

# Toxicology Research

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Effect of Base Sequence Context on the Conformational Heterogeneity of Aristolactam-I Adducted DNA: Structural and Energetic Insights into Sequence-Dependent Repair and Mutagenicity

Preetleen Kathuria,<sup>1</sup> Purshotam Sharma<sup>1,2</sup> and Stacey D. Wetmore<sup>1\*</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta Canada T1K 3M4

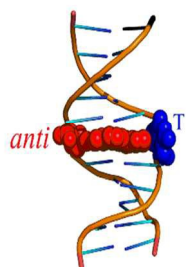
<sup>2</sup>Centre for Computational Sciences, Central University of Punjab, Bathinda, Punjab India 151001

**ABSTRACT:**

Aristolochic acids (AAs) are nephrotoxic and potentially carcinogenic plant mutagens that form bulky DNA adducts at the exocyclic amino groups of the purines. The present work utilizes classical molecular dynamics simulations and free energy calculations to investigate the role of lesion site sequence context in dictating the conformational outcomes of DNA containing ALI-N<sup>6</sup>-dA, the most persistent and mutagenic adduct arising from the AAs. Our calculations reveal that the *anti* base-displaced intercalated conformer is the lowest energy conformer of damaged DNA in all sequence contexts considered (CXC, CXG, GXC and GXG). However, the experimentally-observed greater mutagenicity of the adduct in the CXG sequence context does not correlate with the relative thermodynamic stability of the adduct in different sequences. Instead, AL-N<sup>6</sup>-dA adducted DNA is least distorted in the CXG sequence context, which points toward a possible differential repair propensity of the lesion in different sequences. Nevertheless, the structural deviations between adducted DNA with different lesion site sequences are small, and therefore other factors (such as interactions between the adducted DNA and lesion-bypass polymerases during replication) are likely more important for dictating the observed sequence-dependent mutagenicity of ALI-N<sup>6</sup>-dA.

## Table of Contents Graphic

5'-CGTAMXNATGC-3'  
X = ALI-N<sup>6</sup>-dA, M or N = G or C



base-displaced intercalated

## 1. INTRODUCTION

Aristolochic acids (AAs) I and II (Figure 1a) are plant toxins commonly found in various species of *Aristolochia* and *Aasrum* genera.<sup>1</sup> The consumption of traditional herbal medicines containing extracts of *Aristolochia fangchi* has been linked to a chronic renal disease known as Chinese Herbal Nephropathy<sup>2</sup> (more commonly referred to as Aristolochic Acid Nephropathy (AAN)),<sup>3</sup> which can further develop into urothelial carcinoma.<sup>4-6</sup> In addition, a closely related renal condition called Balkan Endemic Nephropathy (BEN)<sup>7-9</sup> has been associated with the development of upper urinary tract carcinoma,<sup>10</sup> and traced to the consumption of bread prepared from wheat locally grown in the Balkan region and contaminated with seeds of *Aristolochia clematitis*.<sup>11, 12</sup> Based on sufficient evidence of the involvement of AAs in the etiology of AAN, BEN and related carcinogenesis,<sup>13</sup> the International Agency for Research on Cancer (IARC) has categorized herbal medicines containing plant species of the *Aristolochia* genus as Group 1 (human) carcinogens.<sup>14</sup>

The genotoxic mechanism of action of AA-induced carcinogenesis involves the formation of DNA adducts (addition products), which has been supported by their identification in the renal tissue of AA-exposed humans.<sup>12, 15</sup> AAs are metabolically activated in cells to form N-hydroxylaristolactams (N-hydroxyl ALs),<sup>16, 17</sup> which hydrolyze to yield nitrenium ions that react with the amino groups of the purines.<sup>18</sup> Among the resulting AL-N<sup>6</sup>-dA and AL-N<sup>2</sup>-dG adducts, the adenine lesions (Figure 1b) are more persistent and hence more mutagenic in the affected tissues.<sup>19</sup> Furthermore, despite differing by a single methoxy group, the ALI-N<sup>6</sup>-dA adduct is more abundant,<sup>20-22</sup> persistent<sup>23</sup> and nephrotoxic<sup>24</sup> compared to the ALII-N<sup>6</sup>-dA adduct. The mutagenicity of the AL-N<sup>6</sup>-dA adducts has been associated with AA-induced “signature” A → T mutations that are dominant in the TP53 tumor suppressor gene,<sup>25</sup> the FGFR3 and HRAS oncogenes,<sup>26</sup> and T/CXG (X = adduct) trinucleotide sequence motifs within the genome.<sup>27, 28</sup>

In order to understand the structural basis for the observed differential biological outcomes associated with the mutagenic AL-N<sup>6</sup>-dA adducts, it is vital to understand the conformational preferences of (damaged) DNA containing these lesions. In this context, our previous computational work studied both ALI-N<sup>6</sup>-dA and ALII-N<sup>6</sup>-dA in the 5'-CGTACCXCATGC 11-mer sequence,<sup>29</sup> which contains the CXC motif that experiments predict to yield the most stable ALII-N<sup>6</sup>-dA adducted duplexes.<sup>30</sup> Our unrestrained molecular dynamics (MD) simulations revealed that both ALI and ALII-N<sup>6</sup>-dA adducted DNA prefer a base-displaced intercalated structure, which contains an *anti* adduct glycosidic orientation (Figure 1b), displaces the opposing thymine into the major groove, and completely disrupts the lesion site Watson-Crick hydrogen-bonding.<sup>29</sup> These results agree with the most stable ALII-N<sup>6</sup>-dA conformation deduced from NMR based studies in the same sequence,<sup>30</sup> thereby verifying the robustness of our computational approach. Overall, the conformational preferences of the AL-N<sup>6</sup>-dA adducts markedly differ from those of the well-studied N6-linked dA adducts derived from polycyclic aromatic hydrocarbons (PAHs) within the same (CXC) sequence context.<sup>31-35</sup> Specifically, whereas the linker atom in the PAH adducts is tetrahedral, which allows the adduct to maintain either partial or complete hydrogen bonding with the opposing base, the carcinogenic moiety is directly attached to the exocyclic amino group in the AL-N<sup>6</sup>-dA adducts and the lesions cannot form interactions with the opposing base. Thus, the differences in the flexibility within the AL-nucleobase linker bond leads to differences in the intercalated conformations of the PAH and AL-N<sup>6</sup>-dA adducts.

Although the currently available (computational) structural details of ALI-N<sup>6</sup>-dA adducted DNA have only been obtained in a single (CXC) sequence context,<sup>29</sup> previous studies on a variety of carcinogenic adducts, including those formed from aromatic amines,<sup>36-39</sup> PAHs,<sup>35, 40-43</sup> and ochratoxin A,<sup>44, 45</sup> point towards the sequence dependence of the conformational preference of damaged DNA. Specifically, changes in the sequence at the lesion site can affect key structural features and alter the stability of the preferred conformational theme of adducted DNA.<sup>35, 37, 38, 40, 44</sup> For example, although the

10S (+)-*trans-anti-B[a]P-N<sup>2</sup>-dG* PAH adduct preferentially induces a minor groove conformation at G<sup>6</sup> or G<sup>7</sup> in the 5'-CATGCG<sup>6</sup>G<sup>7</sup>CCTAC oligonucleotide, greater distortion occurs with the adduct at G<sup>7</sup> because of intrastrand steric clashes with the 5'-flanking guanine.<sup>40</sup> Furthermore, within the 5'-CTCG<sup>1</sup>G<sup>2</sup>CG<sup>3</sup>CCATC oligonucleotide containing the *NarI* mutational hotspot sequence (underlined),<sup>46</sup> the base-displaced intercalated conformer (the bulky moiety in the *syn* damaged base stacks in the helix and renders the opposing base extrahelical) of AF-C<sup>8</sup>-dG aromatic amine adducted DNA is most stabilized with the lesion at G<sup>3</sup> due to better stacking between the bulky moiety and the flanking bases in the opposing strand than at G<sup>1</sup> or G<sup>2</sup>.<sup>37, 38</sup> Additionally, due to the formation of discrete hydrogen-bonding interactions at the lesion site, DNA containing the (monoanionic) C-linked C8-bonded adduct of dechlorinated ochratoxin A (OTB) and dG (OTB-dG adduct) at G<sup>3</sup> in the *NarI* sequence exhibits greater stabilization of the minor groove conformer compared to the major groove and base-displaced intercalated conformations.<sup>44</sup> In addition to affecting the structure and stability of the preferred conformational theme, changing the flanking bases can even completely alter the conformational preference of damaged DNA. For example, the *NarI* sequence-containing adducted DNA acquires a minor groove conformation when the IQ-C<sup>8</sup>-dG aromatic amine adduct is located at G<sup>1</sup> or G<sup>2</sup>, but a base-displaced intercalated conformation with the lesion at G<sup>3</sup>.<sup>36</sup>

Since changing the sequence context can alter the preferred conformation of adducted DNA, it is not surprising that the sequence can also influence the biological consequences of adduct formation. Specifically, the sequence context has been shown to play an important role in modulating the repair propensity of DNA lesions.<sup>40-43, 47-52</sup> Bulky DNA lesions, including those formed from aristolochic acids, are most commonly repaired through the nucleotide excision repair (NER) pathway,<sup>53</sup> which may proceed either through a transcription-coupled repair (TCR) or a global genome repair (GGR) mechanism.<sup>54-56</sup> Whereas TCR only deals with transcriptionally-active regions,<sup>56</sup> GGR operates on lesions spanning the entire genome.<sup>55</sup> Furthermore, although lesion recognition in TCR is associated with

inhibition of transcription (stalling of an RNA polymerase), GGR recognition involves specific damage recognition factors (e.g., eukaryotic XPC-RAD23B). Nevertheless, the subsequent repair steps are similar in both mechanisms. Although a number of bulky DNA adducts are commonly repaired by GGR,<sup>57-59</sup> experimental studies have determined that the AL-N<sup>6</sup>-dA adducts are resistant to GGR, and are exclusively repaired by TCR.<sup>60</sup> However, the structural basis of this GGR resistance and the related sequence dependence of the repair propensity of the AL-N<sup>6</sup>-dA adducts is not well understood.

An abundance of experimental and computational evidence has led to proposals that the primary GGR recognition step, including binding of the damage recognition factor and opening of the damaged DNA (flipping of the base opposing the lesion into the active site), involves identification of structural perturbations at the lesion site.<sup>42, 51, 61-68</sup> As a result, the relative repair efficiency of DNA adducts in different sequences can be altered by differences in distortions and/or dynamics at the lesion site. For example, the increased repair susceptibility of the 10S (+)-*trans-anti-B[a]P-N<sup>6</sup>-dA* PAH adduct in the CXC compared to the CXA sequence context has been attributed to the formation of a sequence-specific hydrogen bond that increases the roll and bend at the CXC lesion site.<sup>41</sup> Similarly, the 10S (+)-*trans-anti-B[a]P-N<sup>2</sup>-dG* PAH adduct induces a flexible kink when located at G<sup>7</sup> in 5'-CATGCG<sup>6</sup>G<sup>7</sup>CCTAC due to steric interactions with the 5'-flanking guanine, which leads to better recognition by NER machinery at G<sup>7</sup> compared to G<sup>6</sup>.<sup>40</sup> Alternatively, the 10S (+)-*trans-anti-B[a]P-N<sup>2</sup>-dG* PAH adduct exhibits greater helical dynamics and associated repair propensity in the TXT than CXC sequence due to the reduced intrinsic stability of T:A compared to C:G flanking base pairs, as well as the reduced stability of the T-G compared to the C-G base steps involving damaged G.<sup>42, 47</sup>

In addition to influencing the GGR repair propensity, the identity of the bases flanking the lesion site,<sup>69-71</sup> as well as the next nearest neighbors,<sup>72</sup> has been shown to affect the mutagenic profile of DNA lesions. For example, the mutational frequency of the PhIP-C<sup>8</sup>-dG aromatic amine adduct is greater in

the 5'-TCCTCCTNXCCTCTC oligonucleotide when N = C or G compared to A or T.<sup>71</sup> Additionally, the greater stability of the base-displaced intercalated conformer associated with AF-dG in the CXA over the TXA sequence context causes enhanced misincorporation in the CXA motif.<sup>69</sup> Alternatively, the slippage prone active site of the model translesion synthesis (TLS) polymerase Dpo4 leads to different mutagenic outcomes for the 10S (+)-*trans-anti-B[a]P-N<sup>2</sup>*-dG PAH adduct, which exhibits dGMP insertion opposite the lesion in the CXG sequence context, and dAMP insertion in the TXG motif.<sup>70</sup>

Despite observed variations in adduct conformational preferences, repair propensities, and mutational outcomes with sequence context, it is currently unclear how these effects depend on the chemical structure of the adduct. In the case of the AL-N<sup>6</sup>-dA lesions, a previous study indicates that AAlI-damaged DNA is thermodynamically more stable when the adduct is present in the CXC sequence context compared to T/CXG motifs.<sup>30</sup> Nevertheless, recent studies on patients exposed to AAs have revealed that signature A→T mutations most commonly occur in T/CXG motifs.<sup>27, 28</sup> Unfortunately, no other information is available to date regarding the sequence effects on the stability of AL-adducted DNA or the associated biological consequences of AA-induced damage. For example, the conformation and stability of DNA strands containing AL-N<sup>6</sup>-dA in a purine-X-purine or purine-X-pyrimidine sequence is not currently available. Thus, it is not surprising that the molecular basis of the observed sequence dependence associated with AL-N<sup>6</sup>-dA formation has yet to be explored.

In the present work, we utilize classical (unrestrained) MD simulations and post-processing free energy calculations to investigate the role of sequence context in dictating the conformational outcomes of ALI-N<sup>6</sup>-dA, the most persistent and mutagenic DNA adduct arising from the AAs. Specifically, ALI-N<sup>6</sup>-dA (X) is incorporated into three 11-mer oligonucleotides (i.e., 5'-CGTACXGATGC, 5'-CGTAGXCATGC, and 5'-CGTAGXGATGC), which differ in the identity of the bases flanking the lesion (underlined). Two of the oligonucleotides considered place a purine (G) and pyrimidine (C) on either the 5' or 3' side of the



lesion, while a purine (G) flanks both the 5' and 3' sides of the lesion in the third oligonucleotide. The conformational outcomes within these sequences are compared to our previous study on the 5'-CGTACXCATGC oligonucleotide, which places a pyrimidine (C) on both sides of the lesion.<sup>29</sup> We focus on different purine and pyrimidine sequence contexts generated by varying the combination of G:C flanking base pairs, since a previous experimental study determined that the TXG and CXG motifs containing the related ALII-N<sup>6</sup>-dA adduct have similar stabilities, while the CXC motif leads to the most stable helices.<sup>30</sup> Due to the similarity in the conformational preferences of ALII-N<sup>6</sup>-dA and ALI-N<sup>6</sup>-dA predicted in our previous study,<sup>29</sup> it is reasonable to assume that the sequence-dependent conformational preferences of ALI-N<sup>6</sup>-dA will not be wildly different from ALII-N<sup>6</sup>-dA in the absence of experimental data on the sequence-dependent stability of ALI-N<sup>6</sup>-dA adducted DNA. Our detailed comparison of the preferred conformations of ALI-N<sup>6</sup>-dA adducted DNA in different lesion site sequence contexts provides valuable clues about the origin of the experimentally-observed sequence-dependent strand stability. Furthermore, our work explores the relationship between the sequence context and potential biological outcomes of adduct formation, including the mutagenicity and repair propensity, and thereby provides insight into the sequence-dependent persistence of this potent lesion.

## 2. COMPUTATIONAL DETAILS

The ALI-N<sup>6</sup>-dA adduct was considered in four 11-mer DNA sequences: (i) 5'-CGTAGXCATGC, (ii) 5'-CGTAGXGATGC and (iii) 5'-CGTACXGATGC, as well as the previously studied (iv) 5'-CGTACXCATGC (X = adduct).<sup>29</sup> The B-form of canonical DNA was initially built for the 5'-CGTACXCATGC (X = A) sequence using the NAB<sup>73</sup> module of AMBER 11.<sup>74</sup> The ALI moiety was attached to the N6 position of A at the X position using GaussView.<sup>75</sup> The previously characterized lowest energy *anti* and *syn* nucleotide orientations (with  $\theta \sim 0^\circ$ , Figure 1b) were incorporated into DNA,<sup>29</sup> while avoiding steric clashes with the surrounding nucleotides. The damaged base was paired against complementary T. Initial structures for

simulations were built by replacing the flanking base pairs in the representative structures obtained from the simulation trajectories of each of the six conformers belonging to the previously studied 5'–CGTACXCATGC (X = ALI-N<sup>6</sup>-dA) adducted DNA oligonucleotide.<sup>29</sup> All force field parameters, as well as the simulation protocol, were adopted from our previous study of ALI-N<sup>6</sup>-dA adducted DNA.<sup>29</sup> Final unrestrained production MD simulations were subsequently run at 300 K for 20 ns using the PMEMD<sup>76</sup> module of AMBER 12.<sup>77</sup> The choice of room temperature for simulations ensures the viability of the harmonic approximation in the bonded terms of the force field, while remaining close to physiological conditions. The simulation time is justified by our previous study of AL-N<sup>6</sup>-dA adducted DNA in the CXC sequence, which illustrated that extending the simulation time to 320 ns does not significantly change lesion site structural parameters (Tables S1 and S2, Supporting Information).<sup>29</sup> Indeed, the root-mean-square deviation (RMSD) in the position of all heavy atoms in the damaged base pair, and the 5' and 3'–flanking base pairs (including the associated sugar-phosphate backbone), from the 20 and 320 ns simulations on the CXC sequence in the (most stable) base-displaced conformations falls between 0.8 and 1.2 Å (Table S1, Supporting Information). Furthermore, each production simulation for each sequence context gave rise to a single stable conformation, which did not significantly deviate from the initial structure. Specifically, the maximum standard deviation in the backbone RMSD for each final production simulation is approximately 1.5 Å (Table S2, Supporting Information). Additionally, unimodal distributions were obtained with respect to the  $\chi$ ,  $\theta$  and  $\phi$  dihedral angles of the adducted nucleotide (Figure 1 b) throughout each simulation (Figures S1 – S3, Supporting Information). Therefore, a single representative structure was extracted by clustering each simulation trajectory with respect to the position of the atoms forming  $\chi$ ,  $\theta$ , and  $\phi$  using the PTRAJ module of AMBER 12.

To analyze the structural features at the lesion site for each distinct adducted DNA conformation, base step parameters were calculated using a pseudostep, which consists of the base pairs 5' and 3' with respect to the lesion. Free energy calculations were performed on each simulation

trajectory using the molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) method.<sup>78</sup> The entropy term was evaluated using the normal mode approximation at a temperature of 298.15 K, which is close to the temperature used for simulations. Snapshots for the free energy calculations were taken from each simulation at 50 ps intervals (400 frames in total). The relative total free energy ( $G_{rel}$ ) is reported with respect to the adducted DNA conformation with the lowest (most negative)  $G_{total}$ . Furthermore, although all adducted DNA oligonucleotides studied in the present work have different nucleobase sequences, the chemical composition of these strands is the same (i.e., the lesion pair is always flanked by a G:C base pair). Hence, the reported relative stabilities of the damaged DNA strands are compared using the relative total free energies ( $\Delta G_{total}$ ) for the energetically accessible (base-displaced intercalated) conformer with respect to the conformation with the lowest  $G_{total}$  (regardless of the sequence context). We emphasize that these free energy differences do not reflect the strand destabilization due to the AL1 moiety, which cannot be extracted from our data.

### 3. RESULTS AND DISCUSSION

#### 3.1. The Base-Displaced Intercalated Conformational Theme Leads to the Smallest Helical Distortion

**Regardless of the Sequence Context.** Three distinct conformational themes, namely the base-displaced intercalated, 5'-intercalated and 3'-intercalated conformations, were characterized for both the *anti* and *syn* adducted nucleotide orientations in each lesion site sequence context (denoted as GXC, GXG, CXG and CXC). Salient features associated with the three conformational themes are outlined below. In this discussion, all references to the 5' and 3' directions are made with respect to the damaged base.

**3.1.1. The Base-displaced Intercalated Conformers.** The base-displaced intercalated conformation is characterized by the slight displacement of the damaged base towards the minor (major) groove in the *anti* (*syn*) nucleotide orientation (Figure 2). In the base-displaced intercalated conformers, both the damaged adenine and AL1 moieties stack between the 5' and 3'-flanking base pairs, which renders the

opposing T extrahelical. Nevertheless, interactions are maintained between the lesion and the opposing T, primarily in the form of a C–H••• $\pi$  interaction between the methylene group in ring V of the ALI moiety (Figure 1b) and the  $\pi$ -system of T. With the exception of the *syn* base-displaced intercalated conformation in the CXG motif (in which the extrahelical T acquires a *syn* ( $\chi \sim 35^\circ$ ) glycosidic orientation with), the extrahelical T maintains the *anti* orientation in all sequences ( $\chi = \sim 200 - 240^\circ$ ) for both the *anti* and *syn* adduct orientations. Since the T opposing the lesion is displaced out of the helix, few perturbations occur to the 3' and 5'-flanking bases, which maintain Watson-Crick hydrogen bonding (i.e., > 98% occupancy for each G:C hydrogen bond; Table S3, Supporting Information). In this conformational theme, the translational pseudostep parameters (shift, slide and rise) change by  $\sim 0.3 - 2.0$  Å, while the rotational parameters (tilt, roll, twist) change by  $\sim 1 - 14^\circ$ , with respect to the corresponding unmodified DNA (Table S4, Supporting Information). Overall, irrespective of the sequence context, this conformational theme causes minimal distortions to DNA upon adduct incorporation.

**3.1.2. The 5'-intercalated Conformers.** In the 5'-intercalated conformational theme, the ALI moiety stacks between the opposing T and the 5'-base in the opposing strand (Figure 3). Regardless of the sequence considered, hydrogen bonding in the 5'-flanking base pair remains intact (> 98% occupancy for each G:C hydrogen bond; Table S3, Supporting Information). Except in the CXG motif, where both rings I and V are solvent exposed in the major groove, only ring I (Figure 1b) of the polycyclic ALI moiety is solvent exposed in the major groove in the *anti* 5'-intercalated conformation. When the *anti* or *syn* adducted nucleotide occurs in the GXC or CXC motif, the T opposing the lesion is displaced in the 3' direction, while the 3'-C twists to simultaneously hydrogen bond with the displaced T ( $\sim 50\%$  occupancy of the N4–H4•••O4 and  $\sim 40\%$  occupancy of the N3•••H2–N2 hydrogen bond; Table S3, Supporting Information) and its opposing G (> 95% occupancy of the O2•••H2–N2 hydrogen bond; Table S3, Supporting Information). In contrast, both 5'-intercalated conformers maintain hydrogen bonding in the 3'-flanking base pair for the CXG sequence context (Table S3, Supporting Information). Although the

hydrogen bonds in the 3'-flanking base pair are maintained in the *anti* 5'-intercalated conformer for GXG (Table S3, Supporting Information), the corresponding *syn* conformer completely disrupts the 3'-base pair such that the 3'-guanine forms a (O2•••H1-N1) hydrogen bond with the T opposing the lesion (~ 47% occupancy; Table S3, Supporting Information), which displaces the opposing 3'-C in the 3' direction. Overall, this conformational theme is more distorting than the base-displaced intercalated theme. Specifically, the translational pseudostep parameters change by ~ 1.5 – 3.0 Å, and the rotational parameters deviate by ~ 4 – 24°, with respect to the corresponding unmodified DNA (Table S3, Supporting Information).

**3.1.3. The 3'-intercalated Conformers.** In the 3'-intercalated conformational theme, the ALI moiety stacks between the opposing T and the 3'-base with respect to the adduct in the opposing strand (Figure 4). Similar to the 5'-intercalated conformation, the 3'-intercalated conformation adopts different hydrogen-bonding patterns in the lesion region in each sequence context. Specifically, Watson-Crick hydrogen bonding in both the 3' and 5'-flanking base pairs is maintained with the *anti* or *syn* adduct orientation in GXC, and the *syn* orientation in CXG, GXG and CXC (Table S3, Supporting Information). Instead, only the hydrogen bonding in the 3'-flanking base pair is maintained for the *anti* conformer in CXG and GXG, while the hydrogen bonds in the 5'-flanking base pair are partially disrupted (Table S3, Supporting Information). Finally, Watson-Crick hydrogen bonding is completely disrupted in the 5'-flanking base pair when the lesion adopts an *anti* 3'-intercalated conformation in the CXC motif, and a (N4-H4•••O4) hydrogen bond is formed between the 5'-cytosine and the T opposing the lesion.<sup>29</sup> Similar to the 5'-intercalated conformation, this conformational theme is more distorting than the base-displaced intercalated conformation. Specifically, the translational pseudostep parameters change by ~ 0.9 – 3.5 Å, while the rotational parameters change by ~ 5 – 20°, with respect to the corresponding

unmodified DNA (Table S4, Supporting Information). However, the relative distortion of the 5'–intercalated and 3'–intercalated conformations depends on the sequence considered.

Overall, the base-displaced intercalated conformer adopts a similar structure regardless of the sequence considered, with the only exception being the *syn* T opposing the *syn* adduct in the CXG sequence. In contrast, the structural features of the 5'–intercalated and 3'–intercalated conformers can vary with sequence context, mainly in the hydrogen–bonding patterns of the flanking bases and the magnitude of helical distortion. More importantly, regardless of the sequence considered, the ALI-N<sup>6</sup>-dA lesion consistently causes fewer helical perturbations when adducted DNA adopts the base-displaced intercalated orientation. This is largely due to the extrahelical position of the opposing T in this conformational theme, which allows the flanking pairs to remain intact, compared to the intrahelical opposing T in the other (5'–intercalated and 3'–intercalated) conformational themes, which disrupts one or two of the flanking base pairs depending on the sequence context.

**3.2. The *anti* Base-displaced Intercalated Orientation is the Most Stable ALI-N<sup>6</sup>-dA Adducted DNA Conformational Theme Regardless of the Flanking Bases.** Despite the range of structurally distinct conformers isolated from the MD simulations, free energy calculations reveal that the 5'–intercalated and 3'–intercalated conformational themes are  $\sim 25 - 70 \text{ kJ mol}^{-1}$  less stable than the base-displaced intercalated orientation of ALI-N<sup>6</sup>-dA adducted DNA for all sequence contexts (Table 1). This finding correlates with the greatest helical distortion for these adducted DNA conformations (Table S4, Supporting Information). Furthermore, these structural distortions in the 3' or 5'–intercalated conformers decrease the van der Waals (stacking) interactions involving the damaged base by  $\sim 10 - 45 \text{ kJ mol}^{-1}$  compared to the base-displaced intercalated conformers (Table 1). As a result, the 5'–intercalated and 3'–intercalated conformers are likely energetically inaccessible to ALI-N<sup>6</sup>-dA adducted DNA for each sequence considered.

On the other hand, the base-displaced intercalated conformers are stabilized by intact hydrogen bonds in the 5' and 3'-flanking base pairs (Table S3, Supporting Information), significant van der Waals interactions between the lesion and the flanking bases (Table 1), lesion pair interactions (primarily C-H••• $\pi$ , Table S5, Supporting Information), and minimal distortions to the DNA step parameters at the lesion site (Table S4, Supporting Information). Since these factors stabilize the base-displaced conformation for both the *anti* and *syn* glycosidic orientations of the adduct, both conformers lie very close in energy (deviating by a maximum of  $\sim 22 \text{ kJ mol}^{-1}$  depending on the sequence, Table 1). This suggests that both base-displaced intercalated conformers may be thermodynamically accessible regardless of the bases flanking the lesion. Nevertheless, the stabilizing effects are greater for the *anti* than the *syn* base-displaced intercalated conformation in each sequence. For example, the interactions between the ALI moiety and the opposing thymine are  $\sim 8 - 10 \text{ kJ mol}^{-1}$  stronger for the *anti* than the *syn* adduct conformation (Table S5, Supporting Information), which primarily reflects changes in the C-H••• $\pi$  contact. Additionally, the van der Waals interactions between the adduct and the flanking base pairs are (up to  $\sim 6 \text{ kJ mol}^{-1}$ ) greater in the *anti* base-displaced intercalated conformers (Table 1). Most importantly, the experimentally-observed non-accessibility of the *syn* base-displaced intercalated conformation for ALII-N<sup>6</sup>-dA adducted DNA suggests that the *syn* conformer is kinetically prohibited due to a large *anti/syn* energy barrier for rotation about the adduct glycosidic angle, a possibility that remains viable even in the case of ALI-N<sup>6</sup>-dA adducted DNA. Regardless, the *anti* base-displaced intercalated adducted DNA conformer is the most stable among the six conformations characterized in the present work irrespective of the sequential motif.

**3.3. Calculated Relative Free Energies of the Most Stable *anti* Base-displaced Intercalated Adducted DNA Conformer Correlate with Experimentally-Predicted Strand Stabilities.** A previous experimental study carried out by Lukin *et al.* examined the stability of ALII-N<sup>6</sup>-dA adducted DNA in three sequence contexts,<sup>30</sup> two of which (CXG and CXC) were also considered in the present work. It was found that

adducted DNA has a melting temperature ( $T_m$ ) of 39.9°C with the lesion in the CXG sequence context, but a  $T_m$  of 45.2°C when the adduct is in the CXC sequence, and the Gibbs free energy of duplex formation is  $\sim 7 \text{ kJ mol}^{-1}$  more stable for CXC. Furthermore, Lukin *et al.* hypothesized that ALI-N<sup>6</sup>-dA and ALII-N<sup>6</sup>-dA adducted DNA will have similar conformational preferences due to the solvent exposure of ring I of the AL moiety in the most stable *anti* base-displaced intercalated conformer, which contains the methoxy group that differentiates the lesions.<sup>30</sup> This proposal was substantiated by our previous computational work comparing the structure of the ALI and ALII-adducted DNA.<sup>29</sup> Therefore, in the absence of analogous experimental data on the ALI-N<sup>6</sup>-dA adduct, it is anticipated that DNA containing either ALI-N<sup>6</sup>-dA or ALII-N<sup>6</sup>-dA will have a similar dependence of the strand stabilities on the sequence context. Although it is not possible to quantitatively compare the MM-PBSA free energies determined in the present study to the previously reported free energies of duplex formation, our calculations similarly predict that ALI-N<sup>6</sup>-dA adducted DNA is  $\sim 9 \text{ kJ mol}^{-1}$  more stable in the CXC than CXG sequence context (Table 2). This direct correlation between the trends in the sequence dependence of the strand stabilities illustrates the robustness of our computational approach.

### **3.4. Biological Consequences of the Effects of Sequence Context on the Conformational Preferences of ALI-N<sup>6</sup>-dA Adducted DNA.**

**3.4.1. The Relative Stability of the Energetically-Accessible Conformations of Adducted DNA does not Explain the Experimentally-Predicted Sequence-Dependent Mutagenicity of ALI-N<sup>6</sup>-dA.** As discussed in the Introduction, recent experimental studies on AA associated mutagenesis demonstrate that ‘signature’ A→T transversion mutations most commonly occur in T/CXG motifs in the genome.<sup>27, 28</sup> Since ALI-N<sup>6</sup>-dA is the most abundant,<sup>21, 22</sup> nephrotoxic and mutagenic<sup>24</sup> AA adduct, it is reasonable to hypothesize that this adduct plays an important role in these mutations. Although the reason for this sequence preference is not clear from empirical data to date, the mutagenicity of other bulky adducts has been shown to depend on the sequence context<sup>69-71</sup> and be related to enhanced stabilization of the



mutagenic conformation.<sup>69</sup> For example, since the base-displaced intercalated conformer of AF-dG adducted DNA thermodynamically favors the insertion of A over C and is more stable in the CXA than TXA sequential motif, AF-associated mutagenicity is higher in the CXA sequence context.<sup>69</sup> Furthermore, previous studies on DNA adducts suggest that, barring occasional overriding by polymerase interactions,<sup>79</sup> the lesion site conformations observed in damaged DNA are often manifested in polymerase active sites.<sup>79, 80</sup> Thus, the adducted DNA conformations can provide important information regarding the sequence-dependent mutagenicity of the ALI-N<sup>6</sup>-dA lesion.

Since the overall composition of the ALI-N<sup>6</sup>-dA adducted DNA remains the same irrespective of the sequence context considered in the present work, the relative stabilities of the adducted DNA conformations across different sequences can be directly compared ( $\Delta G_{\text{total}}$ , Table 2). Our calculations reveal that the stability of the *anti* base-displaced intercalated conformer varies with sequence context according to GXG > GXC > CXC > CXG. Hence, the *anti* base displaced intercalated conformer of ALI-N<sup>6</sup>-dA adduct is not the most stable in the CXG sequence, but rather falls at least  $\sim 14 \text{ kJ mol}^{-1}$  higher in energy than the most stable GXG sequence. Therefore, the experimentally-observed higher frequency of A $\rightarrow$ T mutations associated with the CXG sequence is not thermodynamically driven by the stability of the mutagenic ALI-N<sup>6</sup>-dA adducted DNA conformation(s). Instead, other factors must dictate the sequence-dependent mutagenicity of ALI-N<sup>6</sup>-dA. For example, interactions between the adduct and the TLS polymerase could be important. Indeed, such interactions were determined to be responsible for the sequence-dependent mutagenicity of PhIP-dG.<sup>70</sup> Alternatively, a higher repair resistance of the most abundant ALI-N<sup>6</sup>-dA adduct in the CXG sequence may play a role in the sequence-dependent mutagenicity, which is considered in the next section.

**3.4.2. Sequence-Dependent Distortions in the Most Persistent Adducted DNA Conformers Point Towards a Possible Sequence-Dependent Repair Propensity of ALI-N<sup>6</sup>-dA.** Although very few correlations between the structural characteristics of damaged DNA and the relative repair propensities

of lesions repaired by TCR are available in literature,<sup>81-84</sup> it is generally believed DNA lesions that can substantially stall RNA polymerases are good substrates for TCR.<sup>56, 85</sup> However, the precise nature of the TCR signal is still not well understood.<sup>56</sup> On the other hand, lesion recognition, as well as lesion repair propensity, in the GGR pathway that operates over the entire genome is believed to be facilitated by changes in a number of local structural features of the helix upon DNA damage, including increases in the helical<sup>42, 62-64</sup> and backbone dynamics,<sup>61</sup> decreases in the stacking interactions at the lesion site<sup>65-67</sup> and perturbations to the helical parameters (such as minor groove dimension, rise and twist).<sup>61, 65, 86-89</sup> Since these structural features can depend on the adducted DNA conformation adopted, it is not surprising that adducted DNA can exhibit conformation-dependent repair propensities.<sup>48, 90</sup> Furthermore, the bases flanking the lesion have been shown to affect the helical structure at the damaged site, and therefore the sequence context affects the repair propensity of some adducts.<sup>40-43, 47-52</sup> On these lines, we critically compare the helical perturbations caused by the lesion in the most energetically-accessible (*anti* base-displaced intercalated) adducted ALI-N<sup>6</sup>-dA DNA conformer for each sequence below.

Although greater lesion site dynamics and an associated higher GGR propensity in the TXT than CXC sequence context has been reported for 10S (+)-*trans-anti*-B[a]P-N<sup>2</sup>-dG PAH adducted DNA,<sup>42, 47</sup> the lesion site dynamics remain unaltered in all base-displaced intercalated conformers of ALI-N<sup>6</sup>-dA adducted DNA compared to the corresponding unmodified DNA regardless of the sequence (Table S4, Supporting Information). In addition, although a decrease in the lesion van der Waals (stacking) energy when AF-dG is at G<sup>2</sup> compared to G<sup>1</sup> or G<sup>3</sup> in the *NarI* sequence explains the observed higher adduct repair propensity at G<sup>2</sup>,<sup>48</sup> the lesion van der Waals interactions in the most persistent *anti* base-displaced intercalated conformer of ALI-N<sup>6</sup>-dA adducted DNA is only 0.5 kJ mol<sup>-1</sup> more stabilizing for the most mutagenic CXG sequence compared to the non-mutagenic GXG sequence (Table 1). Hence, neither dynamics nor the lesion site van der Waals energy likely afford a sequence-dependent GGR propensity

for ALI-N<sup>6</sup>-dA. Nevertheless, sequence-dependent distortions have also been shown to affect the GGR propensity of DNA adducts, including AAF-dG,<sup>48</sup> 10S (+)-*trans-anti*-B[*a*]P-dA<sup>41</sup> and 10S (+)-*trans-anti*-B[*a*]P-dG.<sup>47</sup> In the case of ALI-N<sup>6</sup>-dA adducted DNA, the magnitude of the distortion at the lesion site also varies with the sequence (Figure 5). Specifically, the most persistent *anti* base-displaced intercalated conformer in the most mutagenic CXG sequence<sup>27, 28</sup> exhibits the least lesion site distortions, with the minor groove width changing by 0.4 Å and the helix untwisting by ~ 6° relative to undamaged DNA. In contrast, the most persistent adducted DNA conformers in the other sequence contexts show minor groove widening of up to 0.8 Å and helix untwisting of up to 9° (Figure 5 and Table S4, Supporting Information). These differences point toward the possibility of a differential repair propensity that is consistent with the experimentally-observed greater mutagenicity in the CXG sequence. Nevertheless, the structural differences between the adducted DNA conformers are small, falling within the dynamics of the natural helix (Table S4, Supporting Information), and may not be sufficient to lead to different repair susceptibilities. Therefore, other factors (such as interactions between the adducted DNA and lesion-bypass polymerases during replication) likely have a greater influence on the observed mutagenic profile of the ALI-N<sup>6</sup>-dA adduct and additional biochemical experiments are required to further explore the toxicity of this important lesion.

#### 4. CONCLUSIONS

The present computational study provides structural details of the possible conformers of ALI-N<sup>6</sup>-dA adducted DNA in different sequence contexts (namely, GXC, GXG, CXG and CXC). Our calculations reveal that the conformational outcomes of ALI-N<sup>6</sup>-dA adducted DNA are independent of the identity of the flanking bases. Specifically, the *anti* base-displaced intercalated conformer is the lowest energy conformer in all sequences considered. Our analysis reveals that the experimentally-observed greater mutagenicity of the CXG sequence is not thermodynamically driven by the stability of the adducted DNA as reported for other adducts. However, comparison of the structural perturbations upon lesion

formation in different sequence contexts indicates that less stable lesion site van der Waals (stacking) interactions and greater distortions likely lead to a dependence of the adduct repair propensity on the lesion site sequence context. Most importantly, ALI-N<sup>6</sup>-dA adducted DNA in the CXG sequence context is the least distorted helix, which might contribute to a greater repair resistance that is consistent with the experimentally-observed higher mutagenicity in this sequential motif. Nevertheless, since the differences in distortions are small across different sequence contexts, other factors (such as interactions with the lesion-bypass polymerases) likely more greatly contribute to the experimentally-observed sequence-dependent mutagenicity of this adduct.

## ASSOCIATED CONTENT

### Supporting Information:

Radar plots for the probability distributions of  $\chi$ ,  $\theta$  and  $\phi$  (Figures S1-S3); RMSD of all heavy atoms in the lesion site trimer in the base-displaced intercalated adducted DNA conformations during 20 and 320 ns MD simulations in the CXC sequence context (Table S1); Backbone RMSD for all adducted DNA conformations in different sequence contexts (Table S2); The hydrogen-bonding occupancies in the trimers composed of the damaged base pair and the 3' and 5'-flanking base pairs in different sequence contexts (Table S3); Pseudostep parameters for all adducted DNA conformers in different sequence contexts (Table S4); Interaction energies between the damaged base and its opposing thymine for different base-displaced intercalated conformations of ALI-N<sup>6</sup>-dA adducted DNA (Table S5). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\* Phone: 403-329-2323. Fax: 403-329-2057. E-mail: [stacey.wetmore@uleth.ca](mailto:stacey.wetmore@uleth.ca).

## Funding Sources

Financial support for this research was provided by Canada Foundation for Innovation (CFI), Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canada Research Chair (CRC) program, and the University of Lethbridge Research Fund. P.K. thanks the University of Lethbridge for financial support. PS thanks the Department of Science and Technology (DST), Government of India for financial support under the DST-INSPIRE Faculty scheme (No. DST/INSPIRE/04/2014/001855).

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Calculations were conducted on the New Up-scale Cluster for Lethbridge to Enable Innovative Chemistry (NUCLEIC). Additional computational resources were provided by WestGrid and Compute/Calcul Canada.

## ABBREVIATIONS

dA, 2'-deoxyadenosine; dG, 2'-deoxyguanosine; MD, molecular dynamics; AA aristolochic acid; AL, aristolactam; ALII-N<sup>6</sup>-dA, aristolactam II-dA adduct; ALI-N<sup>6</sup>-dA, aristolactam I-dA adduct; RMSD, root mean square deviation; QM, quantum mechanical; DFT, density functional theory; AAF-dG, acetylaminofluorene-dG adduct; AF-dG, aminofluorene-dG adduct; PAH, polycyclic aromatic hydrocarbons; OTB-dG, C-linked ochratoxin-dG adduct; PhOH-dG, phenoxl-dG adduct; GAFF, general AMBER force field; IARC, International Agency for Research on Cancer; NER, nucleotide excision repair; AMBER, Assisted Model Building and Energy Refinement; B[a]P, benzo[a]pyrene; DB[a,l]P, dibenzo[a,l]pyrene; IQ, 2-Amino-3-methylimidazo[4,5-f]quinolone; PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.

## REFERENCES

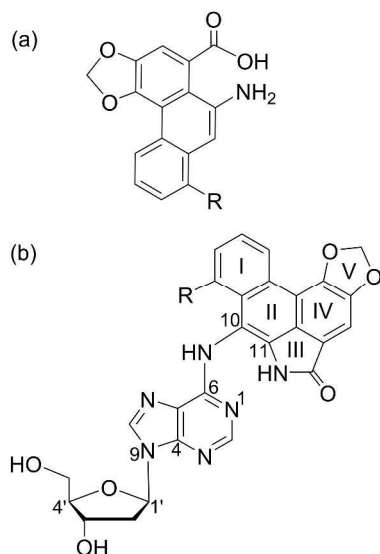
- (1) V. M. Arlt, M. Stiborova, , and H. H. Schmeiser, *Mutagenesis*, 2002, **17**, 265-277.
- (2) J. L. Vanherweghem, C. Tielemans, D. Abramowicz, M. Depierreux, R. Vanhaelen-Fastre, M. Vanhaelen, M. Dratwa, C. Richard, D. Vandervelde, D. Verbeelen, and M. Jadoul, *The Lancet*, 1993, **341**, 387-391.
- (3) F.D. Debelle, J. L. Vanherweghem, and J. L. Nortier, Aristolochic acid nephropathy: A worldwide problem. *Kidney Int.*, 2008, **74**, 158-169.
- (4) J. P. Cosyns, M. Jadoul, J. P. Squifflet, F. X. Wese, and C. van Ypersele de Strihou, *Am. J. Kidney Dis.* 1999, **33**, 1011-1017.
- (5) J. L. Nortier, and J.L. Vanherweghem ,*Toxicology* , 2002, **181–182**, 577-580.
- (6) J. L. Nortier, M.C. M. Martinez, H.H. Schmeiser, V.M. Arlt, C.A. Bieler, M. Petein, M. F. Depierreux, L. De Pauw, D. Abramowicz, P. Vereerstraeten, and J.L. Vanherweghem, *N. Engl. J. Med.*, 2000, **342**, 1686-1692.
- (7) V. Batuman, *Kidney Int.*, 200, **669**, 644-646.
- (8) L. Djukanović, and Z. Radovanović, (2003) Balkan endemic nephropathy, In *Clinical Nephrotoxins* (de Broe, M., Porter, G., Bennett, W., and Verpooten, G., Eds.) pp 587-601, Springer Netherlands.
- (9) M. E. De Broe, *Kidney Int.*, 2012, **81**, 513-515.
- (10) V. Stefanovic, and Z. Radovanovic, *Nat. Clin. Pract. Urol.*, 2008, **5**, 105-112.
- (11) M. Ivić, *Lijecnicki vjesnik* ,1969, **91**, 1273.
- (12) A.P. Grollman, S. Shibutani, M. Moriya, F. Miller, L. Wu, U. Moll, N. Suzuki, A. Fernandes, T. Rosenquist, Z. Medverec, K. Jakovina, B. Brdar, N. Slade, R. Turesky, A. K. Goodenough, R. Rieger, M. Vukelić, and B. Jelaković, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 12129-12134.
- (13) V. M. Arlt, V. Alunni-Perret, G. Quatrehomme, P. Ohayon, L. Albano, H. Gaïd, J. F. Michiels, A. Meyrier, E. Cassuto, M. Wiessler, H. H. Schmeiser, and J.P. Cosyns, *Int. J. Cancer*, 2004, **111**, 977-980.
- (14) *IARC monographs-100 A Plants containing aristolochic acids*, 2002, 347-361.
- (15) H.H. Schmeiser, C.A. Bieler, M. Wiessler, C. van Ypersele de Strihou, and J. P. Cosyns, *Cancer Res.*, 1996, **56**, 2025-2028.
- (16) V. S. Sidorenko, S. Attaluri, I. Zaitseva, C.R. Iden, K. Dickman, F. Johnson, and A.P. Grollman, *Carcinogenesis*, 2014, **35**, 1814-1822.
- (17) M. Stiborová, J. Mareš, E. Frei, V. M. Arlt, V. Martínek, and H. H. Schmeiser, *Environ. Mol. Mutagen.*, 2011, **52**, 448-459.
- (18) W. Pfau, H. H. Schmeiser, and M. Wiessler, *Carcinogenesis*, 1990, **11**, 313-319.
- (19) S. Attaluri, R. R. Bonala, I. Y. Yang, M.A. Lukin, Y. Wen, A. P. Grollman, M. Moriya, C. R. Iden, and F. Johnson, *Nucleic Acids Res.*, 2010, **38**, 339-352.
- (20) V. Martinek, B. Kubickova, V. M. Arlt, E. Frei, H. H. Schmeiser, J. Hudecek, and M. Stiborova, *Neuroendocrinol. Lett.*, 2010, **32**, 57-70.
- (21) C. A. Bieler, M. Stiborova, M. Wiessler, J. P. Cosyns, C. van Ypersele de Strihou, and H. H. Schmeiser, *Carcinogenesis*, 1997, **18**, 1063-1067.
- (22) W. Pfau, H. H. Schmeiser, and M. Wiessler, *Carcinogenesis*, 1990, **11**, 1627-1633.
- (23) H. H. Schmeiser, J. L. Nortier, R. Singh, G. Gamboa da Costa, J. Sennesael, E. Cassuto-Viguier, D. Ambrosetti, S. Rorive, S., A. Pozdzik, A., and D. H. Phillips, *Int. J. Cancer*, 2014, **135**, 502-507.
- (24) S. Shibutani, H. Dong, N. Suzuki, S. Ueda, F. Miller, and A. P. Grollman, *Drug Metab. Dispos.*, 2007, **35**, 1217-1222.

- (25) M. Moriya, N. Slade, B. Brdar, Z. Medverec, K. Tomic, B. Jelaković, L. Wu, S. Truong, A. Fernandes, and A. P. Grollman, *Int. J. Cancer*, 2011, **129**, 1532-1536.
- (26) C. H. Chen, K. G. Dickman, M. Moriya, J. Zavadil, V.S. Sidorenko, K. L. Edwards, D. V. Gnatenko, L. Wu, R. J. Turesky, X. R. Wu, Y. S. Pu, and A. P. Grollman, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 8241-8246.
- (27) S. L. Poon, S. T. Pang, J. R. McPherson, W. Yu, K. K. Huang, P. Guan, W. H. Weng, E. Y. Siew, Y. Liu, H. L. Heng, S. C. Chong, A. Gan, S. T. Tay, W. K. Lim, I. Cutcutache, D. Huang, L.D. Ler, M. L. Nairismägi, M. H. Lee, Y. H. Chang, K. J. Yu, W. Chan-on, , B. K. Li, Y. F. Yuan, , C. N Qian, K. F .Ng, C. F. Wu, , C. L. Hsu, R. M. Bunte, M. R. Stratton, P. A. Futreal, W. K. Sung, , C. K .Chuang, C. K. Ong, S. G. Rozen, P. Tan, and B. T. Teh, *Sci. Transl. Med.*, 2013 **5**, 197ra101.
- (28) M. L. Hoang, C.H. Chen, V. S. Sidorenko, J. He, K.G. Dickman, B. H. Yun, M. Moriya, N. Niknafs, C. Douville, R. Karchin, R. J. Turesky, Y. S. Pu, B. Vogelstein, N. Papadopoulos, A. P. Grollman, K. W. Kinzler, and T. A. Rosenquist, *Sci. Transl. Med.*, 2013, **5**, 197ra102.
- (29) P. Kathuria, P. Sharma, M. Abendong, and S. D. Wetmore, *Biochemistry*, 2015, **54**, 2414-2428.
- (30) M. Lukin, T. Zaliznyak, F. Johnson, and C. de los Santos, *C., Nucleic Acids Res.*, 2012, **40**, 2759-2770.
- (31) Y. Cai, S. Ding, N. E. Geacintov, and S. Broyde, *Chem. Res. Toxicol.*, 2011, **24**, 522-531.
- (32) Z. Li, H. Y. Kim, P. J. Tamura, C. Harris, T. M. Harris, and M. P. Stone, *Biochemistry*, 1999, **38**, 16045-16057.
- (33) Z. Li, H. Mao, H. Y. Kim, P. J. Tamura, C. M. Harris, T. M. Harris, and M. P. Stone, *Biochemistry*, 1999, **38**, 2969-2981.
- (34) Z. Li, P. J. Tamura, A. S. Wilkinson, C. M. Harris, T. M. Harris, and M. P. Stone, *Biochemistry*, 2001, **40**, 6743-6755.
- (35) I. S. Zegar, P. Chary, R. J. Jabil, P. J. Tamura, T. N. Johansen, R. S. Lloyd, C. M. Harris, T. M. Harris, and M. P. Stone, *Biochemistry*, 1998, **37**, 16516-16528.
- (36) F. Wang, C. E. Elmquist, J. S. Stover, C. J. Rizzo, and M. P. Stone, *Biochemistry*, 2007, **46**, 8498-8516.
- (37) B. Mao, B. E. Hingerty, S. Broyde, and D. J. Patel, *Biochemistry*, 1998, **37**, 95-106.
- (38) B. Mao., B. E. Hingerty, S. Broyde, and D. J. Patel, *Biochemistry*, 1998, **37**, 81-94.
- (39) S. Patnaik, and B. P. Cho., *Chem. Res. Toxicol.*, 2010, **23**, 1650-1652.
- (40) K. Kropachev, M. Kolbanovskii, Y. Cai, F. Rodríguez, A. Kolbanovskii, Y. Liu, L. Zhang, S. Amin, D. Patel, S. Broyde, and N. E. Geacintov, *J. Mol. Biol.*, 2009, **386**, 1193-1203.
- (41) S. Yan, M. Wu, T. Buterin, H. Naegeli, N. E. Geacintov, and S. Broyde, *Biochemistry*, 2003, **42**, 2339-2354.
- (42) Y. Cai, D. J. Patel, N. E. Geacintov, and S. Broyde, *J. Mol. Biol.*, 2007, **374**, 292-305.
- (43) F. A. Rodríguez, Y. Cai, C. Lin, Y. Tang, A. Kolbanovskiy, S. Amin, D. J. Patel, S. Broyde, and N. E. Geacintov, *Nucleic Acids Res.*, 2007, **35**, 1555-1568.
- (44) P. Sharma, R. A. Manderville, and S. D. Wetmore, *Nucleic Acids Res.*, 2014, **42**, 11831-11845.
- (45) P. Sharma, R. A. Manderville, and S. D. Wetmore, *Chem. Res. Toxicol.*, 2013, **26**, 803-816.
- (46) R. P. P. Fuchs, N. Schwartz, and M. P. Daune, *Nature*, 1981, **294**, 657-659.
- (47) Y. Cai, D. J. Patel, S. Broyde, and N. E. Geacintov, *J. Nucleic Acids*, 2010, **2010**, 9.
- (48) H. Mu, K. Kropachev, L. Wang, L. Zhang, A. Kolbanovskiy, M. Kolbanovskiy, S. Broyde, and N. E. Geacintov, *Nucleic Acids Res.*, 2012, **40**, 9675-9690.
- (49) Jain, V., Hilton, B., Patnaik, S., Zou, Y., Chiarelli, M. P., and B. P. Cho, *Nucleic Acids Res.*, 2012, **40**, 3939-3951.
- (50) Y. Cai, N. E. Geacintov, and S. Broyde, *Biochemistry*, 2012, **51**, 1486-1499.
- (51) V. Jain, B. Hilton, B. Lin, S. Patnaik, F. Liang, E. Darian, Y. Zou, A. D. MacKerell, and B. P. Cho, *Nucleic Acids Res.*, 2013, **41**, 869-880.

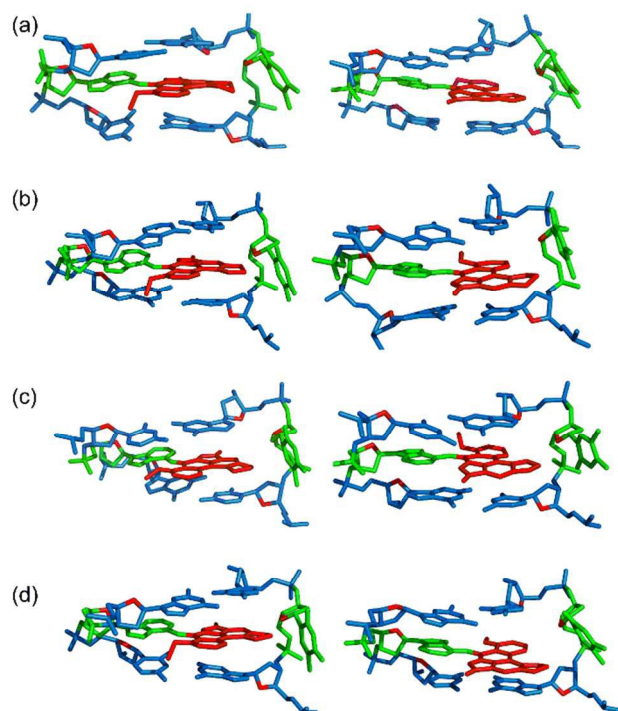
- (52) Meneni, S., S. M. Shell, Y. Zou, and B. P. Cho, *Chem. Res. Toxicol.*, 2007, **20**, 6-10.
- (53) C. P. Rubbi, and J. Milner, *Carcinogenesis*, 2001 **22**, 1789-1796.
- (54) W. L. de Laat, N. G. J. Jaspers, and J. H. J. Hoeijmakers, *Genes Dev.*, 1999, **13**, 768-785.
- (55) L. C. J. Gillet, , and O. D. Schärer, *Chem. Rev.*, 2005, **106**, 253-276.
- (56) P. C. Hanawalt, , and G. Spivak, Transcription-Coupled DNA repair: Two Decades of Progress and Surprises. *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 958-970.
- (57) K. Kropachev, M. Kolbanovskiy, , Z. Liu, Y. Cai, L. Zhang, A. G. Schwaid, A. Kolbanovskiy, S. Ding, S. Amin, S. Broyde, and N. E. Geacintov, *Chem. Res. Toxicol.*, **2013**, **26**, 783-793.
- (58) Y. Cai, D. J. Patel, S. Broyde, and N. E. Geacintov, *J. Mol. Biol.*, 2009, **385**, 30-44.
- (59) B. Hang, *J. Nucleic Acids*, 2010, **2010**, 29.
- (60) V. S. Sidorenko, J. E. Yeo., R. R. Bonala, F. Johnson , O. D. Schärer, and A. P. Grollman, *Nucleic Acids Res.*, 2012, **40**, 2494-2505.
- (61) K. Kropachev, S. Ding, M. A. Terzidis, A. Masi, Z. Liu, Y. Cai, M. Kolbanovskiy, C. Chatgililoglu, S. Broyde, and N. E. Geacintov, and V. Shafirovich, *Nucleic Acids Res.*, 2014, **42**, 5020-5032.
- (62) O. Maillard, U. Camenisch, F. C. Clement, K. B. Blagoev, and H. Naegeli, *Trends Biochem. Sci.*, 2007, **32**, 494-499.
- (63) K. B. Blagoev, B. S. Alexandrov, E. H. Goodwin, and A. R. Bishop, *DNA Repair* , 2006, **5**, 863-867.
- (64) R. J. Isaacs, and H. P. Spielmann, *DNA Repair*, 2004, **3**, 455-464.
- (65) X. Chen, Y. Velmurugu, G. Zheng, , B. Park, Y. Shim, Y. Kim, L. Liu, B. Van Houten, C. He, A. Ansari, , and J. H. Min, *Nat. Commun.* 2015, **6**, 10.1038/ncomms6849.
- (66) D. A. Reeves, H. Mu, K. Kropachev, Y. Cai, S. Ding, A. Kolbanovskiy, M. Kolbanovskiy, Y. Chen, J. Krzeminski, and S. Amin, *Nucleic Acids Res.*, 2011, **39**, 8752-8764.
- (67) W. Yang, *DNA Repair*, 2006, **5**, 654-666.
- (68) A. Janičević, K. Sugawara, Y. Shimizu, F. Hanaoka, N. Wijgers, M. Djurica, J. H. J. Hoeijmakers, and C. Wyman, *DNA Repair*, 2003, **2**, 325-336.
- (69) V. G. Vaidyanathan, and B. P. Cho, *Biochemistry*, 2012, **51**, 1983-1995.
- (70) P. Xu, L. Oum, Y. C. Lee, N. E. Geacintov , and S. Broyde, *Biochemistry*, 2009, **48**, 4677-4690.
- (71) S. Shibutani, A. Fernandes, N. Suzuki, L. Zhou, F. Johnson, and A. P. Grollman, *J. Biol. Chem.*, 1999, **274**, 27433-27438.
- (72) V. Jain, V. G. Vaidyanathan, S. Patnaik, S. Gopal, and B. P. Cho, *Biochemistry*, 2014, **53**, 4059-4071.
- (73) D. A. Case, T. A. Darden, T.E. Cheatham, I., C. L., Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, R. C., W. Zhang, K. M. Merz, B. Wang, , S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, , F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P. A. Kollman, 2008, AMBER Tools, University of California, San Francisco.
- (74) D. A. Case, T. A. Darden, T.E. Cheatham, I., C. L., Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, R. C., W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, , F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P. A. Kollman, 2010, *AMBER 11*, *AMBER 11* University of California, San Francisco, CA.
- (75) D. T. Roy, J. M. Keith. GaussView Version 5. *GaussView Version 5*, 2009
- (76) D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *J. Comput. Chem.* , 2005, **26**, 1668-1688.
- (77) D. A. Case, T. A. Darden, T.E. Cheatham, I., C. L., Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M.



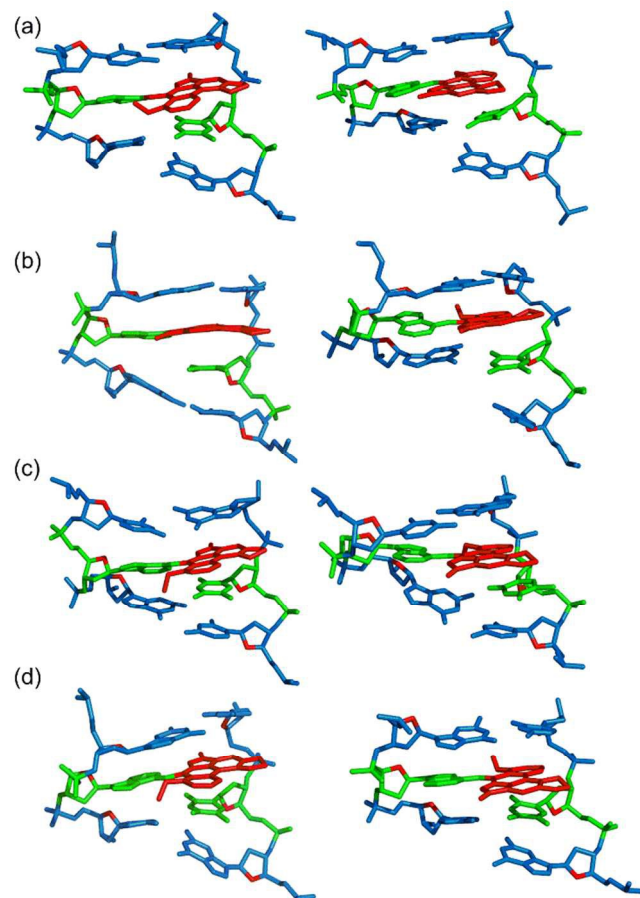
- G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, *AMBER 12*, *AMBER 12*, 2012, University of California, San Francisco, CA.
- (78) B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314-3321.
- (79) S. Broyde, L. Wang, L. Zhang, O. Rechkoblit, N. E. Geacintov, and D. J. Patel, *Chem. Res. Toxicol.*, 2007, **21**, 45-52.
- (80) M. Sproviero, A. M. R. Verwey, K. M. Rankin, A. A. Witham, D. V. Soldatov, R. A. Manderville, M. I. Fekry, S. J. Sturla, P. Sharma, and S. D. Wetmore, *Nucleic Acids Res.*, 2014, **42**, 13405-13421.
- (81) J. S. Mei Kwei, I. Kuraoka, K. Horibata, M. Ubukata, E. Kobatake, S. Iwai, H. Handa, and K. Tanaka, *Biochemical and Biophysical Research Communications*, 2004, **320**, 1133-1138.
- (82) A. Dimitri, A. K. Goodenough, F. P., Guengerich, S. Broyde, and D. A. Scicchitano, *J. Mol. Biol.*, 2008, **375**, 353-366.
- (83) G. E. Damsma, A. Alt, F. Brueckner, T. Carell, and P. Cramer, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1127-1133.
- (84) F. Brueckner, U. Hennecke, T. Carell, and P. Cramer, *Science*, 2007, **315**, 859-862.
- (85) W. Vermeulen, M. Foustari, Mammalian Transcription-Coupled Excision Repair. *Cold Spring Harbor Perspectives in Biology*, 2013, **5**, a012625.
- (86) N. E. Geacintov, S. Broyde, T. Buterin, H. Naegeli, M. Wu, S. Yan, and D.J. Patel, *Biopolymers*, 2002, **65**, 202-210.
- (87) O. Rechkoblit, A. Kolbanovskiy, L. Malinina, N. E. Geacintov, S. Broyde, and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2010, **17**, 379-388.
- (88) O. D. Schäfer, *Mol. Cell.*, 2007, **28**, 184-186.
- (89) K. Sugawara, T. Okamoto, Y. Shimizu, C. Masutani, S. Iwai, and F. Hanaoka, *Genes Dev.*, 2001, **15**, 507-521.
- (90) R. Jankowiak, F. Ariese, A. Hower, A. Luch, D. Zamzow, N. C. Hughes, D. Phillips, A. Seidel, K.-L. Platt, and D. Oesch, *Chem. Res. Toxicol.*, 1998, **11**, 674-685.



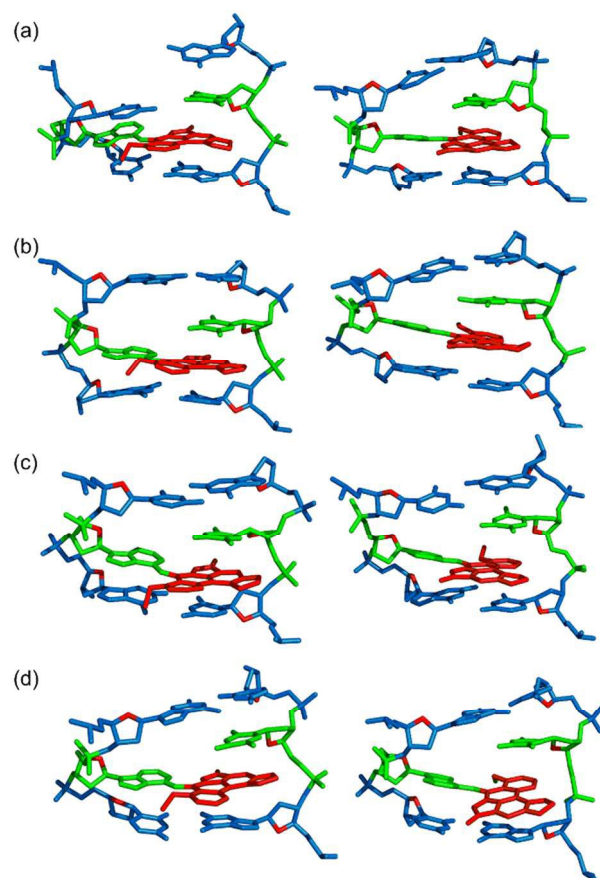
**Figure 1:** Structure of the (a) aristolochic acid and (b) AL-N<sup>6</sup>-dA adduct (R = OCH<sub>3</sub> for ALI and H for ALII). The  $\theta$  ( $\angle(N1C6N6C10)$ ) and  $\phi$  ( $\angle(C6N6C10C11)$ ) dihedral angles determine the orientation of the AL moiety with respect to A, while the  $\chi$  dihedral angle ( $\angle(O4'C1'N9C4)$ ) dictates the glycosidic bond orientation to be *syn* ( $\chi = 0 \pm 90^\circ$ ) or *anti* ( $\chi = 180 \pm 90^\circ$ ).



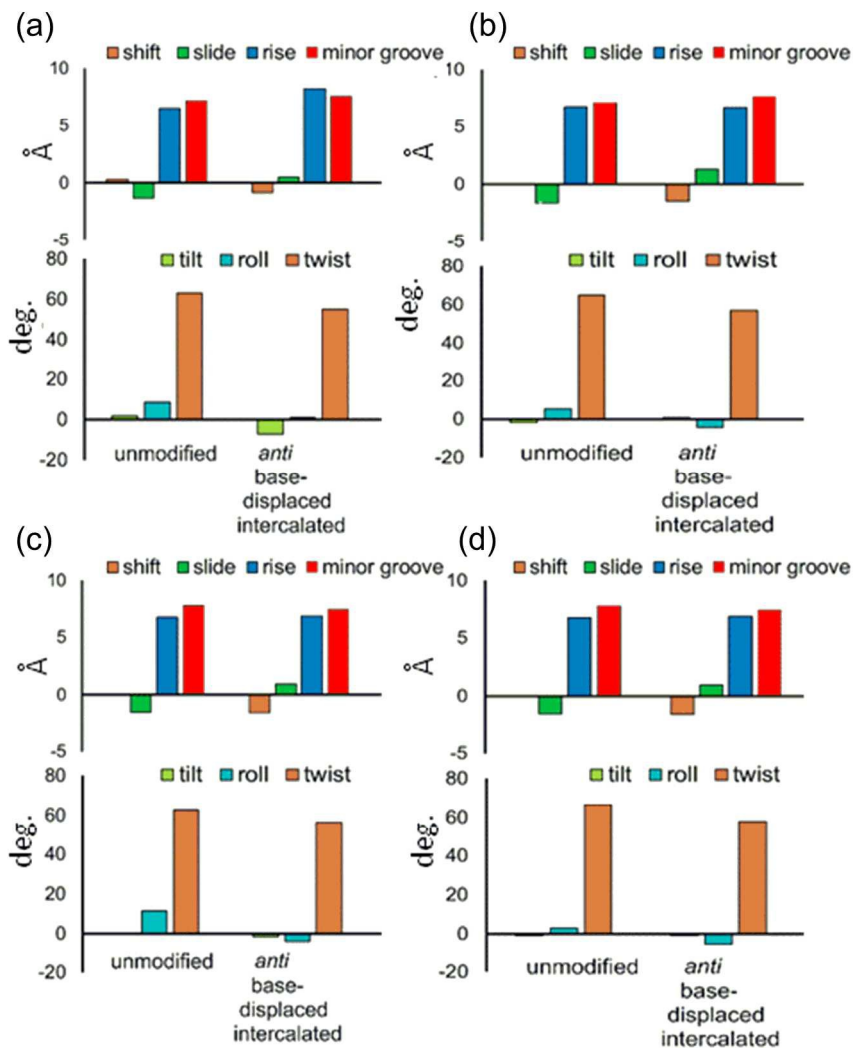
**Figure 2:** Representative structures from MD simulations with the *anti* (left) and *syn* (right) orientation of the ALI-N<sup>6</sup>-dA adduct paired opposite thymine in the (a) CXC, (b) GXG, (c) CXG and (d) GXC sequence contexts for the base-displaced intercalated DNA conformation.



**Figure 3:** Representative structures from MD simulations with the *anti* (left) and *syn* (right) orientation of the ALI-N<sup>6</sup>-dA adduct paired opposite thymine in the (a) CXC, (b) GXG, (c) CXG and (d) GXC sequence contexts for the 5'-intercalated DNA conformation.



**Figure 4:** Representative structures from MD simulations with the *anti* (left) and *syn* (right) orientation of the ALI-N<sup>6</sup>-dA adduct paired opposite thymine in the (a) CXC, (b) GXG, (c) CXG and (d) GXC sequence context for the 3'-intercalated DNA conformation.



**Figure 5:** The pseudostep parameters and minor groove dimensions for the *anti* base-displaced intercalated conformers of ALI-N<sup>6</sup>-dA adducted DNA in the (a) CXC, (b) GXG, (c) CXG and (d) GXC sequence contexts, as well as the values for the corresponding unmodified strand.

**Table 1:** Relative MM-PBSA Free Energy ( $G_{rel}$ ) and the van der Waals stacking energies ( $E_{vdW}$ ) for different conformations of ALI-N<sup>6</sup>-dA adducted DNA derived from 20 ns MD simulations ( $\text{kJ mol}^{-1}$ ).

Conformation	Sequential Motif							
	CXC		GXG		CXG		GXC	
	$G_{rel}^a$	$E_{vdW}^b$	$G_{rel}^a$	$E_{vdW}^b$	$G_{rel}^a$	$E_{vdW}^b$	$G_{rel}^a$	$E_{vdW}^b$
<i>anti</i> base-displaced	0.0	-171.8±6.8	0.0	-173.0±6.3	0.0	-173.5±6.7	0.0	-167.2±6.3
<i>anti</i> 5'-intercalated	36.1	-148.1±10.9	54.0	-153.4±10.9	27.9	-134.4±7.1	44.3	-155.4±8.8
<i>anti</i> 3'-intercalated	51.6	-148.9±12.4	38.7	-135.2±7.1	56.5	-127.7±6.7	52.1	-139.0±9.2
<i>syn</i> base-displaced	2.1	-166.3±7.1	16.2	-169.2±7.5	14.8	-167.6±7.1	22.1	-163.8±7.1
<i>syn</i> 5'-intercalated	42.0	-144.8±10.5	69.4	-149.5±18.5	46.5	-127.3±11.3	56.4	-146.1±9.2
<i>syn</i> 3'-intercalated	49.1	-135.6±7.6	47.5	-139.9±10.9	52.2	-126.0±11.3	47.8	-136.5±7.1

<sup>a</sup>Relative total free energy calculated with respect to the most stable conformation of adducted DNA in each sequence context. <sup>b</sup>Total van der Waals interaction energy between the adduct and both flanking base pairs.

**Table 2:** The components of the total free energy ( $\text{kJ mol}^{-1}$ ) for the *anti* base–displaced intercalated conformations in different sequence contexts.

Components of Free Energy	CXC	GXG	CXG	GXC
$E_{\text{bond}}^{\text{a}}$	808.2	807.9	812.4	807.5
$E_{\text{angle}}^{\text{b}}$	1727.1	1723.6	1720	1722.3
$E_{\text{dihedral}}^{\text{c}}$	2085.5	2087.0	2100.8	2086
$E_{\text{vdW}}^{\text{d}}$	-813.0	-830.8	-821.6	-814.8
$E_{\text{elec}}^{\text{e}}$	1257.1	1351.8	1364.6	1227.6
$G_{\text{sol}}^{\text{f}}$	-22535.5	-22617.8	-22641.2	-22502.2
$-TS^{\text{g}}$	-2464.6	-2461.3	-2461.4	-2464.0
$G_{\text{total}}^{\text{h}}$	-19935.1	-19939.8	-19926.2	-19937.7
$\Delta G_{\text{total}}^{\text{i}}$	4.7	0.0	13.6	2.1

<sup>a</sup>Internal energy emerging from deviation of bonds ( $E_{\text{bond}}$ ). <sup>b</sup>Internal energy emerging from deviation of angles ( $E_{\text{angle}}$ ). <sup>c</sup>Internal energy emerging from deviation of dihedral angles ( $E_{\text{dihedral}}$ ). <sup>d</sup>van der Waals interaction energy ( $E_{\text{vdW}}$ ). <sup>e</sup>Electrostatic interaction energy ( $E_{\text{elec}}$ ). <sup>f</sup>Solvation free energy ( $G_{\text{sol}}$ ). <sup>g</sup>Entropy term ( $-TS$ ). <sup>h</sup>Total free energy ( $G_{\text{total}}$ ). <sup>i</sup>Relative total free energies with respect to the conformation with the lowest (most negative)  $G_{\text{total}}$  regardless of the sequence context ( $\Delta G_{\text{total}}$ ).