Supporting Information

Differences in the Access of Lesions to the Nucleotide Excision Repair Machinery in Nucleosomes

Yuqin Cai[1], Konstantin Kropachev[2], Michael A. Terzidis[3], Annalisa Masi[3], Chryssostomos Chatgilialoglu[3,4], Vladimir Shafirovich,[2] Nicholas E. Geacintov[2,*] and Suse Broyde[1,*]

Department of [1]Biology and [2]Chemistry, New York University, New York, N.Y. 10003, U.S.A, [3]Istituto per la Sintesi Organica e la Fotoreattività, Consiglio Nazionale delle Ricerche, Bologna, 40129, Italy. [4]Institute of Nanoscience and Nanotechnology, National Center for Scientific Research Demokritos, 15341 Agia, Praskevi, Athens, Greece.

Corresponding Authors: *Email: ng1@nyu.edu; *Email: broyde@nyu.edu

Table of Contents

List of Supporting Tables

List of Supporting Figures

List of Supporting Movies

Movie S1, bi5b00564_si_002.avi: Central CG*C duplex trimer of most representative structure from MD simulation of *R*-cdG [1]
Movie S2, bi5b00564_si_003.avi: Central CG*C duplex trimer of most representative structure from MD simulation of *S*-cdG [1]
Movie S3, bi5b00564_si_004.avi: Central CG*C duplex trimer of NMR solution structure of *cis*-B[*a*]P-dG (Cosman, de los Santos et al. 1993)

Movie S4, bi5b00564_si_005.avi: Best representative structure from MD simulation of *R*-cdG in the nucleosome with zoom-in on lesion.
Movie S5, bi5b00564_si_006.avi: Best representative structure from MD simulation of *S*-cdG in the nucleosome with zoom-in on lesion.
Movie S6, bi5b00564_si_007.avi: Best representative structure from MD simulation of *cis*-B[*a*]P-dG in the nucleosome with zoom-in on lesion.

Movie S7, bi5b00564_si_008.avi: Best representative structure from MD simulation of *R*-cdG in the nucleosome showing the electrostatic potential and DNA-histone interactions for the duplex 5-mer containing the centrally located lesion.
Movie S8, bi5b00564_si_009.avi: Best representative structure from MD simulation of *S*-cdG in the nucleosome showing the electrostatic potential and DNA-histone interactions for the duplex 5-mer containing the centrally located lesion.
Movie S9, bi5b00564_si_010.avi: Best representative structure from MD simulation of *cis*-B[*a*]P-dG in the nucleosome showing the electrostatic potential and DNA-histone interactions for the duplex 5-mer containing the centrally located lesion.
Movie S10, bi5b00564_si_011.avi: Best representative structure from MD simulation of the unmodified nucleosome showing the electrostatic potential and DNA-histone interactions. The duplex 5-mer that corresponds to the modified duplex 5-mers containing the centrally located lesions is shown.

1. Methods

1.1. Preparation of the initial lesion-containing and unmodified NCP structure for MD simulations

Molecular modeling of the unmodified and lesion-containing NCPs. We placed our lesions near the dyad axis in our selected nucleosome (Figure 1). We investigated the *R*-cdG, *S*-cdG and *cis*-B[*a*]P-dG lesions and an unmodified control in the nucleosome with PDB[11] ID: 2NZD[12]. The base sequence context around the lesion site (at SHL ~ 0) near the dyad (Figure 2) is 5'-… A67-G68-C69-T70-G71-A72-[A]73-T74-C75-A76-G77-C78-T79…-3' in the crystal structure (Figure S1). The numbering scheme along the modified strand of the DNA is from 1 to 145 from the 5'-end to the 3'-end. We replaced the [A] nucleotide with lesion-containing G* (G* = *R*-cdG, *S*-cdG or *cis*-B[*a*]P-dG) for the modified nucleosomes, and with G for the unmodified control**.** For the cyclopurine lesions in the nucleosome, we utilized the most representative structures in free DNA derived from our MD simulations[1] that were based on NMR solution structures [3, 4] as detailed in. [1] For the *cis*-B[*a*]P-dG adduct in the nucleosome, we used the NMR solution structure.[5] Modeling software InsightII and Discovery Studio (Accelrys Inc.) were used to build the initial models.

Force field. The Cornell et al. force field[7] with modifications[13, 14] and the parm99 parameter set [8] modified by the parmbsc0 DNA parameters[9] were employed for all simulations. Partial charges for the *R*-cdG, *S*-cdG and *cis*-B[*a*]P-dG adducts were taken from our prior publications.[1, 10]

Protonation. The protonation states for the histones were determined in our earlier publication [15] for the same nucleosome with details of the methods given. Only HIS amino acids had to be assigned and the results are: HIE: 487, 538, 746 and 923; HIP: 594, 654 and 979; HID: 627, 672, 770, 827, 872, 1012 and 1039.

1.2. MD protocols

The AMBER 11 simulation package[16] was utilized to carry out all minimization and MD simulations. The LEaP module was used to add hydrogen atoms and neutralize the system with K^+ counterions. The systems were reoriented using SIMULAID [17] to minimize the number of water molecules needed to solvate the system. A periodic rectangular box of TIP3P water[18] with 10.0 Å buffer was created around the nucleosome with the LEaP module Box sizes and numbers of waters added were shown in Table S1.

Minimization, heating, MD equilibration and production protocols. To begin with, the counterions and water molecules were minimized for 50 steps of steepest descent and 50 steps of conjugate gradient, with 50 kcal/mol restraints on the solute atoms. Then, 90 ps initial MD at 10K with 25 kcal/mol restraints on solute were performed to allow the solvent to relax. Next, the

system was heated up from 10K to 300K at constant volume over 30 ps with 10 kcal/mol restraint on the NCP. The restraints on the solute were relaxed from 10 kcal/mol (for 30 ps) to 1 kcal/mol (for 40 ps) to 0.1 kcal/mol (50 ps). Subsequently, production MD was conducted at 1 atm, 300K for 80 ns, with 1 ps coupling constant for both pressure and temperature. These simulation protocols are similar to those given in earlier publication from our group.[19]

In all MD simulations, the Particle-Mesh Ewald[20, 21] method with 9.0 Å cutoff for the non-bonded interactions was used. A 2.0 fs time step and the SHAKE algorithm[22] were applied in the MD simulations. All other parameters were default values in the AMBER 11 simulation package.[6] The stability of the MD simulation was evaluated. For each system, the root mean square deviation (RMSD) of each snapshot in the trajectory relative to its respective initial structure was plotted as a function of time and is shown in Figure S2. Three levels of the local region near the lesion were analyzed, namely within 5, 10 and 15 Å of the damaged base G73*. For all cases, the MD achieved stability fluctuating around the mean after 10.0 ns of simulation, and we employed the structural ensembles from the 10 – 80.0 ns time frame at 10 ps intervals for further structural analyses.

1.3. Structural analyses

The PTRAJ module of the AMBER 11 package[6] was employed for structural analyses. Frames were selected at 10 ps intervals from the last 70.0 ns of simulation. The DNA duplex helicoidal parameters and groove dimensions were analyzed using the CPPTRAJ program in AMBER 14.[23] For the helicoidal parameter Twist, the 3DNA program embedded in the AMBER CPPTRAJ module, skips the bulky and greatly distorting *cis*-B[*a*]P-dG lesion step and only gives a two-step, 72A:219T-74T:217A, result. Therefore used the Curves+ program for this case to obtain step by step results (Table S3). Groove dimensions were obtained by measuring pairwise phosphorus-phosphorus distance (Figure 2), less 5.8 Å to account for the van der Waals radius of the P atoms. [24]

The best representative structure, the frame from the ensemble that is the closest to all other frames, was obtained using the clustering command in the PTRAJ module of AMBER 11. The rms distance metric and K-means clustering method [25] were utilized to include all frames into one cluster and obtain the best representative structure of the ensemble. Block averages and standard deviations of block averages were computed in the same fashion as detailed previously ([1] Supporting Data). Electrostatic surfaces were computed on the PDB2PQR Server (http://nbcr-222.ucsd.edu/pdb2pqr_L8/) with APBS (Adaptive Poisson-Boltzmann Solver) calculation and displayed with PyMOL.

1.4. MM-PBSA binding energy computation

The energy difference between the modified and the unmodified nucleosomes were computed. The PB module embedded in MM-PBSA is used to solve the Poisson-Boltzmann equation. We utilized an internal dielectric constant of 1.0, external dielectric

constant of 80.0, maximum number of iterations of the linear Poisson Boltzmann equation to try of 1000, solvent probe radius of 1.4 Å, temperature of 300 K, surface tension value of 0.00542 kcal/mol/^{2} and offset to correct the value of the non-polar contribution to the solvation free energy term of 0.92. These terms do not include the vibrational entropy term which has been shown to contribute very little to free energy differences with the AMBER force field [26] and is the most uncertain to compute.[27] We utilized the snapshots from the simulation window of 50-80 ns at 10 ps intervals. We computed the binding energy ΔG of the lesion-containing duplex 5-mer, as well as the corresponding unmodified duplex 5-mer, with the histones within a range of 20 Å from any atom of the 5-mer, namely H3 and H2A. For these MM-PBSA binding energy calculations, Ligand is the 5-mer duplex containing the lesion at the center or the corresponding unmodified duplex 5-mer; Receptor is histones within a range of 20 Å from any atom of the 5-mer (namely, H3 and H2A); Complex is Ligand and Receptor; and Delta, the binding energy, which is ΔG = G(Complex) – G(Ligand) - G(Receptor). The impact of each lesion on the binding energy is ΔΔG, which is ΔG(lesion) – ΔG(unmodified). The higher the ΔΔG for a lesion, the weaker the local DNA-histone interactions.

Table S 1. Box sizes and numbers of waters added to the MD simulation starting models.

	R-cdG	*S-cdG*	*cis-B[a]P-dG*	*Unmod-dG*
Box size, $Å^3$	*137 x 134 x 97*	*137 x 134 x 97*	*136 x 142 x 94*	*136 x 142 x 94*
Number of waters added	*40675*	*40674*	*41837*	*40674*

Table S 2. Ensemble average minor groove widths (Å) for the *R*-cdG, *S*-cdG, *cis*-B[*a*]P-dG and the corresponding unmodified control. Standard deviations of the block averages are given.

	R-cdG		*S-cdG*		*Unmod-dG*		*cis-B[a]P-dG*	
	Mean	*block_std*	*Mean*	*block_std*	*Mean*	*block_std*	*Mean*	*block_std*
P70-P225	*2.9*	*0.2*	*3.8*	*0.7*	*2.8*	*0.1*	*2.9*	*0.1*
P71-P224	*5.1*	*0.4*	*4.6*	*0.7*	*5*	*0.8*	*5.2*	*0.3*
P72-P223	*5.5*	*0.6*	*7.8*	*0.6*	*5.5*	*0.4*	*6.9*	*0.5*
P73-P222	*7.8*	*0.7*	*8.3*	*0.3*	*7.5*	*0.3*	*10.2*	*0.6*
P74-P221	*8.7*	*0.9*	*8.7*	*0.4*	*6*	*0.4*	*10.1*	*0.2*
P75-P220	*7.6*	*0.7*	*8.2*	*0.3*	*5.8*	*0.5*	*5.1*	*0.2*
P76-P219	*6.3*	*0.7*	*7.6*	*0.3*	*6.5*	*0.6*	*6.7*	*0.5*
P77-P218	*6.4*	*0.9*	*6.3*	*0.6*	*6.8*	*0.4*	*6.9*	*0.4*

Note that 5.8 Å was subtracted to account for the van der Waals radius of the P atoms. [24]

Table S 3. Ensemble average helicoidal Twist (°) for the *R*-cdG, *S*-cdG, *cis*-B[*a*]P-dG and the corresponding unmodified control.

Twist (°)	*R-cdG (3DNA)*		*S-cdG (3DNA)*		*Unmod-dG (3DNA)*	
Step	*mean*	*block_std*	*mean*	*block_std*	*mean*	*block_std*
70T:221A-71G:220C	*32.6*	*3.9*	*33.1*	*2.4*	*31.9*	*2.2*
71G:220C-72A:219T	*24.7*	*6.7*	*30.2*	*2.1*	*34.5*	*2.4*
72A:219T-73G:C218	*53.5*	*7.6*	*45.3*	*0.4*	*36.8*	*2.2*
73G:C218-74T:217A	*25.8*	*2.1*	*24.5*	*0.5*	*33.3*	*1.1*
74T:217A-75C:216G	*37.2*	*3.7*	*40.2*	*1.2*	*38.3*	*1.4*
75C:216G-76A:215T	*34.9*	*3.4*	*33*	*3.1*	*33.4*	*2.4*

Twist (°)					
cis-B[a]P-dG (Curves+)			*cis-B[a]P-dG (3DNA)*		
Step	*mean*	*block_std*	*Step*	*mean*	*block_std*
70T:221A-71G:220C	*41.3*	*2.2*	*70T:221A-71G:220C*	*40.5*	*2.7*
71G:220C-72A:219T	*30*	*2.7*	*71G:220C-72A:219T*	*30.9*	*2.6*
72A:219T-73G:C218	*7.9*	*2.1*	*72A:219T-74T:217A*	*63.2*	*0.9*
73G:C218-74T:217A	*54.9*	*2.1*			
74T:217A-75C:216G	*35.6*	*2.9*	*74T:217A-75C:216G*	*36.6*	*2.9*
75C:216G-76A:215T	*32.8*	*4.1*	*75C:216G-76A:215T*	*31.8*	*4.1*

Note that the 3DNA program, embedded in the AMBER CPPTRAJ module, skips the lesion step and gives a two-step, 72A:219T-74T:217A, result. On the other hand, the Curves+ program gives step by step results for lesion-containing DNA. The summed Twist value for the two steps, 72A:219T-73G:C218 and 73G:C218-74T:217A, from the Curves+ program is 7.9 (°) + 54.9 (°)= 62.8 (°), roughly the same as the two-step result from the 3DNA program, which is 63.2(°).

Table S 4. MM-PBSA binding energies ΔG (kcal/mol) for the *R*-cdG, *S*-cdG, *cis*-B[*a*]P-dG and the corresponding unmodified control. [1]

	R-cdG	*S-cdG*	*cis-B[a]P-dG*	*Unmod-dG*
40-50 ns	*-25.7 (0.1)*	*-26.5 (0.1)*	*-27 (0.1)*	*-26.8 (0.1)*
50-60 ns	*-27.1 (0.1)*	*-26.9 (0.1)*	*-25.2 (0.1)*	*-26.4 (0.1)*
60-70 ns	*-26.9 (0.1)*	*-27.3 (0.1)*	*-25.8 (0.1)*	*-27.3 (0.1)*
70-80 ns	*-26.8 (0.1)*	*-26.4 (0.1)*	*-26 (0.1)*	*-27.3 (0.1)*
60-80 ns	*-26.9 (0.1)*	*-26.9 (0.1)*	*-25.9 (0.1)*	*-27.3 (0.1)*
50-80 ns	*-26.9 (0.1)*	*-26.9 (0.1)*	*-25.7 (0.1)*	*-27.0 (0.1)*

[1] The 50-80 ns window for the energy calculation was selected.

Table S 5. Energy components for MM-PBSA binding energies ΔG (kcal/mol) for the *R*-cdG, *S*-cdG, *cis*-B[*a*]P-dG and the corresponding unmodified control for the 50 to 80 ns simulation window. [2]

#	COMPLEX (R-cdG)		
#	MEAN	STD	STD. ERR. of MEAN
ELE	-6781.23	92.93	1.70
VDW	-648.62	21.03	0.38
INT	4869.32	43.95	0.80
GAS	-2560.54	101.13	1.85
PBSUR	104.81	0.88	0.02
PBCAL	-5136.04	84.01	1.53
PBSOL	-5031.23	83.79	1.53
PBELE	-11917.28	25.38	0.46
PBTOT	-7591.77	44.67	0.82
#	RECEPTOR		
#	MEAN	STD	STD. ERR. of MEAN
ELE	-4124.55	78.87	1.44
VDW	-580.09	20.19	0.37
INT	4408.80	42.06	0.77
GAS	-295.83	85.10	1.55
PBSUR	94.22	0.86	0.02
PBCAL	-5354.01	68.60	1.25
PBSOL	-5259.80	68.56	1.25
PBELE	-9478.56	24.24	0.44
PBTOT	-5555.63	42.03	0.77
#	LIGAND		
#	MEAN	STD	STD. ERR. of MEAN
ELE	-494.81	19.48	0.36
VDW	-59.31	7.09	0.13
INT	460.52	13.58	0.25
GAS	-93.61	21.58	0.39
PBSUR	13.60	0.23	0.00
PBCAL	-1929.21	16.25	0.30
PBSOL	-1915.60	16.28	0.30
PBELE	-2424.02	8.44	0.15
PBTOT	-2009.21	13.34	0.24
#	DELTA		
#	MEAN	STD	STD. ERR. of MEAN
ELE	-2161.87	49.83	0.91
VDW	-9.22	1.81	0.03
INT	0.00	0.00	0.00
GAS	-2171.09	50.19	0.92
PBSUR	-3.01	0.26	0.00
PBCAL	2147.18	48.79	0.89
PBSOL	2144.17	48.60	0.89
PBELE	-14.69	5.43	0.10
PBTOT	-26.92	4.61	0.08

[2] Ligand is 5-mer duplex containing the lesion at the center or the corresponding unmodified duplex 5-mer; receptor is histones within a range of 20 Å from any atom of the 5-mer (namely, H3 and H2A); complex is ligand + receptor; delta is ΔG = G(Complex) – G(Ligand) - G(Receptor).

```
#                 COMPLEX   (S-cdG)
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -6871.04      78.85       1.44
VDW             -648.34      21.17       0.39
INT             4873.77      42.07       0.77
GAS            -2645.61      86.13       1.57
PBSUR            102.79       0.83       0.02
PBCAL          -5082.06      71.88       1.31
PBSOL          -4979.26      71.63       1.31
PBELE         -11953.10      26.93       0.49
PBTOT          -7624.88      42.66       0.78
#          ======================
#                   RECEPTOR
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -4165.15      75.48       1.38
VDW             -574.30      20.19       0.37
INT             4414.42      40.02       0.73
GAS             -325.02      82.17       1.50
PBSUR             92.54       0.78       0.01
PBCAL          -5307.35      67.55       1.23
PBSOL          -5214.81      67.42       1.23
PBELE          -9472.50      25.26       0.46
PBTOT          -5539.84      40.70       0.74
#          ======================
#                   LIGAND
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE             -557.78      16.94       0.31
VDW              -63.77       6.20       0.11
INT              459.34      13.20       0.24
GAS             -162.20      19.79       0.36
PBSUR             13.44       0.10       0.00
PBCAL          -1909.38      14.90       0.27
PBSOL          -1895.95      14.94       0.27
PBELE          -2467.16       7.89       0.14
PBTOT          -2058.15      12.46       0.23
#          ======================
#                   DELTA
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -2148.12      45.56       0.83
VDW              -10.26       1.98       0.04
INT                0.00       0.00       0.00
GAS            -2158.39      45.92       0.84
PBSUR             -3.18       0.22       0.00
PBCAL           2134.68      44.06       0.80
PBSOL           2131.49      43.91       0.80
PBELE            -13.45       5.36       0.10
PBTOT            -26.89       4.64       0.08
```

```
#                  COMPLEX   (unmod-dG)
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -6869.27      73.18       1.34
VDW             -643.18      20.18       0.37
INT             4857.76      43.31       0.79
GAS            -2654.69      82.02       1.50
PBSUR            105.98       0.85       0.02
PBCAL          -5024.85      66.96       1.22
PBSOL          -4918.87      66.94       1.22
PBELE         -11894.12      24.69       0.45
PBTOT          -7573.56      44.81       0.82
#          ======================
#                   RECEPTOR
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -4246.33      70.28       1.28
VDW             -562.47      18.99       0.35
INT             4397.92      40.79       0.74
GAS             -410.88      78.61       1.44
PBSUR             95.64       0.79       0.01
PBCAL          -5234.30      64.46       1.18
PBSOL          -5138.66      64.63       1.18
PBELE          -9480.63      22.76       0.42
PBTOT          -5549.54      42.38       0.77
#          ======================
#                   LIGAND
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE             -504.18      16.40       0.30
VDW              -70.89       6.23       0.11
INT              459.84      13.54       0.25
GAS             -115.23      18.63       0.34
PBSUR             13.50       0.11       0.00
PBCAL          -1895.31      13.55       0.25
PBSOL          -1881.81      13.60       0.25
PBELE          -2399.49       7.75       0.14
PBTOT          -1997.04      12.88       0.24
#          ======================
#                   DELTA
#          ----------------------
#                  MEAN        STD      STD. ERR. of MEAN
ELE            -2118.77      47.71       0.87
VDW               -9.82       1.92       0.04
INT                0.00       0.00       0.00
GAS            -2128.58      47.89       0.87
PBSUR             -3.16       0.23       0.00
PBCAL           2104.76      45.84       0.84
PBSOL           2101.60      45.68       0.83
PBELE            -14.00       5.78       0.11
PBTOT            -26.98       5.03       0.09
```

```
#                 COMPLEX   (cis-B[a]P-dG)
#          ----------------------
#                  MEAN         STD      STD. ERR. of MEAN
ELE            -6713.58       75.05        1.37
VDW             -655.49       20.67        0.38
INT             4692.00       43.35        0.79
GAS            -2677.07       83.15        1.52
PBSUR            104.82        0.83        0.02
PBCAL          -5044.72       68.53        1.25
PBSOL          -4939.90       68.36        1.25
PBELE         -11758.30       25.70        0.47
PBTOT          -7616.97       44.58        0.81
#          ======================
#                   RECEPTOR
#          ----------------------
#                  MEAN         STD      STD. ERR. of MEAN
ELE            -4206.46       67.00        1.22
VDW             -577.61       19.30        0.35
INT             4183.63       41.05        0.75
GAS             -600.44       75.63        1.38
PBSUR             94.26        0.76        0.01
PBCAL          -5254.51       59.85        1.09
PBSOL          -5160.25       59.83        1.09
PBELE          -9460.97       23.61        0.43
PBTOT          -5760.69       42.68        0.78
#          ======================
#                   LIGAND
#          ----------------------
#                  MEAN         STD      STD. ERR. of MEAN
ELE             -382.08       23.34        0.43
VDW              -67.94        6.55        0.12
INT              508.37       14.30        0.26
GAS               58.36       25.27        0.46
PBSUR             13.70        0.16        0.00
PBCAL          -1902.66       20.59        0.38
PBSOL          -1888.96       20.63        0.38
PBELE          -2284.74        8.20        0.15
PBTOT          -1830.60       13.86        0.25
#          ======================
#                   DELTA
#          ----------------------
#                  MEAN         STD      STD. ERR. of MEAN
ELE            -2125.05       53.47        0.98
VDW               -9.94        1.81        0.03
INT                0.00        0.00        0.00
GAS            -2134.99       53.81        0.98
PBSUR             -3.14        0.25        0.00
PBCAL           2112.46       51.47        0.94
PBSOL           2109.32       51.29        0.94
PBELE            -12.59        5.45        0.10
PBTOT            -25.67        4.84        0.09
```

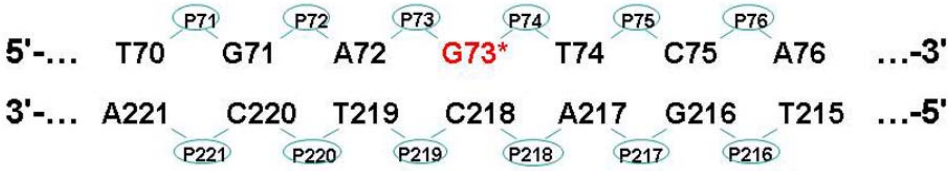

Figure S 1. Sequence context around the lesion, with G* designating the lesion modified base.

In the crystal structure with PDB ID 2NZD[12], Chain I, from 5'-side to 3'-side, P(-2), P(-1),P(0),P(1),P(2),P(3) correspond to P71, P72, P73, P74, P75, P76, respectively. Chain J, from 5'-side to 3'-side, P(-2), P(-1),P(0),P(1),P(2),P(3), correspond to P216, P217, P218, P219, P220, P221, respectively. Note that in the crystal, our G73:C218 pair was an A:T pair. We remodeled the A:T pair to create the lesions and the unmodified control.

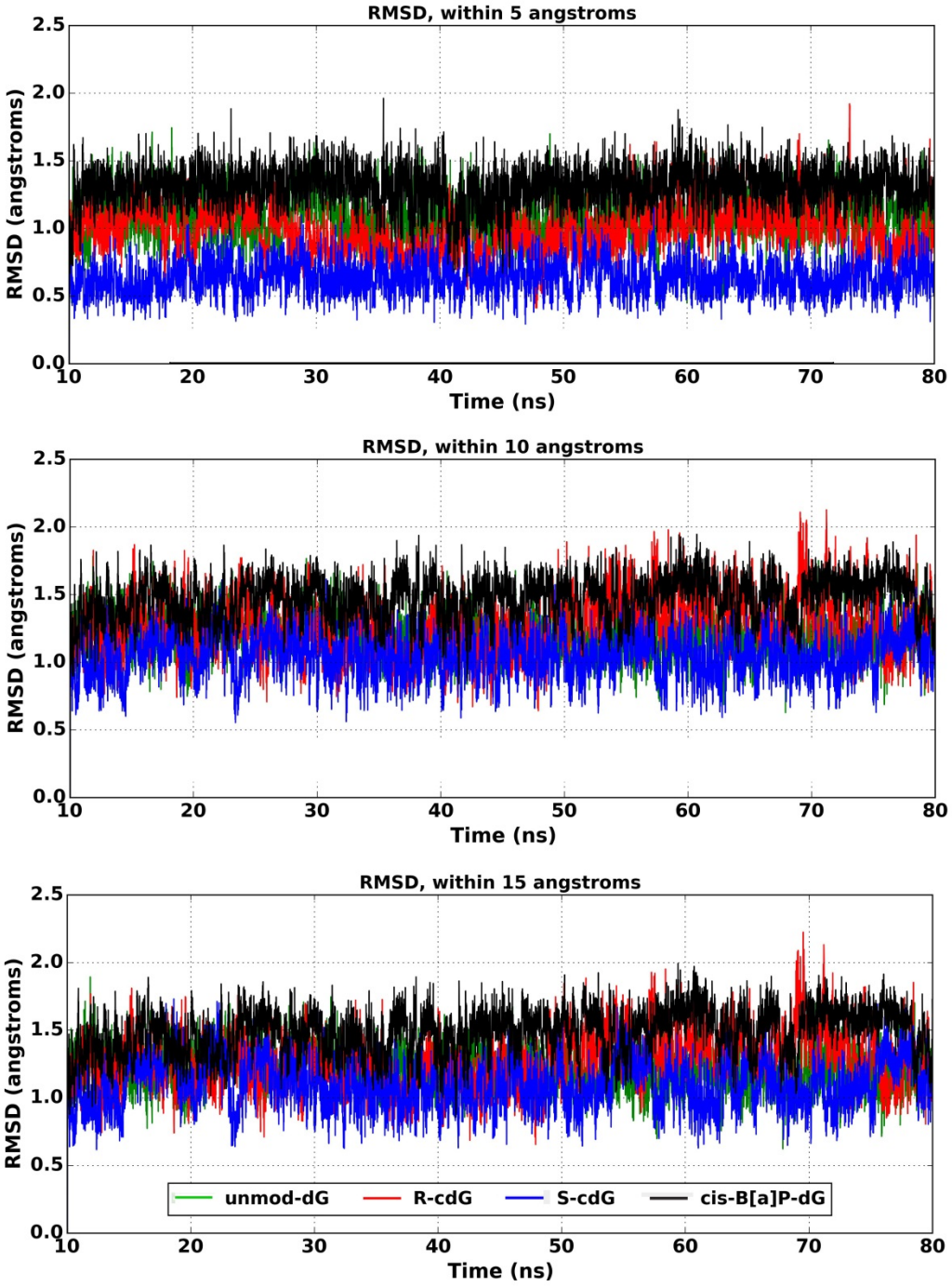

Figure S 2. Time-dependence of RMSDs for all models over the 80.0 ns MD simulations.

The RMSDs were calculated relative to their respective initial structures for regions within 5, 10 and 15 Å of the *R*-cdG. The same residues are selected for computing for all other cases.

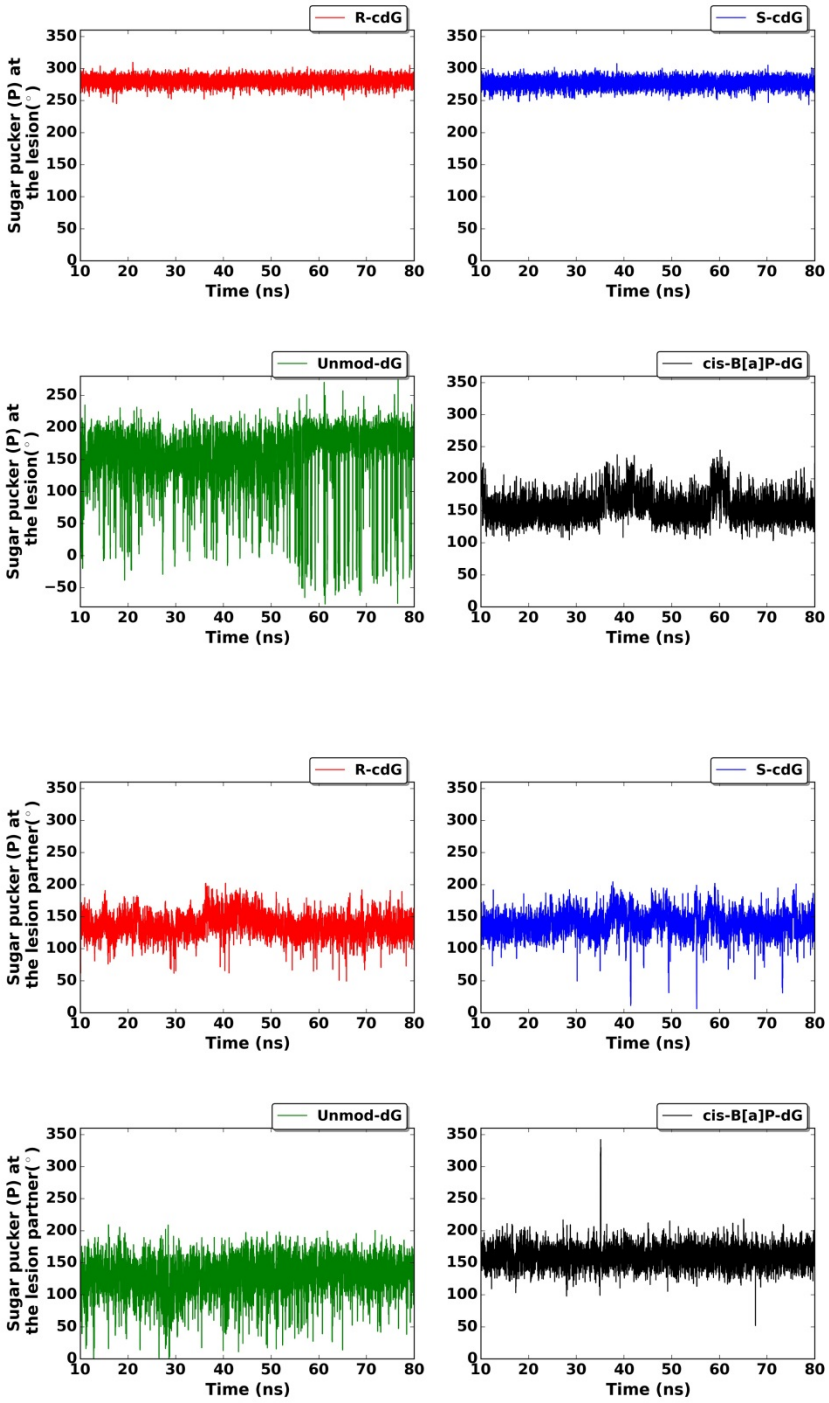

Figure S 3. Sugar pucker at the lesion and at the lesion partner base.

At the lesion, the means and standard deviations of block averages are 281.0 (0.9), 277.8 (0.9), 162.5 (18.5), 155.2 (5.8) for the *R*-cdG, *S*-cdG, unmodified control and *cis*-B[*a*]P-dG, respectively. At the lesion partner base, the means and standard deviations of block averages are 136.1 (7.3), 138.6 (5.6), 127.2 (6.9), 159.5(2.5) for the *R*-cdG, *S*-cdG, unmodified control and *cis*-B[*a*]P-dG, respectively.

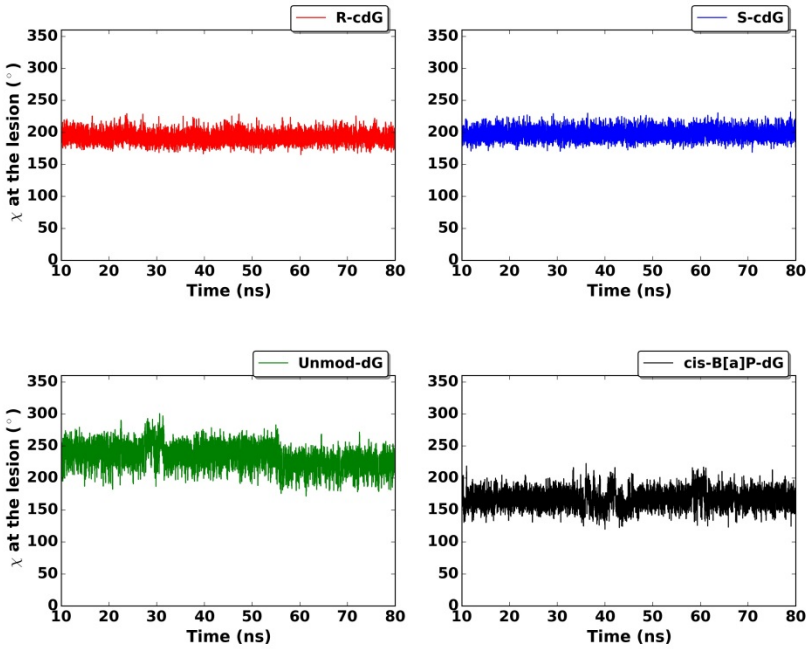

Figure S 4. Glycosidic torsion χ at the lesion.

The means and standard deviation of block averages are 192.3 (1.7), 197.8 (1.0), 233.1 (10.0), 167.8 (3.0) for the *R*-cdG, *S*-cdG, unmodified control and *cis*-B[*a*]P-dG, respectively.

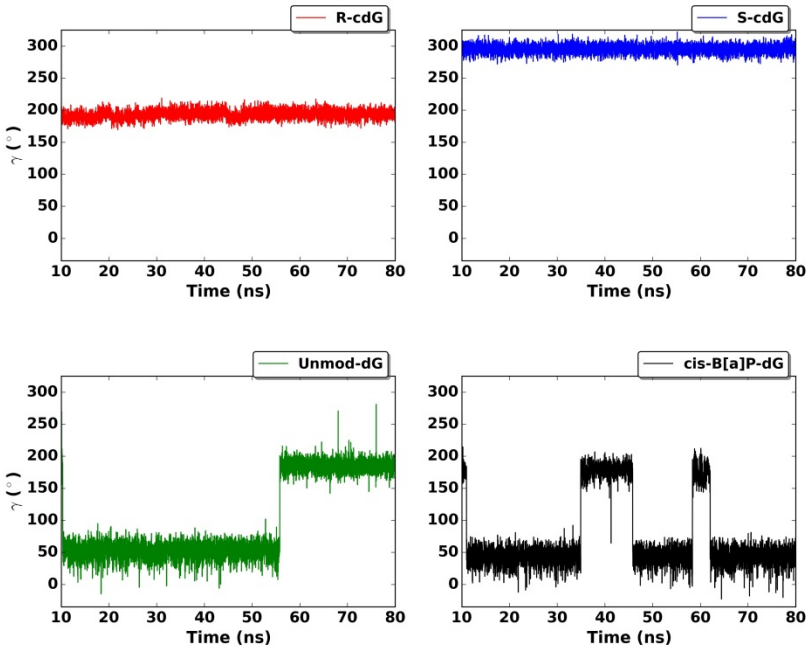

Figure S 5. Time dependence of the backbone torsion angle γ at the lesion.

The means and standard deviation of block averages are 193.4 (2.0) °, 295.5 (0.8)°, 52.1 (1.8) ° (cluster 1) and 185.6 (1.7)° (cluster 2) ,and 44.2 (1.8) ° (cluster 1) and 181.0 (3.1) ° (cluster 2) for the *R*-cdG, *S*-cdG, unmodified control and *cis*-B[*a*]P-dG, respectively.

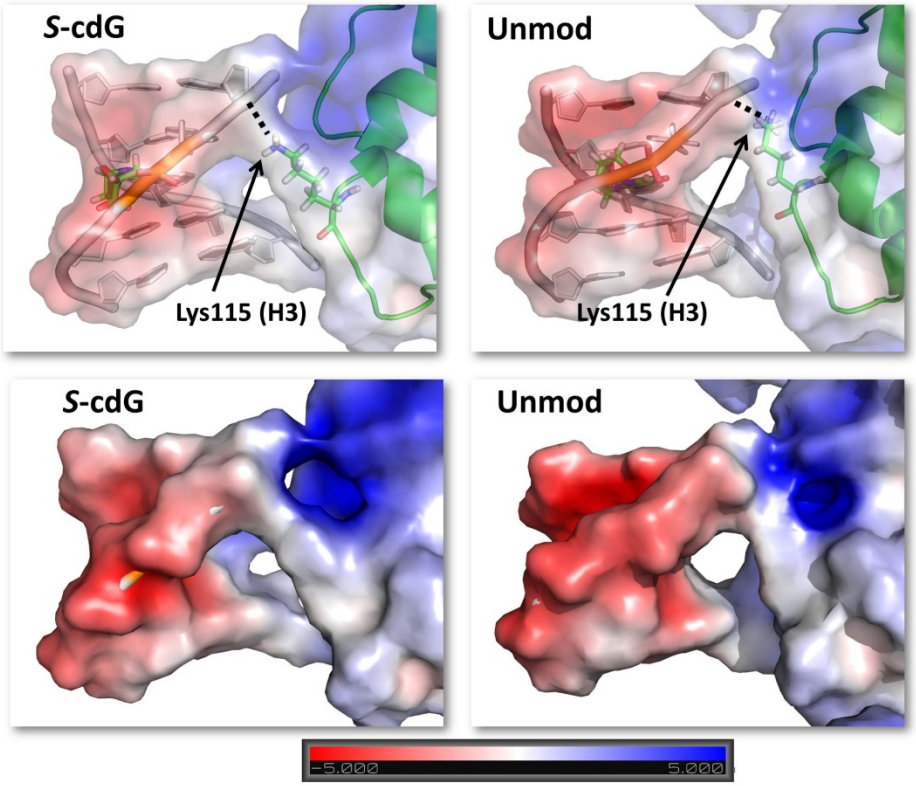

Figure S 6. The most representative structures from the MD simulation for *S*-cdG and the corresponding unmodified sequence showing DNA-histone interactions and electrostatic surface.

Shown are the structures for the duplex 5-mer containing the centrally located lesion or corresponding unmodified duplex 5-mer. The hydrogen bond between Lys 115 of histone H3 and the DNA backbone is present in both *S*-cdG and the corresponding unmodified sequence. The upper panel is rendered in half transparency surface to reveal the molecular interactions while the lower panel without transparency emphasizes the electrostatic surface difference. See also Supporting Movies S8 and S10.

References

1. Kropachev, K., Ding, S., Terzidis, M. A., Masi, A., Liu, Z., Cai, Y., Kolbanovskiy, M., Chatgilialoglu, C., Broyde, S., Geacintov, N. E., and Shafirovich, V. (2014) Structural basis for the recognition of diastereomeric 5',8-cyclo-2'-deoxypurine lesions by the human nucleotide excision repair system, *Nucleic Acids Res. 42*, 5020-5032.
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution, *Nature 389*, 251-260.
3. Huang, H., Das, R. S., Basu, A. K., and Stone, M. P. (2011) Structure of (5'S)-8,5'-cyclo-2'-deoxyguanosine in DNA, *J. Am. Chem. Soc. 133*, 20357-20368.
4. Zaliznyak, T., Lukin, M., and de los Santos, C. (2012) Structure and stability of duplex DNA containing (5'S)-5',8-cyclo-2'-deoxyadenosine: an oxidatively generated lesion repaired by NER, *Chem. Res. Toxicol. 25*, 2103-2111.
5. Cosman, M., de los Santos, C., Fiala, R., Hingerty, B. E., Ibanez, V., Luna, E., Harvey, R., Geacintov, N. E., Broyde, S., and Patel, D. J. (1993) Solution conformation of the (+)-*cis*-*anti*-[BP]dG adduct in a DNA duplex: intercalation of the covalently attached benzo[*a*]pyrenyl ring into the helix and displacement of the modified deoxyguanosine, *Biochemistry 32*, 4145-4155.
6. Case, D. A., Darden, T. A., Cheatham, T. E., 3rd, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) AMBER 11, 11 ed., University of California, San Francisco, San Francisco, CA.
7. 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.
8. Cheatham, T. E., Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat, *J Biomol Struct Dyn 16*, 845-862.
9. Perez, A., Marchan, 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 alpha/gamma conformers, *Biophys. J. 92*, 3817-3829.
10. Mocquet, V., Kropachev, K., Kolbanovskiy, M., Kolbanovskiy, A., Tapias, A., Cai, Y., Broyde, S., Geacintov, N. E., and Egly, J. M. (2007) The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[*a*]pyrenyl-DNA lesions, *EMBO J. 26*, 2923-2932.
11. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-242.
12. Ong, M. S., Richmond, T. J., and Davey, C. A. (2007) DNA stretching and extreme kinking in the nucleosome core, *J. Mol. Biol. 368*, 1067-1074.
13. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins 65*, 712-725.
14. 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.
15. Cai, Y., Wang, L., Ding, S., Schwaid, A., Geacintov, N. E., and Broyde, S. (2010) A bulky DNA lesion derived from a highly potent polycyclic aromatic tumorigen stabilizes nucleosome core particle structure, *Biochemistry*.49, 9943-9945
16. Case, D. A., Darden, T. A., Cheatham III, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2006) AMBER 9, University of California, San Francisco, CA.
17. Mezei, M. (1997) Optimal position of the solute for simulations., *J. Comp. Chem. 18*, 812-815.
18. Jorgensen, W. L., Chandreskhar, J., Madura, J. D., Imprey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.
19. Donny-Clark, K., and Broyde, S. (2009) Influence of local sequence context on damaged base conformation in human DNA polymerase iota: molecular dynamics studies of nucleotide incorporation opposite a benzo[a]pyrene-derived adenine lesion, *Nucleic Acids Res. 37*, 7095-7109.
20. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an *N* log(*N*) method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089-10092.
21. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pederson, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.
22. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, *J. Comp. Phys. 23*, 327-341.

23. Case, D. A., Darden, T. A., Cheatham, T. E., 3rd, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2014) AMBER 14, University of California, San Francisco.
24. Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG, *J. Biol. Chem. 257*, 14686-14707.
25. Shao, J. Y., Tanner, S. W., Thompson, N., and T.E., C. (2007) Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms, *3*, 2312-2334.
26. Brice, A. R., and Dominy, B. N. (2011) Analyzing the robustness of the MM/PBSA free energy calculation method: application to DNA conformational transitions, *J. Comput. Chem. 32*, 1431-1440.
27. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E., 3rd. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res. 33*, 889-897.