



**HAL**  
open science

# Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics

Emmanuelle Bignon, Tomáš Dršata, Christophe Morell, Filip Lankaš, Elise Dumont

## ► To cite this version:

Emmanuelle Bignon, Tomáš Dršata, Christophe Morell, Filip Lankaš, Elise Dumont. Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics. *Nucleic Acids Research*, 2017, 45 (4), pp.2188-2195. 10.1093/nar/gkw1253 . hal-01449663

**HAL Id: hal-01449663**

**<https://hal.science/hal-01449663v1>**

Submitted on 30 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interstrand cross-linking implies contrasting structural consequences for DNA: insights from molecular dynamics

Emmanuelle Bignon<sup>1,2</sup>, Tomáš Dršata<sup>3</sup>, Christophe Morell<sup>1</sup>, Filip Lankaš<sup>3,4,\*</sup> and Elise Dumont<sup>2,\*</sup>

<sup>1</sup>Univ Lyon, CNRS, Université Claude Bernard Lyon 1, Ens de Lyon, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, F-69100 Villeurbanne, France, <sup>2</sup>Univ Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F69342, Lyon, France, <sup>3</sup>Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic and <sup>4</sup>Laboratory of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, 166 28 Praha 6, Czech Republic

Received September 06, 2016; Revised November 18, 2016; Editorial Decision November 28, 2016; Accepted December 05, 2016

## ABSTRACT

**Oxidatively-generated interstrand cross-links rank among the most deleterious DNA lesions. They originate from abasic sites, whose aldehyde group can form a covalent adduct after condensation with the exocyclic amino group of purines, sometimes with remarkably high yields. We use explicit solvent molecular dynamics simulations to unravel the structures and mechanical properties of two DNA sequences containing an interstrand cross-link. Our simulations palliate the absence of experimental structural and stiffness information for such DNA lesions and provide an unprecedented insight into the DNA embedding of lesions that represent a major challenge for DNA replication, transcription and gene regulation by preventing strand separation. Our results based on quantum chemical calculations also suggest that the embedding of the ICL within the duplex can tune the reaction profile, and hence can be responsible for the high difference in yields of formation.**

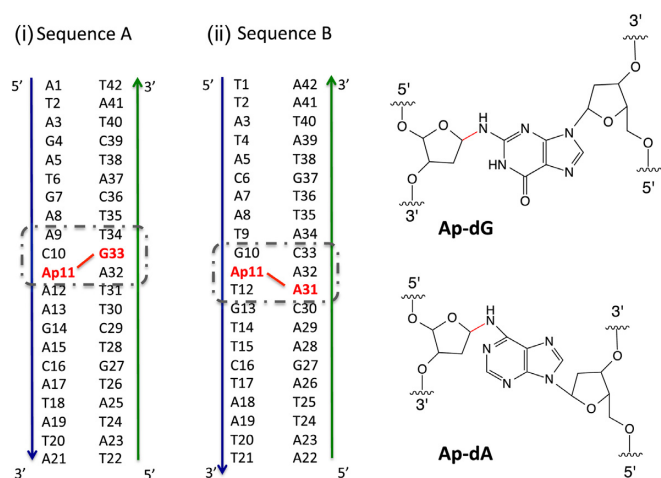
## INTRODUCTION

The large variety of DNA defects that can be formed upon initial attack of reactive oxygen and nitrogen species is indicative of a rich and complex underlying chemistry (1). Amongst the ~80 oxidatively-generated lesions that have been characterized so far, abasic (Ap) sites and 8-oxoguanine are the prominent single-nucleotide defects, but more complex lesions involving two nucleobases that become covalently tethered have also been identified (2) and

are widely investigated (3–11). One possible degradation pathway implies a subsequent reactivity of Ap, which can condensate with the exocyclic amino group of purines. This generates interstrand cross-link (ICL) lesions, which have been firmly identified over the last decade (7,8,12,13): this proves in passing that the B-DNA helix is flexible enough to host the formation of a carbon–nitrogen linkage anchoring the two strands. A single ICL defect can trigger unyielding obstruction to DNA replication, because the strand separation is blocked, and can cause cellular apoptosis (14). Price *et al.* have recently reported (7,8) remarkably high yields (15–70%) for the formation of such defects upon attack of the N<sup>6</sup>-amino group of adenine (Ap-dA or CL(A)). They also evidenced a more rare Ap-dG (or CL(G)) defect (ca. 2–3%), involving the exocyclic N<sup>2</sup>-amino group of guanine (see Figure 1). It is particularly intriguing that such lesions were found to proceed selectively with a purine opposite a 3'-adjacent nucleotide, rather than directly with the orphan adenine (dA32). It was suggested by the authors that the formation of ICLs reflects the ease with which the two partners can approach under the mechanical constraint of a B-helical environment (7). But very recently Greenberg *et al.* have challenged the idea that thermal destabilization can determine the various yields of formation of these highly mutagenetic lesions (15). Covalently tethering the two strands inevitably implies a loss of B-helicity via an alteration of Watson–Crick hydrogen bond (HB) network, possibly ranging over several base pairs, and a redefinition of stacking interactions. The field suffers from the absence of structural data that are crucial to interpret the lack of repair.

Indeed, approaches using synthesized oligonucleotides and nucleotides (16) are not compatible with the produc-

\*To whom correspondence should be addressed. Tel: +33 4 72 72 88 46; Fax: +33 4 72 72 88 60; Email: elise.dumont@ens-lyon.fr  
Correspondence may also be addressed to Filip Lankaš. Tel: +420 220 44 4392; Fax: +420 220 410 320; Email: filip.lankas@vscht.cz



**Figure 1.** The 21-bp DNA sequences A and B featuring an interstrand cross-link between Ap and a purine (displayed with a red line). The yield of formation was experimentally reported to be of 2–3% versus 15–70% for sequences A and B, respectively (7).

tion of milligram quantities or high-quality crystals. Only the chemical structure for the isolated lesion, obtained after enzymatic digestion, can be proposed after careful inspection based on tandem mass spectrometry coupled to high pressure (or high performance) liquid chromatography (13) and/or nuclear magnetic resonance (8). One irremediably loses the structural information concerning the way the ICL redefines the oligonucleotide. A crystal structure of a nucleoside model for the Ap-dG cross-link was recently reported (17), but experimental structures of an ICL within B-DNA are not known.

Moreover, any experimental information concerning the mechanical flexibility or stiffness of these lesions is missing. The peculiar lesion mechanics may be decisive in recognizing the damage for repair. The complex repair mechanisms of ICLs are a subject of intense research (18–20). To palliate the lack of experimental information, molecular modeling can prove useful by analogy to what has been proposed for DNA-drug adducts (21,22) but also for covalently-tethered intrastrand lesions (23–25). In this work, relying on a computational approach, we resolve the structure of two 21-bp oligonucleotides featuring an ICL. We then derive the mechanical properties induced by the cross-link and we determine the stabilization or destabilization of the ICL within the B-DNA local environment. The stabilization energies and entropies can be related to the very different yields of formation of Ap-dA and Ap-dG as the ICL induction obeys a thermodynamic control.

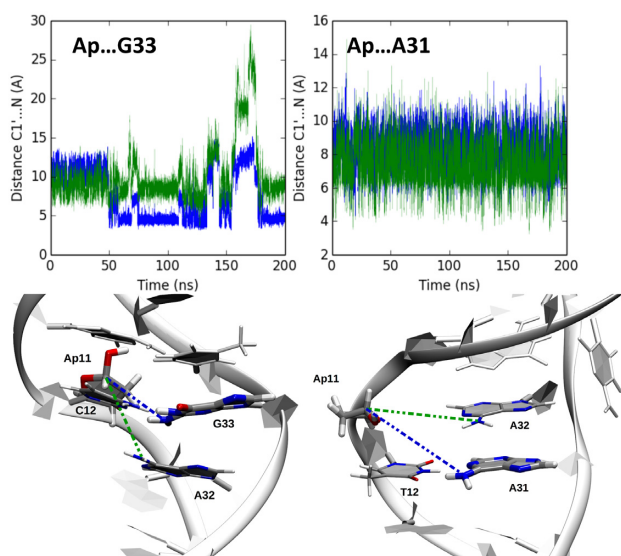
## MATERIALS AND METHODS

We focused on two 21 base pair (bp) duplex DNA sequences, A and B (Figure 1), investigated experimentally by Gates *et al.* (7). All classical molecular dynamics (MD) simulations were performed using the Amber 12 suite of programs (26). We started with reference DNA duplexes featuring an abasic site at the 11th nucleobase position (Ap11). Since no experimental structure has been reported for the ICL-containing duplex DNA, the two duplexes were built

using the nab module of Amber, with conformational parameters characteristic of B-DNA. The parmbsc0 force field was used for DNA, the force field parameters for the Ap site were generated using Antechamber and atomic point charges were assigned using the RESP procedure. They are comparable to parameters similarly derived in the literature (27–29), and we verified that a similar behavior is obtained with the parmbsc1(30) force field, which is the current reference and most tested set of parameters for undamaged DNA (31).

Each duplex was neutralized with 40 potassium ions using Dang parameters, and solvated with TIP3P water molecules in an octahedral simulation box with a 10 Å buffer. After 10 000 steps of energy minimization including 5000 steps of steepest descent and 5000 steps of conjugate gradient, the temperature was increased to 300 K in a 30 ps thermalization run. During the rest of the simulation, the temperature was kept constant using the Langevin thermostat with a collision frequency  $\gamma_{\text{in}}$  of 1 ps<sup>-1</sup>. Then, each system was equilibrated during 1 ns in NPT conditions. Structures obtained after a 200 ns production run provide starting geometries bringing close together the Ap site and the offset purine forming the ICL identified by Gates *et al.* MD simulations were also performed with the open form of the abasic site over 200 ns, which albeit more rare is responsible for the ICL formation alongside a multi-step pathway (32).

As the C1'...N distance gets close to 4–5 Å, it becomes possible to enforce a covalent linkage between Ap11 and G33 or A31 by introducing a harmonic force constant between these two atoms (atom types were changed accordingly still using the parmbsc0 force field, see Supplementary Data). Hybrid quantum mechanics/molecular dynamics (QM/MD), where the ICL is described at the semi-empirical AM1 level, were performed over 1 ns to further validate the classical parameters that were assigned. No strong deviation was observed between the structures obtained by hybrid and unrestrained 500 ns MM-MD simulations (see Supplementary Data). Also the distribution of interatomic distances is comparable at the MM and semi-empirical level. From the final MM-MD structures, a fragment encompassing the ICL lesion and four vicinal nucleobases (as depicted in round dashed box in Figure 1) was extracted, and energies were evaluated within the framework of density functional theory (DFT) at the BLYP-D3BJ/6-31G(d) level of theory. This level of theory has been calibrated on DNA systems featuring non-covalent interactions (33) and in particular the use of a hybrid functional such as B3LYP leads to the formation of spurious hydrogen bonds for the isolated ICL-containing cluster (34). These DFT calculations were performed using the Gaussian09 suite of programs (35). As a control, two B-DNA duplexes analogous to sequences A and B but with no lesion were built, equilibrated and simulated for 200 ns as described above. Representative snapshots were found using the cpp-traj module of AMBER, the rigid base and basepair coordinates were computed using the 3DNA program (36).



**Figure 2.** Cartoon representations for the Ap-containing duplexes before formation of the interstrand cross-link (ICL). For Sequence A, the structures are fully consistent with an attack with an offset, the approaching distances are fully consistent with an attack with an offset, the approaching distances  $d(C1' \dots N)$  being close to 4–5 Å. The green line corresponds to the interatomic distance between Ap/C1' and A32/N6; the blue line refers to the distances between Ap/C1' and G33/N2 or A31/N6.

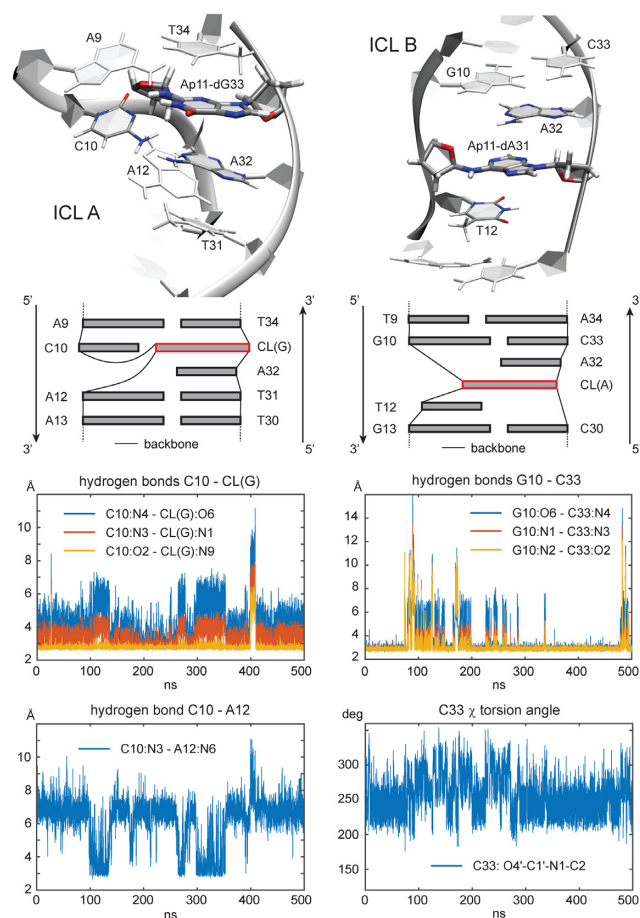
## RESULTS

### Simulations of the Ap-containing duplexes

The simulations of the two Ap-containing sequences allow to monitor the structural changes that take place when a covalent linkage is about to form between the Ap site and the base (G33 or A31, respectively, Figure 1), providing insight into the reactant structures. The structures are flexible enough to undergo structural adaptation, with formation of new hydrogen bond(s) replacing the disrupted Watson–Crick pairing and deviation of the backbone. Figure 2 shows cartoon representations of the damaged sites and time evolutions of atomic distances, indicating that the Ap site seeks to fill up the empty space left by the loss of thymine (27–29). For the oligonucleotide A (left), it can be seen that the distance between the carbon C1' of the AP site and the nitrogen of the amino group of G33 is  $4.5 \pm 0.3$  Å that contrasts to the distance of  $8.4 \pm 0.5$  Å to A32 (blue and green lines, respectively). Our simulations thus corroborate a higher propensity for the ICL to form with a nucleobase adjacent to the orphan base, as reported experimentally (7). For the oligonucleotide B (right), one can see that the Ap site approaches A32 and A31 up to 4–5 Å, yet more transiently, as the distance fluctuates on average around 7.4 and 8.0 Å (green and blue lines, respectively). This behavior holds when the transient open form of the Ap site is considered (see Supplementary Data). In what follows we carefully inspect the structural reorganization and flexibility of the ICL-containing oligomers.

### Simulations of the ICL-containing duplexes

Figure 3 shows atomic structures of parts of Sequences A and B surrounding the lesion, with the covalent linkage al-



**Figure 3.** Cartoon representations and interatomic distances of ICL-containing duplexes A and B. The structures (top) are representative snapshots from the molecular dynamics (MD) conformational ensemble (the pdb files are in Supplementary Data). The figure exposes contrasting structural and dynamical features of the two lesions.

ready present. The structures were obtained as representative snapshots, in the sense of minimal sum of RMSD, from the MD ensembles of the ICL-containing duplexes (the pdb files are in Supplementary Data). Schematic representations and time evolution of selected hydrogen bonds are also shown.

It is seen that for Sequence A (Figure 3 left), the cross-linked site CL(G) forms three Watson–Crick HBs to the C10 base in the opposite strand for most of the time. However, the system occasionally undergoes a conformational switch to an alternative substate in which the C10–CL(G) hydrogen bonding is lost and C10 forms one HB to the A12 base in the same strand instead (around 100 ns and 300 ns). Besides that, there is one short time interval in which C10 loses all hydrogen bonding and tends to flip out of the helix (around 400 ns). We exclude this major, but isolated and short lived perturbation from all subsequent analysis. As for Sequence B (Figure 3 right), there is only one, somewhat unstable Watson–Crick HB triplet, namely that between G10 and C33. Thus, the HB patterns in the two lesions are quite different. The same applies to the stacking patterns: while in Sequence A there is just one unpaired

base (A32) sandwiched between the C10-CL(G) and A12-T31 pairs, the CL(A) in Sequence B forms a stack together with the flanking unpaired bases T12 and A32.

These local differences are reflected in different overall structures and flexibilities of the two damaged sites. In order to quantify them, we define a virtual basepair step consisting of base pairs A9-T34 and A13-T30 for Sequence A, and the analogous pairs T9-A34 and G13-C30 for Sequence B (Figure 3). They are intact WC pairs flanking the lesion. The structure of the lesion as a whole is then defined by the six standard local basepair step coordinates (tilt, roll, twist, shift, slide and rise) associated with the virtual step. By analyzing each snapshot of the MD trajectory using the 3DNA software (36), we obtain time series of the step coordinates. The histograms of the coordinates (Figures S7 and S8 in Supplementary Data) indicate nearly Gaussian distributions. This may be surprising given the non-trivial local conformational dynamics involving two substates in Sequence A and an unstable base pair in Sequence B (Figure 3). A possible explanation is that the global structure results from multiple contributions, both harmonic and anharmonic that, thanks to the Central Limit Theorem, add up to an effective harmonic behavior.

The nearly harmonic probability distribution of the coordinates enables us to apply the model of equilibrium shape and harmonic stiffness (37) to the virtual step. The step coordinates, assembled in a six-dimensional vector  $\mathbf{w}$ , are associated with the effective harmonic deformation energy  $U$  of the form

$$U = \frac{1}{2} (\mathbf{w} - \widehat{\mathbf{w}}) \cdot \mathbf{K} (\mathbf{w} - \widehat{\mathbf{w}}). \quad (1)$$

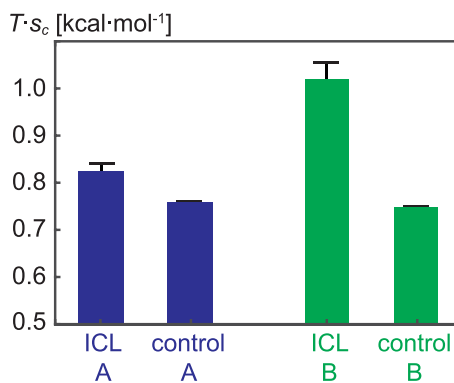
The model contains two sets of parameters, namely the equilibrium coordinate values assembled in vector  $\widehat{\mathbf{w}}$  and the symmetric, positive definite stiffness matrix  $\mathbf{K}$ . These parameters are related to the moments of the coordinate distribution as

$$\widehat{\mathbf{w}} = \langle \mathbf{w} \rangle, \quad \mathbf{K} = k_B T \mathbf{C}^{-1} \quad (2)$$

where  $\langle \mathbf{w} \rangle$  is the coordinate mean (first moment) and  $\mathbf{C}$  is the coordinate covariance matrix (second moment). In our analysis, we estimate the moments using the coordinate time series obtained from the unrestrained MD simulations.

Table 1 shows the equilibrium coordinates of the virtual step in the two damaged sequences together with the same virtual step taken from the control, undamaged DNA duplexes. The structure of ICL A exhibits increased local bending (a tilt of ca.  $-21^\circ$ ) but is otherwise quite similar to the control. In contrast, ICL B affects bending only slightly but shows other important structural deviations from the control. In particular, it is undertwisted by  $20^\circ$ , its lateral displacement of the base pairs (shift and especially slide) change and the vertical displacement (rise) increases by  $>4$  Å, reflecting the peculiar stacking pattern (Figure 3).

The effective mechanical flexibility as defined by the stiffness matrix  $\mathbf{K}$  is also different for the two lesions. Each stiffness matrix contains 21 independent entries. A convenient way to capture the overall stiffness by one single number is to compute the entropy per one basepair step coordinate  $s_c$



**Figure 4.** Entropy per coordinate of the virtual basepair step comprising the lesion, computed from the covariance of the rigid basepair coordinates (Equation (3)). Error bars are mean differences between values for the whole trajectory and for its halves. While the duplex involving the Ap-dG cross-link (ICL A) remains nearly as rigid as a regular B-DNA duplex, the Ap-dG cross-link (ICL B) introduces a flexible hotspot into the helix. More detailed entropy models (Table 3) fully support this finding. The simulation temperature  $T = 300$  K.

(38), given by

$$s_c = \frac{1}{2N} k_B \ln[(2\pi e)^N \det \mathbf{C}] \quad (3)$$

where  $N = 6$  is the number of coordinates,  $e = 2.718\dots$  is the base of the natural logarithm and  $\det$  denotes the matrix determinant. The values of  $s_c$  are shown in Figure 4. While ICL A is only slightly more flexible (has higher entropy) than the control undamaged duplex, ICL B exhibits dramatically higher flexibility than the control. A more detailed insight can be obtained by inspecting the stiffness matrix diagonal. The diagonal entries are stiffness constants for deforming just that one coordinate, while the other coordinates remain unchanged. The values (not shown) reveal that ICL B is more flexible than the control in all the coordinates.

The virtual step model captures the overall flexibility of the ICL site. To obtain more insight, we also consider a more detailed model in which each base within the ICL site (Figure 3) is treated as a rigid body. The configuration is defined by a set of intra-basepair, basepair step and single-strand inter-base coordinates as introduced in the 3DNA program (36). The MD time series of the coordinate vector are then treated exactly as above to obtain the stiffness matrix and the entropy. Here, however, the coordinate distributions are significantly anharmonic in some cases (Supplementary Figure S9), meaning that the use of the harmonic model involves a certain coarse-graining of the probability distribution, namely replacing the actual (possibly complicated) distribution by an effective Gaussian with the same mean and covariance. Finally, we computed the entropy of the ICL sites (Figure 3) using the atomic-resolution quasi-harmonic scheme of Andricioaei and Karplus (39). Notice that this model is also based on the harmonic approximation.

We complement the structural description by measuring the overall bending of the oligomers, using the method described in (40). Briefly, three groups of bases are defined within the oligomer, one at each end and one in the middle.

**Table 1.** Mean basepair step coordinates of the virtual steps comprising the lesions (Figure 3) and the corresponding virtual steps of the control, intact B-DNA. Error estimates in parentheses are mean differences between values for the whole trajectory and for its halves

	tilt (°)	roll (°)	twist (°)	shift (Å)	slide (Å)	rise (Å)
ICL A	−20.7 (0.4)	12.1 (1.4)	125.2 (2.7)	−1.48 (0.14)	−2.47 (0.12)	12.84 (0.08)
control A	−2.6 (0.2)	13.2 (1.1)	125.7 (0.3)	−0.57 (0.13)	−3.11 (0.12)	13.13 (0.01)
ICL B	−1.5 (3.1)	23.6 (7.4)	107.6 (8.6)	−1.54 (0.11)	3.02 (0.24)	17.37 (0.08)
control B	1.5 (0.0)	16.4 (0.2)	127.6 (0.0)	0.12 (0.01)	−3.83 (0.08)	13.09 (0.00)

Base-fixed coordinate frames in each group are averaged to yield one mean frame per group. Bending coordinates in two perpendicular directions (global roll and global tilt) are averaged over the MD trajectory to define the mean bending. It is then expressed in terms of bending magnitude (angle between the z-axes of the end frames) and bending direction angle, defined as the angle between the bending direction and the x-axis of the middle frame pointing into the major groove. Thus, a direction of 0° indicates bending toward the major groove, 180° means bending toward the minor groove in the middle of the oligomer.

Here, we chose bases in pairs 4-5 and 17-18 as the end groups, and bases in the two pairs defining the virtual step (see above) as the middle group. The results are summarized in Table 2. Both control sequences are slightly bent toward the minor groove. Introducing the lesion in Sequence A changes the bend direction toward the backbones, which is consistent with the negative tilt introduced by the lesion (Table 1). The lesion in Sequence B straightens the oligomer. It can be understood as the effect of increased roll (i.e. local bending towards the major groove, Table 1) at the lesion site that compensates for the minor groove bending in the rest of the oligomer.

Finally, we compute global bending and twist stiffness of the damaged and undamaged oligomers. The calculations are done exactly as in (38). Briefly, a quadratic model analogous to Equations (1) and (2) is used, with coordinates involving bending angles toward the grooves and toward the backbone together with total twist. The first two diagonal entries of the  $3 \times 3$  stiffness matrix are stiffness constants for groove and backbone bending, respectively, from which the equivalent isotropic bending stiffness is computed as their harmonic average. The results in Table 2 again indicate exceptionally small stiffness of the oligomer containing ICL B.

Overall, our data indicate that the two lesions exhibit quite different hydrogen bonding and stacking patterns and affect the DNA properties in a different way. In particular, ICL A introduces moderate local bending toward the backbones and is only slightly more flexible than a piece of regular duplex DNA, while ICL B introduces undertwisting as well as a lateral displacement and creates a hotspot of high flexibility.

## DISCUSSION

### Approach between the abasic site and vicinal purines

Our simulation of Ap-containing 21-bp duplexes provide a dynamic view of the approach between Ap11 and the vicinal purines A31 and G33. It differs significantly from the distances inferred from X-ray structures of undamaged

DNA (e.g. PDB ID code 3BSE (7)) or on ‘frozen’ molecular models, which both neglect the rearrangement the duplex undergoes upon formation of an Ap site. Our simulations go beyond these inspections, as we can infer approaching distances between the Ap site and proximal purines, as we also reported for other interstrand cross-links such as dCyd341 (13,41). The two reactants exhibit different eases of approach: Ap11 gets closer to G33 spontaneously within ~50 ns, prefiguring the cross-link Ap-dG. Interestingly, Ap and A31 in sequence B do not approach whereas the corresponding lesion Ap-dA was found experimentally to present a much higher yield of formation.

The selective formation of the ICL Ap-dA31 (15–70%) versus Ap-dA32 hence cannot be explained by the reactant structure. Furthermore, Ap approaches dG33 easily, but its yield of formation is reported to be only 2–3% (7). Our simulations demonstrate that not only the positioning of the two reactive partners within the oligonucleotide but also, and perhaps more importantly, the final structure of the product and its thermal stability determine the ease of formation of the ICL adduct: this also can be proposed since the reaction has been reported to be a reversible process (6).

### Comparison of interstrand cross-links to other lesions

The two interstrand cross-links we characterize by computational means also single out compared to other intrastrand lesions implying a covalent linkage. For instance, we do not observe the exclusion of proximal pyrimidines, in spite of the marked narrowing of the minor groove. The value of the bend angle remains within the thermal fluctuations, again in contrast to what is known for photolesions such as (6-4)PP (42) and intrastrand cross-links (23–25). Other single-nucleotide lesions do not affect the B-helix so markedly. The most distinctive feature of this class of oxidatively-induced lesions is to prevent DNA strand separation: our simulations reveal more finely contrasted structural behavior. We evaluated notably the bending, which can be compared to interstrand lesions induced by cross-linking agents (43): diamminedichloroplatinum induces a bend by 35° to 45° and an unwinding by 79° for the *cis* conformer, but a less marked distortion for the *trans* (unwound by 12° and bent by 26°). Nitrogen mustard only weakly perturbs B-helicity, with a static bend of ca. 14° in DNA, whereas other agents induce no bending such as mitomycin and psoralens. It appears that each interstrand cross-link lesion can be associated to a structural signature, and our simulations for oxidatively-induced ICLs provide such data that are crucially missing.

**Table 2.** Bending magnitudes and directions of the duplexes, their bending stiffness and twist stiffness. The direction of 0° indicates bending toward the major groove in the oligomer centre, 180° means bending into the minor groove. Errors in parentheses were computed as in Table 1

	Bending magnitude (°)	Bending direction (°)	Bending stiffness (nm)	Twist stiffness (nm)
ICL A	15.2 (1.0)	129 (1)	68 (3)	64 (7)
control A	9.1 (1.7)	177 (3)	69 (2)	99 (2)
ICL B	2.6 (3.7)	190 (80)	45 (4)	39 (7)
control B	10.3 (0.5)	204 (2)	77 (3)	109 (2)

### Energetic consequences of the formation of ICL

It is legitimate to ask whether or not the contrasting structure and flexibility of Sequences A and B may be related to the different yields of ICL formation measured experimentally. Indeed, within the hypothesis of a thermodynamic control, and the ICL formation being a reversible process (6), the ICLs thermal stability can be a decisive factor. We can provide an estimate of the reaction energies based on density functional theory (DFT, see Materials and Methods). To this end we compute the ICL formation (de)stabilization energies  $\Delta E^{ICL} = E_{\text{prod}} - E_{\text{react}}$  between the reactants and the products for a fragment featuring the reactive system and the two vicinal base pairs. They account for +5.4 and -14.5 kcal/mol for Sequences A and B, whereas the values for the isolated Ap...A or Ap...G systems are ca. +6–7 kcal/mol (see Supplementary Figure S1). Hence the proximal nucleobases tune the thermal stability of the products, which in turn affects the energy profile, giving rise to contrasted yields. Indeed, from our unconstrained simulations of the Ap-containing oligonucleotides, the Ap sites spontaneously approach purines on the opposite strand at 4–5 Å (see Figure 2). Our DFT analysis suggests that the thermal stability of products is decisive for the selective formation of ICLs, whereas the reactant valley can be more shallow. Our results reinforce the postulate of a correlation between structural distortion and thermal stabilities of DNA interstrand cross-links (15). These overall energies reflect a subtle interplay between different non-covalent interactions. Ap-dA is characterized as a flexible hotspot with an intrinsic chemical stability, which suggests that this lesion in duplex DNA may have the power to block DNA-processing enzymes involved in transcription and replication (9). The ICL induction is more favored for Ap-dA (Sequence B) than for Ap-dG (Sequence A), as the former is associated with a more favorable reaction energy. Moreover, the Ap-dG ICL is further stabilized by the higher entropy of the lesion site (Table 3). To reach a more quantitative assessment, free energies and sequence effects should be tackled, which constitutes a perspective of this work.

### Toward a cross-talk between simulations and experiments on ICLs

The structural and mechanical properties we derived agree fairly well with available experimental data, but an in-depth understanding and validation of our modellization findings can only arise from additional experiments. Our simulations hence call for an experimental proof: whereas the obtention of nuclear magnetic resonance or X-ray structures would be the most definitive probe and validate the calibration of

the timescale of our simulations, it is a considerably difficult task. Another key information accessible experimentally is the yield of induction of oxidatively-induced ICLs, and also sometimes kinetic measurements (44). Hence, free energies should be determined with a high accuracy to reach a synergetic view with classical MD simulations: this would require accurate free energy calculations, since the scarce data available indicate differences that can be as low as 0.3–1.0 kcal/mol (44). Our DFT cluster analysis is probably only capable to characterize ICLs that differ importantly in energy. An equally important information that is more easily accessible consists of experimental measurements of oligonucleotides flexibility. The method of DNA cyclization can provide the magnitude and direction of bending as well as torsional and bending stiffness of a short, intact DNA sequence motif (45). Thus, if the lesion is incorporated in the middle of a short DNA sequence, then the same method should be able to measure the effect of the lesion on these properties. The measurements can then be directly compared to our computed values in Table 2. Only several copies of the lesion would be needed: for instance, no more than three copies of the sequence motif were used in (45). Measures of AP sites and interstrand cross-link half lifetime is also a valuable experimental information (46) characterizing the contrasted mechanical properties DNA lesions experience once in B-DNA or nucleosomal DNA.

### CONCLUSION

Interstrand cross-links are particularly deleterious lesions, yet a deep understanding of their formation remains to be laid on a firm basis. In this work, explicit-solvent MD simulations combined with quantum chemical DFT calculations were performed to unravel the structure, dynamics and mechanical properties of two experimentally characterized ICL-containing oligonucleotides. The results refine the initial rigid model provided by experimentalists, who invoke a higher reactivity for the purine presenting the shortest proximal approaching distance to the electrophilic Ap site. This criterion implies the selectivity onto a purine adjacent to the orphan nucleobase rather than onto the orphan nucleobase itself, yet is not sufficient to explain the lower yield of Ap-dG versus Ap-dA.

Our simulations allow to build up oligonucleotides containing these two defects, and we evidence contrasted structural and mechanical properties that probably induce a different free energy profile. A natural perspective would be to determine a free energy profile for the multi-step reaction, or alternatively free energies of destabilization. This

**Table 3.** Entropy differences per coordinate with respect to the Control B simulation using the virtual step model (Figure 4), the rigid base model and the atomistic quasi-harmonic model. Errors in parentheses were obtained as in Table 1

	$T\Delta s_c$ (kcal/mol)		
	Virtual step	Rigid base	Atomistic
control A	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
ICL A	0.09 (0.02)	0.10 (0.01)	0.05 (0.00)
ICL B	0.29 (0.03)	0.28 (0.05)	0.13 (0.01)

would corroborate a nascent correlation between structure and thermal destabilization for interstrand cross-links (15).

Furthermore, the contrasting mechanical rigidity of the two lesion sites may affect the supercoiling properties of the damaged DNA and play a role in recognizing the ICLs during the initial stages of their repair. The replication-dependent repair requires one or possibly two replication forks to meet at the ICL and thus assumes a large-scale deformation well beyond the harmonic approximation adopted here. The same is true for the transcription-dependent repair that assumes a transcription bubble. Besides that, however, ICL repair in quiescent DNA also takes place, possibly to relieve the negative impact of the tethered strands on DNA topology and supercoiling (18). In this case, the peculiar stiffness of the lesion site may play a role in the ICL recognition, just as the sequence-dependent stiffness of undamaged DNA does for many DNA-binding proteins. The specific mechanisms of the ICL repair have only just started to be unravelled. A clear and systematic assessment of sequence effects, coupled to free energy calculations, would be helpful and will be a topic for future work.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

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). Calculations were performed using the local HPC resources of PSMN at ENS-Lyon. E.B. is grateful for a PhD fellowship from the French Ministry of Higher Education and Research. T.D. and F.L. were supported by the Grant Agency of the Czech Republic (14-21893S).

## FUNDING

Université de Lyon [LABEX PRIMES (ANR-11-LABX-0063)]; French National Research Agency (ANR) ['Investissements d'Avenir' (ANR-11-IDEX-0007)]; Grant Agency of the Czech Republic [14-21893S]. Funding for open access charge: LABEX PRIMES [ANR-11-LABX0063 of Université de Lyon].

*Conflict of interest statement.* None declared.

## REFERENCES

- Cooke, M.S., Evans, M.D., Dizdaroglu, M. and Lunec, J. (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.*, **17**, 1195–1214.
- Box, H.C., Budzinski, E.E., Dawidzik, J.B., Gobey, J.S. and Freund, H.G. (1997) Free radical-induced tandem base damage in DNA oligomers. *Free Rad. Biol. Med.*, **23**, 1021–1030.
- Dutta, S., Chowdhury, G. and Gates, K.S. (2007) Interstrand cross-links generated by abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **129**, 1852–1853.
- Ding, H., Majumdar, A., Tolman, J.R. and Greenberg, M.M. (2008) Multinuclear NMR and kinetic analysis of DNA interstrand cross-link formation. *J. Am. Chem. Soc.*, **130**, 17981–17987.
- Wang, J., Cao, H., You, C., Yuan, B., Bahde, R., Gupta, S., Nishigori, C., Niedernhofer, L.J., Brooks, P.J. and Wang, Y. (2012) Endogenous formation and repair of oxidatively induced G[8-5m]T intrastrand cross-link lesion. *Nucleic Acids Res.*, **40**, 7368–7374.
- Johnson, K.M., Price, N.E., Wang, J., Fekry, M.I., Dutta, S., Seiner, D.R., Wang, Y. and Gates, K.S. (2013) On the formation and properties of interstrand DNA–DNA cross-links forged by reaction of an abasic site with the opposing guanine residue of 5'-CAP sequences in duplex DNA. *J. Am. Chem. Soc.*, **135**, 1015–1025.
- Price, N.E., Johnson, K.M., Wang, J., Fekry, M.I., Wang, Y. and Gates, K.S. (2014) Interstrand DNA–DNA cross-link formation between adenine residues and abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **136**, 3483–3490.
- Price, N.E., Catalano, M.J., Liu, S., Wang, Y. and Gates, K.S. (2015) Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA. *Nucleic Acids Res.*, **43**, 3434–3441.
- Catalano, M.J., Liu, S., Andersen, N., Yang, Z., Johnson, K.M., Price, N.E., Wang, Y. and Gates, K.S. (2015) Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **137**, 3933–3945.
- Yang, Z., Price, N.E., Johnson, K.M. and Gates, K.S. (2015) Characterization of interstrand DNA–DNA cross-links derived from abasic sites using bacteriophage  $\phi$ 29 DNA polymerase. *Biochemistry*, **54**, 4259–4266.
- Zhang, X., Price, N.E., Fang, X., Yang, Z., Gu, L.-Q. and Gates, K.S. (2015) Characterization of interstrand DNA–DNA cross-links using the  $\alpha$ -hemolysin protein nanopore. *ACS Nano*, **9**, 11812–11819.
- Szczepanski, J.T., Jacobs, A.C., Majumdar, A. and Greenberg, M.M. (2009) Scope and mechanism of interstrand cross-link formation by the C4'-oxidized abasic site. *J. Am. Chem. Soc.*, **131**, 11132–11139.
- Regulus, P., Duroux, B., Bayle, P.-A., Favier, A., Cadet, J. and Ravanat, J.-L. (2007) Oxidation of the sugar moiety of DNA by ionizing radiation or bleomycin could induce the formation of a cluster DNA lesion. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14032–14037.
- Noll, D.M., Mason, T.M. and Miller, P.S. (2006) Formation and repair of interstrand cross-links in DNA. *Chem. Rev.*, **106**, 277–301.
- Ghosh, S. and Greenberg, M.M. (2015) Correlation of thermal stability and structural distortion of DNA interstrand cross-links produced from oxidized abasic sites with their selective formation and repair. *Biochemistry*, **54**, 6274–6283.
- Kamiya, H. (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary. *Nucleic Acids Res.*, **31**, 517–531.
- Catalano, M.J., Rudraraju, V., Barnes, C.L. and Gates, K.S. (2016) Crystal structure of a nucleoside model for the interstrand cross-link



- formed by the reaction of 2'-deoxyguanosine and an abasic site in duplex DNA. *Acta Cryst.*, **E72**, 624–627.
18. Hlavín, E.M., Smeaton, M.B. and Miller, P.S. (2010) Initiation of DNA interstrand cross-link repair in mammalian cells. *Environ. Mol. Mutagen.*, **51**, 604–624.
  19. Clauson, C., Schäfer, O.D. and Niedernhofer, L. (2013) Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012732.
  20. Williams, H.L., Gottesman, M.E. and Gautier, J. (2013) The differences between ICL repair during and outside of S phase. *Trends Biochem. Sci.*, **38**, 386–393.
  21. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., Hynes, J.T. and Lavery, R. (2012) Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.*, **134**, 8588–8596.
  22. Dumont, E. and Monari, A. (2013) Benzophenone and DNA: Evidence for a double insertion mode and its spectral signature. *J. Phys. Chem. Lett.*, **4**, 4119–4124.
  23. Churchill, C. D.M., Eriksson, L.A. and Wetmore, S.D. (2011) Formation mechanism and structure of a guanine–uracil DNA intrastrand cross-link. *Chem. Res. Toxicol.*, **24**, 2189–2199.
  24. Dumont, E., Dršata, T., Fonseca Guerra, C. and Lankaš, F. (2015) Insights into the structure of intrastrand cross-link DNA lesion-containing oligonucleotides: G[8–5m]T and G[8–5]C from molecular dynamics simulations. *Biochemistry*, **54**, 1259–1267.
  25. Churchill, C. D.M., Eriksson, L.A. and Wetmore, S.D. (2016) DNA Distortion Caused by Uracil-Containing Intrastrand Cross-Links. *J. Phys. Chem. B*, **120**, 1195–1204.
  26. Case, D.A., Darden, T.A., Cheatham, T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. et al. (2012) AMBER 12. University of California, San Francisco.
  27. Ayadi, L., Coulombeau, C. and Lavery, R. (1999) Abasic sites in duplex DNA: molecular modeling of sequence-dependent effects on conformation. *Biophys. J.*, **77**, 3218–3226.
  28. Barsky, D., Foloppe, N., Ahmadi, S., Wilson III, D.M. and MacKerell, A.D. Jr (2000) New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.*, **28**, 2613–2626.
  29. Chen, J., Dupradeau, F.-Y., Case, D.A., Turner, C.J. and Stubbe, J. (2008) DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Res.*, **36**, 253–262.
  30. Ivani, I., Dans, P.D., Noy, A., Perez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A. et al. (2016) Parmbsc1: a refined force-field for DNA simulations. *Nat. Methods*, **13**, 55–58.
  31. Galindo-Murillo, R., Robertson, J.C., Zgarbová, M., Šponer, J., Otyepka, M., Jurečka, P. and Cheatham, T.E. (2016) Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.*, **12**, 4114–4127.
  32. Sviatenco, L., Gorb, L., Hovorun, D. and Leszczynski, J. (2012) Interaction of 2-Deoxyadenosine with cis-2-Butene-1,4-dial: Computational approach to analysis of multistep chemical reactions. *J. Phys. Chem. A*, **116**, 2333–2342.
  33. Fonseca Guerra, C., van der Wijst, T., Poater, J., Swart, M. and Bickelhaupt, F.M. (2010) Adenine versus guanine quartets in aqueous solution: dispersion-corrected DFT study on the differences in  $\pi$ -stacking and hydrogen-bonding behavior. *Theor. Chem. Acc.*, **125**, 245–252.
  34. Dupont, C., Patel, C. and Dumont, E. (2011) Improved DFT description of intrastrand cross-link formation by inclusion of London dispersion corrections. *J. Phys. Chem. B*, **115**, 15138–15144.
  35. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A. et al. (2009) *G.A.P. Gaussian~09 Revision D.01*. Gaussian Inc. Wallingford.
  36. Lu, X.-J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
  37. Dršata, T. and Lankaš, F. (2013) Theoretical models of DNA flexibility. *WIREs Comput. Mol. Sci.*, **3**, 355–363.
  38. Dršata, T., Spackova, N., Jurečka, P., Zgarbová, M., Šponer, J. and Lankaš, F. (2014) Mechanical properties of symmetric and asymmetric DNA A-tracts: implications for looping and nucleosome positioning. *Nucleic Acids Res.*, **42**, 7383–7394.
  39. Andricioaei, I. and Karplus, M. (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, **115**, 6289–6292.
  40. Lankaš, F., Spackova, N., Moakher, M., Purevjav, E. and Šponer, J. (2010) A measure of bending in nucleic acids structures applied to A-tract DNA. *Nucleic Acids Res.*, **38**, 3414–3422.
  41. Patel, C., Garrec, J., Dupont, C. and Dumont, E. (2013) What singles out the G[8–5]C intrastrand DNA cross-link? mechanistic and structural insights from quantum mechanics/molecular mechanics simulations. *Biochemistry*, **52**, 425–431.
  42. Kim, J.-K., Patel, D. and Choi, B.-S. (1995) Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6-4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. *Photochem. Photobiol.*, **62**, 44–50.
  43. Noll, D.M., Mason, T.M. and Miller, P.S. (2006) Formation and Repair of Interstrand Cross-Links in DNA. *Chem. Rev.*, **106**, 277–301.
  44. Fridman, A.S., Brabec, V., Haroutiunian, S.G., Wartell, R.M. and Lando, D.Y. (2003) Melting of cross-linked DNA V. Cross-linking effect caused by local stabilization of the double helix. *J. Biomol. Struct. Dyn.*, **20**, 533–545.
  45. Roychoudhury, M., Sitlani, A., Lapham, J. and Crothers, D.M. (2000) Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13608–13613.
  46. Catalano, M.J., Liu, S., Andersen, N., Yang, Z., Johnson, K.M., Price, N.E., Wang, Y. and Gates, K.S. (2015) Chemical structure and properties of interstrand cross-links formed by reaction of guanine residues with abasic sites in duplex DNA. *J. Am. Chem. Soc.*, **137**, 3933–3945.