

# Molecular Dynamics Simulation of 8-Oxoguanine Containing DNA Fragments Reveals Altered Hydration and Ion Binding Patterns

Aymeric Naômé,\*†‡ Patric Schyman,‡ Aatto Laaksonen,‡ and Daniel P. Vercauterens†

Laboratoire de Physico-Chimie Informatique, Unité de Chimie Physique Théorique et Structurale, University of Namur (FUNDP), Namur, Belgium, and Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

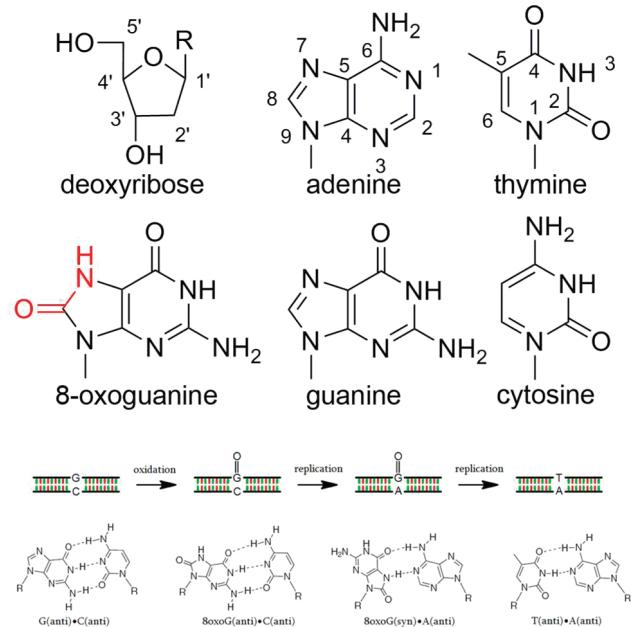
Received: January 4, 2010

We present results of molecular dynamics (MD) simulations of a double-stranded DNA fragment of sequence 5'-d(CGCGAGTCGCG)-3' and 3'-d(GCGCTCAAGCGC)-5' in its native form and altered with guanine in the **G6/C19** base pair replaced by 8-oxoguanine (8oxoG). DNA was simulated in solution with  $\text{Na}^+$  counterions and explicit water molecules. Trajectories were analyzed for radial distribution functions (RDFs), spatial distribution functions (SDFs), residence times, and occupancies, for both water and ions. Structural parameters of DNA fragments were also considered. Particularly, we report sensible differences in the statistical and dynamical behavior of water and ions between the healthy and lesion containing models. Structural features such as sugar–phosphate backbone torsion angles also significantly differ. We conclude that 8oxoG noticeably modifies its close environment and that the observed changes in the vicinity of the lesion site might help the enzymes in charge of cleansing 8oxoG from human DNA to locate the aberration in its intrahelical context. This work supports the existence of a determinant early recognition mode of 8oxoG for hOGG1.

## Introduction

The integrity of the genetic information is constantly threatened by oxidizing agents. Keeping the level of damaged sites in DNA as low as possible and maintaining an error-prone replicative system are constant challenges for living organisms. The endogenous reactive oxygen species, a byproduct of the cell respiration and the inflammatory response metabolism, are the main source of oxidative lesion affecting DNA. One of the most common and among the most mutagenic is 7,8-dihydro-8-oxoguanine (8oxoG). This species results from the oxidation of guanine C8–N7 bond<sup>1</sup> (an O atom in place of an H atom at C8 and an H atom at N7) (Figure 1). During replication of the flawed strand, 8oxoG can be misread by polymerases as a thymine and associated an adenine in the newly synthesized strand. This is because of the extra ability 8oxoG has to form a Hoogsten base pair with adenine besides the normal Watson Crick pair with cytosine. Subsequent to an additional replication cycle, the G/C → T/A transversion surfaces<sup>2</sup> (Figure 1). The repeated occurrence of that transversion whose incidence in the genome has been directly associated to forms of lung and colorectal cancer<sup>3–5</sup> is also linked to a dysfunction of the enzyme hOGG1.<sup>6,7</sup> That protein is the first component of the system entrusted with the mission of cleansing 8oxoG from human DNA.<sup>8</sup>

The enzyme achieves its lesion excision action by inserting in its active site pocket the damaged base, beforehand unpaired, unstacked, and flipped out from the double helix.<sup>9</sup> The structural bases for recognition of the substrate are now well understood but only in the late stage when the base is displaced on the rim of the catalytic site: basically, a healthy base is blocked at the entrance of the active site, while 8oxoG can penetrate.<sup>10</sup> The



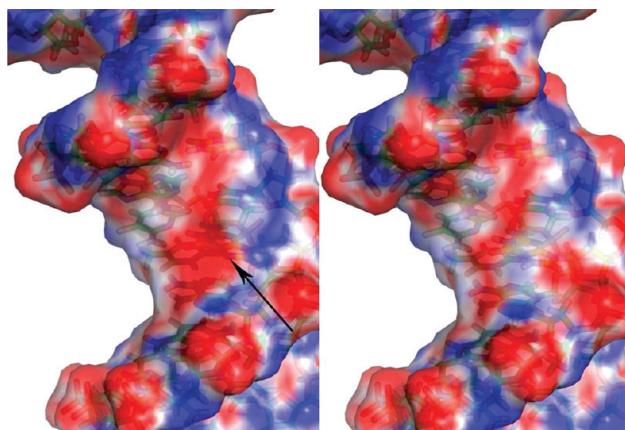
**Figure 1.** Simplified molecular formula of the four natural nucleobases (A, T, G, C) and deoxyribose including atom numbering (top). Structural changes in 8oxoG are highlighted in red. Scheme of the oxidized lesion 8oxoG mutagenic induction resulting in overall G/C → T/A transversion mutation (bottom).

early mechanism by which the specific docking to DNA is triggered and 8oxoG extracted is for its part still obscure. A global mechanism uniquely founded on a discrimination mode at the active site would imply a systematic complete eversion of every guanine for sampling at the active site (and the reverse process to reinser the base in the double helix) until a lesion eventually slips into the active site. The puzzling ability of hOGG1 to stalk and efficiently excise its cognate substrate, structurally very similar to a guanine, amidst the enormous

\* To whom correspondence should be addressed. E-mail: anaome@fundp.ac.be.

† University of Namur.

‡ Stockholm University.



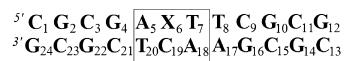
**Figure 2.** Electrostatic potential (ESP) at the molecular surface of healthy (left) and damaged DNA (right) viewed in the major groove, the arrow pointing to the mutation site. ESP values range from  $-16$  (red) to  $-8 \text{ } kT/e$  (blue). The very negative zone present in healthy DNA in the middle of the groove is not visible in damaged DNA. There is a build-up of negative charge close to the rim of the groove, at the sugar phosphate backbone. ESP profiles were calculated from the force field atomic charges with APBS.<sup>20</sup>

excess of healthy bases rules out this possibility. Stopped flow DNA–protein association kinetics<sup>11</sup> and single molecule microscopy experiments<sup>12</sup> go in that sense. In addition, the equivalent stability of the G/C and 8oxoG/C pairs in terms of H bonds and stacking energy<sup>13</sup> discard the possibility for hOGG1 to grab spontaneously unpaired 8oxoG preferentially; “on the fly” grabbing is anyway very improbable considering the three-dimensional (3D) diffusion limit of the protein and the lifetime of an open state.<sup>14–16</sup> Furthermore, no major structural changes are observed in naked damaged DNA that could favor association.<sup>17,18</sup> Hence comes the need to appeal to an early intrahelical recognition mode that allows hOGG1 to pick over DNA for a first selection before the final discrimination step at the active site.

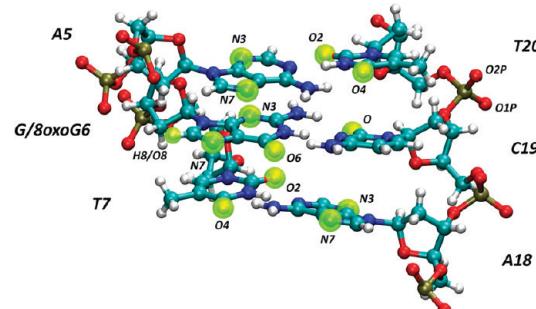
Several distinct physicochemical contributions lead to the formation of a protein DNA complex, i.e., H-bonds, electrostatic interactions, direct and indirect contacts between amino acids and phosphate, sugars, and bases, water-mediated contacts, hydrophobic effects, ion release, mutual conformation rearrangement, bending, distortion, etc.<sup>19</sup> Water- and cation-mediated indirect interactions predominate in the protein DNA association process. Since DNA is highly negatively charged, it strongly binds those species. The role of water and cations must therefore be considered while studying the docking of a protein to DNA and attempting to reveal putative recognition modes. The fact that the electrostatic profile of DNA is sensibly modified upon insertion of 8oxoG compared to healthy DNA suggests that we can observe perceivable differences of water and ion behavior in the vicinity of the lesion site (Figure 2). We follow this reasoning and suggest the possibility of an early intrahelical recognition mode of 8oxoG for hOGG1 in which water and cations play an important role. We therefore focused on the statistical and dynamical behavior of water and cations around fragments of healthy and 8oxoG containing DNA by applying classical molecular dynamics (MD) methodologies.

## Methods

**Molecular Dynamics.** Our system of interest consists of dodecamer duplexes of sequence 5'-d(CGCGAAXTCGCG)-3' and 3'-d(GCGCTCAAGCGC)-5' with X = G and 8oxoG (Figure 3).



**Figure 3.** Sequence of the model DNA fragments, X standing for G in the healthy model and 8oxoG in the damaged one. The central base pairs for which RDFs, SDFs, residence times, and occupancies are analyzed are boxed.



**Figure 4.** 3D structure of the three central base pairs from our models with electronegative sites highlighted in yellow. O2P and O1P are pointing forward (major groove side) and backward (minor groove side), respectively.

The healthy canonical B-DNA structure was generated using the fiber tool from the software for DNA structure analysis 3DNA.<sup>21</sup> Hydrogen atoms were added with the pdb2gmx utility in GROMACS.<sup>22</sup> To create the damaged DNA fragment, we changed H8 at G6<sub>C8</sub> for O8 and removed H7 bound to G6<sub>N7</sub>. Both systems were then prepared following the same procedure. A cubic box of side length 6.25 nm (244.1 nm<sup>3</sup>) with periodic boundary conditions in all directions was filled with the double-stranded DNA fragment, 7683 water molecules (7682 for the damaged system), and 22 Na<sup>+</sup> counterions, keeping the systems neutral by balancing the negative charges of the DNA phosphate groups. The box size was chosen such that any DNA atom is at least 1.0 nm from a box face, preventing unwanted interactions of the solute molecule with its images in translated unit cells. Na<sup>+</sup> were added by replacing a solvent molecule at ca. 6 Å from the phosphorus in the direction of the bisector of the angle  $\angle O1P-P-O2P$  (O1P and O2P being the phosphate nonbridging oxygen atoms)<sup>23</sup> (Figure 4). The system contains a total of 31 511 atoms (including the dummy atoms in the four-particle water model). The damaged system contains one water molecule less and one extra DNA atom for a total of 31 508 atoms.

We used the AMBER-Parm99<sup>24</sup> force field (FF) to model the system, an extensively used FF for proteins and nucleic acids simulation, as ported for the GROMACS simulation package.<sup>22</sup> The FF parameters for 8oxoG absent from Parm99 were taken as such from ref 25. The partial charges on atoms were computed according to the RESP method developed and applied by the developers of the AMBER FFs.<sup>26</sup> Nonexisting bending angles and dihedrals were approximated to existing Parm99 parameters. Regarding water, we employed the TIP4P model.<sup>27</sup> The long-range electrostatic interactions were treated by the particle mesh Ewald (PME) summation method<sup>28</sup> with a cutoff distance of 10 Å. The same cutoff distance was used for the van der Waals interactions. A steepest descents energy minimization on the whole system was performed with a maximum force tolerance of 100 kJ·mol<sup>-1</sup>·nm<sup>-1</sup>. The minimization converged after 150–160 steps. A time step of 2 fs was used for the integration of Newton's equations, since all bond lengths were constrained to their equilibrium values using the LINCS algorithm.<sup>29</sup> Solvent and ions were relaxed for 100 ps at constant volume while keeping the DNA atomic positions restrained (except hydrogen atoms) by a harmonic potential of 1000

$\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ . Two rounds of 50 ps equilibration MD were then performed in the ( $N,P,T$ ) ensemble with positional restraints of 1000 and 500  $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$  on the DNA heavy atoms. Restraints were then removed for the production run. The temperature and pressure were kept at 310 K and 1 atm by turning on the Berendsen thermostat and barostat.<sup>30</sup> The center of mass motion and the angular momentum of the solute DNA were removed at every step as to avoid rotational and translational movements eventually driving the solute out of the box. This removal facilitates the trajectory analysis. Coordinates were recovered every 2 ps for 150 ns. We carried out two independent sets of simulations to check consistency; we thus acquired a total of 600 ns of DNA MD trajectory ( $2 \times 150$  ns for the healthy system +  $2 \times 150$  ns for the damaged one).

## Analysis

**Radial and Spatial Distribution Functions.** Radial distribution functions (RDFs) provide useful information about how a given type of particle is distributed around another set of particles. It is defined as the probability of finding a pair of atoms a distance  $r$  apart, relative to the probability expected for a completely random distribution at the same density. The volume around each particle A is divided into concentric spherical shells of thickness  $dr$ , and the number of particles B in each shell is counted and divided by the shell volume  $4\pi r^2 dr$ , to obtain the local density. The densities at each distance are then averaged over all particles A and over time to provide  $\langle \rho_B(r) \rangle$ . When it is normalized with the overall density  $\langle \rho_B \rangle$ , it results in the RDF for the particle pair AB:

$$g_{AB}(r) = \frac{1}{N_A \langle \rho_B \rangle} \sum_{i=1}^{N_A} \frac{n_B(r)}{4\pi r^2 dr} = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle} \quad (1)$$

where  $n_B(r)$  is the number of particles B such that  $r < r_{AB} < r + dr$ . RDFs provide data such as occupation numbers (or occupancies), a measure of how many particles are bound to a given site, by integrating a peak up to the next minimum; they give the position, the magnitude, and the number of local extrema in the distribution of a species.

As an RDF is orientationally averaged over the angular coordinates, many details of the local solution structure can be lost because of the cancellation of contributions from regions of high and low probability at the same distance but different parts of the local solution structure. To overcome the limitations of RDFs, spatial distribution functions (SDFs) are introduced.<sup>31</sup> They span both the radial and angular coordinates, which allow a good description of 3D density distributions of water and ions in a local coordinate system linked to the solute. In the case of large and very flexible molecules such as DNA, that system must be carefully chosen so as not to smear out hydration and ion binding features situated far from the origin. The full representation of an SDF, a 4D function, is best represented as iso-density surfaces drawing volumes of high density probabilities.

We calculated RDFs for  $\text{Na}^+$  and water oxygen atom (Ow) around various sites on DNA: the phosphate O1P and O2P atoms and the electronegative atoms in the grooves, A<sub>N3</sub>, A<sub>N7</sub>, T<sub>O2</sub>, T<sub>O4</sub>, C<sub>O</sub>, G<sub>N3</sub>, G<sub>N7</sub>, G<sub>O6</sub>, and 8oxoG<sub>O8</sub>. Because of the high flexibility of the DNA fragment extremities, SDFs are only well resolved around the three central base pairs containing the lesion site (Figure 3) on which the local coordinate system was defined. (Atom 1 = G/8oxoG6<sub>N9</sub>; 2 = A18<sub>N9</sub>; 3 = T20<sub>N1</sub>. Atom 1 is the origin, the  $x$ -axis, the bisector of the  $\angle 312$  angle, and the  $z$ -axis, the normal to the 1–2–3 plane). We thereby mainly focused

on those nucleobases in the analysis. A 3D structure of those nucleotides and the sites around which RDFs, SDFs, occupancies, and residence times were obtained is visible in Figure 4. RDFs and SDFs were calculated using the trajectory analysis tools within the MD software M.DynaMix.<sup>32</sup>

**Ion Mobility.** The mobility of the ions in the presence of DNA was characterized by their averaged mean squared displacements (MSDs) and self-diffusion coefficients ( $D$ ) derived from the Einstein equation (eq 2). This requires one to calculate the MSD for every ion  $i$  over a series of time intervals.  $D$  is then taken as the slope of the averaged MSD versus time in the linear region:

$$D = \lim_{t \rightarrow \infty} \frac{\langle |x_i(t_0 + t) - x_i(t_0)|^2 \rangle_{i,t_0}}{6t} \quad (2)$$

We calculated the averaged MSD for time intervals in steps of 10 ps up to 150 ns, i.e., the total time of simulation. The self-diffusion coefficients were calculated by fitting the averaged MSD using intervals from 1 to 10 ns. Attempts to calculate  $D$  on longer time intervals lead to a growing uncertainty because of the breaking of linearity arising from the poorer sampling statistics as  $t$  becomes comparable to the simulation time.

**Residence Time.** How long water molecules and ions stay in the coordination shell around a given site is important to understand the dynamics of water/ion exchange and to assess the stability of bound species within the coordination shell. Residence time is calculated from a standard time correlation function:<sup>33</sup>

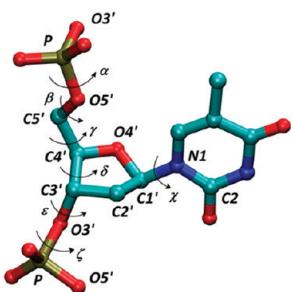
$$p(t) = \langle P_j(t, t_i; t^*) \rangle_{i,j} / \langle P_j(0, t_i; t^*) \rangle_{i,j} \quad (3)$$

$P_j(t, t_i; t^*)$  gives the probability of a particle  $j$  to remain during a time  $t$  within the shell (the limit of which is taken as the first minimum of the corresponding RDF). It takes the value 1 if a molecule  $j$  lies within the shell at both time steps  $t_i$  and  $t_i + t$  and does not cross the boundary for a period longer than  $t^*$ . Otherwise, it takes the value 0. The function is averaged over all particles  $i$  and over all periods of time  $t$ . This correlation function can be fitted by a decaying exponential whose specific time  $\tau$  is a definition of the residence time:

$$p(t) \sim e^{-t/\tau} \quad (4)$$

For our purpose, we chose  $t^*$  equal to 8 ps (four time steps). We computed residence times around the sites listed above in the three central base pairs. In addition to the specific time  $\tau$ , we also give the maximum residence times  $t_{\max}$  and simple average residence times  $\langle \tau \rangle$ . Residence times were calculated using the trajectory analysis tools within the MD software M.DynaMix.<sup>32</sup>

**Nucleotide Conformers.** The canonical B-DNA conformation is divided into two subsets of conformers: B<sub>I</sub> and B<sub>II</sub>. A nucleotide can be classified in one of these categories according to the values of the torsion angles  $\varepsilon$  and  $\zeta$ <sup>34,35</sup> (B<sub>I</sub>:  $\varepsilon - \zeta < 0^\circ$ ,  $\varepsilon = 120 - 210^\circ$ ,  $\zeta = 235 - 295^\circ$ ; B<sub>II</sub>:  $\varepsilon - \zeta > 0^\circ$ ,  $\varepsilon = 210 - 300^\circ$ ,  $\zeta = 150 - 210^\circ$ ). We also characterized the nucleotides in term of their  $\alpha/\gamma$  conformer populations. A definition of the torsion angles is given in Figure 5. Torsion angles were computed with 3DNA.<sup>21</sup> Figures were made with VMD<sup>36</sup> and gOpenmol.<sup>37</sup>



**Figure 5.** Representation of a unit nucleotide with torsion angles for the sugar-phosphate backbone given as  $\alpha = \text{O}3'\text{-P-O}5'\text{-C}5'$ ,  $\beta = \text{P-O}5'\text{-C}5'\text{-C}4'$ ,  $\gamma = \text{O}5'\text{-C}5'\text{-C}4'\text{-C}3'$ ,  $\delta = \text{C}5'\text{-C}4'\text{-C}3'\text{-O}3'$ ,  $\epsilon = \text{C}4'\text{-C}3'\text{-O}3'\text{-P}$ ,  $\zeta = \text{C}3'\text{-O}3'\text{-P-O}5'$ , and  $\chi = \text{O}4'\text{-C}1'\text{-N}1/\text{N}9\text{-C}2/\text{C}4$ , with N1 and C2 being for pyrimidines and N9 and C4 for purines.

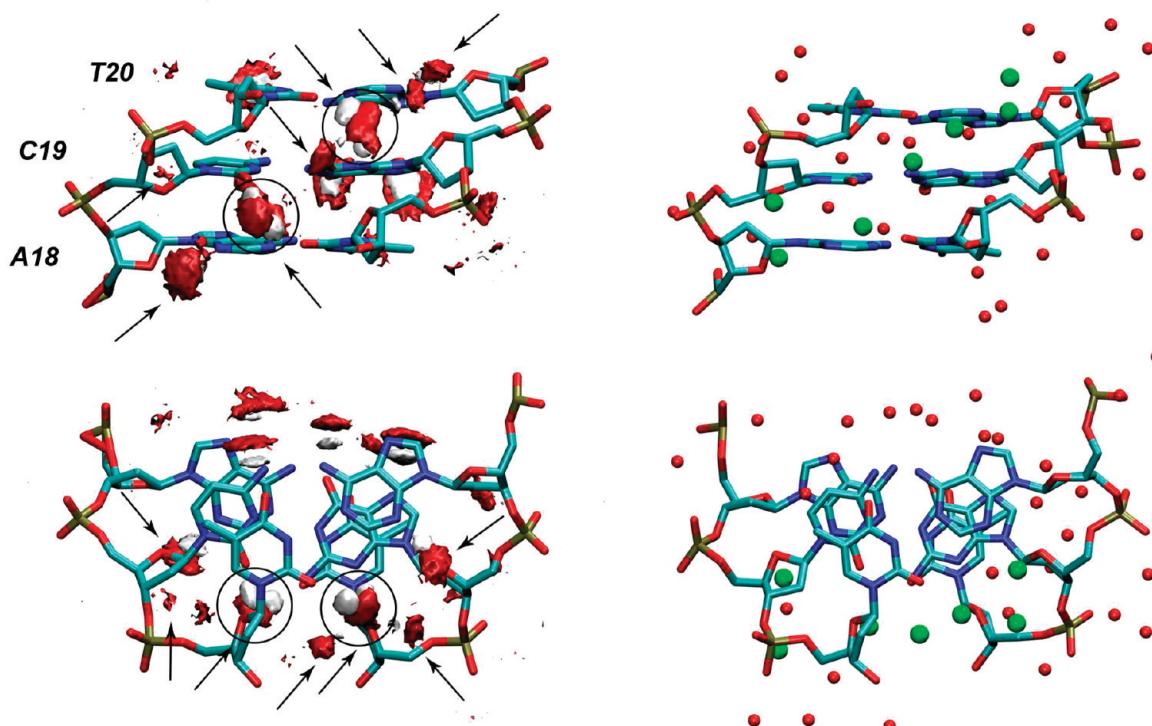
## Results and Discussion

**Convergence of the Properties.** MD simulation trajectories can be considered as reliable provided (i) the configurational space sampling is independent of the starting configuration and (ii) the configurational space sampled corresponds to experimental geometries. The root-mean-square deviations (rmsd's) relative to the starting structures of both sets of independent simulations stabilize quickly after the release of the constraints on the heavy atoms. Their values for healthy and damaged DNA are  $3.2 \pm 0.5$  and  $3.3 \pm 0.8$  Å, respectively. The criteria to decide when to stop were dictated by the convergence of the water and  $\text{Na}^+$  RDFs. When water RDFs converged rapidly after 5–8 ns, the proper statistical sampling of the 22  $\text{Na}^+$  required at least 100 ns. During the whole course of the simulations, DNA retained its native B-DNA conformation. Only local and partial unfoldings were detected in the extremity base pairs of the double-helices: long-lived base opening and base stacking

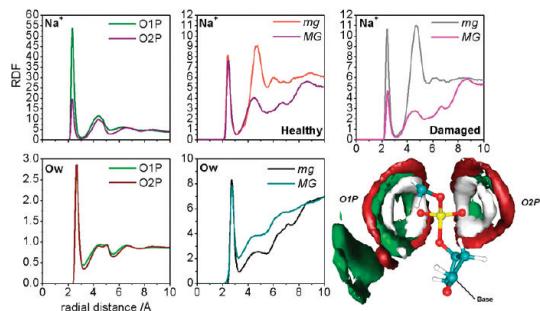
events occurred in all four simulations. The four base pairs of the extremities were not considered in the analyses.

**Comparison with Crystal Structure.** The hydration patterns around the three central base pairs in our healthy DNA model (AGT/TCA) were compared with the distribution of X-ray diffraction (XRD) resolved water molecules in the crystal structure of a DNA fragment containing the same sequence.<sup>38</sup> It turns out that the hydration motif drawn by the SDFs in the minor groove (mg) matches very well the positions of long-range ordered water molecules in the crystal (Figure 6). The electronegative sites in the mg,  $\text{A}5_{\text{N}3}$ ,  $\text{G}6_{\text{N}3}$ ,  $\text{T}7_{\text{O}2}$ ,  $\text{A}18_{\text{N}3}$ ,  $\text{C}19_{\text{O}}$ , and  $\text{T}20_{\text{O}2}$  (Figure 4), describe a scaffold organizing water in a continuous chain referred as the spine of hydration. This belt is arranged by direct H-bonding to DNA and by forming secondary H-bonds with water from the next solvation shells<sup>39</sup> (not shown for clarity). Some hydration sites are so well-defined that the SDFs of Ow and water hydrogens (Hw) draw “frozen images” of the water molecules (circled in Figure 6). The retention of a highly structured solvation pattern from the crystalline state to solution is a sign of the strength of the co-operative forces involved in the hydration of DNA. Solvent is moreover often considered as an integral part of DNA structure.<sup>39</sup> The damaged DNA model shows very similar hydration motifs (not shown here).

**General Binding Features of Water and Ions.** Distribution functions for water and ions calculated around the  $\text{PO}_4$  groups of the DNA backbone look alike for both healthy and damaged systems. The local concentration of  $\text{Na}^+$  in the close vicinity of  $\text{PO}_4$  is about 30 times higher than in the bulk but still barely accounts for 2–3% of the occupancy of water at that site. While water is evenly distributed around  $\text{O}1\text{P}$  and  $\text{O}2\text{P}$ ,  $\text{Na}^+$  shows a clear preference for  $\text{O}1\text{P}$  (Figure 7). RDFs show an average of 0.10  $\text{Na}^+$  bound to  $\text{O}1\text{P}$ , whereas 0.04  $\text{Na}^+$  is found close to  $\text{O}2\text{P}$  (Table 1). This difference is likely due to ions captured in



**Figure 6.** 3D representations of water around the AGT/TCA central base pairs. Front view of the mg and view along the helical axis seen from the top of the structure averaged over 150 ns, as obtained from MD simulations, and the related SDFs of water (left). SDF iso-surfaces encompass densities higher than 140 particles/ $\text{nm}^3$ . Corresponding views of the crystal structure (right). Arrows indicate which blobs correspond to resolved water molecules (green spheres) in the crystal structure. Stable cross-strand bridging water molecules are encircled.



**Figure 7.** RDFs of Na<sup>+</sup> and Ow around O1P and O2P (left) and around all electronegative sites of the MG and mg calculated for the healthy DNA model (middle). The RDF of Na<sup>+</sup> in the grooves is also drawn for the damaged system (top right). Averaged SDFs of water and ions calculated for the PO<sub>4</sub> groups (right). SDFs of Ow (red) and Hw (white) are drawn for densities higher than 110 particles/nm<sup>3</sup>; SDF of Na<sup>+</sup> (green), for densities higher than 5 particles/nm<sup>3</sup>.

the confined space of the *mg*, closed by the PO<sub>4</sub> groups. The O1Ps indeed form a claw retaining the ions just above the surface of the *mg* (ca. 4.5 Å). This high density of ions is visible as the second peak of the Na<sup>+</sup> RDFs in the *mg* (Figure 7, top middle and right). Note that the second Na<sup>+</sup> RDF maximum in the MG lies lower above its floor (4.7 Å) than for the *mg* (4.8 Å) because of the absence of hindrance from the O2Ps and the larger groove width. Water in the grooves is more evenly distributed. The RDF for the narrower *mg*, though, has its minimum lying lower because of the restricted space available to accommodate water molecules. Averaged SDFs around PO<sub>4</sub> display a pictorial image of the solvation and coordination shell (Figure 7, bottom right). The same trends apply for damaged DNA except for the Na<sup>+</sup> distribution in the grooves. While the first coordination shell in both grooves of healthy DNA is almost equally populated, there is a net unbalance favoring the *mg* in damaged DNA. The second high ion density just above the *mg* is also more important. Ion distribution in the grooves will be discussed in more detail in the section “Differences in water and ions distribution between healthy and damaged DNA”.

**Ion Mobility.** Although they are not in an ideal diffusive regime because of binding events with DNA, movements of ions and water can be characterized by their self-diffusion coefficient *D*. The self-diffusion coefficients for Na<sup>+</sup> calculated from the slope of the MSD in the linear region for the healthy and damaged systems are, respectively,  $0.98 \times 10^{-9}$  and  $0.99 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup> (the standard error is less than 0.1%). Those values are consistent with the experimental values for Na<sup>+</sup> solutions at infinite dilution,  $1.33 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup>.<sup>40</sup> They are also in correct agreement with self-diffusion coefficients derived from simulation data with various water and ion models at high ionic strength, ranging from  $0.67 \times 10^{-9}$  to  $1.2 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup>.<sup>41–43</sup> Ions are slightly more mobile in the damaged system.

It is interesting to split the MSD into its directional components (*xy*-MSD and *z*-MSD) in order to study the ion dynamics around DNA. Since the DNA fragments are kept with their helical axis oriented along the *z*-direction during the course of the simulation, one should indeed observe a certain anisotropy in the relative contribution to the total MSD of the MSD components. The *z*-direction along which the grooves stretch out is discernible from the *x*-*y*-directions that are equivalent. We thus break up the total MSD into *z*-MSD and *xy*-MSD. While the lateral self-diffusion coefficient is almost identical for both systems ( $D_{xy} = 0.62 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup> and  $D_{xy} = 0.63 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup> for the healthy and damaged systems, respectively), the *z*-MSD component for the damaged system shows a steeper

slope than that for the healthy system, yielding  $D_z = 0.37 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup> instead of  $D_z = 0.35 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup>. We thus learn that the difference between the self-diffusion coefficients is mainly caused by a slight increase in ion mobility along the helical axis of damaged DNA (Figure 8). To interpret the greater mobility of Na<sup>+</sup> along the *z*-axis compared to the other directions and to explain its higher value in the damaged system, one needs to inspect ion occupancies and residence times at the phosphates and grooves.

From Table 2, one notices that Na<sup>+</sup> ions bound in the grooves are much less mobile than those bound to the PO<sub>4</sub> groups. For both systems, the maximum period of time ions are found to be bound in the coordination sphere of PO<sub>4</sub> is at least 60% shorter than in the grooves. The fact that the quantities  $\tau$ ,  $\langle t \rangle$ , and  $t_{\max}$  are all three higher for the MG in the healthy system is reliable enough to affirm that ions are actually more bound to healthy DNA at that site. The trend is opposite in the *mg* even though the difference is less marked. There is no significant difference in what concerns the PO<sub>4</sub> groups. Whereas ions coordinated to phosphates are transient and free to return to the bulk solution, ions in the grooves are more bound and constrained in the *x*- and *y*-directions, confined in the restricted space closed by the rims of the sugar-phosphate backbone. Occupancies of Na<sup>+</sup> in the coordination sphere of PO<sub>4</sub> oxygen atoms and in the grooves are listed in Table 3, together with the ratios grooves/PO<sub>4</sub> and damaged/healthy systems. There is about 20% fewer ions coordinated in the grooves of damaged DNA, mainly lost from the MG. Despite the build-up in the *mg* (+59%), the loss is not fully compensated. The expelled ions go swell the ranks around the phosphates of damaged DNA (+15%). There is an obvious shift of ions from the grooves to the PO<sub>4</sub> (52% vs 35% of the total occupancy).

Different binding modes of ions on DNA contribute differently to the directional MSDs. Movements at the surface of the double helix can be of two types: (i) ions hopping from one phosphate group to another on the backbone and (ii) ions translating in the grooves. The general trend of the *z*-MSD, steeper than the *x* and *y* contributions ( $(0.35 - 0.37) \times 10^{-9}$  vs  $0.62 - 0.63/2 = 0.31 \times 10^{-9}$  m<sup>2</sup>·s<sup>-1</sup>), can be rationalized as follows. The ions coordinated to the grooves preferentially slide on the groove floor from one electronegative site to another (in the *z*-direction) and then return to the bulk. They are blocked in the *mg* space or above it, retained by the O1Ps, and thus do not contribute much to the *xy*-MSD. If the ions coordinated to the PO<sub>4</sub> are, for their part, free to return to the bulk and thus participate in the *x*- and *y*-MSDs, they can also jump back and forth along the strings of phosphate groups adding up to the *z*-MSD. The net result is therefore an enhanced mobility in the direction parallel to the helical axis. The greater *z*-direction mobility in damaged DNA relative to healthy DNA originates from the shortened residence times in the grooves and by the larger population just above the grooves (Figure 7, top middle and right); more ions are hopping from one phosphate group to the other, crossing the grooves, and climbing the base pair steps.

**Differences in Water and Ion Distributions between Healthy and Damaged DNA.** The first major differences in species distribution appear around the critical nucleobase G6/8oxoG6. It ensues from the appearance of an H-bond acceptor, the oxo group at 8oxoG6<sub>C8</sub>, and the modification of G6<sub>N7</sub> from an H-bond acceptor to a donor. Two new hydration sites hence appear at 8oxoG6<sub>O8</sub> in the MG between which water molecules are shared (red arrows in Figure 9). Water can bridge between 8oxoG6<sub>O8</sub> and a PO<sub>4</sub> oxygen atom, either O5' or O2P. Some Na<sup>+</sup> density is also found at the electronegative sites within

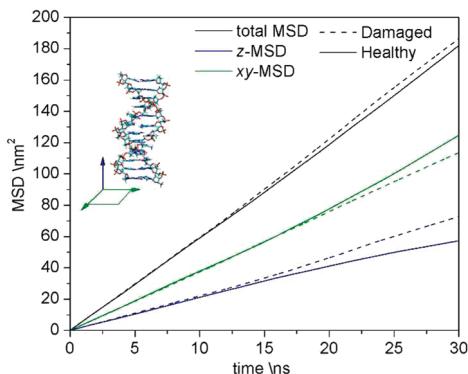
**TABLE 1: Intensities of the First Maximum and Corresponding Distances in Å (in Parentheses) of RDFs of Na<sup>+</sup> and Ow around Selected Sites at DNA, Occupancies, and Residence Times (ps) for Both Healthy and Damaged Systems**

site	RDF		occupancy		residence time ( $\tau$ , $\langle t \rangle$ , $t_{\max}$ )	
	Na <sup>+</sup>	Ow	Na <sup>+</sup>	Ow	Na <sup>+</sup>	Ow
<b>PO<sub>4</sub></b>						
O1P						
damaged	59.0 (2.3)	2.8 (2.7)	0.11	2.9	8.6, 8.9, 110	8.6, 8.6, 384
healthy	53.8 (2.3)	2.8 (2.7)	0.10	2.9	8.6, 8.9, 116	8.6, 8.6, 316
O2P						
damaged	21.9 (2.3)	2.9 (2.7)	0.04	3.0	8.6, 8.9, 108	8.6, 8.6, 408
healthy	19.6 (2.3)	2.9 (2.7)	0.04	3.0	8.6, 8.9, 98	8.6, 8.6, 346
G6 <sub>O1P</sub> + T7 <sub>O1P</sub>						
damaged	43.0 (2.3)	2.8 (2.7)	0.08	3.0	8.3, 8.8, 66	8.6, 8.6, 216
healthy	48.2 (2.3)	2.8 (2.7)	0.09	3.0	8.5, 8.9, 78	8.5, 8.6, 304
G6 <sub>O2P</sub> + T7 <sub>O2P</sub>						
damaged	35.7 (2.3)	2.8 (2.7)	0.07	3.2	8.5, 8.8, 66	8.6, 8.6, 210
healthy	22.5 (2.3)	2.8 (2.7)	0.04	3.0	8.6, 9.1, 76	8.5, 8.6, 228
C19 <sub>O1P</sub> + T20 <sub>O1P</sub>						
damaged	73.6 (2.3)	2.7 (2.7)	0.14	2.9	8.5, 9.0, 76	8.6, 8.7, 316
healthy	58.7 (2.3)	2.8 (2.7)	0.11	3.0	8.7, 9.0, 96	8.6, 8.7, 252
C19 <sub>O2P</sub> + T20 <sub>O2P</sub>						
damaged	15.3 (2.3)	2.9 (2.7)	0.03	3.0	8.6, 9.0, 76	8.6, 8.7, 354
healthy	14.8 (2.3)	2.9 (2.7)	0.03	3.0	9.0, 9.1, 86	8.6, 8.7, 254
<b>mg damaged</b>	16.5 (2.5)	0.77 (2.7)	0.046	1.02	14.9, 14.3, 146 <sup>a</sup>	13.9, 13.2, 628 <sup>a</sup>
<b>healthy</b>	10.4 (2.5)	0.83 (2.7)	0.029	1.05	14.7, 14.2, 182 <sup>a</sup>	13.9, 13.2, 642 <sup>a</sup>
<b>A5<sub>N3</sub></b>						
damaged	0.47 (2.6)	0.87 (2.8)	0.003	1.09	9.9, 10.4, 68	9.2, 9.2, 216
healthy	1.10 (2.5)	0.86 (2.8)	0.001	1.02	10.2, 9.8, 86	9.2, 9.1, 288
<b>G6<sub>N3</sub></b>						
damaged		0.48 (2.9)		0.84		9.3, 9.3, 196
healthy		0.81 (2.8)		1.02		9.5, 9.4, 220
<b>T7<sub>O2</sub></b>						
damaged	65.7 (2.4)	1.01 (2.7)	0.12	0.99	8.4, 9.1, 68	8.6, 8.7, 168
healthy	28.1 (2.4)	1.10 (2.7)	0.05	1.06	8.6, 9.4, 68	8.6, 8.7, 364
<b>A18<sub>N3</sub></b>						
damaged	40.3 (2.5)	0.75 (2.8)	0.11	0.96	8.9, 9.3, 72	9.1, 9.1, 158
healthy	32.3 (2.5)	0.80 (2.8)	0.09	0.98	9.1, 9.4, 68	9.1, 9.1, 456
<b>C19<sub>O</sub></b>						
damaged	5.7 (2.4)	1.09 (2.7)	0.01	1.03	9.6, 9.7, 78	8.9, 8.9, 162
healthy	14.5 (2.4)	1.08 (2.7)	0.03	1.01	9.4, 9.5, 106	8.9, 8.9, 228
<b>T20<sub>O2</sub></b>						
damaged	7.0 (2.4)	1.15 (2.7)	0.014	1.2	9.1, 9.0, 66	8.0, 8.1, 142
healthy	0.6 (2.4)	1.12 (2.7)	0.001	1.08	9.4, 9.6, 128	8.8, 8.9, 154
<b>MG damaged</b>	5.9 (2.5)	0.88 (2.7)	0.015	1.19	14.6, 14.1, 174 <sup>a</sup>	13.8, 13.2, 732 <sup>a</sup>
<b>healthy</b>	15.1 (2.5)	0.79 (2.7)	0.044	1.15	15.0, 14.4, 194 <sup>a</sup>	13.9, 13.2, 1036 <sup>a</sup>
<b>A5<sub>N7</sub></b>						
damaged	4.9 (2.5)	0.96 (2.8)	0.013	1.09	9.4, 9.6, 86	8.5, 8.6, 148
healthy	36.7 (2.5)	0.91 (2.8)	0.098	1.19	8.9, 9.4, 64	8.6, 8.7, 158
<b>G6<sub>N7</sub></b>						
damaged		0.72 (2.9)		1.19		9.1, 9.1, 288
healthy	26.3 (2.5)	0.76 (2.8)	0.08	1.03	9.0, 9.4, 88	8.8, 8.9, 242
<b>G6<sub>O6</sub></b>						
damaged	12.2 (2.4)	1.04 (2.7)	0.03	1.14	8.3, 8.8, 86	8.6, 8.7, 190
healthy	28.8 (2.4)	0.99 (2.7)	0.07	1.16	8.7, 9.1, 88	8.6, 8.7, 242
<b>8oxoG6<sub>O8</sub>/G6<sub>H8</sub></b>						
damaged	9.4 (2.4)	1.37 (2.7)	0.02	1.65	7.4, 7.8, 74	8.8, 8.8, 162
healthy		0.40 (2.7)				8.1, 8.2, 184
<b>T7<sub>O4</sub></b>						
damaged	1.7 (2.4)	1.00 (2.7)	0.004	1.12	9.8, 9.6, 82	8.9, 9.0, 340
healthy	6.4 (2.4)	0.97 (2.7)	0.016	1.10	8.4, 9.1, 76	8.9, 8.9, 180
<b>A18<sub>N7</sub></b>						
damaged	8.1 (2.5)	0.93 (2.8)	0.020	1.16	9.3, 9.5, 98	8.8, 8.9, 150
healthy	10.1 (2.5)	0.93 (2.8)	0.024	1.20	10.4, 9.9, 96	9.0, 9.0, 256
<b>T20<sub>O4</sub></b>						
damaged	10.5 (2.4)	0.97 (2.7)	0.023	1.06	9.2, 9.5, 74	8.8, 8.9, 218
healthy	11.7 (2.3)	1.00 (2.7)	0.025	1.12	9.5, 9.6, 88	9.0, 8.9, 866

<sup>a</sup> Average residence times for all electronegative sites in the grooves were evaluated for a shell radius of 5 Å.

the first hydration shell and is competing with water for the complexation of 8oxoG6<sub>O8</sub>. The same competition takes place

at the electronegative sites 8oxoG/G6<sub>O6</sub> and G6<sub>N7</sub>. The Na<sup>+</sup> distribution is particularly affected at the O6 and N7 sites. While



**Figure 8.** MSD of  $\text{Na}^+$  in the damaged and healthy systems. Total MSD and projections on the  $xy$  plane and  $z$ -directions are plotted. The inset shows the orientation of the DNA fragments with color correspondence between the system of axes and the MSD traces. Self-diffusion coefficients were obtained by linear regression from 1 to 10 ns.

**TABLE 2:**  $\text{Na}^+$  Residence Times (ps) at the Phosphate Groups and in the Grooves at the Level of the Central Base Pairs AXT/TCA ( $X = \text{G}$  or 8oxoG)<sup>a</sup>

	site	$\tau$	$\langle t \rangle$	$t_{\max}$
<b>PO<sub>4</sub></b>	damaged	8.6	8.9	110
	healthy	8.6	8.9	116
<b>MG</b>	damaged	14.6	14.1	174
	healthy	15.0	14.4	194
<b>mg</b>	damaged	14.9	14.3	146
	healthy	14.7	14.2	182

<sup>a</sup> Nonbridging oxygens of phosphate groups from G6, T7, C19, and T20 are considered.

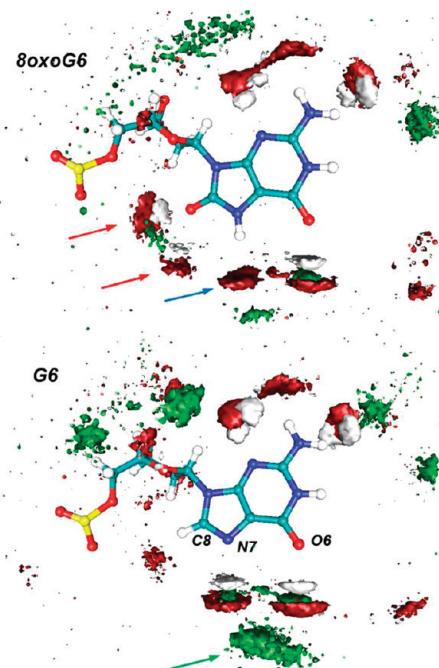
**TABLE 3:**  $\text{Na}^+$  Occupancies at Grooves and Phosphates at the Level of the Central Base Pairs AXT/TCA ( $X = \text{G}$  or 8oxoG)<sup>a</sup>

	grooves	PO <sub>4</sub>	occup. ratio
damaged	0.029 (0.015/0.046)	0.082 (0.115/0.049)	35%
	0.037 (0.044/0.029)	0.071 (0.103/0.038)	
occup. ratio	78% (34%/159%)	115% (112%/129%)	

<sup>a</sup> Average occupancies for one electronegative site in the grooves and one phosphate group oxygen. Values between parentheses are for (MG/mg) and (O1P/O2P).

one observes a localized well-defined  $\text{Na}^+$  high density zone bound to the Ow of the first hydration shell in G6 (green arrow in Figure 9), it is greatly weakened in 8oxoG6 because water molecules bound to 8oxoG6<sub>N7</sub> are turned over. Ow is involved in H-bonds with N7 and  $\text{Na}^+$  ions are dispersed by the Hw atoms then oriented toward the bulk.

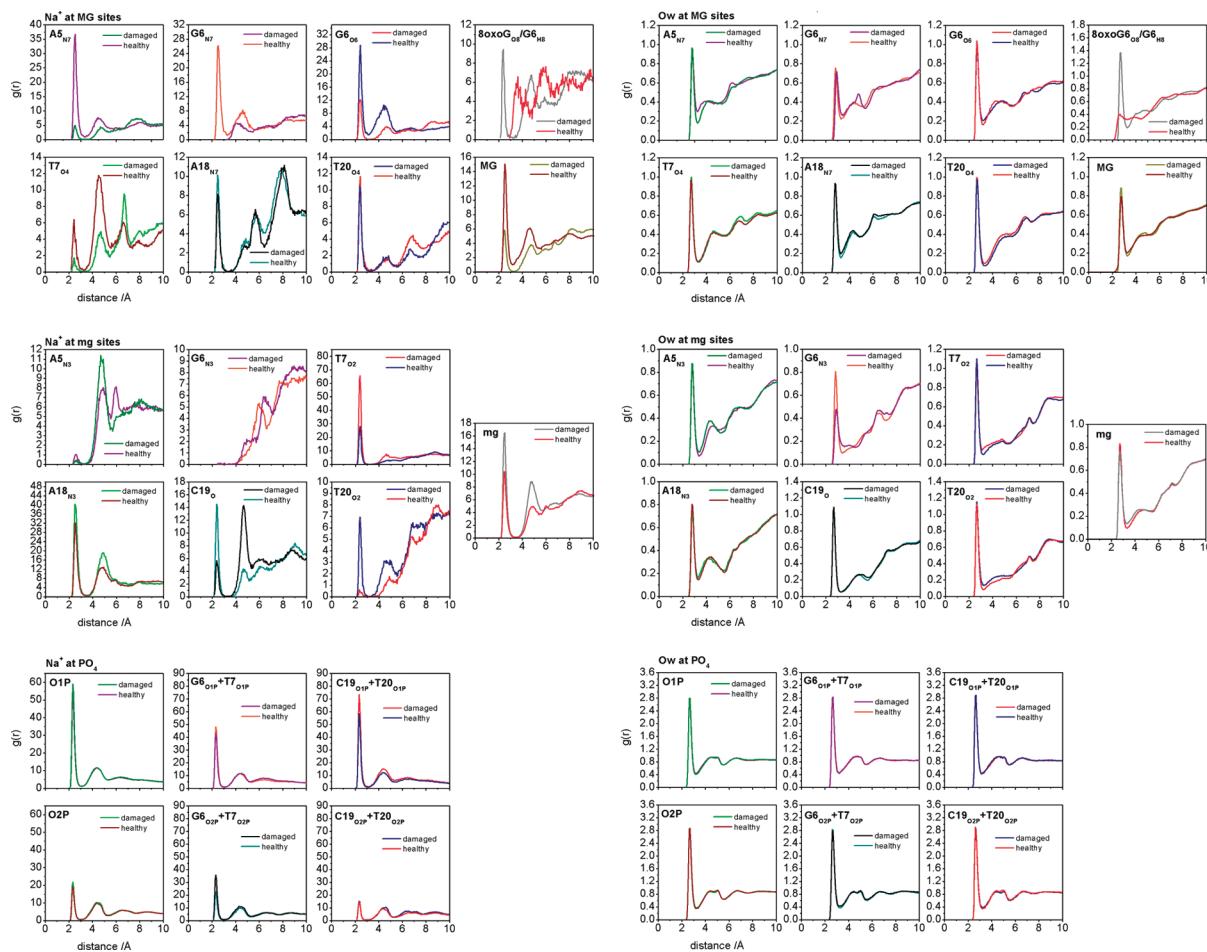
The shift of an electronegative site in 8oxoG (N7 → O8) closer to the sugar phosphate backbone has the larger scale consequence of moving the  $\text{Na}^+$  density from the middle of the MG to the rim of the strand bearing the lesion. There are twice fewer ions at 8oxoG6<sub>O6</sub> (0.07/0.03), about 8 times fewer at A5<sub>N7</sub> (0.098/0.013), and 4 times fewer at T7<sub>O4</sub> (0.016/0.004). Sites on the opposite strand are less affected (A18<sub>N7</sub>, -25%; T20<sub>O4</sub>, -8%) (RDFs and related values are gathered in Figure 10 and in Table 1). While the MG is getting robbed of its ions, there is a build-up at 8oxoG6<sub>O8</sub> and at 8oxoG6<sub>O2P</sub> + T7<sub>O2P</sub> pointing toward the MG side of the double helix (Figure 11A). There is a general slight increase of ion density at the O1P and



**Figure 9.** SDF of  $\text{Na}^+$  (green) and water (red, Ow; white, Hw) around 8oxoG6 (top) and G6 (bottom) viewed from the top, parallel to the helical axis. Iso-surfaces encompass densities higher than 10 and 110 particles/nm<sup>3</sup> for  $\text{Na}^+$  and for water, respectively. Arrows indicate the new and modified hydration sites at 8oxoG6<sub>O8</sub> (red) and 8oxoG6<sub>N7</sub> (blue) and the  $\text{Na}^+$  second coordination shell at G6<sub>N7</sub> and G6<sub>O6</sub> (green).

O2P atoms when one goes from healthy to damaged DNA, yet this increase is more marked at 8oxoG6 and T7. From healthy to damaged, the ratio of  $\text{Na}^+$  occupancies at G/8oxoG6<sub>O2P</sub> + T7<sub>O2P</sub> is 1.75, while it is unchanged at the opposite strand for C19<sub>O2P</sub> + T20<sub>O2P</sub>. In response to the deprivation of ions, the MG is slightly more hydrated. The Ow RDF minimum at A5<sub>N7</sub> and 8oxoG6<sub>O6</sub> also lies lower.

Ions on the opposite side, in the mg, feel the effect of the shuffle of ions in the MG. The growth of ion density at 8oxoG6<sub>O2P</sub> + T7<sub>O2P</sub> disrupts ions in the vicinity of A5<sub>N3</sub>, and the occupancy at 8oxoG6<sub>O1P</sub> + T7<sub>O1P</sub> drops by 10%. The high ion density region lying 4.8 Å from A5<sub>N3</sub> is more developed in damaged DNA (the second peak in healthy DNA corresponds to ions in the MG). Ions are more spread and situated higher above A5 (arrow in Figure 11A). In damaged DNA, this trail of ions stretches toward the center of the mg, repelled by ions closer to the rim of the MG, whereas it essentially follows the path of the backbone in healthy DNA (not clearly visible in Figure 11 due to the high isosurface value). This altered ion distribution in damaged DNA affects hydration; indeed, more ions penetrate the first hydration shell. All longest binding events of Ow in the mg are shorter in damaged DNA. It appears that a proper ion distribution is required to arrange and lock in place water molecules making up the spine of hydration. Circles in the damaged DNA view and corresponding arrows in healthy DNA visible in Figure 11B and C emphasize the void left by one disrupted organized water molecule present in healthy DNA at the level of A5, in the second hydration shell. Water molecules as important as those bridging the two strands, hence co-operatively adding up to native interbase H-bonds and contributing to the stability of the double-helix, are also affected. Those highly structured water molecules called cross-strand bridging water molecules (CSBWs) are fully H-bonded: they “donate” two H-bonds to successive base steps C/T 2-keto(s) and/or A/G N3(s) and “receive” two H-bonds from secondary



**Figure 10.** RDFs of  $\text{Na}^+$  (left panel) and Ow (right panel) around electronegative sites in the MG (top), mg (center), and nonbridging  $\text{PO}_4$  oxygens (bottom) of the damaged and healthy systems.

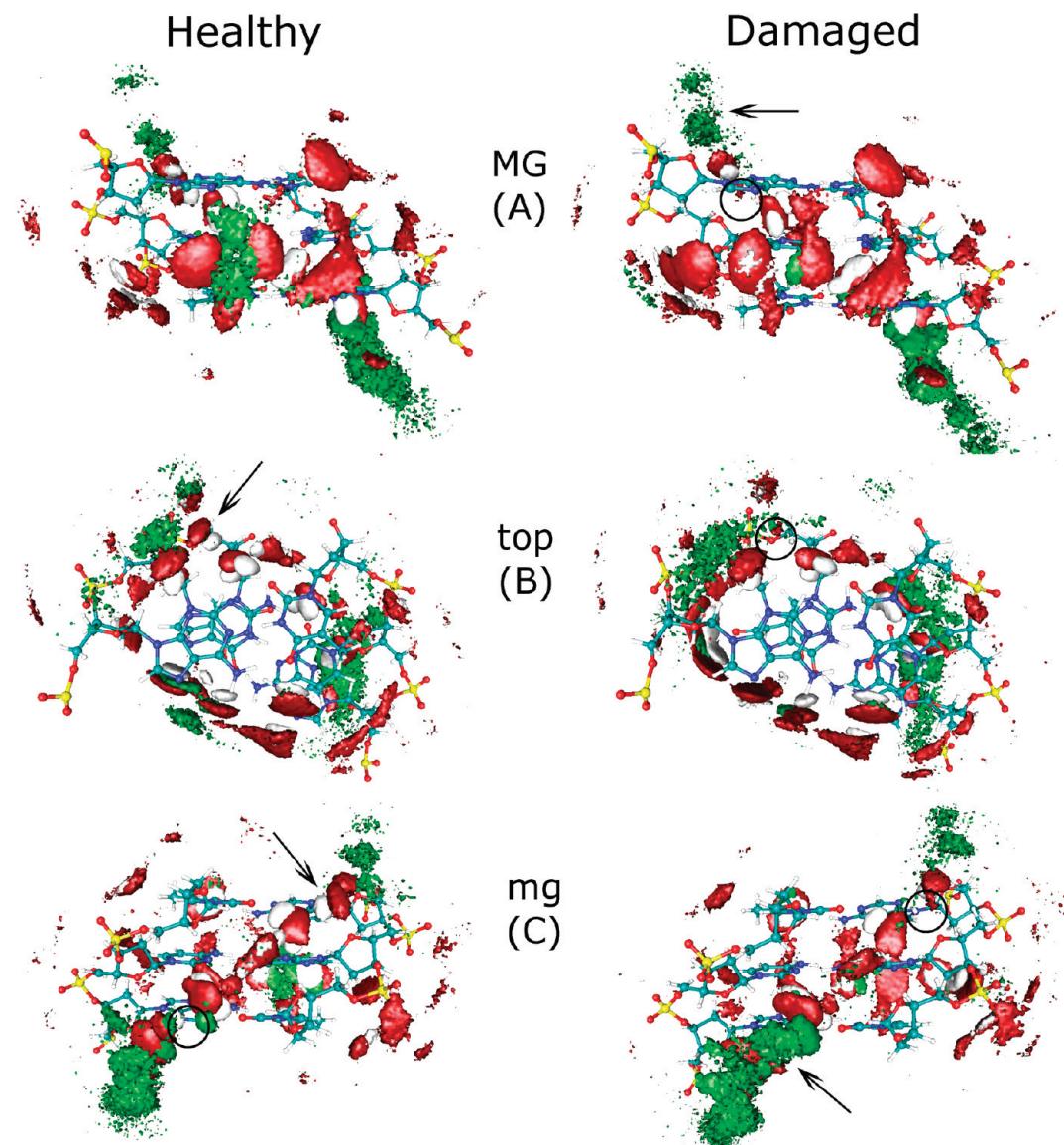
hydration water. An effect of the disorder induced by the more intrusive  $\text{Na}^+$  is visible at the CSBW binding 8oxoG/G6<sub>N3</sub> and T20<sub>O2</sub>. As one of the structured water molecules giving a H-bond to the CSBW in healthy DNA is greatly disturbed (Figure 11 and circle in Figure 12), the CSBW previously locked in position is held more loosely and the H-bond with 8oxoG6<sub>N3</sub> is more fragile. This is shown by the SDF of Hw that takes smaller values there (arrow in Figure 12). The minimum distance between 8oxoG6<sub>N3</sub> and water is shifted toward longer lengths. The minimum distance distribution between T20<sub>O2</sub> and water is similar for both systems (Figure 12). The Ow RDF at 8oxoG6<sub>N3</sub>, occupancies, and residence times confirm that the density is more diffuse, that fewer water molecules (1.02/0.84) are found there, and that they remain for a shorter time ( $t_{\max} = 220$  vs 196 ps). The RDF of  $\text{Na}^+$  around T20<sub>O2</sub> makes the  $\text{Na}^+$  density shift toward the center of the mg perceptible. The second CSBW, H-bonded to T7<sub>O2</sub> and C19<sub>O</sub>, is affected to a lesser extent by the upheaval all around. Its binding to T7<sub>O2</sub> is weakened:  $t_{\max}$  falls from 364 to 168 ps, and the Ow occupancy drops because of enhanced competition with  $\text{Na}^+$ . The minimum distance distribution of Ow is also spread somewhat farther (not shown here).

Another noticeable change in the mg occurs at the level of A18 and C19.  $\text{Na}^+$  ions are forced outside the first hydration shell of C19<sub>O</sub> and concentrate above in the second coordination sphere. This density increase is visible in the  $\text{Na}^+$  RDF at C19<sub>O1P</sub> + T20<sub>O1P</sub>.

Our observations can finally be summarized as follows. In the MG of damaged DNA, ions are moved from the middle to

the rim. The ion density in the mg at the level of the lesion bearing strand becomes more scattered and stretches towards the center of the grooves. It also penetrates deeper what disarranges one well organized hydration site. In turn, the CSBW at 8oxoG6<sub>N3</sub> and T20<sub>O2</sub> becomes destabilized as a consequence of the loss of an H-bond. The second bridging water molecule at T7<sub>O2</sub> and C19<sub>O</sub> is also more labile. The transit time of water in the mg perceptibly drops. The global effect is a partial unstitching and weakening of the spine of hydration and ion condensation in the first and second coordination shells.

**Influence of 8-Oxoguanine on DNA Structure.** The torsion angles  $\varepsilon$  and  $\zeta$  of the sugar-phosphate backbone (Figure 5) are strongly correlated as attested by the clustering in the scattergrams of Figure 13 for the eight central base pairs. The average B<sub>I</sub> conformer population is in the range 80–85% in all simulations compared to the experimental 98 and 87% from NMR and XRD (averaged on all available data for the Dickerson's dodecamer). The values obtained are in agreement with the level observed in 400 ns of Dickerson's dodecamer MD simulations with the Parm99 force field.<sup>44</sup> If one looks at the population in B conformers in the context of the sequence, it appears that the extremities of the dodecamers are more prone to adopt a B<sub>II</sub> conformation. Nucleotides A5 to T8 on the lesion-bearing strand present a low level of B<sub>II</sub> (>15%), similarly do nucleotides C21 to A17 on the complementary strand with the exception of A18 in damaged DNA. One also notices from Figure 14 that a high B<sub>II</sub> population is followed, with few exceptions, by a low B<sub>II</sub> population. Healthy and damaged DNA are very alike in terms of their conformer population on the



**Figure 11.** Views in the *MG* (A), from top (B), and in the *mg* (C) of the central base pairs AXT/TCA ( $X = G$  or 8oxoG) 3D structure for the healthy and damaged systems with the SDF of water (red;  $O_w$ ; white;  $H_w$ ;  $SDF > 110$  particles/ $\text{nm}^3$ ) and  $\text{Na}^+$  (green;  $SDF > 10$  particles/ $\text{nm}^3$ ). The arrows' and circles' signification is explained in text.

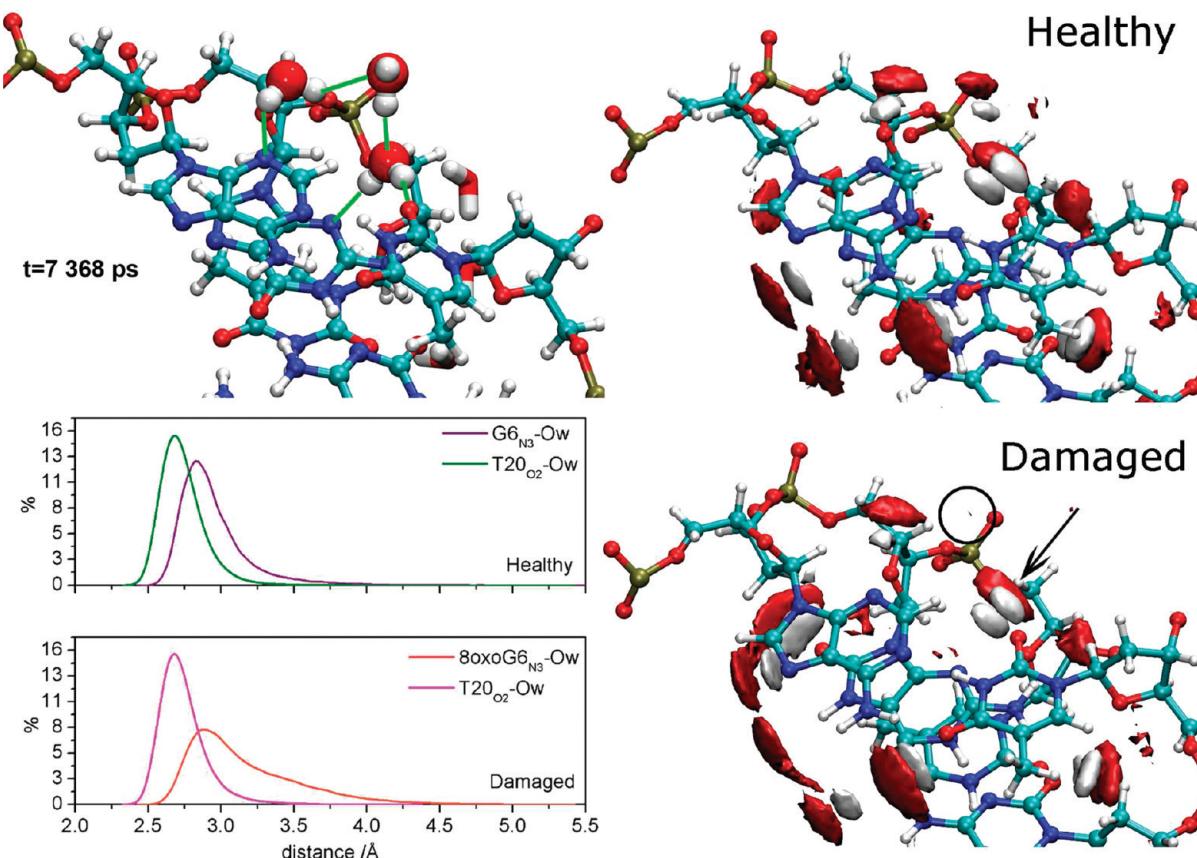
lesion-bearing strand except the case of 8oxoG/G6 which has an increased propensity toward  $B_{II}$  conformations (0.9 vs 12.5%). From a systematic study of all available high resolution DNA XRD structures, it has been shown that  $B_{II}/B_{II}$  clustering is very rare and that the  $B_I$  conformer is by far the most frequent both in naked and complexed DNA.  $B_{II}$  conformers tend to be isolated.<sup>34</sup>

We also checked the  $\alpha/\gamma$  geometries, more to ascertain whether the MD trajectories are well-behaved than for any conclusive purpose. The AMBER forcefields for nucleic acids (Parm94, Parm99) indeed lead to structural deviation artifacts like massive  $\alpha/\gamma$  transitions from the canonical *gauche<sup>-</sup>,gauche<sup>+</sup>* ( $g^-g^+$ ) to the *gauche<sup>+</sup>,trans* ( $g^+$ ) geometry for extended simulation times ( $>20$  ns).<sup>45</sup> Where NMR and XRD data indicate a nearly exclusive  $g^-g^+$  geometry, our simulations achieve an average population of 60–66% similar to other long Parm99 simulations.<sup>44</sup> Nevertheless, the central nucleotides A5–T8 show an acceptable percentage of canonical  $g^-g^+$  geometries. The complementary strand is less satisfactory, though. However, the fact that the population profiles are similar between damaged

and healthy DNA keeps intact the value of a comparative analysis of the structural features.

The three central base pairs were also monitored for the interbase H-bond lengths in Watson–Crick pairs. We consider an H-bond as disrupted when the distance between donor and acceptor atoms exceeds 4.0 Å. We observed more H-bond disruptions in damaged DNA: 872 versus 522 breaking events for healthy DNA. The most affected base pair is A5/T20 for both systems, with 600 and 318 events for damaged and healthy DNA, respectively. It is the site of the lesion that is less affected, with 33 and 42 events. The T7/A18 pair undergoes 249 and 161 events. Those observations are in line with the fact that the spine of hydration is weakened. As reported in other MD simulations, the 8oxoG/C pair shows no special instability in terms of Watson Crick pairing compared to the G/C pair.<sup>46,47</sup>

Minor groove width can be defined as the distance between the phosphorus atom of nucleotide  $i$  and nucleotide  $i+3$  in the opposite strand subtracted by 5.8 Å (twice the van der Waals radius of the  $\text{PO}_4$  group). The mean *mg* widths are alike for both systems. The interesting differences are found in the



**Figure 12.** 3D views of the structure of the three central base pairs magnified in the region of the cross-strand bridging water molecules. Snapshot from the MD simulation of the healthy system (top left). The three structured waters of interest are in space filling representation. Green lines represent H-bonds. Averaged 3D structure of healthy and damaged DNA, respectively, with Ow SDF for densities  $>130$  particles/nm<sup>3</sup> (top and bottom right). Circles in the damaged representation show where structured water molecules lie in healthy DNA. Arrow points at the lowered Hw density vs healthy DNA. Graphs depict the minimum distance distribution between water and the two electronegative sites involved in the cross-strand bridging, T20<sub>O2</sub> and G6<sub>N3</sub>.

standard deviations of the three central base pair step groove widths, A5X6/T20C19, X6T7/C19A18, and T7T8/A18A17 ( $X = G6$  and 8oxoG6), that are larger for damaged DNA, i.e., 40, 25, and 27%, respectively. The more spread groove width distributions for these base pair steps betray a higher flexibility of the backbone in the neighborhood of the lesion, likely due to the loss of the structuring action of the weakened spine of hydration. Note that if the second Na<sup>+</sup> RDF peak for all electronegative sites in the mg (Figure 7, top middle and right) is higher than that for the central base pairs only (Figure 10), this is because those base pairs are situated in an AT rich context (AT tract) that tends to narrow the mg width. The O1Ps capturing ions above the groove are in consequence closer to one another and less ions can accommodate the reduced space.

## Conclusion

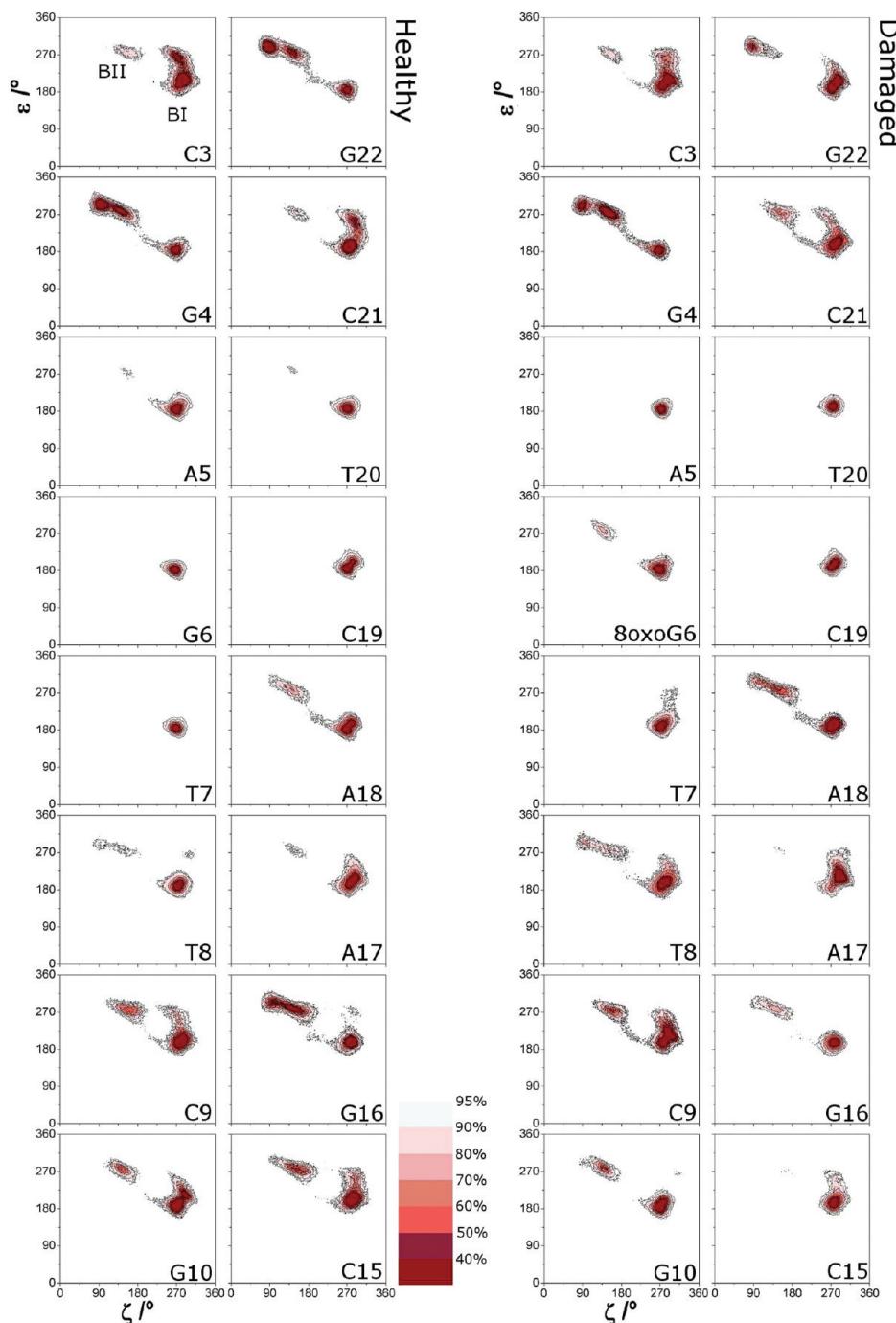
In this work, we have investigated the behavior of water and Na<sup>+</sup> ions around a healthy and a lesion-containing DNA fragment via statistical tools applied to classical MD simulation trajectories. We also explored some conformational aspects of DNA. The lesion consists of an oxidized form of guanine (G), 8-oxoguanine (8oxoG), that is highly mutagenic. The DNA glycosylase hOGG1 takes care of its elimination from human DNA. We proved that the tiny structural difference in 8oxoG underlies significant changes at the level of the Na<sup>+</sup> and water organization, at the scale of the whole fragment as well as at the more local level of the lesion site and its neighboring base pairs. Likewise, we showed that the local conformation of DNA

is sensibly modified in terms of the backbone torsion angles, groove width distributions.

Having recourse to time dependent mean square displacement and self-diffusion coefficients for Na<sup>+</sup>, we have shown that ions are more mobile in the direction parallel to the helical axis compared to the perpendicular directions. This is due to the phosphate groups hampering the escape in the bulk of ions situated in or just above the grooves. They more easily slide on the floor of the grooves or leap above it from phosphate to phosphate. The higher z-direction mobility shown in the damaged DNA system originates in the more fugitive character of the ions on the groove floor and the more numerous ions above the minor groove.

Analysis of SDFs and RDFs, together with occupancies, residence times, and minimum distance distributions, tells us about how Na<sup>+</sup> and water rearrange upon exchange of G to 8oxoG. The implication on DNA stability of the shuffle of ions perturbing water molecules on their way is to be found at the level of the spine of hydration. This one indeed suffers a substantial weakening at the level of the lesion and at the level of the opposite strand, though not as large. The partial unzipping of the water belt might impair the base pairing stability and the double helix stiffness. The solvation pattern in the minor groove of course greatly depends on the electronegative site scaffold, and thus on the sequence, but it is very likely that 8oxoG induces a similar ion shift and structured water disruption in any sequence environment.

The number of breaking events for interbase H-bonds can be interpreted in the frame of the water and Na<sup>+</sup> distributions.



**Figure 13.** Scattergrams of torsion angles  $\epsilon$  and  $\zeta$  for nucleotides in healthy (left) and damaged (right) DNA.

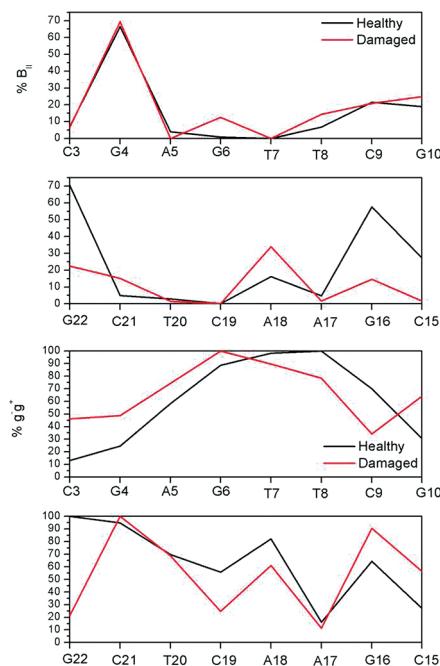
The base pairing destabilization at A5/T20 and T7/A18 in damaged DNA relative to healthy DNA can be related to the altered spine of hydration at those locations.

Minor groove width distributions at the level of the central base pairs display a wider spread in damaged DNA. The increased fluctuations of the minor groove width are likely due to the weakening of the spine of hydration there and the reinforced ion density above the minor groove.

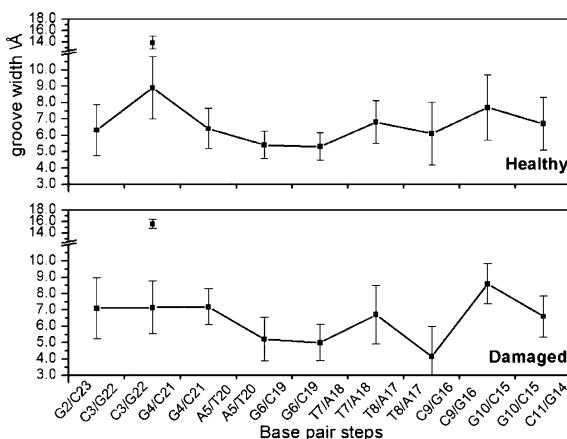
The question addressing whether hOGG1 could locate 8oxoG through an intrahelical recognition mechanism finds here a partial answer. Considering the modifications 8oxoG brings about in DNA conformation, hydration, and ion coordination, as negligible they might be, it is very likely the enzyme builds on them in some way to complete its task successfully. The

protein docks in the minor groove and flips out the lesion in the major groove. The fact that the spine of hydration is weakened at some points in the minor groove could ease intimate contacts to form between hOGG1 and DNA during the docking, and even facilitate base pair opening. This is true on the kinetic point of view, the docking being thermodynamically less favored because of the reduced release of entropy contained in already less structured water molecules. The transition of certain nucleotides around the lesion site from a  $B_I$  to a  $B_{II}$  conformer could also trigger DNA–protein association.

To go further into that questioning, it would be promising to simulate DNA together with the enzyme in solution with ions. For that, a classical all-atom molecular dynamics simulation might however not be the most adequate, as it implies



**Figure 14.** Population profile for  $B_{II}$  and  $\alpha/\gamma$   $g^- g^+$  conformers for healthy (black) and damaged (red) DNA.



**Figure 15.** Profile of  $mg$  width along the healthy (top) and damaged (bottom) DNA fragments sequence. Error bars represent the standard deviation associated to a normal distribution fitted to the groove width distribution.

phenomenological time scales in the order of the microsecond and a much vaster system. Protein and DNA indeed interact through extensive conformational rearrangements that take place in that time range. A coarse-grained-type representation of the supramolecular system is then a good candidate to reach such a goal. We are planning to develop a coarse-grained model that has the crucial characteristic of retaining the sequence details of DNA and proteins needed to study specific DNA–protein association. This will be done using a Newton inversion method developed by one of the authors (Aatto Laaksonen).<sup>48</sup> Observing *in silico* the enzyme interrogating and docking DNA is already within reach.

**Acknowledgment.** All authors acknowledge the Parallel-Datorcentrum (PDC) in Stockholm. The authors also thank the FUNDP, FNRS-F.R.F.C., and the Loterie Nationale (convention 2.4578.02) for the use of the Interuniversity Scientific Computing Facility (iSCF) in Namur. A.N. thanks

the FUNDP for his Erasmus fellowship and the F.R.I.A. for his doctoral scholarship.

## References and Notes

- (1) Lindahl, T. *Nature* **1993**, *362*, 709–715.
- (2) Shibutani, S.; Takeshita, M.; Grollman, A. P. *Nature* **1991**, *349*, 431–434.
- (3) Greenblatt, M. S.; Bennett, W. P.; Hollstein, M.; Harris, C. C. *Cancer Res.* **1994**, *54*, 4855–4878.
- (4) Hainaut, P.; Hernandez, T.; Robinson, A.; Rodriguez-Tome, P.; Flores, T.; Hollstein, M.; Harris, C. C.; Montesano, R. *Nucleic Acids Res.* **1998**, *26*, 205–213.
- (5) Dahle, J.; Brunborg, G.; Svensrud, D. H.; Stokke, T.; Kvam, E. *Cancer Lett.* **2008**, *267*, 18–25.
- (6) Al-Tassan, N.; Chmiel, N. H.; Maynard, J.; Fleming, N.; Livingston, A. L.; Williams, G. T.; Hodges, A. K.; Davies, D. R.; David, S. S.; Sampson, J. R.; Cheadle, J. P. *Nat. Genet.* **2002**, *30*, 227–232.
- (7) Park, J.; Chen, L.; Tockman, M. S.; Elahi, A.; Lazarus, P. *Pharmacogenetics* **2004**, *14*, 103–109.
- (8) David, S. S.; O’Shea, V. L.; Kundu, S. *Nature* **2007**, *447*, 941–950.
- (9) Scherer, O. D.; Jiricny, J. *Bioessays* **2001**, *23*, 270–281.
- (10) Banerjee, A.; Yang, W.; Karplus, M.; Verdine, G. L. *Nature* **2005**, *434*, 612–618.
- (11) Kuznetsov, N. A.; Koval, V. V.; Nevinsky, G. A.; Douglas, K. T.; Zharkov, D. O.; Federova, O. S. *J. Biol. Chem.* **2007**, *282*, 1029–1038.
- (12) Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5752–5757.
- (13) Plum, G. E.; Grollman, A. P.; Johnson, F.; Breslauer, K. J. *Biochemistry* **1995**, *34*, 16148–16160.
- (14) Bhattacharya, P. K.; Cha, J.; Barton, J. K. *Nucleic Acids Res.* **2002**, *30*, 4740–4750.
- (15) Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37–65.
- (16) Stivers, J. T. *Chem.—Eur. J.* **2008**, *14*, 786–793.
- (17) Oda, Y.; Uesugi, S.; Ikebara, M.; Nishimura, S.; Kawase, Y.; Ishikawa, H.; Inoue, H.; Ohtsuka, E. *Nucleic Acids Res.* **1991**, *19*, 1407–1412.
- (18) Lipscomb, L. A.; Peek, M. E.; Morningstar, M. L.; Verghis, S. M.; Miller, E. M.; Rich, A.; Essigmann, J. M.; Williams, L. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 719–723.
- (19) Spyrosakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, N.; Kellogg, G. E.; Mozzarelli, A. *BMC Struct. Biol.* **2007**, *7*, 1–18.
- (20) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (21) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (22) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (23) Ravishankar, G.; Auffinger, P.; Langley, D. R.; Jayaram, B.; Young, M. A.; Beveridge, D. L. Treatment of counterions in computer simulations of DNA. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: 1997; New York, Vol. 11, pp 317–372.
- (24) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (25) Miller, J. H.; Fan-Chiang, C. C. P.; Straatsma, T. P.; Kennedy, M. A. *J. Am. Chem. Soc.* **2003**, *125*, 6331–6336.
- (26) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (30) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (31) Kusalik, P. G.; Svischnev, I. M. *Science* **1994**, *265*, 1219–1221.
- (32) Lyubartsev, A. P.; Laaksonen, A. *Comput. Phys. Commun.* **2000**, *128*, 565–589.
- (33) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Chem. Phys.* **1983**, *87*, 5071–5083.
- (34) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113–124.
- (35) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. *Nucleic Acids Res.* **2008**, *36*, 3690–3706.
- (36) Humphrey, W.; Dalke, A.; Schulter, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (37) Laaksonen, L. *J. Mol. Graphics* **1992**, *10*, 33–34.
- (38) Shakhed, Z.; Guzikovich-Guerstein, G.; Frolov, F.; Rabinovich, D.; Joachimiak, A.; Sigler, P. B. *Nature* **1994**, *368*, 469–473.
- (39) Feig, M.; Pettitt, B. M. *J. Mol. Biol.* **1999**, *286*, 1075–1095.
- (40) *CRC Handbook of Chemistry and Physics*, 86th ed.; Lide, D. R. , Ed.; CRC Press: Boca Raton, FL, 2005.

- (41) Berkowitz, M.; Wan, W. *J. Chem. Phys.* **1987**, *86*, 376–382.  
(42) Lee, S. H.; Rasaiah, J. C. *J. Phys. Chem.* **1996**, *100*, 1420–1425.  
(43) Lyubartsev, A. P.; Laaksonen, A. *J. Phys. Chem.* **1996**, *100*, 16410–16418.  
(44) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.  
(45) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.  
(46) Cheng, X.; Kelso, C.; Hornak, V.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2005**, *127*, 13906–13918.  
(47) Fujimoto, H.; Pinak, M.; Nemoto, T.; Bunta, J. K. *Cent. Eur. J. Phys.* **2007**, *5*, 49–61.  
(48) Lyubartsev, A.; Mirzoev, A.; Chen, L. J.; Laaksonen, A. *Faraday Discuss.* **2010**, *144*, 43–56.

JP1000539