



OPEN

Nucleosomal embedding reshapes the dynamics of abasic sites

Emmanuelle Bignon^{1,2}✉, Victor E. P. Claerbout¹, Tao Jiang¹, Christophe Morell²,
Natacha Gillet¹ & Elise Dumont^{1,3}✉

Apurinic/aprimidinic (AP) sites are the most common DNA lesions, which benefit from a most efficient repair by the base excision pathway. The impact of losing a nucleobase on the conformation and dynamics of B-DNA is well characterized. Yet AP sites seem to present an entirely different chemistry in nucleosomal DNA, with lifetimes reduced up to 100-fold, and the much increased formation of covalent DNA-protein cross-links leading to strand breaks, refractory to repair. We report microsecond range, all-atom molecular dynamics simulations that capture the conformational dynamics of AP sites and their tetrahydrofuran analogs at two symmetrical positions within a nucleosome core particle, starting from a recent crystal structure. Different behaviours between the deoxyribo-based and tetrahydrofuran-type abasic sites are evidenced. The two solvent-exposed lesion sites present contrasted extrahelicities, revealing the crucial role of the position of a defect around the histone core. Our all-atom simulations also identify and quantify the frequency of several spontaneous, non-covalent interactions between AP and positively-charged residues from the histones H2A and H2B tails that prefigure DNA-protein cross-links. Such an *in silico* mapping of DNA-protein cross-links gives important insights for further experimental studies involving mutagenesis and truncation of histone tails to unravel mechanisms of DPCs formation.

Apurinic/aprimidinic (AP) sites are the most frequent spontaneous DNA lesions, amounting to 5000–10,000 lesions per day per mammal cell^{1,2}. They are generated by hydrolysis of the N-glycosylic bond between a nucleobase and a deoxyribose³, spontaneously or as intermediates in the excision repair of damaged bases. Their repair by dedicated enzymes turns out to be most efficient by base excision repair (BER) but also nucleotide excision repair (NER)⁴. Yet accumulation of AP sites, notably triggered by exogenous oxidative stress, represents a threat for genome integrity. Notably AP sites can act as precursors towards much more deleterious lesions such as interstrand cross-links, strand breaks and DNA-protein cross-links (DPCs)⁵, through the reactive aldehyde moiety of their open conformation⁶ (see Fig. 1A). Hence understanding AP sites chemistry and DPCs formation is of utmost importance for the development of anticancer therapies^{7,8}. Structural and thermodynamic features of AP sites as well as their repair have been intensively studied for oligonucleotides and G-quadruplexes^{9–11}: the structural reorganization of the “hole” left within the B-helix by the nucleobase removal and the non-covalent interactions involving the orphan nucleobase on the opposite strand remain to be elucidated. Recent chemical methods have been proposed to sequence them in DNA with single-nucleotide resolution¹². Yet until recently, insights into the chemistry of AP sites within nucleosomal DNA were scarce. Greenberg and coworkers have reported vastly shortened lifetimes for AP sites embedded within nucleosome core particles (NCPs). A recent X-Ray structure, shown in Fig. 1B, has been obtained by Osakabe and coworkers¹³. This structure features a segment of ds-DNA, comprising 145 base pairs (bp) wrapped onto an octameric core of histone proteins: H2A, H2B, H3 and H4. It harbors two tetrahydrofuran analogs of AP site (THF, see Fig. 1A) at symmetrical superhelical location SHL4.5.

Little is known about *in vivo* DPCs involving histone tails as they rapidly undergo hydrolysis towards the formation of strand breaks, rendering them rather unstable and difficult to observe. Nevertheless, the use of experimental techniques such as chromatin immunoprecipitation seems promising for their efficient detection in the native chromatin environment¹⁴. Recently H2B and H2A have been suggested to form Schiff base conjugates with 5-Formylcytosine *in vitro* and *in vivo*^{15,16}, but there is a lack of data about histone cross-linking with AP sites *in vivo*. Histones possess short tails which play a significant role, notably their interactions with AP sites

¹Univ. Lyon, ENS de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F69342 Lyon, France. ²Institut des Sciences Analytiques, UMR 5280, Université de Lyon 1 (UCBL) CNRS, Lyon, France. ³Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France. ✉email: emmanuelle.bignon@univ-cotedazur.fr; elise.dumont@ens-lyon.fr

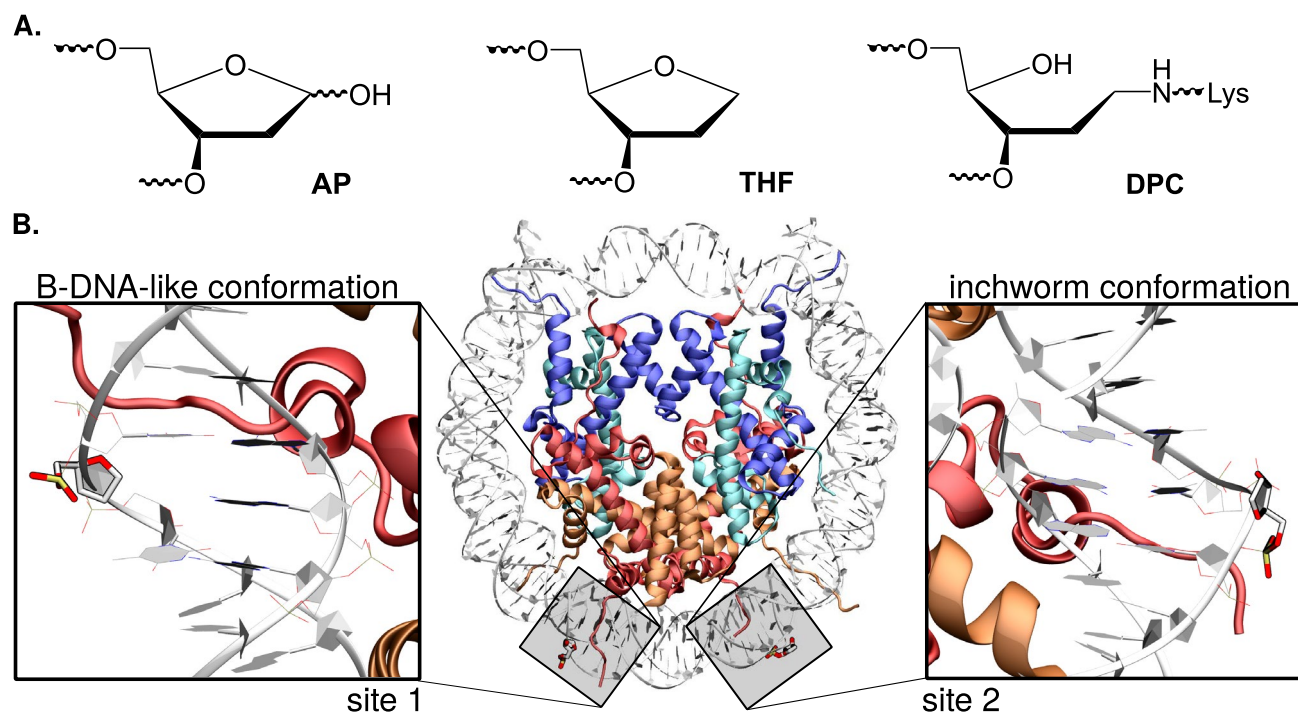


Figure 1. (A) Chemical structure of an AP site, its THF analog, and DNA-protein cross-link (DPC) formed upon reaction between a lysine and an AP site. (B) Cartoon representation of the crystal structure of a nucleosome core particle harboring two symmetrical solvent-exposed tetrahydrofuran residues (sites 1 and 2; PDB ID 5JRG). Site 1 exhibits a classical B-DNA-like conformation, while site 2 adopts the so-called inchworm conformation following the exclusion of THF and of an adenine from the opposite strand. Biomolecule pictures were rendered with VMD 1.9.4a37 (<https://www.ks.uiuc.edu/Research/vmd/>)^{19,20}. All the figures were rearranged using Inkscape 0.91 (<https://inkscape.org/>).

give rise to a much enhanced reactivity compared to naked DNA (100-fold more reactive) depending on the exact positioning of the AP site on the NCP^{17,18}. Therefore, understanding the molecular mechanisms behind such histone-DNA interactions remains a current scientific challenge.

Molecular dynamics has become a dominant method to shed insights on: structural²¹ and mechanical properties²², reactivity and repair of AP sites^{23–25}, and effects of clustered AP sites^{26,27}. A few all-atom molecular dynamics simulations have been performed on NCPs, undamaged^{28–30}, stacked³¹, with double-strand breaks³², or harboring acetylated histone tails³³. As it has become possible to reach the microsecond range for such systems³⁴, they can serve as a computational microscope³⁵ to probe the conformational and mechanical cross-talk between an abasic site and its N-DNA environment. In this study, microsecond range all-atom simulations were performed to unravel the structural outcome and time evolution of AP-containing and THF-containing NCP, starting from the X-ray structure obtained by Osakabe et al.¹³. A different structural behavior exhibited by AP and THF sites within the NCP is detailed, and interactions with residues of the H2A tail are mapped.

Results

The nucleobases numbering used corresponds to the labelling shown in Fig. 2B.

The two abasic site analogs exhibit contrasted dynamical behavior. Molecular modeling techniques were employed to probe the structural behavior of NCP harboring deoxyribose- and tetrahydrofuran-type abasic sites, hereafter referred to as AP and THF. Three unbiased classical MD simulation replicates of 1 μ s each (MD1, MD2, and MD3) were carried out on the NCP structure exhibiting either AP sites or THF analogs at symmetrical superhelical location (SHL) of 4.5, as positioned in the crystal structure by Osakabe et al.¹³. We describe hereafter the local structural features of the two 15-bp sections embedding the damaged sites 1 and 2, as shown in Fig. 2. A first overall measure of the structural impact of an abasic site (or more generally a DNA lesion) on the B-helicity is the bend angle, which was monitored along our MD simulations. For AP sites, the bend angle remains stable around $55 \pm 9^\circ$ in the NCP vs. $27 \pm 13^\circ$ for the control AP-containing sequence (see Figure S1). The bend angle reflects the mechanical constraint exerted along the histone core as the DNA sequence wraps around it³⁶, and is found to be rather constant including in the vicinity of AP sites. However the higher curvature may have an impact on the ease to flip out the AP site or a proximal nucleobase compared to naked DNA. This leads us to investigate more locally the structural distortion in the vicinity of the AP sites.

Our simulations allow us to delineate significant differences of local deformation between AP sites. Extrahelicity of AP sites is an important structural signature that can also be monitored along our simulations^{24,27}. The extrahelicity criteria is based on the distance between the AP site C1' atom and its facing nucleotide C1'

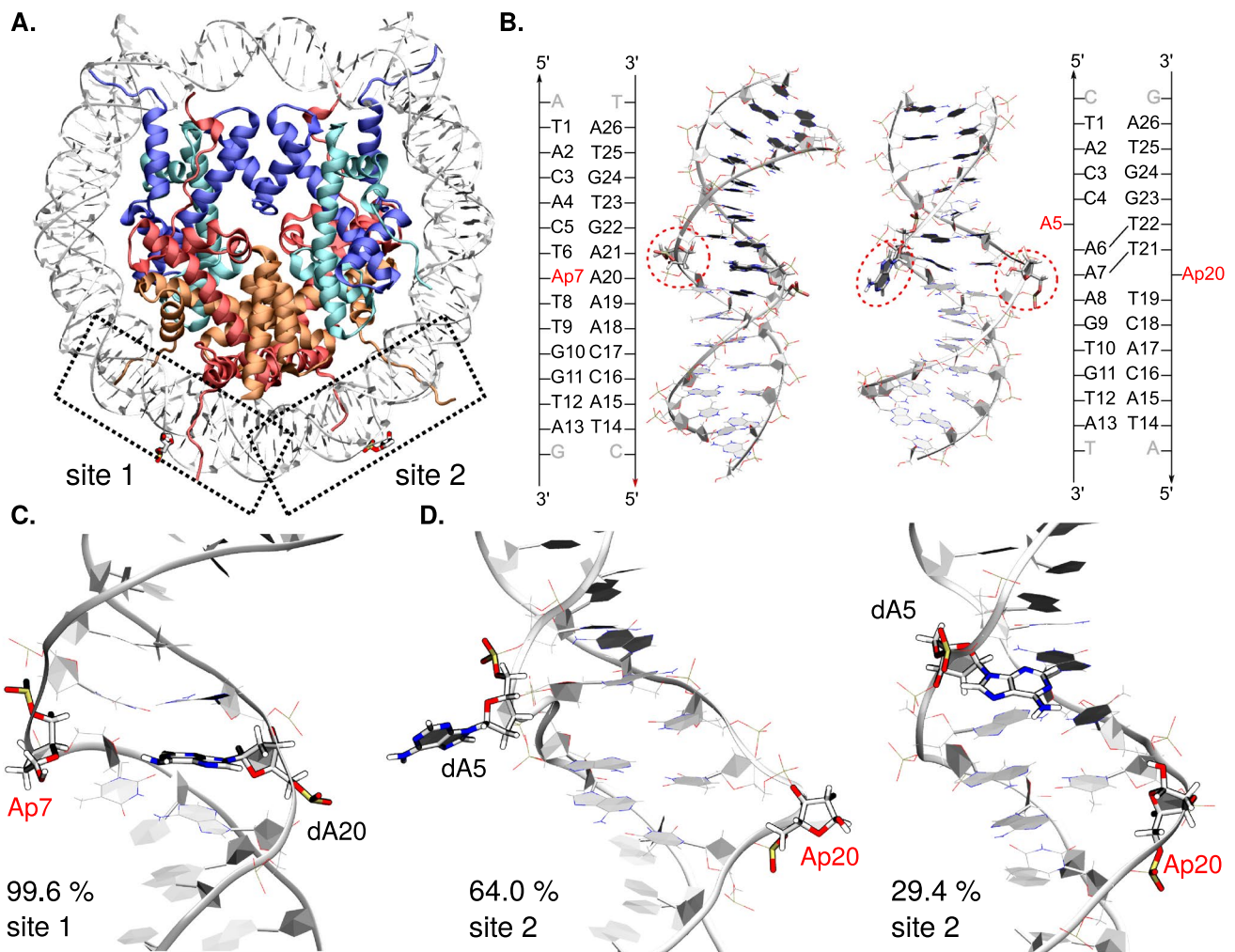


Figure 2. (A) Structure of the THF-containing NCP (PDB ID 5JRG¹³). Histones H2A, H2B, H3, and H4 appear in red, orange, blue, and cyan, respectively. The two symmetric lesion sites are depicted in licorice, site 1 on the left (B-DNA like conformation) and site 2 on the right (in-chworm conformation). (B) Cartoon representation of the NCP 15-bp sections harboring the lesion sites, and their corresponding sequences. The damaged sites (here AP), as well as the in silico modeled dA5 for site 2, are encircled in red. Structural analyses were performed on these 15-bp sections only. Reference MD simulations of naked DNA were performed on the same sequences, with one extra base pair at each extremity to avoid edge effects. (C) and (D) Representative structures of the main clusters and their frequency of occurrence for site 1 (C) and site 2 (D) in AP-containing NCP. The lesion sites, the facing base for site 1 (dA20) and the ejected dA5 for site 2 are depicted in licorice. Conformations sampled for site 1 mostly belong to one dominant cluster (99.6% of the simulation time) while site 2 exhibits two main different conformations (64.0% and 29.4% of the simulation time). Pictures were rendered with VMD 1.9.4a37 (<https://www.ks.uiuc.edu/Research/vmd/>)^{19,20}.

atom. The threshold value above which the AP site is considered as extrahelical is 14 Å²². The extrahelicity is hereafter expressed as a percentage of time, not necessarily consecutive, during which the lesion is found to be extrahelical in a given simulation. The intrinsic propensity for an AP site to flip out of the B-helix within an oligonucleotide is typically of 30%, as is indeed observed in the control 15-bp sequence simulations. Averaged extrahelicity values reported in Table 1 highlight two very different behaviors for AP sites at site 1 and site 2 within the NCP, with respective values of 19.0% and 88.2%. The NCP environment can induce either a slightly lower extrahelical character at site 1 or, on the contrary, an exacerbated flipping of the lesion at site 2. The initial in-chworm conformation of site 2, as well as the ejected dA5 on the facing strand might favor the extrahelical character of the lesion, which is conserved in all our simulations. Noteworthy, the same trend is observed for THF-containing NCP (average average 5.3% site 1 and 85.3% site 2), yet THF exhibits a higher extrahelicity in the control 15-bp (up to 72.8%) suggesting an enhanced lability in naked DNA.

Experimentally, THF is often invoked as an experimental model for deoxyribo-type AP sites, since this analog offers a higher stability^{11,21,37}. Yet it presents chemical and structural differences, with consequences in terms of repair³⁸. Simulations for the THF-containing nucleosome were performed to analyze the dynamic behavior of THF within a NCP. Unlike AP sites, THF moieties lack the chemical features to form hydrogen bonds within

System	All (%)	MD1 (%)	MD2 (%)	MD3 (%)	Control 15-bp (%)
AP (Site 1)	19.0	14.4	12.8	29.7	32.1
AP (Site 2)	88.2	67.0	99.6	97.9	39.1
THF (Site 1)	5.3	3.5	1.0	11.5	42.7
THF (Site 2)	85.3	56.6	99.5	99.7	72.8

Table 1. Averaged extrahelicities over the whole simulation time and detailed values for each 1 μ s MD replicate, for AP and THF analogs at site 1 and site 2 in the NCP and in the 1 μ s 15-bp control simulations.

the double helix. They tend to show a higher extrahelicity than canonical AP sites in the reference 15-bp oligonucleotides but not in the NCP environment—see Table 1.

At site 1 in NCP, THF remains within the double helix, often slightly rotated towards the major groove—Figure S3. No significant reorganization of the surroundings is observed, with the facing dA20 only transiently interacting with dT6 and dT8 flanking THF. Contrary to the control 15-bp, its extrahelicity is found to be very low, i.e. 5.3% in average.

At site 2, the inchworm conformation is well conserved for THF—see Figure S3. dA5 remains ejected and interacts with the H2B tail as seen with deoxyribo-AP—see Figure S5. As dA5 and THF remain ejected from the double helix, a reorganization of the surrounding nucleobases is observed. A strong Watson-Crick pairing and π -stacking is formed between the adjacent base pairs, as could be found in a standard B-DNA conformation. In MD1, however, THF tends to slightly rotate back within the double helix. As a consequence, its extrahelicity drops to 56.6%. Interestingly, the lower extrahelical character of THF does not favor dA5 rotation back within the helix as can be observed with AP sites—see Figure S3.

In the 15-bp control oligomer, the dA-dT base pairs adjacent to the lesion site tend to swing among each other. It triggers the reorganization of the internal Watson-Crick pairing as the abasic or THF site is ejected and re-inserted within the double helix along the simulations. THF can be more labile than AP with an extrahelicity up to 72.8% against 30–40% for AP, unlike what is observed in the NCP embedding—see Table 1. Interestingly, the inchworm conformation is spontaneously observed for site 1 control sequence harboring AP, with the ejection of an adenine two base-pairs above on the facing strand. For the artificial 145-bp DNA sequence from Osakabe and coworkers was chosen as an alternation of dA-dT and dC-dG tetramers, the reorganisation of the dA-dT pairing around the lesion site stays easy. As such, ejection of one nucleobase on the facing strand, concomitantly to an extrahelical lesion site, allows the stacking of the remaining trimer in a canonical B-DNA shape, as was observed for clustered abasic sites in previous works^{21,27}.

Cluster and PCA analyses of the MD ensembles. Our thorough structural analysis along the MD ensembles reveals a complex structural behavior with several regimes for the AP-sites conformation as shown for main representative clusters of the three MD replicates in Fig. 2. At site 1, a characteristic deformation of the backbone induced by AP sites, so-called bulge, is identified. The cluster analysis shows that the conformational behavior of this site is rather monotonous, with an extrahelicity of the main cluster of the three MD replicates up to 77%, 83%, and 99%, respectively—see Figure S3. The AP site is not involved in stable contacts within the double-helix. It is interacting only transiently with the facing dA20 and the adjacent dT6 and dT8. In MD1 and MD3, structural fluctuations sometimes allow weak interactions between dA20 and dT8.

AP at site 2 exhibits a much more complex structural behavior. One can visualize on Fig. 2D the representative conformations of the two main clusters identified from the simulations. Although the main cluster (observed in 64.0% of the simulation time) shows the AP site in its inchworm conformation, the second one reveals a B-DNA-like structure triggered by the rotation of the initially ejected AP and dA5 back within the double-helix. This is especially observed in the MD1 and MD3 replicates as shown in Figure S3. In MD1, the two most representative clusters (representing 43% and 35% of the simulation time) exhibit a marked return of the ejected adenine to its canonical position within the major groove. Similarly, dA5 starts to flip back within the major groove in the last 100 ns of MD3, corresponding to the third cluster (11% of the simulation time). The latter could be more populated if the trajectories were spanning several microseconds, allowing sufficient time for dA5 to fully retrieve its canonical positioning in the double helix. Interestingly, the extrahelical character of the AP site seems to be positively correlated to the rotation of dA5 out of the helix—see Figure S4. The evolution of the distance between centers of mass of dA5 and its complementary nucleobase (dT22) was chosen as a descriptor of dA5 ejection from the double helix along the trajectory. The threshold distance corresponding to dA5 ejection was fixed at a lower bound of 10 Å by observation of the MD conformational ensembles, the reference distance in canonical B-DNA lying at 6.8 Å—see Figure S4. The evolution of the AP site extrahelicity and dA5 ejection along the MD trajectories indeed shows that the adenine dA5 movement back within the double helix is promoted by a prior decrease of AP extrahelicity. In MD1, the overall extrahelicity of the AP site rapidly decreases in the first 100 ns, allowing dA5 to rapidly flip back within the major groove and form stable canonical interactions until the end of the simulation. In MD3, we observe the same trend in the last 100 ns with the AP site extrahelicity drop concomitant to dA5 rotation towards the major groove—see the third cluster structure in Figure S3. The so-called inchworm conformation of AP at site 2, as found in the crystal structure for THF, remains strongly pronounced in MD2 and in the first 350 ns of MD3 gathering most of the population of cluster 1 (48% of the simulation time).

To overcome the tediousness and potential lack of reproducibility of the visual inspection of the AP sites with the nucleosomal DNA environment, we performed PCA on the whole nucleosome core for each MD (see Fig. 3).

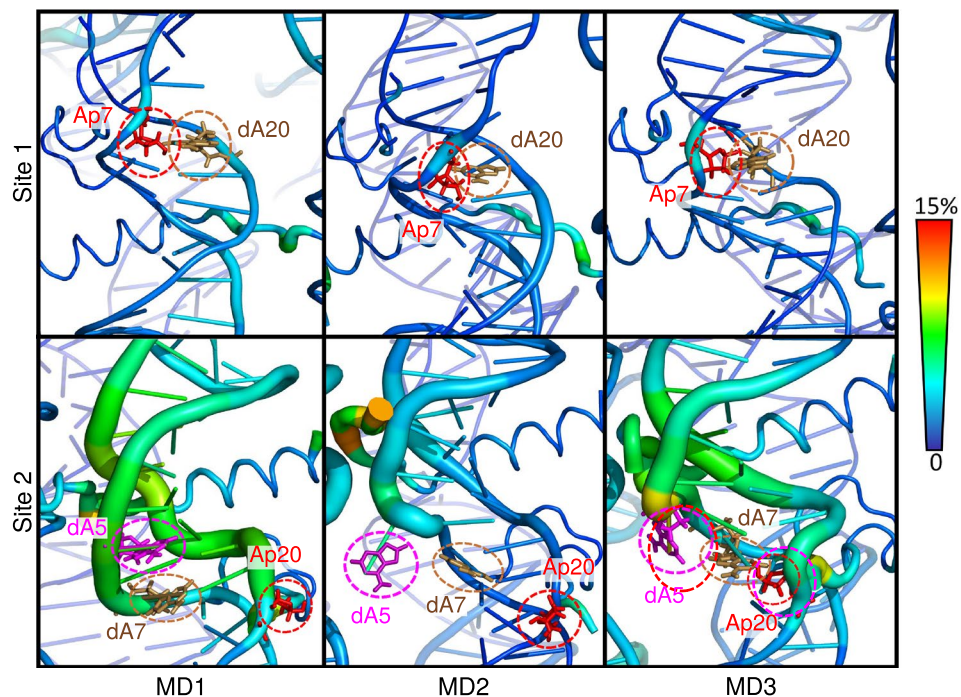


Figure 3. Representation of the per-residue relative contribution to the 10 first principal components around site 1 (top) and site 2 (bottom), detailed for the three MD replicates with AP sites. Only DNA and the N-term tails of H2A and H2B are represented. The percentage of contribution of each residue ranges from 0% (blue) to 15% (red). Abasic sites, its facing orphan base, and the ejected dA5 at site 2 are depicted and encircled in red, magenta and brown, respectively. For a representation of the full NCP, see Figure S6. PCA pictures were rendered using open-source Pymol Version 1.8 (<https://pymol.org>)³⁹.

The overall picture highlights the localization of salt bridges within the histone core or between histones and DNA (see Figure S6). The motion of the histone tails is also captured. Consistently with our previous observations, the inchworm conformation around site 2 is the most important DNA region with respect to PCs, with a predominance of the ejected dA5 and its neighbours. Its relative importance is higher for MD1 and MD3 where dA5 can flip back in the double helix. On the contrary, the contribution of the surrounding region of site 1 is low in all MD (but slightly higher in MD1). However, the contribution of the abasic site itself is similar in both sites (1.9 to 3.6% of the PC total weight), suggesting that its flexibility is not highly correlated to its initial conformation. These observations on the DNA sites can be correlated with the behaviour of H2A and H2B N-term tails: H2A interacts with the abasic site and does not contribute a lot to PCs whereas the H2B tail, in close contact with dA5 and the inchworm region, present a relatively high mobility. However, the per-residue importance of these tails differs depending on the MD simulation and probably the interaction between its numerous positively-charged residues and the DNA.

Interaction with histone tails. Beyond the structural distortion of the DNA duplex, the nucleosomal environment offers a high number of residues prone to interact with the phosphate backbone, as well-known, but also and perhaps competitively with the abasic sites and the orphan nucleobase. Our simulations reveal the mobility of the proximal H2A and H2B N-term tails to the damaged sites located at SHL 4.5, as highlighted by the PCA analysis—see Fig. 3. In-depth inspection of histone tail interactions shows that positively-charged lysines and N-term residues are close to damaged sites, prone to react with AP site towards the formation of DNA-protein cross-links.

Notably, H2A K13 lies within 6 Å of both damaged sites during 6.0–9.2% and 12.2–12.8% of the total simulation time with AP and THF, respectively—see Table 2. Interaction between AP20 and the positively-charged N-term A12 amino acid of the truncated H2A tail is also observed (see Fig. 4), more frequently with the canonical AP site (~20% of the time) than with its THF analog (~7.3% of the time). The H2A tail interacts most of the time in the minor groove nearby both sites, often through insertion of a positively-charged residue, typically R11 two base pairs above site 1 and K13 near site2.

The ejected orphan adenine at site 2 (dA5) also interacts with histone tails. It mainly forms cation-pi interactions with H2B basic residues (K27, K28, R29, R33) and eventually interacts through its amino group with the negatively-charged backbone of the second turn double-helix of DNA—see Figure S5. The nature of the ejected nucleobase might probably not perturb this interaction pattern, for dC, dG and dT can be involved in the same type of contacts as dA. In simulations with AP sites, these interactions fade as the adenine slides back within the DNA helix, which is interestingly not observed with THF. It suggests that the presence of AP sites instead

Interaction	AP (all), %	AP (MD1), %	AP (MD2), %	AP (MD3), %
K13(H2A)-AP7	6.0	4.5	6.1	7.5
K13(H2A)-AP20	9.2	8.3	7.9	11.3
A12(H2A)-AP20	19.9	5.0	30.0	24.8
Interaction	THF (all), %	THF (MD1), %	THF (MD2), %	THF (MD3), %
K13(H2A)-THF7	12.8	1.8	6.2	6.6
K13(H2A)-THF20	12.2	4.8	14.5	17.4
A12(H2A)-THF20	7.3	3.2	10.6	8.2

Table 2. Frequencies of residence of H2A charged amino groups in a radius lower than to 6 Å of the damaged site. Percentages are given for AP and THF in the three 1 μ s MD replicates (MD1, MD2, MD3) as well as averaged over the replicate (all).

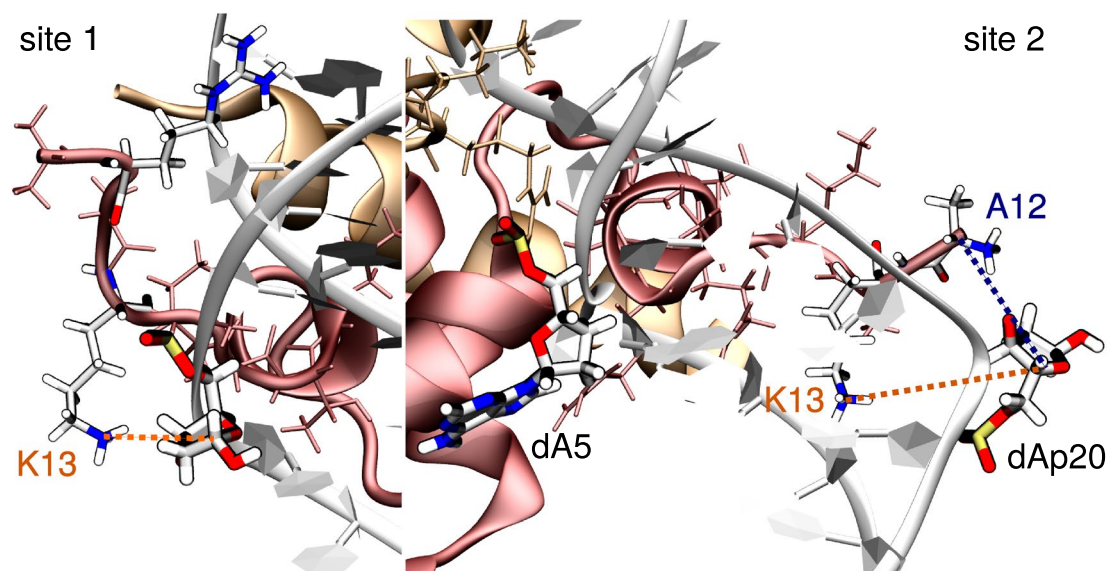


Figure 4. Representation of the main interactions between the damaged site (here AP) and H2A K13 near site 1 (left) and H2A K13 and N-term A12 near site 2 (right). H2B appears in orange and H2A in red, with their tail residues in thin licorice. The abasic sites (dAp7 at site 1 and dAp20 at site 2), the ejected adenine at site 2 (dA5), and the interacting tail residues (K13, N-term A12, and R11) are depicted in thick licorice, colored by element. Pictures were rendered with VMD 1.9.4a37 (<https://www.ks.uiuc.edu/Research/vmd/>)^{19,20}.

of THF destabilizes the inchworm conformation by favoring the B-DNA-like structure and the reorganization of the interaction network.

Discussion

In direct line with the recent experimental X-ray investigation¹³, our simulations allow us to test at the micro-second range the stability on the so-called “inchworm” conformation of tetrahydrofuran-type abasic sites. THF sites are confirmed to be stable along our simulations, ruling out a possible crystal packing effect.

A direct comparison between canonical vs. THF-type abasic sites is absent from the literature, especially within a NCP, yet THF analogs are often invoked experimentally to mimic AP sites^{11,37}. Our MD investigation allows to “restore” a deoxyribo-type abasic site with exactly the same nucleosomal embedding. It also offers a same-footing comparison with control B-DNA 15-bp oligonucleotides. We observe that the nucleosomal environment reshapes the dynamic behavior of the AP site-bearing double helix with respect to the control naked DNA, by imposing structural constraints through strong interactions with the histone octamer. These structural constraints are reflected in the value of the bend angle around the histone core⁴⁰. The bend angle of the damaged NCP oligonucleotide lies at around 55°, consistent with characteristic value of 41.3° calculated at SHL 4.5 from an undamaged NCP crystal structure (PDB ID 1KX5), much higher than in naked B-DNA—see Figure S2.

Formation of an abasic site leaves a “hole” in the double-stranded helix, which can evolve towards flipping out of the AP site and/or of the orphan nucleobase. The structural outcome of the orphan base, here an adenine, is highly driven by sequence effects^{27,41}. The sequence used by Osakabe et al. is a succession of dG-dC and dA-dT tetramers. Consequently, the ejection of the AP-site leads to the rotation of an adenine on the facing strand to allow the stabilization of the tetramer in a canonical trimer-like conformation. This reorganization is observed

at site 2 in NCP, but also in the control 15-bp oligomers simulations. This kind of behavior has previously been reported for clustered abasic sites and THF-bearing B-DNA^{21,27}.

Inspection of the crystallographic data¹³ suggested that site 1 cannot exhibit an inchworm conformation because of the stretching of the double helix around the histone core induced by the helix shrinking at site 2. In our simulation, site 1 indeed remains in a B-DNA-like conformation. Nevertheless, while THF remains in an inchworm-like conformation at site 2, deoxyribo-AP site shows a different structural behavior. Our simulations reveal that the AP site tends to rotate back within the double helix, concomitantly with the ejected dA5 flipping back to its canonical position. According to our PCA analysis, these motions at site 2 are correlated to an enhanced flexibility at site 1. The propensity of dA5 to flip back within the double-helix appears to be favored by the extrahelicity drop of the facing lesion (see Figure S4), but also to its interaction with the histone core. Within THF-harboring NCP, the ejected adenine interacts strongly with the H2B tail and the backbone of the second turn DNA helix of the NCP (Figure S5), the inchworm conformation remaining conserved in our simulations. On the contrary, these interactions are destabilized upon presence of an AP site instead of THF. The additional hydroxyl group of the AP site might enhance interactions with its surroundings. Therefore, by promoting the classical B-DNA conformation against the inchworm one, the presence of the AP site might trigger the destabilization of ejected dA5 initial interactions to favor its rotation back within the double helix—see Figure S5. As revealed by the PCA analysis, the contribution of the adenine in the overall motion of the system is much more pronounced for the simulation in which site 2 retrieves a classical B-DNA conformation (MD1 and MD3), favored by the disruption of dA5 initial interactions with H2B and the proximal DNA backbone.

Interactions between histone tails and abasic sites are also most important, as they can lead to the formation of DNA-protein cross-links via a Schiff base intermediate between AP and lysines or N-terminal residues, ultimately resulting in deleterious strand breaks. It has been experimentally established that abasic sites and oxidized derivatives undergo rapid strand scission at SHL4.5 and SHL4.7 via DPCs formation with histone tails *in vitro*^{42–44}. H2A and H2B have been reported to be responsible for ~ 77% and ~ 12% of the DPCs with AP sites at SHL4.7, respectively⁴⁴. Interestingly, the basic residues of H2B and H2A tails have also been proposed to react *in vitro* and *in vivo* with 5-Formylcytosine^{15,16}. Yet the mapping of the amino acids reacting with AP sites at SHL4.5 have not been performed. Our simulations reveal the close proximity of H2A K13 to the damaged site at this position. THF and AP sites are within 6 Å of either K13 or the positively-charged N-term A12 during up to ~ 20% of the simulation time, suggesting the possibility of DPCs formation. Mapping these interactions offers perspectives for experimental studies involving mutagenesis and/or truncation of the H2A tail to assess DPC formation mechanisms and specificity. Moreover, the histone tails are prone to post-translational modifications such as lysine methylation, acetylation, phosphorylation or even ubiquitination which participate to the regulation of the chromatin compaction or the interaction with other proteins (transcription regulator, DNA repair complex)⁴⁵. These modifications can alter the interaction between the damage and the tail³³, or increase the reactivity of the abasic site⁴⁶ and should also be considered in future works to describe a wider panorama of DPC formations.

Our simulations provide an accurate measure of the contacts occurring between AP sites and positively-charged residues, consistent with *in vitro* data from the literature. This could be generalized to other nucleosomes, notably more canonical sequences such as α -satellite. Indeed the AP-containing nucleosomal structure obtained by X-ray¹³ features a particular repetition of tetramers which was chosen to promote crystallization, but is somehow less representative of canonical N-DNA. Our simulation protocol restores a dynamical view of AP-containing DNA in interaction with a histone core, which remained elusive from the literature. This offers insights at the atomic scale for further experimental studies involving mutagenesis of the histone tails, as it has already been reported for lesions at SHL1.5^{47,48}. We are currently using the same protocol to explore the dynamics of a NCP in the presence of clustered abasic sites, to rationalize the drastic increase in DPCs formation rates compared to single AP-sites⁴⁹. MD simulations provide important insights into the finely-tuned sequence effects and dynamics of AP-sites within NCP, which are of utmost important to unravel the complicated mechanisms of their processing by DNA repair enzymes⁵⁰.

Methods

The crystal structure of a NCP harboring symmetrical THF was used as the starting structure for our models¹³. Two systems were generated from the crystal structure: one with THF as initially in the crystal, the other one with a canonical abasic site. The missing adenine nearby lesion site 2 in the crystal was re-built *in silico* using `xleap`. Crystallographic water molecules and chloride ions were kept, and potassium ions were added to neutralize the system. Protonation states of histone core residues were assigned using `propka3.1`⁵¹, and histidines were protonated on the ϵ position. The two 15-bp sequences harboring each damaged sites (1 and 2) were used to run reference simulations on control naked B-DNA structures with either AP or THF—see Fig. 1. Each system was soaked in a TIP3P water box (truncated octahedron) with a buffer of 10 Å, resulting in ~ 73,500 atoms for NCPs and ~ 25,000–35,000 atoms for the naked 15-bp.

All MD simulations were performed with the Amber and Ambertools 2018 packages⁵². The ff14SB force field was used⁵³ together with the OL15 corrections for nucleic acids^{54,55}. Force field parameters for non-canonical lesions were generated for THF and taken from previous works for AP sites^{24,27}—see Table S1 and Figure S1. The NCP starting structures were carefully optimized through four 10,000 steps energy minimization runs with restraints on the histones and DNA residues, gradually decreased from 20 to 0 kcal mol⁻¹ Å². The system was then heated from 0 to 300 K in a 60 ps thermalization run. The temperature was subsequently maintained at 300 K along the simulation using the Langevin thermostat with a collision frequency γ/\ln of 1 ps⁻¹. The system was then equilibrated during 2 ns in the NPT ensemble. Finally, a one microsecond production run was carried out. For the two NCP systems (with either AP or THF), three simulation replicates (MD1, MD2, MD3) were generated

using different starting velocities. The Particle Mesh Ewald method with a 9.0 Å cutoff was used to account for electrostatic interactions. Production runs for the B-DNA reference structures were 1 μs long. Post-processing and structural analyses were carried out using AmberTools 2018 and Curves+⁵⁶ programs. They were performed on the 13-bp section surrounding the damaged sites as showed in Fig. 1. Analyses of the 15-bp DNA reference simulations were consistently ran on the corresponding 13-bp to avoid any edge effects from the floppy first and last base pairs of the oligonucleotide. Clustering of MD ensembles was carried out according to the RMSD of the 13-bp sequence harboring the damaged site showed in Fig. 1B.

The sheer size of the dimensionality of the data from molecular dynamics simulations often makes it difficult for visual inspection of the trajectory, especially for such all-atom N-DNA trajectories. Principal components analysis (PCA) can help reducing the dimensionality by reconstructing a new configurational space that contains the most important degrees of freedom, providing a better angle for grasping the essence of the chemical process. In practice, the PCA method creates a covariance matrix from the coordinates of the trajectory as input. A set of eigenvectors are then obtained by diagonalizing the covariance matrix, serving as basis of a new configurational space, with each of them being a direction of motion. The magnitudes of the eigenvalues indicates the data variance in each of these directions of motions, the eigenvector with the largest eigenvalue is called the first principal component, and so on. The PCA was performed with a home brew script utilizing the Scikit-learn⁵⁷ library, the internal coordinates (inverse distance between geometric centers of two residues) of the trajectory as input instead of Cartesian coordinates, due to better performance⁵⁸. To obtain the per residue importance, the sum of the weighted principal components up to certain threshold with the corresponding as weights is taken. The sum was then normalized and mapped back to the residues.

Received: 30 April 2020; Accepted: 31 July 2020

Published online: 14 October 2020

References

- Greenberg, M. M. Abasic and oxidized abasic site reactivity in DNA: Enzyme inhibition, cross-linking, and nucleosome catalyzed reactions. *Acc. Chem. Res.* **47**, 646–655. <https://doi.org/10.1021/ar400229d> (2014).
- Wilson, D. M. & Barsky, D. The major human abasic endonuclease: Formation, consequences and repair of abasic lesions in DNA. *Mutat. Res. DNA Repair* **485**, 283–307. [https://doi.org/10.1016/S0921-8777\(01\)00063-5](https://doi.org/10.1016/S0921-8777(01)00063-5) (2001).
- Drohat, A. C. & Maiti, A. Mechanisms for enzymatic cleavage of the n-glycosidic bond in DNA. *Org. Biomol. Chem.* **12**, 8367–8378. <https://doi.org/10.1039/C4OB01063A> (2014).
- Kitsera, N., Rodriguez-Alvarez, M., Emmert, S., Carell, T. & Khobta, A. Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res.* **47**, 8537–8547. <https://doi.org/10.1093/nar/gkz558> (2019).
- Dutta, S., Chowdhury, G. & Gates, K. S. Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc.* **129**, 1852–1853. <https://doi.org/10.1021/ja067294u> (2007).
- Rachofsky, E. L., Seibert, E., Stivers, J. T., Osman, R. & Ross, J. B. A. Conformation and dynamics of abasic sites in DNA investigated by time-resolved fluorescence of 2-aminopurine. *Biochemistry* **40**, 957–967. <https://doi.org/10.1021/bi001665g> (2001).
- Kidane, D., Murphy, D. L. & Sweasy, J. B. Accumulation of abasic sites induces genomic instability in normal human gastric epithelial cells during helicobacter pylori infection. *Oncogenesis* **3**, e128 (2014).
- Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol.* **18**, 563 (2017).
- Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F. & Breslauer, K. J. Thermodynamic consequences of an abasic lesion in duplex DNA are strongly dependent on base sequence. *Biochemistry* **37**, 7321–7327. <https://doi.org/10.1021/bi9803372> (1998).
- Boiteux, S. & Guillet, M. Abasic sites in DNA: Repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* **3**, 1–12. <https://doi.org/10.1016/j.dnarep.2003.10.002> (2004).
- Esposito, V. *et al.* Effects of abasic sites on structural, thermodynamic and kinetic properties of quadruplex structures. *Nucleic Acids Res.* **38**, 2069–2080. <https://doi.org/10.1093/nar/gkp1087> (2009).
- Liu, Z. J., Cuesta, S. M., van Delft, P. & Balasubramanian, S. Sequencing abasic sites in DNA at single-nucleotide resolution. *Nat. Chem.* **11**, 629–637 (2019).
- Osakabe, A. *et al.* Polymorphism of apyrimidinic DNA structures in the nucleosome. *Sci. Rep.* **7**, 41783 (2017).
- Visa, N. & Jordan-Pla, A. Chip and chip-related techniques: expanding the fields of application and improving chip performance. In *Chromatin Immunoprecipitation*, 1–7 (Springer, Berlin, 2018).
- Raiber, E.-A. *et al.* 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* **10**, 1258–1266 (2018).
- Ji, S., Shao, H., Han, Q., Seiler, C. L. & Tretyakova, N. Y. Reversible DNA–protein cross-linking at epigenetic DNA marks. *Angew. Chem. Int. Ed.* **56**, 14130–14134 (2017).
- Banerjee, S. *et al.* Probing enhanced double-strand break formation at abasic sites within clustered lesions in nucleosome core particles. *Biochemistry* **56**, 14–21. <https://doi.org/10.1021/acs.biochem.6b01144> (2017).
- Szczepanski, J. T., Wong, R. S., McKnight, J. N., Bowman, G. D. & Greenberg, M. M. Rapid DNA–protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proc. Natl. Acad. Sci.* **107**, 22475–22480. <https://doi.org/10.1073/pnas.1012860108> (2010).
- Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
- Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation*. Master's thesis, Computer Science Department, University of Missouri-Rolla (1998).
- Barsky, D., Foloppe, N., Ahmadi, S., Wilson, D. M. III. & MacKerell, A. D. Jr. New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.* **28**, 2613–2626. <https://doi.org/10.1093/nar/28.13.2613> (2000).
- Ayadi, L., Coulombeau, C. & Lavery, R. The impact of abasic sites on DNA flexibility. *J. Biomol. Struct. Dyn.* **17**, 645–653. <https://doi.org/10.1080/07391102.2000.10506555> (2000).
- Amara, P., Serre, L., Castaing, B. & Thomas, A. Insights into the DNA repair process by the formamidopyrimidine-DNA glycosylase investigated by molecular dynamics. *Protein Sci.* **13**, 2009–2021. <https://doi.org/10.1110/ps.04772404> (2004).
- Gattuso, H. *et al.* Repair rate of clustered abasic DNA lesions by human endonuclease: Molecular bases of sequence specificity. *J. Phys. Chem. Lett.* **7**, 3760–3765. <https://doi.org/10.1021/acs.jpcclett.6b01692> (2016).
- Bignon, E., Dršata, T., Morell, C., Lankaš, F. & Dumont, E. Interstrand cross-linking implies contrasting structural consequences for DNA: Insights from molecular dynamics. *Nucleic Acids Res.* **45**, 2188–2195 (2017).
- Fujimoto, H. *et al.* Molecular dynamics simulation of clustered DNA damage sites containing 8-oxoguanine and abasic site. *J. Comput. Chem.* **26**, 788–798. <https://doi.org/10.1002/jcc.20184> (2005).

27. Bignon, E. *et al.* Correlation of bistranded clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion. *Nucleic Acids Res.* **44**, 8588–8599. <https://doi.org/10.1093/nar/gkw773> (2016).
28. Bishop, T. C. Molecular dynamics simulations of a nucleosome and free DNA. *J. Biomol. Struct. Dyn.* **22**, 673–685. <https://doi.org/10.1080/07391102.2005.10507034> (2005).
29. Roccatano, D., Barthel, A. & Zacharias, M. Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* **85**, 407–421. <https://doi.org/10.1002/bip.20690> (2007).
30. Pasi, M., Angelov, D., Bednar, J., Dimitrov, S. & Lavery, R. Extra views on structure and dynamics of DNA loops on nucleosomes studied with molecular simulations. *Nucleus* **7**, 554–559. <https://doi.org/10.1080/19491034.2016.1260800> (2016).
31. Saurabh, S., Glaser, M. A., Lansac, Y. & Maiti, P. K. Atomistic simulation of stacked nucleosome core particles: Tail bridging, the h4 tail, and effect of hydrophobic forces. *J. Phys. Chem. B* **120**, 3048–3060. <https://doi.org/10.1021/acs.jpcc.5b11863> (2016).
32. Cleri, F., Landuzzi, F. & Blossey, R. Mechanical evolution of DNA double-strand breaks in the nucleosome. *PLoS Comput. Biol.* **14**, 1–24. <https://doi.org/10.1371/journal.pcbi.1006224> (2018).
33. Fu, I., Cai, Y., Geacintov, N. E., Zhang, Y. & Brody, S. Nucleosome histone tail conformation and dynamics: Impacts of lysine acetylation and a nearby minor groove benzo[a]pyrene-derived lesion. *Biochemistry* **56**, 1963–1973. <https://doi.org/10.1021/acs.biochem.6b01208> (2017).
34. Pérez, A., Luque, F. J. & Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **45**, 196–205. <https://doi.org/10.1021/ar2001217> (2012).
35. Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G. & Schulten, K. *Structure*. **17**, 1295–1306 (2009).
36. Pasi, M. & Lavery, R. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res.* **44**, 5450–5456. <https://doi.org/10.1093/nar/gkw293> (2016).
37. Takeshita, M., Chang, C. N., Johnson, F., Will, S. & Grollman, A. P. Oligodeoxynucleotides containing synthetic abasic sites model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* **262**, 10171–10179 (1987).
38. Otsuka, C. *et al.* Difference between deoxyribose- and tetrahydrofuran-type abasic sites in the in vivo mutagenic responses in yeast. *Nucleic Acids Res.* **30**, 5129–5135. <https://doi.org/10.1093/nar/gkf666> (2002).
39. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8 (2015).
40. Dršata, T. *et al.* Mechanical properties of symmetric and asymmetric DNA A-tracts: Implications for looping and nucleosome positioning. *Nucleic Acids Res.* **42**, 7383–7394. <https://doi.org/10.1093/nar/gku338> (2014).
41. Ayadi, L., Coulombeau, C. & Lavery, R. Abasic sites in duplex DNA: Molecular modeling of sequence-dependent effects on conformation. *Biophys. J.* **77**, 3218–3226. [https://doi.org/10.1016/S0006-3495\(99\)77152-3](https://doi.org/10.1016/S0006-3495(99)77152-3) (1999).
42. Zhou, C., Szczepanski, J. T. & Greenberg, M. M. Histone modification via rapid cleavage of c4-oxidized abasic sites in nucleosome core particles. *J. Am. Chem. Soc.* **135**, 5274–5277 (2013).
43. Zhou, C. & Greenberg, M. M. Histone-catalyzed cleavage of nucleosomal DNA containing 2-deoxyribonolactone. *J. Am. Chem. Soc.* **134**, 8090–8093 (2012).
44. Szczepanski, J. T., Zhou, C. & Greenberg, M. M. Nucleosome core particle-catalyzed strand scission at abasic sites. *Biochemistry* **52**, 2157–2164 (2013).
45. Bowman, G. D. & Poirier, M. G. Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **115**, 2274–2295. <https://doi.org/10.1021/cr500350x> (2015).
46. Yang, K., Prasse, C. & Greenberg, M. M. Effect of histone lysine methylation on DNA lesion reactivity in nucleosome core particles. *Chem. Res. Toxicol.* <https://doi.org/10.1021/acs.chemrestox.9b00049> (2019).
47. Wang, R., Yang, K., Banerjee, S. & Greenberg, M. M. Rotational effects within nucleosome core particles on abasic site reactivity. *Biochemistry* **57**, 3945–3952 (2018).
48. Greenberg, M. M. Abasic and oxidized abasic site reactivity in DNA: Enzyme inhibition, cross-linking, and nucleosome catalyzed reactions. *Acc. Chem. Res.* **47**, 646–655 (2014).
49. Yang, K. & Greenberg, M. M. Enhanced cleavage at abasic sites within clustered lesions in nucleosome core particles. *ChemBioChem* **19**, 2061–2065 (2018).
50. Singh, V., Kumari, B. & Das, P. Repair efficiency of clustered abasic sites by ape1 in nucleosome core particles is sequence and position dependent. *RSC Adv.* **5**, 23691–23698. <https://doi.org/10.1039/C4RA17101B> (2015).
51. Søndergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p k a values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
52. Case, D. *et al.* Amber 2018: San Francisco (2018).
53. Maier, J. A. *et al.* ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
54. Wang, J., Cieplak, P. & Kollman, P. A. 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 (2000).
55. Zgarbová, M. *et al.* Refinement of the sugar-phosphate backbone torsion beta for amber force fields improves the description of z- and b-DNA. *J. Chem. Theory Comput.* **11**, 5723–5736 (2015).
56. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929. <https://doi.org/10.1093/nar/gkp608> (2009).
57. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
58. Fleetwood, O., Kasimova, M. A., Westerlund, A. M. & Delemotte, L. Molecular insights from conformational ensembles via machine learning. *Biophys. J.* <https://doi.org/10.1016/j.bpj.2019.12.016> (2019).

Acknowledgements

The authors thank the SYSPROD project and AXELERA Pôle de Compétitivité for financial support (PSMN Data Center). V.C. is grateful for an ATOSIM Master research grant. This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program ‘Investissements d’Avenir’ (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

Author contributions

E.D. and E.B. conceived the experiments; E.B., V.C., N.G. and T.J. conducted the experiments; E.B., N.G. and T.J. analyzed the experiments; All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-73997-y>.

Correspondence and requests for materials should be addressed to E.B. or E.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020