6.2)	2.3 (± 0.5)
dG-C8-AF (rotamer 2) [b]	CG_1*G	-31.3 (± 0.7)	34.2 (± 2.9)	2.1 (± 0.9)
	GG_2*C	-30.6 (± 0.3)	41.3 (± 6.2)	1.9 (± 0.6)
	CG_3*C	-30.7 (± 0.4)	37.5 (± 1.9)	2.0 (± 0.7)
Unmodified control		–	64.5 (± 1.0)	0.0

a. The mean values and standard deviations of the block averages (see Supplementary Methods) for the time series data between 20 ns and 50 ns.

b. The ensemble average values for the β' rotamer 1 are 171˚, 158˚, and 166˚, respectively for the G_1*, G_2* and G_3* duplexes; for the β' rotamer 2 the ensemble average values are –10˚, –17˚ and -14˚, respectively.

c. For dG-C8-AF-modified G_2*, there are two twist angle populations (Supplementary Figure S16). The mean and standard deviation values of the twist angle and minor groove enlargement for each population (Supplementary Figure S16) are given.

Table S7. The van der Waals interaction energies between the displaced partner C and adjacent DNA residues (Partner C / DNA) for the dG-C8-AAF and dG-C8-AF base-displaced duplexes.

Adduct	Sequence	Partner C / DNA (kcal/mol)
dG-C8-AAF	CG_1*G	-6.8 (± 0.5)
	GG_2*C	-6.7 (± 0.5)
	CG_3*C	-6.2 (± 1.0)
dG-C8-AF (rotamer 1) [a]	CG_1*G	-3.6 (± 1.0)
	GG_2*C	-5.8 (± 1.0)
	CG_3*C	-4.4 (± 1.4)
dG-C8-AF (rotamer 2) [a]	CG_1*G	-5.1 (± 0.2)
	GG_2*C	-4.9 (± 0.2)
	CG_3*C	-1.7 (± 0.1)

a. The ensemble average values for the β' rotamer 1 are 171˚, 158˚, and 166˚, respectively for the G_1*, G_2* and G_3* duplexes; for the β' rotamer 2 the ensemble average values are –10˚, –17˚ and -14˚, respectively.

Table S8. Population ratios of conformational families (Base-displaced, Major groove and Wedge) of dG-C8-AAF and dG-C8-AF modified *Nar*I duplexes observed using solution NMR methods.

Adduct	Sequence	Base-displaced (%)	Major groove (%)	Wedge (%)	Other (%)
dG-C8-AAF	CG_1*G	34 [a]	46 [a]	20 [a]	–
	GG_2*C	15 [a]	57 [a]	9 [a]	14 and 5 [a]
	CG_3*C	61 [a]	13 [a]	26 [a]	–
dG-C8-AF	CG_1*G	58 [a] 30 [b]	42 [a] 70 [b]	–	–
	GG_2*C	31 [a] 10 [b]	69 [a] 90 [b]	–	–
	CG_3*C	65 [a] 50 [b]	35 [a] 50 [b]	–	–

a. The data are obtained from Jain *et al.* (32)

b. The data are obtained from Patel *et al.* (37)

SUPPLEMENTARY FIGURES

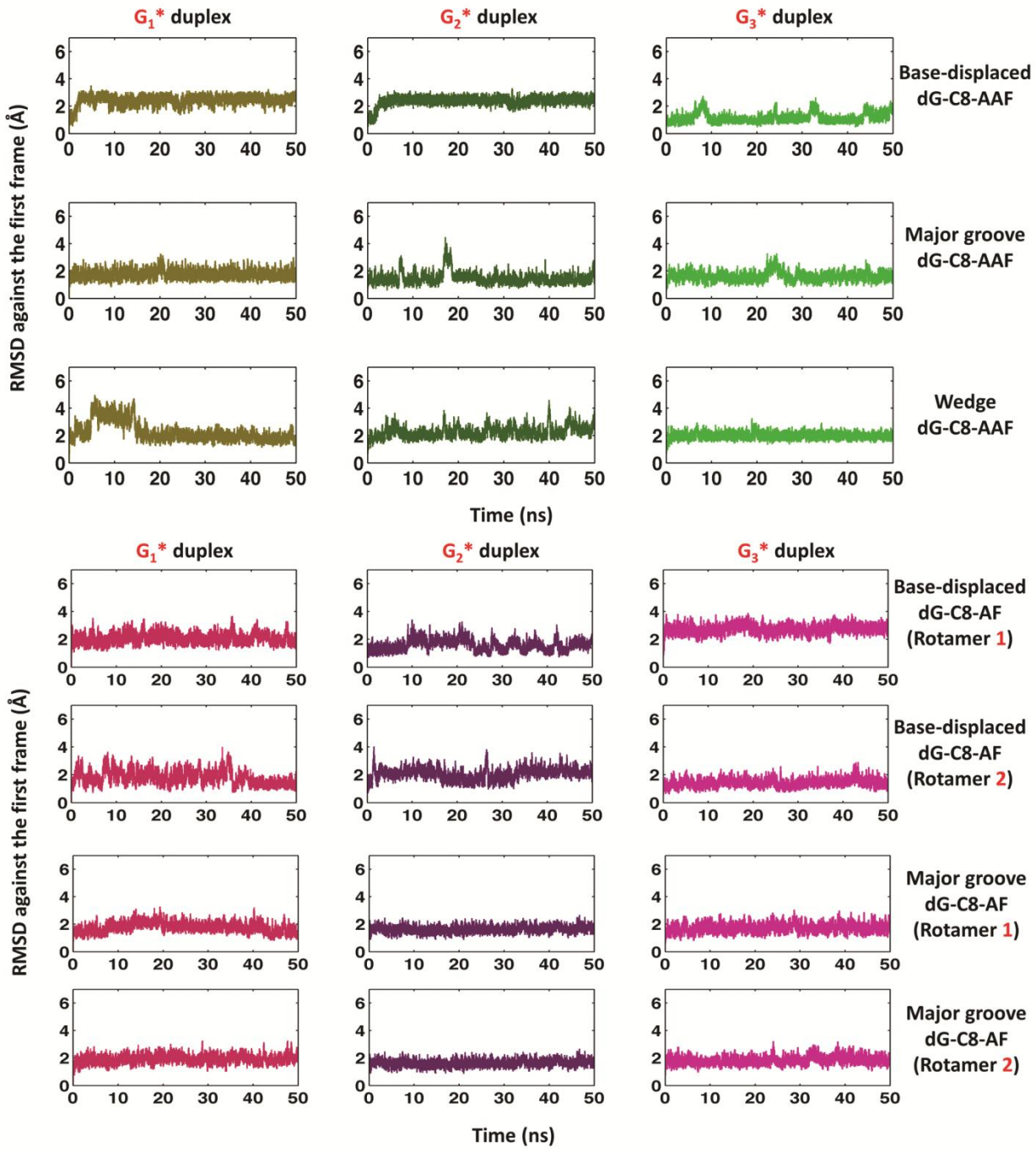


Figure S4. Time dependence of RMSDs (Å) for the central 5-mers of the dG-C8-AAF and dG-C8-AF-modified duplexes. The RMSDs were calculated relative to the first frame in each 50.0 ns trajectory, excluding two base pairs at each end.

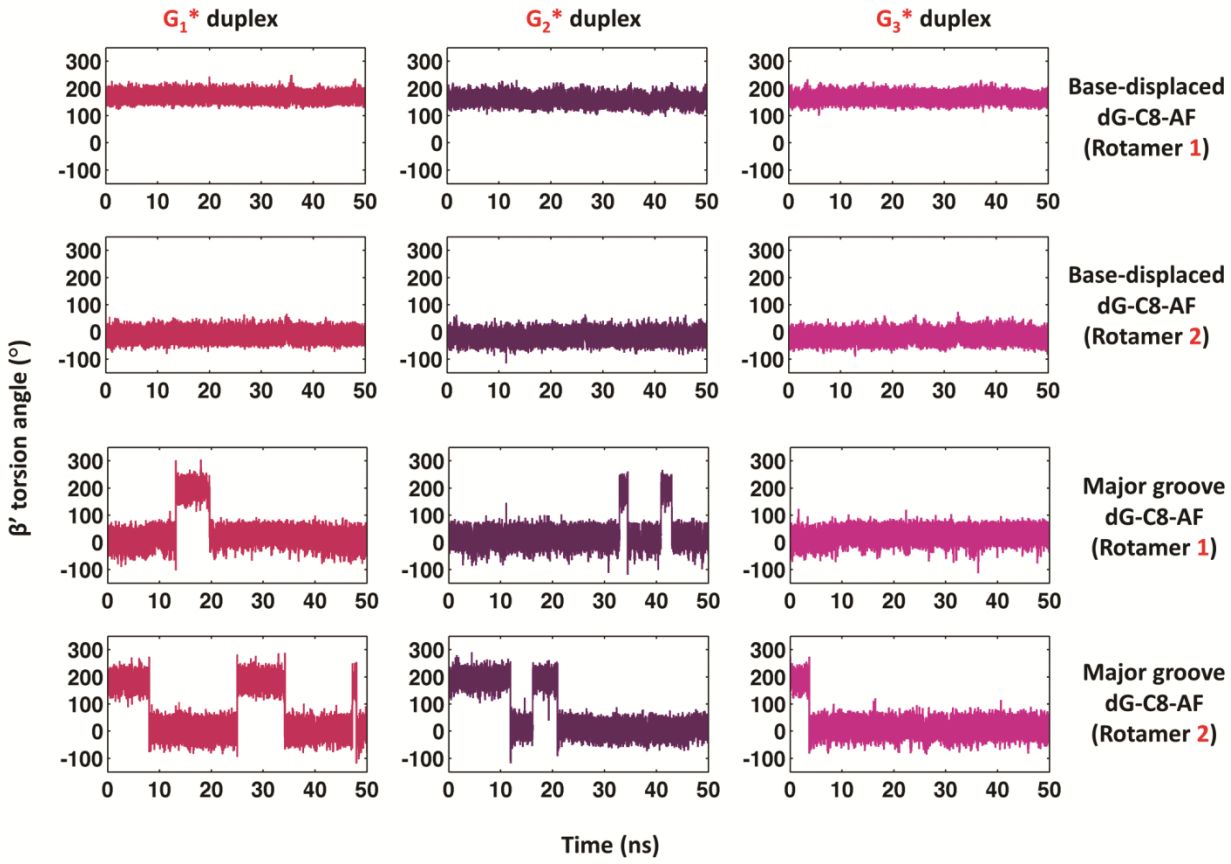


Figure S5. Time dependence of β' torsion angle of all dG-C8-AF-modified duplexes. The base-displaced conformers show a stable β' torsion angle. The major groove conformers show transitions between the two β' rotamers (rotamer 1 and rotamer 2). Note that the β' rotamer 1 of the major groove conformers is more stable in all MD simulations.

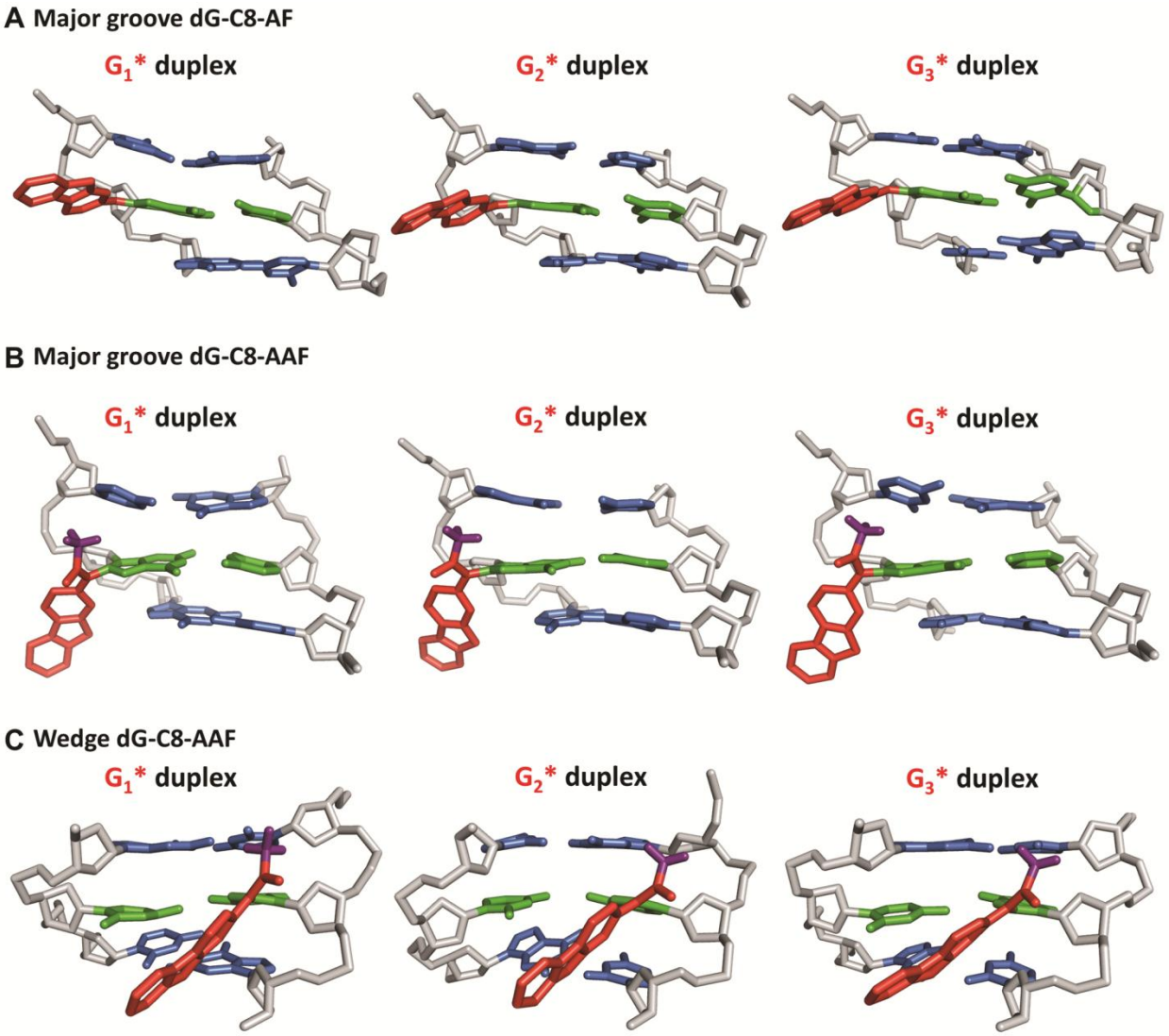


Figure S6. Most representative structures of the major groove and Wedge conformers. (A) Major groove conformers of dG-C8-AF-modified duplexes. (B) Major groove conformers of dG-C8-AAF-modified duplexes. (C) Wedge conformers of dG-C8-AAF-modified duplexes. The central trimers are represented as in Figure 2 of the main text. The major groove conformers are viewed from the major groove side, and the Wedge conformers are viewed from the minor groove side.

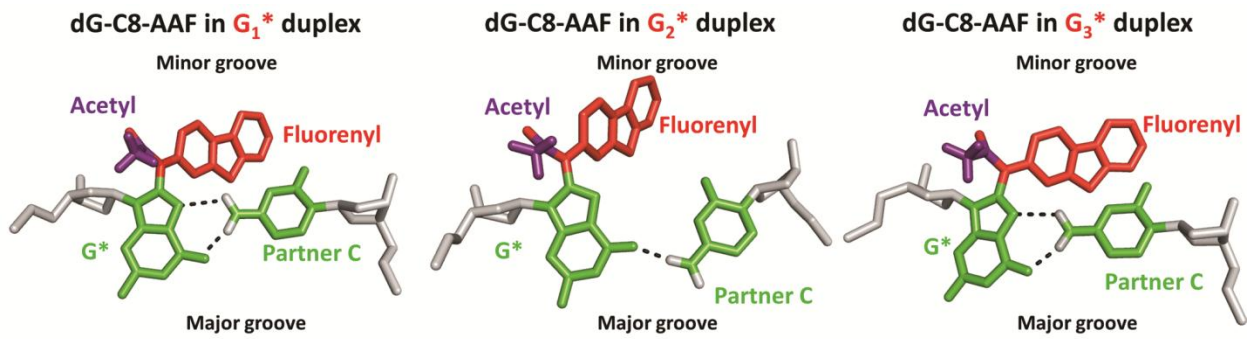


Figure S7. Non-Watson-Crick hydrogen bonds between G* and the partner C in the Wedge conformers of the dG-C8-AAF-modified duplexes. The lesion containing base pairs of the most representative structures for the Wedge conformers of the dG-C8-AAF-modified G_1*, G_2* and G_3* duplexes are viewed along the helix axis and represented as in Figure 2 of the main text, except that the hydrogen atoms of the partner C amino groups are shown in light gray sticks. The hydrogen bonds between G* and the partner C are shown in black dashed lines.

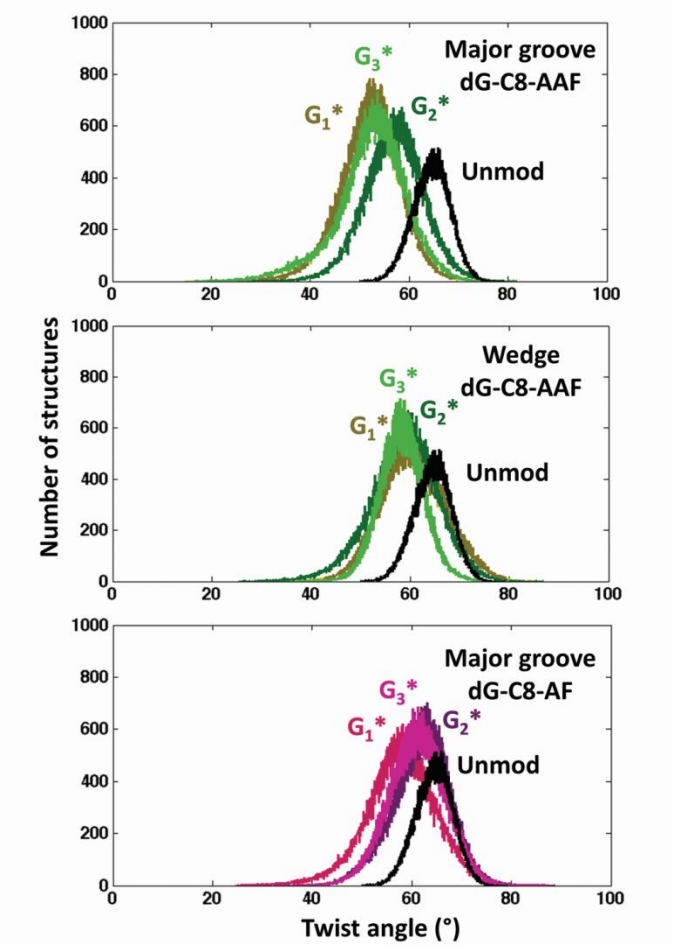


Figure S8. Population distributions of twist angles for the major groove and Wedge conformers vs. that of the unmodified control.

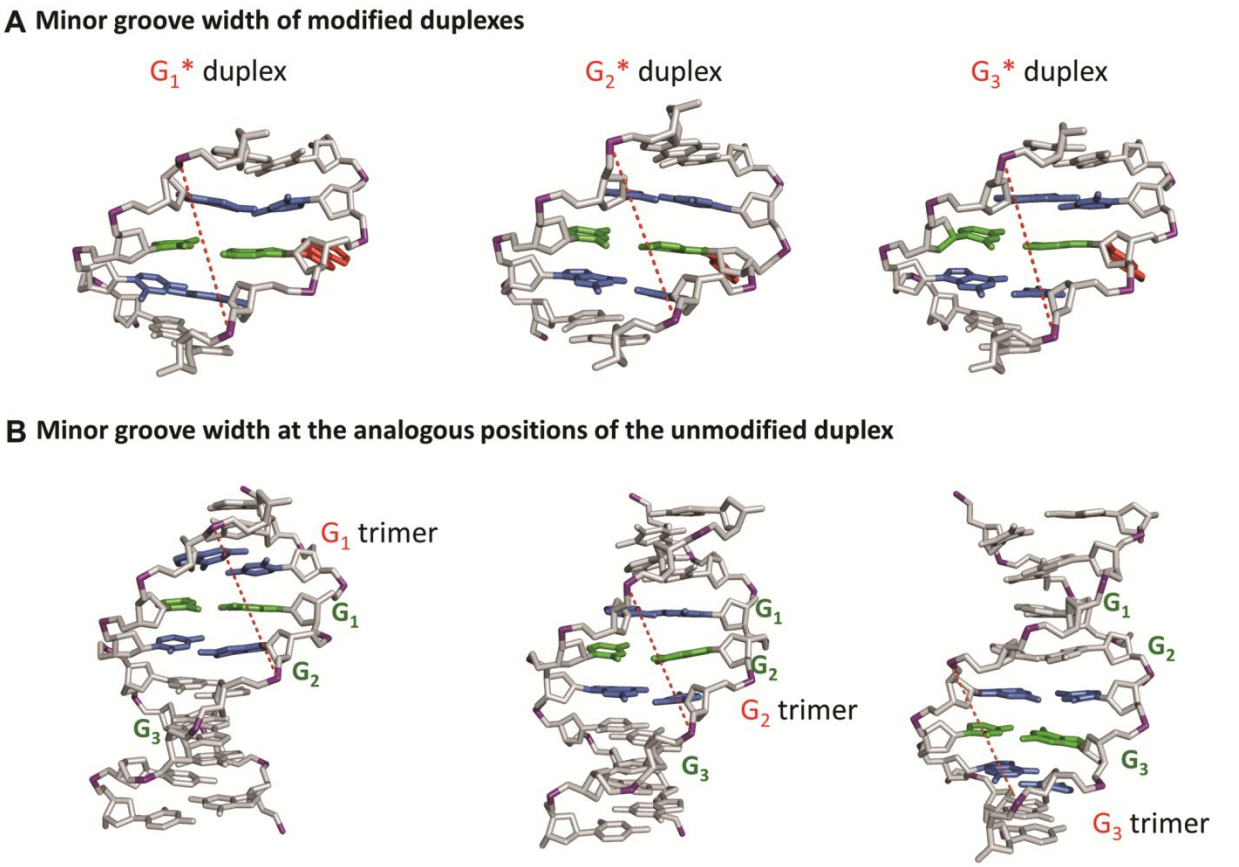


Figure S9. The minor groove widths at the lesion site of the modified duplexes and at the analogous positions of the unmodified controls. The structures are rendered as in Figure 2 of the main text and viewed from the minor groove side. The distances between phosphate atoms (in purple) are shown in red dashed lines to represent the positions for the minor groove width calculations. (A) Minor groove widths for the major groove conformers of dG-C8-AF in all three duplexes. The minor groove widths for the other modified duplexes were calculated between the same phosphate atoms. The central 5-mers are shown. (B) The minor groove widths at the analogous positions (G_1, G_2 and G_3) of the unmodified control. The central 8-mers are shown. For each analogous trimer sequence, the central G:C base pair is green and neighboring base pairs are blue.

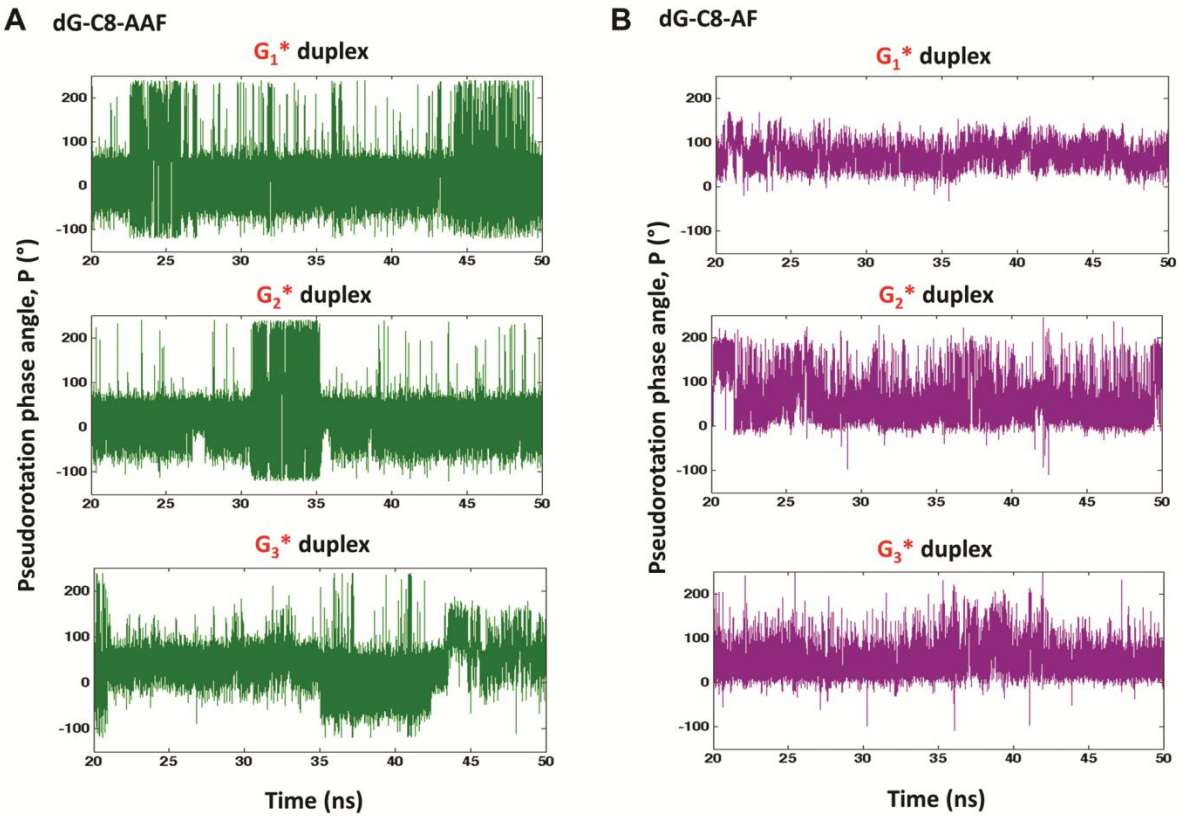


Figure S10. Dynamics of the sugar pucker of the G* in the base-displaced dG-C8-AAF and dG-C8-AF duplexes. (A) Time dependence of the pseudorotation phase angle, ***P*** (94), of the sugar ring of dG-C8-AAF in the G_1*, G_2* and G_3* duplexes. (B) Time dependence of the pseudorotation phase angle, ***P***, of the sugar ring of dG-C8-AF in the G_1*, G_2* and G_3* duplexes. The ***P*** values of lesion nucleotides for the ensembles in the 20 – 50 ns time period are calculated as described in Supplementary Methods and plotted against the time to reveal the dynamics of the sugar pucker.

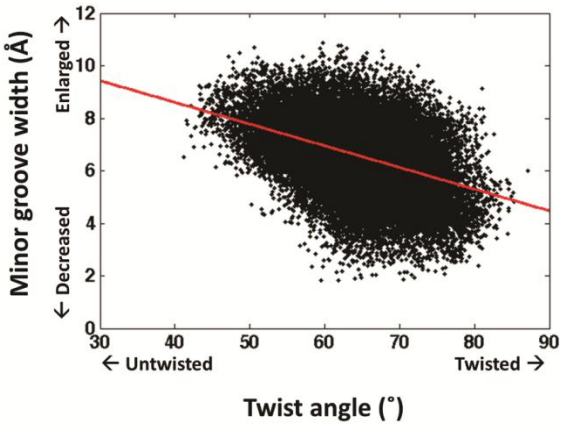


Figure S11. Correlation between duplex untwisting and minor groove width. The minor groove widths at the G_1 position of the unmodified duplex (Supplementary Figure S9B) were plotted against the twist angle for the 5'- …CG_1G_2… -3' trimer (two nucleotide steps) of the unmodified duplex. The red line shows the linear fit of all data points. The figure shows that the more untwisted the duplex, the more enlarged the minor groove is. This correlation is also exhibited at the G_2 and G_3 positions of the *Nar*I sequence in our unmodified duplex (data not shown), and is a known feature of B-DNA (97).

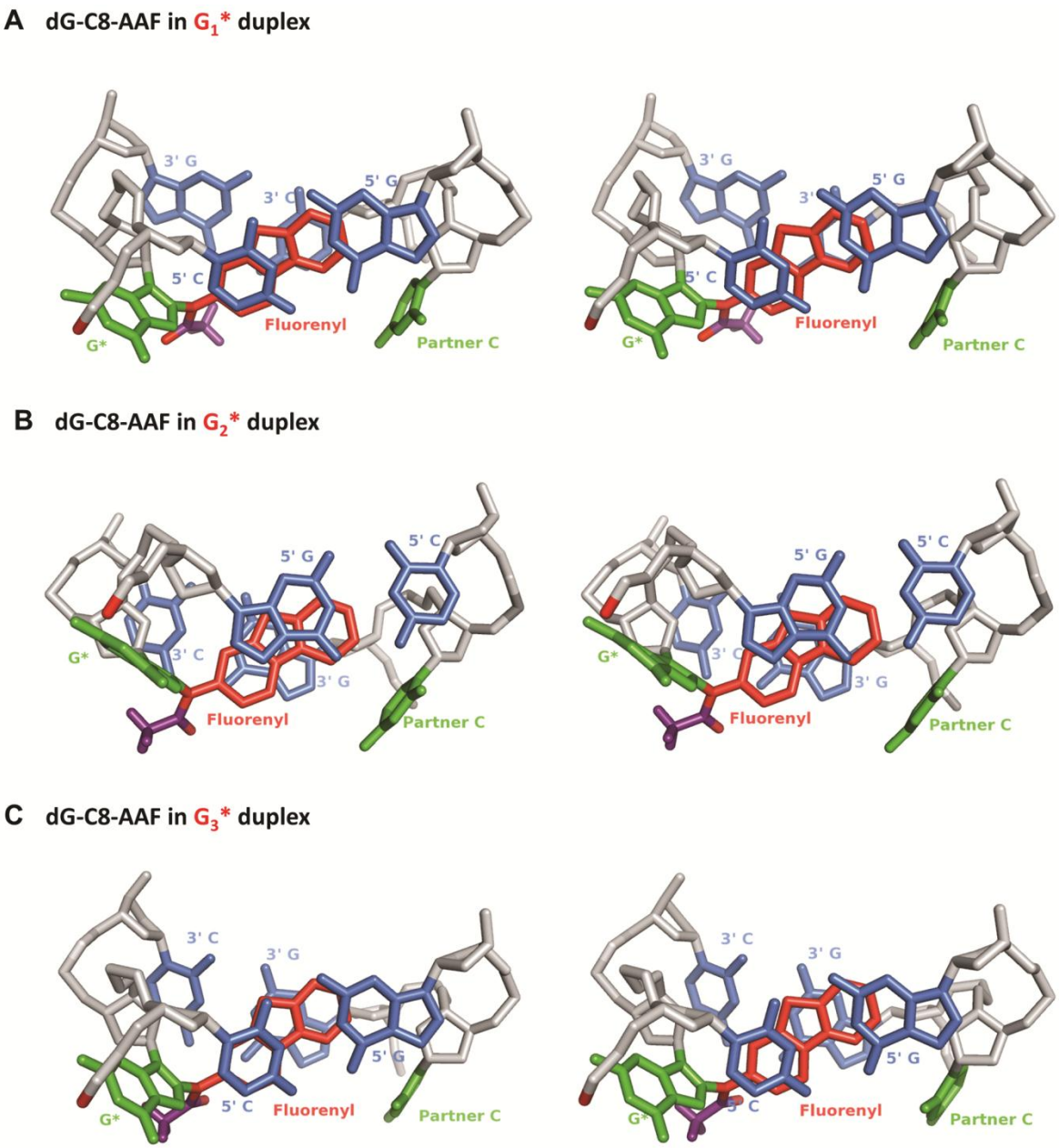


Figure S12. Stereo views along the helix axis of the central trimers of base-displaced dG-C8-AAF-modified duplexes.

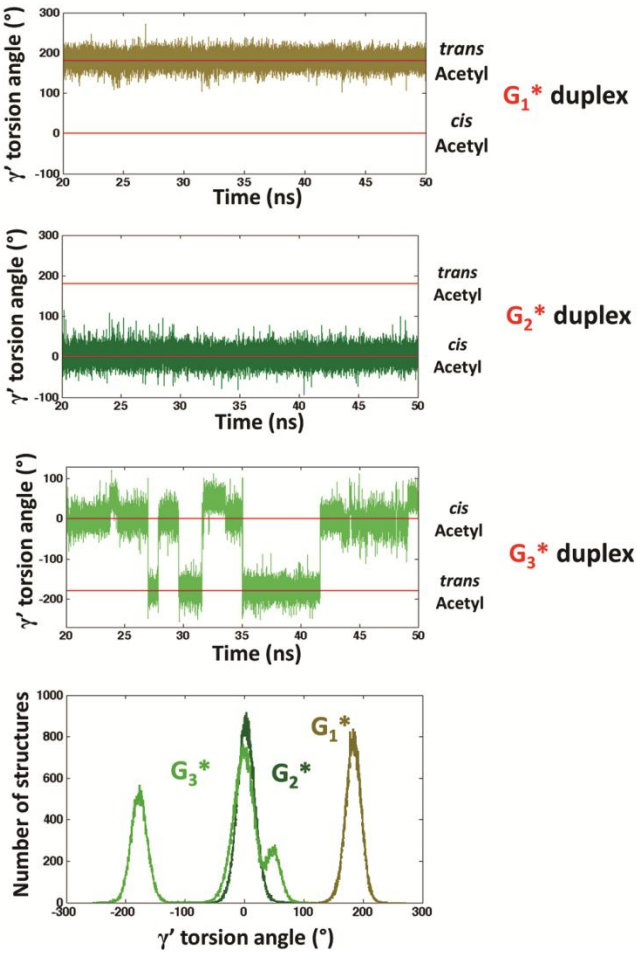
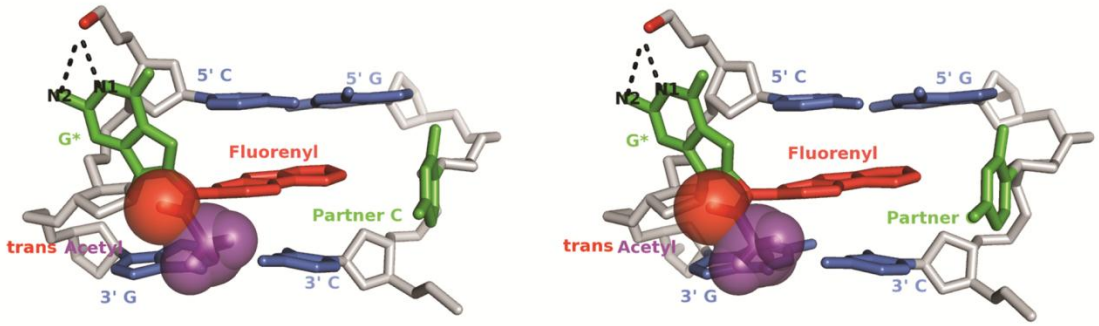
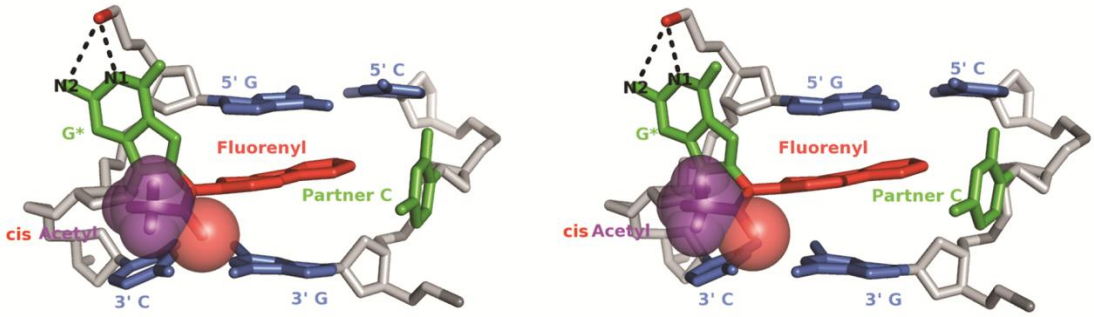


Figure S13. Dynamics and distribution of the γ' torsion angle values for the base-displaced dG-C8-AAF-modified duplexes. The γ' torsion angle values are plotted against time to show the dynamics of the acetyl group. The red lines are used to mark the *trans* (γ' = 180˚) and *cis* (γ' = 0˚) rotamers of the acetyl group. The population distribution of the γ' torsion angle values is shown at the bottom, where the cluster for the G_1* duplex is in olive, the cluster for the G_2* duplex is in dark green, and the cluster for the G_3* duplex is in light green.

A dG-C8-AAF in G_1* duplex



B dG-C8-AAF in G_2* duplex



C dG-C8-AAF in G_3* duplex

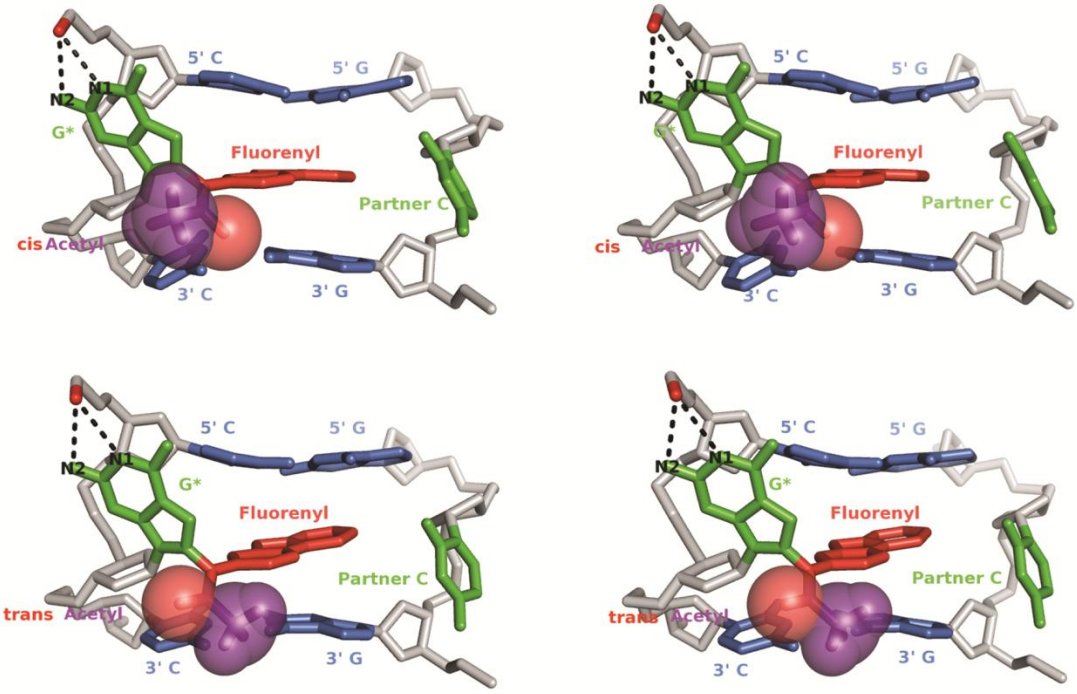


Figure S14. Steric hindrance imposed by the acetyl group in the base-displaced dG-C8-AAF duplexes. (A) Stereo view of central trimer of the most representative structure for the dG-C8-AAF-modified G_1* duplex. (B) Stereo view of the central trimer of the most representative structure for the dG-C8-AAF-modified G_2* duplex. (C) Stereo view of the central trimer of the representative structures for the *cis/trans* rotamers of the acetyl group in the dG-C8-AAF-modified G_3* duplex. The central trimers are viewed from the major groove side and represented as in Figure 7A of the main text.

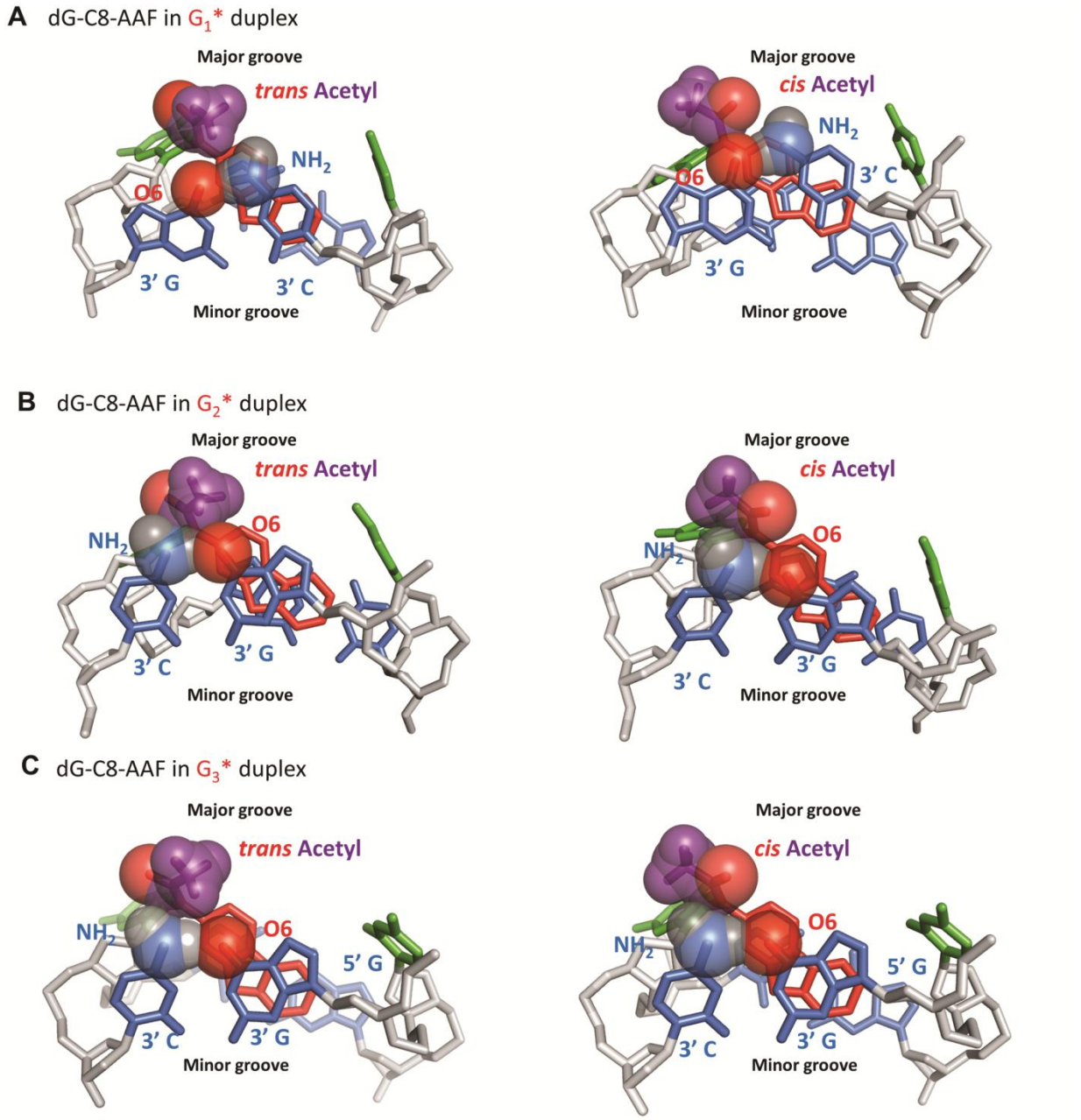


Figure S15. Steric and electrostatic origin for the *cis/trans* rotamer preference of the acetyl group in base-displaced dG-C8-AAF duplexes. The *cis/trans* rotamer preference is determined by the sequence context. The central trimers are viewed along the helix axis and rendered as in Figure 7A of the main text, except that the oxygen (O6, in red) and NH_2 group (N, in blue and H, in gray) forming the major groove side edge of the base pair on the 3' side of the lesion are also shown in CPK. (A) The *cis* rotamer is disfavored in the G_1* duplex. The *trans* rotamer is the most representative structure of the ensemble. The *cis* rotamer, which was not populated in the ensemble, was modeled by rotating the γ' torsion angle of the *trans* rotamer to the *cis* domain (0°), because in the *cis* conformation its 3' G base has its O6 atom on the major groove side near the *cis* acetyl oxygen, a repulsive interaction, while the *trans* rotamer is uncrowded with the

acetyl oxygen in the solvent. (B) The *cis* rotamer is preferred in the G_2* duplex. The *cis* rotamer is the most representative structure of the ensemble. The *trans* rotamer, which was not populated in the ensemble, was modeled by rotating the γ' torsion angle of the *cis* rotamer to the *trans* domain (180˚). The *cis* rotamer is preferred, because its 3' C has an NH_2 group on the major groove side, which provides favorable electrostatic interactions with the acetyl oxygen. However, in the disfavored *trans* domain the bulky methyl group is sterically hindered by the bulky amino group of the 3' C. (C) Both *trans* and *cis* domains are significantly populated for G_3*. The *cis* and *trans* rotamers are the most representative structure of the ensemble for each rotamer. The *cis* rotamer is preferred about 2:1 (Supplementary Figure S13); the 5' C plays the key role: in this case the fluorenyl rings are positioned to optimize stacking with the partner G of the 5' C, while in G_2* the fluorenyl rings optimize stacking with the 5' G. As a result the acetyl group is positioned away from the 3' C with its bulky amino group and hence *trans* is not disallowed.

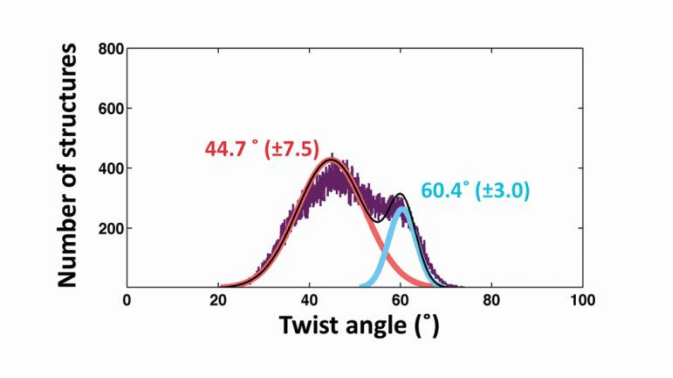


Figure S16. Deconvolution of the twist angle populations of the base-displaced dG-C8-AF G_2* duplex. The actual population distribution of the twist angle for the duplex is purple; the deconvoluted models of the two populations are red and blue respectively, with mean values and standard deviations labeled in the same colors; the curve of the sums of the deconvoluted models in black shows its close fit to the actual population. The red cluster (80%) represents the population with higher untwisting/lower twist, in which the fluorenyl rings are less protruding into the minor groove. The blue cluster (20%) represents the population with lower untwisting/higher twist, in which the fluorenyl rings are more protruding into the minor groove. Most representative structures of these two populations are shown in Figure 10C and 10D in the main text. See Supplementary Methods for deconvolution procedure and method for obtaining the most representative structure of each population.

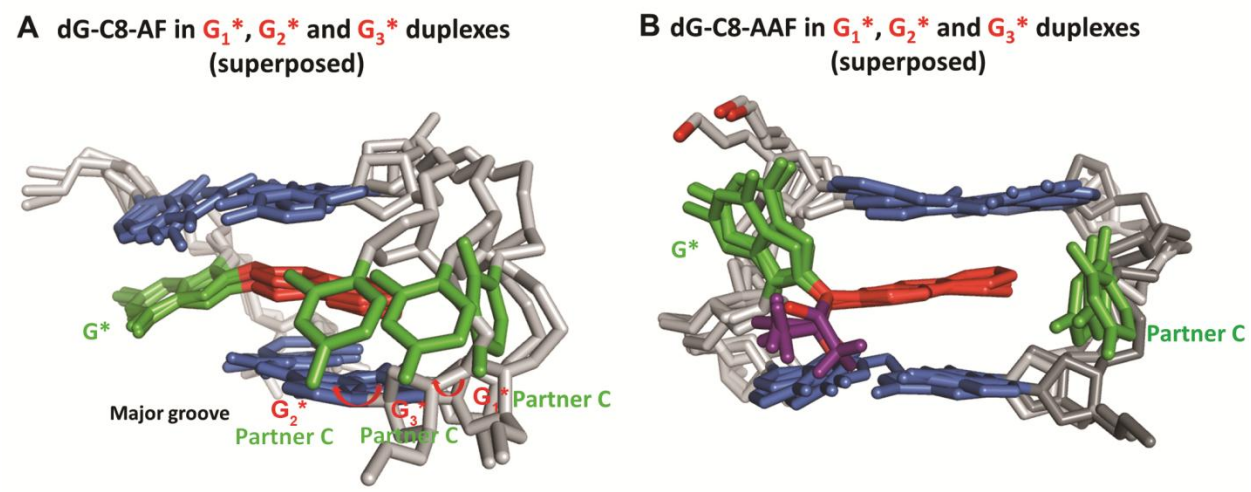
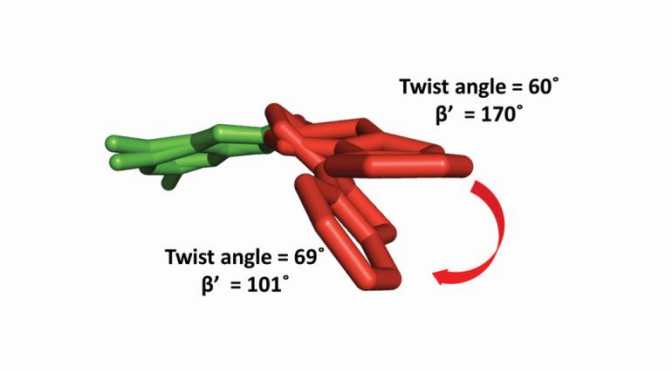


Figure S17. Positions of displaced partner C for base-displaced dG-C8-AF and dG-C8-AAF-modified duplexes. The central trimers of the most representative structures from all three ensembles for each adduct are superposed based on the lesion-containing strand. (A) The position of the displaced partner C to the modified guanine follows the untwisting order of the dG-C8-AF-modified duplexes. The partner C in the G_2* duplex is in closest contact with its 3' and 5' neighboring residues on the major groove side, because this duplex is the least untwisted (Figures 8 in the main text). The magnitudes of the van der Waals interaction energies between the displaced C and adjacent DNA residues (Supplementary Table S7) rank in the same relative order as the degree of untwisting (Figure 8 in the main text): the least untwisted duplex has the greatest van der Waals interaction energy between the displaced partner C and nearby DNA residues. (B) The similar positions of the partner C in the major grooves of the dG-C8-AAF-modified duplexes. In the dG-C8-AAF adducts, the greater untwisting caused by the acetyl group (Figure 8 in the main text), places the displaced partner C into the major groove in a similar orientation in all three sequences; this feature also manifests itself in the similar interaction energies of the displaced C bases with neighboring DNA residues (Supplementary Table S7).The central trimers are viewed from the major groove side and rendered as in Figure 2 of the main text.



Supplementary Figure S18. Partial flipping about the long axis of the fluorenyl rings for the less untwisted population of the dG-C8-AAF-modified G_2* duplex. The fluorenyl rings of this less untwisted population are protruded into the minor groove (Figure 10D in the main text and Supplementary Figure S16). In the minor groove, the fluorenyl rings rotate along the long axis. The damaged guanine of the structure showing the most rotation about the long axis in the minor groove is superposed to that of the most representative structure of the less untwisted population (Figure 10D in the main text) to show the rotation about the long axis. The fluorenyl rings (red) and guanines (green) are shown in sticks.

SUPPLEMENTARY MOVIES

Movie S1. Rocking of central trimer of dG-C8-AAF-modified G_2* duplex around helical axis. Due to de steric effect of the acetyl group, G* is displaced into the major groove, resulting in decreased stacking between the lesion and neighboring bases. Although the stacking pattern varies in different sequence contexts, the orientations of G* and its sugar ring are similar in all duplexes.

Movie S2. Rocking of central trimer of dG-C8-AF-modified G_2* duplex around helical axis. Without an acetyl group, G* is stacked with the adjacent guanine on its 5' side, resulting in extensive stacking between the lesion and neighboring bases. Although the stacking pattern varies in different sequence contexts, the G* is similarly stacked in all duplexes.

Movie S3. Interchange between the *cis/trans* γ' rotamers of the base-displaced dG-C8-AAF in the G_3* duplex.

Movie S5. The base-displaced intercalated dG-C8-AF in the G_2* duplex exhibits populations with different helical twist angles resulting from different extrusions of the fluorenyl rings into the minor groove. Partial flipping about the fluorenyl long axis is observed when these rings are extruded to the minor groove.

REFERENCES

32. Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M.P. and Cho, B.P. (2012) Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.*, **40**, 3939-3951.
35. O'Handley, S.F., Sanford, D.G., Xu, R., Lester, C.C., Hingerty, B.E., Broyde, S. and Krugh, T.R. (1993) Structural characterization of an *N*-acetyl-2-aminofluorene (AAF) modified DNA oligomer by NMR, energy minimization, and molecular dynamics. *Biochemistry*, **32**, 2481-2497.
37. Patel, D.J., Mao, B., Gu, Z., Hingerty, B.E., Gorin, A., Basu, A.K. and Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance. *Chem. Res. Toxicol.*, **11**, 391-407.
40. Mao, B., Hingerty, B.E., Broyde, S. and Patel, D.J. (1998) Solution structure of the aminofluorene [AF]-external conformer of the *anti*-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry*, **37**, 95-106.
41. Mao, B., Hingerty, B.E., Broyde, S. and Patel, D.J. (1998) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn*-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry*, **37**, 81-94.
49. Wang, L. and Broyde, S. (2006) A new *anti* conformation for *N*-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson-Crick pairing in the *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4). *Nucleic Acids Res.*, **34**, 785-795.
50. Shapiro, R., Hingerty, B.E. and Broyde, S. (1989) Minor-groove binding models for acetylaminofluorene modified DNA. *J. Biomol. Struct. Dyn.*, **7**, 493-513.

59. Case, D.A., Darden, T.A., Cheatham, T.E., III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M., *et al.* (2006), AMBER9. University of California San Francisco, San Francisco, CA.
61. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., *et al.* (2002), AMBER7. University of California San Francisco, San Francisco, CA.
62. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917-5929.
83. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197.
84. Wang, J.M., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049-1074.
85. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817-3829.
86. Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157-1174.
87. Cieplak, P., Cornell, W.D., Bayly, C. and Kollman, P.A. (1995) Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins. *J. Comput. Chem.*, **16**, 1357-1377.
88. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery Jr., J.A., Vreven, T., Kudin, K.N., Burant, J.C. *et al.* (2004). Gaussian 03, Revision B.05 ed. Gaussian, Inc., Wallingford, CT.
89. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.*, **97**, 10269-10280.
90. Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A. and Haak, J.R. (1984) Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys*, **81**, 3684-3690.
91. Flyvbjerg, H. and Petersen, H.G. (1989) Error-Estimates on Averages of Correlated Data. *J Chem Phys*, **91**, 461-466.
92. Yang, W., Bitetti-Putzer, R. and Karplus, M. (2004) Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.*, **120**, 2618-2628.
93. Saenger, W. (1984) *Principles of Nucleic Acis Structure*. Springer-Verlag, New York.
94. Altona, C. and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.*, **94**, 8205-8212.
95. Cho, B.P. and Zhou, L. (1999) Probing the conformational heterogeneity of the acetylaminofluorene-modified 2'-deoxyguanosine and DNA by ^{19}F NMR spectroscopy. *Biochemistry*, **38**, 7572-7583.

96. Evans, F.E., Miller, D.W. and Levine, R.A. (1984) Conformation and Dynamics of the 8-Substituted Deoxyguanosine 5'-Monophosphate Adduct of the Carcinogen 2-(Acetylamino)Fluorene. *J. Am. Chem. Soc.*, **106**, 396-401.
97. Johnson, R.C., Stella, S. and Heiss, J.K. (2008) Bending and Compaction of DNA by Proteins. In Rice, P. A. and Correll, C. C. (eds.), *Protein-Nucleic Acid Interactions: Structural Biology*. The Royal Society of Chemistry, Cambridge, pp. 176-220.