CONFORMATIONAL ANALYSIS OF DNA DAMAGED BY ARISTOLOCHIC ACIDS TO ASSESS THE TOXICITY AND REPAIR: A COMPUTATIONAL STUDY

PREETLEEN KATHURIA
Masters of Science, Panjab University, 2009
Bachelors of Science (Hons. School), University of Delhi, 2007

A Thesis
Submitted to the School of Graduate Studies
Of the University of Lethbridge
In Partial Fulfilment of the
Requirements for the Degree

MASTERS OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

CONFORMATIONAL ANALYSIS OF DNA DAMAGED BY ARISTOLOCHIC ACIDS TO ASSESS THE TOXICITY AND REPAIR: A COMPUTATIONAL STUDY

PREETLEEN KATHURIA

Date of Defence: July 21, 2015

Dr. S. Wetmore Supervisor	Professor	Ph.D.
Dr. R. Boeré Thesis Examination Committee Member	Professor	Ph.D.
Dr. P. Dibble Thesis Examination Committee Member	Associate Professor	Ph.D
Dr. A. Brown External Examiner University of Alberta Edmonton, Alberta	Professor	Ph.D.
Dr. U. Kothe Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Abstract

Molecular modeling study is carried out on DNA damaged by aristolochic acids I and II. The carcinogenic effects of these acids have been attributed to the formation of bulky adducts within DNA (specifically with the adenine and guanine nucleobases). The analysis is initiated by considering the conformational preferences of the aristolochic acid adducts in small models (nucleobase, nucleoside and nucleotide) using quantum mechanical calculations. Subsequently, the adducts were studied using molecular dynamics simulations at the oligonucleotide level by incorporating the damaged base in an 11-mer DNA strand. The structural preferences of adenine and guanine adducts are compared, which may explain the experimentally-observed differential repair propensity. Finally, sequence-induced differences in the structural features of adducted DNA containing the ALI-adenine lesion are analyzed to explain the experimentally-reported higher mutagenicity of the lesion in a particular sequence. Results and implications from each of these studies are presented, and future research directions are proposed.

Acknowledgements

First of all, I am grateful to my supervisor Dr. Stacey Wetmore for constant support and guidance throughout my degree. I would like to thank you for all the encouragement and appreciation that kept me going. I would also like to thank my committee members Dr. René Boere and Dr. Peter Dibble for their valuable feedbacks on my research. I would especially like to thank my group members Dr. Purshotam Sharma, Shahin, Stefan and Katie for their constant support and friendliness which made this journey an enjoyable one. I would also like to thank NSERC, CRC and University of Lethbridge Research Fund for funding this project.

A special thanks to my family members, my parents and my parents-in-law, your blessings have helped me sustain thus far. At the end, I would like to thank my husband Jitender Singh for all the love and sacrifices you made to help me get through all the difficult times during the course.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A	Adenine
G	Guanine
C	Cytosine
T	Thymine
dA	2′-deoxyadenosine
dG	2′-deoxyguanosine
MD	Molecular Dynamics
AA	Aristolochic Acid
AL	Aristolactam
ALII-N^6-dA	Aristolactam II-dA adduct
ALI-N^6-dA	Aristolactam I-dA adduct
ALII-N^2-dG	Aristolactam II-dG adduct
ALI-N^2-dG	Aristolactam I-dG adduct
RMSD	Root Mean Square Deviation
QM	Quantum Mechanical
PCM	Polarizable Continuum Model
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
AMBER	Assisted Model Building and Energy Refinement
NMR	Nuclear Magnetic Resonance
GAFF	General AMBER Force Field
PES	Potential Energy Surface
IARC	International Agency for Research on Cancer
NER	Nucleotide Excision Repair
GGR	Global Genomic Repair
TCR	Transcription Coupled Repair
BEN	Balkan Endemic Nephropathy
CHN	Chinese Herb Nephropathy
AAN	Aristolochic Acid Nephropathy
IQ-dG	2-Amino-3-methylimidazo[4,5-f]quinolone-dG adduct
BD	1,3-butadiene
AN-dG	Aniline-dG adduct
Ph-dG	Phenyl-dG adduct
DB[*a,l*]P	Dibenzo[*a,l*]pyrene
B[*a*]P	Benzo[*a*]pyrene
B[*c*]Ph	Benzo[*c*]phenanthrene
AP	Aminopyrene
XPC	Xeroderma pigmentosum C

/Mb	per one million nucleotides

Chapter 1

Thesis Introduction

1.1 Thesis Overview

Deoxyribonucleic acid (DNA) is the hereditary material in all cellular life forms. In eukaryotes, DNA is present in either the nucleus (nuclear DNA) or the mitochondria (mitochondrial DNA) of the cell. The complementary nucleobase pairing pattern within DNA is critical for the transfer of genetic information from one generation to the next. The DNA base-sequence directs various complex biological processes in the cell. However, DNA is not inert. Instead, many environmental contaminants and reactive species produced during cellular processes, such as metabolism, damage DNA. This can lead to mispairing of the DNA nucleobases, which in turn blocks various biological processes, such as DNA replication and transcription. The present thesis uses various computational methods to understand one specific type of DNA damage caused by exposure to plant mutagens called aristolochic acids (AAs), and to explain the structural changes to DNA induced by the AAs, as well as the associated implications for repair and mutagenesis. The following sections give a general overview of DNA structure, the types of DNA damage possible, their biological implications, a general introduction to AA damage, and an outline of each subsequent chapter.

1.2 DNA Structure

DNA has a double stranded structure (Figure 1.1a). The backbones of DNA are composed of sugar and phosphate units, run in an antiparallel direction and wrap

around the periphery of the helix to form two kinds of grooves, the smaller minor groove, and the larger major groove (Figure 1.1a). All DNA-protein interactions that occur during replication and transcription take place in either of these grooves. The core of DNA is composed of purine [guanine (G) and adenine (A)] and pyrimidine [cytosine (C) and thymine (T)] nucleobases (Figure 1.1b). The deoxyribose sugar is attached to the nucleobases via the C–N glycosidic bond involving N1 for the pyrimidines and N9 for the purines (Figure 1.1b). A nucleoside is composed of the nucleobase attached to a sugar, while a nucleotide includes the base, sugar and phosphate. The nucleotide is the repeating unit of a DNA polymer.

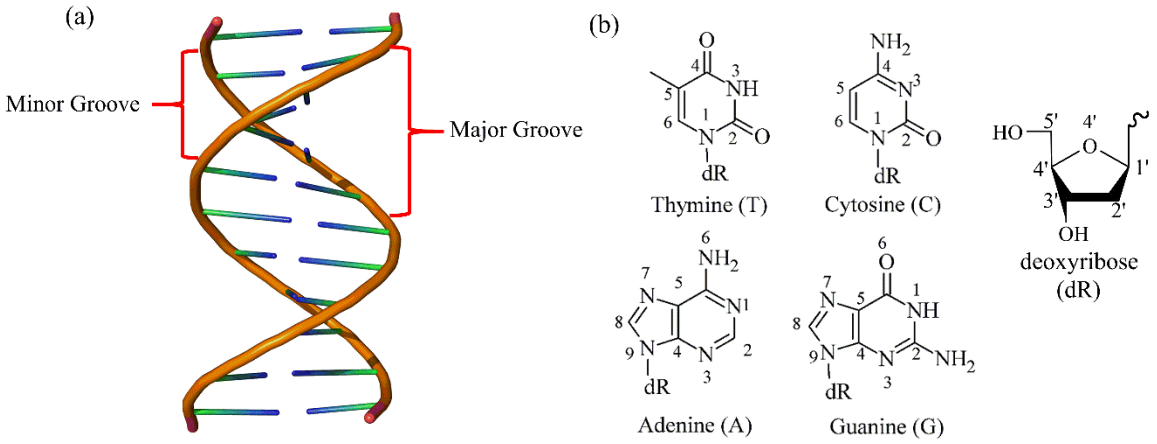


Figure 1.1: (a) Double-stranded DNA showing the major and minor grooves. (b) Structures and numbering of the four DNA nucleobases, namely the pyrimidines (C and T) and purines (A and G) and deoxyribose sugar (dR).

The two strands of DNA are held together by complementary bases bound by Watson-Crick (WC) base pairing (Figure 1.2), where A binds with T and C binds with G, as well as stacking between consecutive base pairs. For WC hydrogen bonding to form, the nucleobases must adopt an *anti* orientation about the glycosidic bond,

which is defined by the dihedral angle $\chi = 180 \pm 90^{\circ}$ (Figure 1.3), where $\chi = \angle(O4'C1'N9C4)$ for the purines and $\angle(O4'C1'N1C2)$ for the pyrimidines. In the DNA double helix, the minor groove exposes the sugar edge of the WC base pair involving the glycosidic bonds of both the bases, and the major groove exposes the Hoogsteen edge of the base pair to the cellular environment (Figure 1.4).

Rotation of the base about the glycosidic bond changes the orientation to *syn* ($\chi = 0 \pm 90^{\circ}$, Figure 1.3). If the nucleoside adopts the *syn* orientation, WC base pairing cannot take place, but rather the Hoogsteen face (Figure 1.3) of the base becomes available for pairing. This may lead to mispairing upon replication, causing mutations within the DNA.

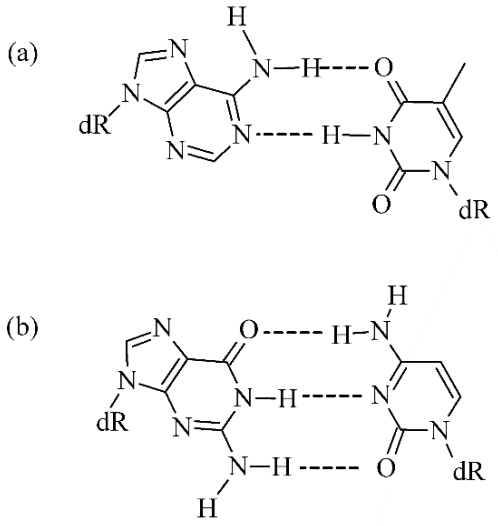


Figure 1.2: Watson-Crick hydrogen bonding in the (a) A:T and (b) G:C base pairs in DNA.

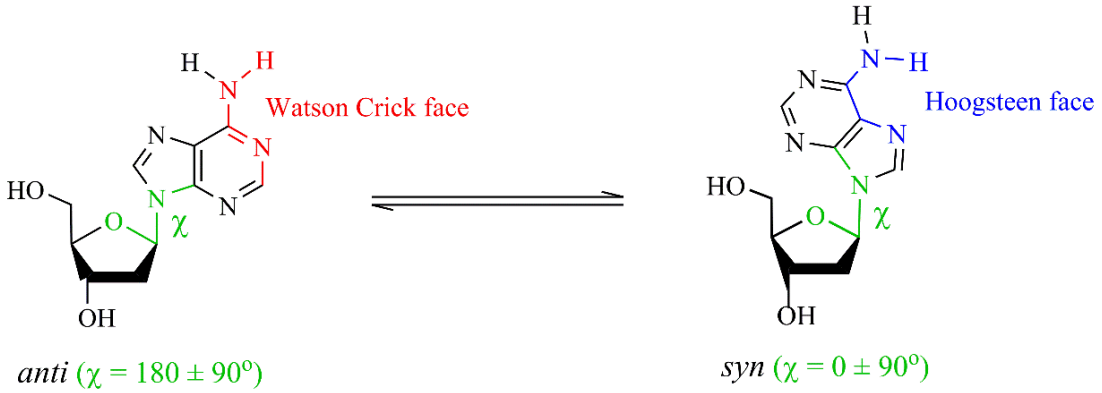


Figure 1.3: The* anti *and* syn *orientations about the glycosidic bond in a nucleoside. The dihedral angle χ dictates the conformation of the base with respect to the sugar.

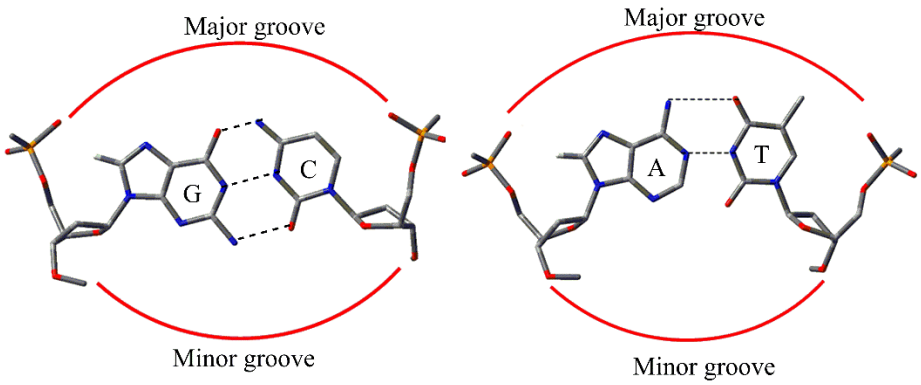


Figure 1.4: Edges of the nucleobase pairs facing the major and minor grooves in the* anti *glycosidic orientation.

1.3 Types of DNA Helices

DNA generally exists in one of three helical structures, namely B-DNA, A-DNA or Z-DNA. B-DNA is biologically predominant. It is a right-handed double helix consisting of Watson-Crick base pairs that are nearly perpendicular to the helical axis. It has a 36˚ average twist and 10 base pairs per turn. A-DNA is also a right-handed helix, but is wider and flatter than B-DNA, with 11.6 base pairs per turn. In contrast, Z-DNA is a left-handed helix that contains 12 base pairs per turn. Furthermore, the base pairs in Z-DNA are flipped by 180˚ with respect to the base pairs in B-DNA.

Therefore, alternating purine and pyrimidine residues in the same strand are *anti* and *syn* in Z-DNA.

1.4 DNA Mutations

Upon replication, there are generally two types of mutations possible. The first category is base substitution mutations, which alter one base in a DNA sequence. These can be transition mutations, which substitute a pyrimidine for another pyrimidine or a purine for another purine. Alternatively, transversion mutations can occur, which substitute a purine for a pyrimidine or vice versa (Figure 1.5). A second category of DNA mutations is frameshift mutations. These mutations result from either insertion or deletion of nucleotides within the DNA, causing a shift in the reading frame of the codon. This may lead to an alteration in the amino acid sequence during protein synthesis (Figure 1.6).

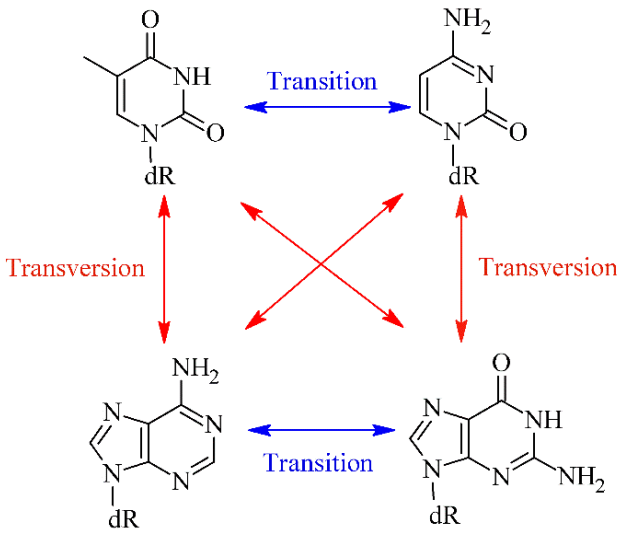


Figure 1.5: Base-substitution transition (blue) and transversion (red) mutations that can occur upon DNA replication.

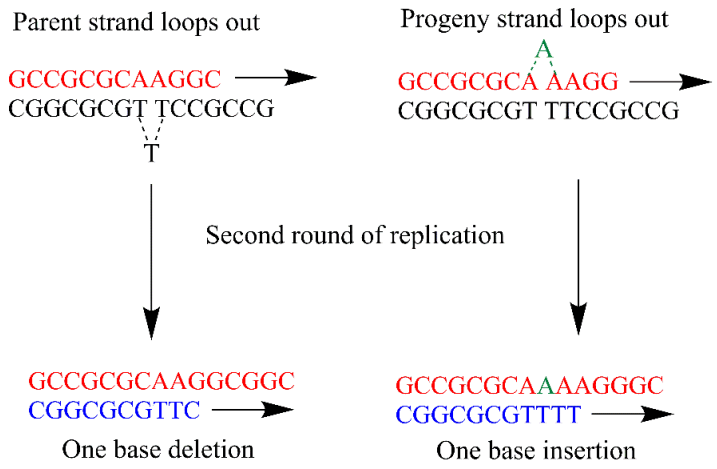


Figure 1.6: One base deletion (left) or insertion (right), leading to frameshift mutations.

1.5 Overview of DNA Damage

In cells, the genetic integrity of DNA is constantly assaulted by various endogenous and exogenous sources. Common forms of nucleobase damage include dimerization,[1] oxidation,[2] alkylation,[3] and the formation of bulky adducts (addition products).[4-39] Such damage, if unrepaired, may lead to mispairing of the DNA nucleobases upon replication, causing mutations that disrupt the genetic integrity of the cell, and possibly block DNA replication and transcription. For example, the dimerization of the cytosine base occurs on sunlight exposure, which (if unrepaired) may lead to CC → TT mutations.[1] Additionally, oxidation of guanine is a common occurrence in the cell. Specifically, hydroxyl radicals formed during metabolism commonly attack the C8 position of G and form 8-oxoguanine (8OG).[2] Due to stabilization of the *syn* orientation, 8OG preferentially mispairs with A upon replication through Hoogsteen hydrogen bonding. Hence, 8OG may lead to G → T transversion mutations if unrepaired prior to replication.[40] Another frequent type of

damage to the DNA nucleobases is alkylation. For example, vinyl chloride, an environmental carcinogen, interacts with adenine to form N-ethenoadenine, which may lead to A→T transversion mutations upon replication.[3] Additionally, DNA nucleobases can be modified by the formation of bulky DNA adducts,[4-39] which are discussed in detail in the next section and are the focus of the present thesis.

1.5.1 Bulky Adduct Formation

DNA adducts are formed when reactive intermediates containing one or more aromatic ring(s) attack a DNA nucleobase. Various chemical substances found in the environment have been identified by the International Agency for Research on Cancer (IARC) to be carcinogenic to humans. The carcinogenic activity of these compounds has been completely or partly associated with the genotoxicity caused by the formation of bulky adducts.

In cells, bulky DNA adduct formation involves the addition of diverse carcinogenic species to various sites in a DNA nucleobase, leading to different mutagenic consequences.[9,26-28,32-39] For example, bulky lesions formed from aryl hydrazines found in mushrooms[28] can lead to the removal of the (damaged) base due to a decrease in the stability of the glycosidic bond upon adduct formation, which may further lead to transversion mutations.[27] Similarly, adducts formed upon exposure to polycyclic aromatic hydrocarbons (PAHs) found in tobacco smoke lead to depurination, and subsequently transversion mutations.[26] On the other hand, bulky adducts formed from a variety of aromatic amines, which can be found in pesticides, insecticides, and charred meat, have been implicated in base substitution,[37,38] as well as frameshift,[37,39] mutagenesis. Furthermore, adducts formed by exposure to

ubiquitous phenolic toxins, including chlorophenols[33] and the well-known food toxin ochratoxin A,[34-36] cause deletion[32] and base-substitution[9] mutations.

In the literature, most bulky DNA adducts have been reported to form with the purine nucleobases,[4-39] with damage to the pyrimidine bases being less prevalent.[41-43] Hence, this thesis and the subsequent sections of this chapter focus on DNA adducts formed by the attachment of a carcinogen moiety at different positions in a purine and the conformational preferences of adducted DNA containing purine lesions.

1.5.2 Bulky Adduct Formation with the Purine Bases

For A, the bulky moiety binds at either the N3, C8, N7, N1 or N6 position (Figure 1.1b), while adducts are formed at the N3, C8, N7, N1, O6 or N2 position of G (Figure 1.1b). Based on the site to which the bulky moiety binds, carcinogens or mutagens can be classified into three main categories. The first category binds to same positions in both purines. For example, a potent antitumor antibiotic, CC-1065, alkylates the N3 position of both A[11] and G,[10] and has been established to be cytotoxic.[10,11] A nitro-polycyclic aromatic hydrocarbon (nitro-PAH), 2-nitropyrene, which is a genotoxic environmental pollutant found as an ambient particulate in the air, forms adducts at the C8 position of A and G.[8] Another chemical compound, benzo[*a*]pyrene-7,8-dione (B[*a*]P-7,8-dione), which is formed from oxidation of the PAHs present in car exhaust and cigarette smoke, forms adducts at N2 or N7 of guanine[24] and N6 or N7 of adenine.[7]

The second category of carcinogens or mutagens bind to both A and G, but at different positions. For example, styrene oxide, which is the main metabolite of styrene (used in the polymer industry) in humans, forms adducts at the N7 position of G and N1 or N6 position of A.[6] 1,3-butadiene (BD), another chemical that is widely used in the polymer industry, is also known to form adducts at N1[5] or N2[21] of A and N7 of G.[20] Dibenzo[*a,l*]pyrene (DB[*a,l*]P), which is present in vehicle exhaust and cigarette smoke, a synthetic estrogen 2-hydroxyestrone (2-OHE_1),[25] and aristolochic acids, which are plant mutagens,[44] bind to the exocyclic amino groups of A (at N6) and G (at N2).

The third category is carcinogens or mutagens that predominantly bind to either the G or A nucleobase. For example, the acridine-targeted nitrogen mustards used in cancer chemotherapy predominantly form N1 adenine adducts.[4] Alternatively, other carcinogenic chemical species, such as the aromatic amines,[15,22,37-39] aryl hydrazines,[28] and ochratoxin A,[34] are known to predominantly form adducts at the C8 position of guanine.

It has been observed that chemically distinct adducts have different conformational preferences, which can lead to different biological consequences. Hence, to understand the exact mechanism of carcinogenicity of the adducts, it is vital to look at the possible conformational outcomes of the associated adducted DNA (discussed in detail in the next section).

1.5.3 Adducted DNA Conformations and Biological Implications

The bulky moiety attached to the nucleobase in a DNA adduct can broadly be present at two sites in adducted DNA. Specifically, the bulky group can be either

extrahelical (outside the helix) or intrahelical (intercalated within the helix, a position stabilized by stacking with the flanking bases). In the extrahelical location, the bulky moiety can be present either in the major[45] or minor[46] groove (Figure 1.7). In the intrahelical position, the bulky moiety can either stack in the helix, rendering the base opposing the lesion extrahelical (called base-displaced intercalated conformer),[47] or intercalate at a position 5′ (called the 5′–intercalated conformer)[48] or the 3′ (called the 3′–intercalated conformer)[48] with respect to the damaged base (Figure 1.7). Regardless of the conformer adopted, all bulky moiety orientations alter the structure of DNA.

In order to maintain the genetic integrity of the cell, it is crucial for the DNA structure to stay intact. Therefore, our body has intricate repair mechanisms to counter the effects of DNA damage. Bulky DNA lesions are most commonly repaired through the nucleotide excision repair (NER) pathway, which removes a 24–32 nucleotide long section of the lesion-containing strand.[49] NER recognition in eukaryotes is believed to involve identification of structural perturbations at the lesion site by XPC-RAD23B sensors, followed by binding of these sensors with the damaged site to form a stable DNA recognition complex.[50-52] Studies on a variety of bulky DNA adducts have revealed that a greater NER propensity correlates with perturbations in a number of local structural features upon DNA damage.[53-56]

In the event that DNA damage escapes repair, a lesion can induce permanent changes within the DNA structure, which can lead to mutations upon replication. Replication of DNA adducts is mainly carried out by two types of polymerases: the

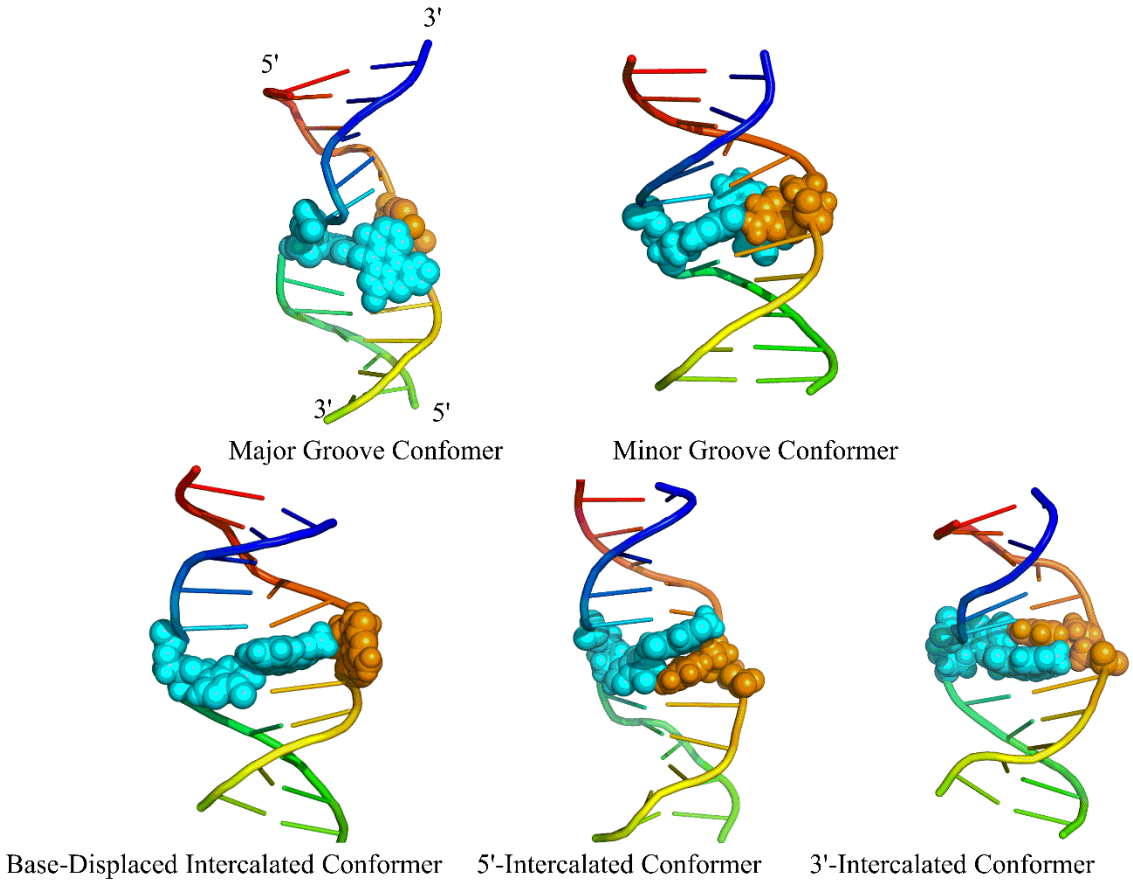


Figure 1.7: Different possible conformations of adducted DNA (the adduct is shown in cyan and the opposing base is shown in yellow).

replicative polymerases (which replicate the natural or damaged DNA nucleobases) and the translesion synthesis (TLS) polymerases (which replicate a damaged base if the replicative polymerases fail). These polymerases have different active site structures and hence can lead to different mutagenic outcomes for a particular lesion.[57,58] Specifically, replication by a TLS polymerase is more error-prone due to the flexible active site of these enzymes.[59] Since these polymerases utilize sterics, as well as hydrogen-bonding, interactions to insert the correct nucleotide opposite a

lesion, a modification in the DNA structure can alter these interactions, and hence cause mutations.[60-62]

Since each adduct may perturb the DNA structure to a different extent, looking at the conformational outcomes of adducted DNA is crucial for understanding the 'repair versus persistence' and mutagenicity of DNA adducts. In fact, previous studies on select bulky purine adducts have revealed that the first step toward understanding the carcinogenic effect of a particular lesion is to identify the associated changes to DNA structure.[15,22,31,48,60,63-77] The following section therefore highlights the various factors that can bias the conformational outcome of adducted DNA. Although the list of the effects discussed in the next section is comprehensive, select examples have been chosen to illustrate the potential effect, with additional examples discussed in the appropriate context throughout the thesis.

1.6 Factors that Affect the Conformational Preferences and Stability of Adducted DNA

1.6.1 Ionization State

Although the adduct ionization state does not entirely alter the conformation adopted by adducted DNA, it can lead to structural changes within the same conformational theme, which alter the stability of adducted DNA. This occurs because the electrostatic interactions of the adduct with the flanking base pairs vary with changes in adduct ionization state. To the best of my knowledge, only one adduct (i.e., the C8-linked ochratoxin-A guanine adduct, OTB-dG, Figure 1.8) has been studied to elucidate the effect of the ionization state on the conformational outcome. At the G^2 position of the *Nar*I sequence ($CG^1G^2CG^3CC$), the neutral form of the OTB-dG adduct

has the highest lesion site stacking energies compared to the monoanionic (–COOH group ionized) or dianionic (both the –COOH and –OH ionized) forms in the minor groove conformer. This is attributed to greater electrostatic repulsion between the bulky moiety and the opposing C in the anionic forms. Hence, the cytosine opposing the damage is better penetrated into the helix in the neutral form compared to the anionic analogues.[23] This leads to increased stacking and greater stabilization of the minor groove conformer for the neutral adduct compared to the anionic forms, even though this conformation is accessible to all ionization states of OTB-dG adducted DNA.

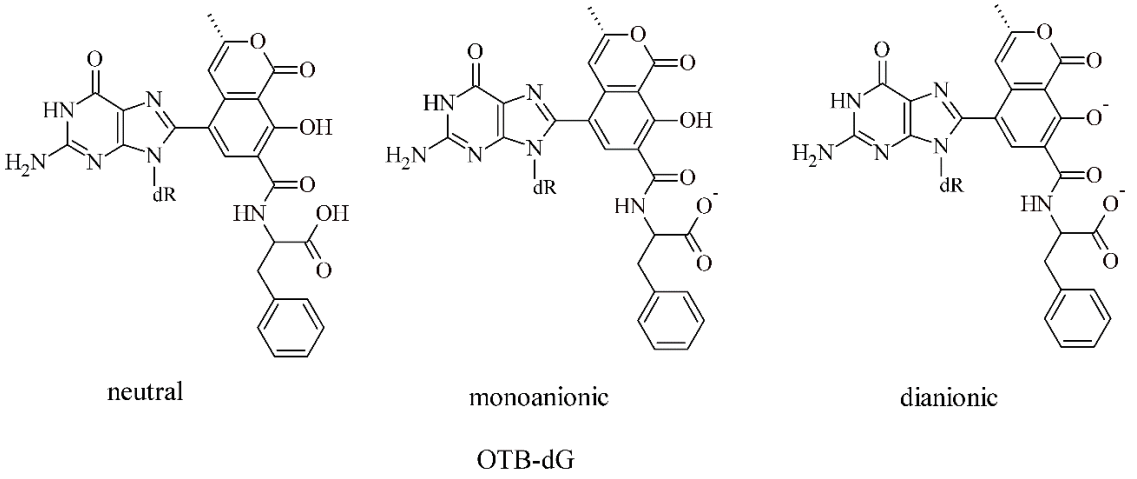


Figure 1.8: Example of an adduct (OTB-dG) that shows differential conformational stability of adducted DNA with a change in the ionization state.

1.6.2 Presence and Location of Functional Group

A functional group on the carcinogen-purine linkage can alter the discrete interactions between the adduct and the flanking nucleobases, and hence affect the stability of a particular conformer. For example, the AAF and AF aromatic amine

adducts (Figure 1.9) differ by an acetyl substituent attached to the linker N atom, and three conformations of AAF-dG adducted DNA are stable (namely, major groove, base-displaced stacked and minor-groove wedge) versus two conformations of AF-dG adducted DNA (namely, major groove and base-displaced stacked).[78] In the predominant base-displaced stacked conformation of AAF-dG adducted DNA, steric repulsion due to the acetyl group distorts the orientation of the deoxyribose sugar in the adducted nucleotide such that the damaged guanine becomes perpendicular to the helix axis. This diminishes the stacking energies between the adducted nucleotide and the flanking bases compared to AF-dG, where the damaged guanine is nearly coplanar with respect to the nucleobase and contributes to stacking.[78] Hence, the base-displaced intercalated conformer associated with AF-dG is more stable than AAF-dG.

The same functional group at two distinct locations in the bulky moiety can also be vital for determining the conformational preferences of the adducted DNA. For example, in the major groove conformations associated with C-linked *o*-phenoxyl-dG (*o*-PhOH-dG, Figure 1.9) adducted DNA, the damaged nucleotide adopts only the high *anti* ($-60^\circ > \chi > -90^\circ$) orientation because of a stabilizing O–H···O hydrogen bond between the hydroxyl group in the bulky moiety and O in the phosphate attached to the 5′ base.[79] On the other hand, the major groove conformer for C-linked *p*-phenoxyl-dG (*p*-PhOH-dG, Figure 1.9) adducted DNA spans both the *anti* and high *anti* orientations about the glycosidic bond in the adduct, since the hydroxyl group at the *para* position is not in proximity of the O atom of the phosphate to form a similar stabilizing interaction.[79]

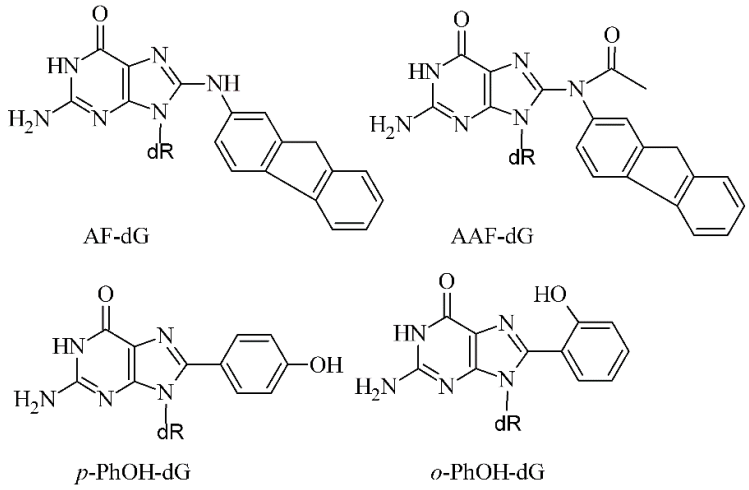


Figure 1.9: Examples of adducts that show differential conformational preference and stability of adducted DNA with the presence or change in location of a functional group.

1.6.3 Identity of the Flanking Bases

The effect of the flanking bases on the conformational preferences of adducted DNA vary with the adduct under consideration. For most adducts, changing the sequence simply alters the stability of the adducted DNA.[23,80-83] For example, the intercalated conformer of AF-dG (Figure 1.10) in the *Nar*I sequence is most stabilized with the lesion at G^3 due to better stacking between the bulky moiety and the flanking bases in the opposing strand than at G^1 or G^2.[81,82] Furthermore, changing the flanking bases has also been shown to completely alter the conformational preference of damaged DNA.[84,85] For example, the *Nar*I sequence-containing adducted DNA acquires a minor groove conformation when the IQ-C^8-dG (2-Amino-3-methylimidazo[4,5-f]quinolone, Figure 1.10), aromatic amine adduct is located at G^1 or G^2, but a base-displaced intercalated conformation with the lesion at G^3.[85]

However, the structural reasons for these distinct conformational preferences remain unclear.

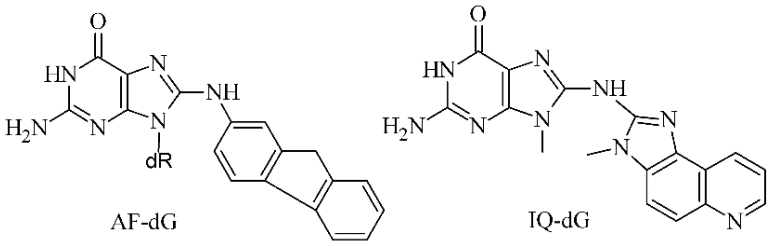


Figure 1.10: Examples of adducts that show differential conformational preference and stability of adducted DNA with a change in the identity of the flanking bases.

1.6.4 Size of the Bulky Moiety

Most carcinogens are composed of one or more aromatic ring systems containing a delocalized π-electron cloud.[21,23,29,79,83,84,86-89] Adducts having a larger polycyclic ring system tend to better stabilize an intercalated conformer compared to an extrahelical conformation. This is because the π-electrons of the aromatic rings can form stable π-π interactions with the flanking bases, which compensate for the helix destabilization induced by the intrahelical position of the bulky moiety. For example, the N6-linked tetrahydrobenzo[*a*]pyrenyl PAH adduct of adenine (B[*a*]P-dA, Figure 1.11) prefers an intercalated conformation,[83] while the single-ringed N6-linked styrene oxide adenine adduct (SO-dA, Figure 1.11) prefers the major groove conformer.[45] Similarly, in the case of the aniline (AN-dG), aminofluorene (AF-dG) and aminopyrene (AP-dG) guanine adducts (Figure 1.11), the intercalated conformer of adducted DNA becomes increasingly stable in the order of AN-dG < AF-dG < AP-dG.[89]

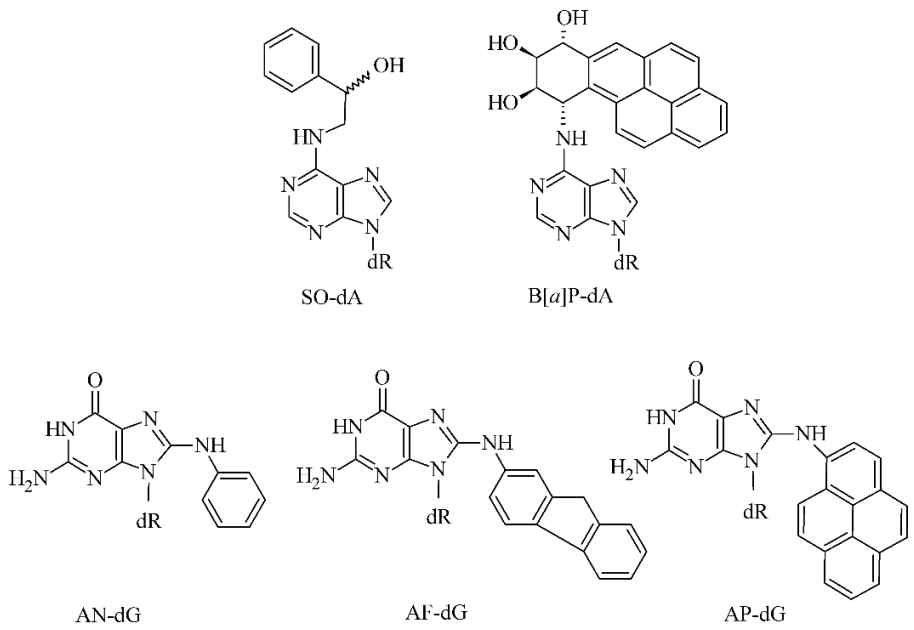


Figure 1.11: Examples of adducts that show differential conformational preference and stability of adducted DNA with a change in the size of the bulky moiety.

1.6.5 Stereochemistry

Stereochemically isomeric adducts can have different conformational preferences of adducted DNA due to variable steric interactions with the flanking base pairs within each conformer. For example, the N6-linked PAH-adenine adducts of benzo[*c*]phenanthrene (B[*c*]Ph)[75,86] and dibenzo[*a,l*]pyrene (DB[*a,l*]P),[48] as well as the benzo[*a*]pyrene adduct (B[*a*]P)[83,88] (Figure 1.12), prefer intercalated conformations for both stereochemically isomeric adducts. However, the bulky moiety with the S absolute configuration is intercalated on the 3′ side of the damaged base, while the isomer with the R absolute configuration is intercalated on the 5′ side of the damaged base. Notably, no structural explanation is known for these differential conformational preferences. In the different conformations adopted, the S isomer of the adduct imposes more unfavorable steric interactions in DNA compared to R. Therefore, the S isomer causes greater structural perturbations

within DNA and is more destabilizing than the R isomer. This has been termed the 'S destabilization effect'.[13] Similarly, for the N2-linked PAH-guanine adducts of 14R and 14S stereoisomeric dibenzo[*a,l*]pyrene (DB[*a,l*]P-dG) (Figure 1.12), the R isomer adopts an intercalated conformation, where the bulky moiety stacks on the 3′ side of the modified base. In contrast, the bulky moiety of the S stereoisomer stays in the minor groove, with the aromatic rings oriented on the 5′ side of the modified base because of unfavorable steric interactions with the flanking base pairs.[90]

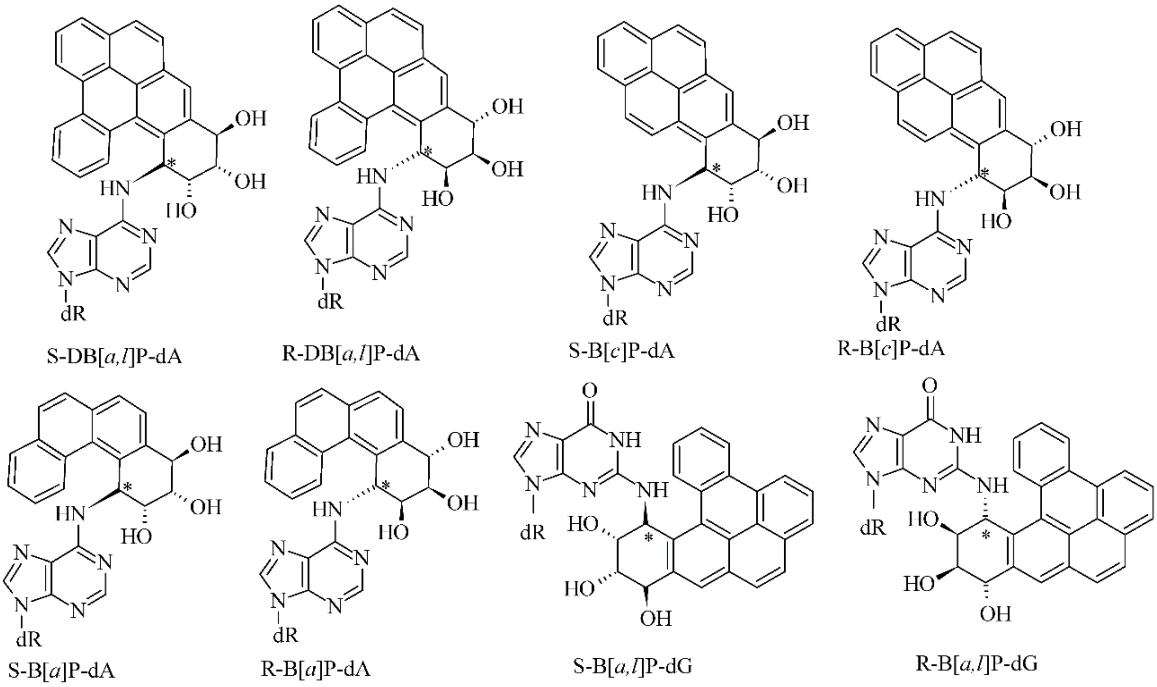


Figure 1.12: Examples of adducts that show differential conformational preference and stability of adducted DNA with a change in the stereochemistry

1.6.6 Linkage Type

The bulky moiety can bind to the nucleobase via different linker atoms (usually N, O or C).[79] The identity of the linker atom can change the discrete interactions within adducted DNA and hence affect the relative conformational

stabilities. Specifically, the N or O linker atoms are electronegative, and therefore, these linkers are capable of forming hydrogen bonds in adducted DNA conformations, which can stabilize a particular conformer and render other conformations inaccessible to DNA. In contrast, the linker C atom is unable to form such interactions and would likely be conformationally heterogeneous. For example, the C8-guanine adducts of aniline (AN-dG), phenoxyl (PhO-dG) and phenyl (Ph-dG) all contain a single bulky ring system (Figure 1.13), but differ in the type of linker atom. Although all three prefer a major groove conformation of adducted DNA, the minor groove conformation is also close in energy to the most stable conformer for C-linked phenyl (Ph-dG) adducted DNA.[79] Nevertheless, studies of polycyclic adducts needs to be performed on, for example, the N, O and C-linked phenanthrene and pyrene to gain better insight into the effect of linkage on the conformational preference and stability of adducted DNA.

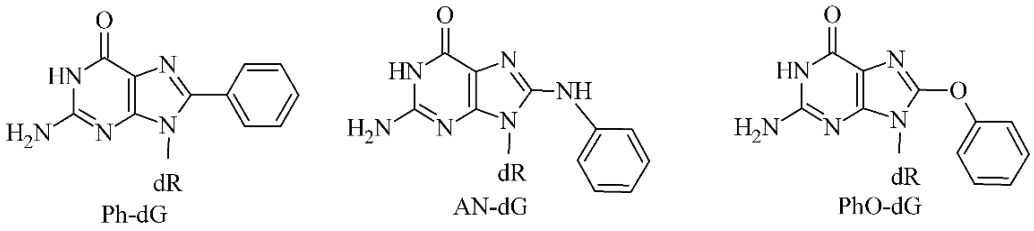


Figure 1.13: Examples of adducts that show differential conformational preference of adducted DNA with a change in the linker atom.

1.6.7 Site of Attachment

As highlighted in Section 1.5.2, the bulky moiety can attach to different positions of the nucleobases. Depending on the site of attachment and orientation about the adduct glycosidic bond, the position of the bulky moiety within the helix can change. For example, the aromatic amine IQ binds to both C8 and N2 of guanine

(Figure 1.14). Within the same sequence, both C8 and N2-linked IQ-dG adducted DNA prefer the base-displaced intercalated conformation. However, due to differences in the site of IQ attachment, the C8-linked IQ-dG adduct adopts a *syn* glycosidic orientation, while an *anti* glycosidic orientation is observed for the N2-linked IQ-dG adduct.[47]

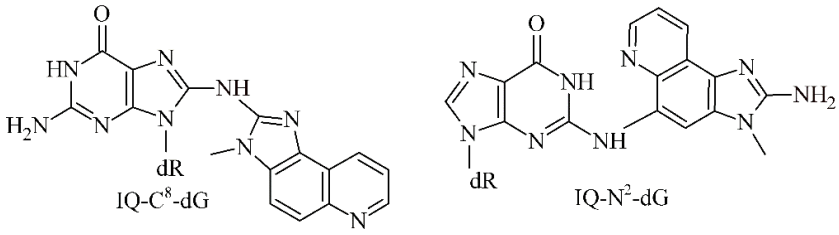


Figure 1.14: Example of adducts that show differential conformational preference of adducted DNA with a change in the site of bulky moiety attachment.

In summary, the adducted DNA conformation can be affected by different factors, but the magnitude of the effect of each factor depends on the adduct considered. Despite observed variations in adduct conformational preferences and the stability of the associated adducted DNA, it is currently unclear how the effects depend on the chemical structure of the adduct. Therefore, in order to understand the conformational preferences of adducted DNA, each adduct has to be extensively studied through incorporation into a DNA helix. Although many of such studies have been carried out for a variety of adducts, it is important to emphasize that there are other known adducts for which experimental and computational data is lacking. These include adducts formed from the naturally-occurring plant carcinogens, AAs. Therefore, the present thesis focuses on the DNA damage caused by exposure to

aristolochic acids. The subsequent section discusses the occurrence of AAs, as well as the mutagenicity and repair of adducts formed by these carcinogens.

1.7 Aristolochic Acids: Potent Plant Mutagens

1.7.1 Occurrence

AAs are produced by species of the *Aristolochia* and *Asarum* plants,[91] and have been used in traditional herbal medicines since ancient times.[91,92] Indeed, these plants have been reported to contain antibacterial, antiviral, antifungal and antitumor properties.[93] As a result, the plant roots have traditionally been used to treat conditions such as stomach aches, hypertension, tooth aches, eczema and poisonous snake bites.[91]

In 1991, AAs were determined to be responsible for Chinese Herb Nephropathy (CHN), a rapidly progressive kidney fibrosis associated with the prolonged intake of Chinese herbs contaminated with *Aristolochia fangchi*.[94] In addition, exposure to AAs has been linked to Balkan endemic nephropathy (BEN) through the consumption of locally grown wheat contaminated with seeds from *Aristolochia clematitis*.[95,96] Due to the striking similarities in the clinical expression and pathology of lesions associated with CHN and BEN, the term aristolochic acid nephropathy (AAN) is now proposed to cover both conditions.[97] In addition to their nephrotoxic potential, exposure to AAs has been associated with urothelial carcinoma in the upper urinary tract (UUC).[98-100] Due to their potential carcinogenic effects, the International Agency for Research on Cancer (IARC) categorized herbal medicines containing the plant species of *aristolochia* genus as Group 1 (human) carcinogens.[101] AAs have a strikingly high somatic mutation rate (150 mutations/Mb), exceeding

smoking-associated lung cancer (8 mutations/Mb) and ultraviolet radiation–associated melanoma (111 mutations/Mb).[102] Chemically, the extracts of the these plants contain a mixture of structurally similar nitrophenanthrene carboxylic acids, mainly AAI and AAII, which differ in the presence or absence of a methoxy group attached to the polycyclic system (Figure 1.15).[103] These compounds are metabolically activated within the body and damage DNA by forming adducts. The details of activation and adduct formation by these carcinogens are discussed in the next section.

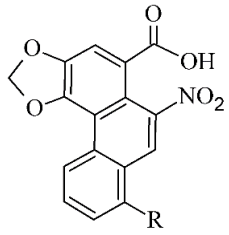


Figure 1.15: Structure of aristolochic acids (R = OCH_3 for AAI and H for AAII).

1.7.2 Aristolochic Acid Associated Mutagenesis

Metabolic activation of AAs has been shown to involve the NAD(P)H:quinone oxidoreductase catalyzed reduction of the nitro group to yield N-hydroxyaristolactams (N-hydroxy ALs), which further hydrolyze to form nitrenium ions.[104-106] These intermediates specifically attack the purine DNA nucleobases to form the AL-N^6-dA and AL-N^2-dG adducts (Figure 1.16),[105,107] which were identified in the renal tissues of patients affected by BEN.[108] The AA-initiated carcinogenesis is associated with 'signature' A→T mutations,[102,109,110] which are dominant in the TP53 tumor suppressor gene[108] and the FGFR3 or HRAS protooncogene.[99] This carcinogenesis has been observed to be sequence dependent. Specifically, recent

experimental whole exome sequencing studies have shown that the A→T mutations are most commonly observed in the YXG (where Y = T or C and X = AL-N^6-dA) sequence.[102,111]

Furthermore, despite similarities in structure, the adenine adducts formed from AAI have been observed to have a higher abundance[106,112,113] and greater nephrotoxic potential[114] compared to ALII adducts in cells. In addition, *in vitro* studies of the replication of these adducts indicate that both adenine and guanine lesions misincorporate A, leading to A→T or G→T tranversion mutations for the dA or dG adduct, respectively.[115] However, the adenine adducts have greater overall persistence[16,116,117] and mutagenicity [115,118] compared to guanine adducts in cells. This greater persistence and mutagenicity of the adenine adducts has been related to their higher repair resistance compared to the corresponding guanine adduct. The details of experimental studies carried out on the repair of these lesions are discussed in the next section.

1.7.3 Repair of the AL-DNA Adducts

As stated in Section 1.5.3 of the present thesis, bulky DNA adducts, including those formed by aristolochic acids, are primarily repaired by the NER pathway. This pathway operates by two different mechanisms, *viz*. transcription-coupled repair (TCR) and global genomic repair (GGR),[119-121] which primarily differ in two ways. First, TCR only repairs transcriptionally active regions, while GGR repairs lesions in the entire genome. Second, damage recognition in TCR is associated with transcription inhibition (i.e., stalling of an RNA polymerase), while GGR involves

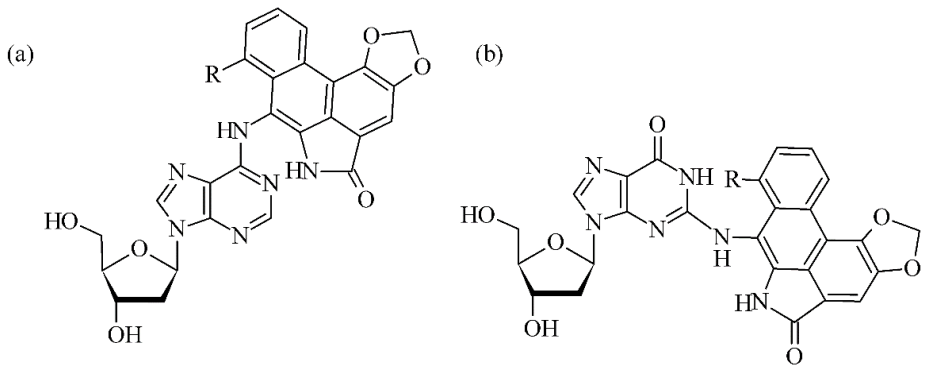


Figure 1.16: Structure of the (a) AL-N^6-dA and (b) AL-N^2-dG adducts (R = OCH$_3$ for ALI and H for ALII).

specific damage recognition factors (i.e., XPC-RAD23B in eukaryotes). Nevertheless, the subsequent repair steps are similar in both mechanisms.

For adducts formed by AAs, a recent experimental study by Sidorenko and coworkers observed a small, but significant, change in the levels of the ALII-N^2-dG adduct in cells deficient in either TCR or GGR compared to control cells.[117] This implies that the ALII-N^2-dG adduct is repaired by both TCR and GGR, which correlates with the high rate of AL-N^2-dG removal in target tissues.[16] On the other hand, investigations on the toxicity of ALII in cell lines with deficiencies in TCR or GGR or both pathways point towards the GGR resistance of ALII-N^6-dA.[117] Furthermore, a number of recent studies on patients exposed to AA[99,102,110,111] have revealed that the signature A→T mutation of AL-N^6-dA adduction is predominantly located in non-transcribed DNA strands.[102,111] This conspicuous lack of A→T mutations in transcriptionally active regions of the genome suggests that the associated AL-N^6-dA adduct is actively repaired by the TCR pathway.[117] However, *in vitro* binding studies of the eukaryotic GGR damage sensing factor (XPC-RAD23B) to oligonucleotides

containing ALII-N^6-dA reveal that these adducts lack GGR recognition.[117] The proposed GGR resistance of the ALII-N^6-dA adduct is in synchrony with its greater overall persistence[16,116,117] and mutagenicity[115,118] compared to ALII-N^2-dG in cells.

Although many experimental studies have been carried out to study the carcinogenicity of aristolochic acids, to the best of my knowledge only one structural study using NMR has been accomplished for the ALII-N^6-dA adduct.[122] Hence, the structural information about these adducts to date is inadequate to explain their experimentally-observed differential mutagenicity of the AAI versus AAII adenine adducts, the difference in the repair propensity of the adenine versus guanine adduct of the aristolochic acids, as well as the sequence dependent mutagenicity of the adenine adducts. Due to the complexity synthesizing large biomolecules and hazards handling carcinogenic substances, computational chemistry has served as an important tool to analyze the conformational outcomes of adducted DNA,[31,60,63,87,89,123,124] establish the sequence effect on the conformational preferences,[23,46,80-82,85,123,125,126] and provide a structural basis to explain the relative repair propensities of adducts.[13,18,56,74,78,127,128] Computational approaches are used in the present thesis, to analyze the molecular level details of adducted DNA conformations, and thereby explain the 'repair versus persistence' and the mutagenicity associated with the aristolochic acid adducts.

1.8 Thesis Objectives and Methodology

The focus of the present thesis is to analyze in detail the structure of the aristolochic acid purine adducts and explain the experimentally-observed mutagenicity of the AAI versus AAII adducts, the relative repair propensity of the

guanine and adenine adducts, and the sequence dependent mutagenicity of the adenine adducts. To accomplish this goal, the conformational preferences of the AL-adducts will be analyzed by systematically increasing the model size and studying each model at different levels of computational theory. The adenine and guanine adducts are initially studied with nucleobase, nucleoside and nucleotide models using quantum mechanical (QM) calculations, specifically density functional theory (DFT). DFT has comparable accuracy to *ab initio* QM methods, but is less computationally expensive. Previous studies on DNA adducts using DFT agreed well with experimental findings for the isolated nucleobase, nucleoside and nucleotide models.[71-73] In fact, DFT methods and small models have provided important information about DNA adducts, such as the relative conformational preference of the bulky moiety with respect to the nucleobase, and the effect of the sugar and phosphate on the conformational preference of the adduct. [71-73,129,130]

Following the study of small DNA models, the computational model is grown to consider an 11-mer DNA oligonucleotide with molecular dynamics (MD) simulations. MD employs empirical molecular mechanics designed from classical physics to calculate the forces between two interacting particles. These methods are less computationally expensive (faster) than QM methods. In MD simulations, the trajectories of molecules and atoms are determined by numerically solving the Newton's equations of motion for a system of interacting particles and can be used to calculate time averaged properties of the system. The accuracy of MD simulations comes from carefully parameterizing the molecular mechanics force field using quantum mechanical and/or experimental data. Due to computational efficiency, MD

simulations can be used to study the conformation of macromolecules in solvent environments. Therefore, the present thesis uses this approach to analyze the conformational preferences of the adenine and guanine adducts of AAs, as well as DNA containing these adducts.

1.9 Thesis Outline

The study was initiated by looking at the more mutagenic adenine adducts formed by aristolochic acids in Chapter 2 of this thesis. This chapter focuses on comparing the structural features of the adenine adducts formed by the two predominantly occurring aristolochic acids (AAI and AAII). The structural study investigates the nucleobase, nucleoside and nucleotide models using DFT calculations. MD simulations and free energy calculations reveal the conformational preferences for these adducts within DNA. The different conformer(s) adopted by the adducted DNA are compared to other N6-linked adenine lesions, including those derived from polycyclic aromatic hydrocarbons. The smaller calculated energy difference in the accessible conformations of ALI than ALII adducted DNA provides insight into the higher mutagenicity of AAI compared to AAII. In addition, the structural characteristics of the preferred conformations of adducted DNA are explained and provide a structural basis for the resistance of these adducts to repair.

Chapter 3 focuses on the intrinsic structural preferences of the aristolochic acid guanine adducts studied at the nucleobase, nucleoside and nucleotide models using DFT calculations. These intrinsic structural preferences are then compared to the adenine adducts studied in the previous chapter. Furthermore, the relative conformational preferences of ALII-N^6-dA and ALII-N^2-dG adducted DNA are

compared by carrying out MD simulations with the guanine adduct incorporated into the analogous DNA sequence studied in Chapter 2. Comparing the structural perturbations induced to DNA by the guanine and adenine adducts provides a structural basis for their experimentally-observed differential repair propensities.

Chapter 4 of my thesis focuses on understanding how the DNA base sequence modulates the conformation of adducted DNA and its biological consequence by incorporating the ALI-N^6-dA adduct (X) into three sequential motifs, namely GXC, GXG and CXG**,** opposite a complementary T. These sequences differ in the identity of the bases flanking the lesion. The results from this study are compared to ALI-N^6-dA adducted DNA in the CXC sequence context studied in Chapter 2. This study explain the effect of sequence dependent changes in the structural features of adducted DNA, and provides clues to its implications to the differential repair and mutagenicity of the ALI-N^6-dA adduct in different sequences.

Finally, in Chapter 5, the main conclusions and contributions of the present work to the field of the toxicity of aristolochic acids are summarized. The Future Work section highlights standing questions in the field and explains how these can be resolved in future studies using computational chemistry.

1.10 References

(1) Goodsell, D. S. *The Oncologist* **2001**, *6*, 298.

(2) van Loon, B.; Markkanen, E.; Hübscher, U. *DNA Repair* **2010**, *9*, 604.

(3) Long, C.; Kovvali, G.; Marion, M.J.; Brandt-Rauf, P.; Monaco, R.; Li, Y. *Plastics and carcinogenesis: The example of vinyl chloride*, **2012**; Vol. 11.

(4) Boritzki, T. J.; Palmer, B. D.; Coddington, J. M.; Denny, W. A. *Chem. Res. Toxicol.* **1994**, *7*, 41.

(5) Rodriguez, D. A.; Kowalczyk, A.; Ward, J. B.; Harris, C. M.; Harris, T. M.; Lloyd, R. S. *Environ. Mol. Mutagen.* **2001**, *38*, 292.

(6) Koskinen, M.; Hemminki, K. *Org. Lett.* **1999**, *1*, 1233.

(7) Huang, M.; Liu, X.; Basu, S. S.; Zhang, L.; Kushman, M. E.; Harvey, R. G.; Blair, I. A.; Penning, T. M. *Chem. Res. Toxicol.* **2012**, *25*, 993.

(8) Fu, P. P.; Miller, D. W.; Von Tungeln, L. S.; Bryant, M. S.; Lay, J. O.; Huang, K.; Jones, L.; Evans, F. E. *Carcinogenesis* **1991**, *12*, 609.

(9) Akman, S. A.; Adams, M.; Case, D.; Park, G.; Manderville, R. A. *Toxins* **2012**, *4*, 267.

(10) Mitchell, M. A.; Weiland, K. L.; Aristoff, P. A.; Johnson, P. D.; Dooley, T. P. *Chem. Res. Toxicol.* **1993**, *6*, 421.

(11) Lee, C. S.; Sun, D.; Kizu, R.; Hurley, L. H. *Chem. Res. Toxicol.* **1991**, *4*, 203.

(12) Baird, W. M.; Hooven, L. A.; Mahadevan, B. *Environ. Mol. Mutagen.* **2005**, *45*, 106.

(13) Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2012**, *51*, 1486.

(14) Cavalieri, E. L.; Rogan, E. G.; Li, K.-M.; Todorovic, R.; Ariese, F.; Jankowiak, R.; Grubor, N.; Small, G. J. *Chem. Res. Toxicol.* **2005**, *18*, 976.

(15) Cho, B. In *The Chemical Biology of DNA Damage*; Wiley-VCH Verlag GmbH & Co. KGaA: **2010**, p 217.

(16) Fernando, R.; Schmeiser, H.; Scherf, H.; Wiessler, M. *IARC Sci. Publ.* **1992**, 167.

(17) Geacintov, N. E.; Cosman, M.; Hingerty, B. E.; Amin, S.; Broyde, S.; Patel, D. J. *Chem. Res. Toxicol.* **1997**, *10*, 111.

(18) Hang, B. *J. Nucleic Acids* **2010**, *2010*, 29.

(19) Hinrichs, B.; Zahid, M.; Saeed, M.; Ali, M. F.; Cavalieri, E. L.; Rogan, E. G. *J. Steroid Biochem. Mol. Biol.* **2011**, *127*, 276.

(20) Koivisto, P.; Peltonen, K. *Chem. Biol. Interact* **2001**, *135–136*, 363.

(21) Kowal, E. A.; Seneviratne, U.; Wickramaratne, S.; Doherty, K. E.; Cao, X.; Tretyakova, N.; Stone, M. P. *Chem. Res. Toxicol.* **2014**, *27*, 805.

(22) Patel, D. J.; Mao, B.; Gu, Z.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 391.

(23) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(24) Shou, M.; Harvey, R. G.; Penning, T. M. *Carcinogenesis* **1993**, *14*, 475.

(25) Wang, L.; Hingerty, B. E.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **2004**, *17*, 311.

(26) Chakravarti, D.; Pelling, J. C.; Cavalieri, E. L.; Rogan, E. G. *Proc. Natl. Acad. Sci.* **1995**, *92*, 10422.

(27) Schlitt, K. M.; Sun, K. M.; Paugh, R. J.; Millen, A. L.; Navarro-Whyte, L.; Wetmore, S. D.; Manderville, R. A. *J. Org. Chem.* **2009**, *74*, 5793.

(28) Gannett, P. M.; Lawson, T.; Miller, M.; Thakkar, D. D.; Lord, J. W.; Yau, W. M.; Toth, B. *Chem. Biol. Interact.* **1996**, *101*, 149.

(29) Millen, A. L.; Sharma, P.; Wetmore, S. D. *Future Med. Chem.* **2012**, *4*, 1981.

(30) Lukin, M.; de los Santos, C. *Chem. Rev.* **2006**, *106*, 607.

(31) Cho, B. P. *J. Environ. Sci. Health., C* **2004**, *22*, 57.

(32) Kuroda, K.; Hibi, D.; Ishii, Y.; Takasu, S.; Kijima, A.; Matsushita, K.; Masumura, K.-i.; Watanabe, M.; Sugita-Konishi, Y.; Sakai, H.; Yanai, T.; Nohmi, T.; Ogawa, K.; Umemura, T. *Mutagenesis* **2014**, *29*, 27.

(33) Dai, J.; Sloat, A. L.; Wright, M. W.; Manderville, R. A. *Chem. Res. Toxicol.* **2005**, *18*, 771.

(34) Dai, J.; Wright, M. W.; Manderville, R. A. *J. Am. Chem. Soc.* **2003**, *125*, 3716.

(35) Pfohl-Leszkowicz, A.; Manderville, R. A. *Chem. Res. Toxicol.* **2011**, *25*, 252.

(36) Al-Anati, L.; Petzinger, E. *J. Vet. Pharmacol. Ther.* **2006**, *29*, 79.

(37) Bichara, M.; Fuchs, R. P. P. *J. Mol. Biol.* **1985**, *183*, 341.

(38) Shibutani, S.; Fernandes, A.; Suzuki, N.; Zhou, L.; Johnson, F.; Grollman, A. P. *J. Biol. Chem.* **1999**, *274*, 27433.

(39) Melchior, W. B.; Marques, M. M.; Beland, F. A. *Carcinogenesis* **1994**, *15*, 889.

(40) Kalam, M. A.; Basu, A. K. *Chem. Res. Toxicol.* **2005**, *18*, 1187.

(41) Ding, S.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2005**, *44*, 14565.

(42) Plna, K.; Segerbäck, D.; Schweda, E. K. H. *Carcinogenesis* **1996**, *17*, 1465.

(43) Giussani, A.; Serrano-Andrés, L.; Merchán, M.; Roca-Sanjuán, D.; Garavelli, M. *J. Phys. Chem. B* **2013**, *117*, 1999.

(44) Pfau, W.; Schmeiser, H. H.; Wiessler, M. *Carcinogenesis* **1990**, *11*, 313.

(45) Hennard, C.; Finneman, J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **2001**, *40*, 9780.

(46) Rodríguez, F. A.; Cai, Y.; Lin, C.; Tang, Y.; Kolbanovskiy, A.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E. *Nucleic Acids Res.* **2007**, *35*, 1555.

(47) Stavros, K. M.; Hawkins, E. K.; Rizzo, C. J.; Stone, M. P. *Nucleic Acids Res.* **2014**, *42*, 3450.

(48) Cai, Y.; Ding, S.; Geacintov, N. E.; Broyde, S. *Chem. Res. Toxicol.* **2011**, *24*, 522.

(49) Rubbi, C. P.; Milner, J. *Carcinogenesis* **2001**, *22*, 1789.

(50) Clement, F. C.; Camenisch, U.; Fei, J.; Kaczmarek, N.; Mathieu, N.; Naegeli, H. *Mutat. Res-Fund. Mol. M.* **2010**, *685*, 21.

(51) Hess, M. T.; Schwitter, U.; Petretta, M.; Giese, B.; Naegeli, H. *Proc. Natl. Acad. Sci.* **1997**, *94*, 6664.

(52) Maillard, O.; Camenisch, U.; Blagoev, K. B.; Naegeli, H. *Mutat. Res-Rev. Mutat.* **2008**, *658*, 271.

(53) Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S. *Nucleic Acids Res.* **2011**, *39*, 8752.

(54) Yang, W. *DNA Repair* **2006**, *5*, 654.

(55) Janićijević, A.; Sugasawa, K.; Shimizu, Y.; Hanaoka, F.; Wijgers, N.; Djurica, M.; Hoeijmakers, J. H. J.; Wyman, C. *DNA Repair* **2003**, *2*, 325.

(56) Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D.; Cho, B. P. *Nucleic Acids Res.* **2013**, *41*, 869.

(57) Dutta, S.; Li, Y.; Johnson, D.; Dzantiev, L.; Richardson, C. C.; Romano, L. J.; Ellenberger, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16186.

(58) Wang, L.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785.

(59) Prakash, S.; Johnson, R. E.; Prakash, L. *Annu. Rev. Biochem.* **2005**, *74*, 317.

(60) Kozack, R.; Seo, K.-Y.; Jelinsky, S. A.; Loechler, E. L. *Mutat. Res-Fund. Mol. M.* **2000**, *450*, 41.

(61) Milhe, C.; Fuchs, R. P. P.; Lefevre, J. F. *Eur. J. Biochem.* **1996**, *235*, 120.

(62) Vaidyanathan, V. G.; Liang, F.; Beard, W. A.; Shock, D. D.; Wilson, S. H.; Cho, B. P. *J. Biol. Chem.* **2013**, *288*, 23573.

(63) Tang, Y.; Liu, Z.; Ding, S.; Lin, C. H.; Cai, Y.; Rodriguez, F. A.; Sayer, J. M.; Jerina, D. M.; Amin, S.; Broyde, S.; Geacintov, N. E. *Biochemistry* **2012**, *51*, 9751.

(64) Wu, M.; Yan, S.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *Chem. Res. Toxicol.* **2001**, *14*, 1629.

(65) Cho, B. P.; Zhou, L. *Biochemistry* **1999**, *38*, 7572.

(66) Milhé, C.; Dhalluin, C.; Fuchs, R. P. P.; Lefèvre, F.F. *Nucleic Acids Res.* **1994**, *22*, 4646.

(67) Kohda, K.; Tsunomoto, H.; Kasamatsu, T.; Sawamura, F.; Terashima, I.; Shibutani, S. *Chem. Res. Toxicol.* **1997**, *10*, 1351.

(68) Heavner, S.; Gannett, P. M. *J. Biomol. Struct. Dyn.* **2005**, *23*, 203.

(69) Millen, A. L.; Churchill, C. D. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 12995.

(70) Millen, A. L.; Kamenz, B. L.; Leavens, F. M. V.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2011**, *115*, 12993.

(71) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373.

(72) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A.* **2008**, *112*, 3742.

(73) Omumi, A.; Millen, A. L.; Wetmore, S. D.; Manderville, R. A. *Chem. Res. Toxicol.* **2011**, *24*, 1694.

(74) Kropachev, K.; Kolbanovskiy, M.; Liu, Z.; Cai, Y.; Zhang, L.; Schwaid, A. G.; Kolbanovskiy, A.; Ding, S.; Amin, S.; Broyde, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2013**, *26*, 783.

(75) Li, Z.; Tamura, P. J.; Wilkinson, A. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **2001**, *40*, 6743.

(76) Mao, B.; Gu, Z.; Gorin, A.; Chen, J.; Hingerty, B. E.; Amin, S.; Broyde, S.; Geacintov, N. E.; Patel, D. J. *Biochemistry* **1999**, *38*, 10831.

(77) Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J. *Biochemistry* **1989**, *28*, 7462.

(78) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2012**, *40*, 9675.

(79) Sharma, P.; Majdi Yazdi, M.; Merriman, A.; Manderville, R. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2015**, *28*, 782.

(80) Kropachev, K.; Kolbanovskii, M.; Cai, Y.; Rodríguez, F.; Kolbanovskii, A.; Liu, Y.; Zhang, L.; Amin, S.; Patel, D.; Broyde, S.; Geacintov, N. E. *J. Mol. Biol.* **2009**, *386*, 1193.

(81) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 81.

(82) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95.

(83) Zegar, I. S.; Chary, P.; Jabil, R. J.; Tamura, P. J.; Johansen, T. N.; Lloyd, R. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1998**, *37*, 16516.

(84) Elmquist, C. E.; Wang, F.; Stover, J. S.; Stone, M. P.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 445.

(85) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498.

(86) Li, Z.; Kim, H.-Y.; Tamura, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1999**, *38*, 16045.

(87) Wang, F.; DeMuro, N. E.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *J. Am. Chem. Soc.* **2006**, *128*, 10085.

(88) Zegar, I. S.; Kim, S. J.; Johansen, T. N.; Horton, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1996**, *35*, 6212.

(89) Sproviero, M.; Verwey, A. M. R.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 13405.

(90) Rodríguez, F. A.; Liu, Z.; Lin, C. H.; Ding, S.; Cai, Y.; Kolbanovskiy, A.; Kolbanovskiy, M.; Amin, S.; Broyde, S.; Geacintov, N. E. *Biochemistry* **2014**, *53*, 1827.

(91) Arlt, V. M.; Stiborova, M.; Schmeiser, H. H. *Mutagenesis* **2002**, *17*, 265.

(92) Rucker, G.; Chung, B. S. *Planta Med.* **1975**, *27*, 68.

(93) Kupchan, S. M.; Doskotch, R. W. *J. Med. Chem.* **1962**, *5*, 657.

(94) Vanherweghem, J. L.; Tielemans, C.; Abramowicz, D.; Depierreux, M.; Vanhaelen-Fastre, R.; Vanhaelen, M.; Dratwa, M.; Richard, C.; Vandervelde, D.; Verbeelen, D.; Jadoul, M. *The Lancet* **1993**, *341*, 387.

(95) Ivić, M. *Lijecnicki vjesnik* **1969**, *91*, 1273.

(96) Djukanović, L.; Radovanović, Z. In *Clinical Nephrotoxins*; de Broe, M., Porter, G., Bennett, W., Verpooten, G., Eds.; Springer Netherlands: **2003**, p 587.

(97) De Broe, M. E. *Kidney Int.* **2012**, *81*, 513.

(98) Arlt, V. M.; Stiborová, M.; vom Brocke, J.; Simões, M. L.; Lord, G. M.; Nortier, J. L.; Hollstein, M.; Phillips, D. H.; Schmeiser, H. H. *Carcinogenesis* **2007**, *28*, 2253.

(99) Chen, C. H.; Dickman, K. G.; Moriya, M.; Zavadil, J.; Sidorenko, V. S.; Edwards, K. L.; Gnatenko, D. V.; Wu, L.; Turesky, R. J.; Wu, X. R.; Pu, Y. S.; Grollman, A. P. *Proc. Natl. Acad. Sci.* **2012**, *109*, 8241.

(100) Cosyns, J. P.; Jadoul, M.; Squifflet, J. P.; Wese, F. X.; van Ypersele de Strihou, C. *Am J. Kidney Dis.* **1999**, *33*, 1011.

(101) *IARC monographs-100 A Plants containing aristolochic acids* **2002**, 347.

(102) Poon, S. L.; Pang, S. T.; McPherson, J. R.; Yu, W.; Huang, K. K.; Guan, P.; Weng, W. H.; Siew, E. Y.; Liu, Y.; Heng, H. L.; Chong, S. C.; Gan, A.; Tay, S. T.; Lim, W. K.;

Cutcutache, I.; Huang, D.; Ler, L. D.; Nairismägi, M. L.; Lee, M. H.; Chang, Y. H.; Yu, K. J.; Chan-on, W.; Li, B. K.; Yuan, Y. F.; Qian, C. N.; Ng, K. F.; Wu, C. F.; Hsu, C. L.; Bunte, R. M.; Stratton, M. R.; Futreal, P. A.; Sung, W. K.; Chuang, C. K.; Ong, C. K.; Rozen, S. G.; Tan, P.; Teh, B. T. *Sci. Transl. Med.* **2013**, *5*, 197ra101.

(103) Joshi, B. D.; Srivastava, A.; Gupta, V.; Tandon, P.; Jain, S. *Spectrochim. Acta Mol. Biomol. Spectros.* **2013**, *116*, 258.

(104) Sidorenko, V. S.; Attaluri, S.; Zaitseva, I.; Iden, C. R.; Dickman, K. G.; Johnson, F.; Grollman, A. P. *Carcinogenesis* **2014**, *35*, 1814.

(105) Stiborová, M.; Mareîs, J.; Frei, E.; Arlt, V. M.; Martínek, V.; Schmeiser, H. H. *Environ. Mol. Mutagen.* **2011**, *52*, 448.

(106) Martinek, V.; Kubickova, B.; Arlt, V. M.; Frei, E.; Schmeiser, H. H.; Hudecek, J.; Stiborova, M. *Neuro. Endocrinol. Lett.* **2010**, *32*, 57.

(107) Schmeiser, H. H.; Frei, E.; Wiessler, M.; Stiborova, M. *Carcinogenesis* **1997**, *18*, 1055.

(108) Grollman, A. P.; Shibutani, S.; Moriya, M.; Miller, F.; Wu, L.; Moll, U.; Suzuki, N.; Fernandes, A.; Rosenquist, T.; Medverec, Z.; Jakovina, K.; Brdar, B.; Slade, N.; Turesky, R. J.; Goodenough, A. K.; Rieger, R.; Vukelić, M.; Jelaković, B. *Proc. Natl. Acad. Sci.* **2007**, *104*, 12129.

(109) Hoang, M. L.; Chen, C. H.; Sidorenko, V. S.; He, J.; Dickman, K. G.; Yun, B. H.; Moriya, M.; Niknafs, N.; Douville, C.; Karchin, R.; Turesky, R. J.; Pu, Y. S.; Vogelstein, B.; Papadopoulos, N.; Grollman, A. P.; Kinzler, K. W.; Rosenquist, T. A. *Sci. Transl. Med.* **2013**, *5*, 197ra102.

(110) Moriya, M.; Slade, N.; Brdar, B.; Medverec, Z.; Tomic, K.; Jelaković, B.; Wu, L.; Truong, S.; Fernandes, A.; Grollman, A. P. *Int. J. Cancer* **2011**, *129*, 1532.

(111) Hoang, M. L.; Chen, C. H.; Sidorenko, V. S.; He, J.; Dickman, K. G.; Yun, B. H.; Moriya, M.; Niknafs, N.; Douville, C.; Karchin, R.; Turesky, R. J.; Pu, Y. S.; Vogelstein, B.; Papadopoulos, N.; Grollman, A. P.; Kinzler, K. W.; Rosenquist, T. A. *Sci. Transl. Med.* **2013**, *5*, 197ra102.

(112) Bieler, C. A.; Stiborova, M.; Wiessler, M.; Cosyns, J. P.; van Ypersele de Strihou, C.; Schmeiser, H. H. *Carcinogenesis* **1997**, *18*, 1063.

(113) Pfau, W.; Schmeiser, H. H.; Wiessler, M. *Carcinogenesis* **1990**, *11*, 1627.

(114) Shibutani, S.; Dong, H.; Suzuki, N.; Ueda, S.; Miller, F.; Grollman, A. P. *Drug Metab. Dispos.* **2007**, *35*, 1217.

(115) Attaluri, S.; Bonala, R. R.; Yang, I. Y.; Lukin, M. A.; Wen, Y.; Grollman, A. P.; Moriya, M.; Iden, C. R.; Johnson, F. *Nucleic Acids Res.* **2010**, *38*, 339.

(116) Schmeiser, H. H.; Nortier, J. L.; Singh, R.; Gamboa da Costa, G.; Sennesael, J.; Cassuto-Viguier, E.; Ambrosetti, D.; Rorive, S.; Pozdzik, A.; Phillips, D. H.; Stiborova, M.; Arlt, V. M. *Int. J. Cancer* **2014**, *135*, 502.

(117) Sidorenko, V. S.; Yeo, J. E.; Bonala, R. R.; Johnson, F.; Schärer, O. D.; Grollman, A. P. *Nucleic Acids Res.* **2012**, *40*, 2494.

(118) Broschard, T. H.; Wiessler, M.; von der Lieth, C. W.; Schmeiser, H. H. *Carcinogenesis* **1994**, *15*, 2331.

(119) de Laat, W. L.; Jaspers, N. G. J.; Hoeijmakers, J. H. J. *Genes Dev.* **1999**, *13*, 768.

(120) Gillet, L. C. J.; Schärer, O. D. *Chem. Rev.* **2005**, *106*, 253.

(121) Hanawalt, P. C.; Spivak, G. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 958.

(122) Lukin, M.; Zaliznyak, T.; Johnson, F.; de los Santos, C. *Nucleic Acids Res.* **2012**, *40*, 2759.

(123) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(124) Cosman, M.; Fiala, R.; Hingerty, B. E.; Laryea, A.; Lee, H.; Harvey, R. G.; Amin, S.; Geacintov, N. E.; Broyde, S.; Patel, D. *Biochemistry* **1993**, *32*, 12488.

(125) Patnaik, S.; Cho, B. P. *Chem. Res. Toxicol.* **2010**, *23*, 1650.

(126) Yan, S.; Wu, M.; Buterin, T.; Naegeli, H.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2003**, *42*, 2339.

(127) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2009**, *385*, 30.

(128) Jankowiak, R.; Ariese, F.; Hewer, A.; Luch, A.; Zamzow, D.; Hughes, N. C.; Phillips, D.; Seidel, A.; Platt, K. L.; Oesch, F. *Chem. Res. Toxicol.* **1998**, *11*, 674.

(129) Wilson, K. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2014**, *27*, 1310.

(130) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2013**, *26*, 803.

Chapter 2

Conformational Preferences of ALI and ALII-N^6-dA Adducted DNA[a,b,c]

2.1 Introduction

As stated in the Introduction, despite the known mutagenicity of the AL-N^6-dA lesions, only one recent study has analyzed the structural characteristics of damaged DNA containing the ALII-N^6-dA adduct using NMR spectroscopy.[1] Surprisingly, unlike the structurally similar N6-dA linked PAH lesions, which intercalate the bulky moiety into DNA without displacing the opposing base,[2-5] intercalation of the ALII moiety displaces the opposing thymine into the major groove.[1] This raises an important question regarding the structural and energetic features that drive the differential conformational outcomes of the ALII-N^6-dA lesion compared to the PAH adducts. Furthermore, despite the higher abundance[6-8] and greater nephrotoxic potential[9] of AAI compared to AAII in cells, recent studies have focused on the properties of the ALII-N^6-dA lesion.[1,10,11] However, previous studies on other DNA adducts (e. g., the C8-AF-dG[12-14], C8-AAF-dG,[12,13,15-17] and PAH[3-5,18] adducts) have shown that subtle changes in the chemical structure of the lesion may have profound implications on its conformation and mutagenicity within DNA (Section 1.5.3). Therefore, the effect of the methoxy group present in ALI-N^6-dA on the conformational outcomes of adducted DNA must be investigated.

---

[a]This chapter is reproduced with permission from: Kathuria, P.; Sharma, P.; Abendong, M.; Wetmore, S. D. *Biochemistry* **2015**, *54,* 2414.
[b]My contribution to the publication was running all the calculations, analysis of the data and writing the preliminary manuscript.
[c]I thank Minnette P. Abendong for preliminary calculations on the nucleobase and nucleoside models.

This chapter undertakes a systematic molecular modeling study, including QM (DFT) methods, MD simulations, and free energy analysis, to explore and compare the conformational space of the ALI-N^6-dA and ALII-N^6-dA lesions. Specifically, the nucleobase, nucleoside, and nucleotide adducts are considered to determine the preferred orientation of the bulky moiety with respect to the base and the *anti*/*syn* preference. Subsequently, the adducts are incorporated into an 11-mer oligonucleotide opposite complementary thymine, and MD simulations are carried out to reveal the effects of the helical environment on the *anti/syn* conformational preference. The computational approach is validated through careful comparison to NMR structural data previously published for the ALII-N^6-dA lesion.[1] As a result, this study reliably predicts the effect of the additional methoxy group present in the ALI-N^6-dA lesion on the conformational preferences of AL-adducted DNA. Furthermore, the preferred conformations of adducted DNA are analyzed in detail, which provides a structural basis for the resistance of the AL-N^6-dA lesions to be repaired by NER machinery. Overall, the present chapter yields valuable structural insights into the experimentally observed differences in the mutagenic potentials and persistence of the AL-N^6-dA adducts.

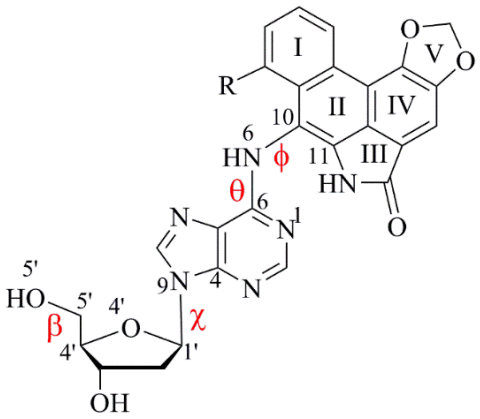
2.2 Computational Details

2.2.1 Nucleobase Model

To build the nucleobase model, the AL moiety formed from AAI or AAII was attached to the N6 position of adenine. Although both AL-N^6-A adducts can exist in either the amino or imino tautomeric forms at the N6 position,[19] a recent combined

experimental and modeling study concluded that the amino form predominates in DNA.[1] For this reason, only the amino form is considered in the present study. The relative orientation of the AL moiety with respect to the nucleobase is dictated by rotation about two dihedral angles, θ (∠(N1C6N6C10) and ϕ (∠(C6N6C10C11), Figure 2.1a). Therefore, the B3LYP/6-31G(d) potential energy surface (PES) was initially mapped as a function of θ and ϕ. Full optimizations were subsequently performed on all minima identified from the PES using B3LYP/6-31G(d) and the dispersion-corrected B3LYP-D3/6-31G(d). Finally, B3LYP and B3LYP-D3 were used in conjunction with the 6-311+G(2df,p) basis set to obtain more accurate relative energies. Although other approaches for including dispersion corrections in DFT functionals are available in the literature,[20-22] the inclusion of the empirical dispersion correction using the DFT-D3 approach has been shown to perform reasonably well for systems where noncovalent interactions are important.[23] Since a previous study on the S66 and S66x8 datasets of noncovalent interactions indicate that adding the dispersion correction strongly diminishes the performance differences between functionals,[24] hence the B3LYP-D3 functional was chosen. In the present study. Single-point energy calculations were performed using B3LYP-D3 and the larger 6-311+G(2df,p) basis set to obtain more accurate relative energies. The choice of the 6-31G(d) basis set for the geometry optimizations and 6-311+G(2df,p) basis set for single-point calculations is justified based on a previous studies on DNA adducts.[25-28]

(a)



(b)

5'-C^1 G^2 T^3 A^4 C^5 X^6 C^7 A^8 T^9 $G^{10}C^{11}$-3'

3'-$G^{22}C^{21}A^{20}T^{19}G^{18}T^{17}G^{16}T^{15}A^{14}C^{13}G^{12}$-5'

Figure 2.1: (a) Structure of the AL-N⁶-dA adducts (R = OCH_3 for ALI and H for ALII). The θ (∠(N1C6N6C10)) and ϕ (∠(C6N6C10C11)) dihedral angles determine the orientation of the AL moiety with respect to base, while χ (∠(O4′C1′N9C4)) dictates the glycosidic bond orientation to be* syn *(χ = 0±90°) or* anti *(χ = 180±90°), and β (∠(C4′N5′O5′H5′)) governs the sugar–phosphate backbone orientation. (b) The 11-mer DNA sequence used for MD simulations with the adduct at the X⁶ position.

2.2.2 Nucleoside Model

The nucleoside model was built by adding 2′-deoxyribose in the B-DNA relevant C2′-endo puckering to the lowest-energy conformation obtained from the nucleobase model. This build-up approach for the conformational scans has been validated in a previous study on phenolic DNA adducts,[27] where DFT results correlate with experimental spectroscopic data. Addition of the sugar moiety increases the complexity of the model since the dihedral angle χ (∠(O4′C1′N9C4), Figure 2.1a) must also be considered, which determines the orientation of the base about the glycosidic bond. The gas-phase B3LYP/6-31G(d) PES of the nucleoside adduct was scanned with respect to θ and χ. The ε = ∠(C4′C3′O3′H5′) torsion angle (Figure 2.1a) was initially

set to approximately 180˚, which is the average value obtained from previous MD simulations on the Dickerson-Drew dodecamer.[29] The minima obtained from the PES were fully optimized using B3LYP and B3LYP-D3 with the 6-31G(d) basis set, and the same functionals were used with 6-311+G(2df,p) to obtain more accurate relative energies. Furthermore, B3LYP-D3/6-31G(d) optimizations, followed by B3LYP-D3/6-311+G(2df,p) single-point calculations were performed on the lowest energy *anti* and *syn* conformations while geometrically constraining the β dihedral angle to 180˚, which permits analysis of DNA-relevant conformations of the adduct.

2.2.3 Nucleotide Model

To generate the initial nucleotide model, a 5′–monophosphate group was added to the lowest energy *anti* and *syn* conformers identified from the nucleoside model. A previous study indicates that the DNA-relevant geometries of the nucleotide model are obtained only when the solvent (water) environment and a Na^+ counter ion are used during the optimization routine.[26] Therefore, the nucleotide model was optimized in water (ϵ = 78.4) using IEF-PCM-B3LYP/6-31G(d) and the UFF radii to build the molecular cavity for solvation. Single-point calculations were performed with IEF-PCM-B3LYP-D3/6-311+G(2df,p).

Due to good agreement between the results obtained from B3LYP and B3LYP-D3, the B3LYP-D3 data is presented in the main chapter for the nucleobase and nucleoside models, while B3LYP results can be found in the Appendix A (Tables A1 – A3). Coordinates of all nucleobase minima, and the lowest energy *anti* and *syn* minima for the nucleoside, β-constrained nucleoside and nucleotide models are provided in the Appendix A (Tables A4 – A27). All reported DFT (B3LYP and B3LYP-D3) relative

energies include a scaled (0.9806) zero-point vibrational energy correction. All quantum mechanical calculations were performed using Gaussian 09 (Revision C.01 or D.01).[30,31]

2.2.4 DNA model

2.2.4.1 DNA sequence. The 11-mer oligonucleotide 5′–CGTACXCATGC (X = adduct; Figure 2.1b) was used in the MD simulations. This sequence was chosen based on previous experimental studies that suggest ALII-N^6-dA adducted DNA is most stable when cytosine flanks the lesion.[1] In addition, this sequence was used in a previous NMR study of ALII-N^6-dA adducted DNA,[1] which allows direct comparison to the simulations done in the present study. Moreover, other structurally similar N6-linked adenine PAH adducts have also been studied in a similar sequence context,[18,32,33] which allows us to determine the effect of the chemical composition of the bulky moiety on the conformational preferences. In addition to the (CXC) sequence context studied in this chapter, two other (T/CXG) mutational hotspots are known in the context of aristolochic acids.[34,35] However, the main aim of the present study is to carefully characterize all possible conformations accessible to ALI-N^6-dA and ALII-N^6-dA adducted DNA, and to compare relative flexibilities of the two damaged strands. To best achieve this goal, the present chapter focuses on only one sequence, which is the experimentally determined most stable sequence for ALII-N^6-dA adducted DNA.[1]

2.2.4.2 Starting Structures. The initial model for Arnott B-form DNA was built using the NAB*[46]* module of the AMBER package.[36,37] Adenine initially present at the X position in the natural DNA sequence was modified at N6 using GaussView[38] by

attaching the ALI or ALII moiety formed from AAI or AAII, respectively. Initial conformations of the adducts within DNA were built based on the lowest energy *anti/syn* structures obtained from the nucleotide model, while avoiding steric clashes with the surrounding nucleotides, and the damaged base was paired against complementary T. To adequately sample the conformational space of the adducted DNA, trial simulations were carried out with various starting structures, which differ in the location of the bulky moiety in the helical environment. Final production simulations were subsequently initiated from representative conformations obtained from these trial simulations.

2.2.4.3 Force Field. The adducted DNA was neutralized with 20 sodium ions and then solvated in an 8 Å octahedral box of TIP3P water,[39] which corresponds to near physiological conditions (Table A28, Appendix A). Although force fields are available that afford improved descriptions of monovalent ions,[40-42] simulations at low ion concentration or at near physiological conditions are robust regardless of the ion model implemented.[43] For this reason, parm99 parameters for sodium ions were used in the present study, which are the default choice in AMBER 11 or 12. Partial atomic charges for ALI-N^6-dA and ALII-N^6-dA were calculated with the RED.v.III.4[44] program, using the HF/6-31G* and the RESP-A1 scheme. ANTECHAMBER 1.4[45] was used to assign atom types (Tables A29 – A30, Appendix A). The parambsc0[46] modification of the parm99[47] force field was used to simulate the natural and adducted nucleotides, while missing parameters for the N6-moiety of the adducted nucleotides were taken from the General Amber Force Field (GAFF).[48] Although the χ_{OL} parameter set has recently been developed to improve the description of *syn* nucleotides in RNA (i.e., in

loops or unpaired regions),[49] the use of χ_{OL} worsens the description of B-DNA (i.e., distorts the helical parameters and sugar pucker compared to X-ray crystal structures).[49,50] Since the present conformational study focuses on damaged B-DNA helices, the χ_{OL} parameters were not used in order to avoid such distortions. Furthermore, the parm99 parameters provide an adequate description of the AL-N^6-dA adducts in DNA since: 1) the bulky moiety is positioned away from the glycosidic bond, which ensures that the χ profile of natural adenine does not change upon adduct formation (Figure A1, Appendix A); the stacking of the bulky moiety in DNA locks the χ-value of the adduct in MD simulations close to the value obtained from the DFT models (Table A31, Appendix A).

2.2.4.4 Simulation Details. Each DNA system was initially minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} Å^{-2} restraint on DNA. Subsequently, minimization was performed using 1000 steps of steepest descent followed by 1500 steps of conjugate gradient minimization with no restraints. The system was then heated to 300 K with the DNA restrained using a force constant of 10 kcal mol^{-1} Å^{-2} over a 20 ps constant volume MD simulation. Production simulations were subsequently run for 20 ns starting from DNA conformations obtained from trial simulations using the PMEMD[62] module of AMBER 12. All final production simulations gave rise to stable structures with a maximum standard deviation of 0.9 Å in the backbone RMSD (Figure A2, Appendix A).

To analyze the structural features at the lesion site for each distinct conformation of adducted DNA, the step parameters (Figure A3, Appendix A) were

calculated using a pseudostep consisting of the base pairs 5′ and 3′ with respect to the lesion. A pseudostep was used for analysis since the disrupted AL-N^6-dA:T pair cannot be used to calculate the step parameters at the lesion site due to intercalation of the AL moiety. The pseudostep parameters thus calculated markedly differ from the step parameters defined for natural DNA according to adjacent base pairs. For example, the twist angle calculated using a pseudostep is almost double (~ 60˚) (Table A32, Appendix A) the average twist in natural DNA (~ 30˚)[51] calculated using adjacent base pairs.

In order to confirm convergence of key structural parameters at the lesion site, the production simulations of the *syn*/*anti* base-displaced intercalated conformations, as well as the natural (control) helix were extended by 300 ns. The structures do not deviate significantly between the 20 ns and 320 ns simulations (Table A32, Appendix A), which provides confidence in the reported results. Most importantly, the distribution in the pseudotwist angle, which affects the stacking energies at the lesion site and therefore the relative stabilities of the adducted DNA conformations, agree very well between the 20 and 320 ns simulations (Figure A4, Appendix A).

A unimodal distribution was obtained with respect to the χ, θ, and ϕ dihedral angles of the adduct throughout each simulation (Figure A5 – A8, Appendix A). Therefore, a single representative structure was obtained from each simulation by clustering the entire trajectory with respect to the position of the atoms forming χ, θ, and ϕ using the PTRAJ module of AMBER 11.[18,36] Van der Waals interaction energies at the lesion site were calculated using the 'lie' command of CPPTRAJ module of

AMBER 12. This command is used to calculate the non-bonded interactions of one atom or group of atoms with other atom or groups of atoms. For each simulation frame the non-bonded interactions between all atoms of the adduct and all the atoms in the flanking bases were calculated. A default cutoff of 12 Å is assigned for both the electrostatic and van der Waals interactions. The model uses explicit solvation of the solute.

2.2.4.5 Free Energy Calculations. Free energies were calculated from each simulation trajectory using the molecular mechanics Poisson Boltzmann surface area (MM-PBSA)[52] method. Although free energy methods that are theoretically more rigorous than end-point free energy methods such as the MM-PBSA method have been used in the literature,[53,54] a previous study has shown that MM-PBSA free energies and experimental binding affinities show an excellent correlation.[55] Moreover, MM-PBSA has been the most commonly used method in previous studies of damaged DNA.[56-59] The total free energy (G_{total}) of each adducted DNA conformation is estimated from the molecular mechanics energy (E_{MM}), solvation energy (E_{sol}) and the entropy term (TS) as:

$$G_{total} = E_{MM} + G_{sol} - TS$$

where E_{MM} is the sum of the internal energy emerging from deviation of bonds (E_{bond}), angles (E_{angle}) and dihedral angles ($E_{dihedral}$) from their equilibrium values deduced according to the force field, the van der Waals (E_{vdW}) and the electrostatic (E_{elec}) energy without water, while G_{sol} is the solvation energy calculated as the sum of two components:

$$G_{sol} = G_{PB} + G_{SA}$$

G_{PB} is the term arising from the polar interactions between the solvent and solute. This component of the solvation energy is equal to the change in the electrostatic energy for the transfer of the solute from a vacuum to the solvent, which is effectively evaluated by the Poisson Boltzmann (PB) equation:

$$G_{PB} = \frac{1}{2} \Sigma\, q_i\, (V_{solvent} - V_{vacuum})$$

where q_i is the solute charge and $V_{solvent} - V_{vacuum}$ is the difference between the potential in the presence of the solvent and in vacuum. G_{SA} is the non-polar component of the solvation energy, which is proportional to the solvent accessible surface area (SASA), given as:

$$G_{SA} = \gamma\, (SASA)$$

γ is the microscopic surface free energy for formation of cavity in water. This equation assumes that the contribution from dispersion interactions between the solute and solvent are negligible in comparison to the energy required to displace water molecules to create a cavity.

The entropy component is calculated assuming the biological system obeys a rigid rotor model, where the rotational and translational entropy can be calculated using standard statistical mechanical formulae and the vibrational entropy is evaluated using normal mode calculations. T is the temperature (298.15 K). The relative total free energy (G_{rel}) is reported with respect to the adducted DNA conformation with the lowest (most negative) G_{total}.In these calculations, snapshots were taken at 50 ps intervals over the entire simulation (400 frames in total), and the entropy term was estimated using the normal mode analysis method. In these calculations, snapshots were taken at 50 ps intervals over the entire simulation (400

frames in total), and the entropy term was estimated using the normal mode analysis method.

2.3 Results

2.3.1 Flexibility of the Nucleobase Adduct

The ALI and ALII nucleobase adducts were investigated to determine the preferred orientation of the bulky moiety with respect to the base. The PES with respect to θ and ϕ for both ALI-N^6-A and ALII-N^6-A (Figures 2.2a and b) indicate that for a given ϕ, a planar orientation about the C6–N6 bond (θ ~ 0˚ or 180˚) is more stable than the orthogonal orientation (θ ~ 90˚ or 270˚). The two minima at θ ~ 0˚ (ϕ ~ 30˚ or 330˚) are separated by a low (< 10 kJ mol^{-1}) rotational barrier with respect to ϕ (Figures 2.2a and b). On the other hand, the two minima at θ ~ 180˚ (ϕ ~ 60˚ or 300˚) are separated by a higher rotational barrier (~ 20 kJ mol^{-1}). The interconversion between a minimum at θ ~ 0˚ and a minimum of θ ~ 180˚ is associated with a large energetic penalty (~ 40 – 60 kJ mol^{-1}), with ALII-N^6-A exhibiting greater flexibility about θ compared to ALI-N^6-A. Thus, the ϕ dihedral angle is inherently more flexible than θ.

Upon full optimization (Figures 2.2c and d), the global minimum for ALI-N^6-A (ALII-N^6-A) occurs at θ = 12.4˚ (5.2˚) and ϕ = 327.0˚ (322.4˚). In addition, a nearly isoenergetic minimum is obtained through rotation about ϕ by ~ 70˚. On the other hand, two isoenergetic minima (θ ~ 180˚ and ϕ ~ 50–60˚ or 295–305˚) are ~ 8 – 13 kJ mol^{-1} less stable than the global minimum. The larger energetic difference in the case of ALI-N^6-A is due to the loss of two hydrogen bonds in ALI-N^6-A versus a single

hydrogen bond in ALII-N^6-A upon rotation about θ. Nevertheless, the energy difference is less than the strength of the hydrogen bond(s) since the loss of the stabilizing interactions in the higher energy minima (θ ~ 180˚) is partially compensated by reduced steric repulsion at θ ~ 180˚ compared to the global minimum (θ ~ 0˚). Overall, the small energetic differences between minima indicate that the AL-N^6-A adducts possess significant conformational flexibility at the nucleobase level.

2.3.2 Flexibility of the Nucleoside Adduct

The ALI-N^6-dA and ALII-N^6-dA nucleoside adducts were built by adding a deoxyribose to the corresponding nucleobase global minimum to determine the conformational flexibility at the glycosidic bond and the *anti*/*syn* energetic preference. Seven minima can be located on the PES of the ALI-N^6-dA and ALII-N^6-dA adducts when considered as a function of the χ and θ dihedral angles (Figures 2.3a and c). For both adducts, the global minimum adopts the *syn* orientation, while the corresponding *anti* conformers are approximately 20 – 30 kJ mol^{-1} higher in energy. Nevertheless, interconversion between the lowest energy *anti* and *syn* minima (rotation with respect to χ) is energetically more favourable than between two *syn* minima (rotation with respect to θ). This is due to regions of high energy encountered at θ ~ 270˚ and 90˚. Thus, the AL-N^6-dA nucleoside adducts are more flexible with respect to χ than θ.

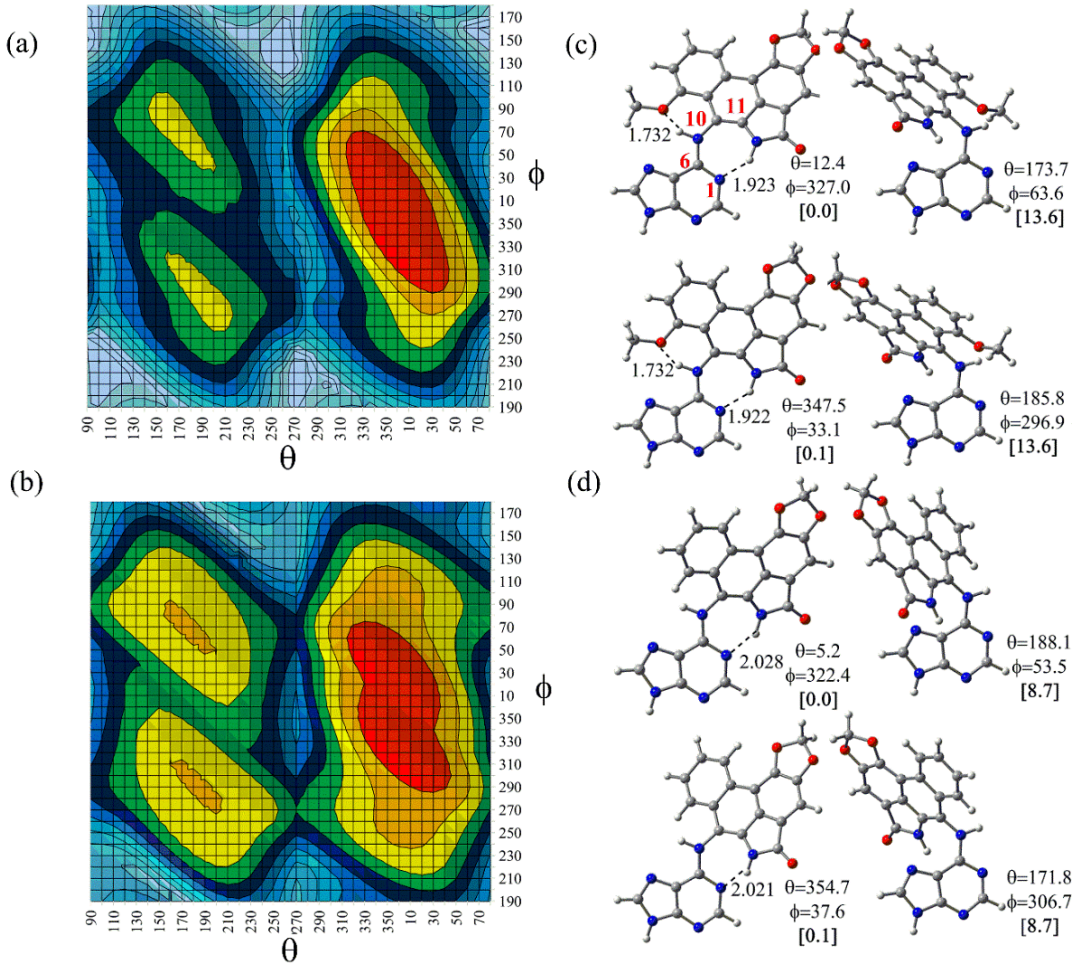


Figure 2.2: B3LYP/6-31G(d) potential energy surfaces (PES) for the (a) ALI-N⁶-A and (b) ALII-N⁶-A nucleobase adducts. B3LYP-D3/6-31G(d) optimized minima, select bond lengths (Å), dihedral angles (θ (∠(N1C6N6C10) and ϕ (∠(C6N6C10C11), deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol⁻¹) for the (c) ALI-N⁶-A and (d) ALII-N⁶-A nucleobase adducts are provided.

The nucleobase model indicates that two nearly isoenergetic ϕ values are possible for a given θ. Therefore, for each of the seven minima on the χ versus θ PES, a second conformer is possible that differs in ϕ, such that ϕ = 30 – 60˚ or 330 – 300˚. Thus, a total of 14 fully-optimized minima were characterized for each adduct (Figures A9 – A12, Appendix A), where the lowest energy *anti* and *syn* minima for

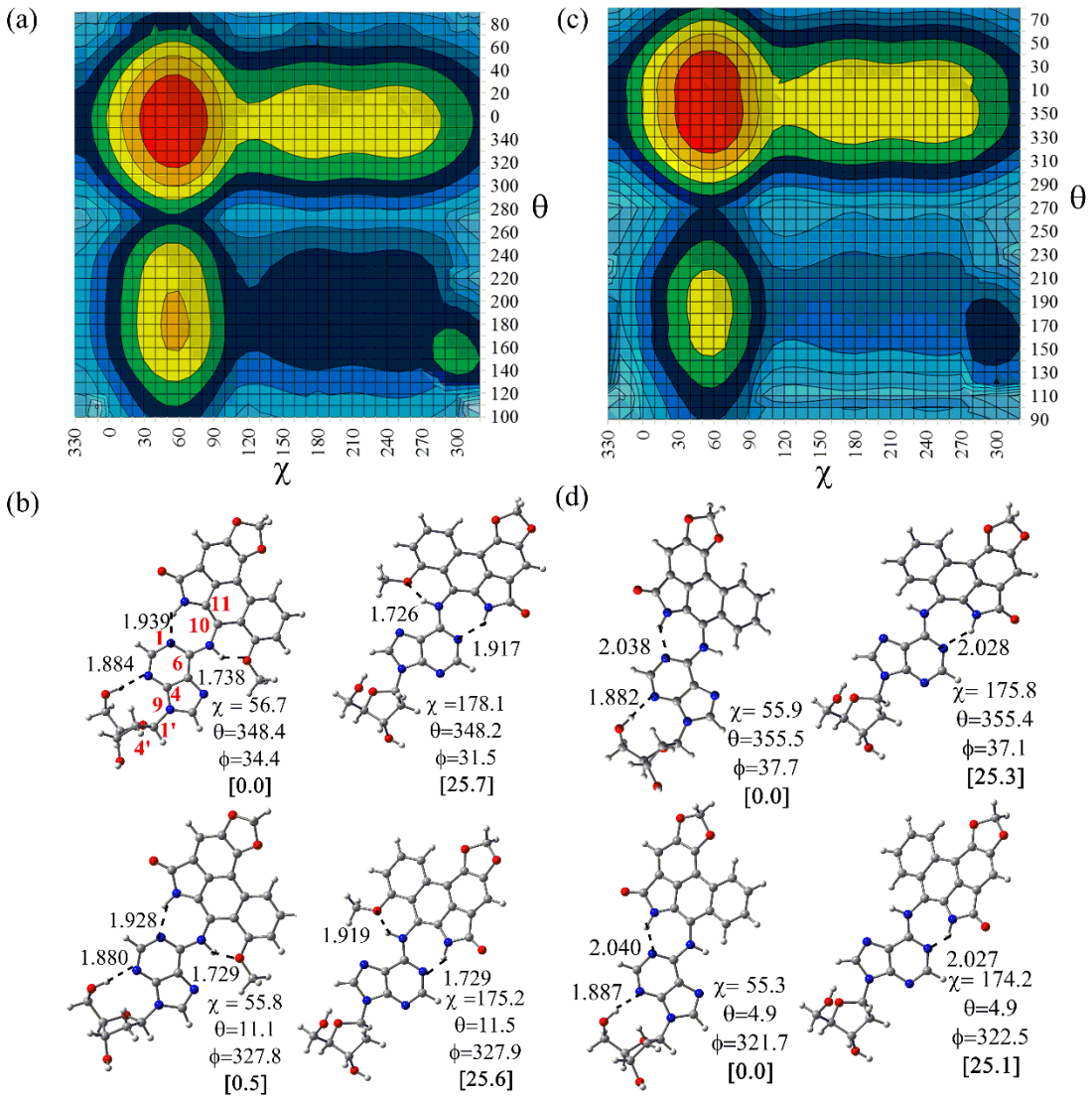


Figure 2.3: B3LYP/6-31G(d) potential energy surface (PES) (a) ALI-N⁶-dA and (c) ALII-N⁶-dA. B3LYP-D3/6-31G(d) fully-optimized lowest energy* anti *and* syn *minima with the two possible ϕ values for the (b) ALI-N⁶-dA and (d) ALII-N⁶-dA adduct. Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol⁻¹) are provided. (θ (∠(N1C6N6C10), ϕ (∠(C6N6C10C11) and χ (∠(O4′C1′N9C4)).

each possible ϕ value are shown in Figures 2.3b and d. Regardless of the ϕ value considered, the fully optimized global (*syn*) minimum for both adducts (χ ~ 55˚ and θ ~ 350˚) is ~ 8 – 14 kJ mol^{-1} lower in energy that the *syn* minimum related by rotation about θ (χ ~ 55˚ and θ ~ 180˚), primarily due to differences in intramolecular hydrogen bonding. On the other hand, the lowest energy *anti* minimum (χ ~ 180˚ and

θ ~ 350°) is ~ 25 kJ mol^{-1} less stable than the global *syn* minimum. This *anti* minimum is nearly isoenergetic to a second *anti* minimum that primarily differs in χ (~ 255°), which reflects the fact that changes in χ does not affect interactions between adenine and the AL moiety. The remaining three *anti* minima are (~ 40 kJ mol^{-1}) less stable than the global *syn* minimum due to the lack of hydrogen bonds between the sugar, nucleobase and/or the AL moiety. Since the structures with ϕ = 30 – 60° are lower (if not equal) in energy to those with ϕ = 330 – 300° for a given combination of χ and θ, subsequent models only consider initial structures with ϕ ranging between 30 and 60°.

Closer examination of the lowest energy *syn* and *anti* conformations reveals that the global *syn* conformer possesses an additional hydrogen bond between 5′–OH of the sugar and N3 of the nucleobase that is absent in the *anti* conformation (Figures 2.3b and d). Although this interaction may affect the *anti*/*syn* relative stability, it is only relevant when the adduct is located at the 5′–terminal end of DNA since 5′–OH is replaced by a 5′–phosphate group at all other helical positions. To analyze the conformation of the adduct at other helical positions, the lowest energy *anti* and *syn* conformations (with ϕ = 30 – 60°) were reoptimized while geometrically constraining the β dihedral angle (Figure 2.1a) to 180°. The most significant structural changes on optimization occur in *syn* due to repulsions between O5′ of the sugar with the nucleobase moiety. More importantly, removal of the O5′–H•••N3 interaction increases the energy of the *syn* conformer such that the *anti* orientation is ~ 11 kJ mol^{-1} more stable (Figure 2.4). Nevertheless, this small *anti*/*syn* energy difference

indicates that the adducted nucleosides may possess flexibility at the glycosidic bond within DNA.

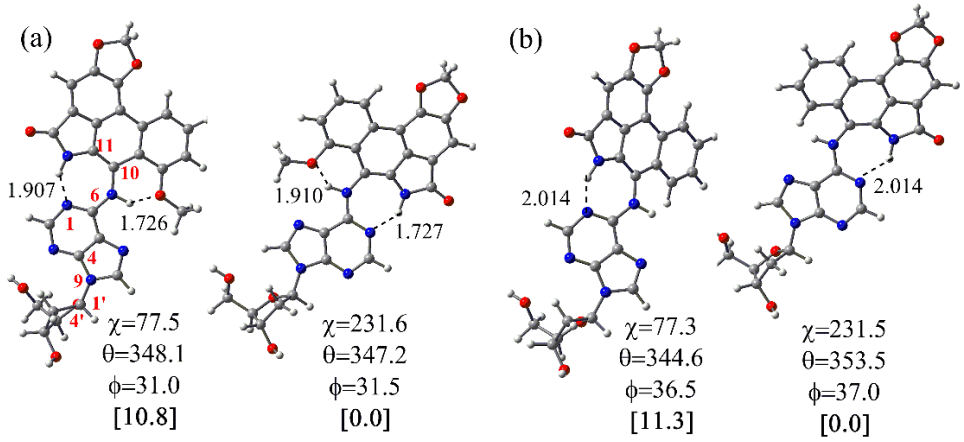


Figure 2.4: B3LYP-D3/6-31G(d) lowest energy* anti *and* syn *minima for the (a) ALI-N^6-dA and (b) ALII-N^6-dA β-constrained nucleoside adducts. Select bond lengths (Å), dihedral angles (θ (∠(N1C6N6C10), ϕ (∠(C6N6C10C11) and χ (∠(O4′C1′N9C4), deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol^{-1}) are provided.

2.3.3 Flexibility of the Nucleotide Adduct

Since the small difference in the *anti*/*syn* stability for the nucleoside model may be altered by the 5′–phosphate group in DNA, a model phosphate group was added to the lowest energy *anti* and *syn* β-constrained nucleoside conformations (Figure 2.5). In the previously studied nucleotide models of the C8-OTB-dG[25] and C8-phenoxyl-dG[26] adducts, the phosphate moiety interacts with either the bulky group or the nucleobase, which alters the *anti*/*syn* energy difference. In contrast, calculations indicate that such base–phosphate or phosphate–bulky moiety interactions are absent in the AL-N^6-dA adducts. The incorporation of a phosphate group decreases the *anti*/*syn* energy difference such that the *anti* conformation is marginally more stable (by ~ 6 kJ mol^{-1}). Nevertheless, there is no significant effect

on key structural features for the AL-N^6-dA nucleotide models compared to the DNA-relevant β-constrained nucleoside model. This small *anti*/*syn* energy difference may be further affected by the presence of flanking nucleotides, and therefore both the *anti* and *syn* conformations must be considered in DNA.

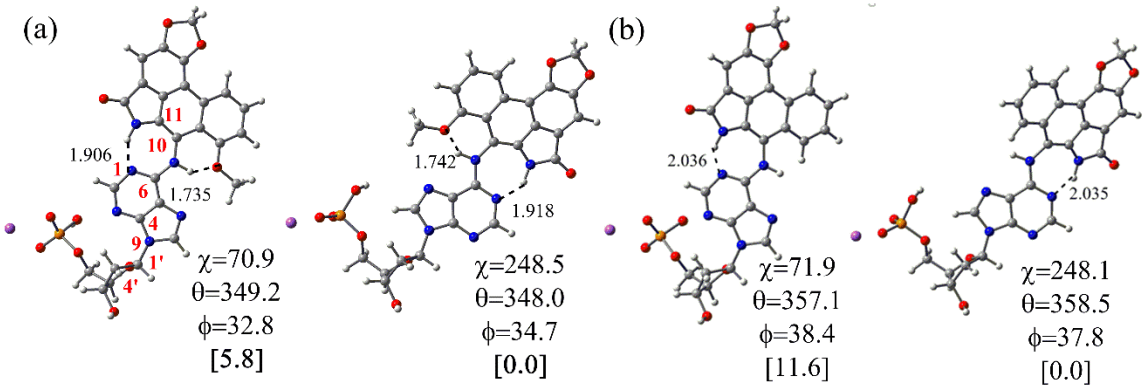


Figure 2.5: PCM-B3LYP-D3/6-31G(d) lowest energy* anti *and* syn *minima for the (a) ALI-N^6-dA and (b) ALII-N^6-dA nucleotide adducts. Select bond lengths (Å), dihedral angles (θ (∠(N1C6N6C10), ϕ (∠(C6N6C10C11) and χ (∠(O4′C1′N9C4), deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol^{-1}) are provided.

2.3.4 Flexibility of Adducted DNA

2.3.4.1 Structural Features of Possible Conformations of Adducted DNA

MD simulations of the ALI and ALII-N^6-dA adducts (X) in the 5′–CGTACXCATGC 11-mer paired against T (Figure 2.1b) were initiated from various input structures to sample the conformational space. As a result, three distinct conformations for each of the *anti* and *syn* adducted nucleotide orientations were characterized for both adducts (Figures 2.6 and 2.7; throughout the thesis, the damaged base pair is shown in green, the AL moiety is shown in red and 5′ and 3′–flanking bases pairs are shown

in blue). The first structure is the base-displaced intercalated conformer in which the damaged base is slightly displaced towards the minor groove (Figure A13, Appendix A), the bulky AL moiety stacks between the 5′ and 3′–flanking base pairs (Figures 2.6a,d and 2.7a,d) and the opposing T becomes extrahelical. In the second conformation, the AL moiety stacks between the opposing T and the 5′–base with respect to the adduct in the opposite strand (denoted as 5′–intercalated, Figures 2.6b,e and 2.7b,e). Finally, in the third conformation, the AL moiety stacks between the opposing T and the 3′–base with respect to the adduct in the opposite strand (denoted as 3′–intercalated, Figures 2.6c,f and 2.7c,f). In the above conformers, the damaged adenine maintains stacking interaction with the 5′ and 3′–cytosines. Despite the symmetry of the sequence considered, the 3′ and 5′–intercalated conformations are unique due to the helical twist. The structural features of each conformation are discussed below in relation to the natural helix, where references to the (3′ and 5′) nucleotide positions of flanking bases are made with respect to the adducted nucleotide. All structural features are reported as a change with respect to the unmodified DNA sequence.

The anti *base-displaced intercalated conformer.* In the base-displaced intercalated conformer containing the *anti* orientation of the lesion (Figure 2.6a and d), the bulky AL moiety intercalates into the helix. With the exception of ring I (Figure 2.1a), which is solvent exposed in the major groove, all rings of the polycyclic AL moiety are well stacked within the helix between the 5′ and 3′–guanines in the opposing strand. Although the opposing T maintains the *anti* orientation ($\chi \sim 230^\circ - 260^\circ$), it is displaced out of the helix (Figure 2.6a and d). However, the flanking bases are

unperturbed, and maintain strong Watson-Crick (WC) hydrogen bonding (~ 100% occupancy; Table A33, Appendix A). As a result, the pseudostep parameters at the lesion site change very little (shift, slide, and rise change by 1.9 Å, and the tilt, roll, and

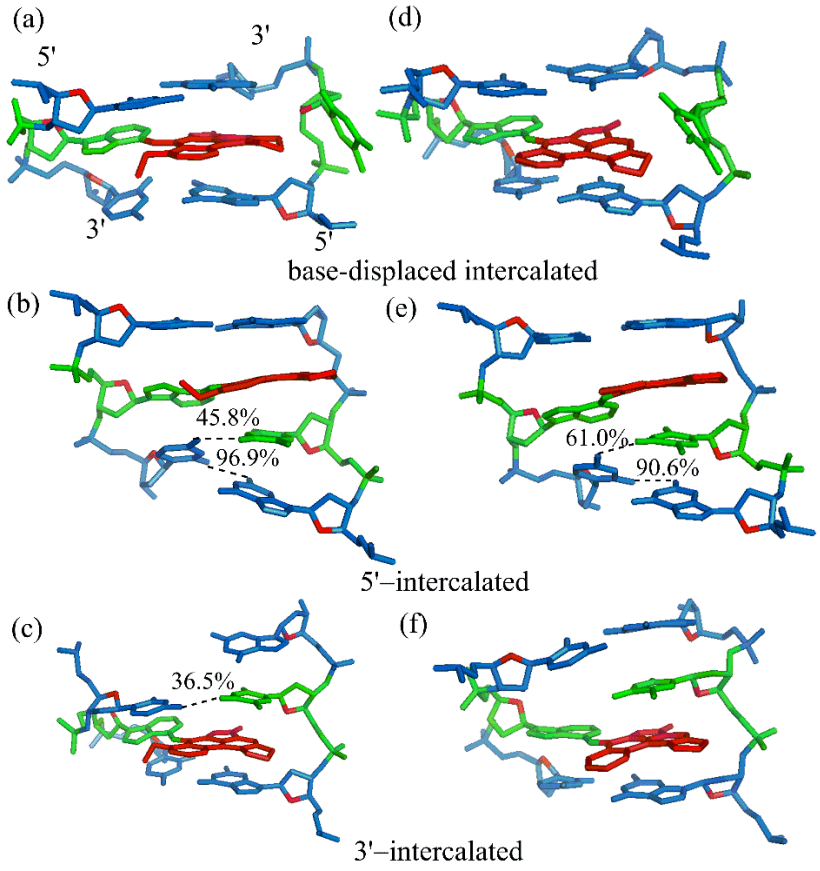


Figure 2.6: Representative structures from MD simulations with the* anti *orientation of the ALI-N⁶-dA (a, b and c) or ALII-N⁶-dA (d, e and f) adduct paired opposite thymine in three adducted DNA conformations. Select hydrogen-bond occupancies are provided (hydrogen-bond distance cut off within 3.4 Å heavy atom separation and 120° X–H–X angle).

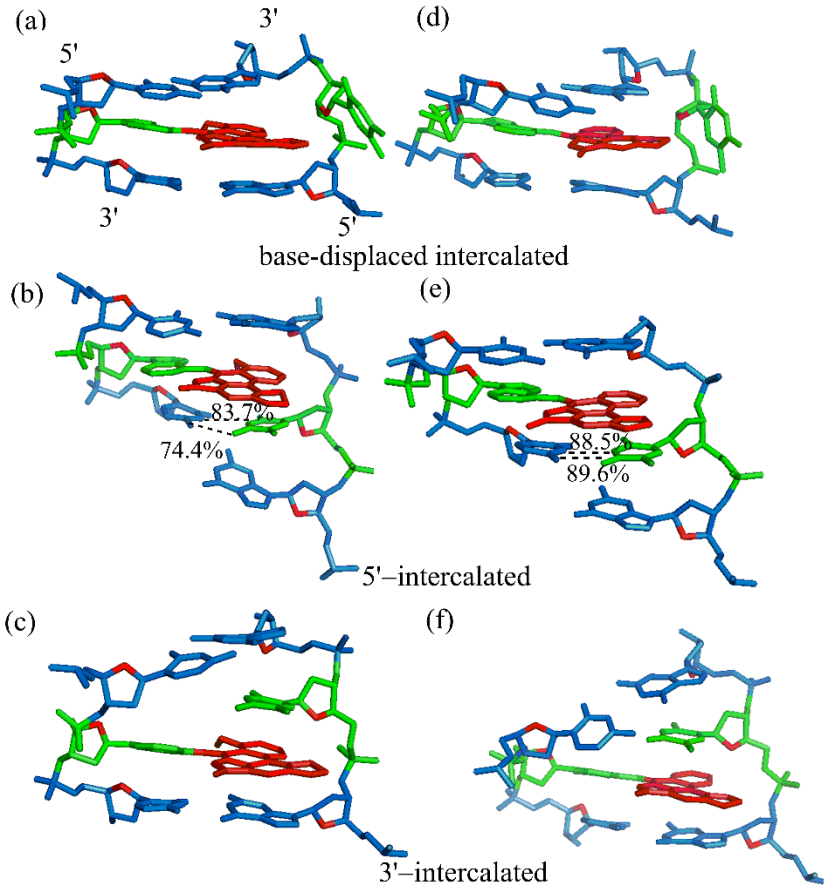


Figure 2.7: Representaive structures from MD simulations with the ALI-N⁶-dA (a, b and c) or ALII-N⁶-dA (d, e and f) adduct in the* syn *conformation paired opposite thymine in three adducted DNA conformations. Select hydrogen-bond occupancies are provided (hydrogen-bond distance cut off within 3.4 Å heavy atom separation and 120˚ X–H–X angle).

twist change by a maximum of 13˚ relative, to the natural DNA, Figure 2.8a). However, there is a slight (0.4 – 0.5 Å) widening of the minor groove compared to the natural helix (Figure 2.8a). Nevertheless, the DNA distortions are localized to the lesion site, with minimum perturbation to the remainder of the helix.

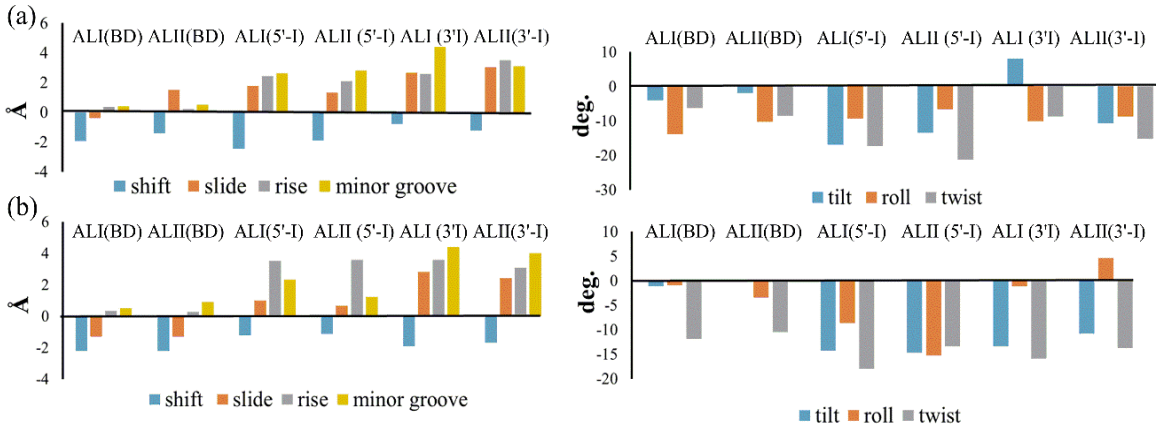


Figure 2.8: The change in the pseudostep parameters and minor groove dimensions relative to unmodified DNA for the (a) anti AL-N^6-dA or (b) syn AL-N^6-dA adducted DNA helices in the base-displaced intercalated (BD), 5′–intercalated (5′–I) and 3′–intercalated (3′–I) conformers.

The anti *5′– intercalated conformer.* In the 5′–intercalated conformer (Figure 2.6b and e), all rings of the polycyclic AL moiety (except ring I) stack within the helix between the complementary T and 5′–G in the opposing strand. Although the 5′–base pair is unperturbed (~ 100% occupancy; Table A33, Appendix A), C in the 3′–base pair rotates to simultaneously hydrogen bond with its complementary G and T opposing the adduct. Specifically, one (O2(C)•••H–N2(G)) hydrogen bond remains in the original C:G pair for most of the simulation time (91 – 97% occupancy; Table A33, Appendix A), and a new sporadic interaction (N4–H(C)•••O4(T)) forms between C and T (46% occupancy for ALI-N^6-dA and 61% occupancy for ALII-N^6-dA; Table A33, Appendix A). The increase in minor groove dimension in the 5′–intercalated conformer is greater than observed in the base-displaced intercalated conformer (*vide supra*), which corresponds to greater untwisting of the DNA strand (Figure 2.8a). In addition, the shift, slide, and rise parameters increases by ~ 2.5 Å and the

angular step parameters (tilt, roll, and twist) change by ~ 21˚ (Figure 2.8a). Thus, DNA is overall more distorted in this conformation than the base-displaced intercalated structure.

The anti *3′– intercalated conformer.* In the 3′–intercalated conformer, the AL moiety stacks (with ring I solvent exposed) between the T opposing the adduct and 3′–G in the opposing strand (Figure 2.6c and f). For ALI-N^6-dA, although hydrogen bonding is maintained in the 3′–flanking base pair (~ 100% occupancy; Table A33, Appendix A), the WC pairing of the 5′–flanking pair is disrupted. Specifically, the T opposing the adduct adopts a *syn* orientation (χ ~ 303˚, Figure 2.6c) and forms a (N4–H(C)•••O4(T)) hydrogen bond with the 5′–C (36.5% occupancy; Table A33, Appendix A). Whereas the disruption of the 5′–flanking base pair to accommodate the ALI-N^6-dA lesion widens the minor groove (~ 4.4 Å), the ALII-N^6-dA lesion perturbs the step parameters (e.g. the rise, Figure 2.8a) at the lesion site without disrupting the WC hydrogen bonding in the flanking pairs (~ 100% occupancy; Table A33, Appendix A). Thus, ALI-N^6-dA and ALII-N^6-dA adducted DNA possess slightly different structural features in the *anti* 3′–intercalated conformer.

The syn *base-displaced intercalated conformer*. Similar to the corresponding *anti* conformation, the AL moiety intercalates into the helix in the *syn* base-displaced intercalated conformation (Figure 2.7a and d), while simultaneously displacing the opposing T (χ ~ 210˚ – 270˚) into an extrahelical position. In contrast to the corresponding *anti* conformer, where ring I of the AL moiety is solvent exposed, ring IV partially stacks with the flanking bases in the *syn* base-displaced conformation. However, the additional methoxy group present in ALI-N^6-dA is solvent exposed in

both the *anti* and *syn* base-displaced intercalated conformations. Regardless, the flanking bases remain unperturbed and maintain WC hydrogen bonding (~ 100% occupancy; Table A33, Appendix A). As a result, the angular step parameters change by a maximum of 12° (Figure 2.8b), and the change in the minor groove width (0.5 – 0.9 Å) is almost similar to the observed widening in the corresponding *anti* conformation. The rise (~ 0.5 Å), shift (~ 2.2 Å) and slide (~ 0.8 Å) parameters exhibit a slightly greater perturbation than in the *anti* base-displaced intercalated conformation. Overall, the lesions are well accommodated in DNA in the base-displaced intercalated conformation, with little perturbation to the helix.

The syn *5′–intercalated conformer.* Similar to the corresponding *anti* conformer, the AL moiety stacks between the opposing thymine and 5′–G in the opposite strand in the *syn* 5′–intercalated conformer. This intercalation moves the opposing T in the 3′–direction, which allows it to hydrogen bond with C in the 3′–base pair. In contrast to the corresponding *anti* conformer, which partially maintains a hydrogen bond in the 3′–neighbouring base pair, the 3′–base pair is completed disrupted in the *syn* 5′–intercalated conformer (Figure 2.7b and e). Instead, two hydrogen bonds are formed between 3′–C and T opposing the adduct, one of which (N4–H(C)•••O4(T)) largely remains intact (84 – 90% occupancy for both adducts, Figure 2.7b and e) and one (N3–H(T)•••N3(C)) that is more persistent for ALII-N^6-dA (89% occupancy) than ALI-N^6-dA (75% occupancy). The complete disruption of the 3′–base pair in the *syn* 5′–intercalated conformer leads to a larger change in the rise parameter (~ 3.5 Å), and a narrower minor groove (~ 1.3–2.3 Å) relative to the natural helix than the corresponding *anti* conformer (Figure 2.8b). In addition, as seen in the *syn* base-

displaced intercalated conformer, ring IV of the AL moiety (Figure 2.1a) does not stack within the helix, but rather is essentially solvent exposed.

The syn *3′–intercalated conformer.* Similar to the corresponding *anti* conformer, the AL moiety stacks between the opposing thymine and the 3′–G in the opposing strand in the *syn* 3′–intercalated conformation (Figure 2.7c and f). As discussed for the *syn* 5′–intercalated structure, ring IV remains solvent exposed. However, in contrast to the *syn* 5′–intercalated structure, WC bonding of the 5′ and 3′–flanking base pairs remain intact for both lesions (~ 100% occupancy; Table A33, Appendix A), as observed for the *anti* 3′–intercalated conformer of ALII-N^6-dA. In order to simultaneously accommodate the damaged adenine and AL moieties between the 5′ and 3′–flanking pairs, the minor groove widens by ~ 4.0 Å, and the rise parameter increases by 3.1 – 3.5 Å, compared to natural DNA (Figure 2.8b). In fact, this minor groove widening is the largest among all three conformations with either (*anti* or *syn*) adduct orientation. In contrast, the shift and slide parameters change by a maximum of 3.5 Å, and the angular step parameters change by a maximum of 15˚ compared to natural DNA.dx

2.3.4.2 Relative Stabilities of Adducted DNA Conformers

ALI-N^6-dA Adducted DNA. The *anti* base-displaced intercalated conformer is the most energetically preferred for ALI-N^6-dA adducted DNA (Table 2.1) due to unperturbed WC hydrogen bonding in the flanking base pairs, as well as C–H•••π interactions between ring V (Figure 2.1a) of the AL moiety and the extrahelical T (Figure 2.7a). The next most stable conformer is the *syn* base-displaced intercalated conformation,

which is 2.1 kJ mol^{-1} less stable than the corresponding *anti* conformer. This energy difference is partly due to weaker stacking interactions of the damaged base and AL moiety with the 5′ and 3′–flanking base pairs (E_{vdw}, Table 2.1). The *anti* 5′–intercalated structure is the third most stable conformer, lying 36.1 kJ mol^{-1} above the most stable conformer (Table 2.1) due to disruption of WC hydrogen bonding in the 5′–flanking base pair, which is partially compensated by enhanced stacking interactions (Table 2.1) and additional hydrogen bonding between T opposing the adduct and the 3′–C (~ 46% occupancy). The *syn* 5′–intercalated conformer is ~ 6 kJ mol^{-1} less stable than the corresponding *anti* 5′–intercalated structure due to weaker stacking interactions, despite a higher occupancy of the 3′–C:T hydrogen bond (Figure 2.7b). The *syn* 3′–intercalated conformer is 49.1 kJ mol^{-1} less stable than the lowest energy conformer, but 2.3 kJ mol^{-1} more stable than the corresponding *anti* conformer. Despite weaker stacking interactions (Table 2.1), the *syn* 3′–intercalated conformer is more stable due to intact hydrogen bonding in both flanking base pairs, while the 5′–flanking base pair is disrupted in the *anti* conformation.

ALII-N^6-dA Adducted DNA. The *anti* base-displaced intercalated structure is energetically preferred for ALII-N^6-dA adducted DNA (Table 2.1) due to significant stacking stabilization and persistent WC hydrogen bonding in both flanking base pairs (Table A33, Appendix A). In the absence of the methoxy group, the relative stability of the remaining conformations aligns with the stabilization provided by stacking interactions at the lesion site. Specifically, the corresponding *syn* conformer

Table 2.1: Relative MM-PBSA free energies (G_{rel}) and the van der Waals energies (E_{vdw}) for different conformations of AL-N^6-dA adducted DNA derived from 20 ns MD simulations (kJ mol^{-1}).

Adduct	Orientation	Conformation	G_{rel}[a]	E_{vdw}[b]
unmodified DNA	-	-	-	–70.1±5.3
ALI-N^6-dA	*anti*	base-displaced	0.0	–171.8±6.8
		5′–intercalated	36.1	–148.1±10.9
		3′–intercalated	51.6	–148.9±12.4
	syn	base-displaced	2.1	–166.3±7.1
		5′–intercalated	42.0	–144.8±10.5
		3′–intercalated	49.1	–135.6±7.6
ALII-N^6-dA	*anti*	base-displaced	0.0	–161.6±6.7
		5′–intercalated	33.9	–144.7±11.3
		3′–intercalated	48.9	–132.1±6.8
	syn	base-displaced	8.0	–159.4±6.3
		5′–intercalated	42.7	–136.0±10.5
		3′–intercalated	38.5	–127.5±8.8

[a]Relative total free energy calculated with respect to the most stable conformation for each adduct. [b]Total van der Waals interaction energy between the adducted (or natural) base present at the X^6 position in the 11-mer DNA (Figure 2.1b) and the flanking base pairs.

is ~ 8 kJ mol^{-1} less stable due to slightly weaker stacking at the lesion site, while the next most stable (*anti* and *syn* 5′–intercalated) conformations are much less stable (~ 34 – 42 kJ mol^{-1} above the most stable conformer) due to significantly weaker stacking interactions. Similarly, despite intact hydrogen bonding in the 5′ and 3′–flanking base pairs, the *syn* and *anti* 3′–intercalated conformers are ~ 39 kJ mol^{-1} and ~ 49 kJ mol^{-1} less stable than the lowest energy structure, respectively, due to weaker stacking interactions.

2.4 Discussion

2.4.1 Conformational Heterogeneity of the AL-N^6-dA Adducts

The AL-N^6-A nucleobase models indicate that the AL and adenine moieties prefer an almost planar relative orientation in the absence of the sugar, which maximizes the number of stabilizing intramolecular hydrogen bonds. The addition of the sugar ring does not affect the orientation of the base with respect to the AL moiety, which is attached to a position far from the glycosidic bond. Although the *anti* nucleoside conformer is ~ 25 kJ mol^{-1} less stable at the 5′–terminus of DNA, the β-constrained nucleoside and nucleotide models predict the *anti* conformer to be slightly more stable than the *syn* orientation in other helical positions (by ~ 5 – 10 kJ mol^{-1}). This small difference in stability indicates that both the *anti* and *syn* conformations must be considered in DNA helices.

Three distinct conformations were isolated for both the *anti* and *syn* lesion orientations in ALI and ALII-N^6-dA adducted DNA, namely the base-displaced intercalated, 5′–intercalated and 3′–intercalated conformers. In the base-displaced intercalated conformer, intercalation of the bulky moiety pushes the opposing T out of the helix, which disrupts the Watson-Crick hydrogen bonding between the lesion and the opposing T. This conformation parallels those presented in the literature for guanine N-linked aromatic amines adducts,[60,61] with the exception that the damaged base is displaced to a lesser extent, and therefore maintains stacking interactions with the 5′ and 3′–flanking bases, in the AL-N^6-dA adducted DNA. In contrast, the 5′ and 3′–intercalated conformers of AL-N^6-dA adducted DNA parallel those previously observed for PAH adenine adducts,[3,4,18] where the bulky moiety intercalates into the

helix between the opposing base and the base pair on the 5′ or 3′ side of the adducted pair. Since the opposing base is not displaced in the 5′ and 3′–intercalated conformers, the base-step parameters are altered at the lesion site in order to accommodate the bulky moiety, with the degree of variation depending on both the adducted nucleotide (*anti/syn*) orientation and the (5′/3′) intercalation site.

Despite the range of conformations isolated from MD simulations on adducted DNA, the *anti* base-displaced intercalated conformer is the most stable and hence the most likely orientation to be adopted. Nevertheless, the *syn* base-displaced intercalated conformer is close in energy to the *anti* conformer, especially in the case of ALI-N^6-dA, indicating that a subtle change in the cellular environment may dictate the conformation of AL-adducted DNA. Interestingly, the small energetic difference in the *anti/syn* orientation of the adducts predicted by the small (nucleoside and nucleotide) models extrapolates to adducted DNA. Furthermore, due to greater energetic differences, other (5′ and 3′–intercalated) conformations will likely not be adopted regardless of the cellular environment. Interestingly, the higher energy *anti*/*syn* 5′–intercalated and *syn* 3′–intercalated conformations have similar structural characteristics for both AL-N^6-dA adducts; however, the *anti* 3′–intercalated conformations of the two adducts have similar relative energies despite some structural differences at the lesion site due to the presence of select mutually compensating discrete (hydrogen–bonding and stacking) interactions involving the adduct.

2.4.2 Comparison to the Previous NMR Structure of ALII-N^6-dA

A recent NMR based study of ALII-N^6-dA adducted DNA predicted an *anti* base-displaced intercalated conformation.[1] Within the same sequence, unrestrained MD simulations carried out in the present study isolated six stable conformers of ALII-N^6-dA adducted DNA. However, post-processing free energy calculations reveal the same most stable (*anti* base-displaced intercalated) conformation as predicted from NMR.[1] Although complete assignments of the NOE cross peaks were not reported in the previous NMR study,[1] lesion site structural parameters of representative MD structure match the reported average NMR structure (Figure 2.9a), which dictates NOE values. Indeed, comparison of the structural parameters at the lesion site (Table A34, Appendix A) indicates that the χ dihedral of the modified base and opposing thymine, as well as 3′–C and G, match very well between the two structures. Additionally, the sugar puckering at the T base opposing the lesion, as well as 3′–G and 5′–C, is similar to that observed from restrained MD. Furthermore, the sugar puckering of 3′–C observed using NMR is similar to the second most common sugar puckering observed in MD simulations (Table A34, Appendix A). Nevertheless, some differences in the sugar puckering of the adducted nucleotide and 5′–G are observed, and the χ dihedral of the 5′–C and 5′–G bases deviate by $\sim 50^\circ$ from the NMR value (Table A34, Appendix A). In addition to the most stable conformation, present study provides information about other possible conformations for the carcinogenic AL-N^6-dA adducts. Interestingly, calculations carried in the present study indicate that the *syn* base-displaced conformation, which was not derived from the NMR data of ALII-N^6-dA adducted DNA, may also be energetically accessible with slight perturbations

to the surrounding environment (e.g., within the active site of lesion-bypass polymerases).[62]

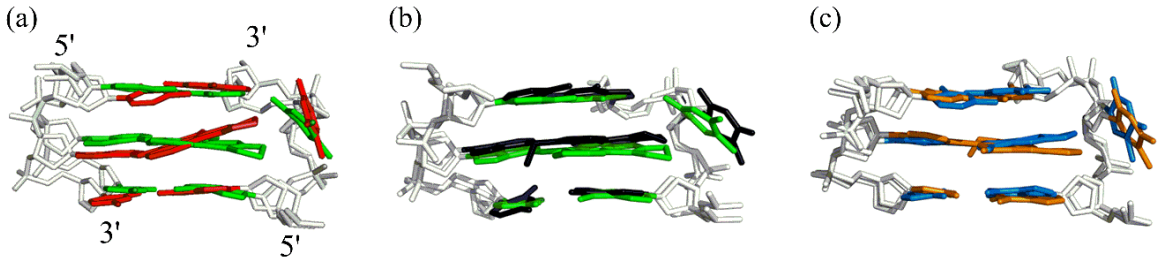


Figure 2.9: Overlay of trimers of the base pairs containing the adducted pair comparing (a) the NMR[1] (red) and the representative (unrestrained) MD structure of the most stable ALII-N^6-dA adducted DNA conformation (green) (RMSD=1.3 Å), the representative (unrestrained MD) structure of the most stable (b) anti conformation of ALII-N^6-dA (green) and ALI-N^6-dA (black) adducted DNA (RMSD=0.8 Å), and (c) syn conformation of ALII-N^6-dA (blue) and ALI-N^6-dA (orange) adducted DNA (RMSD=1.0 Å).

2.4.3 Comparison of the ALI-N^6-dA and ALII-N^6-dA Adducts

Although NMR data is available for ALII-N^6-dA, no such structural study has been carried out to date for ALI-N^6-dA. Nevertheless, the close agreement between most stable conformation of ALII-adducted DNA determined from the present study and the one derived from NMR[1] suggests that reliable information can be obtained from MD data. Free energy calculations indicate that the most stable conformation is the *anti* base-displaced intercalated structure for both ALI-N^6-dA and ALII-N^6-dA adducted DNA. Furthermore, an overlay of the representative structures for both adducts indicates these conformers are very similar (Figure 2.9b). Both adducts can also lead to the *syn* base-displaced intercalated conformation of adducted DNA. However, the *syn* base-displaced intercalated conformer is associated with a greater energetic penalty in the case of the ALII (8.0 kJ mol^{-1}) than ALI-N^6-dA (2.1 kJ mol^{-1})

adduct. Interestingly, when the van der Waals energy is calculated over the simulation trajectories of ALI-N^6-dA adducted DNA with the methoxy group removed, the methoxy group is determined to increase the van der Waals energies equally (by ~ 10 kJ mol^{-1}) in the *anti* and *syn* based-displaced intercalated conformations (Table A35, Appendix A). Thus, the smaller energy difference between the *anti* and *syn* base-displaced intercalated conformations for ALI-N^6-dA compared to ALII-N^6-dA does not directly arise due to van der Waals interactions involving the methoxy group. Instead, the methoxy group exerts a spatial effect only in the *anti* base-displaced intercalated conformation, which is most significantly reflected in the displacement of the AL moiety relative to ALII-N^6-dA (overlays in Figure A14, Appendix A). The synergistic effect of several geometric changes is ultimately one factor that alters the stability of the *anti* base-displaced intercalated conformation relative to the *syn* counterpart. In addition, the *syn* base-displaced intercalated conformations are more hydrated, and therefore the location of the bulky moiety is more stabilized, than the *anti* orientation for both ALI-N^6-dA and ALII-N^6-dA (Figure 2.10). However, the difference in hydration is greater for ALI-N^6-dA, which further reduces the *anti/syn* energy difference.

The greater energetic difference between the *anti* and *syn* base-displaced intercalated conformers for ALII-N^6-dA may explain why a detectable fraction of the *syn* base-displaced intercalated conformer was not observed in the previous NMR study.[1] In contrast, the relative energies suggest that the fraction of the *syn* base-displaced intercalated conformation may be greater for ALI-N^6-dA adducted DNA.

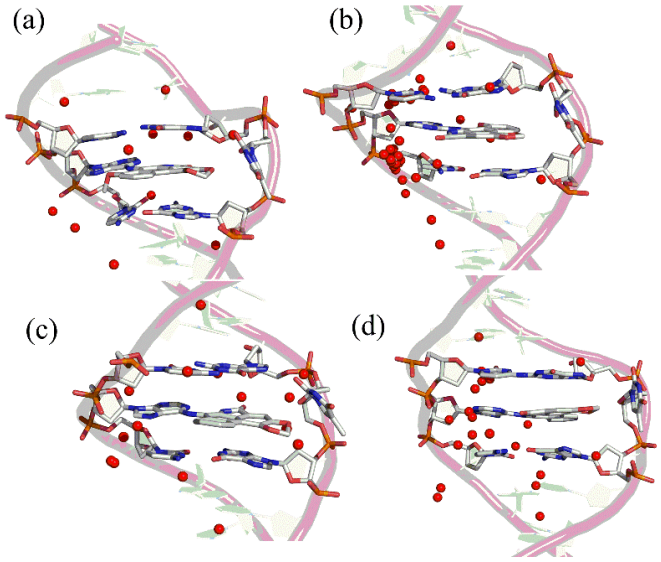


Figure 2.10: The hydration pattern surrounding the* anti *(a,c) and* syn *(b,d) base- displaced intercalated conformer of ALI-N^6-dA (a,b) and ALII-N^6-dA (c,d) adducted DNA. The water molecules present within 5 Å of the lesion for more than 80% of the simulation time are shown (red spheres).

Indeed, the calculations undertaken in the present study indicate that this conformation may be detectable in NMR experiments, and therefore will likely play an important role in determining the mutagenic potential of the ALI-N^6-dA adduct. Although the mutagenic activity of the lesion is dependent on the adduct conformation at the single strand–double strand junction[63] and the active site of polymerases,[64,65] the intrinsic adduct conformation of adducted dsDNA in solution has been shown to persist in the polymerase active site for select PAH and aromatic amine adducts.[66] Furthermore, a greater flexibility in the active site of a lesion-bypass polymerase has been associated with increased misincorporation upon replication.[67] Therefore, if the smaller energetic difference between the *syn* and *anti* base-displaced intercalated conformations for ALI-N^6-dA than ALII-N^6-dA adducted DNA persists in the polymerase, then the ALI-N^6-dA adduct will likely have a greater tendency to

induce base mismatches. This proposal correlates with the experimentally observed greater mutagenic potential of ALI-N^6-dA over ALII-N^6-dA.[9,68]

DNA adducts that are structurally similar, but differ in the chemical composition of the bulky group via stereochemistry[18] or the presence of additional rings,[28,69] display different conformational preferences. However, the ALI-N^6-dA and ALII-N^6-dA adducts provide only the second example of lesions that potentially exhibit unique conformational outcomes and chemically differ by a substituent. Specifically, the composition of AAF (acetylaminoflourene) and AF (aminoflourene) deviate by an acetyl substituent on the N-linkage, and lead to three conformations for AAF-dG adducted DNA (namely, major groove, base-displaced stacked and minor-groove wedge) versus two conformations of AF-dG adducted DNA (namely, major groove and base-displaced stacked).[70] This difference occurs since steric repulsion involving the acetyl group attached causes the damaged guanine to become nonplanar, which diminishes the stacking energy at the lesion site compared to (a nearly co-planar) AF-dG.[70] In the case of the ALI and ALII-N^6-dA adducts, the distinguishing methoxy substituent is located on the bulky ring rather than the linkage, and therefore does not sterically interact with the damaged base. Nevertheless, these adducts exhibit diverse conformational flexibility because the additional methoxy group in ALI-N^6-dA changes the relative geometries and hydration of the *syn* and *anti* base-displaced intercalated conformers at the lesion site. This example further underscores that subtle changes in the chemical composition of bulky adducts can affect the conformational equilibrium of the associated dsDNA.

2.4.4 Comparison to Other N6-linked dA Adducts

A number of other carcinogens form adducts at the N6 site of dA, which can broadly be divided into two main categories: (i) adducts lacking a conjugated π-ring (e.g., DHB-dA[2] formed from 1,3-butadiene, and endogenous estrogen adducts, including 2-OHE$_1$-6α-N^6-dA and 2-OHE$_1$-6β-N^6-dA[71]); and (ii) adducts containing a conjugated polycyclic π–system (e.g., PAH adducts).[3-5,18,72-75] Adducts that lack conjugated π–rings prefer non-intercalated conformations with the bulky moiety in either the minor or major groove.[2,71] On the other hand, N6-dA adducts with polycyclic π–systems preferentially form intercalated conformations that are stabilized through π–π interactions between the bulky moiety and the flanking bases in the DNA helix.[3-5,18,60-63]

Similar to the PAH adducts, the AL lesions contain a conjugated π–ring and the associated adducted DNA preferentially adopts an intercalated conformation. However, the AL-N^6-dA adducts induce a base-displaced intercalated conformation, while most PAH adducts lead to either the 5′ or 3′–intercalated conformations. Furthermore, although the preferred intercalated structures associated with PAH adducts either partially or completely maintain Watson-Crick hydrogen bonding in the damaged base pair,[3,4,18,32] the hydrogen bonding in the lesion pair is entirely disrupted in all conformations of AL-N^6-dA adducted DNA. This variance arises due to differences in the conjugation of the carbon atom in the bulky moiety attached to N6 of adenine. Specifically, the AL adducts adopt a planar orientation of the bulky moiety with respect to the base due to extended conjugation of N6 in A with an sp^2 hybridized carbon in the AL moiety, while the carbon attached to N6 in A is sp^3

hybridised in the PAH adducts.[76] The absence of extended conjugation and intramolecular hydrogen bonding imparts flexibility to the PAH lesions that allows the bulky moiety to bend and maintain hydrogen bonding with the opposing base upon intercalation.

Interestingly, the preferred DNA conformations of AL-N^6-dA are less structurally distorting than conformations of DNA containing PAH adducts. In the preferred base-displaced intercalated conformations of AL-N^6-dA adducted DNA, the lesion causes minimal changes to the pseudostep parameters and the minor groove width. In contrast, in the 5′ or 3′–intercalated conformations, the PAH adducts significantly change the pseudostep parameters and the minor groove width with respect to the natural DNA.[18,77] Although energetically inaccessible (5′–intercalated and 3′–intercalated) conformations of ALI-N^6-dA and ALII-N^6-dA adducted DNA exhibit greater helical distortions, these conformers are still less distorted than the PAH induced conformations. Thus, the planar conformations of the AL-N^6-dA adducts are better accommodated in DNA than the non-planar PAH adducts.

2.4.5 NER Recognition of AA Adducts

Previous studies based on ^{32}P post labeling assays have revealed that ALI-N^6-dA is more persistent in cells than ALII-N^6-dA.[7,68] However, it is currently unclear whether this difference arises because of differences in the repair propensity of the two lesions. Bulky DNA lesions are generally repaired by the nucleotide excision repair (NER) pathway, which removes a short segment (~ 24–32 nucleotides) of the lesion-containing DNA strand.[78] One of the most versatile mechanisms for NER of bulky lesions is the global genomic repair (GGR) pathway, which involves recognition

factors (e.g., XPC-RAD23B in eukaryotes) that constantly scan the genome. Studies on other bulky adducts (mainly the PAH and aromatic amine purine adducts) have shown that the damage is typically recognized in the GGR pathway by thermodynamic destabilization induced by the lesion.[72,79-81] Changes in the thermodynamic stability have been directly connected to alterations in structural features such as distortion to the helix at the lesion site,[80,82-84] increase in helical[85] and backbone[84] dynamics, enhanced stacking interactions involving the damaged base,[86,87] and the identity of the partner base.[88] On the other hand, the transcription coupled repair (TCR) pathway of NER is more efficient in transcriptionally-active regions of the genome, where lesion-induced stalling of RNA polymerases serves as the lesion recognition signal.[89]

Previous studies have shown that the ALII-N^6-dA adduct is resistant to GGR and is exclusively repaired by the TCR pathway.[10] The calculations carried out in the present study may provide the missing structural explanations for the resistance of the ALII-N^6-dA adducts to GGR and the greater relative persistence of ALI-N^6-dA. First, close examination of the preferred (*anti* base-displaced intercalated) conformations of both AL-N^6-dA adducts indicates that the displacement of the opposing thymine into the major groove results in the loss of stacking interactions with the flanking base pairs compared to the natural helix. However, this destabilization is effectively compensated by additional stacking stabilization between the AL moiety and the flanking base pairs at the lesion site (Table 2.1). In fact, stacking interactions are enhanced in all conformations of AL-adducted DNA compared to the natural helix (Table 2.1). Second, distortions in the pseudostep

parameters relative to the natural helix for both adducts indicate that these planar DNA lesions do not substantially distort DNA, which may prevent the GGR recognition step. Third, the dynamics of both modified DNA strands are similar to the natural strand (Figure A15, Appendix A). Therefore, the enhanced stacking, lack of helical distortion, and minimal changes in helical dynamics relative to the natural helix together might explain the resistance of these lesions to GGR.

In addition to the above factors, a previous crystal structure of the yeast orthologue of the initial GGR recognition factor (XPC-RAD23B) by Min *et al*[79] reveals that the initial steps in lesion recognition involve insertion of a β-hairpin via the major groove side of DNA, and flipping of the lesion and opposing base(s) out of the helix into the DNA minor groove. Since the AL-N^6-dA adducts intercalate via the major groove in the most stable *anti* base-displaced intercalated structure, these conformers may obstruct β-hairpin insertion into DNA. In addition, displacement of the opposing T into the major groove in the most stable conformation could block interactions between the enzyme and the lesion site in DNA. Thus, as indicated for PAH-N^6-dA lesions,[72] unfavorable steric interactions between the lesion intercalating through the DNA major groove and the β-hairpin of the XPC-RAD23B recognition factor may also prevent the AL-N^6-dA lesions from being effectively recognized by GGR.

Most importantly, the factors that potentially affect the repair propensities (e.g., stacking energies, distortions at the lesion site, the helical dynamics and the helical position of the AL moiety) are very similar for DNA containing ALII-N^6-dA and ALI-N^6-dA. Therefore, it is likely that these lesions are equally resistant to the GGR

pathway. This implies that the greater persistence of ALI-N^6-dA in cells is not because of differential repair propensities, but rather must be due to other factors, such as the experimentally observed greater activation and binding of AAI than AAII to DNA.[8]

2.5 Conclusions

The present conformational study of the ALI and ALII-N^6-dA adducts revealed that their intrinsic flexibility about the glycosidic bond holds true both at the nucleoside/nucleotide units and within DNA, which suggests that small computational models can provide important insights into the conformational preference of DNA adducts. Unrestrained MD simulations predict the *anti* base-displaced intercalated conformation of adducted DNA to be the most stable, which agrees with previously published NMR data for ALII-N^6-dA adducted DNA and thereby validates the computational methodology. Using the same approach, structural information for the more prevalent ALI-N^6-dA adduct was obtained. Most importantly, free energy calculations reveal that the ALI-N^6-dA adduct possesses greater *anti*/*syn* conformational flexibility. If this increased flexibility persists in the active site of lesion-bypass polymerases, this may explain the experimentally observed greater nephrotoxic potential of the ALI-N^6-dA over ALII-N^6-dA adduct. Furthermore, this chapter provides structural clues for the NER resistance of both AL adenine adducts, including enhanced stacking, lack of helical distortions and minimal changes in helical dynamics relative to natural DNA, as well as possible blockage of β-hairpin insertion of the GGR recognition factor. Hence, this computational approach can be extended to reliably study the adducts for which experimental data is lacking, such as the AL-guanine adducts (studied in Chapter 3), as well as to examine the effect

of sequence context on the conformational preferences of AL-N^6-dA adducted DNA (studied in Chapter 4).

2.6 References

(1) Lukin, M.; Zaliznyak, T.; Johnson, F.; de los Santos, C. *Nucleic Acids Res.* **2012**, *40*, 2759.

(2) Kowal, E. A.; Seneviratne, U.; Wickramaratne, S.; Doherty, K. E.; Cao, X.; Tretyakova, N.; Stone, M. P. *Chem. Res. Toxicol.* **2014**, *27*, 805.

(3) Li, Z.; Kim, H. Y.; Tamura, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1999**, *38*, 16045.

(4) Li, Z.; Mao, H.; Kim, H. Y.; Tamura, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1999**, *38*, 2969.

(5) Li, Z.; Tamura, P. J.; Wilkinson, A. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **2001**, *40*, 6743.

(6) Fernando, R.; Schmeiser, H.; Scherf, H.; Wiessler, M. *IARC Sci. Publ.* **1992**, 167.

(7) Pfau, W.; Schmeiser, H. H.; Wiessler, M. *Carcinogenesis* **1990**, *11*, 1627.

(8) Martinek, V.; Kubickova, B.; Arlt, V. M.; Frei, E.; Schmeiser, H. H.; Hudecek, J.; Stiborova, M. *Neuroendocrinol. Lett.* **2010**, *32*, 57.

(9) Shibutani, S.; Dong, H.; Suzuki, N.; Ueda, S.; Miller, F.; Grollman, A. P. *Drug Metab. Dispos.* **2007**, *35*, 1217.

(10) Sidorenko, V. S.; Yeo, J. E.; Bonala, R. R.; Johnson, F.; Schärer, O. D.; Grollman, A. P. *Nucleic Acids Res.* **2012**, *40*, 2494.

(11) Attaluri, S.; Bonala, R. R.; Yang, I. Y.; Lukin, M. A.; Wen, Y.; Grollman, A. P.; Moriya, M.; Iden, C. R.; Johnson, F. *Nucleic Acids Res.* **2010**, *38*, 339.

(12) Milhé, C.; Dhalluin, C.; Fuchs, R. P. P.; Lefèvre, F. F. *Nucleic Acids Res.* **1994**, *22*, 4646.

(13) Tan, X.; Suzuki, N.; Grollman, A. P.; Shibutani, S. *Biochemistry* **2002**, *41*, 14255.

(14) Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J. *Biochemistry* **1989**, *28*, 7462.

(15) Milhe, C.; Fuchs, R. P. P.; Lefevre, J. F. *Eur. J. Biochem.* **1996**, *235*, 120.

(16) Patel, D. J.; Mao, B.; Gu, Z.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 391.

(17) Schorr, S.; Carell, T. *ChemBioChem* **2010**, *11*, 2534.

(18) Cai, Y.; Ding, S.; Geacintov, N. E.; Broyde, S. *Chem. Res. Toxicol.* **2011**, *24*, 522.

(19) Pfau, W.; Schmeiser, H. H.; Wiessler, M. *Chem. Res. Toxicol.* **1991**, *4*, 581.

(20) Zhao, Y.; Truhlar, D. *Theor. Chem. Account* **2008**, *120*, 215.

(21) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

(22) Torres, E.; DiLabio, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 3342
.

(23) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456.

(24) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421.

(25) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2013**, *26*, 803.

(26) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373.

(27) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A.* **2008**, *112*, 3742.

(28) Millen, A. L.; Sharma, P.; Wetmore, S. D. *Future Med. Chem.* **2012**, *4*, 1981.

(29) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham III, T. E.; Otyepka, M.; Jurečka, P. *J. Chem. Theory Comput.* **2013**, *9*, 2339.

(30) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision C.01 ed.; Gaussian, Inc.: Wallingford CT, **2010**.

(31) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.;

Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavchari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision D.01 ed.; Gaussian, Inc.: Wallingford CT, **2013**.

(32) Mao, B.; Gu, Z.; Gorin, A.; Chen, J.; Hingerty, B. E.; Amin, S.; Broyde, S.; Geacintov, N. E.; Patel, D. J. *Biochemistry* **1999**, *38*, 10831.

(33) Kropachev, K.; Kolbanovskiy, M.; Liu, Z.; Cai, Y.; Zhang, L.; Schwaid, A. G.; Kolbanovskiy, A.; Ding, S.; Amin, S.; Broyde, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2013**, *26*, 783.

(34) Hoang, M. L.; Chen, C. H.; Sidorenko, V. S.; He, J.; Dickman, K. G.; Yun, B. H.; Moriya, M.; Niknafs, N.; Douville, C.; Karchin, R.; Turesky, R. J.; Pu, Y. S.; Vogelstein, B.; Papadopoulos, N.; Grollman, A. P.; Kinzler, K. W.; Rosenquist, T. A. *Sci. Transl. Med.* **2013**, *5*, 197ra102.

(35) Poon, S. L.; Pang, S. T.; McPherson, J. R.; Yu, W.; Huang, K. K.; Guan, P.; Weng, W. H.; Siew, E. Y.; Liu, Y.; Heng, H. L.; Chong, S. C.; Gan, A.; Tay, S. T.; Lim, W. K.; Cutcutache, I.; Huang, D.; Ler, L. D.; Nairismägi, M. L.; Lee, M. H.; Chang, Y. H.; Yu, K. J.; Chan on, W.; Li, B. K.; Yuan, Y. F.; Qian, C. N.; Ng, K. F.; Wu, C. F.; Hsu, C. L.; Bunte, R. M.; Stratton, M. R.; Futreal, P. A.; Sung, W. K.; Chuang, C.-K.; Ong, C. K.; Rozen, S. G.; Tan, P.; Teh, B. T. *Sci. Transl. Med.* **2013**, *5*, 197ra101.

(36) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; *AMBER 11* University of California, San Francisco, CA: **2010**.

(37) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; *AMBER 12* University of California, San Francisco, CA: **2012**.

(38) Roy, D. T., Keith; John, Millam *GaussView Version 5* **2009**.

(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(40) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.

(41) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100*, 3757.

(42) Joung, I. S.; Cheatham III, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020.

(43) Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596.

(44) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821.

(45) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247.

(46) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J .* **2007**, *92*, 3817.

(47) Cheatham III, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

(48) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(49) Zgarbová, M.; Otyepka, M.; Šponer, J. i.; Mládek, A. t.; Banáš, P.; Cheatham III, T. E.; Jurečka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886.

(50) Šponer, J.; Cang, X.; Cheatham III, T. E. *Methods* **2012**, *57*, 25.

(51) Dans, P. D.; Perez, A.; Faustino, I.; Lavery, R.; Orozco, M. *Nucleic Acids Res.* **2012**, *40*, 10668.

(52) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314.

(53) Beveridge, D. L.; DiCapua, F. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(54) Brice, A. R.; Dominy, B. N. *J. Comput. Chem.* **2011**, *32*, 1431.

(55) Hou, T.; Wang, J.; Li, Y.; Wang, W. *J. Chem. Inf. Model.* **2011**, *51*, 69.

(56) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(57) Sproviero, M.; Verwey, A. M. R.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 13405.

(58) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(59) Yan, S.; Wu, M.; Buterin, T.; Naegeli, H.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2003**, *42*, 2339.

(60) Stavros, K. M.; Hawkins, E. K.; Rizzo, C. J.; Stone, M. P. *Nucleic Acids Res.* **2014**, *42*, 3450.

(61) Mao, B.; Gorin, A.; Gu, Z.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1997**, *36*, 14479.

(62) Vaidyanathan, V. G.; Liang, F.; Beard, W. A.; Shock, D. D.; Wilson, S. H.; Cho, B. P. *J. Biol. Chem.* **2013**, *288*, 23573.

(63) Feng, B.; Gorin, A.; Hingerty, B. E.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1997**, *36*, 13769.

(64) Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Nat Struct Mol Biol* **2010**, *17*, 379.

(65) Lior-Hoffmann, L.; Ding, S.; Geacintov, N. E.; Zhang, Y.; Broyde, S. *Biochemistry* **2014**, *53*, 5683.

(66) Broyde, S.; Wang, L.; Zhang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J. *Chem. Res. Toxicol.* **2007**, *21*, 45.

(67) Perlow-Poehnelt, R. A.; Likhterov, I.; Wang, L.; Scicchitano, D. A.; Geacintov, N. E.; Broyde, S. *J. Biol. Chem.* **2007**, *282*, 1397.

(68) Bieler, C. A.; Stiborova, M.; Wiessler, M.; Cosyns, J. P.; van Ypersele de Strihou, C.; Schmeiser, H. H. *Carcinogenesis* **1997**, *18*, 1063.

(69) Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S. *Chem. Res. Toxicol.* **1998**, *11*, 335.

(70) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2012**, *40*, 9675.

(71) Wang, L.; Hingerty, B. E.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **2004**, *17*, 311.

(72) Cai, Y.; Zheng, H.; Ding, S.; Kropachev, K.; Schwaid, A. G.; Tang, Y.; Mu, H.; Wang, S.; Geacintov, N. E.; Zhang, Y.; Broyde, S. *Chem. Res. Toxicol.* **2013**, *26*, 1115.

(73) Cosman, M.; Fiala, R.; Hingerty, B. E.; Laryea, A.; Lee, H.; Harvey, R. G.; Amin, S.; Geacintov, N. E.; Broyde, S.; Patel, D. *Biochemistry* **1993**, *32*, 12488.

(74) Zegar, I. S.; Chary, P.; Jabil, R. J.; Tamura, P. J.; Johansen, T. N.; Lloyd, R. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1998**, *37*, 16516.

(75) Zegar, I. S.; Kim, S. J.; Johansen, T. N.; Horton, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1996**, *35*, 6212.

(76) Baird, W. M.; Hooven, L. A.; Mahadevan, B. *Environ. Mol. Mutagen.* **2005**, *45*, 106.

(77) Yan, S.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *J. Am. Chem. Soc.* **2001**, *123*, 7054.

(78) Rubbi, C. P.; Milner, J. *Carcinogenesis* **2001**, *22*, 1789.

(79) Min, J. H.; Pavletich, N. P. *Nature* **2007**, *449*, 570.

(80) Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J. *Biopolymers* **2002**, *65*, 202.

(81) Gunz, D.; Hess, M. T.; Naegeli, H. *J. Biol. Chem.* **1996**, *271*, 25089.

(82) Sugasawa, K.; Okamoto, T.; Shimizu, Y.; Masutani, C.; Iwai, S.; Hanaoka, F. *Genes Dev.* **2001**, *15*, 507.

(83) Schärer, O. D. *Mol. Cell.* **2007**, *28*, 184.

(84) Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgilialoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V. *Nucleic Acids Res.* **2014**, *42*, 5020.

(85) Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H. *Trends Biochem. Sci.* **2007**, *32*, 494.

(86) Yang, W. *DNA Repair* **2006**, *5*, 654.

(87) Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S. *Nucleic Acids Res.* **2011**, *39*, 8752.

(88) Mu, H.; Kropachev, K.; Chen, Y.; Zhang, H.; Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2013**, *52*, 5517.

(89) Hanawalt, P. C.; Spivak, G. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 958.

Chapter 3

Conformational Preferences of Adenine versus Guanine DNA Adducts of Aristolochic Acid-II[a,b]

3.1 Introduction

In contrast to the adenine adducts studied in the previous chapter, the guanine lesions of aristolochic acids have been observed to be repair prone.[1] As discussed in Section 1.7.2, bulky DNA lesions including those formed by aristolochic acids are repaired by GGR pathway. Studies on a variety of bulky DNA adducts have revealed that GGR propensity correlates with changes in a number of local structural features upon DNA damage, including significant perturbations to helical parameters (such as minor groove dimension, rise and twist),[2-6] increases in helical[7-10] and backbone dynamics,[2] and decreases in the stacking interactions at the lesion site,[11,12] as well as global changes to the DNA helix, such as helix bending.[13] Furthermore, a crystal structure of the yeast orthologue (Rad4) of XPC-RAD23B bound to DNA containing the cyclobutane-pyrimidine (CPD) dimer indicates that GGR recognition is initiated by the insertion of a β-hairpin through the major groove side of damaged DNA and subsequent flipping of the bases opposing the lesion out of the helix.[14] To complement this information, a recent crystal structure of Rad4 bound to undamaged DNA led to speculations that lesion recognition may depend on the time required to open a (damaged) DNA site (opening time) and the time the recognition factor spends at the

---

[a]This chapter is reproduced with permission from: Kathuria, P.; Sharma, P.; Wetmore, S. D. *Nucleic Acids Research* **2015**, 10.1093/nar/gkv701.

[b]My contribution to the publication was running all the calculations, analysis of the data and writing the preliminary manuscript.

damaged site (residence time).[15] In addition, the presence of distortions at the lesion site have been hypothesized to decrease the opening time and increase the residence time of the recognition factor.[15] Most importantly, this literature suggests that the structural characteristics of damaged DNA are crucial for understanding the issue of the 'repair versus persistence' of bulky lesions.

In this chapter, the structural properties of the repair-prone ALII-N^2-dG adduct are compared with the repair-resistant ALII-N^6-dA adduct previously studied using NMR[16] and computational methods (Chapter 2). Specifically, within the same 11-mer oligonucleotide, both NMR and MD simulations reveal a preference for an intercalated ALII-N^6-dA adducted DNA conformation with the opposing thymine displaced into the major groove. However, no analogous data has been reported for the ALII-N^2-dG adduct. Thus, the ALII-adducted nucleobases, nucleosides and nucleotides are initially considered in the present chapter using DFT in order to determine intrinsic conformational differences between the damaged purines in terms of the orientation of the ALII and deoxyribose moieties with respect to the nucleobase. Subsequently, the conformational space of ALII-adducted DNA is explored using MD simulations on an 11-mer damaged DNA sequence. In the absence of experimental NMR data in the literature, this study provides structural details of ALII-N^2-dG adducted DNA with the help of molecular modeling. Furthermore, the calculations reveal key differences in the structural features of ALII-N^2-dG and ALII-N^6-dA, as well as the associated adducted DNA helices, that likely play a critical role in dictating the observed differential GGR recognition of these adducts.

3.2 Computational Details

Following the a similar computational protocol as discussed in Chapter 2, the B3LYP/6-31G(d) potential energy surface (PES) was considered as a function of the θ and ϕ dihedral angles for the ALII-N^2-G nucleobase adduct, and as a function of χ and θ for the nucleoside adduct (dihedral angles are defined in Figure 3.1b). The minima identified on the PESs were subsequently optimized using B3LYP-D3/6-31G(d). The ALII-N^2-dG nucleotide was built by adding a (Na^+ ion neutralized) 5′–monophosphate unit to the lowest energy *anti* and *syn* nucleoside conformers, and the structures were energy minimized in water using PCM-B3LYP-D3/6-31G(d). Single-point energy calculations were carried out using B3LYP-D3/6-311+G(2df,p) for the nucleobase and nucleoside minima, and PCM-B3LYP-D3/6-311+G(2df,p) for the nucleotide minima. MD simulations were performed by incorporating the adduct into the 11-mer DNA sequence (5′–CGTACXCATGC, X = adduct) previously studied for ALII-N^6-dA (Chapter 2).[16] All simulations were carried out in water containing 20 Na^+ ions for 20 ns (excluding trial simulations) using the AMBER 11[17] or 12 program.[18] MM-PBSA free energy (Section 2.2.4.5, Chapter 2) calculations were performed on simulation trajectories.[19] Full details of the computational protocol are provided in Appendix B (Section B1).

3.3 Results and Discussion

3.3.1 ALII-N^2-dG intrinsically prefers a twisted conformation, while ALII-N^6-dA prefers a planar conformation, about the carcinogen-purine linkage.

DFT energies plotted as a function of θ and ϕ for the ALII-N^2-G nucleobase adduct reveal two low-energy conformations with a twisted orientation of ALII with

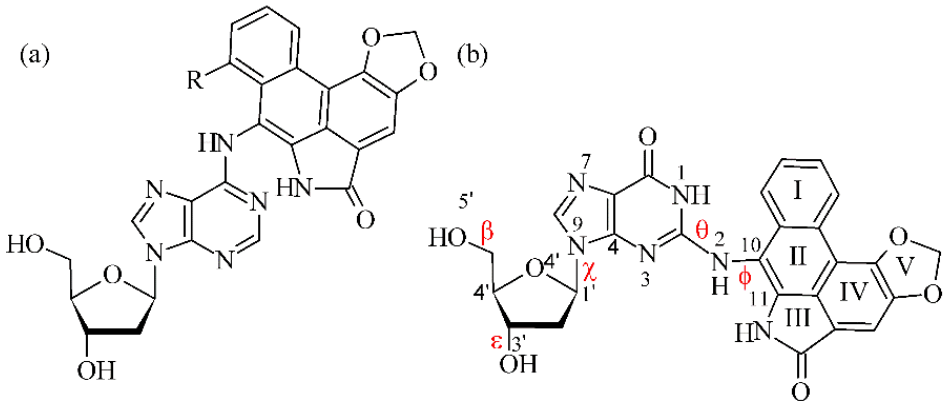


Figure 3.1: Structure of (a) AL-N^6-dA and (b) AL-N^2-dG adducts (R = OCH_3 for ALI and H for ALII). Definitions are provided for the θ (∠(N1C2N2C10)) and ϕ (∠(C2N2C10C11)) dihedral angles, which determine the orientation of the ALII moiety with respect to the base, and χ (∠(O4′C1′N9C4)), which dictates the glycosidic bond orientation to be* syn *(90° > χ > –60°), high* syn *(120° > χ > 90°),* anti *(120° > χ >–90°) or high* anti *(–60° > χ >–90°).[20] Dihedral angles β (∠(C4′C5′O5′H)) and ε (∠(C4′C3′O3′H)) govern the DNA sugar–phosphate backbone orientation.

respect to G (Figures 3.2a (right) and B1, and Section B2, Appendix B). This twist decreases steric repulsion between N1–H of damaged G and the lactam N–H group of ALII. In contrast, the lowest energy ALII-N^6-A nucleobase minimum is planar (Figure 3.2a (left)) because of stabilizing hydrogen bonding between N1 of A and the lactam N–H of ALII (Chapter 2). DFT nucleoside energies plotted as a function of χ and θ reveal that both ALII-N^2-dG (Figures B3 and B2 and Section B2, Appendix B) and ALII-N^6-dA (Chapter 2) intrinsically prefer a *syn* glycosidic orientation (by ~ 14 and 25 kJ mol^{-1}, respectively), mainly due to an interaction between the 5′–OH group and N3 (Figure 3.2b). Although such a conformational preference is relevant when the adduct is located at the 5′-terminal position (where the 5′–OH is not capped by a phosphate group), the presence of the 5′-phosphate group may alter the preferences at non-terminal DNA positions. Therefore, in order to neglect such spurious interactions, 5′–

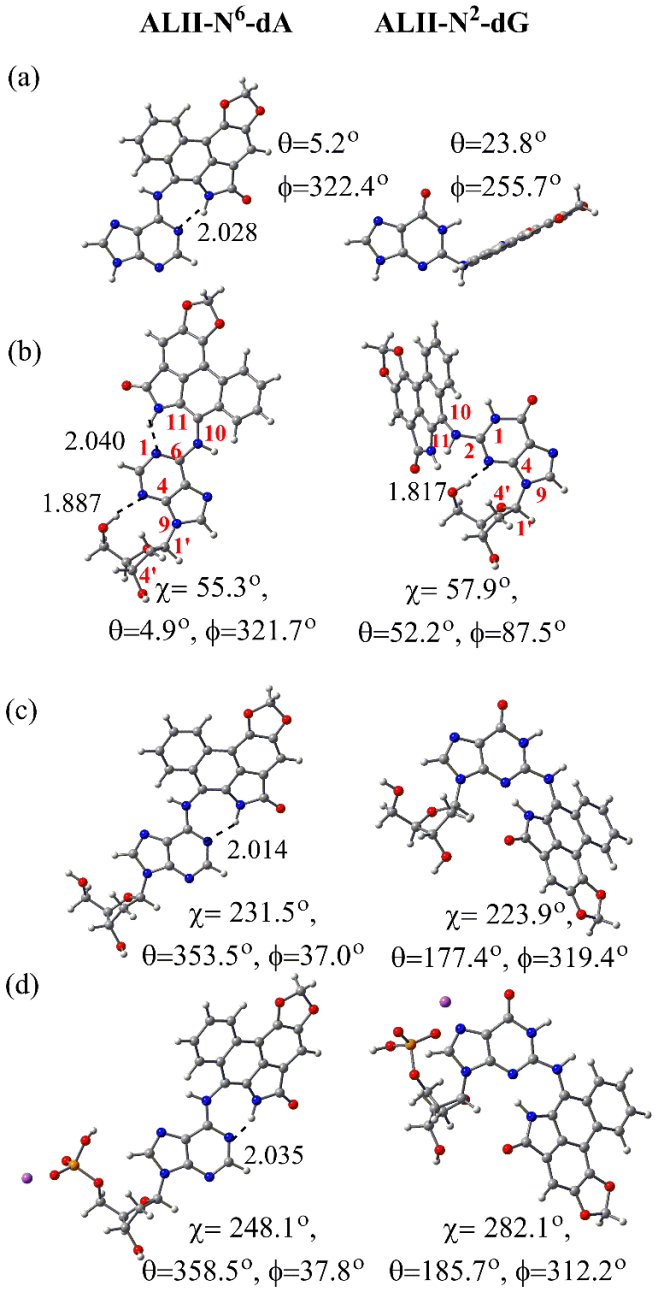


Figure 3.2: DFT (B3LYP-D3/6-31G(d)) minimum energy conformations of ALII-N⁶-dA (left, Chapter 2) and ALII-N²-dG (right) according to (a) nucleobase, (b) nucleoside, (c) 5′–OH constrained nucleoside and (d) nucleotide models. (θ = ∠(N1C2N2C10)), ϕ = ∠(C2N2C10C11) and χ = ∠(O4′C1′N9C4) for ALII-N²-dG and θ = ∠(N1C6N6C10)), ϕ = ∠(C6N6C10C11) and χ = ∠(O4′C1′N9C4) for ALII-N⁶-dA)

OH was subsequently directed (and fixed) away from the nucleobase (Figures 3.2c, and B4a and b, and Section B2, Appendix B). This results in the *anti* conformation being slightly more stable than the *syn* for both ALII-N^2-dG (by ~ 7 kJ mol^{-1}) and ALII-

N^6-dA (by ~ 11 kJ mol^{-1}, Chapter 2). However, steric repulsion arising from the close proximity of the 5′–phosphate and ALII moieties in the *syn* orientation renders the *anti* ALII-N^2-dG nucleotide 27 kJ mol^{-1} more stable than the *syn* conformer (Figures 3.2d, and B4c and d, and Section B2, Appendix B). In contrast, since the ALII moiety of the ALII-N^6-dA nucleotide does not interact with the 5′–phosphate in either the *syn* or *anti* conformation, *anti* ALII-N^6-dA is the most stable conformer by only ~ 10 kJ mol^{-1} (Chapter 2). Overall, irrespective of the DFT model considered, ALII-N^2-dG prefers a twisted conformation, while ALII-N^6-dA adopts a planar conformation, at the carcinogen-purine linkage.

3.3.2 Twisted conformation about the carcinogen-purine linkage facilitates greater conformational heterogeneity of ALII-N^2-dG compared to ALII-N^6-dA adducted DNA.

3.3.2.1 ALII-N^2-dG adducted DNA is conformationally heterogeneous. Six distinct conformers were isolated from MD simulations of ALII-N^2-dG adducted DNA, which differ in the location of the ALII moiety and the (*anti*/*syn*) glycosidic orientation of the adducted nucleotide (Figure 3.3a). Two of these conformations (i.e., *anti* base-displaced intercalated and 3′–intercalated) correspond to the *anti* adduct glycosidic orientation, one (i.e., minor groove stacked) involves the high *anti* orientation, and three (i.e., *syn* base-displaced intercalated, 5′–intercalated and 3′,5′–intercalated) correspond to the *syn* glycosidic orientation. Calculated (MM-PBSA) free energies indicate a close energetic separation between all six conformations (< 21 kJ mol^{-1}; Figure 3.3a). Salient features of each conformation are described below, with details provided in Appendix B (Section B2).

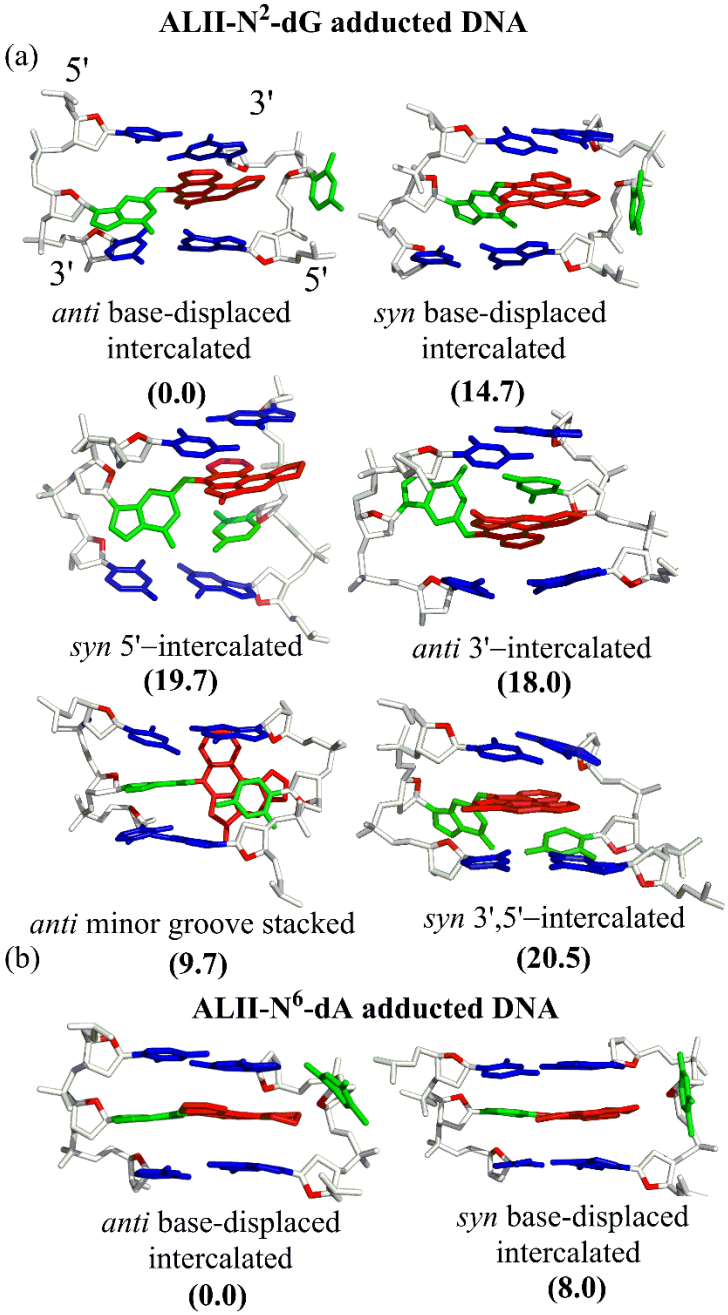


Figure 3.3: Base-pair trimers containing the lesion site in representative MD structures of energetically-accessible conformations of (a) ALII-N^2-dG and (b) ALII-N^6-dA adducted DNA. The relative free energies (kJ mol^{-1}) with respect to the lowest energy conformation for a given adduct are provided in parantheses.

3.3.2.2 Five of the accessible ALII-N^2-dG adducted DNA conformations are analogous to those previously observed for other DNA adducts. The *anti* base-displaced intercalated conformation of ALII-N^2-dG adducted DNA is characterized by stacking of the ALII moiety between the 5′–base pair and 3′–flanking base in the opposing strand, the extrahelical (major groove) location of the opposing C and complete loss of Watson- Crick hydrogen bonding in the damaged base pair (Figure 3.3a). This structure is similar to that previously reported for DNA containing the two stereoisomeric *cis*-*anti*-[BP]-N^2-dG polycyclic aromatic hydrocarbon (PAH) adducts,[21,22] which exhibit (major groove) displacement of the opposing cytosine and a twisted conformation at the carcinogen-dG linkage. Nevertheless, the origin of the twist differs between the PAH and ALII adducts. Specifically, the twist in ALII-N^2-dG alleviates steric repulsion between N1–H of G and the lactam N–H group, while the twist in the PAH adducts arises due to the absence of extended conjugation between the damaged guanine and PAH moieties (i.e., sp^3 hybridization of the linker C of the carcinogen). In contrast, DNA damaged by a heterocyclic aromatic amine carcinogenic adduct (i.e., IQ-N^2-dG) adopts a base-displaced intercalated conformation with a nearly planar carcinogen-guanine linkage due to extended conjugation (i.e., sp^2 hybridization of the linker C of the carcinogen) and the absence of intrinsic steric repulsion.[23]

Although the damaged guanine retains stacking interactions with its 3′–flanking pair in the *syn* base-displaced intercalated conformation of ALII-N^2-dG adducted DNA, only the ALII moiety stacks with the 5′–flanking pair (Figure 3.3a). To the best of my knowledge, no example exists in the literature of N2-dG adducted DNA

in such a conformation. Nevertheless, this structure is analogous to the *syn* base-displaced intercalated conformation observed for several C8-dG adducts[24-27] in terms of the (major groove) extrahelical opposing cytosine. However, damaged guanine is displaced towards the major groove in C8-dG adducted DNA versus the minor groove in ALII-N^2-dG adducted DNA.

In the *syn* 5′–intercalated conformation, the ALII moiety stacks between the opposing (intrahelical) cytosine and the 5′–base pair with respect to the lesion (Figure 3.3a). Although this conformation has not been reported in the literature for other N2-dG adducts, some N6-linked PAH adducts acquire an analogous conformation.[28-30] In contrast, the *anti* 3′–intercalated conformer (i.e., the ALII moiety stacks between the opposing (intrahelical) cytosine and the 3′–base pair with respect to the lesion; Figure 3.3a) has been observed for 14R (+)-*trans*-*anti*-DB[*a,l*]P-N^2-dG PAH adducted DNA.[31] However, the PAH adducted DNA conformation involves simultaneous stacking of the damaged guanine and PAH moieties with the flanking bases (due to flexibility imparted by the sp^3 hybridization of the linker C of the PAH moiety), while only the ALII ring stacks with the flanking bases in ALII-N^2-dG adducted DNA.

In the *anti* minor-groove stacked conformation, cytosine opposing the lesion twists to form van der Waals (π–π stacking) interactions with the ALII moiety that is located in the minor groove (Figure 3.3a). A variant of this conformation was previously reported for 14S *cis*-*anti*-DB[*a,l*]N^2-dG PAH adducted DNA.[32] However, the minor-groove stacked conformation associated with this PAH adduct ruptures WC pairing in both the damaged and flanking base pairs, while the ALII-N^2-dG adducted

DNA conformer retains WC pairing in the flanking pairs, and a N4–H•••O6 hydrogen bond in the lesion pair (Table B1, Appendix B).

3.3.2.3 ALII-N^2-dG adducted DNA adopts a unique (syn 3′,5′–intercalated) conformation. In addition to the conformations discussed above, ALII-N^2-dG adducted DNA adopts a unique conformation that has not been reported in the literature for other DNA adducts to date. In this conformation, the modified guanine and opposing cytosine shift towards the minor groove, while the ALII moiety simultaneously stacks with the 5′ and 3′–flanking pairs, as well as the opposing cytosine (Figure 3.3a). The opposing cytosine maintains a N4–H•••O6 hydrogen bond with the damaged guanine, and forms additional hydrogen bonds with the 3′–base pair with respect to the lesion (Table B1, Appendix B).

3.3.2.4 Structural differences at the carcinogen-purine linkage of the ALII-N^2-dG and ALII-N^6-dA adducts alter the conformational heterogeneity of damaged DNA. The *anti* base-displaced intercalated conformer is the most stable orientation for both ALII-N^2-dG and ALII-N^6-dA adducted DNA (Figure 3.3). However, the total number of energetically-accessible conformations differs between the two types of adducted DNA. Specifically, although MM-PBSA suggests five conformations lie within 21 kJ mol^{-1} of the most stable conformer for ALII-N^2-dG adducted DNA (Figure 3.3a), only one other conformation, namely the *syn* base-displaced intercalated orientation, falls within this energetic separation for ALII-N^6-dA adducted DNA (Figure 3.3b, Chapter 2). The reduced conformational heterogeneity of ALII-N^6-dA arises since the planar adduct forms highly stabilizing van der Waals (stacking) interactions with the flanking base pairs when the opposing C is extrahelical (i.e., the base-displaced

intercalated conformations). However, since the opposing C remains inside the helix in all other possible ALII-N^6-dA adducted DNA conformations, the rise between the flanking pairs increases, stacking with the lesion decreases and the associated adducted DNA conformations become energetically unfavorable (> approximately 35 kJ mol^{-1}, Chapter 2). In contrast, since ALII-N^2-dG is twisted, optimal stacking at the lesion site cannot be achieved in the corresponding base-displaced intercalated conformations, which prevents these conformers from becoming highly stabilized as for ALII-N^6-dA adducted DNA. In fact, the conformational heterogeneity for ALII-N^2-dG adducted DNA arises due to a complicated interplay between mutually compensating interactions arising from the twisted lesion, including van der Waals, steric and hydrogen-bonding interactions at the damaged site (Section B2, Appendix B). Thus, differences in the intrinsic (planar versus twisted) conformation at the carcinogen-purine linkage lead to differences in the conformational heterogeneity of ALII-N^6-dA and ALII-N^2-dG adducted DNA.

3.3.3 Increased lesion site distortions, diminished stacking and enhanced dynamics likely contribute to the greater propensity for GGR recognition of ALII-N^2-dG compared to ALII-N^6-dA.

3.3.3.1 Twisted conformation of ALII-N^2-dG leads to greater DNA distortions at the lesion site compared to ALII-N^6-dA. Structural distortions at the lesion site are believed to be one of the important factors in GGR recognition.[2,3,27] For example, the enhanced repair susceptibility of the (aromatic amine) AAF-C^8-dG adduct has been partly attributed to its greater (structurally distorting) impact on DNA compared to the corresponding nonacetylated AF-C^8-dG adduct.[27] In the case of the ALII adducts,

DFT calculations indicate that the ALII-N^2-dG nucleotide intrinsically prefers a twisted conformation at the carcinogen-purine linkage. This twist distorts DNA at the lesion site according to the MD pseudostep parameters (i.e., step parameters calculated between the base step consisting of the 3′ and 5′–base pairs with respect to the lesion; see Section B1, Appendix B) and minor groove dimensions. Specifically, each accessible conformation of ALII-N^2-dG adducted DNA significantly changes at least one of the lesion site pseudostep parameters compared to unmodified DNA (i.e., shift (up to 3 Å), slide (up to 6 Å), rise (up to 4 Å), tilt (up to 10°), roll (up to 27°) or twist (up to 27°)) or the minor groove dimensions (up to 6 Å, Figure 3.4). On the other hand, ALII-N^6-dA prefers an intrinsically planar conformation of the ALII moiety with respect to the damaged A, and therefore ALII-N^6-dA adducted DNA exhibits only minimal changes in the pseudostep parameters compared to unmodified DNA (i.e., shift (up to 1.4 Å), slide (up to 1.5 Å), rise (up to 0.2 Å), tilt (up to 2°), roll (up to 10°) or twist (up to 10°)) and minor groove dimensions (up to 1 Å; Figure 3.4, Chapter 2). This suggests that planar ALII-N^6-dA is better accommodated in DNA than the (distorting) twisted ALII-N^2-dG adduct.

3.3.3.2 Twisted conformation of ALII-N^2-dG diminishes lesion site stacking interactions. Previous experimental and computational studies indicate that DNA adducts prone to GGR recognition typically exhibit diminished stacking at the lesion site compared to GGR resistant adducts.[2,27,32,33] In the ALII adducts, the intrinsic twist in ALII-N^2-dG hinders simultaneous stacking of the damaged guanine and ALII moieties with the flanking bases in all accessible conformations of ALII-N^2-dG

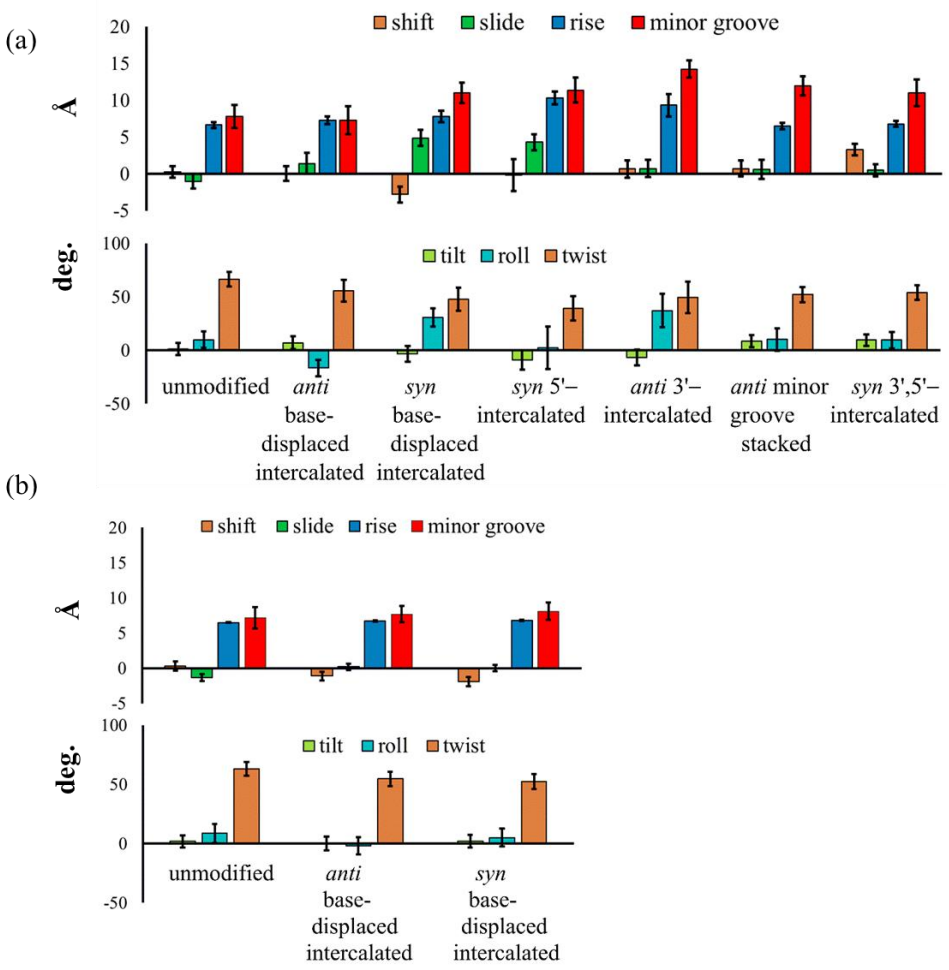


Figure 3.4: The pseudostep parameters and minor groove dimensions for different conformations of (a) ALII-N^2-dG or (b) ALII-N^6-dA (Chapter 2) adducted DNA relative to the corresponding unmodified helix. The helical dynamics are indicated by error bars.

adducted DNA. In contrast, the intrinsic planar conformation of ALII-N^6-dA facilitates simultaneous stacking of both the ALII and modified adenine moieties between the flanking base pairs in DNA. As a result, the calculated van der Waals (stacking) interactions involving the lesion are consistently smaller (less negative) for the repair prone ALII- N^2-dG than the repair resistant ALII-N^6-dA in DNA (Figure 3.5). Although the trend in relative NER propensities of the two ALII-purine adducts is similar to that observed for the 14R (+)-*trans*-*anti*-DB[*a*,*l*]P-N^2-dG and 14R (+)-*trans*-*anti*-DB[*a*,*l*]P-

N^6-dA adducts (i.e., the dG adduct is more repair prone than the dA adduct), the structural origin of the differential NER propensity differs between the two types of adducts. Specifically, the differential NER propensity of the PAH adducts has been interpreted in terms of structural distortions in the 14R (+)-*trans*-*anti*-DB[*a*,*l*]P-N^2-dG lesion originating from the loss of Watson-Crick hydrogen bonding between the damaged pair, and relatively weaker stacking interactions between the intercalated bulky moiety and the adjacent bases. On the other hand, the 14R (+)-*trans*-*anti*-DB[*a*,*l*]P-N^6-dA adduct retains Watson-Crick hydrogen-bonding within the damaged pair and is associated with greater lesion site stacking interactions, which makes this adduct more repair resistant.[32]

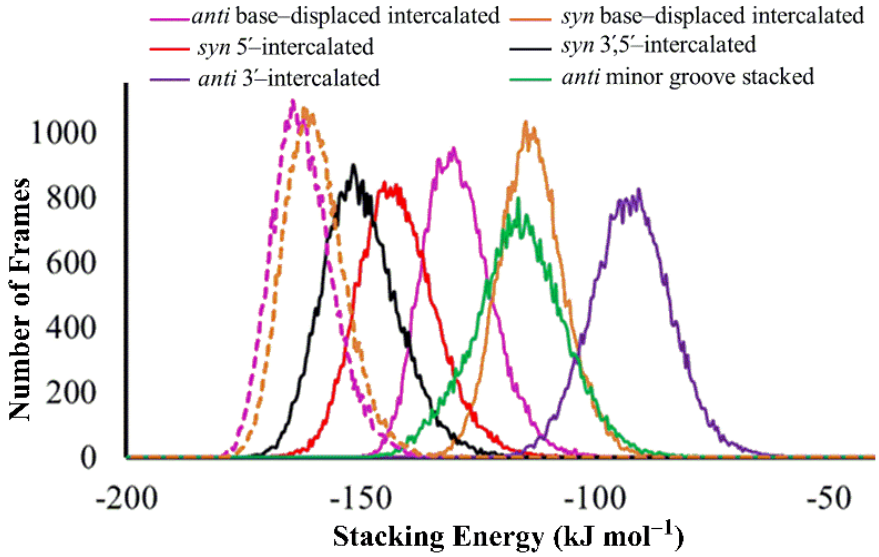


Figure 3.5: Comparison of lesion van der Waals (stacking) energies for* anti *or* syn *ALII-N^2-dG (solid lines) and ALII-N^6-dA (dashed lines) in different adducted DNA conformations over the 20 ns MD simulations.

3.3.3.3 DNA containing ALII-N^2-dG is structurally more dynamic than DNA containing ALII-N^6-dA. Previous studies have shown that damage recognition by

XPC-RAD23B is facilitated by increased lesion site dynamics.[2,7,8] For example, the greater repair propensity of the (+)-*trans*-*anti*-[BP]N^2-dG adduct in the TG*T over the CG*C sequence has been partly attributed to a more dynamic roll, bending and greater minor groove flexibility in the former sequence.[8] Similarly, greater lesion site backbone dynamics of the 5′*R* compared to the 5′*S* enantiomer of the 5′,8-cyclo-2′-deoxypurine lesion correlates with the greater NER susceptibility of the 5′*R* enantiomer.[2] For the ALII adducts, the standard deviations in the pseudostep parameters at the lesion site in ALII-N^6-dA adducted DNA (maximum of 1.4 Å for the translational parameters (i.e., shift, slide and rise), 1.2 Å for the minor groove width and 6.3˚ for the rotational parameters (i.e., tilt, roll and twist; Figure 3.4) are similar to those for the corresponding unmodified DNA (maximum of 1.0 Å for the translational parameters, 1.5 Å for the minor grove width and 6.0˚ for the rotational parameters). However, all accessible conformations of ALII-N^2-dG adducted DNA typically exhibit greater standard deviations in the lesion site parameters (maximum of 2.2 Å for the translational parameters, 1.9 Å for the minor grove width and 15˚ for the rotational parameters) than the corresponding unmodified DNA (maximum of 1.0 Å for the translational parameters, 1.6 Å for the minor grove width and 7.6˚ for the rotational parameters).

3.3.3.4 Differences in structural features of ALII-N^2-dG and ALII-N^6-dA adducted DNA may lead to better recognition of ALII-N^2-dG in cells. The above discussion highlights that, due to an intrinsically twisted carcinogen-purine linkage, ALII-N^2-dG induces greater structural distortions to DNA, leads to smaller lesion site van der Waals stacking interactions, and induces significantly greater helical flexibility at the

lesion site compared to ALII-N^6-dA adduct. These factors likely combine to contribute to its enhanced recognition of ALII-N^2-dG over ALII-N^6-dA by the GGR pathway. This provides a correlation between a twisted carcinogen-purine linkage and GGR propensity, which expands the current list of factors that affect the GGR of DNA adducts (e.g., substitution at the carcinogen-base linkage,[27] stereochemistry,[2,32,34] adduct ionization state,[24] sequence context,[8,27] and the identity of the partner base[35]).

The newly identified differences in the structural features of ALII-N^2-dG and ALII-N^6-dA adducted DNA also provide information about XPC-RAD23 binding to adducted DNA. Specifically, the significant structural distortions and diminished van der Waals (stacking) interactions at the lesion site in all accessible conformations of ALII-N^2-dG adducted DNA will likely decrease the opening time and increase the residence time of the recognition factor at the damage site, which will in turn facilitate lesion recognition. In contrast, the smaller structural distortions and enhanced stacking interactions at the lesion site in ALII-N^6-dA adducted DNA, due to the planar ALII–dA linkage (Chapter 2),[16] will likely increase the opening time and decrease the residence time of the recognition factor, and thereby render this adduct resistant to GGR. Furthermore, the major-groove intercalation of the bulky moiety may block β-hairpin insertion through the major-groove side of DNA in the most stable (*anti* base-displaced intercalated) ALII-N^6-dA adducted DNA conformation, while the minor-groove intercalation of the bulky moiety in the most stable ALII-N^2-dG adducted DNA conformer may facilitate repair. Nevertheless, the conformational heterogeneity of ALII-N^2-dG found in the present study allows the bulky moiety to adopt a number of helical positions and therefore underscores that differences in key structural features

at the lesion site are likely the primary explanation for the enhanced GGR of ALII-N^2-dG over ALII-N^6-dA.

3.4 Conclusions

The present computational study provides comprehensive structural details of the conformational preferences of ALII-N^2-dG adducted DNA with the help of molecular dynamics simulations. The calculations reveal an intrinsic twist at the linkage between the damaged G and ALII moieties, which leads to conformational heterogeneity of ALII-N^2-dG adducted DNA. Comparison of the structural features of ALII-N^2-dG and ALII-N^6-dA adducted DNA provides a structural explanation for the observed differential GGR recognition propensities of these purine adducts. Specifically, the intrinsic twisted conformation of ALII-N^2-dG induces destabilizing distortions to DNA at the lesion site, reduces van der Waals (stacking) interactions with the neighboring base pairs, and enhances the helical dynamics at the damaged site, which may collectively facilitate GGR recognition. In contrast, smaller distortions, greater stacking stabilization, and decreased dynamics at the lesion site likely render ALII-N^6-dA adducted DNA resistant to GGR. This GGR resistance is likely the primary factor responsible for the greater abundance and nephrotoxic potential of ALII-N^6-dA compared to ALII-N^2-dG in cells affected by aristolochic acids. The present study highlights the critical role of the conformation at the carcinogen-DNA linkage in determining GGR recognition and excision propensities, and thereby extends the list of established effects on GGR recognition.

3.5 References

(1) Sidorenko, V. S.; Yeo, J. E.; Bonala, R. R.; Johnson, F.; Schärer, O. D.; Grollman, A. P. *Nucleic Acids Res.* **2012**, *40*, 2494.

(2) Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgilialoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V. *Nucleic Acids Res.* **2014**, *42*, 5020.

(3) Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J. *Biopolymers* **2002**, *65*, 202.

(4) Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2010**, *17*, 379.

(5) Schärer, O. D. *Mol. Cell.* **2007**, *28*, 184.

(6) Sugasawa, K.; Okamoto, T.; Shimizu, Y.; Masutani, C.; Iwai, S.; Hanaoka, F. *Genes Dev.* **2001**, *15*, 507.

(7) Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H. *Trends Biochem. Sci.* **2007**, *32*, 494.

(8) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(9) Blagoev, K. B.; Alexandrov, B. S.; Goodwin, E. H.; Bishop, A. R. *DNA Repair* **2006**, *5*, 863.

(10) Isaacs, R. J.; Spielmann, H. P. *DNA Repair* **2004**, *3*, 455.

(11) Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S. *Nucleic Acids Res.* **2011**, *39*, 8752.

(12) Yang, W. *DNA Repair* **2006**, *5*, 654.

(13) Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D.; Cho, B. P. *Nucleic Acids Res.* **2013**, *41*, 869.

(14) Min, J.-H.; Pavletich, N. P. *Nature* **2007**, *449*, 570.

(15) Chen, X.; Velmurugu, Y.; Zheng, G.; Park, B.; Shim, Y.; Kim, Y.; Liu, L.; Van Houten, B.; He, C.; Ansari, A.; Min, J. H. *Nat. Commun.* **2015**, *6*, 10.1038/ncomms6849.

(16) Lukin, M.; Zaliznyak, T.; Johnson, F.; de los Santos, C. *Nucleic Acids Res.* **2012**, *40*, 2759.

(17) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; *AMBER 11* University of California, San Francisco, CA: **2010**.

(18) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; *AMBER 12* University of California, San Francisco, CA: **2012**.

(19) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314.

(20) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, NY, **1984**, 9.

(21) Cosman, M.; Hingerty, B. E.; Luneva, N.; Amin, S.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1996**, *35*, 9850.

(22) Cosman, M.; Fiala, R.; Hingerty, B. E.; Laryea, A.; Lee, H.; Harvey, R. G.; Amin, S.; Geacintov, N. E.; Broyde, S.; Patel, D. *Biochemistry* **1993**, *32*, 12488.

(23) Stavros, K. M.; Hawkins, E. K.; Rizzo, C. J.; Stone, M. P. *Nucleic Acids Res.* **2014**, *42*, 3450.

(24) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(25) Wang, F.; DeMuro, N. E.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *J. Am. Chem. Soc.* **2006**, *128*, 10085.

(26) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 81.

(27) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2012**, *40*, 9675.

(28) Cai, Y.; Ding, S.; Geacintov, N. E.; Broyde, S. *Chem. Res. Toxicol.* **2011**, *24*, 522.

(29) Li, Z.; Mao, H.; Kim, H. Y.; Tamura, P. J.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1999**, *38*, 2969.

(30) Li, Z.; Tamura, P. J.; Wilkinson, A. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **2001**, *40*, 6743.

(31) Tang, Y.; Liu, Z.; Ding, S.; Lin, C. H.; Cai, Y.; Rodriguez, F. A.; Sayer, J. M.; Jerina, D. M.; Amin, S.; Broyde, S.; Geacintov, N. E. *Biochemistry* **2012**, *51*, 9751.

(32) Kropachev, K.; Kolbanovskiy, M.; Liu, Z.; Cai, Y.; Zhang, L.; Schwaid, A. G.; Kolbanovskiy, A.; Ding, S.; Amin, S.; Broyde, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2013**, *26*, 783.

(33) Yan, S.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *J. Am. Chem. Soc.* **2001**, *123*, 7054.

(34) Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2012**, *51*, 1486.

(35) Mu, H.; Kropachev, K.; Chen, Y.; Zhang, H.; Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2013**, *52*, 5517.

Chapter 4

Effect of Base Sequence Context on Aristolactam-I Adducted DNA Conformations[a,b]

4.1 Introduction

In addition to being repair resistant, the nephrotoxic potential of the AL-adenine adduct has also been experimentally-observed to be sequence dependent.[1, 2] However, all the currently available (computational) structural details of ALI-N^6-dA adducted DNA have only been obtained in a single (CXC) sequence context (Chapter 2), previous studies on a variety of carcinogenic adducts, including those formed from aromatic amines,[3-6] PAHs,[7-11] and ochratoxin A,[12, 13] 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.[4, 5, 7, 11, 12] For example, although the 10*S* (+)-*trans*-*anti*-B[*a*]P-N^2-dG PAH adduct preferentially induces a minor groove conformation at G^6 or G^7 in the 5′–CATGCG^{6}G^{7}CCTAC oligonucleotide, greater distortion occurs with the adduct at G^7 because of intrastrand steric clashes with the 5′–flanking guanine.[7] Furthermore, within the 5′–CTCG^{1}G^{2}CG^{3}CCATC oligonucleotide containing the *Nar*I mutational hotspot sequence (underlined),[14] the base-displaced intercalated conformer (the bulky moiety in the *syn* damaged base stacks in the helix and renders the opposing base extrahelical) of

---

[a]This chapter is adapted from submitted manuscript: Kathuria, P.; Sharma, P.; Wetmore, S. D. *Chemical Research and Toxicology* **2015**, Manuscript id: tx-2015-00240s.
[b]My contribution to the publication was running all the calculations, analysis of the data and writing the preliminary manuscript.

AF-C^{8}-dG aromatic amine adducted DNA is most stabilized with the lesion at G^{3} due to better stacking between the bulky moiety and the flanking bases in the opposing strand than at G^{1} or G^{2}.[4, 5] 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^{3} in the *Nar*I sequence exhibits greater stabilization of the minor groove conformer compared to the major groove and base-displaced intercalated conformations.[12] 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 *Nar*I sequence-containing adducted DNA acquires a minor groove conformation when the IQ-C^{8}-dG aromatic amine adduct is located at G^{1} or G^{2}, but a base-displaced intercalated conformation with the lesion at G^{3} position.[3]

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.[7-10, 15-20] Bulky DNA lesions, including those formed from aristolochic acids, are most commonly repaired through the nucleotide excision repair (NER) pathway,[21] which may proceed either through a transcription-coupled repair (TCR) or a global genome repair (GGR) mechanism.[22-24] Whereas TCR only deals with transcriptionally-active regions,[24] GGR operates on lesions spanning the entire genome.[23] 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,[25-27] experimental studies have determined that the AL-N^6-dA adducts are resistant to GGR, and are exclusively repaired by TCR.[28] However, the structural basis of this GGR resistance and the related sequence dependence of repair propensity of AL-N^6-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.[9, 19, 29-36] 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^6-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.[8] Similarly, the 10S (+)-*trans*-*anti*-B[*a*]P-N^2-dG PAH adduct induces a flexible kink when located at G^7 in 5′–CATGCG^{6}G^{7}CCTAC due to steric interactions with the 5′–flanking guanine, which leads to better recognition by NER machinery at G^7 compared to G^6.[7] Alternatively, the 10S (+)-*trans*-*anti*-B[*a*]P-N^2-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.[9, 15]

In addition to influencing the GGR repair propensity, the identity of the bases flanking the lesion site,[37-39] as well as the next nearest neighbors,[40] has been shown to affect the mutagenic profile of DNA lesions. For example, the mutational frequency of the PhIP-C^8-dG aromatic amine adduct is greater in the 5′–TCCTCCTNXCCTCTC oligonucleotide when N = C or G compared to A or T.[39] 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.[37] 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^2-dG PAH adduct, which exhibits dGMP insertion opposite the lesion in the CXG sequence context, and dAMP insertion in the TXG motif.[38]

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^6-dA lesions, a previous study indicates that AAII-damaged DNA is thermodynamically more stable when the adduct is present in the CXC sequence context compared to T/CXG motifs.[41] Nevertheless, recent studies on patients exposed to AAs have revealed that signature A→T mutations most commonly occur in T/CXG motifs.[1, 2] 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^6-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^6-dA formation has yet to be explored.

In the present work, classical (unrestrained) MD simulations and post-processing free energy calculations are utilized to investigate the role of sequence context in dictating the conformational outcomes of ALI-N^6-dA, the most persistent and mutagenic DNA adduct arising from the AAs. Specifically, ALI-N^6-dA (X) is incorporated into three 11-mer oligonucleotides (i.e., 5′–CGTA<u>CXG</u>ATGC, 5′–CGTA<u>GXC</u>ATGC, and 5′–CGTA<u>GXG</u>ATGC), 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 the 5′–CGTA<u>CXC</u>ATGC oligonucleotide studied in Chapter 2, which places a pyrimidine (C) on both sides of the lesion. This present chapter focuses 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^6-dA adduct have similar stabilities, while the CXC motif leads to the most stable helices.[41] Due to the similarity in the conformational preferences of ALII-N^6-dA and ALI-N^6-dA predicted in Chapter 2, it is reasonable to assume that the sequence dependent conformational preferences of ALI-N^6-dA will not be wildly different from ALII-N^6-dA in the absence of experimental data on the

sequence dependent stability of ALI-N^6-dA adducted DNA. The detailed comparison of the preferred conformations of ALI-N^6-dA adducted DNA in different lesion site sequence contexts provides valuable clues about the origin of the experimentally-observed sequence-dependent strand stability. Furthermore, the present 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.

4.2 Computational Details

The ALI-N^6-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) (Chapter 2). The B-form of canonical DNA was initially built for the 5′–CGTACXCATGC (X = A) sequence using the NAB[42] module of AMBER 11.[43] The ALI moiety was attached to the N6 position of A at the X position using GaussView.[44] The previously characterized lowest energy *anti* and *syn* nucleotide orientations (with $\theta \sim 0^{\circ}$, Figure 2.1a) were incorporated into DNA (Chapter 2), 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^6-dA) adducted DNA oligonucleotide (Chapter 2). All force field parameters, as well as simulation protocol were adopted from ALI-N^6-dA adducted DNA studied in Chapter 2. Final unrestrained

production MD simulations were subsequently run at 300 K for 20 ns using the PMEMD[45] module of AMBER 12.[46] The simulation time is justified by study of AL-N^6-dA adducted DNA in the CXC sequence (Chapter 2), which illustrated that extending the simulation time to 320 ns does not significantly change lesion site structural parameters (Tables C1 and C2, Appendix C). 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 C1, Appendix C). 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 C2, Appendix C). Additionally, unimodal distributions were obtained with respect to the χ, θ and ϕ dihedral angles of the adducted nucleotide (Figure 1 b) throughout each simulation (Figures C1 – C3, Appendix C). Therefore, a single representative structure was extracted by clustering each simulation trajectory with respect to the position of the atoms forming χ, θ, and ϕ 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.[47] 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, since all adducted DNA oligonucleotides studied in the present work have the same chemical composition (i.e., the lesion pair is always flanked by a G:C base pair), the relative total free energies (ΔG_{total}) are also reported for the energetically accessible (base-displaced intercalated) conformer with respect to the conformation with the lowest G_{total} (regardless of the sequence context).

4.3 Results and Discussion

4.3.1 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.

4.3.1.1 The Base-displaced Intercalated Conformers. The base-displaced intercalated conformation is characterized by the slight displacement of the damaged base the minor (major) groove in the *anti* (*syn*) nucleotide orientation

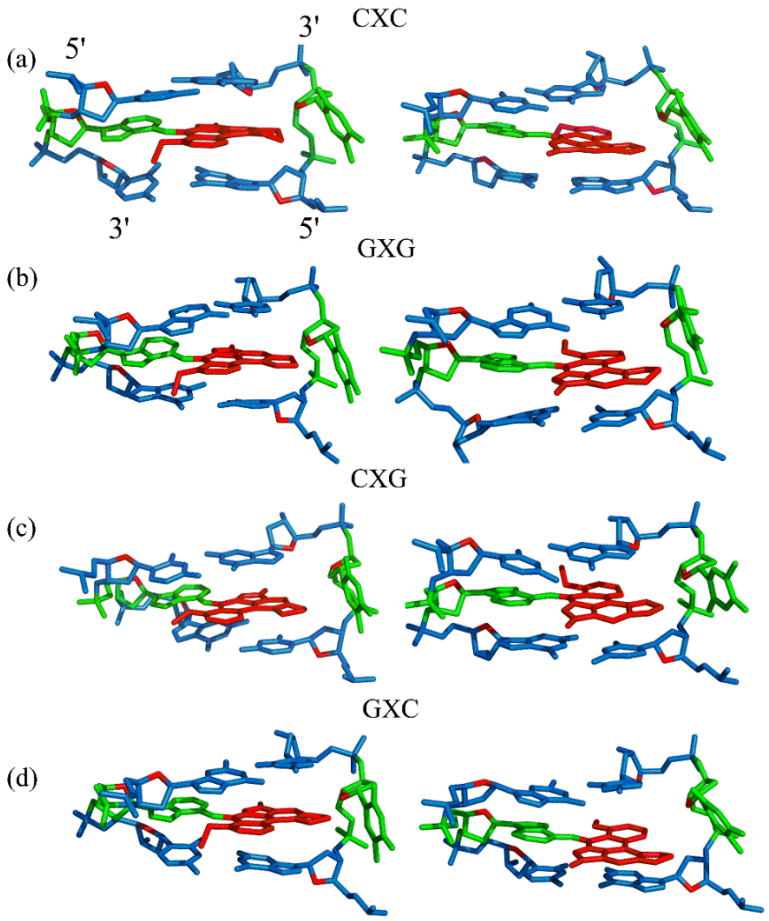


Figure 4.1: Representative structures from MD simulations with the* anti *(left) and* syn *(right) orientation of the ALI-N^6-dA adduct paired opposite thymine in the different sequence contexts for the base-displaced intercalated DNA conformation.

(Figure 4.1). In the base-displaced intercalated conformers, both the damaged adenine and ALI 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•••π interaction between the methylene group in ring V of the ALI moiety (Figure 2.1a) and the π–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 (χ

= ~ 200 – 240°) 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 C3, Appendix C). In this conformational theme, the translational pseudostep parameters (shift, slide and rise) change by ~ 0.3 – 2.0 Å, while the rotational parameters (tilt, roll, twist) change by ~ 1 – 14°, with respect to the corresponding unmodified DNA (Table C4, Appendix C). Overall, irrespective of the sequence context, this conformational theme causes minimal distortions to DNA upon adduct incorporation.

4.3.1.2 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 4.2). Regardless of the sequence considered, hydrogen bonding in the 5′–flanking base pair remains intact (> 98% occupancy for each G:C hydrogen bond; Table C3, Appendix C). Except in the CXG motif, where both rings I and V are solvent exposed in the major groove, only ring I (Figure 2.1a) 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 (~ 50% occupancy of the N4–H4•••O4 and ~ 40% occupancy of the N3•••H2–N2 hydrogen bond; Table C3, Appendix C) and its opposing G (> 95% occupancy of the O2•••H2–N2 hydrogen bond; Table C3, Appendix C). In contrast, both 5′–intercalated conformers maintain hydrogen

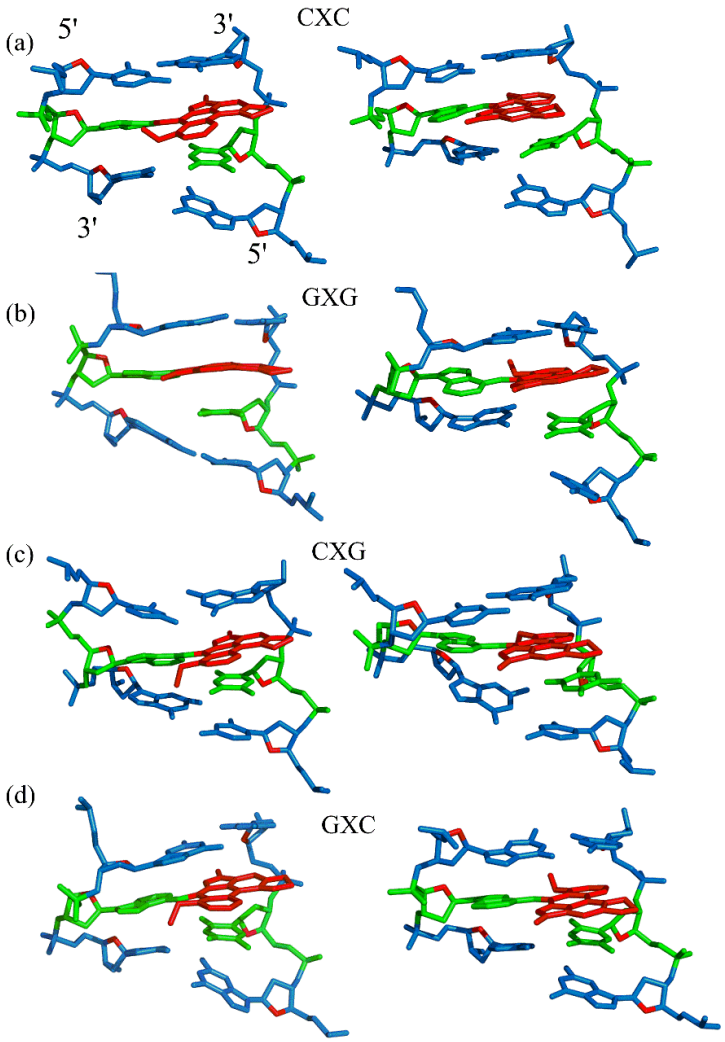


Figure 4.2: Representative structures from MD simulations with the* anti *(left) and* syn *(right) orientation of the ALI-N^6-dA adduct paired opposite thymine in the different sequence contexts for the 5′–intercalated DNA conformation.

bonding in the 3′–flanking base pair for the CXG sequence context (Table C3, Appendix C). Although the hydrogen bonds in the 3′–flanking base pair are maintained in the *anti* 5′–intercalated conformer for GXG (Table C3, Appendix C), 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 C3, Appendix C), 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 C3, Appendix C).

4.3.1.3 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 4.3). 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 C3, Appendix C). 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 C3, Appendix C). Finally, the 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 (Chapter 2). Similar to 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

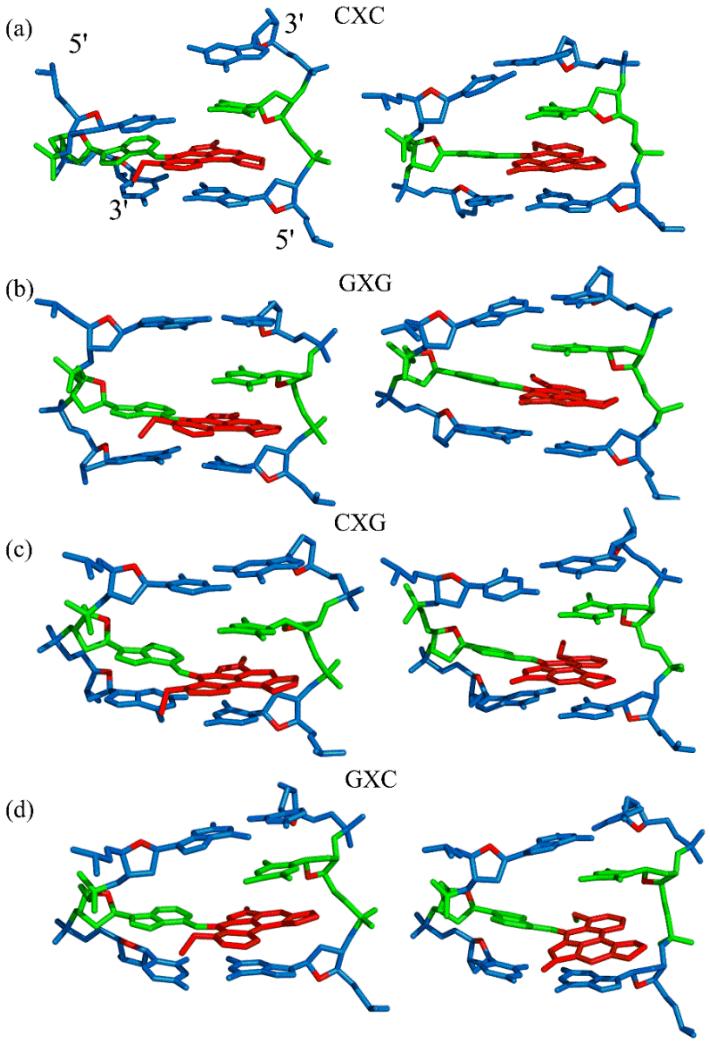


Figure 4.3: Representative structures from MD simulations with the* anti *(left) and* syn *(right) orientation of the ALI-N^6-dA adduct paired opposite thymine in the different sequence context for the 3′–intercalated DNA conformation.

unmodified DNA (Table C4, Appendix C). 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^6-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.

4.3.2 The *anti* Base-displaced Intercalated Orientation is the Most Stable ALI-N^6-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 ~ 25 – 70 kJ mol^{-1} less stable than the base-displaced intercalated orientation of ALI-N^6-dA adducted DNA for all sequence contexts (Table 4.1). This finding correlates with the greatest helical distortion for these adducted DNA conformations (Table C4, Appendix C). Furthermore, these structural distortions in the 3′ or 5′–intercalated conformers decrease the van der Waals (stacking) interactions involving the damaged base by ~ 10 – 45 kJ mol^{-1} compared to the base-displaced intercalated conformers (Table 4.1). As a result, the 5′–intercalated and 3′–intercalated conformers are likely energetically inaccessible to ALI-N^6-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 C3, Appendix C), significant van der Waals interactions between the lesion and the flanking bases

(Table 4.1), lesion pair interactions (primarily C–H•••π, Table C5, Appendix C), and minimal distortions to the DNA step parameters at the lesion site (Table C4, Appendix C). 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 ~ 22 kJ mol^{-1} depending on the sequence, Table 4.1). Therefore, both base-displaced intercalated conformers are likely energetically 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 ~ 8 – 10 kJ mol^{-1} stronger for the *anti* than the *syn* adduct conformation (Table C5, Appendix C), which primarily reflects changes in the C–H•••π contact. Furthermore, the van der Waals interactions between the adduct and the flanking base pairs are (up to ~ 6 kJ mol^{-1}) greater in the *anti* base-displaced intercalated conformers (Table 4.1). As a result, the *anti* base-displaced intercalated adducted DNA conformer is the most stable among the six conformations characterized in the present work regardless of the sequential motif.

4.3.3 Calculated Relative Free Energies of the Most Stable *anti* Base-displaced Intercalated Adducted DNA Conformer Correlates with Experimentally Predicted Strand Stabilities.

A previous experimental study carried out by Lukin *et al*. examined the stability of ALII-N^6-dA adducted DNA in three sequence contexts,[41] 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, which translates into the Gibbs free energy of duplex formation being ~ 7 kJ mol^{-1} more stable for CXC. Furthermore, Lukin *et al.* hypothesized that ALI-N^6-dA and ALII-N^6-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.[41] This proposal was substantiated by computational work comparing the structure of the ALI and ALII-adducted DNA (Chapter 2). Therefore, in the absence of analogous experimental data on the ALI-N^6-dA adduct, it is anticipated that DNA containing either ALI-N^6-dA or ALII-N^6-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, the calculations similarly predict that ALI-N^6-dA adducted DNA is ~ 9 kJ mol^{-1} more stable in the CXC than CXG sequence context (Table 4.2). This direct correlation between the trends in the sequence dependence of the strand stabilities illustrates the robustness of the computational approach.

4.3.4. Identity of the Flanking Bases Alters the Conformational Heterogeneity of ALI-N^6-dA Adducted DNA.

As discussed above, the simulations predict that both the *anti* and *syn* base-displaced intercalated conformers are energetically accessible for ALI-N^6-dA adducted DNA in all studied sequence contexts. However, the energy difference between the *anti* and *syn* base-displaced intercalated conformers varies with the sequence context according to CXC > CXG > GXG > GXC (Table 4.2). As discussed in the

Introduction, changes in the conformational stability with sequence context have been reported for other adducts and correlated to differences in discrete lesion site interactions. For example, deviations in the adducted DNA conformational heterogeneity have been reported for OTB-dG due to changes in discrete hydrogen-bonding interactions at the lesion site,[12] AF-dG due to distinctive stacking patterns,[4,5] and 10S (+)-*trans*-*anti*-B[*a*]P-N^2-dG due to intrastrand steric interactions.[7]

In contrast to other adducts, no obvious structural features explain the differences in the conformational heterogeneity of ALI-N^6-dA adducted DNA in the four sequences. Instead, the sequence dependence of the ALI-N^6-dA induced conformational heterogeneity arises due to a complex interplay between all components of the free energy (Table 4.2). In particular, the major discrepancies in the free energies for the *anti* and *syn* base-displaced intercalated conformations between the four sequence contexts mainly occur in the solvation (ΔG_{sol}), van der Waals (ΔE_{vdw}) and electrostatic repulsion (ΔE_{elec}) components. For example, CXC has the largest ΔG_{sol} among the four sequences, which favors the *anti* conformation, and the largest ΔE_{elec}, which favors the *syn* conformer. These factors effectively cancel to yield the smallest energy difference between the *syn* and *anti* base-displaced intercalated conformations (G_{rel}(*syn*–*anti*) = 2.1 kJ mol^{-1}). Similarly, for the CXG sequence, ΔG_{sol} and ΔE_{elec} are the largest components of the free energy difference, but G_{sol} dominates and the *anti* conformer becomes more stabilized in this sequence (G_{rel}(*syn*–*anti*) = 14.8 kJ mol^{-1}). Although ΔG_{sol} and ΔE_{elec} are mutually compensating for GXG, the difference in the van der Waals component of the free energy is the

largest for this sequence among the four sequence contexts, which stabilizes the *anti* conformer more than seen for CXC and CXG (G_{rel}(*syn*–*anti*) = 16.3 kJ mol^{-1}). Finally, despite similar solvation of the *anti* and *syn* base-displaced intercalated conformers in GXC, this is the only sequence in which the *anti* conformer has less electrostatic repulsion compared to the *syn* orientation, which further increases G_{rel}(*syn*–*anti*) to 21.9 kJ mol^{-1}. Although these examples qualitatively explain the predicted differential conformational heterogeneity, the calculated energy differences arise from the sum of the differences in all free energy components for the *anti* and *syn* conformers in each sequence. Regardless, the present data illustrates that the sequence context influences the relative stability of the *syn* base-displaced intercalated conformer, which is more favoured in the CXC sequence context compared to the CXG, GXG and GXC sequences.

4.3.5 Biological Consequences of the Effects of Sequence Context on the Conformational Preferences of ALI-N^6-dA Adducted DNA.

4.3.5.1 The Relative Stability of the Energetically Accessible Conformations of Adducted DNA does not Explain the Experimentally-Predicted Sequence-Dependent Mutagenicity of ALI-N^6-dA. As discussed in Chapter 1, recent experimental studies on AA associated mutagenesis demonstrate that 'signature' A→T transversion mutations most commonly occur in T/CXG motifs in the genome.[1,2] Since ALI-N^6-dA is the most abundant,[48,49] nephrotoxic and mutagenic[50] 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[37-39] and be related to enhanced stabilization of the mutagenic conformation.[37] 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.[37] Furthermore, previous studies on DNA adducts suggest that, barring occasional overriding by polymerase interactions,[51] the lesion site conformations observed in damaged DNA are often manifested in polymerase active sites.[51,52] Thus, the adducted DNA conformations can provide important information regarding the sequence dependent mutagenicity of the ALI-N^6-dA lesion.

Since the overall composition of the ALI-N^6-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 (ΔG_{total}, Table 4.2). The calculations reveal that the stability of the energetically-accessible base-displaced intercalated conformer varies with sequence context (and adduct glycosidic orientation) according to GXG (*anti*) > GXC (*anti*) > CXC (*anti*) > CXC (*syn*) > CXG (*anti*) > GXG (*syn*) > GXC (*syn*) > CXG (*syn*). Hence, the ALI-N^6-dA adduct is not the most stable in the CXG sequence in either the *anti* or *syn* conformer, but rather falls at least ~ 14 kJ mol^{-1} higher in energy than the most stable GXG (*anti*) sequence. Therefore, the experimentally-observed higher frequency of A→T mutations associated with the CXG sequence is not thermodynamically driven by the stability of the mutagenic ALI-N^6-dA adducted DNA conformation(s). Instead, other factors must dictate the sequence-dependent mutagenicity of AL-N^6-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.[38] Alternatively, a higher repair resistance of the most abundant ALI-N^6-dA adduct in the CXG sequence may play a role in the sequence dependent mutagenicity, which is considered in the next section.

4.3.5.2 Structural Differences in the Energetically Accessible ALI-N^6-dA Adducted DNA Conformers Suggest that the Repair Propensity may depend on the Glycosidic Orientation of the Adducted Nucleoside and the Lesion Site Sequence Context. 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,[53-56] it is generally believed that DNA lesions that can substantially stall RNA polymerases are generally good substrates for TCR.[24,57] However, the precise nature of the TCR signal is still not well understood.[24] 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 increase in helical[9,30-32] and backbone dynamics,[29] decrease in the stacking interactions at the lesion site[33-35] and perturbations to the helical parameters (such as minor groove dimension, rise and twist).[29,33,58-61] 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.[16,62] Furthermore, the bases flanking the lesion have been shown to affect the helical

Table 4.1: Relative MM-PBSA free energies, G_{rel} and the van der Waals stacking energies, E_{vdW} for different conformations of ALI-N^6-dA adducted DNA derived from 20 ns MD simulations (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]
anti 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
anti 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
anti 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
syn 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
syn 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
syn 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

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

Table 4.2: The components of the total free energy (kJ mol^{-1}) for the anti and syn base-displaced intercalated conformations in different sequence contexts.

Conformer		E_{bond}[a]	E_{angle}[b]	$E_{dihedral}$[c]	E_{vdW}[d]	E_{elec}[e]	G_{sol}[f]	$-TS$[g]	G_{total}[h]	ΔG_{total}[i]
	anti	808.2	1727.1	2085.5	−813.0	1257.1	−22535.5	−2464.6	−19935.1	4.7
CXC	***syn***	807.1	1727.5	2054.2	−800.2	1153.8	−22423.0	−2452.4	−19933.0	6.8
	Δ[j]	−1.1	0.4	−31.2	12.8	−103.3	112.4	12.2	2.1	
	anti	807.9	1723.6	2087.0	−830.8	1351.8	−22617.8	−2461.3	−19939.8	0.0
GXG	***syn***	812.2	1732.8	2065.8	−810.4	1331.8	−22596.8	−2458.7	−19923.5	16.3
	Δ[j]	4.3	9.2	−21.3	20.4	−20.0	21.0	2.6	16.3	
	anti	812.4	1720.0	2100.8	−821.6	1364.6	−22641.2	−2461.4	−19926.2	13.6
CXG	***syn***	809.8	1735.5	2069.0	−807.2	1294.9	−22556.1	−2457.4	−19911.4	28.4
	Δ[j]	−2.6	15.5	−31.8	14.4	−69.7	85.1	3.9	14.8	
	anti	807.5	1722.3	2086.0	−814.8	1227.6	−22502.2	−2464.0	−19937.7	2.1
GXC	***syn***	806.4	1736.8	2062.7	−808.7	1247.4	−22500.1	−2460.2	−19915.6	24.2
	Δ[j]	−1.1	14.5	−23.3	6.1	19.8	2.1	3.8	22.1	

[a]Internal energy emerging from deviation of bonds (E_{bond}). [b]Internal energy emerging from deviation of angles (E_{angle}). [c]Internal energy emerging from deviation of dihedral angles ($E_{dihedral}$). [d]van der Waals interaction energy (E_{vdW}). [e]Electrostatic interaction energy (E_{elec}). [f]Solvation free energy (G_{sol}). [g]Entropy term (–TS). [h]Total free energy (G_{total}). [i]Relative total free energies with respect to the conformation with the lowest (most negative) G_{total} regardless of the sequence context (ΔG_{total}). [j]The *syn* – *anti* difference (Δ) for each energy component.

structure at the damaged site, and therefore the sequence context affects the repair propensity of some adducts.[7-10,15-20] On these lines, comparison of the helical perturbations caused by the lesion in the energetically-accessible (base-displaced intercalated) adducted ALI-N^6-dA DNA conformers are critically compared for each sequence below. 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,[53-56] it is generally believed that DNA lesions that can substantially stall RNA polymerases are generally good substrates for TCR.[24, 57] However, the precise nature of the TCR signal is still not well understood.[24] 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[9, 30-32] and backbone dynamics,[29] decreases in the stacking interactions at the lesion site[33-35] and perturbations to the helical parameters (such as minor groove dimension, rise and twist).[29, 33, 58-61] 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.[16, 62] 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.[7-10, 15-20] On these lines, comparison of the helical perturbations caused by the lesion in the energetically-accessible (base-displaced

intercalated) adducted ALI-N^6-dA DNA conformers are critically compared for each sequence below.

Disparities in van der Waals Stacking and Helical Distortions at the Lesion Site Point Towards A Greater Repair Propensity of the syn *than the Corresponding* anti *Adduct Conformation.* For both the *anti* and *syn* base-displaced intercalated conformers, there is little change in the dynamics at the lesion site with respect to the corresponding unmodified DNA regardless of the sequence, with standard deviations being < 0.5 Å for translational step parameters and < 5˚ for rotational step parameters (Table C4, Appendix C). This suggests that the lesion site dynamics likely does not dictate the GGR repair propensity of ALI-N^6-dA. In contrast, the van der Waals (stacking) interaction between the adduct and the flanking base pairs is less stable for the *syn* base-displaced intercalated ALI-N^6-dA adducted DNA conformer than for the *anti* analogue in all sequences (Table 4.1). Furthermore, the *syn* base-displaced intercalated conformer is more distorted than the corresponding *anti* base-displaced intercalated conformer in GXG, GXC and CXG (Figure 4.4 and Table C4, Appendix C). This is primarily seen in the minor groove width and helix untwisting, where the minor groove of the *syn* base-displaced intercalated conformers widens by an average of 1 – 2 Å and the helix untwists by ~ 13 – 14˚ with respect to undamaged DNA, while the corresponding parameters for the *anti* conformers change by 0.4 – 0.8 Å and ~ 6 – 9˚, respectively. Interestingly, greater distortions and less stabilizing stacking at the

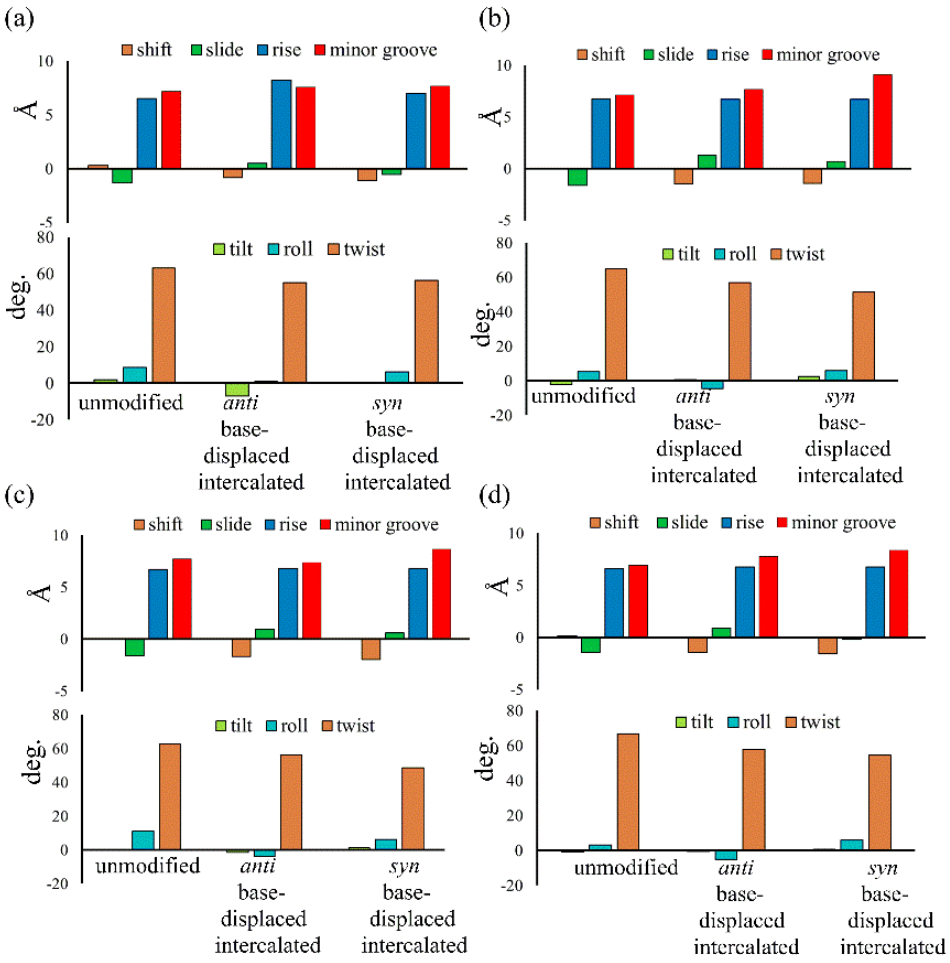


Figure 4.4: The pseudostep parameters and minor groove dimensions for the* anti *and* syn *base-displaced intercalated confomers of ALI-N^6-dA adducted DNA in the (a) CXC, (b) GXG, (c) CXG and (d) GXC sequence contexts, as well as the values for the corresponding unmodified strand.

lesion site in the *syn* base-displaced intercalated conformer of AAF-dG and AF-dG adducted DNA compared to the major and minor groove conformers has been proposed to result in a greater GGR repair propensity for the base-displaced intercalated conformer.[16] By analogy, although the differences with respect to the ALI-N^6-dA nucleoside orientation are small, the decreased stacking and increased distortion may lead to a smaller ALI-N^6-dA GGR repair resistance in the *syn* over the *anti* base-displaced intercalated conformers in the GXG, GXC and CXG sequential motifs.

In contrast to the other sequence contexts, although the helix untwists by ~ 7° for the *anti* and ~ 12° for *syn* base-displaced conformations, the *syn* base-displaced intercalated conformer has similar major groove widening as the corresponding *anti* orientation in CXC relative to canonical DNA (~ 0.4 – 0.5 Å, Figure 4.4 and Table C4, Appendix C). Thus, the less stable van der Waals stacking energy in the *syn* compared to the *anti* base-displaced intercalated conformation (Table 4.2) might facilitate the smaller lesion GGR resistance in the CXC sequence context, while the smallerdistortion in the *syn* conformer in CXC compared to other sequences points towards greater lesion GGR resistance. Since it is currently unclear which effect predominates during the GGR recognition step, it is difficult to definitively conclude based on the present data whether the *syn* base-displaced conformer will persist in the CXC sequence in cells. Indeed, biochemical experiments probing the interactions between ALI-N^6-dA adducted DNA and GGR enzymes are required to confirm the repair resistance of the adduct in the CXC sequence context. Furthermore, experiments probing the interactions between ALI-N^6-dA adducted DNA and RNA polymerases are required to understand the sequence dependence of the TCR repair propensity of the ALI-N^6-dA adduct in DNA. In the absence of this information, it can be proposed that the *anti* base-displaced intercalated conformation will be the most GGR repair-resistant ALI-N^6-dA adducted DNA conformer in all the studied sequence contexts, when the lesion is present primarily in nontranscribed regions of the genome in cells. Additionally, depending on the *anti-syn* energy barrier, an *anti-syn* conformational exchange may occur, due to which the *syn* base-displaced intercalated conformer may also persist in the CXC sequential motif.

Sequence-Dependent Distortions in the Most Persistent Adducted DNA Conformers Point Towards a Possible Sequence-Dependent Repair Propensity of ALI-N^6-dA. 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^2-dG PAH adducted DNA,[9, 15] the lesion site dynamics remain unaltered in all base-displaced intercalated conformers of ALI-N^6-dA adducted DNA compared to the corresponding unmodified DNA regardless of the sequence (Table C4, Appendix C). In addition, although a decrease in the lesion van der Waals (stacking) energy when AF-dG is at G^2 compared to G^1 or G^3 in the *Nar*I sequence explains the observed higher adduct repair propensity at G^2,[16] the lesion van der Waals interactions in the most persistent *anti* base-displaced intercalated conformer of ALI-N^6-dA adducted DNA is only 0.5 kJ mol^{-1} more stabilizing for the most mutagenic CXG sequence compared to the non-mutagenic GXG sequence (Table 4.1). Hence, neither dynamics nor the lesion site van der Waals energy likely afford a sequence-dependent GGR propensity for ALI-N^6-dA. Nevertheless, sequence-dependent distortions have also been shown to affect the GGR propensity of DNA adducts, including AAF-dG,[16] 10S (+)-*trans*-*anti*-B[*a*]P-dA[8] and 10S (+)-*trans*-*anti*-B[*a*]P-dG.[15] In the case of ALI-N^6-dA adducted DNA, the magnitude of the distortion at the lesion site also varies with the sequence (Figure 4.4). Specifically, the most persistent *anti* base-displaced intercalated conformer in the most mutagenic CXG sequence[1, 2] 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 4.4 and Table C4, Appendix C). 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 C4, Appendix C), 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^6-dA adduct and additional biochemical experiments are required to further explore the toxicity of this important lesion.

4.4 Conclusions

The present computational study provides structural details of the possible conformers of ALI-N^6-dA adducted DNA in different sequence contexts (namely, GXC, GXG, CXG and CXC). The calculations reveal that the conformational outcomes of ALI-N^6-dA adducted DNA are independent of the identity of the flanking bases. Specifically, the base-displaced intercalated conformers with either the *anti* or *syn* adduct orientations are energetically accessible in all sequences considered. However, the degree of adducted DNA conformational heterogeneity depends on the sequence context, with the CXC sequence having the smallest calculated free energy difference between the *anti* and *syn* base-displaced intercalated conformers. Analysis of the energetically accessible conformations indicates 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 glycosidic orientation of the adducted nucleoside (with the exception of the CXC sequence) and possibly the lesion site sequence context. Most importantly, ALI-N^6-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.

In this chapter, a conformational analysis of DNA containing the AL-N^6-dA lesion is carried out to provide a possible structural reason for the experimentally-observed differential mutagenicity of AL-adenine adducts in different sequences. However, additional experimental and molecular modelling studies are required to unveil the exact mechanism of the sequence-dependent mutagenicity of ALI-N^6-dA. Further future directions in the field of aristolochic acids toxicity and the conclusions from each chapter of the thesis are highlighted in the next chapter.

4.5 References

(1) Hoang, M. L.; Chen, C. H.; Sidorenko, V. S.; He, J.; Dickman, K. G.; Yun, B. H.; Moriya, M.; Niknafs, N.; Douville, C.; Karchin, R.; Turesky, R. J.; Pu, Y. S.; Vogelstein, B.; Papadopoulos, N.; Grollman, A. P.; Kinzler, K. W.; Rosenquist, T. A. *Sci. Transl. Med.* **2013**, *5*, 197ra102.

(2) Poon, S. L.; Pang, S. T.; McPherson, J. R.; Yu, W.; Huang, K. K.; Guan, P.; Weng, W. H.; Siew, E. Y.; Liu, Y.; Heng, H. L.; Chong, S. C.; Gan, A.; Tay, S. T.; Lim, W. K.; Cutcutache, I.; Huang, D.; Ler, L. D.; Nairismägi, M. L.; Lee, M. H.; Chang, Y. H.; Yu, K. J.; Chan-on, W.; Li, B. K.; Yuan, Y. F.; Qian, C. N.; Ng, K. F.; Wu, C. F.; Hsu, C. L.; Bunte, R. M.; Stratton, M. R.; Futreal, P. A.; Sung, W. K.; Chuang, C. K.; Ong, C. K.; Rozen, S. G.; Tan, P.; Teh, B. T. *Sci. Transl. Med.* **2013**, *5*, 197ra101.

(3) Wang, F.; Elmquist, C. E.; Stover, J. S.; Rizzo, C. J.; Stone, M. P. *Biochemistry* **2007**, *46*, 8498.

(4) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 95.

(5) Mao, B.; Hingerty, B. E.; Broyde, S.; Patel, D. J. *Biochemistry* **1998**, *37*, 81.

(6) Patnaik, S.; Cho, B. P. *Chem. Res. Toxicol.* **2010**, *23*, 1650.

(7) Kropachev, K.; Kolbanovskii, M.; Cai, Y.; Rodríguez, F.; Kolbanovskii, A.; Liu, Y.; Zhang, L.; Amin, S.; Patel, D.; Broyde, S.; Geacintov, N. E. *J. Mol. Biol.* **2009**, *386*, 1193.

(8) Yan, S.; Wu, M.; Buterin, T.; Naegeli, H.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2003**, *42*, 2339.

(9) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(10) Rodríguez, F. A.; Cai, Y.; Lin, C.; Tang, Y.; Kolbanovskiy, A.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E. *Nucleic Acids Res.* **2007**, *35*, 1555.

(11) Zegar, I. S.; Chary, P.; Jabil, R. J.; Tamura, P. J.; Johansen, T. N.; Lloyd, R. S.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Biochemistry* **1998**, *37*, 16516.

(12) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(13) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2013**, *26*, 803.

(14) Fuchs, R. P. P.; Schwartz, N.; Daune, M. P. *Nature* **1981**, *294*, 657.

(15) Cai, Y.; Patel, D. J.; Broyde, S.; Geacintov, N. E. *J. Nucleic Acids* **2010**, *2010*, 9.

(16) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2012**, *40*, 9675.

(17) Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P. *Nucleic Acids Res.* **2012**, *40*, 3939.

(18) Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2012**, *51*, 1486.

(19) Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D.; Cho, B. P. *Nucleic Acids Res.* **2013**, *41*, 869.

(20) Meneni, S.; Shell, S. M.; Zou, Y.; Cho, B. P. *Chem. Res. Toxicol.* **2007**, *20*, 6.

(21) Rubbi, C. P.; Milner, J. *Carcinogenesis* **2001**, *22*, 1789.

(22) de Laat, W. L.; Jaspers, N. G. J.; Hoeijmakers, J. H. J. *Genes Dev.* **1999**, *13*, 768.

(23) Gillet, L. C. J.; Schärer, O. D. *Chem. Rev.* **2005**, *106*, 253.

(24) Hanawalt, P. C.; Spivak, G. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 958.

(25) Kropachev, K.; Kolbanovskiy, M.; Liu, Z.; Cai, Y.; Zhang, L.; Schwaid, A. G.; Kolbanovskiy, A.; Ding, S.; Amin, S.; Broyde, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2013**, *26*, 783.

(26) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2009**, *385*, 30.
(27) Hang, B. *J. Nucleic Acids* **2010**, *2010*, 29.

(28) Sidorenko, V. S.; Yeo, J. E.; Bonala, R. R.; Johnson, F.; Schärer, O. D.; Grollman, A. P. *Nucleic Acids Res.* **2012**, *40*, 2494.

(29) Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgilialoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V. *Nucleic Acids Res.* **2014**, *42*, 5020.

(30) Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H. *Trends Biochem. Sci.* **2007**, *32*, 494.

(31) Blagoev, K. B.; Alexandrov, B. S.; Goodwin, E. H.; Bishop, A. R. *DNA Repair* **2006**, *5*, 863.

(32) Isaacs, R. J.; Spielmann, H. P. *DNA Repair* **2004**, *3*, 455.

(33) Chen, X.; Velmurugu, Y.; Zheng, G.; Park, B.; Shim, Y.; Kim, Y.; Liu, L.; Van Houten, B.; He, C.; Ansari, A.; Min, J.-H. *Nat. Commun.* **2015**, *6*, 10.1038/ncomms6849.

(34) Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S. *Nucleic Acids Res.* **2011**, *39*, 8752.

(35) Yang, W. *DNA Repair* **2006**, *5*, 654.

(36) Janićijević, A.; Sugasawa, K.; Shimizu, Y.; Hanaoka, F.; Wijgers, N.; Djurica, M.; Hoeijmakers, J. H. J.; Wyman, C. *DNA Repair* **2003**, *2*, 325.

(37) Vaidyanathan, V. G.; Cho, B. P. *Biochemistry* **2012**, *51*, 1983.

(38) Xu, P.; Oum, L.; Lee, Y.-C.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2009**, *48*, 4677.

(39) Shibutani, S.; Fernandes, A.; Suzuki, N.; Zhou, L.; Johnson, F.; Grollman, A. P. *J. Biol. Chem.* **1999**, *274*, 27433.

(40) Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P. *Biochemistry* **2014**, *53*, 4059.

(41) Lukin, M.; Zaliznyak, T.; Johnson, F.; de los Santos, C. *Nucleic Acids Res.* **2012**, *40*, 2759.

(42) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; Version 1.0 ed.; University of California: San Francisco, **2008**.

(43) Case, D. A.; Darden, T. A.; T.E. Cheatham, III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; *AMBER 11* University of California, San Francisco, CA: **2010**.

(44) Roy, D. T., Keith; John, Millam *GaussView Version 5* **2009**.

(45) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(46) Case, D. A.; Darden, T. A.; T.E. Cheatham III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.;

Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; *AMBER 12* University of California, San Francisco, CA: **2012**.

(47) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314.

(48) Bieler, C. A.; Stiborova, M.; Wiessler, M.; Cosyns, J. P.; van Ypersele de Strihou, C.; Schmeiser, H. H. *Carcinogenesis* **1997**, *18*, 1063.

(49) Pfau, W.; Schmeiser, H. H.; Wiessler, M. *Carcinogenesis* **1990**, *11*, 1627.

(50) Shibutani, S.; Dong, H.; Suzuki, N.; Ueda, S.; Miller, F.; Grollman, A. P. *Drug Metab. Dispos.* **2007**, *35*, 1217.

(51) Broyde, S.; Wang, L.; Zhang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J. *Chem. Res. Toxicol.* **2007**, *21*, 45.

(52) Sproviero, M.; Verwey, A. M. R.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 13405.

(53) Mei Kwei, J. S.; Kuraoka, I.; Horibata, K.; Ubukata, M.; Kobatake, E.; Iwai, S.; Handa, H.; Tanaka, K. *Biochemical and Biophysical Research Communications* **2004**, *320*, 1133.

(54) Dimitri, A.; Goodenough, A. K.; Guengerich, F. P.; Broyde, S.; Scicchitano, D. A. *J. Mol. Biol.* **2008**, *375*, 353.

(55) Damsma, G. E.; Alt, A.; Brueckner, F.; Carell, T.; Cramer, P. *Nat Struct Mol Biol* **2007**, *14*, 1127.

(56) Brueckner, F.; Hennecke, U.; Carell, T.; Cramer, P. *Science* **2007**, *315*, 859.

(57) Vermeulen, W.; Fousteri, M. *Cold Spring Harbor Perspectives in Biology* **2013**, *5*.

(58) Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J. *Biopolymers* **2002**, *65*, 202.

(59) Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2010**, *17*, 379.

(60) Schärer, O. D. *Mol. Cell.* **2007**, *28*, 184.

(61) Sugasawa, K.; Okamoto, T.; Shimizu, Y.; Masutani, C.; Iwai, S.; Hanaoka, F. *Genes Dev.* **2001**, *15*, 507.

(62) Jankowiak, R.; Ariese, F.; Hewer, A.; Luch, A.; Zamzow, D.; Hughes, N. C.; Phillips, D.; Seidel, A.; Platt, K.-L.; Oesch, F. *Chem. Res. Toxicol.* **1998**, *11*, 674.

Chapter 5

Thesis Summary and Future Work

5.1 Thesis Summary and Conclusions

In the present thesis, a thorough computational study was undertaken to analyze the damage to DNA caused by exposure to aristolochic acids. Specifically, long standing questions in the field regarding the mutagenicity and repair of damage caused by AAs was considered from a structural perspective. A comprehensive study of the purine adducts formed by aristolochic acids was carried out by systematically increasing the computational model size from a nucleobase to nucleoside to nucleotide, and finally incorporating the adducts into an 11-mer DNA oligonucleotide. This approach allowed a detailed analysis on the effect of each component of the DNA strand on the preferred adduct conformation. At various stages of analysis, the results were compared to experimental data, which proved the accuracy and robustness of the computational approach.

The first chapter of this thesis highlighted the fact that it is crucial to understand the structural changes induced by the lesion within the adducted DNA strand (i.e., the conformational outcome of the adducted DNA) in order to understand the mutagenicity and repair propensity of a particular lesion. A review of the literature using select examples was done to understand how different factors can bias the conformational outcome of adducted DNA and can lead to different biological consequences. This discussion was followed by an introduction of the occurrence and sources of cellular exposure to aristolochic acids, various experiments that have been

done so far to analyze the activation of these carcinogens, and the repair and mutagenicity of the adducts formed.

The second chapter of this thesis focused on understanding the structural basis of the nephrotoxicity induced by AAs through studying the ALI-N^6-dA and ALII-N^6-dA lesions. DFT calculations indicated that the aristolactam moiety intrinsically prefers a planar conformation with respect to adenine. Calculations on the nucleoside and nucleotide adducts suggested that the *anti* and *syn* orientations about the glycosidic bond are isoenergetic for both adducts. MD simulations and free energy calculations revealed that the *anti* base-displaced intercalated conformation is the most stable conformer for both types of AL-N^6-dA adducted DNA, which agreed with previous experimental work on the ALII-N^6-dA adduct and thereby validated the computational approach. Interestingly, this conformer differs from the dominant conformations adopted by other N6-linked adenine lesions, including those derived from polycyclic aromatic hydrocarbons. Furthermore, the second most stable *syn* base-displaced intercalated conformation is close in energy to the *anti* base-displaced intercalated conformation for ALI-N^6-dA compared to ALII-N^6-dA. This indicates that a mixture of conformations may be detectable for ALI-N^6-dA in DNA. If this enhanced conformational flexibility persists when bound adducted DNA is bound to a lesion-bypass polymerase, this provided a possible structural explanation for the previously observed greater nephrotoxic potential for the ALI over ALII-N^6-dA adduct. In addition, the structural characteristics of the preferred conformations of adducted DNA explained the resistance of these adducts to repair and thereby added to the current understanding of the toxicity of AAs within living cells.

In the third chapter of the present thesis, molecular modeling was used to determine the structural characteristics and conformational preferences of ALII-N^2-dG adducted DNA. DFT calculations revealed an intrinsic twist at the linkage between the damaged G and ALII moieties. This twist at the carcinogen-purine linkage leads to greater conformational heterogeneity of ALII-N^2-dG than the ALII-N^6-dA adducted DNA studied in Chapter 2. Comparison of the structural features of ALII-N^2-dG damaged DNA and ALII-N^6-dA adducted DNA provided a structural explanation for the observed differential GGR recognition, and thereby repair propensities of these ALII adducts. As a result, this chapter highlighted the critical role of the conformation at the carcinogen-DNA linkage in determining GGR recognition and excision propensities, which significantly contributes to the literature on DNA adducts by expanding the list of previously established effects (e.g., carcinogen-base linkage,[1] stereochemistry,[2-4] adduct ionization state,[5] sequence context,[1,6] and the identity of the partner base[7]) on GGR recognition of DNA adducts.

In Chapter 4, MD simulations were used to model ALI-N^6-dA in different adducted DNA sequences. Specifically, the adduct was considered in four 11-mer DNA sequences, two of the oligonucleotides place a purine and pyrimidine on either the 5′ or 3′–side of the lesion (CXG and GXC), while a purine (GXG) or a pyrimidine (CXC) flanks both the 5′ and 3′–sides of the lesion in the other two oligonucleotides. This study revealed that the conformational preference of ALI-N^6-dA adducted DNA are independent of the identity of the flanking bases, where the base-displaced intercalated conformer in both the *syn* and *anti* adduct orientations is energetically accessible regardless of the sequence considered. However, the energetic separation

between two conformers depends on the sequence context. Regardless, the experimentally-observed greater mutagenicity of the CXG sequence context does not correlate with the calculated thermodynamic stability in different sequences. Instead, AL-N^6-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^6-dA.

Despite many studies in the field of AA-associated carcinogenesis, including the present thesis, there are still many aspects of the toxicity of these carcinogens that remain a mystery. A few such studies and possible outcomes are highlighted in the following section.

5.2 Future Work

AA adducts have been shown to cause A→T transversion mutations, which arise due to the incorporation of a dAMP nucleotide against the lesion upon DNA replication.[8] However, the molecular basis for the preferential incorporation of dAMP over dGTMP, dTMP and dCMP remains unexplored. In a previous study, molecular modelling has been successfully used to explain the experimentally observed preferential incorporation of a specific nucleobase opposite a DNA adduct.[9] MD study of the AL adenine adduct incorporated in DNA and paired opposite each of the natural DNA nucleobases, as well as post processing free energy calculations can provide

valuable insights into the preferential pairing of the lesion with dAMP over the other nucleotides.

Furthermore, as discussed in Chapter 1, bulky DNA adducts are mainly replicated by two types of polymerases, either the replicative or Y-family polymerases, which differ in their active site structure and hence can lead to different mutagenic outcomes for a particular lesion.[10,11] Specifically, replication by the Y-family polymerases (translesion synthesis, TLS) is more erroneous. Hence, in order to understand the mutagenicity of a bulky lesion, it is important to analyze the conformational preferences of the adduct in the TLS polymerase active site. The experimental translesion bypass synthesis of the aristolochic acid adducts shows that the adenine lesion misincorporates dAMP with a 22% frequency, while the guanine lesion misincorporates dAMP with a 9% frequency.[8] Previous studies of known bulky DNA adducts have shown that MD simulations can be successfully used to model an adducted DNA template bound within the polymerase active sites to determine which nucleobase is preferentially incorporated opposite the lesion during replication.[12-15] Therefore, modelling with the adducted DNA bound to the polymerases may help determine the mechanism for the formation of the 'signature' A→T transversion mutations associated with AA toxicity. Furthermore, this would help verify the hypotheses put forward in the present thesis (Chapter 2), that the higher mutagenicity of the ALI-N^6-dA adducts caused by the greater conformational flexibility of the associated adducted dsDNA compared to ALII-N^6-dA.

The mutagenicity of a DNA adduct is directly related to its persistence in the cell, which is in turn dependent on the repair propensity of the adduct. In Chapter 3

of this thesis, it is predicted that the AL-guanine adduct will be better recognized by XPC-RAD23B (the primary recognition factor of the NER pathway) based on the structural perturbations in six different conformers of adducted DNA. Additionally, In Chapter 4, the sequence-dependent structural perturbations of the adducted DNA point towards its differential repair propensity for the ALI-N^6-dA lesion in different sequence context. Although the AL-adenine and guanine adducts have been experimentally-observed to have differential repair propensity,[16] no experimental repair studies have been done for the ALI-N^6-dA adducts in different sequences. Hence, experiments with TCR and GGR enzymes would help verify the hypothesis of sequence-dependent repair propensity of the ALI-N^6-dA lesion in different sequence context.

Furthermore, the crystal structure of XPC-RAD23B also reveals that the bases opposing the lesion (thymine dimer) are flipped out in the active site of the protein during the recognition step.[17] Previous studies of adducted DNA containing the PAH lesion have modelled adducted DNA in the active site of XPC-RAD23B and calculated the barrier for the base flipping, which explains the experimentally-observed repair propensity of the lesion.[18,19]. However, no such studies have been done for the AL-adducts to date. Hence, modelling adducted DNA in the active site of XPC-RAD23B can provide important insights into the relative repair propensity of the adducts formed and supplement the present work.

Finally, although the various thermodynamically possible conformations for AL-adducted DNA were studied in the present thesis (Chapter 2, 3 and 4), further computational work is required to analyze the barriers for interconversion between

the different possible conformations. Previous studies on other conformationally heterogenetic adducts, such as AF-dG and AAG-dG have used potential mean force calculations to estimate the barrier for interconversion from one conformer to the other.[20] Therefore, a similar methodology can be applied to estimate the conversion barrier about the *anti/syn* glycosidic bond within the double-stranded DNA, which will provide further insight into the kinetic feasibility of the formation of different adducted DNA conformations and thereby serve as an important tool for deducing conformational outcomes.

5.3 Contributions of the Present Work

AL-adduct formation constitutes the principal aspect of the genotoxicity and resulting carcinogenicity associated with aristolochic acids. However, the molecular basis of the carcinogenesis, as well as the repair resistance of these lesions, is not well understood. This thesis undertook an extensive molecular modeling study, including quantum mechanics, MD simulations and post-processing free energy calculations, to determine the conformational preferences of AL-adducted DNA. Furthermore, this study was used to explain the experimentally-observed mutagenicity and repair of these lesions. Specifically, the toxicity of AAI versus AAII, the repair of AL-adenine versus guanine adducts, and the sequence dependent mutagenicity of the AL-adenine lesions were analyzed. Overall, this body of work provides a structural basis that answers persistent questions related to the toxicity of these harmful carcinogens.

5.4 References

(1) Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S. *Nucleic Acids Res.* **2012**, *40*, 9675.

(2) Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2012**, *51*, 1486.

(3) Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgilialoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V. *Nucleic Acids Res.* **2014**, *42*, 5020.

(4) Kropachev, K.; Kolbanovskiy, M.; Liu, Z.; Cai, Y.; Zhang, L.; Schwaid, A. G.; Kolbanovskiy, A.; Ding, S.; Amin, S.; Broyde, S.; Geacintov, N. E. *Chem. Res. Toxicol.* **2013**, *26*, 783.

(5) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(6) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(7) Mu, H.; Kropachev, K.; Chen, Y.; Zhang, H.; Cai, Y.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2013**, *52*, 5517.

(8) Attaluri, S.; Bonala, R. R.; Yang, I. Y.; Lukin, M. A.; Wen, Y.; Grollman, A. P.; Moriya, M.; Iden, C. R.; Johnson, F. *Nucleic Acids Res.* **2010**, *38*, 339.

(9) Wilson, K. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2014**, *27*, 1310.

(10) Wang, L.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785.

(11) Dutta, S.; Li, Y.; Johnson, D.; Dzantiev, L.; Richardson, C. C.; Romano, L. J.; Ellenberger, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16186.

(12) Perlow, R. A.; Broyde, S. *J. Mol. Biol.* **2002**, *322*, 291.

(13) Zhang, L.; Shapiro, R.; Broyde, S. *Chem. Res. Toxicol.* **2005**, *18*, 1347.

(14) Wang, L. H.; Broyde, S. *Nucleic Acids Res.* **2006**, *34*, 785.

(15) Benitez, B. A. S.; Arora, K.; Balistreri, L.; Schlick, T. *J. Mol. Biol.* **2008**, *384*, 1086.

(16) Sidorenko, V. S.; Yeo, J. E.; Bonala, R. R.; Johnson, F.; Schärer, O. D.; Grollman, A. P. *Nucleic Acids Res.* **2012**, *40*, 2494.

(17) Min, J. H.; Pavletich, N. P. *Nature* **2007**, *449*, 570.

(18) Cai, Y.; Zheng, H.; Ding, S.; Kropachev, K.; Schwaid, A. G.; Tang, Y.; Mu, H.; Wang, S.; Geacintov, N. E.; Zhang, Y.; Broyde, S. *Chem. Res. Toxicol.* **2013**, *26*, 1115.

(19) Zheng, H.; Cai, Y.; Ding, S.; Tang, Y.; Kropachev, K.; Zhou, Y.; Wang, L.; Wang, S.; Geacintov, N. E.; Zhang, Y.; Broyde, S. *Chem. Res. Toxicol.* **2010**, *23*, 1868.

(20) Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D.; Cho, B. P. *Nucleic Acids Res.* **2013**, *41*, 869.

APPENDIX A

Conformational Preferences of ALI and ALII-N^6-dA Adducted DNA

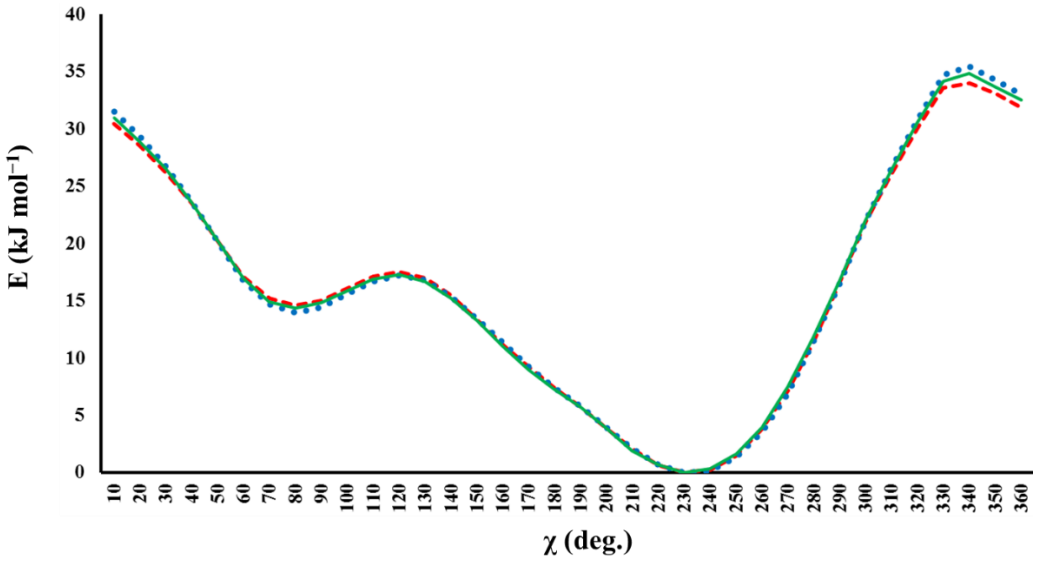


Figure A1: Comparison of the χ potential energy curves of natural adenine (blue), ALI-N⁶-dA (green) and ALII-N⁶-dA (red).

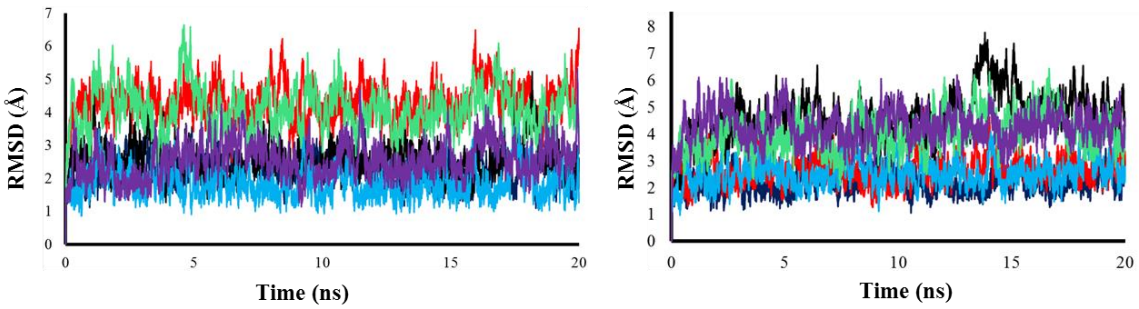


Figure A2: Backbone RMSD versus time for all (a) ALI-N^6-dA and (b) ALII-N^6-dA in all adducted DNA conformations, including the* anti *base-displaced (light blue), 5′–intercalated (green), 3′–intercalated (purple),* syn *base-displaced (dark blue), 5′–intercalated (red), and 3′–intercalated (black) orientations.

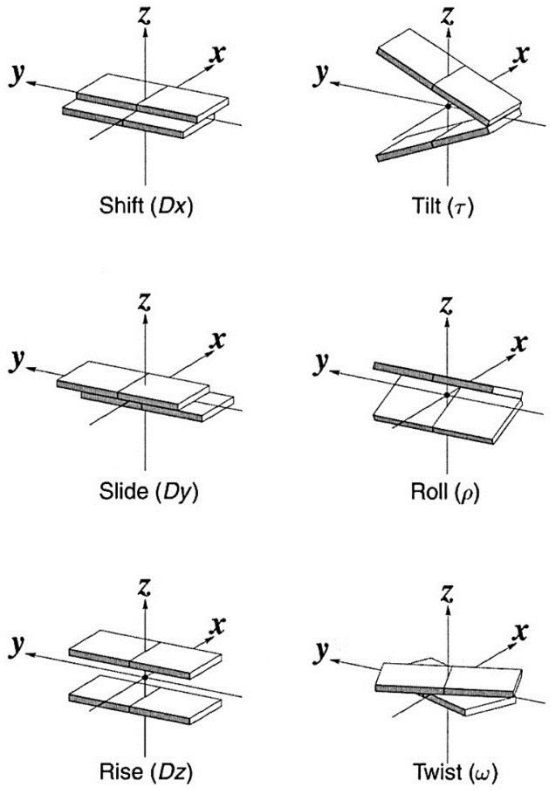


Figure A3: Pictorial definitions of basestep parameters traditionally used to describe the geometry of sequential base pair steps (adapted with permission from Lu, Xiang-Jun Olson, Wilma K. Nucleic Acids Res. 2003, 31, 5108)

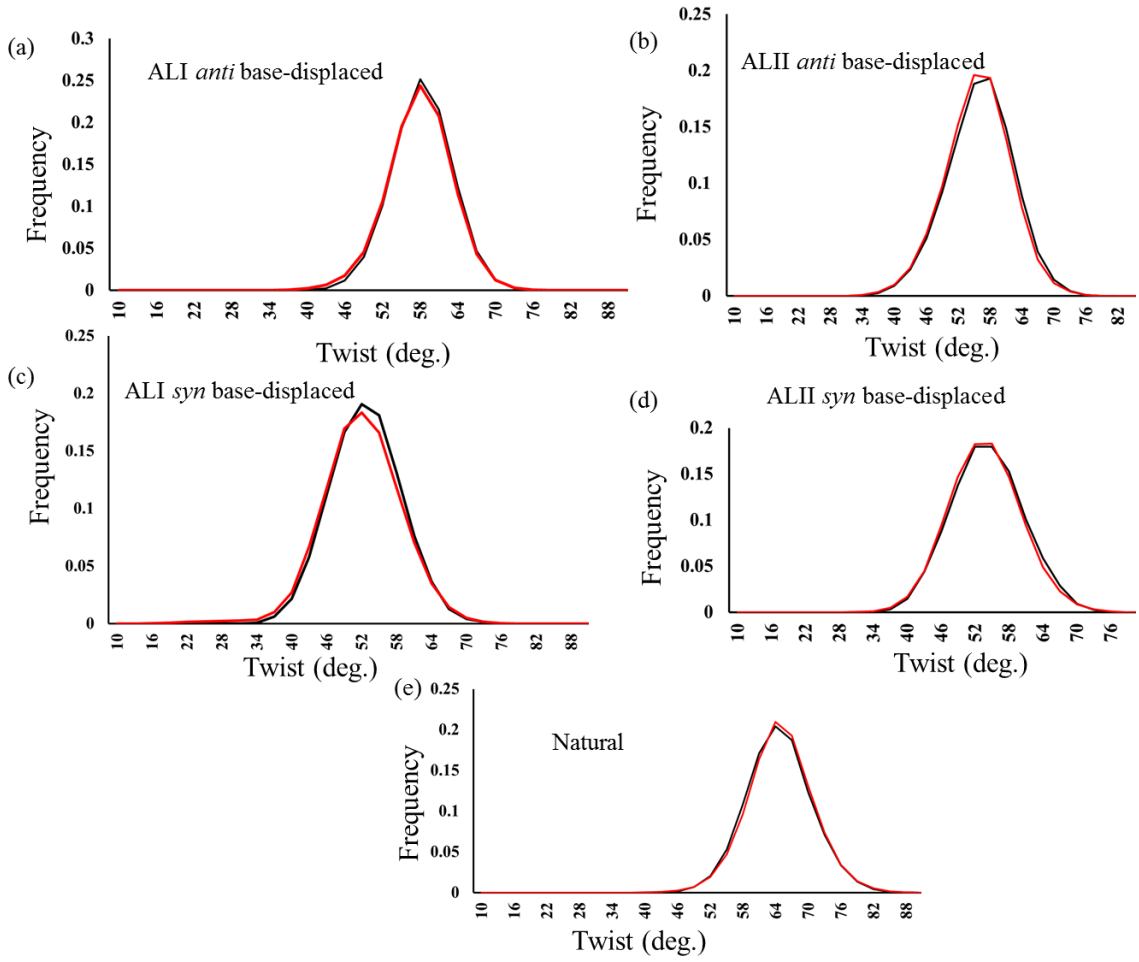


Figure A4: Comparison of the distribution of the twist angle for the psuedostep obtained from 20 ns (black) and 320 ns (red) simulations.

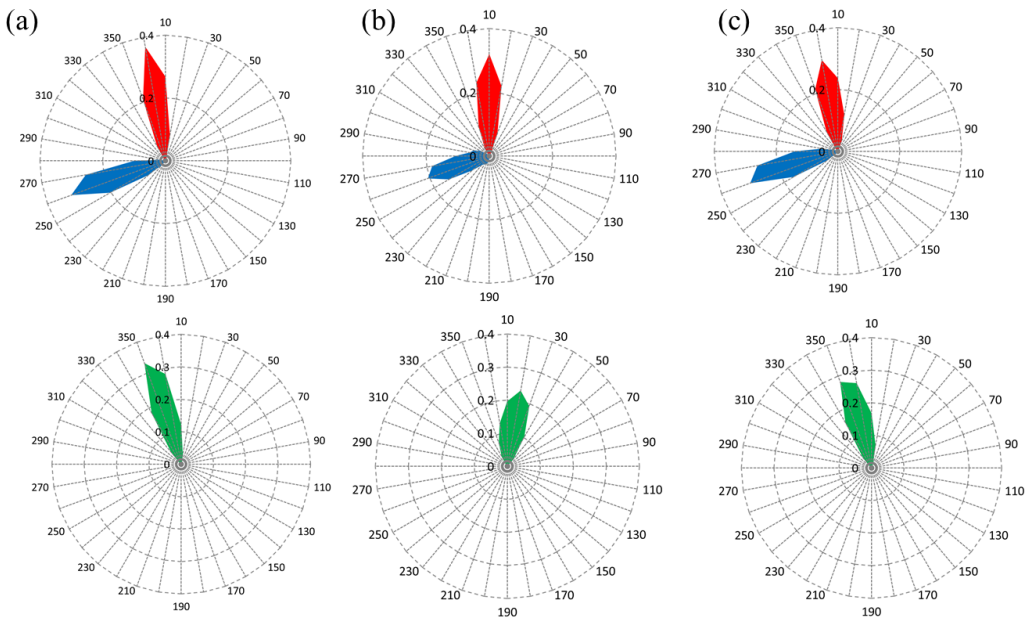


Figure A5: Radar plots for the probability distribution of χ (blue), θ (red) and ϕ (green) for the* anti *ALI-N^6-dA adducted DNA (a) base-displaced, (b) 5′–intercalated and (c) 3′–intercalated conformations.

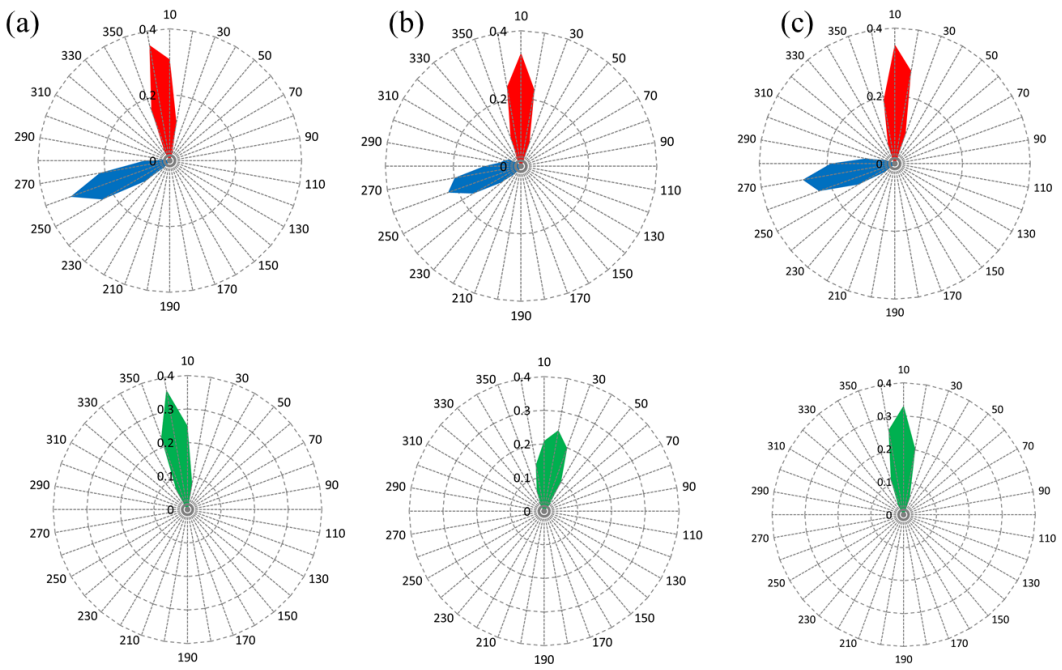


Figure A6: Radar plots for the probability distribution of χ (blue), θ (red) and ϕ (green) for the* anti *ALII-N⁶-dA adducted DNA (a) base-displaced, (b) 5′–intercalated and (c) 3′–intercalated conformations.

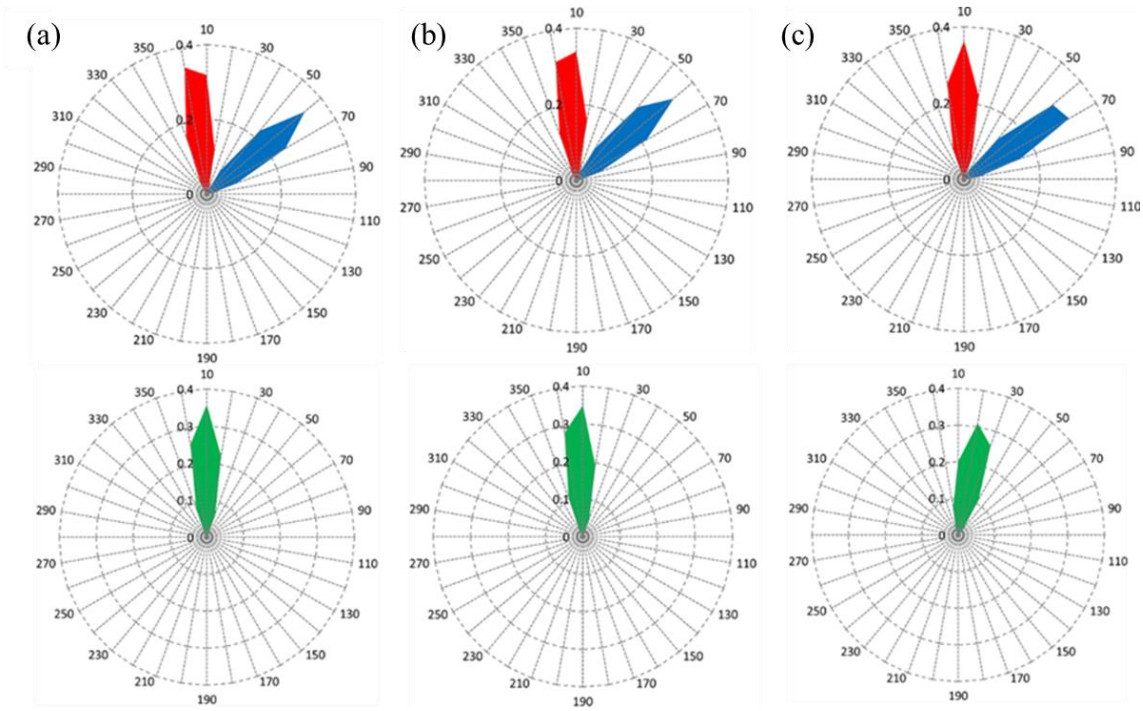


Figure A7: Radar plots for the probability distribution of χ (blue), θ (red) and ϕ (green) for the* syn *ALI-N⁶-dA adducted DNA (a) base-displaced, (b) 5′–intercalated and (c) 3′–intercalated conformations.

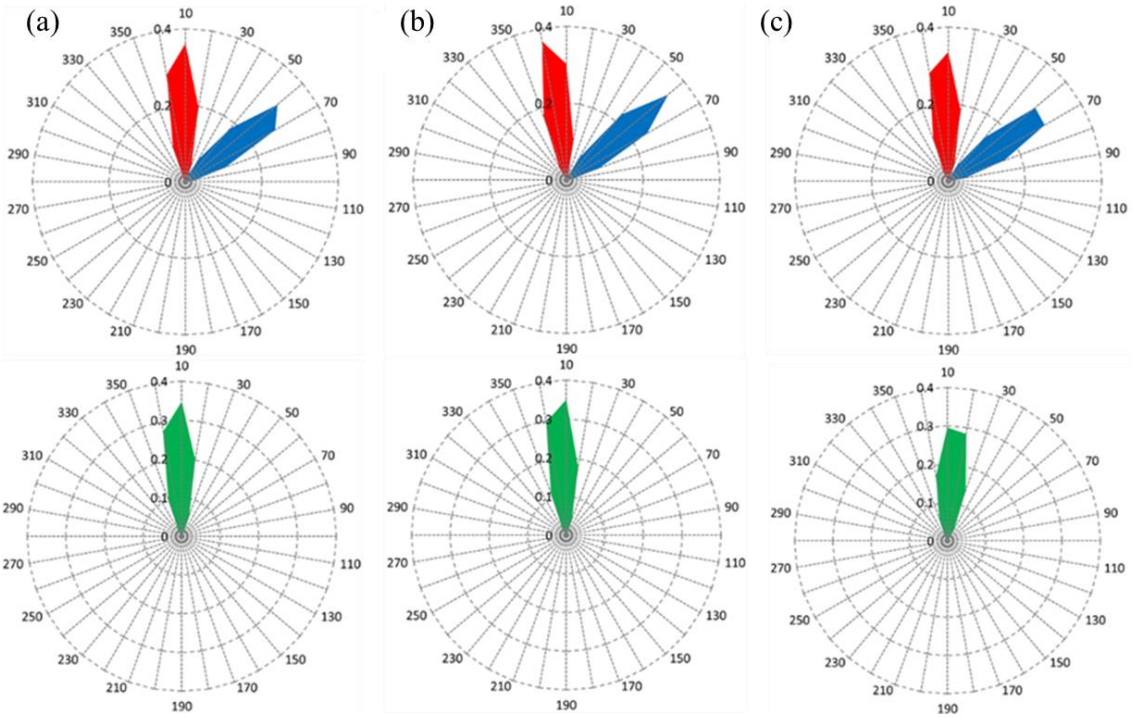


Figure A8: Radar plots for the probability distribution of χ (blue), θ (red) and ϕ (green) for the* syn *ALII-N⁶-dA adducted DNA (a) base-displaced, (b) 5′–intercalated and (c) 3′–intercalated conformations.

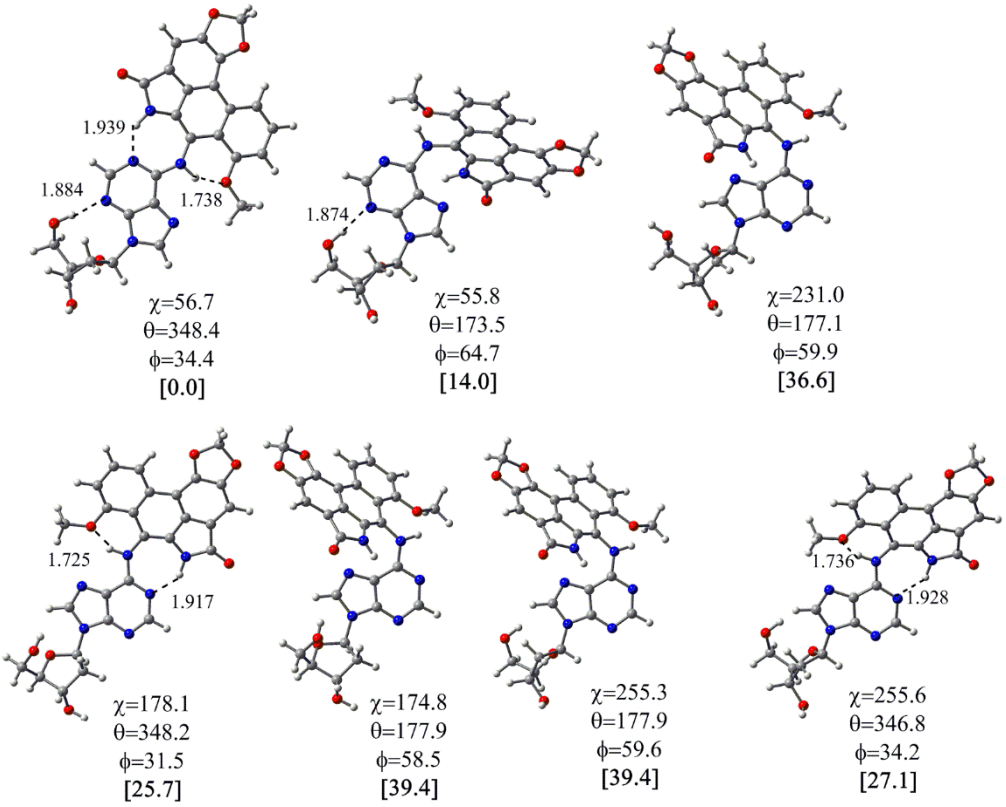


Figure A9: B3LYP-D3/6-31G(d) optimized minima for the ALI-N^6-dA nucleoside (ϕ = 30 – 60°). Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol^{-1}, in square brackets) are provided.

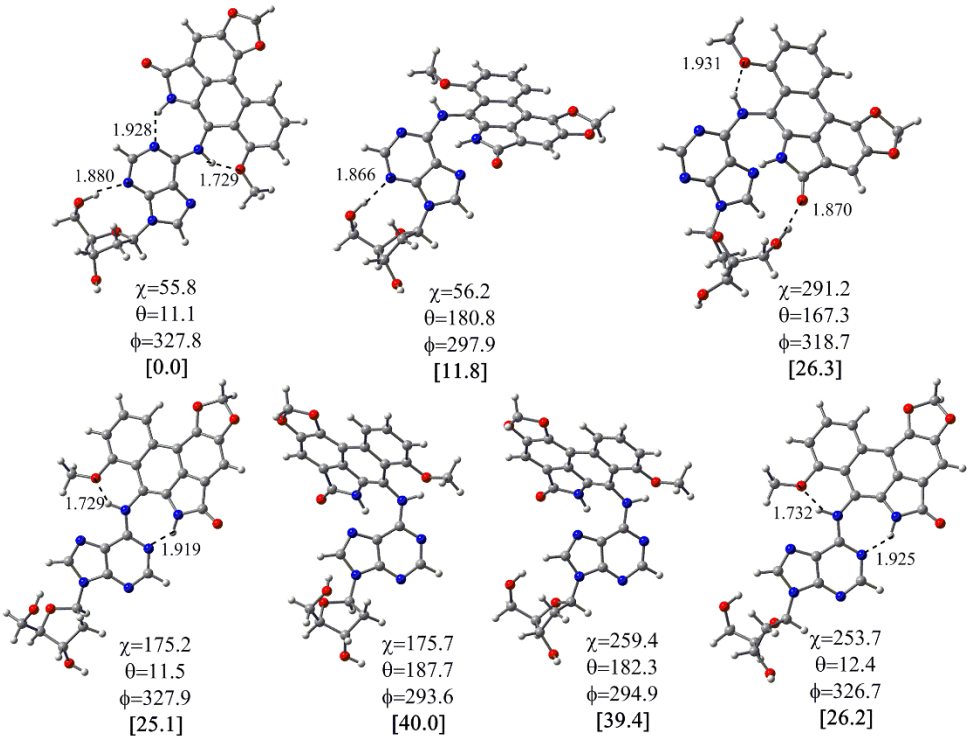


Figure A10: B3LYP-D3/6-31G(d) optimized minima for the ALI-N⁶-dA nucleoside (ϕ = 330˚– 300˚). Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol⁻¹, in square brackets) are provided.

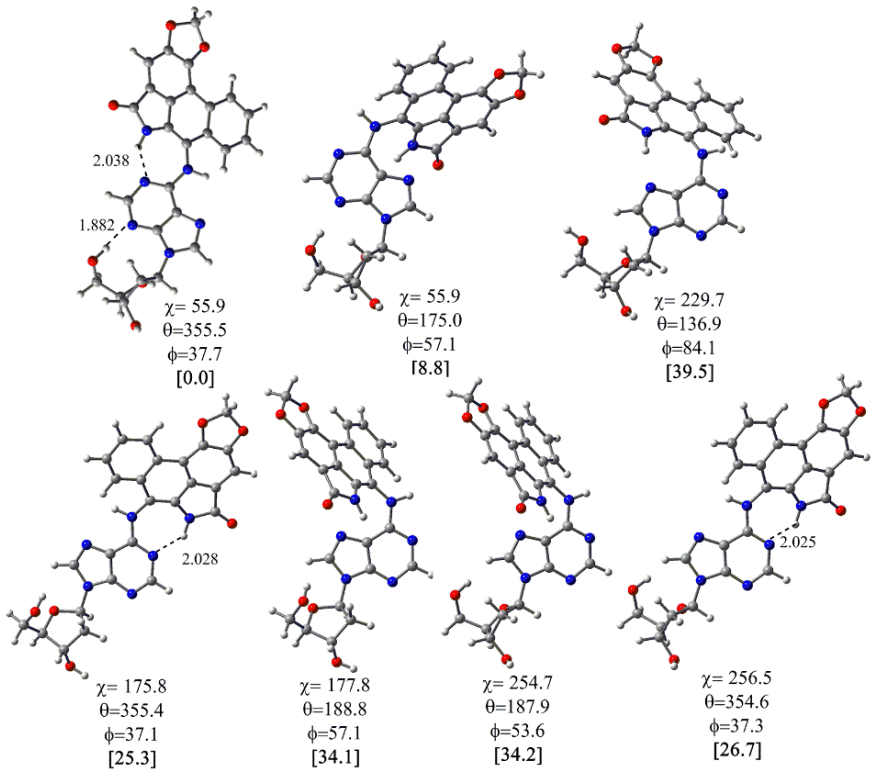


Figure A11: B3LYP-D3/6-31G(d) optimized minima for the ALII-N^6-dA nucleoside (ϕ = 30 – 60˚). Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6311+G(2df,p) relative energies (kJ mol^{-1}, in square brackets) are provided.

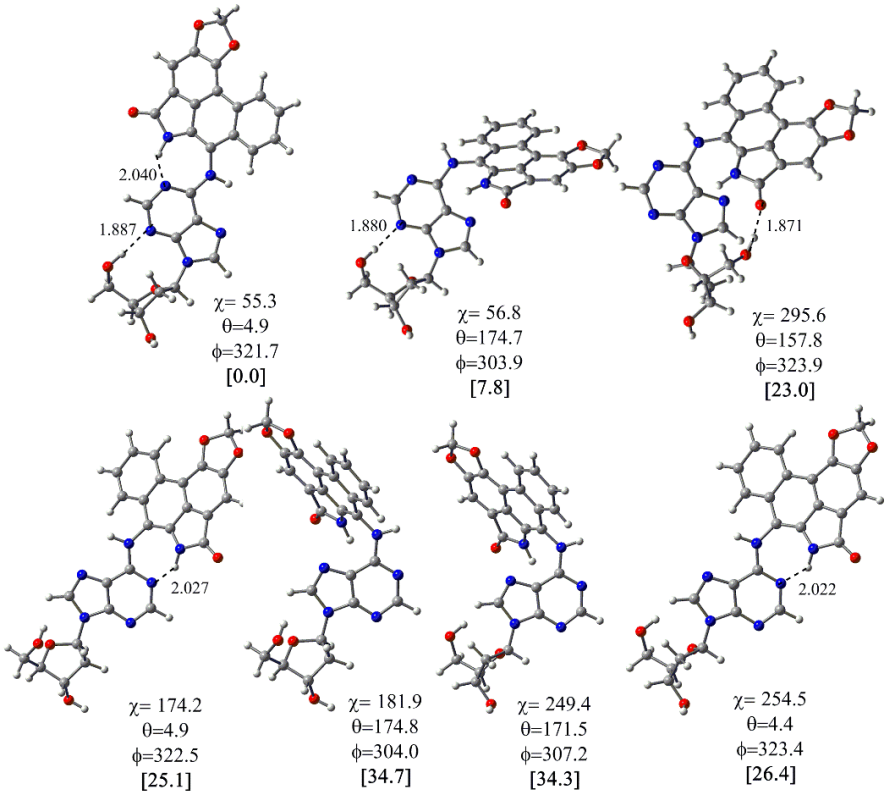


Figure A12: B3LYP-D3/6-31G(d) optimized minima for the ALII-N⁶-dA nucleoside (ϕ = 330˚– 300˚). Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol⁻¹, in square brackets) are provided.

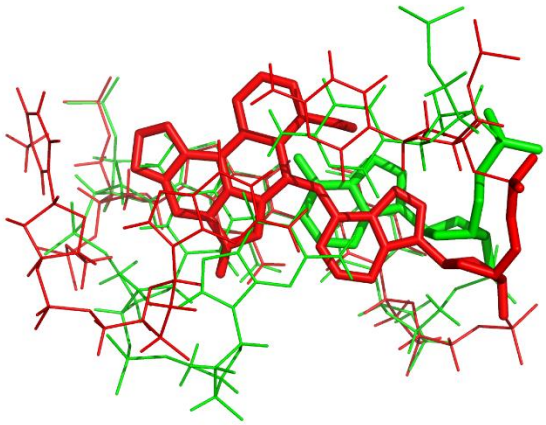


Figure A13: Overlay of the modified (red) and natural (green) strands, including the 5′and 3′flanking base pairs (wireframe), which shows the displacement of the modified adenine (tube) towards the minor groove relative to natural adenine (tube).

(a) (b)

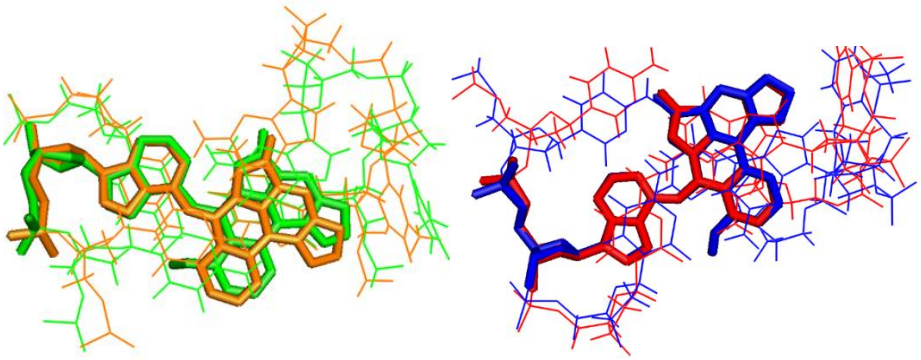


Figure A14: Overlay of the damaged adenine in the (a)* anti *(ALI-N^6-dA in green and ALII-N^6-dA in orange) and (b)* syn *(ALI-N^6-dA in blue and ALII-N^6-dA in red) base-displaced intercalated conformers of ALI-N^6-dA and ALII-N^6-dA adducted DNA.

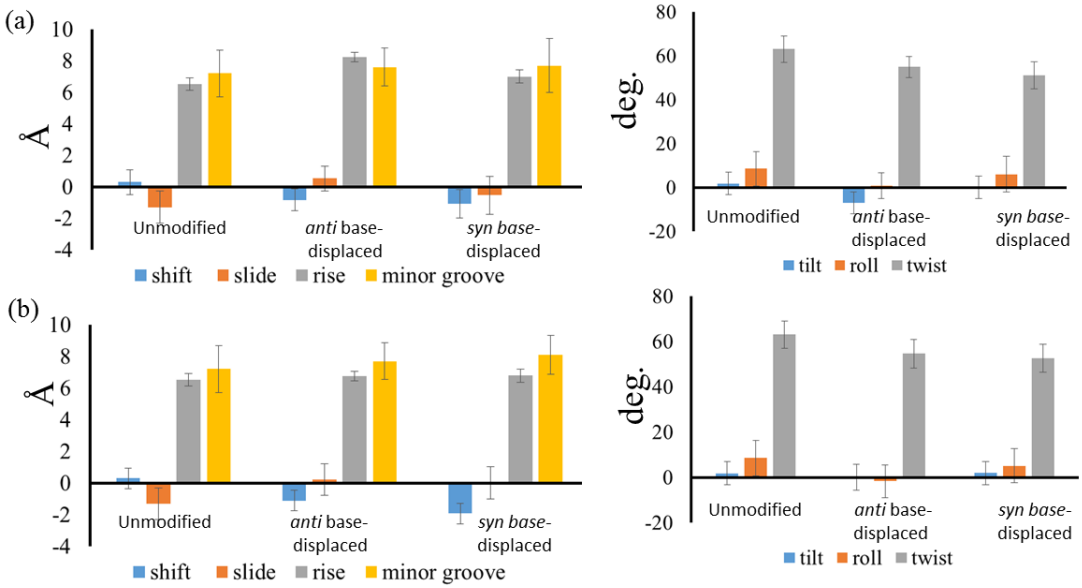


Figure A15: The pseudostep parameters and minor groove dimensions for (a) ALI-N⁶-dA and (b) ALII-N⁶-dA adducted DNA in the* anti *base-displaced intercalated and* syn *base-displaced intercalated conformers, as well as the unmodified strand. The helical dynamics are indicated by error bars. Pseudostep parameters were calculated using a pseudostep consisting of the base pairs 5′ and 3′ with respect to the lesion.

Table A1: Comparison of key B3LYP/6-31G(d) and B3LYP-D3/6-31G(d) structural parameters (deg.), and B3LYP/6-311+G(2df,p) and B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol^{-1}) for the nucleobase model.[a]

Adduct	B3LYP-D3			B3LYP		
	θ	**ϕ**	**ΔE[b]**	**θ**	**Φ**	**ΔE[b]**
ALI-N^6-A	12.4	327.0	0.0	8.5	331.9	0.0
	347.5	33.01	0.1	351.4	28.2	0.1
	173.7	63.6	13.6	172.6	65.9	17.6
	185.8	296.9	13.6	172.9	294.3	17.5
ALII-N^6-A	5.2	322.4	0.1	4.3	34.9	0.1
	354.7	37.6	0.0	355.7	35.1	0.0
	188.1	53.5	8.7	181.5	61.4	12.9
	171.8	306.5	8.8	178.2	298.9	13.0

[a]See Figure 2.1a for definitions of the θ and ϕ dihedral angles. [b]Relative energies include ZPVE (zero-point vibrational energy) corrections and are reported with respect to the lowest energy conformation of each adduct

Table A2: Comparison of key B3LYP/6-31G(d) and B3LYP-D3/6-31G(d) structural parameters (deg.), and B3LYP/6-311+G(2df,p) and B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol^{-1}) for the nucleoside model (ϕ = 30 – 60°).[a]

Adduct	B3LYP-D3				B3LYP			
	χ	**θ**	**ϕ**	**ΔE[b]**	**χ**	**θ**	**ϕ**	**ΔE[b]**
	56.7	348.4	34.4	0.0	57.1	352.0	30.4	0.0
	55.8	173.5	64.7	14.1	56.3	171.8	68	18.1
	231.0	177.1	59.9	36.7	232.2	175.3	62.5	35.6
ALI-N^6-dA	178.1	348.2	31.5	25.8	178	351.5	27.7	19.9
	174.8	177.9	39.4	39.4	180.6	175.8	62.1	38.4
	255.3	177.9	39.4	39.4	249.1	175.8	63.1	35.2
	255.6	346.8	27.1	27.1	246.9	351.7	27.1	18.7
	55.9	355.5	37.7	0.0	56.1	356.1	35.0	0.0
	55.9	175	57.1	8.8	56.4	179.6	63.5	13.3
	229.7	136.9	84.1	39.6	231.1	147.1	73.1	36.8
ALII-N^6-dA	175.8	355.4	37.1	25.3	174.7	355.5	34.4	20.1
	177.8	188.8	57.1	34.1	178.5	178.9	63.8	33.6
	254.7	187.9	53.6	34.2	248.9	182.1	60.9	30.6
	256.5	354.6	37.3	26.7	247.1	354.8	35.5	17.7

[a]See Figure 2.1a for the definitions of the χ, θ and ϕ dihedral angles. [b]Relative energies include ZPVE corrections and are reported with respect to the lowest energy conformation of each adduct.

Table A3: Comparison of key B3LYP/6-31G(d) and B3LYP-D3/6-31G(d) structural parameters (deg.), and B3LYP/6-311+G(2df,p) and B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol^{-1}) for the nucleoside model (ϕ = 330 – 300°).[a]

Adduct	B3LYP-D3				B3LYP			
	χ	**θ**	**Φ**	**ΔE[b]**	**χ**	**θ**	**ϕ**	**ΔE[b]**
	55.8	11.1	327.8	0.0	56.2	7.7	332.2	0.0
	56.2	180.8	297.9	11.8	56.5	182.6	297.9	16.3
	291.2	167.3	318.7	26.3	292.9	165.1	322.1	36.4
ALI-N^6-dA	175.2	11.5	327.9	25.1	177.5	7.3	333.9	19.6
	175.7	187.7	293.6	40.0	179.9	187.7	293.6	38.8
	259.4	182.3	294.9	39.5	250.8	173	294.9	35.9
	253.7	12.4	326.7	26.2	247.4	8.5	331.3	17.7
	55.3	4.9	321.7	0.0	55.5	4.0	324.6	0.0
	56.8	174.7	303.9	7.8	56.9	179.9	297.1	12.5
	295.6	157.8	323.9	23.0	296.3	156.7	326.4	34.4
ALII-N^6-dA	174.2	4.9	322.5	25.1	176.2	4.1	325.4	19.6
	181.9	174.8	304.0	34.8	180.2	179.5	298.1	33.6
	249.4	171.5	307.2	34.4	249.4	177.3	299.3	30.7
	254.5	4.4	323.4	26.4	246.7	3.9	324.9	27.0

[a]See Figure 2.1a for the definitions of the χ, θ and ϕ dihedral angles. [b]Relative energies include ZPVE corrections and are reported with respect to the lowest energy conformation of each adduct.

Table A4: Cartesian coordinates of the nucleobase minimum for the ALI-N^6-A adduct with θ = 12.4° and ϕ = 327.0°.

Calculated energy (in Hartrees) = -1476.25481800

Atom	X	Y	Z
N	6.21374500	-0.50450500	0.14198000
C	5.89316400	0.76321900	-0.29472800
H	6.65705100	1.50503400	-0.48692100
N	4.60219500	0.95003200	-0.44169800
C	4.04315100	-0.26570200	-0.08429700
C	2.69823700	-0.71202300	-0.01111000
N	1.71658400	0.16250800	-0.35887500
H	2.09314600	1.08469900	-0.55650500
N	2.48121500	-1.97459300	0.40146000
C	3.53510300	-2.75031300	0.72434600
H	3.28183100	-3.75789400	1.04530100
N	4.83222000	-2.44750400	0.70226800
C	5.02575700	-1.18708800	0.28641300
C	0.31394200	0.16853000	-0.20701300
C	-0.45650700	-0.97006600	-0.23687000
C	-0.37883000	1.47118600	-0.06421700
C	-1.87402400	-0.88845800	-0.13125200
C	-1.81578700	1.53923200	0.04452700
C	0.30132200	2.73574200	-0.03096700
C	-2.42819700	-2.17060600	-0.21152200
C	-2.58549300	0.31014400	0.00362700
C	-2.47806800	2.77598300	0.19176600
C	-0.37720300	3.93828300	0.11713100
C	-3.80707300	-2.35514700	-0.14605500
C	-3.97038400	0.09565000	0.07398500
C	-1.77073700	3.95873100	0.23141400
H	-3.55722700	2.78364600	0.27377400
H	0.17452600	4.86969000	0.14139800
C	-4.53818700	-1.18054500	0.00021300
H	-4.27811800	-3.32954800	-0.20675300
H	-2.28791600	4.90676400	0.34666700
N	-0.14682700	-2.31515500	-0.41713100
H	0.79693600	-2.65411200	-0.23066300
C	-1.29285000	-3.12567300	-0.34960500
O	-1.30430800	-4.34454800	-0.39236500
O	-5.91194100	-1.07366600	0.05727600
O	-4.96983400	1.03097000	0.18817600
C	-6.18093800	0.28550900	0.38944100
H	-6.47920200	0.36264600	1.44570200
H	-6.95918200	0.67972100	-0.26961900

H	7.13405800	-0.87759000	0.32690200
O	1.66839700	2.71626900	-0.15890800
C	2.43316200	3.91318600	-0.15008100
H	2.14773000	4.57530000	-0.97746300
H	2.32245900	4.44599600	0.80264300
H	3.46411400	3.57934100	-0.27479400

Table A5: Cartesian coordinates of the nucleobase minimum for the ALI-N^6-A adduct with θ = 347.5° and ϕ = 33.1°.

Calculated energy (in Hartrees) = -1476.25482693

Atom	X	Y	Z
N	-6.21564000	-0.49907200	0.11206600
C	-5.89065800	0.76903000	-0.32026300
H	-6.65234600	1.51225300	-0.51569500
N	-4.59855700	0.95430400	-0.45902300
C	-4.04339900	-0.26289000	-0.10050400
C	-2.69949600	-0.71101900	-0.01956100
N	-1.71438300	0.16273300	-0.35952100
H	-2.08871400	1.08564500	-0.55806000
N	-2.48688300	-1.97462300	0.39219900
C	-3.54377000	-2.74950300	0.70718600
H	-3.29382900	-3.75800600	1.02785500
N	-4.84030900	-2.44499900	0.67729700
C	-5.02951900	-1.18356700	0.26244100
C	-0.31298700	0.16756900	-0.19640100
C	0.45688300	-0.97160900	-0.22118100
C	0.37968500	1.46958800	-0.04710600
C	1.87332000	-0.89140500	-0.10121500
C	1.81592600	1.53655300	0.07128400
C	-0.30001700	2.73445500	-0.01583700
C	2.42709100	-2.17406100	-0.17668100
C	2.58518900	0.30716600	0.03216700
C	2.47799400	2.77263800	0.22470700
C	0.37838400	3.93639400	0.13809100
C	3.80551300	-2.35957200	-0.10466500
C	3.97001400	0.09241700	0.10212600
C	1.77120400	3.95582000	0.26084800
H	3.55650900	2.77942500	0.31498000
H	-0.17297800	4.86805400	0.16079100
C	4.53726500	-1.18456400	0.03487200
H	4.27548400	-3.33514800	-0.15378600
H	2.28827800	4.90333900	0.38073800
N	0.14755400	-2.31643800	-0.40425700

H	-0.79799500	-2.65450000	-0.22491100
C	1.29229800	-3.12800700	-0.32645200
O	1.30315900	-4.34683200	-0.37018500
O	5.90696000	-1.08291200	0.16198800
O	4.96727900	1.02361200	0.26223000
C	6.19884000	0.30881800	0.07511300
H	6.60416400	0.54665200	-0.91988800
H	6.90069100	0.58891900	0.86490500
H	-7.13755300	-0.87118700	0.29088100
O	-1.66620700	2.71592400	-0.15174300
C	-2.43086300	3.91289400	-0.14139600
H	-2.32388500	4.44197700	0.81382700
H	-2.14191100	4.57828000	-0.96494400
H	-3.46131400	3.57973400	-0.27176300

Table A6: Cartesian coordinates of the nucleobase minimum for the ALI-N^6-A adduct with θ = 173.7° and ϕ = 63.6°.

Calculated energy (in Hartrees) = -1476.24646563

Atom	X	Y	Z
N	-3.66816800	-2.47611500	1.95373900
C	-2.28985900	-2.44913700	2.02492800
H	-1.74512700	-3.09459700	2.70145500
N	-1.74358600	-1.58786600	1.20598600
C	-2.81969000	-1.00691800	0.54440700
C	-2.94778200	0.00182000	-0.44057600
N	-1.93051700	0.67485400	-1.06044100
H	-2.25238100	1.52816900	-1.49805300
N	-4.18851300	0.36276900	-0.83477600
C	-5.24861500	-0.25084800	-0.29619000
H	-6.21735700	0.08851100	-0.65730200
N	-5.27369200	-1.22005300	0.62563800
C	-4.03567100	-1.54956000	1.00558300
C	-0.55013600	0.59468400	-0.76304900
C	0.09629000	-0.59659300	-0.93796600
C	0.22495700	1.74323900	-0.27795500
C	1.45351900	-0.75294500	-0.55348200
C	1.60828000	1.59361700	0.09822600
C	-0.33023200	3.05668100	-0.16007200
C	1.85750800	-2.07458100	-0.76111600
C	2.23498500	0.28990200	-0.03851200
C	2.33969900	2.69132400	0.59670500
C	0.41300800	4.11717000	0.34435900
C	3.15618200	-2.47271800	-0.45001900

C	3.53711000	-0.13828800	0.26122600
C	1.74712000	3.93082600	0.72660100
H	3.37513800	2.54813900	0.87908700
H	-0.03062600	5.10108600	0.43159200
C	3.96098900	-1.45680600	0.05779900
H	3.51755400	-3.48580300	-0.58413200
H	2.31439800	4.77157700	1.11527300
N	-0.32373200	-1.82439600	-1.43502600
H	-1.28420100	-2.07640300	-1.61270900
C	0.67775000	-2.80900300	-1.30089100
O	0.54916500	-3.98675600	-1.58206400
O	5.26890400	-1.58401500	0.47479300
O	4.56640100	0.59253400	0.79910000
C	5.72191900	-0.25893200	0.74098700
H	6.37992600	0.08339700	-0.07141900
H	6.23746400	-0.23174700	1.70464700
H	-4.30409900	-3.05384800	2.48448400
O	-1.62200600	3.22254000	-0.59124100
C	-2.27480700	4.46805300	-0.39313900
H	-2.29355600	4.74467700	0.66874600
H	-1.79818100	5.27065000	-0.97123000
H	-3.29635700	4.32147400	-0.74811600

Table A7: Cartesian coordinates of the nucleobase minimum for the ALI-N^6-A adduct with θ = 185.8° and ϕ = 296.9°.

Calculated energy (in Hartrees) = -1476.25482693

Atom	X	Y	Z
N	3.63600600	-2.49576100	1.96361400
C	2.25751100	-2.45990700	2.02593900
H	1.70415200	-3.10522900	2.69556800
N	1.72230100	-1.59093600	1.20783000
C	2.80658600	-1.01394100	0.55609900
C	2.94799400	-0.00116200	-0.42303400
N	1.93964400	0.68179600	-1.04629400
H	2.26879100	1.53596400	-1.47674900
N	4.19381000	0.35411900	-0.80631700
C	5.24617400	-0.26881600	-0.26337100
H	6.21956500	0.06646700	-0.61568000
N	5.25860600	-1.24323000	0.65322400
C	4.01587300	-1.56681000	1.02269100
C	0.55577400	0.60286900	-0.76640400
C	-0.09071000	-0.58674900	-0.95271600
C	-0.22310900	1.75139600	-0.28741100
C	-1.45268600	-0.74177300	-0.58470300

C	-1.60994800	1.60215100	0.07581400
C	0.33138900	3.06460000	-0.16402200
C	-1.85627500	-2.06232800	-0.80001900
C	-2.23563000	0.29864200	-0.06703000
C	-2.34523000	2.69971000	0.56894400
C	-0.41561200	4.12481700	0.33538300
C	-3.15649600	-2.46132200	-0.49643300
C	-3.53742500	-0.13187100	0.23071400
C	-1.75295300	3.93856800	0.70619000
H	-3.38358100	2.55715600	0.84070400
H	0.02740000	5.10861900	0.42689500
C	-3.96111100	-1.44905000	0.01882700
H	-3.51964700	-3.47205400	-0.64293900
H	-2.32320400	4.77918100	1.09078800
N	0.33248000	-1.81367800	-1.44923600
H	1.29382200	-2.06610000	-1.62136700
C	-0.67185500	-2.79708600	-1.32895700
O	-0.54199400	-3.97419100	-1.61230400
O	-5.28868500	-1.56291300	0.37286200
O	-4.58088900	0.60717900	0.72839900
C	-5.62614000	-0.33566600	1.01473200
H	-5.68686000	-0.48523200	2.10286800
H	-6.57082800	0.04242800	0.61489200
H	4.26463900	-3.08040700	2.49547500
O	1.62644700	3.23075400	-0.58533800
C	2.27648400	4.47724900	-0.38422900
H	1.80250900	5.27864800	-0.96611200
H	2.28817600	4.75489100	0.67749000
H	3.30047100	4.33172600	-0.73255800

Table A8: Cartesian coordinates of the nucleobase minimum for the ALII-N^6-A adduct with θ = 354.7° and ϕ = 37.6°.

Calculated energy (in Hartrees) = -1361.72857366

Atom	X	Y	Z
N	-6.37851200	0.14615100	-0.14437200
C	-5.96973400	1.31540900	-0.75316400
H	-6.68395300	2.03780500	-1.12570200
N	-4.66473000	1.43724500	-0.82055100
C	-4.19204900	0.28209000	-0.22564500
C	-2.88200300	-0.18251000	0.02951000
N	-1.82538600	0.58976900	-0.34956000
H	-2.11280700	1.43981900	-0.81932400
N	-2.73499100	-1.36755600	0.64637700
C	-3.84206500	-2.05111300	1.00426200

H	-3.65470600	-3.00152300	1.49804000
N	-5.12134500	-1.71954300	0.83244200
C	-5.23646100	-0.53913800	0.20691800
C	-0.43650200	0.45235100	-0.14127600
C	0.24055200	-0.73868300	-0.15903900
C	0.33380300	1.68763000	0.03916800
C	1.66631900	-0.74479100	-0.06995500
C	1.77086700	1.68340700	0.07965400
C	-0.32007500	2.93336800	0.18094900
C	2.13391000	-2.05946600	-0.15365400
C	2.46121100	0.40954900	0.00821400
C	2.46470700	2.90261900	0.20693400
H	-1.40352300	2.97430800	0.23071300
C	0.38777400	4.11839800	0.30910100
C	3.50110800	-2.33187100	-0.14076000
C	3.82905500	0.10722500	0.01914800
C	1.78883500	4.10801600	0.31027000
H	3.54829300	2.88456500	0.23019300
H	-0.15226800	5.05464900	0.41990800
C	4.31131700	-1.20440000	-0.05222800
H	3.90590400	-3.33564800	-0.19930300
H	2.34454500	5.03622400	0.40750100
N	-0.16124200	-2.05935300	-0.32084000
H	-1.11137300	-2.33659400	-0.08639100
C	0.93663500	-2.94257600	-0.25459200
O	0.86713800	-4.15834400	-0.27474100
O	5.68795600	-1.18653800	0.01113600
O	4.88485500	0.97791700	0.11831600
C	6.06429200	0.18491800	-0.09319200
H	6.45921000	0.39007800	-1.09908000
H	6.80193700	0.42452000	0.67729400
H	-7.32638300	-0.16206400	0.02024100

Table A9: Cartesian coordinates of the nucleobase minimum for the ALII-N^6-A adduct with θ = 5.2° and ϕ = 322.4°.

Calculated energy (in Hartrees) = -1361.72855858

Atom	X	Y	Z
N	6.37930700	0.14246400	-0.11711400
C	5.97534300	1.31099000	-0.73052900
H	6.69249300	2.03201000	-1.10008900
N	4.67089400	1.43395400	-0.80599300
C	4.19351300	0.28031700	-0.21190300
C	2.88148900	-0.18265900	0.03596500

N	1.82794500	0.58989900	-0.35084600
H	2.11890200	1.43859700	-0.82088000
N	2.72955600	-1.36650200	0.65400700
C	3.83381000	-2.05040500	1.01987100
H	3.64257800	-2.99980500	1.51411800
N	5.11443200	-1.72031400	0.85535600
C	5.23448900	-0.54110500	0.22846400
C	0.43754100	0.45322100	-0.15216200
C	-0.23975500	-0.73761000	-0.17339100
C	-0.33353400	1.68897400	0.02155500
C	-1.66626400	-0.74294500	-0.09704400
C	-1.77084900	1.68524000	0.05297500
C	0.31981900	2.93479500	0.16515500
C	-2.13380500	-2.05740100	-0.18375100
C	-2.46108200	0.41133200	-0.01879200
C	-2.46506500	2.90492500	0.17375800
H	1.40296900	2.97555000	0.22097100
C	-0.38842400	4.12022500	0.28712600
C	-3.50114900	-2.32934700	-0.17649600
C	-3.82899300	0.10893000	-0.00733500
C	-1.78946900	4.11022300	0.27988100
H	-3.54879100	2.88743100	0.18911100
H	0.15121100	5.05655700	0.39922900
C	-4.31136300	-1.20224100	-0.08304900
H	-3.90618000	-3.33241800	-0.24467400
H	-2.34549500	5.03876900	0.37196800
N	0.16259000	-2.05850800	-0.33161100
H	1.11091400	-2.33617400	-0.09064100
C	-0.93617500	-2.94120000	-0.27353300
O	-0.86701900	-4.15702600	-0.29182300
O	-5.68920300	-1.18093400	-0.08413100
O	-4.88520000	0.98267200	0.04833300
C	-6.06003400	0.16795900	0.19325700
H	-6.42981500	0.25051000	1.22567300
H	-6.81660900	0.49744900	-0.52420500
H	7.32586700	-0.16642800	0.05365100

Table A10: Cartesian coordinates of the nucleobase minimum for the ALII-N^6-A adduct with θ =188.1° and ϕ = 53.5°.

Calculated energy (in Hartrees) = -1361.72333809

Atom	X	Y	Z
N	-4.35317000	-1.46023100	1.87024400
C	-3.00626900	-1.52804300	2.16541300
H	-2.63490800	-2.14279800	2.97484300

N	-2.26252000	-0.79418600	1.37754000
C	-3.16665100	-0.20352600	0.50343600
C	-3.04599100	0.70367200	-0.57466000
N	-1.88032600	1.24488700	-1.03808200
H	-2.02603300	1.93572200	-1.76617200
N	-4.16505700	1.10665400	-1.21096100
C	-5.35004000	0.63230600	-0.80466700
H	-6.21284800	0.99844600	-1.35679000
N	-5.60858900	-0.22937900	0.18498300
C	-4.48327900	-0.60725600	0.79881800
C	-0.55216200	0.90729400	-0.68804900
C	-0.08685700	-0.37451300	-0.75889900
C	0.36947800	1.96479600	-0.28413600
C	1.26796000	-0.66194700	-0.41927900
C	1.74991300	1.69098200	0.00780400
C	-0.08194700	3.29759200	-0.17004000
C	1.50664300	-2.03265000	-0.54736600
C	2.21001800	0.31585700	-0.06125500
C	2.60256300	2.75431700	0.36367800
H	-1.13090800	3.51592200	-0.34187800
C	0.77894200	4.32447600	0.18350400
C	2.77298100	-2.55974200	-0.29668500
C	3.47223700	-0.24185800	0.18307800
C	2.13028800	4.05510800	0.44471400
H	3.64301000	2.54014400	0.58033000
H	0.39971900	5.33949900	0.26477500
C	3.72602800	-1.61341100	0.06765400
H	3.00490200	-3.61544300	-0.37724200
H	2.80516600	4.86048800	0.72026500
N	-0.66926400	-1.57780000	-1.13506400
H	-1.66141400	-1.73241100	-1.23440500
C	0.22300500	-2.66144400	-0.96857800
O	-0.05588800	-3.83223000	-1.14784300
O	5.03969700	-1.86225300	0.39896400
O	4.61545700	0.40346700	0.57811500
C	5.65852400	-0.58412200	0.53624400
H	6.30629100	-0.38636700	-0.32997700
H	6.22611300	-0.54774300	1.47014700
H	-5.11302000	-1.93427400	2.33723200

Table A11: Cartesian coordinates of the nucleobase minimum for the ALII-N^6-A adduct with θ = 171.8° and ϕ = 306.5°.

Calculated energy (in Hartrees) = -1361.72333322

Atom	X	Y	Z
N	4.34039800	-1.46822700	1.88187800
C	2.99212100	-1.53473300	2.17087100
H	2.61624800	-2.15139900	2.97675600
N	2.25300900	-0.79728400	1.38196200
C	3.16197300	-0.20547300	0.51361600
C	3.04757300	0.70499400	-0.56243400
N	1.88500900	1.24932000	-1.02978400
H	2.03542600	1.94135900	-1.75580000
N	4.17022500	1.10830700	-1.19223300
C	5.35261300	0.63101700	-0.78195000
H	6.21849100	0.99761300	-1.32893700
N	5.60531500	-0.23420000	0.20614900
C	4.47663000	-0.61218400	0.81365300
C	0.55436500	0.91081200	-0.69022500
C	0.08930200	-0.37075600	-0.76725200
C	-0.37021700	1.96790000	-0.29192100
C	-1.26919500	-0.65787300	-0.44218400
C	-1.75312100	1.69406000	-0.01187000
C	0.08060300	3.30040200	-0.17201400
C	-1.50785200	-2.02800000	-0.57632900
C	-2.21251500	0.31895800	-0.08501200
C	-2.60841400	2.75708700	0.33859300
H	1.13092000	3.51880600	-0.33523900
C	-0.78282300	4.32691800	0.17637000
C	-2.77598700	-2.55523300	-0.33509600
C	-3.47519400	-0.23960100	0.15501100
C	-2.13630600	4.05750100	0.42623300
H	-3.65092200	2.54310100	0.54523700
H	-0.40405000	5.34174100	0.26210300
C	-3.72933900	-1.61050400	0.03263800
H	-3.00959500	-3.60959900	-0.42741000
H	-2.81320500	4.86264600	0.69748400
N	0.67397200	-1.57349700	-1.14162800
H	1.66668400	-1.72858100	-1.23416700
C	-0.22063500	-2.65684600	-0.98636500
O	0.05908500	-3.82739800	-1.16591900
O	-5.05879800	-1.85224900	0.29982200
O	-4.62993900	0.41025100	0.50673100
C	-5.59197900	-0.62055900	0.78378600
H	-5.75058200	-0.68188100	1.87013000

H	-6.52474200	-0.39190100	0.26098000
H	5.09742500	-1.94499300	2.35067200

Table A12: Cartesian coordinates of* syn *nucleoside minimum for the ALI-N^6-dA adduct with χ = 56.7°, θ = 348.4° and ϕ = 34.4°.

Calculated energy (in Hartrees) = -1897.27905061

Atom	X	Y	Z
N	-4.29017200	0.97941400	-0.36804100
C	-3.66819000	2.16594000	-0.70457800
H	-4.24995600	3.05771700	-0.89952200
N	-2.35990400	2.08690300	-0.76008400
C	-2.09476800	0.77004500	-0.42930300
C	-0.87444200	0.05942400	-0.28813600
N	0.28289700	0.72638200	-0.52425200
H	0.11793200	1.71559100	-0.68808300
N	-0.94101600	-1.23679100	0.07632700
C	-2.14366400	-1.79182000	0.29398300
H	-2.13048700	-2.83894800	0.58528900
N	-3.34775400	-1.22567000	0.20068600
C	-3.27291300	0.06601000	-0.17233400
C	1.64320600	0.43126800	-0.28661100
C	2.16455000	-0.83832300	-0.36438300
C	2.57241300	1.54855000	0.00183100
C	3.55718300	-1.06187300	-0.16741700
C	3.98079800	1.30805000	0.20016200
C	2.16148100	2.92110700	0.09578900
C	3.84400900	-2.42234900	-0.32175000
C	4.48591100	-0.04834300	0.09970600
C	4.86683400	2.36701500	0.48842100
C	3.05654700	3.94330600	0.38329900
C	5.14893200	-2.89306900	-0.19738900
C	5.79123400	-0.54871700	0.22052000
C	4.41245300	3.66534900	0.58266600
H	5.91578100	2.14512200	0.63624100
H	2.70536200	4.96528200	0.45126200
C	6.09196200	-1.90727000	0.07542000
H	5.41509000	-3.93844300	-0.30405200
H	5.10066100	4.47502200	0.80687000
N	1.60186700	-2.07285000	-0.67534800
H	0.60135100	-2.22575000	-0.55841200
C	2.55232400	-3.10798600	-0.60599000
O	2.31737300	-4.29567000	-0.74822400
O	7.44205300	-2.10022000	0.27718500

O	6.94284400	0.14267000	0.50581500
C	8.01506100	-0.79647100	0.32734200
H	8.53139500	-0.57611100	-0.61887800
H	8.69957900	-0.72632600	1.17669500
O	0.83317300	3.19326900	-0.11895000
C	0.32955500	4.52115900	-0.07074800
H	0.46748600	4.96068200	0.92494900
H	0.81087800	5.15735800	-0.82430100
H	-0.73404400	4.41892200	-0.28997800
C	-5.72258500	0.77449000	-0.20518800
O	-5.96870000	0.32881500	1.11106400
H	-6.19097600	1.75348200	-0.37756300
C	-6.33088800	-0.28617000	-1.13103500
C	-7.06545000	-0.61802600	1.12289500
C	-7.56659300	-0.70603000	-0.32845800
H	-5.65087000	-1.13420900	-1.23079700
H	-6.57181800	0.11069800	-2.12208100
C	-6.57143600	-1.95077700	1.68550800
H	-7.85918000	-0.20924100	1.75967500
H	-7.89634200	-1.72237000	-0.57594200
H	-7.44363600	-2.58268000	1.89557900
H	-6.06661500	-1.74919600	2.64390600
O	-5.74993000	-2.66612200	0.78823100
H	-4.90492100	-2.17206700	0.68054400
O	-8.62995100	0.24034700	-0.46155000
H	-8.96413700	0.18209500	-1.37032200

Table A13: Cartesian coordinates of* anti *nucleoside minimum for the ALI-N^6-dA adduct with χ = 178.1°, θ = 348.2° and ϕ = 31.5°.

Calculated energy (in Hartrees) = -1897.26826896

Atom	X	Y	Z
N	-4.24774800	-0.19888200	-0.62614800
C	-3.82171300	1.05517400	-1.02008500
H	-4.53426000	1.81521500	-1.30408800
N	-2.51250300	1.18719100	-1.02247600
C	-2.04867600	-0.05161500	-0.62121800
C	-0.74113300	-0.56147300	-0.40902500
N	0.30983900	0.26946800	-0.63878800
H	-0.00323000	1.20799000	-0.86943700
N	-0.62807200	-1.83681600	0.00801600
C	-1.74356500	-2.56626900	0.20491000
H	-1.57180900	-3.58743200	0.53688000
N	-3.01552300	-2.20305500	0.04723500
C	-3.10672100	-0.92870800	-0.37082300

C	1.68952000	0.21246400	-0.35092900
C	2.40734800	-0.95951600	-0.29708200
C	2.42333800	1.48499100	-0.14858200
C	3.80865200	-0.93963900	-0.04517000
C	3.84393500	1.49031500	0.10251900
C	1.80171100	2.77921500	-0.19248000
C	4.30918100	-2.24616900	-0.05546100
C	4.55742200	0.22794200	0.14727900
C	4.54425900	2.69788900	0.30508200
C	2.51677600	3.95205200	0.01202600
C	5.66449900	-2.49083600	0.15204400
C	5.91800700	-0.04623000	0.35333500
C	3.89166500	3.91161800	0.26346700
H	5.60893000	2.65828200	0.49546400
H	2.00851500	4.90738200	-0.02612500
C	6.43120000	-1.34746700	0.35269200
H	6.09264400	-3.48666400	0.15775700
H	4.43793200	4.83693600	0.42123400
N	2.05778800	-2.29204300	-0.49647800
H	1.08453400	-2.58657100	-0.41444900
C	3.15249500	-3.15153700	-0.30471500
O	3.11296800	-4.37029400	-0.33798500
O	7.78495600	-1.30497400	0.61257700
O	6.93527700	0.84241500	0.60301600
C	8.14646000	0.07198800	0.55274300
H	8.66759800	0.28319300	-0.39301800
H	8.77093200	0.32753500	1.41291600
O	0.45477100	2.81922900	-0.45387100
C	-0.25478000	4.04767000	-0.52954400
H	-0.21261300	4.58788700	0.42446700
H	0.13687200	4.68586000	-1.33180400
H	-1.28231300	3.75540100	-0.75000000
C	-5.61707700	-0.70258700	-0.53542300
O	-6.49116000	0.38405100	-0.78089700
H	-5.74975400	-1.47392100	-1.30369000
C	-5.97390000	-1.23401900	0.85669500
C	-7.64708500	0.30365700	0.09082600
C	-7.48546100	-0.99975400	0.89055900
H	-5.48631700	-0.61906800	1.61754300
H	-5.67474200	-2.27562600	0.99245900
C	-7.68003100	1.57729300	0.93140100
H	-8.54638700	0.23763600	-0.53182700
H	-7.86751400	-0.89262800	1.91489700
H	-8.53684400	1.54944900	1.61260000
H	-7.80995300	2.43984800	0.25965800
O	-6.52077600	1.72125100	1.73982200

H	-5.76753500	1.78361300	1.12955900
O	-8.20000800	-2.00403000	0.17131000
H	-8.03733200	-2.85521800	0.60744500

Table A14: Cartesian coordinates of the* syn *nucleoside minimum for the ALI-N^6-dA adduct with χ = 55.8°, θ = 11.1° and ϕ = 327.8°.

Calculated energy (in Hartrees) = -1897.27888394

Atom	X	Y	Z
N	4.28806600	1.00958600	-0.27592000
C	3.66286600	2.17876300	-0.66313200
H	4.24177300	3.06444300	-0.89141500
N	2.35476900	2.09358400	-0.71544900
C	2.09498100	0.78235700	-0.35902300
C	0.87797300	0.06459700	-0.22132900
N	-0.28032100	0.72104800	-0.48026800
H	-0.11699800	1.70627800	-0.66902700
N	0.95126700	-1.22607500	0.16120800
C	2.15607000	-1.77205100	0.38872500
H	2.14798500	-2.81726600	0.68700400
N	3.35712200	-1.20011800	0.29361400
C	3.27593300	0.08960600	-0.08488300
C	-1.64480100	0.42662200	-0.26827900
C	-2.16054300	-0.84764100	-0.30367200
C	-2.58570000	1.55167900	-0.05356500
C	-3.55732000	-1.06778300	-0.13160700
C	-3.99786100	1.31375500	0.12025200
C	-2.18408700	2.92982600	-0.01428400
C	-3.83597300	-2.43517100	-0.22972300
C	-4.49474900	-0.04855900	0.07588600
C	-4.89628300	2.38056700	0.33078000
C	-3.09137600	3.95990900	0.19618800
C	-5.14075000	-2.90641600	-0.10674700
C	-5.79908500	-0.55007600	0.20328800
C	-4.45104600	3.68475900	0.37147400
H	-5.94801000	2.16008700	0.45952000
H	-2.74631700	4.98586200	0.22324800
C	-6.09166000	-1.91504600	0.11334800
H	-5.40126700	-3.95620900	-0.17810100
H	-5.14901900	4.50062200	0.53475100
N	-1.58814900	-2.09283600	-0.54435700
H	-0.58706700	-2.23431600	-0.41640400
C	-2.53589900	-3.12755800	-0.45286000
O	-2.29428800	-4.31951200	-0.53775600
O	-7.45191400	-2.10017400	0.23933300

O	-6.96360300	0.15156900	0.39660200
C	-7.97957500	-0.83622500	0.63261400
H	-8.22605600	-0.84881000	1.70473700
H	-8.85889600	-0.59884700	0.02801100
O	-0.85059400	3.20026900	-0.19959200
C	-0.36217900	4.53388300	-0.24618900
H	-0.83814800	5.10417500	-1.05389200
H	-0.52107300	5.04744200	0.71026700
H	0.70587900	4.42740400	-0.44024100
C	5.72481900	0.79700300	-0.16369400
O	6.09217500	-0.25970700	-1.02326700
H	6.19679000	1.73711900	-0.48114700
C	6.21049600	0.39397000	1.23457200
C	7.15721800	-1.04296400	-0.42978700
C	7.50807500	-0.34708300	0.89578200
H	5.50069800	-0.29441800	1.69697200
H	6.36390900	1.25758000	1.88919400
C	6.68513900	-2.48722600	-0.25960600
H	8.02141800	-1.00422400	-1.10376200
H	7.78349800	-1.07572800	1.66783200
H	7.55746500	-3.11161600	-0.02845900
H	6.27744500	-2.82769300	-1.22480500
O	5.76586400	-2.65407500	0.79813800
H	4.92499100	-2.20580800	0.54963200
O	8.57786900	0.56096300	0.62164200
H	8.81957800	0.99196000	1.45631400

Table A15: Cartesian coordinates of the* anti *nucleoside minimum for the ALI-N^6-dA adduct with χ = 175.2°, θ = 11.5° and ϕ = 327.9°.

Calculated energy (in Hartrees) = -1897.26832217			
Atom	X	Y	Z
N	-4.24030800	0.17178500	-0.49304000
C	-3.80847700	1.42002800	-0.08730400
H	-4.50812000	2.23439100	0.02815900
N	-2.50882300	1.48730800	0.10821100
C	-2.05747300	0.21563400	-0.19127200
C	-0.76215200	-0.36447600	-0.19407700
N	0.28622600	0.41923900	0.17255000
H	-0.00471100	1.38068800	0.32388000
N	-0.65605300	-1.65499300	-0.56367800
C	-1.76750000	-2.33391100	-0.90902200
H	-1.60267300	-3.37066400	-1.19266200
N	-3.02819400	-1.90500500	-0.94684900
C	-3.11248600	-0.61613200	-0.57437300

C	1.68801700	0.28159300	0.09824100
C	2.33869200	-0.92630600	0.19290700
C	2.51407700	1.50606100	-0.02891800
C	3.76093300	-0.98653100	0.17003100
C	3.95445800	1.43089000	-0.04774600
C	1.96523900	2.82941700	-0.13233100
C	4.17955000	-2.31490900	0.30374200
C	4.59458900	0.13330800	0.05960700
C	4.74352800	2.59359400	-0.17165200
C	2.76714900	3.95650500	-0.25550900
C	5.53487100	-2.63526400	0.32291700
C	5.95303700	-0.21759900	0.07526200
C	4.16026400	3.83848200	-0.27650400
H	5.82124500	2.49460700	-0.18331000
H	2.31269400	4.93610100	-0.33479000
C	6.38617200	-1.54163600	0.20175900
H	5.90262800	-3.64968000	0.42651800
H	4.77491500	4.72869800	-0.37269500
N	1.88750700	-2.23020700	0.37697200
H	0.92520400	-2.47809000	0.14676000
C	2.94908700	-3.15060600	0.39094500
O	2.83660800	-4.36351600	0.45561300
O	7.76478400	-1.57144800	0.22343900
O	7.04521000	0.61236100	0.00611800
C	8.18558700	-0.25258600	-0.11384900
H	8.54803500	-0.22752600	-1.15238900
H	8.96126300	0.07699200	0.58258400
O	0.59769800	2.94493300	-0.10165200
C	-0.04469100	4.21112000	-0.15363600
H	0.25646800	4.84597000	0.68937100
H	0.17145900	4.72641600	-1.09790100
H	-1.10868000	3.97977500	-0.08842000
C	-5.60257300	-0.25433900	-0.81298000
O	-6.47345100	0.81449800	-0.49167200
H	-5.65123700	-0.46786400	-1.88744500
C	-6.06509500	-1.45564500	0.01872400
C	-7.68280800	0.30935400	0.12841000
C	-7.57589800	-1.22340700	0.07548900
H	-5.64255900	-1.38741500	1.02431900
H	-5.77141600	-2.40761600	-0.42896500
C	-7.77076900	0.90247100	1.53231900
H	-8.53917000	0.63443300	-0.47305300
H	-8.03602400	-1.68926000	0.95756000
H	-8.67506300	0.53781000	2.03071100
H	-7.84978900	1.99727000	1.44763000
O	-6.67423900	0.52441100	2.35205900

H	-5.87804000	0.90027300	1.94184200
O	-8.22747000	-1.63080600	-1.12704600
H	-8.09403000	-2.58632600	-1.22881000

Table A16: Cartesian coordinates of the* syn *nucleoside minimum for the ALII-N^6-dA adduct with χ = 55.9°, θ = 355.5° and ϕ = 37.7°.

Calculated energy (in Hartrees) = -1782.75256125

Atom	X	Y	Z
N	-4.25985100	1.14471600	-0.52685100
C	-3.64852900	2.28356600	-1.01803000
H	-4.24006300	3.11937300	-1.36903600
N	-2.33768100	2.23294000	-1.01125400
C	-2.06273500	0.99115000	-0.47324300
C	-0.83891000	0.34660800	-0.18209400
N	0.32452100	0.99426600	-0.45614800
H	0.17448000	1.90268700	-0.87928900
N	-0.87637100	-0.88294400	0.36449200
C	-2.07486000	-1.43520700	0.61995400
H	-2.04856400	-2.42779300	1.06193100
N	-3.28800300	-0.92474800	0.40118700
C	-3.23346200	0.29890100	-0.15661400
C	1.67226200	0.65895200	-0.19954700
C	2.18871100	-0.60728300	-0.28092100
C	2.58447300	1.76601500	0.10545800
C	3.59575600	-0.80870000	-0.13370500
C	4.00490200	1.56841100	0.20391400
C	2.08949800	3.07388400	0.31401000
C	3.89358300	-2.16459200	-0.29673900
C	4.52747200	0.22222200	0.06504900
C	4.84264300	2.67363900	0.45040800
H	1.01913500	3.25339500	0.32285700
C	2.93677100	4.14328400	0.55951700
C	5.21176500	-2.61581700	-0.24657900
C	5.84259300	-0.25839400	0.11082000
C	4.32293900	3.94758700	0.61651000
H	5.91242400	2.51159100	0.51609800
H	2.51710100	5.13276200	0.71819500
C	6.15477800	-1.61428800	-0.03885600
H	5.48688400	-3.65772000	-0.36334900
H	4.98801000	4.78532800	0.80532400
N	1.62890000	-1.84800200	-0.56060500
H	0.64194000	-2.01712800	-0.38644300
C	2.59958300	-2.87200200	-0.51509400
O	2.37336900	-4.06266600	-0.63069400

O	7.51654400	-1.78339200	0.08330000
O	6.99483100	0.45582600	0.31960700
C	8.07231500	-0.47008800	0.10261000
H	8.54610700	-0.24914500	-0.86497400
H	8.79014500	-0.38550000	0.92280100
C	-5.69431000	0.92245000	-0.39937800
O	-5.98981100	0.64417700	0.95171900
H	-6.17612800	1.85709800	-0.71797500
C	-6.23210100	-0.26712900	-1.20420200
C	-7.06341800	-0.32541000	1.04115500
C	-7.49170500	-0.62051200	-0.40655200
H	-5.52473100	-1.09787900	-1.16676300
H	-6.43878300	-0.00833100	-2.24727200
C	-6.56404200	-1.55672000	1.79713600
H	-7.89650000	0.14012900	1.58123800
H	-7.78025400	-1.67125700	-0.53110600
H	-7.42886200	-2.18361200	2.04861800
H	-6.11246800	-1.21643800	2.74274800
O	-5.68090200	-2.35488100	1.03933800
H	-4.84355900	-1.85289300	0.91288400
O	-8.57368300	0.26303500	-0.71036100
H	-8.86553800	0.07178500	-1.61542400

Table A17: Cartesian coordinates of the* anti *nucleoside minimum for the ALII-N^6-dA adduct with χ = 175.8°, θ = 355.5° and ϕ = 37.1°.

Calculated energy (in Hartrees) = -1782.74186900

Atom	X	Y	Z
N	-4.22958500	-0.11799200	-0.66625200
C	-3.81390700	1.05069200	-1.27860500
H	-4.53016900	1.71809300	-1.73412900
N	-2.50967800	1.22233600	-1.25270100
C	-2.04047100	0.10029700	-0.60093200
C	-0.73779500	-0.31383200	-0.24151100
N	0.31603200	0.48917700	-0.55822900
H	0.03290300	1.31995700	-1.06406600
N	-0.59565400	-1.48299100	0.40724700
C	-1.70110900	-2.19940100	0.69849300
H	-1.51735700	-3.13434000	1.22234100
N	-2.97510600	-1.91535600	0.42944300
C	-3.08565900	-0.74888900	-0.22832000
C	1.69209500	0.40069300	-0.25905300
C	2.40665700	-0.76735800	-0.21060200
C	2.40928600	1.66382200	-0.05389500
C	3.82347200	-0.72446400	-0.03211800

C	3.84015300	1.70807700	0.07774600
C	1.70769700	2.88915400	0.02206400
C	4.33762700	-2.02405700	-0.06143000
C	4.57425200	0.45691700	0.07340300
C	4.48465900	2.95166100	0.22648900
H	0.62261100	2.89453300	0.00121800
C	2.36725800	4.09909000	0.17289200
C	5.70947600	-2.25036200	0.04113400
C	5.94777700	0.20072300	0.17442700
C	3.76493600	4.13532500	0.26416500
H	5.56451300	2.97006400	0.31926200
H	1.79137900	5.01857100	0.23104800
C	6.47579000	-1.09510700	0.15790000
H	6.14935400	-3.24089500	0.02641600
H	4.28301600	5.08301500	0.37957300
N	2.05898100	-2.10336000	-0.37161100
H	1.10493300	-2.40873600	-0.19550700
C	3.17820900	-2.94817400	-0.22019500
O	3.14911400	-4.16584300	-0.22145000
O	7.84462800	-1.02975900	0.30541900
O	6.96685300	1.10774900	0.32176400
C	8.18142200	0.35115000	0.19182800
H	8.62385600	0.54935900	-0.79558300
H	8.86734900	0.63068800	0.99574800
C	-5.59312000	-0.63092700	-0.52968400
O	-6.47517200	0.36149100	-1.01930600
H	-5.67765200	-1.54105300	-1.13546400
C	-5.99584000	-0.88785500	0.92682500
C	-7.65043100	0.45146500	-0.17563100
C	-7.50914100	-0.67012000	0.86628600
H	-5.53905800	-0.13200500	1.57060000
H	-5.69767600	-1.88043200	1.27177300
C	-7.70505700	1.86401900	0.40039900
H	-8.53414600	0.26181000	-0.79509300
H	-7.92564500	-0.37125100	1.83787700
H	-8.58313900	1.96520100	1.04680500
H	-7.81214300	2.57918500	-0.42949900
O	-6.57227900	2.16771300	1.20222100
H	-5.80125700	2.14558000	0.61203100
O	-8.19344100	-1.80311900	0.33315500
H	-8.04364500	-2.55113000	0.93250100

Table A18: Cartesian coordinates of the* syn *nucleoside minimum for the ALII-N^6-dA adduct with χ = 55.3°, θ = 4.9° and ϕ = 321.7°.

Calculated energy (in Hartrees) = -1782.75243792

Atom	X	Y	Z
N	-4.20407204	0.87080911	-1.17774638
C	-3.71056818	2.13051469	-0.89218288
H	-4.37376995	2.98457952	-0.84291104
N	-2.41322406	2.16667300	-0.70107508
C	-2.02433883	0.85092614	-0.85814840
C	-0.75899353	0.22948566	-0.75488782
N	0.32452100	0.99426600	-0.45614800
H	0.09339570	1.97593647	-0.35769242
N	-0.68036955	-1.10022706	-0.94889823
C	-1.80982315	-1.77706465	-1.21881782
H	-1.69074739	-2.84735151	-1.36568211
N	-3.05336588	-1.30659019	-1.33016671
C	-3.11319038	0.02528347	-1.14586495
C	1.67226200	0.65895200	-0.19954700
C	2.08818943	-0.51937077	0.36192623
C	2.68949855	1.62130608	-0.63526464
C	3.48505476	-0.80787855	0.44880266
C	4.09189875	1.30855819	-0.59258085
C	2.31790212	2.89829800	-1.11508048
C	3.66683798	-2.07379253	1.01284237
C	4.49953495	0.03207539	-0.03669227
C	5.02621675	2.24908661	-1.06855813
H	1.27647634	3.20322619	-1.09433069
C	3.25919767	3.80688556	-1.57354851
C	4.94480169	-2.61060190	1.16240261
C	5.77192828	-0.53129359	0.12535885
C	4.62127065	3.47877924	-1.56302786
H	6.08022058	1.99799900	-1.03635469
H	2.93370095	4.77978640	-1.93115749
C	5.97001196	-1.79220907	0.69908869
H	5.13142429	-3.58536257	1.59817918
H	5.35940859	4.18851595	-1.92525201
N	1.41370918	-1.62992808	0.85460666
H	0.45925979	-1.54276599	1.19291889
C	2.31271690	-2.60170380	1.34482068
O	1.99645489	-3.63076343	1.91335521
O	7.32026164	-2.06039554	0.75225537
O	6.98531749	0.01712187	-0.20370306
C	7.95321342	-1.02486788	0.00320517
H	8.27721621	-1.41294913	-0.97348752

H	8.79811490	-0.62393267	0.56936836
C	-5.59884546	0.52531254	-1.41974515
O	-5.99000785	-0.45535010	-0.48405686
H	-6.17113661	1.45181246	-1.27379908
C	-5.87950142	-0.07369995	-2.80339080
C	-6.92109520	-1.38534633	-1.09143904
C	-7.15271031	-0.87963431	-2.52565670
H	-5.07066378	-0.74719646	-3.09339714
H	-6.01187901	0.69300577	-3.57301889
C	-6.33507426	-2.79570383	-1.02451256
H	-7.86214364	-1.33782885	-0.53053642
H	-7.27729050	-1.71312361	-3.22752181
H	-7.12508448	-3.51211911	-1.28293743
H	-6.03517592	-2.99132493	0.01745018
O	-5.27586255	-3.00207305	-1.93376354
H	-4.50832904	-2.45969282	-1.64077573
O	-8.31044350	-0.04172091	-2.49120981
H	-8.47803517	0.26724264	-3.39528364

Table A19: Cartesian coordinates of the* anti *nucleoside minimum for the ALII-N^6-dA adduct with χ = 174.2°, θ = 4.9° and ϕ = 322.5°.

Calculated energy (in Hartrees) = -1782.74184170

Atom	X	Y	Z
N	-4.11525211	-0.44508474	-1.28886789
C	-3.85162928	0.90556423	-1.14802385
H	-4.64693996	1.63212944	-1.22246431
N	-2.57987696	1.17423691	-0.94404592
C	-1.97176525	-0.06417562	-0.96558904
C	-0.62704125	-0.46474781	-0.79555871
N	0.31603200	0.48917700	-0.55822900
H	-0.07046288	1.42557943	-0.55739480
N	-0.33733299	-1.77582212	-0.86751049
C	-1.34263493	-2.64924972	-1.08335887
H	-1.04144971	-3.69291305	-1.13180233
N	-2.64236991	-2.40301591	-1.24520979
C	-2.90040251	-1.08593638	-1.18009388
C	1.69209500	0.40069300	-0.25905300
C	2.27924681	-0.63864740	0.41348977
C	2.55273017	1.47312920	-0.77006699
C	3.70144636	-0.68749242	0.54007483
C	3.98577732	1.39849408	-0.68361643
C	1.99316626	2.62535898	-1.36897448
C	4.07038061	-1.85277186	1.21843157
C	4.57977067	0.25900935	-0.01051317

C	4.76833169	2.43360872	-1.23148499
H	0.91604774	2.75682265	-1.38917631
C	2.78719795	3.63160730	-1.89688822
C	5.41394120	-2.16118866	1.42682309
C	5.92169017	-0.07537978	0.21275375
C	4.18368002	3.53284636	-1.84004080
H	5.84777288	2.36124590	-1.16365811
H	2.31837573	4.50314514	-2.34532057
C	6.30565524	-1.23242426	0.89982373
H	5.74326972	-3.05137017	1.95025837
H	4.80688453	4.31934455	-2.25582798
N	1.78037364	-1.79523154	1.00097810
H	0.81359061	-1.83399146	1.31451111
C	2.81058492	-2.56121088	1.58514347
O	2.65139193	-3.57392473	2.24289278
O	7.68053123	-1.27304591	0.98822944
O	7.03966866	0.63053968	-0.15429131
C	8.15444688	-0.22496776	0.14550829
H	8.54627727	-0.64581169	-0.79224139
H	8.92022115	0.35226414	0.67021566
C	-5.40240745	-1.09269901	-1.54344151
O	-6.40353518	-0.09429944	-1.48360439
H	-5.37023219	-1.53319577	-2.54712063
C	-5.76945722	-2.13811657	-0.48420101
C	-7.58087102	-0.60141287	-0.80697094
C	-7.29815131	-2.08560506	-0.52280564
H	-5.41277701	-1.80705124	0.49435929
H	-5.34741272	-3.12083111	-0.70612569
C	-7.81503860	0.25940578	0.43168162
H	-8.43310754	-0.51899540	-1.49075062
H	-7.74953465	-2.40651056	0.42590932
H	-8.69899978	-0.09984177	0.96895281
H	-8.01213073	1.29352423	0.11010080
O	-6.73032709	0.20009482	1.34722393
H	-5.96263213	0.58341644	0.89225358
O	-7.83250017	-2.81547589	-1.62640963
H	-7.58881258	-3.74768814	-1.51339810

Table A20: Cartesian coordinates of the* syn *nucleotide minimum for the ALI-N^6-dA adduct with χ = 70.9°, θ = 349.2° and ϕ = 32.8°.

Calculated energy (in Hartrees) = -2626.84217481

Atom	X	Y	Z
N	2.90830700	-2.26296100	-0.77574700
C	2.07154000	-3.35915700	-0.69106900
H	2.47770600	-4.36241000	-0.69097300
N	0.79658300	-3.05898300	-0.62292300
C	0.78114400	-1.67691700	-0.65489200
C	-0.28664400	-0.74690900	-0.59985900
N	-1.55741600	-1.23376500	-0.49382200
H	-1.56707700	-2.24264400	-0.36361100
N	0.02117000	0.56229700	-0.64539400
C	1.31495200	0.92597100	-0.73896700
H	1.49187400	1.99821100	-0.77601000
N	2.39374500	0.14930600	-0.78980900
C	2.07560100	-1.15557100	-0.74394700
C	-2.81803600	-0.64454200	-0.27404100
C	-3.14149100	0.64067200	-0.64771600
C	-3.88524000	-1.46916500	0.35011700
C	-4.44729600	1.16012200	-0.42762700
C	-5.20527100	-0.92701100	0.57036600
C	-3.70144900	-2.83361200	0.76383900
C	-4.52493700	2.47216000	-0.91109600
C	-5.49256400	0.43584400	0.15960800
C	-6.21958700	-1.69946600	1.17343900
C	-4.72215000	-3.56886000	1.35659400
C	-5.71672400	3.19444900	-0.81860200
C	-6.67645000	1.18542800	0.24232000
C	-5.98169900	-3.00084800	1.56446700
H	-7.19479200	-1.25507100	1.32487300
H	-4.54392200	-4.59281000	1.65851600
C	-6.76732900	2.50164900	-0.22976500
H	-5.82657400	4.21331800	-1.17340900
H	-6.76720300	-3.58961400	2.02852000
N	-2.42606200	1.64548600	-1.30585500
H	-1.40439900	1.62059000	-1.31632000
C	-3.18448200	2.81002100	-1.44909400
O	-2.77615900	3.86726100	-1.92544700
O	-8.03197300	2.98632700	0.00904900
O	-7.87783900	0.81252500	0.77756400
C	-8.80081700	1.88060000	0.48964400
H	-9.50334800	1.54811000	-0.28368200

H	-9.32042300	2.16270400	1.40738900
O	-2.47369700	-3.40286000	0.54922600
C	-2.22669600	-4.76756000	0.89293800
H	-2.34224200	-4.92784300	1.96997400
H	-2.89389600	-5.43904100	0.34202100
H	-1.19217300	-4.94455100	0.59956900
C	4.35839800	-2.34536400	-0.81632800
O	4.88953900	-1.96038800	0.44791500
H	4.58835200	-3.39915300	-1.01815000
C	5.06432400	-1.44057100	-1.82655600
C	6.25917400	-1.54375300	0.26560900
C	6.48949500	-1.42566100	-1.26039900
H	4.62869200	-0.44024600	-1.78617700
H	5.01564600	-1.82335200	-2.84961100
C	6.51955900	-0.27776300	1.06348300
H	6.92429900	-2.32697200	0.65228700
H	7.02792800	-0.50781400	-1.52104300
H	7.59940300	-0.09031600	1.10168200
H	6.15289200	-0.41699700	2.08741500
O	5.85509500	0.82989500	0.44509800
O	7.23911000	-2.57916300	-1.65008600
H	7.40561100	-2.51005700	-2.60426100
P	5.76227000	2.26262600	1.23560900
O	7.12176100	2.71460400	1.73456300
O	5.02202600	3.20088600	0.31145100
O	4.81779100	1.87553700	2.52808400
Na	6.85716300	4.73341900	0.46675500
H	5.32932100	2.00720400	3.34328500

Table A21: Cartesian coordinates of the* anti *nucleotide minimum for the ALI-N^6-dA adduct with χ = 248.5°, θ = 348.0° and ϕ = 34.7°.

Calculated energy (in Hartrees) = -2626.84639669

Atom	X	Y	Z
N	2.81516800	-1.54813200	-0.31607700
C	2.71054300	-0.19503100	-0.06338200
H	3.59670300	0.41753500	0.03112100
N	1.47003800	0.22393200	0.06062900
C	0.71031500	-0.91839800	-0.11392400
C	-0.68751700	-1.14280900	-0.10603900
N	-1.51124800	-0.07254600	0.09564700
H	-0.99816600	0.80535100	0.09997200
N	-1.12126300	-2.40185600	-0.30793300
C	-0.21814600	-3.38183300	-0.51104100

H	-0.64006500	-4.37278100	-0.66157600
N	1.10935600	-3.29184000	-0.55200200
C	1.52388300	-2.03154400	-0.34247400
C	-2.90651200	0.11018700	0.04261800
C	-3.81941300	-0.88129200	0.32329500
C	-3.43389000	1.46200500	-0.27542100
C	-5.21659200	-0.61707200	0.30192300
C	-4.85520500	1.71511900	-0.29680300
C	-2.59984500	2.59517300	-0.56958800
C	-5.92476800	-1.77457600	0.64832700
C	-5.77491100	0.63474100	0.01294000
C	-5.36322400	2.99305300	-0.60981600
C	-3.13191900	3.84291100	-0.87611000
C	-7.31982600	-1.76081200	0.71515300
C	-7.17665400	0.61949100	0.08458300
C	-4.51480600	4.04184100	-0.89886500
H	-6.43529600	3.14157700	-0.61885000
H	-2.47136700	4.67122300	-1.09754600
C	-7.89959500	-0.53265400	0.42190700
H	-7.91395400	-2.63015200	0.97471300
H	-4.91306900	5.02279300	-1.13999900
N	-3.67770900	-2.21469400	0.71955100
H	-2.80654100	-2.71295600	0.52680400
C	-4.91784300	-2.83964200	0.87748000
O	-5.08297100	-4.02760600	1.14694800
O	-9.24586400	-0.25376700	0.38566800
O	-8.04720900	1.64474600	-0.16020900
C	-9.35698700	1.15846000	0.19131300
H	-9.67996500	1.64073100	1.12154300
H	-10.04836200	1.36554600	-0.62770800
O	-1.24235400	2.41088000	-0.53501300
C	-0.34914600	3.50168800	-0.76723800
H	-0.46897700	3.89789500	-1.78105700
H	-0.50339200	4.30037800	-0.03388100
H	0.64703300	3.07555700	-0.65001700
C	4.04150900	-2.32466300	-0.44819100
O	4.82945200	-2.12503500	0.72023800
H	3.72225800	-3.36908700	-0.51810900
C	4.92674300	-1.91885500	-1.63404200
C	6.22489600	-1.99090600	0.37973600
C	6.32830400	-2.27329600	-1.12972900
H	4.85521000	-0.84240300	-1.80829700
H	4.65641700	-2.44469300	-2.55373300
C	6.74915300	-0.63497100	0.82646200
H	6.78440400	-2.75951100	0.92586000
H	7.10806900	-1.66794000	-1.60731400

H	7.83823100	-0.60956700	0.69948500
H	6.51679300	-0.48894400	1.88832000
O	6.14388800	0.40784300	0.04859900
O	6.61090200	-3.66715600	-1.26321800
H	6.63059300	-3.87239500	-2.21215200
P	6.71098500	1.94430400	0.26223000
O	6.28128300	2.48940600	1.61031600
O	8.19220900	1.99944400	-0.04161300
O	5.88746300	2.66849700	-0.95098300
Na	8.55561400	3.02316200	2.09472300
H	5.05081500	3.03254900	-0.61596800

Table A22: Cartesian coordinates of the* syn *nucleotide minimum for the ALI-N^6-dA adduct with χ = 73.3°, θ = 10.5° and ϕ = 328.0°.

Calculated energy (in Hartrees) = -2626.84227771

Atom	X	Y	Z
N	2.95296400	2.36950100	0.06012600
C	2.13346000	3.35418200	-0.45895600
H	2.53694800	4.32285200	-0.72397000
N	0.87630700	2.99772500	-0.57829500
C	0.85964000	1.69068400	-0.12508200
C	-0.19660400	0.75737600	0.01226100
N	-1.44826500	1.15401300	-0.35480600
H	-1.47462000	2.14276400	-0.59091000
N	0.09946200	-0.46124000	0.50088100
C	1.37519000	-0.73218100	0.84120300
H	1.55039700	-1.73534700	1.22202200
N	2.43950300	0.06356500	0.77155800
C	2.13324500	1.27307700	0.27251700
C	-2.73113100	0.60027800	-0.18888000
C	-2.97751700	-0.75221800	-0.15369900
C	-3.89587400	1.51667600	-0.10689400
C	-4.30478900	-1.24913800	-0.04492900
C	-5.23899000	0.99912400	0.00039000
C	-3.78265500	2.94898000	-0.13459200
C	-4.28963600	-2.64910700	-0.06088700
C	-5.44447600	-0.43763100	0.02639400
C	-6.34822900	1.86657900	0.08181000
C	-4.89586800	3.77807900	-0.05380400
C	-5.48084800	-3.37350700	0.01549200
C	-6.62592600	-1.19008900	0.10853800
C	-6.17960800	3.23596000	0.05617300
H	-7.34067900	1.44310200	0.16525200

H	-4.77215300	4.85311300	-0.07616900
C	-6.62673200	-2.59191900	0.10218900
H	-5.52243200	-4.45681800	0.00381800
H	-7.03833400	3.89746000	0.11907100
N	-2.14564700	-1.86921100	-0.27131300
H	-1.14729800	-1.78329400	-0.07090200
C	-2.86767800	-3.06100600	-0.15998600
O	-2.37766000	-4.18811700	-0.14954200
O	-7.92323400	-3.04770300	0.15803600
O	-7.91312700	-0.73578500	0.17674100
C	-8.73906600	-1.89655500	0.39662600
H	-9.08901700	-1.89430500	1.43633700
H	-9.57284800	-1.88212300	-0.30793500
O	-2.52485300	3.47856900	-0.24638000
C	-2.31734100	4.88983000	-0.31864800
H	-2.82568000	5.32000800	-1.18882400
H	-2.66158600	5.38391200	0.59651000
H	-1.23803600	5.00292400	-0.42136200
C	4.39349800	2.48977500	0.18088600
O	5.00628400	1.81686300	-0.91541600
H	4.61200700	3.56400100	0.13449600
C	5.02789700	1.87093900	1.42521300
C	6.34239700	1.42506500	-0.53440800
C	6.48538300	1.75489600	0.96804000
H	4.60650400	0.87930200	1.59548900
H	4.90599300	2.48995000	2.31806500
C	6.56303800	-0.03454000	-0.88885200
H	7.06419000	2.02906800	-1.09866100
H	7.03035100	0.97116800	1.50681300
H	7.63339700	-0.26774800	-0.83507700
H	6.21766900	-0.20557000	-1.91537800
O	5.84072000	-0.86174100	0.03096500
O	7.17380100	3.00616300	1.04717700
H	7.24079400	3.24420500	1.98612400
P	5.51502600	-2.41084600	-0.39112400
O	6.76052500	-3.11047200	-0.90339800
O	4.80504900	-3.03121600	0.78951000
O	4.46665300	-2.20719800	-1.64559200
Na	6.50252500	-4.74789000	0.84370500
H	4.87305800	-2.56616200	-2.45126800

Table A23: Cartesian coordinates of the* anti *nucleotide minimum for the ALI-N^6-dA adduct with χ = 248.1°, θ = 11.5° and ϕ = 327.5°.

Calculated energy (in Hartrees) = -2626.84659381

Atom	X	Y	Z
N	-2.88058800	-1.69288900	0.05927000
C	-2.77237400	-0.36551100	0.42958500
H	-3.65101200	0.24223700	0.58791400
N	-1.52853200	0.04815400	0.53951200
C	-0.77241300	-1.06482700	0.21702400
C	0.62384900	-1.26600900	0.12084500
N	1.43819500	-0.20812500	0.40129000
H	0.90303300	0.64351000	0.55077700
N	1.06069300	-2.48466000	-0.25067300
C	0.15526400	-3.44882900	-0.51446900
H	0.57718400	-4.40803800	-0.80450500
N	-1.17432600	-3.37689400	-0.46737400
C	-1.58951200	-2.15446400	-0.09478900
C	2.80992000	0.03918600	0.21503600
C	3.77368000	-0.93975200	0.27247000
C	3.25768200	1.43838400	0.00502700
C	5.15108700	-0.61531200	0.14349200
C	4.66014800	1.75561700	-0.11997200
C	2.35721900	2.55530300	-0.08213100
C	5.92472200	-1.77578600	0.26791000
C	5.63800100	0.68530600	-0.04165400
C	5.08976400	3.08412400	-0.32019200
C	2.81189900	3.85413200	-0.28019300
C	7.31761100	-1.71191800	0.19365300
C	7.03852900	0.71974100	-0.11899200
C	4.17940200	4.11827000	-0.40007400
H	6.14927300	3.28489900	-0.41214500
H	2.10390600	4.67033000	-0.34384400
C	7.82669000	-0.43408000	-0.00481800
H	7.96101400	-2.57954900	0.28686800
H	4.51763900	5.13835500	-0.55501100
N	3.71139900	-2.31529200	0.51249500
H	2.83687500	-2.82012400	0.35336200
C	4.97873500	-2.90259200	0.46156200
O	5.20679100	-4.10641500	0.56027600
O	9.15573900	-0.08802600	-0.08273600
O	7.84892600	1.80871500	-0.27943300
C	9.18450500	1.29431000	-0.45100600
H	9.47398100	1.39572100	-1.50436800
H	9.86542000	1.83767400	0.20685200

O	1.01806300	2.29494200	0.03691400
C	0.05144000	3.34666800	0.02478200
H	0.23112500	4.05769800	0.83894800
H	0.05526800	3.87210100	-0.93652900
H	-0.90412800	2.84227400	0.17112800
C	-4.10543400	-2.42489900	-0.22861200
O	-4.75014500	-1.82474300	-1.34840900
H	-3.79084400	-3.44121100	-0.48329800
C	-5.12393200	-2.41273100	0.91914300
C	-6.16689000	-1.69984500	-1.11417000
C	-6.45650700	-2.49030900	0.17184000
H	-5.06381200	-1.47460500	1.47411300
H	-4.97470100	-3.24523200	1.61145600
C	-6.56404300	-0.23324800	-1.07789800
H	-6.69539100	-2.16913100	-1.95234200
H	-7.28320000	-2.05064000	0.74408700
H	-7.65738900	-0.15874800	-1.01307100
H	-6.23831600	0.25789600	-1.99978600
O	-5.94709500	0.39412100	0.05507300
O	-6.76191800	-3.82563500	-0.23429900
H	-6.87188300	-4.35916700	0.56945000
P	-6.17992600	2.00786100	0.30308100
O	-6.62113400	2.65947500	-0.99171300
O	-7.04954500	2.22465700	1.51995600
O	-4.65283900	2.45766200	0.71090400
Na	-8.57894200	3.48402100	0.13279900
H	-4.11367400	2.56594700	-0.09132400

Table A24: Cartesian coordinates of the* syn *nucleotide minimum for the ALII-N^6-dA adduct with χ = 71.9°, θ = 357.1°and ϕ = 38.4°.

Calculated energy (in Hartrees) = -2512.31513544

Atom	X	Y	Z
N	2.98473900	-2.33572300	-0.90388700
C	2.21846100	-3.48485500	-0.83914700
H	2.68413300	-4.45971900	-0.90490100
N	0.93168100	-3.26566000	-0.70336600
C	0.83753400	-1.88667900	-0.66312700
C	-0.26985400	-1.02392600	-0.51093300
N	-1.51640400	-1.56036300	-0.36775200
H	-1.51910700	-2.57322300	-0.39268000
N	-0.05400200	0.30328400	-0.49682600
C	1.21300900	0.74713500	-0.62630400
H	1.32979300	1.82797000	-0.61094600

N	2.33099500	0.03850700	-0.76900900
C	2.09399900	-1.28316700	-0.77778500
C	-2.76433500	-0.95019100	-0.13363800
C	-3.16803100	0.23452800	-0.69070900
C	-3.71767900	-1.67254600	0.71865800
C	-4.50243900	0.69885900	-0.49332700
C	-5.07546000	-1.22761800	0.88062600
C	-3.32100700	-2.83670600	1.41593400
C	-4.69194200	1.89395600	-1.19498400
C	-5.48372600	0.00432900	0.23185700
C	-5.96073200	-1.98181600	1.67690100
H	-2.28917400	-3.16818000	1.37437200
C	-4.21190200	-3.55752800	2.19698900
C	-5.93774600	2.52653700	-1.20033200
C	-6.72149200	0.66061800	0.21074000
C	-5.54343200	-3.13645300	2.32087600
H	-6.98358200	-1.64065600	1.78746000
H	-3.86823900	-4.44533700	2.71977800
C	-6.92364800	1.86541800	-0.47684000
H	-6.13378100	3.45408500	-1.72612600
H	-6.24300300	-3.70179500	2.92940400
N	-2.54600100	1.14791200	-1.54213700
H	-1.53089300	1.21866900	-1.55867000
C	-3.39369100	2.22838300	-1.83431200
O	-3.07512800	3.21762400	-2.48518600
O	-8.22106600	2.28040000	-0.29408000
O	-7.87954000	0.28602900	0.82902600
C	-8.89180500	1.21175000	0.38343500
H	-9.56791300	0.69608000	-0.30907800
H	-9.42706800	1.60313600	1.25082600
C	4.43511100	-2.32014100	-0.94324400
O	4.92828800	-2.02049000	0.35901000
H	4.73965600	-3.33157900	-1.23997100
C	5.07848000	-1.27874500	-1.85782500
C	6.24049500	-1.43121600	0.23824100
C	6.49268800	-1.22324900	-1.27115700
H	4.57856400	-0.31797700	-1.72621700
H	5.06229200	-1.57042600	-2.91125700
C	6.30777700	-0.16472200	1.07277000
H	6.98394700	-2.13881700	0.62645200
H	6.99003100	-0.26729500	-1.47178100
H	7.35408100	0.14317600	1.18873400
H	5.88981300	-0.37213100	2.06504700
O	5.56623600	0.87412900	0.42261200
O	7.29764700	-2.32226000	-1.70700400
H	7.44107100	-2.21845100	-2.66171000

P	5.11423700	2.18310100	1.29789100
O	4.39672200	3.10710500	0.34168400
O	6.28896300	2.75549600	2.07001800
O	4.04728600	1.52524900	2.36706900
H	4.40603300	1.62455800	3.26410400
Na	5.96778800	4.84524800	0.91766100

Table A25: Cartesian coordinates of the* anti *nucleotide minimum for the ALII-N^6-dA adduct with χ = 248.1°, θ = 358.5 °and ϕ = 37.8°.

Calculated energy (in Hartrees) = -2512.31894478

Atom	X	Y	Z
N	2.83838900	-1.27032300	-0.61682100
C	2.74225000	0.08189800	-0.34193000
H	3.62884400	0.69047600	-0.23461100
N	1.50260900	0.50577400	-0.21964000
C	0.73802700	-0.63032900	-0.41466600
C	-0.65552800	-0.85010800	-0.42006100
N	-1.49975300	0.20289100	-0.21576200
H	-1.01465500	1.08088100	-0.07340500
N	-1.10839800	-2.09833500	-0.63950200
C	-0.20949700	-3.08101400	-0.85346000
H	-0.63666700	-4.06703600	-1.01955800
N	1.12035200	-2.99965400	-0.88584100
C	1.54469600	-1.74636800	-0.65586200
C	-2.90643000	0.27280200	-0.20926200
C	-3.73167200	-0.69576400	0.29883600
C	-3.52183800	1.50180100	-0.72623200
C	-5.13756000	-0.46746000	0.37823500
C	-4.93280800	1.75073600	-0.60419800
C	-2.73519500	2.48414100	-1.37044900
C	-5.75900200	-1.57013700	0.97374000
C	-5.76986700	0.72373000	-0.01242000
C	-5.46462600	2.96670900	-1.07868800
H	-1.67931800	2.30512400	-1.54276900
C	-3.28554200	3.66999800	-1.83328700
C	-7.13615100	-1.56970900	1.20966800
C	-7.14906500	0.69629900	0.23349700
C	-4.65608700	3.92081700	-1.67710000
H	-6.52840500	3.14740700	-0.97568100
H	-2.64888800	4.39898900	-2.32610500
C	-7.78823100	-0.40627500	0.81766900
H	-7.66396800	-2.39991200	1.66517800
H	-5.08775900	4.85064000	-2.03546500
N	-3.49365500	-1.94399900	0.87463400

H	-2.64833400	-2.45993700	0.63993800
C	-4.69468200	-2.57005600	1.24430700
O	-4.78729800	-3.70842300	1.69078400
O	-9.13662800	-0.15486400	0.90605100
O	-8.07072000	1.66340000	-0.04822500
C	-9.31239700	1.21663600	0.53359400
H	-9.52830100	1.82100900	1.42298500
H	-10.10715400	1.29778800	-0.21045500
C	4.05214000	-2.06519800	-0.73721300
O	4.73694400	-2.05974700	0.51200000
H	3.71730800	-3.08014600	-0.97062100
C	5.04430100	-1.53872300	-1.78223600
C	6.15101100	-1.86178800	0.31519000
C	6.38968700	-1.97507800	-1.19892700
H	5.00063700	-0.44932900	-1.83665000
H	4.85382900	-1.95257700	-2.77576200
C	6.58397600	-0.54651700	0.94132300
H	6.68533500	-2.67122500	0.82617600
H	7.21683100	-1.33492000	-1.53119800
H	7.67613100	-0.45615600	0.86584500
H	6.30646800	-0.53216900	2.00016200
O	5.93686600	0.53435700	0.25581800
O	6.66458600	-3.35211300	-1.46213500
H	6.75129900	-3.45624300	-2.42350700
P	6.32556000	2.07380000	0.69928500
O	4.92562700	2.82277800	0.28916700
O	7.42329100	2.61654600	-0.18799600
O	6.56395800	2.12092800	2.19547300
H	4.29154400	2.75684500	1.02346400
Na	8.77841600	2.95453300	1.77776200

Table A26: Cartesian coordinates of the* syn *nucleotide minimum for the ALII-N^6-dA adduct with χ = 71.2°, θ = 0.9° and ϕ = 320.6°.

Calculated energy (in Hartrees) = -2512.31517407

Atom	X	Y	Z
N	2.95296400	2.36950100	0.06012600
C	2.13346000	3.35418200	-0.45895600
H	2.53694800	4.32285200	-0.72397000
N	0.87630700	2.99772500	-0.57829500
C	0.85964000	1.69068400	-0.12508200
C	-0.19660400	0.75737600	0.01226100
N	-1.44826500	1.15401300	-0.35480600
H	-1.47462000	2.14276400	-0.59091000
N	0.09946200	-0.46124000	0.50088100

C	1.37519000	-0.73218100	0.84120300
H	1.55039700	-1.73534700	1.22202200
N	2.43950300	0.06356500	0.77155800
C	2.13324500	1.27307700	0.27251700
C	-2.73113100	0.60027800	-0.18888000
C	-2.97751700	-0.75221800	-0.15369900
C	-3.89587400	1.51667600	-0.10689400
C	-4.30478900	-1.24913800	-0.04492900
C	-5.23899000	0.99912400	0.00039000
C	-3.78265500	2.94898000	-0.13459200
C	-4.28963600	-2.64910700	-0.06088700
C	-5.44447600	-0.43763100	0.02639400
C	-6.34822900	1.86657900	0.08181000
C	-4.89586800	3.77807900	-0.05380400
C	-5.48084800	-3.37350700	0.01549200
C	-6.62592600	-1.19008900	0.10853800
C	-6.17960800	3.23596000	0.05617300
H	-7.34067900	1.44310200	0.16525200
H	-4.77215300	4.85311300	-0.07616900
C	-6.62673200	-2.59191900	0.10218900
H	-5.52243200	-4.45681800	0.00381800
H	-7.03833400	3.89746000	0.11907100
N	-2.14564700	-1.86921100	-0.27131300
H	-1.14729800	-1.78329400	-0.07090200
C	-2.86767800	-3.06100600	-0.15998600
O	-2.37766000	-4.18811700	-0.14954200
O	-7.92323400	-3.04770300	0.15803600
O	-7.91312700	-0.73578500	0.17674100
C	-8.73906600	-1.89655500	0.39662600
H	-9.08901700	-1.89430500	1.43633700
H	-9.57284800	-1.88212300	-0.30793500
O	-2.52485300	3.47856900	-0.24638000
C	-2.31734100	4.88983000	-0.31864800
H	-2.82568000	5.32000800	-1.18882400
H	-2.66158600	5.38391200	0.59651000
H	-1.23803600	5.00292400	-0.42136200
C	4.39349800	2.48977500	0.18088600
O	5.00628400	1.81686300	-0.91541600
H	4.61200700	3.56400100	0.13449600
C	5.02789700	1.87093900	1.42521300
C	6.34239700	1.42506500	-0.53440800
C	6.48538300	1.75489600	0.96804000
H	4.60650400	0.87930200	1.59548900
H	4.90599300	2.48995000	2.31806500
C	6.56303800	-0.03454000	-0.88885200
H	7.06419000	2.02906800	-1.09866100

H	7.03035100	0.97116800	1.50681300
H	7.63339700	-0.26774800	-0.83507700
H	6.21766900	-0.20557000	-1.91537800
O	5.84072000	-0.86174100	0.03096500
O	7.17380100	3.00616300	1.04717700
H	7.24079400	3.24420500	1.98612400
P	5.51502600	-2.41084600	-0.39112400
O	6.76052500	-3.11047200	-0.90339800
O	4.80504900	-3.03121600	0.78951000
O	4.46665300	-2.20719800	-1.64559200
Na	6.50252500	-4.74789000	0.84370500
H	4.87305800	-2.56616200	-2.45126800

Table A27: Cartesian coordinates of the* anti *nucleotide minimum for the ALII-N^6-dA adduct with χ = 249.1°, θ = 3.5° and ϕ = 321.7°.

Calculated energy (in Hartrees) = -2512.31910091

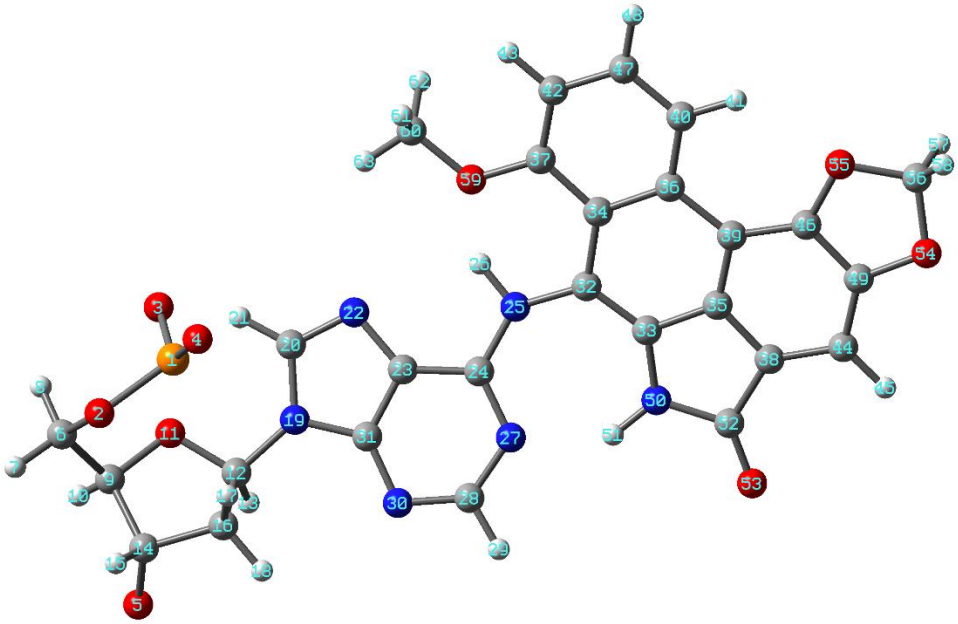
Atom	X	Y	Z
N	-2.88058800	-1.69288900	0.05927000
C	-2.77237400	-0.36551100	0.42958500
H	-3.65101200	0.24223700	0.58791400
N	-1.52853200	0.04815400	0.53951200
C	-0.77241300	-1.06482700	0.21702400
C	0.62384900	-1.26600900	0.12084500
N	1.43819500	-0.20812500	0.40129000
H	0.90303300	0.64351000	0.55077700
N	1.06069300	-2.48466000	-0.25067300
C	0.15526400	-3.44882900	-0.51446900
H	0.57718400	-4.40803800	-0.80450500
N	-1.17432600	-3.37689400	-0.46737400
C	-1.58951200	-2.15446400	-0.09478900
C	2.80992000	0.03918600	0.21503600
C	3.77368000	-0.93975200	0.27247000
C	3.25768200	1.43838400	0.00502700
C	5.15108700	-0.61531200	0.14349200
C	4.66014800	1.75561700	-0.11997200
C	2.35721900	2.55530300	-0.08213100
C	5.92472200	-1.77578600	0.26791000
C	5.63800100	0.68530600	-0.04165400
C	5.08976400	3.08412400	-0.32019200
C	2.81189900	3.85413200	-0.28019300
C	7.31761100	-1.71191800	0.19365300
C	7.03852900	0.71974100	-0.11899200
C	4.17940200	4.11827000	-0.40007400
H	6.14927300	3.28489900	-0.41214500

H	2.10390600	4.67033000	-0.34384400
C	7.82669000	-0.43408000	-0.00481800
H	7.96101400	-2.57954900	0.28686800
H	4.51763900	5.13835500	-0.55501100
N	3.71139900	-2.31529200	0.51249500
H	2.83687500	-2.82012400	0.35336200
C	4.97873500	-2.90259200	0.46156200
O	5.20679100	-4.10641500	0.56027600
O	9.15573900	-0.08802600	-0.08273600
O	7.84892600	1.80871500	-0.27943300
C	9.18450500	1.29431000	-0.45100600
H	9.47398100	1.39572100	-1.50436800
H	9.86542000	1.83767400	0.20685200
O	1.01806300	2.29494200	0.03691400
C	0.05144000	3.34666800	0.02478200
H	0.23112500	4.05769800	0.83894800
H	0.05526800	3.87210100	-0.93652900
H	-0.90412800	2.84227400	0.17112800
C	-4.10543400	-2.42489900	-0.22861200
O	-4.75014500	-1.82474300	-1.34840900
H	-3.79084400	-3.44121100	-0.48329800
C	-5.12393200	-2.41273100	0.91914300
C	-6.16689000	-1.69984500	-1.11417000
C	-6.45650700	-2.49030900	0.17184000
H	-5.06381200	-1.47460500	1.47411300
H	-4.97470100	-3.24523200	1.61145600
C	-6.56404300	-0.23324800	-1.07789800
H	-6.69539100	-2.16913100	-1.95234200
H	-7.28320000	-2.05064000	0.74408700
H	-7.65738900	-0.15874800	-1.01307100
H	-6.23831600	0.25789600	-1.99978600
O	-5.94709500	0.39412100	0.05507300
O	-6.76191800	-3.82563500	-0.23429900
H	-6.87188300	-4.35916700	0.56945000
P	-6.17992600	2.00786100	0.30308100
O	-6.62113400	2.65947500	-0.99171300
O	-7.04954500	2.22465700	1.51995600
O	-4.65283900	2.45766200	0.71090400
Na	-8.57894200	3.48402100	0.13279900
H	-4.11367400	2.56594700	-0.09132400

Table A28: The volume of the 8 Å water box, the number of water residues and sodium ions added in each MD simulation.

Conformation	Volume of water box (in Å^3)		No. of water molecules		No. of Na^+ ions
	ALI-N^6-dA	ALI-N^6-dA	ALI-N^6-dA	ALI-N^6-dA	
anti base-displaced	159764.86	159764.86	4148	3939	20
anti 5′–intercalated	138728.63	138728.63	3525	3568	20
anti 3′–intercalated	184275.24	184275.24	4926	3628	20
syn base-displaced	165984.71	165984.71	4370	3994	20
syn 5′–intercalated	142583.02	142583.02	3654	5573	20
syn 3′–intercalated	184453.93	184453.93	4935	3246	20

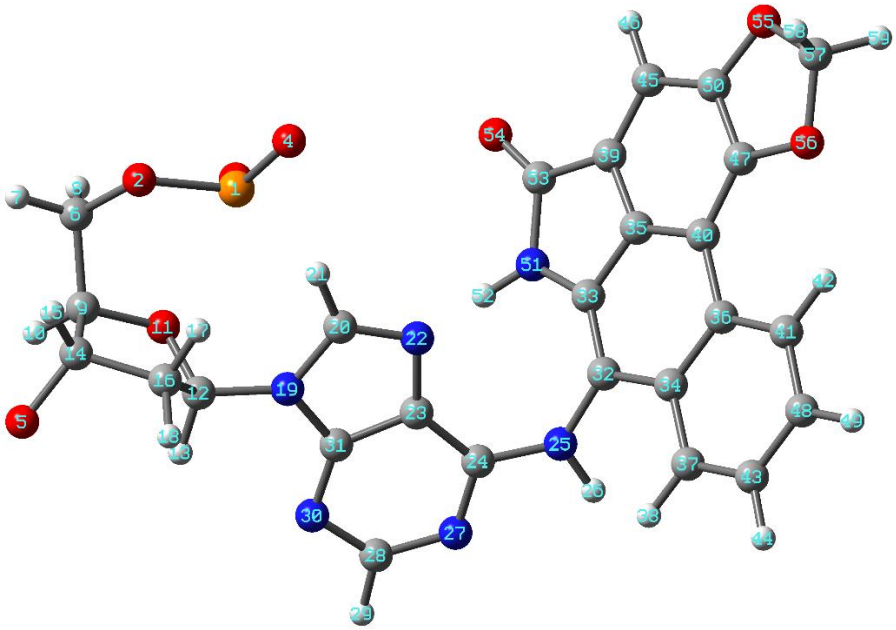
Table A29: Mol2 file of the ALI-N^6-dA adduct generated using the RED program (atom numbering provided in the figure below).



1	P	4.8440	0.5110	-1.3820	P	1 LIG 1.166200
2	O5′	4.1730	1.1990	0.0000	OS	1 LIG -0.487900
3	O1	4.7530	1.4920	-2.5100	O2	1 LIG -0.768300
4	O2	6.0930	-0.1790	-0.9240	O2	1 LIG -0.768300
5	O3′	0.0000	0.0000	0.0000	OS	1 LIG -0.543300
6	C5′	3.0320	2.0210	-0.1970	CI	1 LIG -0.023900
7	H5′1	2.7490	2.4080	0.7870	H1	1 LIG 0.081900
8	H5′2	3.2640	2.8810	-0.8430	H1	1 LIG 0.081900
9	C4′	1.8620	1.2500	-0.7970	CT	1 LIG 0.187300
10	H4′	0.9940	1.9100	-0.9150	H1	1 LIG 0.100600
11	O4′	2.2760	0.7780	-2.1050	OS	1 LIG -0.423100
12	C1′	1.9350	-0.5910	-2.2360	CT	1 LIG 0.233400
13	H1′	0.9210	-0.6970	-2.6440	H2	1 LIG 0.082100
14	C3′	1.4270	-0.0000	-0.0000	CT	1 LIG 0.133700
15	H3′	1.8220	0.0210	1.0240	H1	1 LIG 0.073600
16	C2′	2.0120	-1.1580	-0.8170	CT	1 LIG -0.043600
17	H2′1	3.0610	-1.3140	-0.5460	HC	1 LIG 0.040700
18	H2′2	1.4700	-2.1010	-0.7080	HC	1 LIG 0.040700
19	N9	2.8710	-1.1920	-3.1810	N*	1 LIG -0.128800
20	C8	3.8400	-0.5670	-3.9410	CK	1 LIG 0.066400
21	H8	4.0060	0.4970	-3.8600	H5	1 LIG 0.186100
22	N7	4.4870	-1.3830	-4.7440	NB	1 LIG -0.459500
23	C5	3.9050	-2.6140	-4.5130	CB	1 LIG 0.017700
24	C6	4.1480	-3.9040	-5.0540	CA	1 LIG 0.472200

25 N6	5.1330	-4.0300	-5.9850	N2	1 LIG	-0.337500
26 H61	5.6350	-3.1570	-6.1160	H	1 LIG	0.211400
27 N1	3.3860	-4.9200	-4.6040	NC	1 LIG	-0.669900
28 C2	2.4460	-4.6770	-3.6720	CQ	1 LIG	0.496700
29 H2	1.8700	-5.5460	-3.3620	H5	1 LIG	0.073800
30 N3	2.1330	-3.5230	-3.0860	NC	1 LIG	-0.670400
31 C4	2.8980	-2.5210	-3.5500	CB	1 LIG	0.453500
32 C10	5.7630	-5.1210	-6.6230	CA	1 LIG	0.080500
33 C11	5.1590	-6.3400	-6.8350	CN	1 LIG	0.113000
34 C12	7.1410	-4.9320	-7.1430	CA	1 LIG	0.024800
35 C13	5.8470	-7.3830	-7.5190	CB	1 LIG	-0.049400
36 C14	7.8260	-6.0000	-7.8310	CA	1 LIG	-0.003500
37 C15	7.8830	-3.7080	-7.0220	CA	1 LIG	0.080700
38 C16	5.0210	-8.5080	-7.6220	CB	1 LIG	-0.050500
39 C17	7.1460	-7.2670	-8.0300	CA	1 LIG	-0.087300
40 C18	9.1410	-5.8340	-8.3140	CA	1 LIG	-0.063600
41 H18	9.6180	-6.6640	-8.8180	HA	1 LIG	0.094700
42 C19	9.1780	-3.5750	-7.5060	CA	1 LIG	-0.159500
43 H9	9.7030	-2.6350	-7.3880	HA	1 LIG	0.147900
44 C20	5.4650	-9.6610	-8.2650	CA	1 LIG	-0.318000
45 H10	4.8600	-10.5560	-8.3620	HA	1 LIG	0.225800
46 C21	7.5680	-8.4410	-8.6720	CB	1 LIG	0.212400
47 C22	9.8110	-4.6410	-8.1510	CA	1 LIG	-0.243700
48 H11	10.8240	-4.5210	-8.5240	HA	1 LIG	0.179200
49 C23	6.7550	-9.5740	-8.7770	CB	1 LIG	0.273600
50 N5	3.8880	-6.8290	-6.5470	NA	1 LIG	-0.583900
51 H12	3.3070	-6.3760	-5.8400	H	1 LIG	0.394000
52 C24	3.7390	-8.1670	-6.9440	C	1 LIG	0.696600
53 O5	2.7550	-8.8620	-6.7470	O	1 LIG	-0.584800
54 O6	7.4470	-10.5740	-9.4270	OS	1 LIG	-0.356800
55 O7	8.7820	-8.6990	-9.2620	OS	1 LIG	-0.289500
56 C25	8.6370	-9.9670	-9.9190	CT	1 LIG	0.154900
57 H13	8.5560	-9.8030	-11.0030	H2	1 LIG	0.108700
58 H14	9.4990	-10.5960	-9.6800	H2	1 LIG	0.108700
59 O8	7.2610	-2.6490	-6.4060	OS	1 LIG	-0.136400
60 C26	7.9100	-1.3930	-6.2660	CT	1 LIG	-0.105800
61 H15	8.8160	-1.4810	-5.6550	H1	1 LIG	0.087300
62 H16	8.1650	-0.9660	-7.2440	H1	1 LIG	0.087300
63 H17	7.1790	-0.7600	-5.7610	H1	1 LIG	0.087300

Table A30: Mol2 file of ALII-N^6-dA adduct generated using a RED program (atom number provided in the figure below).



1 P	4.8390	0.5750	-1.4400 P	1 LIG	1.166200
2 O5′	4.1920	1.1610	0.0000 OS	1 LIG	-0.487600
3 O1P	4.7430	1.6410	-2.4880 O2	1 LIG	-0.768400
4 O2P	6.0860	-0.1600	-1.0530 O2	1 LIG	-0.768400
5 O3′	0.0000	0.0000	0.0000 OS	1 LIG	-0.541900
6 C5′	3.0580	2.0060	-0.1180 CI	1 LIG	-0.024100
7 H5′1	2.8050	2.3360	0.8950 H1	1 LIG	0.084100
8 H5′2	3.2850	2.9010	-0.7170 H1	1 LIG	0.084100
9 C4′	1.8630	1.2870	-0.7330 CT	1 LIG	0.171700
10 H4′	1.0030	1.9650	-0.7940 H1	1 LIG	0.105200
11 O4′	2.2380	0.8840	-2.0750 OS	1 LIG	-0.419400
12 C1′	1.9140	-0.4820	-2.2650 CT	1 LIG	0.250100
13 H1′	0.8990	-0.5790	-2.6720 H2	1 LIG	0.077800
14 C3′	1.4270	-0.0000	-0.0000 CT	1 LIG	0.130100
15 H3′	1.8230	-0.0320	1.0230 H1	1 LIG	0.078500
16 C2′	2.0090	-1.1160	-0.8760 CT	1 LIG	-0.048500
17 H2′1	3.0610	-1.2770	-0.6220 HC	1 LIG	0.041400
18 H2′2	1.4740	-2.0650	-0.8040 HC	1 LIG	0.041400
19 N9	2.8540	-1.0260	-3.2430 N*	1 LIG	-0.147200
20 C8	3.7590	-0.3340	-4.0280 CK	1 LIG	0.106700
21 H8	3.8660	0.7360	-3.9340 H5	1 LIG	0.179800
22 N7	4.4300	-1.1000	-4.8600 NB	1 LIG	-0.505200
23 C5	3.9290	-2.3630	-4.6240 CB	1 LIG	0.032200
24 C6	4.2400	-3.6240	-5.1810 CA	1 LIG	0.454400
25 N6	5.2130	-3.6930	-6.1350 N2	1 LIG	-0.312200

26 H6	5.5900	-2.7800	-6.3600	H	1 LIG	0.252400
27 N1	3.5670	-4.7000	-4.7380	NC	1 LIG	-0.682900
28 C2	2.6400	-4.5300	-3.7730	CQ	1 LIG	0.506400
29 H2	2.1310	-5.4390	-3.4600	H5	1 LIG	0.074700
30 N3	2.2680	-3.4070	-3.1620	NC	1 LIG	-0.671000
31 C4	2.9480	-2.3470	-3.6300	CB	1 LIG	0.463900
32 C10	5.8170	-4.7740	-6.8120	CA	1 LIG	0.061900
33 C11	5.1810	-5.9340	-7.1730	CN	1 LIG	0.130700
34 C12	7.2180	-4.5970	-7.2130	CA	1 LIG	0.036300
35 C13	5.8660	-6.8990	-7.9750	CB	1 LIG	-0.037900
36 C14	7.8890	-5.5460	-8.0590	CA	1 LIG	0.019700
37 C15	7.9600	-3.4750	-6.7760	CA	1 LIG	-0.132100
38 H18	7.5200	-2.7710	-6.0770	HA	1 LIG	0.125700
39 C16	5.0100	-7.9690	-8.2520	CB	1 LIG	-0.064200
40 C17	7.1720	-6.7410	-8.4650	CA	1 LIG	-0.103800
41 C18	9.2200	-5.3040	-8.4520	CA	1 LIG	-0.066300
42 H9	9.7110	-6.0270	-9.0950	HA	1 LIG	0.093800
43 C19	9.2700	-3.2640	-7.1750	CA	1 LIG	-0.182500
44 H10	9.8060	-2.3910	-6.8130	HA	1 LIG	0.159000
45 C20	5.4270	-9.0250	-9.0620	CA	1 LIG	-0.325000
46 H11	4.7980	-9.8750	-9.2990	HA	1 LIG	0.227600
47 C21	7.5650	-7.8140	-9.2760	CB	1 LIG	0.215400
48 C22	9.9040	-4.1760	-8.0290	CA	1 LIG	-0.171700
49 H12	10.9300	-4.0110	-8.3460	HA	1 LIG	0.152700
50 C23	6.7230	-8.8970	-9.5520	CB	1 LIG	0.271600
51 N5	3.8920	-6.4160	-6.9780	NA	1 LIG	-0.625500
52 H13	3.3190	-6.0630	-6.2150	H	1 LIG	0.406200
53 C24	3.7230	-7.6920	-7.5510	C	1 LIG	0.742400
54 O5	2.7220	-8.3810	-7.4610	O	1 LIG	-0.590700
55 O6	7.3930	-9.8030	-10.3450	OS	1 LIG	-0.350600
56 O7	8.7750	-8.0080	-9.8940	OS	1 LIG	-0.287100
57 C25	8.6050	-9.1610	-10.7340	CT	1 LIG	0.147400
58 H14	8.5380	-8.8360	-11.7820	H2	1 LIG	0.111400
59 H15	9.4470	-9.8420	-10.5850	H2	1 LIG	0.1114

Table A31: Comparison of χ distributions obtained from MD simulations and preferred χ values from QM nucleoside models.

Adduct	χ (deg.)		
	Conformation	**MD average**	**QM**
ALI-N^6-dA	*syn* base-displaced	55.8±11.6	56.7
	syn 5′–intercalated	58.7±12.3	
	syn 3′–intercalated	60.7±11.5	
ALII-N^6-dA	*syn* base-displaced	55.9±11.4	55.9
	syn 5′–intercalated	56.7±13.4	
	syn 3′–intercalated	60.0±12.8	

Table A32: Comparison of the pseudostep parameters[a] and minor groove dimensions[b] for ALI-N⁶-dA and ALII-N⁶-dA adducted DNA in the* anti *base-displaced intercalated (BD-*anti*) and* syn *base-displaced intercalated (BD-*syn*) conformers, as well as the natural strand, from 20 and 320 ns simulations.

Conformer		Shift (Å)	Slide (Å)	Rise (Å)	Minor groove(Å)	Tilt (°)	Roll (°)	Twist (°)
Unmodified	20ns	0.3±0.8	–1.3±1.0	6.5±0.4	7.2±1.5	1.9±5.2	8.6±7.7	63.1±5.9
	320ns	0.2±0.8	–1.4±0.9	6.6±0.4	7.1±1.6	1.6±5.1	9.4±7.6	63.3±6.0
ALI-BD-*anti*	20ns	–1.6±0.7	0.8±0.8	6.9±0.4	7.6±1.2	–2.1±4.9	–5.2±5.9	56.7±4.8
	320ns	–1.6±0.7	0.8±0.8	6.9±0.4	7.7±1.2	–1.9±5.0	–5.3±6.2	56.7±5.1
ALI-BD-*syn*	20ns	–1.9±0.9	0.00±1.2	6.9±0.4	7.7±1.7	0.7±5.1	0.8±5.3	51.1±6.1
	320ns	–2.1±0.9	0.1±1.26	6.9±0.4	7.9±1.7	0.8±5.3	–6.5±8.2	50.4±6.9
ALII-BD-*anti*	20ns	–1.1±0.9	0.2±1.4	6.7±0.4	7.7±1.2	0.0±5.7	–1.6±7.3	54.6±6.2
	320ns	–1.2±1.0	0.2±0.2	6.8±0.4	7.3±1.3	0.0±5.7	–0.7±7.3	54.2±6.1
ALII-BD-*syn*	20ns	–1.9±0.9	0.0±1.3	6.8±0.4	8.1±1.2	2.0±5.2	–5.2±7.5	52.6±6.3
	320ns	–1.8±0.9	0.4±1.3	6.8±0.4	7.9±1.3	1.5±5.3	–5.6±7.6	53.8±6.3

[a] The pseudostep parameters were calculated using a pseudostep consisting of the base pairs 5′ and 3′ with respect to the lesion. [b]The minor groove dimensions were calculated as the distance between the phosphate atom in residues 7th and 20th residue minus 5.8 Å to account for the radius of two phosphate gropus (Figure 2.1b).

Table A33: The hydrogen-bonding occupancies in the trimers composed of the damaged base pair and the 3′ and 5′–flanking base pairs.[a]

Adduct	*Conformation*	*Base Pair*	*H-Bond*	*Occupancy (%)*
ALI-N^6-dA *(anti)*	base-displaced	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.9
		3′–C(7):G(16)	O2•••H2–N2	99.9
			N3•••H–N1	99.9
			N4–H41•••O6	98.2
	5′–intercalated	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H41•••O6	99.1
		3′–C(7):G(16)	O2•••H2–N2	96.9
		3′–C(7):T(17)	N4–H41•••O4	45.8
	3′–intercalated	5′–C(5):T(17)	N4–H41•••O4	36.5
		3′–C(7):G(16)	O2•••H2–N2	99.7
			N3•••H1–N1	99.9
			N4–H41•••O6	98.6
ALII-N^6-dA *(anti)*	base-displaced	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H–N1	99.9
			N4–H41•••O6	98.7
		3′–C(7):G(16)	O2•••H21–N2	99.9
			N3•••H1–N1	99.9
			N4–H41•••O6	98.7
			O2•••H21–N2	99.8
	5′–intercalated	5′–C(5):G(18)	N3•••H1–N1	99.9
			N4–H41•••O6	98.4
		3′–C(7):G(16)	O2•••H21–N2	90.2
		3′–C(7):T(17)	N4–H41•••O4	61.0
	3′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.7
			N3•••H1–N1	99.7
			N4–H41•••O6	98.5
		3′–C(7):G(16)	O2•••H21–N2	99.7
			N3•••H1–N1	99.9
			N4–H41•••O6	98.6

ALI-N^6-dA *(syn)*	base-displaced	5′–C(5):G(18)	O2•••H21–N2	99.7
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
		3′–C(7):G(16)	O2•••H21–N2	99.8
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
	5′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.9
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
		3′–C(7):T(17)	O4•••H41–N4	83.7
			N3•••H3–N3	74.4
	3′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.5
			N3•••H1–N1	99.1
			N4–H41•••O6	97.1
		3′–C(7):G(16)	O2•••H21–N2	99.5
			N3•••H1–N1	99.9
			N4–H41•••O6	98.7
ALII-N^6-dA *(syn)*	base-displaced	5′–C(5):G(18)	O2•••H21–N2	99.8
			N3•••H1–N1	99.9
			N4–H41•••O6	98.7
		3′–C(7):G(16)	O2•••H21–N2	99.7
			N3•••H1–N1	99.9
			N4–H41•••O6	99.8
	5′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.9
			N3•••H1–N1	99.8
			N4–H41•••O6	98.9
		3′–C(7):T(17)	O4•••H41–N4	89.7
			N3•••H3–N3	88.5
	3′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.4
			N3•••H1–N1	99.2
			N4–H41•••O6	96.7
		3′–C(7):G(16)	O2•••H21–N2	99.6
			N3•••H1–N1	99.2
			N4–H41•••O6	98.9

[a]Hydrogen-bond distance cut-off within 3.4 Å heavy atom separation and 120˚ X–H–X angle.

Table A34: Comparison of the structural parameters for DNA containing ALII-N^6-dA obtained from NMR and the most stable* anti *base-displaced intercalated conformation obtained from MD simulations.

Base pair	Sugar Puckering		χ (deg.)	
	NMR[a]	MD[b]	NMR[a]	MD[c]
5′–C	C2′–*endo*	C2′–*endo* (40.6%), C1′–*exo* (22.6%)	–72	–115
5′–G	C1′–*endo*	C2′–*endo* (28.5%), C1′–*exo* (25.1%)	–158	–118
ALII-N^6-dA	C4′–*endo*	C3′–*exo* (65.0%), C2′–*endo* (26.8%)	–101	–111
T	C2′–*endo*	C2′–*endo* (61.4%), C3′–*exo* (31.7%)	–96	–96
3′–C	C1′–*exo*	O4′–*endo* (37.0%), C1′–*exo* (32.0%)	–119	–132
3′–G	C2′–*endo*	C2′–*endo* (69.3%), C3′–*exo* (15.8%)	–93	–114

[a]Computed from the average minimized structure obtained from NMR (Ref. 1 Section 2.6). [b]Two most common sugar puckerings acquired by the deoxyribose sugar during the MD simulation, with the percentage of time the pucker was adopted given in parentheses. [c]Calculated from the representative MD structure obtained by clustering with respect to χ, θ and ϕ (Figures 2.1a and 2.6b).

Table A35: Effect of the methoxy group on the van der Waals interaction energy (kJ mol^{-1}) between the adducted base at position X in the 11-mer DNA (Figure 2.1b) and the flanking base pairs for DNA containing ALI-N^6-dA in the* anti *and* syn *base-displaced intercalated conformers.

Conformation	With methoxy[a]	Without methoxy[b]	Difference
anti	–171.8±6.8	–161.3±6.1	10.5
syn	–166.3±7.1	–157.5±7.1	9.8

[a]Stacking energies between the damaged base and flanking base pairs. [b]Stacking energies between the damaged base and flanking base pairs calculated without including the atoms of the methoxy group**.**

APPENDIX B

Conformational Preferences of Adenine versus Guanine DNA Adducts of Aristolochic Acid-II

B1. Full Computational Details:

B1.1. Nucleobase Model: The damaged ALII-N^2-G nucleobase was built by adding the ALII moiety formed from AAII to the N2 positon of guanine. Subsequently, the preferred orientation of the ALII moiety with respect to guanine was considered, which depends on rotation about θ (∠(N1C2N2C10)) and ϕ (∠ (C6N6C10C11), Figure 3.1b). Specifically, a two-dimensional potential energy surface (PES) was generated using constrained B3LYP/6-31G(d) optimizations with θ and ϕ fixed in 10˚ increments from 0˚ to 360˚. Subsequently, the minima identified on the surface were fully optimized using the dispersion-corrected B3LYP-D3 functional in combination with 6-31G(d). Although other approaches for including dispersion corrections in DFT functionals are available in the literature,[1-3] the inclusion of the empirical dispersion correction using the DFT-D3 approach has been shown to perform reasonably well for systems where noncovalent interactions are important.[4] Since a previous study on the S66 and S66x8 datasets of noncovalent interactions indicate that adding the dispersion correction strongly diminishes the performance differences between functionals[5], the B3LYP-D3 functional is chosen in the present work in order to allow meaningful comparisons with the structures reported using the same functional in the previous study on the AL-N^6-dA adducts (Chapter 2). Single-point energy calculations were performed using B3LYP-D3 and the larger 6-311+G(2df,p) basis set to obtain more accurate relative energies. The choice of the 6-31G(d) basis set for the geometry optimizations and the 6-311+G(2df,p) basis set for single-point calculations is justified based on the previous studies on DNA adducts.[6-9]

B1.2 Nucleoside Model: The damaged ALII-N^2-dG nucleoside was built by adding 2′–deoxyribose with the C2′–endo puckering to the lowest energy nucleobase conformation. This build-up approach for the DFT conformational scans has been validated in the previous combined experimental and computational studies on phenolic DNA adducts,[8] where spectroscopic data matches the results obtained from the DFT calculations. The ε dihedral angle (∠(C4′C3′O3′H5′), Figure 3.1b) was initially set to approximately 180˚, which is the average value obtained from previous molecular dynamics (MD) simulations on the Dickerson-Drew dodecamer.[10] The B3LYP/6-31G(d) PES of the nucleoside adduct was mapped as a function of θ and χ (∠(O4′C1′N9C4), Figure 3.1b) using constrained optimizations and 10˚ increments from 0˚ to 360˚. The minima on the surface were then fully optimized with B3LYP-D3/6-31G(d), while B3LYP-D3/6-311+G(2df,p) was used to obtain more accurate relative energies. Subsequently, the lowest energy *anti* and *syn* conformations were fully optimized with constraints on β (∠(C4′N5′O5′H) = 180˚, Figure 3.1b) and ε

(∠(C4′C3′O3′H) = 270°), which prevent superficial interactions between the 5′ and 3′–OH of the sugar and the nucleobase or AL moiety in order to better mimic damage at a non-terminal position.

B1.3 Nucleotide Model: The ALII-N^2-dG nucleotide was built by adding a 5′–monophosphate unit to the lowest energy *anti* and *syn* conformers of the constrained nucleoside model. According to a previous study,[7] the structural properties of modified nucleotides are best characterized in solvent (water) with a sodium ion to counter the phosphate charge. Thus, the ALII-N^2-G nucleotide was optimized in water (ϵ = 78.4) using PCM-B3LYP-D3/6-31G(d) and single-point calculations were performed with PCM-B3LYP-D3/6-311+G(2df,p).

Coordinates of the nucleobase minima, and the lowest energy *anti* and *syn* minima for the nucleoside, constrained nucleoside and nucleotide models, are provided in Tables B5–B12. All reported (B3LYP-D3) relative energies include a scaled (0.9806) zero-point vibrational energy correction. All quantum mechanical calculations were performed using Gaussian 09 (Revision C.01 or D.01).[11,12]

B1.4 DNA Model: To analyze the preferred conformations of ALII-N^2-dG adducted DNA, ALII-N^2-dG was incorporated into the 5′–CGTACXCATGC 11-mer DNA sequence (X = adduct; Figure B5) previously used to experimentally and computationally analyze the ALI-N^6-dA and ALII-N^6-dA[13] adenine adducts (Chapter 2). The B-form of the DNA was initially built using the NAB module of AMBER 11.[14] The guanine at position X was modified by adding the ALII moiety to the N2 position using GaussView.[15] The damaged base was paired against complementary C. Partial charges for the modified ALII-N^2-dG bases were calculated using the RED.v.III.4 program[16] and ANTECHAMBER 1.4 was used to assign atom types (Table B13). The parambsc0[17] modification to the parm99[18] force field was used to simulate the natural and adducted nucleotides, while additional parameters for the N2-moiety of the adducted nucleotides were taken from the General Amber Force Field (GAFF).[19] The DNA strand was neutralized with 20 sodium ions and subsequently solvated in an 8 Å octahedral box of TIP3P water.[20] Although improved force field parameters for monovalent ions are available,[21] simulations at low ion concentrations or at near physiological conditions are shown to be robust regardless of the ion model implemented.[21] Thus, the parm94 parameters for sodium ions were used in this work, which are the default choice in AMBER 11 or 12. The system was initially minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ restraint on DNA. Subsequently, an unrestrained minimization was performed using 1000 steps of the steepest descent method followed by 1500 steps of the conjugate gradient method. The system was then heated to 300 K with DNA restrained using a force constant of 10 kcal mol^{-1} $Å^{-2}$ and a 20 ps constant volume MD simulation was performed.

Starting structures were generated by first incorporating the adduct into DNA in the lowest energy *anti* and *syn* orientations obtained from the nucleotide model, while avoiding steric clashes with the surrounding nucleotides. Since the lowest energy structure with $\theta \sim 180^{\circ}$ did not form an intercalated conformation, the next lowest energy $\theta \sim 0^{\circ}$ structure was used to build the initial adducted DNA models. Furthermore, to adequately sample the conformational space of the adducted DNA, trial simulations were carried out with various starting structures, which cover all possible locations of the bulky moiety in the helical environment. Specifically, simulations were started with the ALII moiety located in the major groove (with the *syn* adduct orientation), minor groove (with the *anti* adduct orientation) and intercalated into the helix (with both the *anti* and *syn* adduct orientations). Furthermore, three types of intercalated structures were considered, which differ in the orientation of the ALII moiety with respect to the damaged guanine. In these structures, the ALII moiety was allowed to intercalate either from the 5' or 3' side of the damaged guanine. The third type of intercalated conformer was considered for both the *anti* and *syn* orientations of the adducted nucleoside, which involved intercalation of the ALII moiety with displacement of the cytosine opposing the damaged guanine (the base-displaced intercalated conformer). In the simulated structures, it was ensured that the Watson-Crick hydrogen bonds between both terminal base pairs were persistent in order to allow meaningful comparisons of the relative stabilities of different conformations using free energy calculations. In cases where more than one starting structure reverted to the same conformation after simulation or the desired optimal location of the bulky moiety was not achieved, new starting structures were built and simulated. This procedure was repeated until no new structures were obtained. Overall, ~30 trial simulations were carried out in order to identify stable structures corresponding to each conformational theme reported in this work. These stable structures were then used for the production simulations. It was noted that, despite carrying out a number of trial simulations with different starting structures, stable major groove, *anti* 5'-intercalated and *syn* 3'-intercalated conformations could not be located for ALII-N^2-dG adducted DNA. Furthermore, each simulated conformation remains stable over the duration of the (20 or 120 ns) MD simulations, and no interconversions between conformations are observed.

Final production simulations were subsequently initiated from representative conformations obtained from the trial simulations for 20 ns using the PMEMD[22] module of AMBER 11[14] or AMBER 12.[23] To analyze the structural features at the lesion site for each distinct conformation of adducted DNA, base step parameters were calculated using a pseudostep consisting of the base pairs 5′ and 3′ with respect to the lesion. A pseudostep was used for analysis since the disrupted ALII-N^2-dA:G pair cannot be used to calculate the step parameters at the lesion site due to intercalation of the AL moiety. The van der Waals component of the stacking interaction energy between the damaged guanine and the neighboring bases (including the flanking base

pairs and the opposing base) was calculated in order to estimate the differences in stacking interactions between the adducted DNA conformations. Although the electrostatic component also contributes to stacking (although in relatively smaller proportion compared to the van der Waals component), only the van der Waals component is reported since some of the conformations involved hydrogen bonding (which is mainly an electrostatic interaction) between the damaged pair in addition to stacking (which is predominantly a van der Waals interaction), while other conformations did not exhibit such hydrogen bonding. To quantify the stacking interactions at the lesion site separately from hydrogen bonding, it was found considering only the van der Waals component of the interaction to be the best approach.

Due to the extensive process for selecting initial structures for the production simulations described above, the structural parameters of each distinct conformation obtained from independent 20 ns simulations are sufficiently converged. This was verified by extending each simulation to 120 ns. Specifically, the average and standard deviation in the RMSD of the trimer containing the lesion pair (Table B2), as well as key structural parameters (Table B3), do not change significantly upon extending the simulations. Thus, the analysis focuses on the first 20 ns of the production MD simulations on AL-N^2-dG adducted DNA. Free energy calculations were performed on each simulation using the MM-PBSA (Molecular Mechanics-Poisson Boltzmann surface area) method and snapshots at 50 ps intervals (400 frames in total). The entropy term was calculated using normal mode analysis. Although free energy methods that are theoretically more rigorous than end-point free energy methods such as the MM-PBSA method have been used in the literature,[24,25] a previous study has shown that DNA conformational free energy differences obtained using the MM-PBSA approach are in close agreement with the potentials of mean force determined using the more rigorous umbrella sampling approach.[26] Moreover, MM-PBSA has been the most commonly used method in previous studies of damaged DNA.[27-30] Thus, the MM-PBSA method is chosen in the present work in order to allow meaningful comparisons of the structural properties of AL-N^2-dG adducted DNA with a previous study on AL-N^6-dA adducted DNA (Chapter 2).

B2. Detailed Structural and Free Energy Analysis:

B2.1 Nucleobase Model: To determine the preferred relative orientation of the ALII and guanine moieties, the structure of the ALII-N^2-G nucleobase was considered by plotting the PES as a function of θ and ϕ (Figure B1a). Two minima with θ ~ 20˚ or 340˚ are separated by a small ϕ rotational barrier of ~ 20 kJ mol^{-1}. The next two higher energy minima at θ ~ 170˚ or 190˚ are separated by a larger ϕ barrier (~ 30 – 40 kJ mol^{-1}). Similarly, the rotational barrier between minima with different θ (θ ~

20˚ or 340˚ and θ ~ 170˚ or 190˚) is ~ 20 kJ mol^{-1} at ϕ ~ 260˚ or 310˚ and ~ 40 kJ mol^{-1} at ϕ ~ 50˚ or 100˚. Hence, the flexibility about θ is dependent on ϕ and *vice versa*.

Upon full optimization of all minima identified on the PES (Figure B1b), the two minima with θ ~ 20˚ or 340˚ (ϕ ~ 260˚ or 100˚) become isoenergetic and constitute the global minimum. These structures contain a twist about the ϕ–linkage, which alleviates steric interactions between the guanine and bulky moieties. On the other hand, two higher energy minima at θ ~ 170˚ or 190˚ (ϕ ~ 50˚ or 310˚) are stabilized by a hydrogen bond between N–H in ring III (Figure 3.1b) of the AL moiety and N3 of guanine, and are ~ 6 kJ mol^{-1} higher in energy than the global minimum.

B2.2 Nucleoside Model: To determine the (*anti/syn*) conformational flexibility about the glycosidic bond, the ALII-N^2-dG PES surface was plotted as a function of χ and θ (Figure B2a). Overall, seven minima were located on the PES, which were then fully optimized (Figure B3). Since the nucleobase model indicates that two isoenergetic orientations are possible about ϕ for a given θ, all minima on the nucleoside surface were optimized with both initial ϕ orientations. As a result, 13 minima were fully optimized for the ALII-N^2-dG nucleoside (Figure B3). The high *anti* minimum with χ ~ 300˚ and θ ~ 180˚ is only possible for one ϕ value (~ 40˚) since the stabilizing interaction between 3′–OH of the sugar at the exocyclic oxygen on ring III of the lactam moiety is absent when ϕ ~ 320˚ (Figure B3). The global minimum is the *syn* orientation at χ = 58˚, θ = 40˚ and ϕ = 88˚. Whereas all *syn* minima lie within 14 kJ mol^{-1}, only two of the seven *anti* or high *anti* minima are within 14 kJ mol^{-1} of the global minimum.

For the lowest energy *anti* and *syn* nucleoside structures, either the 3′–OH or 5′–OH of the sugar interacts with the guanine or bulky moiety. In the case of damage to a non-terminal helical position, the hydrogen atoms of these hydroxyl groups are replaced by phosphate groups. Hence, to make the model relevant to other helical positions, the hydroxyl groups were constrained away from the nucleobase (β (∠(C4′N5′O5′H5′)) = 180˚ and ε (∠(C4′C3′O3′H)) = 270˚; Figure 3.1b), and the nucleoside minima were re-optimized. In the resulting structures (Figures B4a and b), the *anti* conformation is ~ 7 kJ mol^{-1} more stable than the *syn* conformer, which can be attributed to reduced steric interactions between the guanine and bulky moieties in the *anti* conformer.

B2.3 Nucleotide Model: To determine the effect of the phosphate backbone on the *anti/syn* stability, a 5′–phosphate group neutralized by a sodium ion was added to the lowest energy *anti* and *syn* conformers of the constrained nucleoside model. Similar to the nonterminal nucleoside model, the *anti* nucleotide is more stable than the *syn*

conformer (Figures B4c and d). However, when the phosphate group is added, the *anti*/*syn* energy difference increases to ~ 27 kJ mol^{-1}.

B2.4 DNA Model: Six distinct conformers of ALII-N^2-dG adducted DNA were isolated from MD simulations (Figure 3.3a). Specifically, the (*anti* and *syn*) base-displaced intercalated, the (*syn*) 5′–base intercalated, the (*anti*) 3′–base intercalated, (*anti*) minor groove and the (*syn*) 3′, 5′–base intercalated conformations were characterized. Key structural features of each conformation are highlighted below (references to the (3′ and 5′) nucleotide positions of flanking bases are made with respect to the adducted nucleotide) mainly in terms of changes with respect to unmodified DNA in the pseudostep parameters (calculated using the pseudostep consisting of the 5′ and 3′– base pairs with respect to the lesion) and changes in the minor groove dimensions (calculated as the distance between the phosphate atom in residues 7 and 20; see Figure B5 for residue numbering).

B2.4.1 The anti *base-displaced intercalated conformer:* In the *anti* base-displaced intercalated conformer, the ALII moiety stacks between the 5′–flanking base pair and the 3′–guanine in the opposing strand (Figure 3.3a). However, owing to the twist at the guanine–ALII linkage, accommodation of the ALII moiety in the helix disrupts the stacking between the damaged guanine moiety and its 5′–cytosine such that the guanine moiety only stacks with the 3′–cytosine. Despite an extrahelical location and loss of hydrogen bonding with the damaged G, the opposing cytosine remains in an *anti* orientation (χ ~ 220°). All three Watson-Crick (WC) hydrogen bonds are maintained in the 5′–base pair with respect to the lesion. However, the hydrogen bonding in the 3′–base pair is partially disrupted due to the twist at the carcinogen–purine linkage (Table B4). Although the minor groove dimensions of this conformation differ very little (0.5 Å) from natural DNA, there are marked changes in the roll (~ 26°) and twist (~ 11°) at the lesion site (Figure 3.4).

B2.4.2 The syn *base-displaced intercalated conformer:* In the *syn* base-displaced intercalated conformer, the *syn* damaged guanine is displaced into the minor groove. The damaged base retains stacking interactions with the 3′–flanking pair, while stacking with the 5′–flanking base pair is completely disrupted. In contrast, the ALII moiety stacks only with the 5′–flanking base pair at the lesion site (Figure 3.3a). The cytosine opposing the lesion becomes extrahelical while maintaining an *anti* orientation. Although the hydrogen bonding in the lesion pair is completely disrupted, the 5′ and 3′–base pairs remain intact (~ 100% occupancy; Table B4). This conformation yields the greatest change in the slide (6 Å) relative to the natural helix compared to all other conformers. In addition, this conformation is characterized by significant changes in the shift (3 Å) and minor groove dimensions (3 Å) compared to natural DNA (Figure 3.4).

B2.4.3 The syn *5′–intercalated conformer:* In the *syn* 5′–intercalated conformer, the ALII moiety stacks between the opposing cytosine and the 5′–flanking pair (Figure 3.3a). However, due to twisting at the guanine–ALII linkage, the stacking interactions involving the damaged guanine are diminished with its 3′–flanking base, and are completely lost with its 5′–flanking base. However, WC hydrogen bonding is maintained in both the 5′ and 3′–flanking base pairs (~ 100% occupancy; Table B4). Additionally, one of the hydrogen bonds between the damaged G and the opposing cytosine (N1–H1 (G6)•••N3 (C17)) remains for ~ 83% of the simulation time. However, the simultaneous accommodation of the ALII moiety and opposing cytosine within the helix results in significant changes in the slide (~ 5 Å) and rise (~ 4 Å) compared to natural DNA. Although the change in roll is smaller compared to the *anti* base-displaced intercalated conformer, the *syn* 5′–intercalated structure has more pronounced duplex untwisting (27°) and minor groove enlargement (4 Å) at the lesion site (Figure 3.4).

B2.4.4 The anti *3′–intercalated conformer:* In the *anti* 3′–intercalated conformer, the ALII moiety stacks between the cytosine opposing the lesion and the 3′–flanking base pair (Figure 3.3a). Uniquely, the damaged guanine moiety twists to become parallel to the helical axis, and therefore is fully located in the minor groove. This disrupts stacking interactions between the damaged G and both the 5′ and 3′–flanking bases. Nevertheless, WC hydrogen bonding is maintained in both the 5′ and 3′–base pairs (~ 100% occupancy; Table B4). Additionally, one hydrogen bond (N4–H41 (C17)•••O6 (G6)) is maintained in the lesion pair (~ 95% occupancy; Table B4). Changes in the stacking pattern at the lesion site result in significant deviations in the slide (1.7 Å) and rise (2.7 Å) compared to natural DNA (Figure 3.4). However, the most significant structural deviation in this conformer is the minor groove width (6 Å), which is greater than for any other conformer. The minor groove enlargement is also accompanied by significant increases in the roll (27°) and duplex untwisting (17°; Figure 3.4).

B2.4.5 The anti *minor groove stacked conformer:* The minor groove stacked conformer differs from all other adducted DNA conformers since the ALII moiety does not intercalate into the helix. Instead, the bulky moiety is fully located in the minor groove, with the carbocyclic rings directed toward the 5′–end of the damaged strand. In this position, both the ALII moiety and the cytosine opposing the lesion become parallel to the helical axis, which permits π–π interactions between the two moieties (Figure 3.3a). Nevertheless, the damaged guanine remains in the helix, stacks with the 5′ and 3′–base pairs, and maintains one (N4–H42 (C17)•••O6 (G6)) hydrogen bond with the opposing cytosine (~ 89% occupancy). In addition, the hydrogen bonding in the 5′ and 3′–base pairs remains intact (Table B4). Since the bulky moiety

stacks in the minor groove, this conformation exhibits a significant increase in the minor groove dimension (4.2 Å; Figure 3.4). Despite noticeable duplex untwisting (14˚) and change in the slide (2 Å) with respect to the natural duplex, changes in other lesion site parameters are smaller compared to the other adducted DNA conformers.

B2.4.6 The syn 3′,*5′–intercalated conformer:* The *syn* 3′,*5′*–intercalated conformer is a particularly unique conformation since the ALII moiety is completely intercalated into the helix, and simultaneously stacks with the cytosine opposing the lesion, the 5′–base pair and the 3′–base pair. The damaged base is displaced into the minor groove and hence loses stacking contacts with both flanking base pairs (Figure 3.3a). Nevertheless, the damaged base forms a (O6 (G6)•••N4–H42 (C17)) hydrogen bond with the opposing base. Additional hydrogen bonds are observed between the cytosine opposing the lesion and the 3′–cytosine (O2 (C7)•••H41–N4 (C17), with ~ 80% occupancy; Table B4). This conformer is characterized by marked changes in the shift (3 Å), tilt (8˚) and twist (13˚) at the lesion site compared to natural DNA (Figure 3.4); however, changes in other helical parameters are less pronounced compared to the other adducted DNA conformers.

B2.4.7 Free Energy Analysis for the ALII-N²-dG adducted DNA conformers: All conformers described above lie very close in energy (within 21 kJ mol^{-1}, Figure 3.3). This can be explained by the effects of mutually compensating interactions, namely the van der Waals (stacking), (repulsive) steric and hydrogen-bonding interactions in the trimer of the base pairs about the damaged base. Specifically, the *anti* base-displaced intercalated conformer of ALII-N^2-dG adducted DNA has significant lesion site stacking due to the intercalation of the ALII moiety into the helix. However, since the canonical rise is maintained, the twist at the carcinogen–purine linkage decreases hydrogen bonding in the 3′–flanking pair because of the displaced damaged G. In contrast, although the canonical rise is also maintained in the *syn* base-displaced intercalated conformer, lesion site stacking decreases since the adduct is not as well accommodated in the helix (i.e., the damaged G and bulky moiety are more displaced towards the minor groove). Therefore, the hydrogen bonds are maintained in the 5′ and 3′–flanking pairs, which compensates for the loss in stacking. For the *anti* 3′–intercalated and the minor groove stacked conformers, the bulky moiety is not well stacked in the helix. However, the twist at the carcinogen–purine linkage displaces the damaged guanine or ALII moiety into the minor groove, respectively, which relieves steric repulsion with the flanking bases. The hydrogen-bonding interactions in the flanking pairs, as well as the one hydrogen bond at the lesion site (Table B4), compensate for the decreased stacking interactions. Although the *syn* 5′–intercalated conformer has significant stacking with the bulky moiety, the rise increases, which maintains hydrogen bonds in the 3′ and 5′–base pairs and one hydrogen bond at the

lesion site (Table B4). However, the twist introduces unfavorable steric interactions between the damaged guanine and the 3′–flanking base pair and the 5′–base. Finally, despite significant stacking, the *syn* 3′,5′– intercalated conformer is destabilised by repulsive interactions involving the C opposing the lesion due to maintenance of the canonical rise. Although this provides a qualitative explanation for the close energetic separation between all six adducted DNA conformers, other factors are likely also important in dictating the final energy rankings, such as differences in the hydration at the lesion site.

B3 References

(1) Zhao, Y.; Truhlar, D. *Theor. Chem. Account* **2008**, *120*, 215.

(2) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

(3) Torres, E.; DiLabio, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 3342.

(4) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem*. **2011**, *32*, 1456.

(5) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421.

(6) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Chem. Res. Toxicol.* **2013**, *26*, 803.

(7) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373.

(8) Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. A.* **2008**, *112*, 3742.

(9) Millen, A. L.; Sharma, P.; Wetmore, S. D. *Future Med. Chem.* **2012**, *4*, 1981.

(10) Zgarbová, M.; Luque, F. J.; Šponer, J.; Cheatham III, T. E.; Otyepka, M.; Jurečka, P. *J. Chem. Theory Comput.* **2013**, 9, 2339.

(11) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Revision C.01 ed.; Gaussian, Inc*.: Wallingford CT, **2010**.

(12) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavchari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman,

J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision D.01 ed.; Gaussian, Inc.: Wallingford CT, **2013**.

(13) Lukin, M.; Zaliznyak, T.; Johnson, F.; de los Santos, C. *Nucleic Acids Res.* **2012**, *40*, 2759.

(14) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; R.C.Walker; Zhang, W.; Merz, K. M.; B.Wang; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F.Wong; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A.; *AMBER 11* University of California, San Francisco, CA: **2010**.

(15) Roy, D. T., Keith; John, Millam *GaussView Version 5* **2009**.

(16) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821.

(17) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J .* **2007**, *92*, 3817.

(18) Cheatham III, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.

(19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(21) Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596.

(22) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(23) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; *AMBER 12* University of California, San Francisco, CA: **2012**.

(24) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. *J. Chem. Theory Comput.* **2013**, *9*, 1885.

(25) Beveridge, D. L.; DiCapua, F. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(26) Brice, A. R.; Dominy, B. N. *J. Comput. Chem.* **2011**, *32*, 1431.

(27) Sharma, P.; Manderville, R. A.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 11831.

(28) Sproviero, M.; Verwey, A. M. R.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 13405.

(29) Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *J. Mol. Biol.* **2007**, *374*, 292.

(30) Yan, S.; Wu, M.; Buterin, T.; Naegeli, H.; Geacintov, N. E.; Broyde, S. *Biochemistry* **2003**, *42*, 2339.

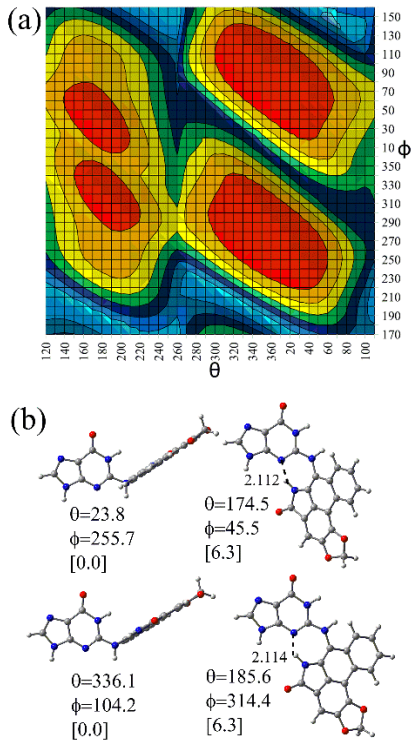


Figure B1: (a) B3LYP/6-31G(d) PES and (b) B3LYP-D3/6-31G(d) optimized minima, including select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol^{-1}), for the ALII-N^2-G nucleobase adduct. In the PES, the relative energy (kJ mol^{-1}) is represented by color, where the lowest energy regions are red, and each change in color represents a 10 kJ mol^{-1} increment.

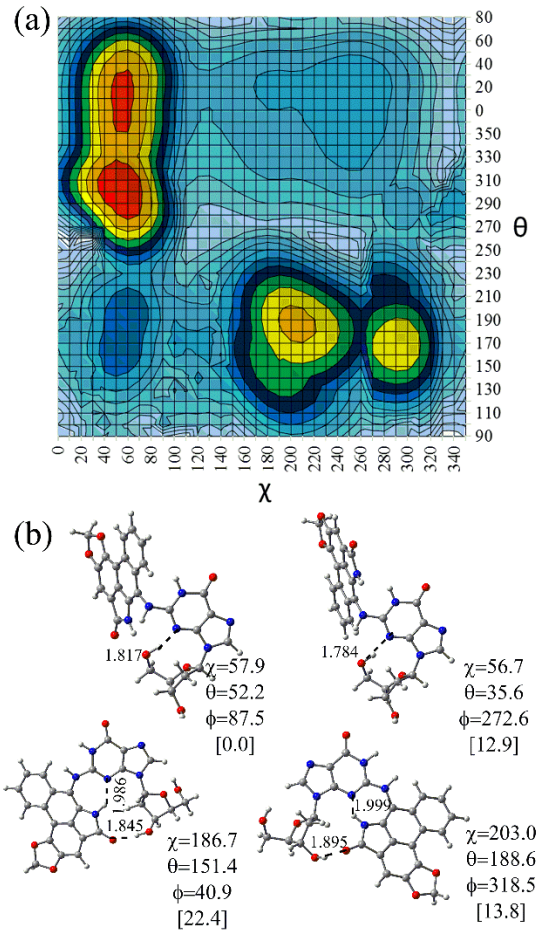


Figure B2: (a) B3LYP/6-31G(d) PES and (b) B3LYP-D3/6-31G(d) optimized minima, including select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol^{-1}), for the ALII-N^2-dG nucleoside adduct. In the PES, the relative energy (kJ mol^{-1}) is represented by color, where the lowest energy regions are red, and each change in color represents a 5 kJ mol^{-1} increment.

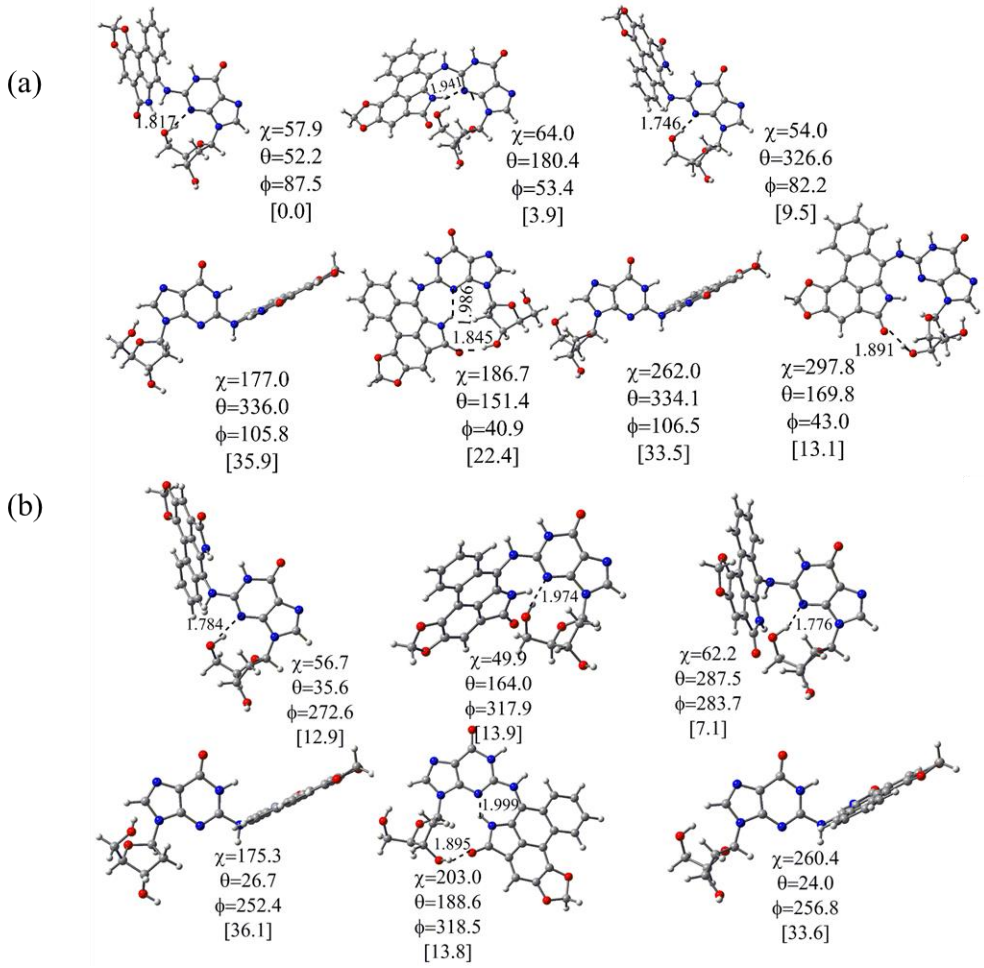


Figure B3: B3LYP-D3/6-31G(d) optimized minima for the ALII-N^2-dG nucleoside adduct obtained (a) from the PES (ϕ ~ 40 °–100 °; Figure B2) and (b) upon reoptimization following rotation about the ϕ-dihedral angle by 180 ° (ϕ ~ 250 °–320 °). Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (kJ mol^{-1}, in square brackets) are provided.

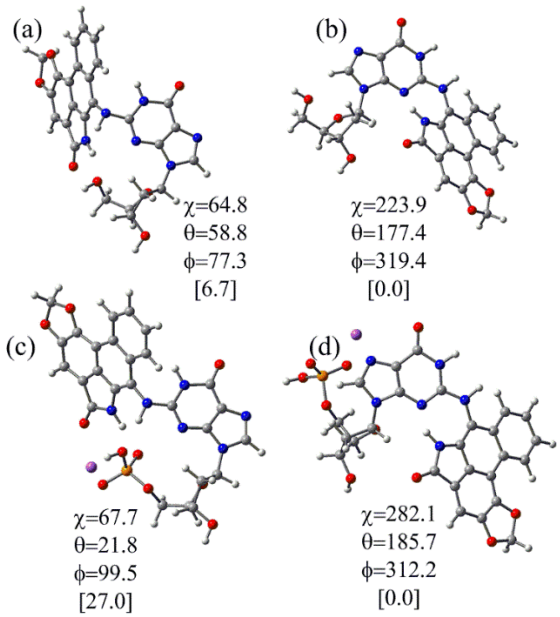


Figure B4: B3LYP-D3/6-31G(d) lowest energy* syn *(a and c) and* anti *(b and d) minima for the constrained nucleoside (a and b)and the nucleotide (c and d) of the ALII-N^2-dG adduct. Select bond lengths (Å), dihedral angles (deg.) and the corresponding B3LYP-D3/6-311+G(2df,p) relative energies (in square brackets, kJ mol^{-1}) are provided.

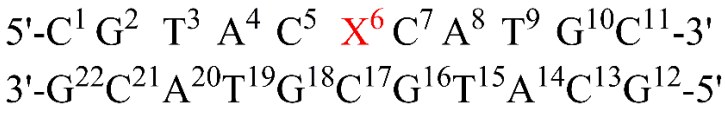


Figure B5: Residue numbering in the 11-mer DNA sequence used for MD simulations with the adduct at the X^6 position.

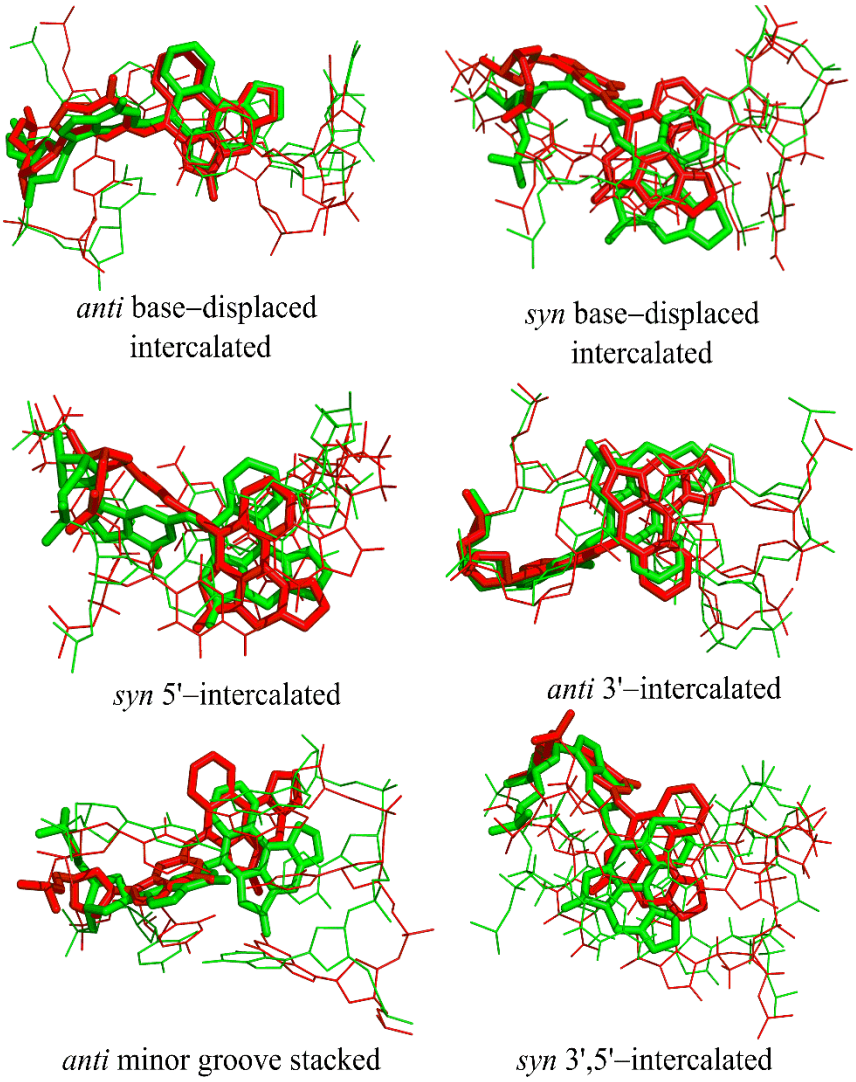


Figure B6: Overlay of the representative (green) and initial (red) MD structures, including the lesion (tube), as well as the opposing cytosine, 5′ and 3′–flanking base pairs (wireframe), for each ALII-N^2-dG adducted DNA conformer.

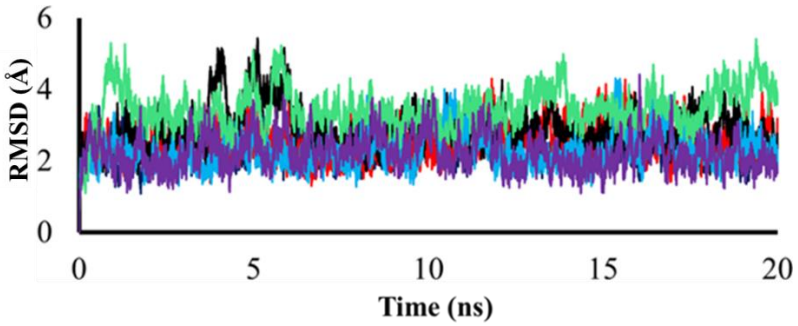


Figure B7: Backbone RMSD versus time for ALII-N²-dG adducted DNA in various conformations, including the* anti *base-displaced intercalated (dark blue),* syn *base-displaced intercalated (light blue),* syn *5′–intercalated (black),* anti *3′–intercalated (red),* anti *minor groove stacked (purple) and* syn *3′,5′–intercalated (green) orientations.

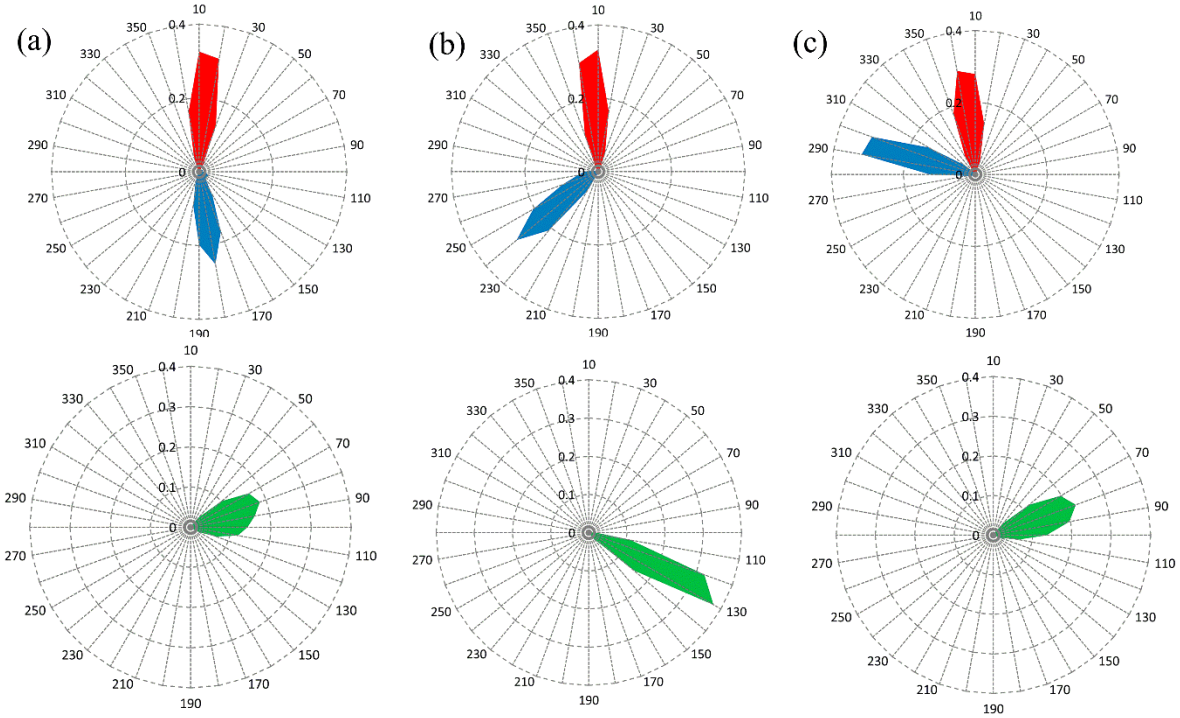


Figure B8: Radar plots of the probability distribution in χ (blue), θ (red) and ϕ (green) for* anti *ALII-N²-dG adducted DNA in the (a) 3′–intercalated, (b) base-displaced intercalated and (c) minor groove stacked conformations.

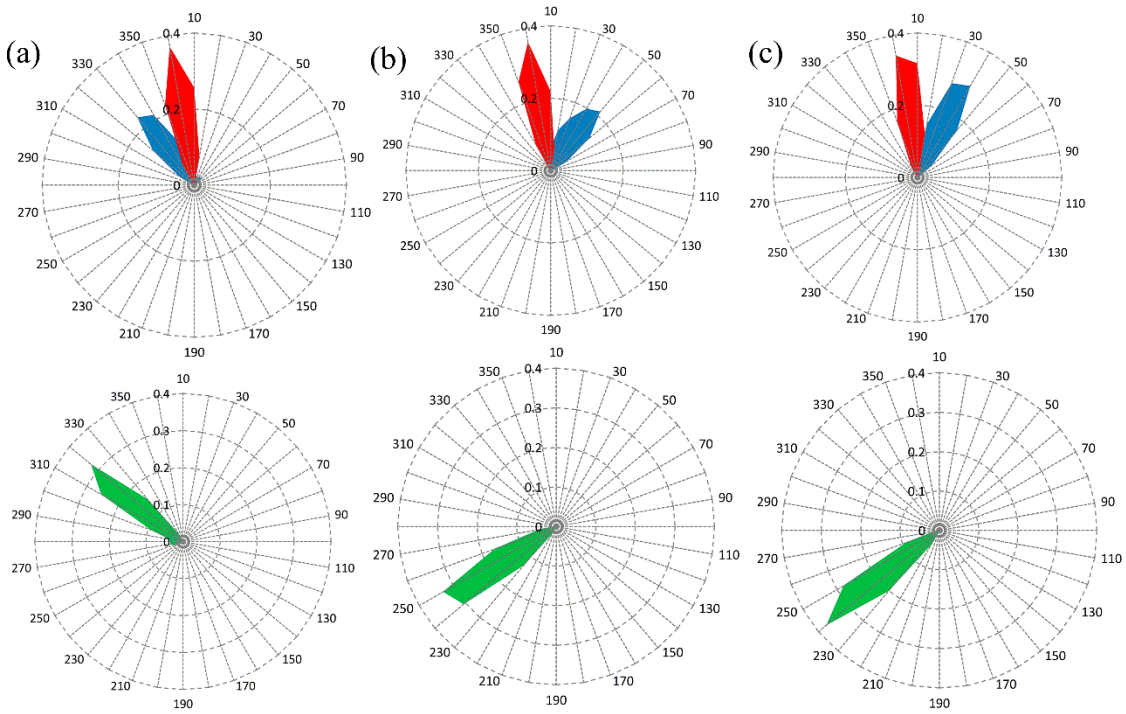


Figure B9: Radar plots of the probability distribution in χ (blue), θ (red) and ϕ (green) for* syn *ALII-N²-dG adducted DNA in the (a) 5′–intercalated, (b) base-displaced intercalated and (c) 3′,5′–intercalated conformations.

Table B1: The χ, θ and ϕ dihedral angles[a] (deg.) in the initial and representative structures obtained from each MD simulation.

Conformation	Initial Structure			Representative Structure		
	χ	**θ**	**ϕ**	**χ**	**θ**	**ϕ**
anti base-displaced intercalated	262.1	353.1	133.7	237.9	11.1	116.7
syn base-displaced intercalated	45.1	358.1	237.2	36.1	352.3	242.0
syn 5′–intercalated	355.8	356.9	286.8	336.1	357.4	300.9
anti 3′–intercalated	196.1	1.2	83.8	167.6	6.8	79.9
anti minor groove stacked	263.1	355.5	87.8	291.5	358.2	77.9
syn 5′,3′–intercalated	37.5	3.4	246.0	37.8	352.5	249.7

[a]See Figure 3.1b for the definitions of the dihedral angles.

Table B2: RMSD for all heavy atoms in the trimer containing the damaged base pair and the flanking base pairs in ALI-N^2-dG adducted DNA during 20 and 120 ns simulations.

Conformation	Simulation Time (ns)	RMSD (Å)
anti base-displaced intercalated	20	0.9±0.3
	120	1.2±0.3
syn base-displaced intercalated	20	1.4±0.2
	120	1.6±0.3
syn 5′–intercalated	20	1.6±0.2
	120	2.2±0.5
anti 3′–intercalated	20	1.4±0.3
	120	1.8±0.4
anti minor groove stacked	20	1.3±0.3
	120	1.4±0.5
syn 5′,3′–intercalated	20	2.2±0.3
	120	2.2±0.2

Table B3: Comparison of the pseudostep parameters[a] and minor groove dimensions[b] in all energetically accessible ALII-N²-dG adducted DNA conformers for 20 and 120 ns simulations.[c]

Conformation	Simulation Time (ns)	Shift (Å)	Slide (Å)	Rise (Å)	Minor groove (Å)	Tilt (°)	Roll (°)	Twist (°)
anti base-displaced	20	–0.1±1.0	1.4±1.4	7.3±0.5	7.3±1.9	–7.2±6.5	–16.5±7.5	55.9±8.4
intercalated	120	–0.6±1.1	–1.8±1.6	7.4±0.5	6.3±2.5	–3.5±7.5	–17.3±7.6	58.7±10.1
syn base-displaced	20	–2.8±1.1	4.9±1.1	7.8±0.8	11.8±1.4	–3.4±7.3	30.7±8.7	48.0±10.8
intercalated	120	–3.4±1.7	3.7±1.5	7.7±1.0	10.8±1.7	–5.1±8.0	34.7±19.0	39.5±14.8
syn 5′–	20	–0.2±2.1	4.3±1.1	10.3±0.9	11.4±1.7	–8.8±9.1	2.4±20.1	39.2±11.2
intercalated	120	–0.2±1.7	1.7±2.2	10.1±0.7	10.6±2.9	–7.3±7.9	–6.8±15.2	44.4±12.2
anti 3′–	20	0.7±1.2	0.7±1.2	9.3±1.5	14.2±1.2	–6.7±7.3	37.2±15.7	49.6±14.8
intercalated	120	1.5±1.8	1.0±1.6	9.0±0.9	12.3±2.4	–6.5±6.9	32.1±9.9	41.4±14.2
anti minor groove	20	0.7±1.1	0.6±1.3	6.5±0.4	12.0±1.3	8.6±5.9	10.1±10.5	52.2±7.1
stacked	120	0.9±1.1	0.8±1.4	6.4±0.5	11.5±1.9	9.9±6.1	15.2±10.8	55.8±7.6
syn 5′,3′–intercalated	20	3.3±0.8	0.5±0.8	6.8±0.4	10.7±1.8	8.8±6.6	7.6±14.9	49.2±10.9
	120	3.2±0.7	0.4±0.8	6.9±0.4	10.4±1.6	8.9±5.2	10.0±7.2	54.1±5.7

[a]The pseudostep parameters were calculated using a pseudostep consisting of the base pairs 5′ and 3′ with respect to the lesion. [b]The minor groove dimensions were calculated as the distance between the phosphate atom in residues 7 and 20 (Figure B5).

Table B4: Hydrogen-bonding occupancies in different conformers of ALII-N^2-dG adducted DNA for the trimers composed of the damaged base pair, and the 3′ and 5′–flanking base pairs.[a]

Conformation	Base Pair	H-Bond	Occupancy (%)
anti base-displaced intercalated	5′–C(5):G(18)	O2•••H21– N2	99.8
		N3•••H1– N1	99.9
		N4– H41•••O6	98.7
	3′–C(7):G(16)	O2•••H21– N2	99.5
		N3•••H1– N1	64.0
		N4– H41•••O6	52.4
syn 5′–intercalated	5′–C(5):G(18)	O2•••H21– N2	99.9
		N3•••H1– N1	99.9
		N4– H41•••O6	99.2
	3′–C(7):G(18)	O2•••H21– N2	99.4
		N3•••H1– N1	99.3
		N4– H41•••O6	94.4
	ALII-N^2-dG:C(17)	N1– H1•••N3	82.5
anti 3′–intercalated	5′–C(5):G(18)	O2•••H21– N2	99.0
		N3•••H1– N1	99.3
		N4– H41•••O6	98.7
	3′–C(7):G(16)	O2•••H21– N2	99.9
		N3•••H1– N1	99.9
		N4– H41•••O6	99.6
	ALI-N^2-dG:C(17)	N4– H41•••O6	93.1
anti minor groove stacked	5′–C(5):G(18)	O2•••H21– N2	99.8
		N3•••H1– N1	99.6
		N4– H41•••O6	95.8
	3′–C(7):G(16)	O2•••H21– N2	97.2
		N3•••H1– N1	99.8
		N4– H41•••O6	97.6
	ALI-N^2-dG:C(17)	N4– H42•••O6	88.7
syn base-displaced intercalated	5′–C(5):G(18)	O2•••H21– N2	99.9
		N3•••H1– N1	99.9
		N4– H41•••O6	98.7
	3′–C(7):G(16)	O2•••H21– N2	99.7
		N3•••H1– N1	99.8
		N4– H41•••O6	98.0

syn 3′,5′–intercalated	5′–C(5):G(18)	O2•••H21– N2	99.9
		N3•••H1– N1	99.9
		N4– H41•••O6	98.0
	3′–C(7):G(16)	O2•••H21–N2	99.9
		N3•••H1– N1	100.0
		N4– H41•••O6	99.7
	ALI-N^2-dG:C(17)	O6••• H42– N4	98.3
	3′–C(7):C(17)	O2••• H41– N4	73.5
	C(17): 3′–G(16)	N3•••H22 –N2	96.5

[a]Hydrogen-bond distance cut-off within 3.4 Å heavy atom separation and 120˚ X–H–X angle.

Table B5: Cartesian coordinates of the nucleobase minimum for the ALII-N²-G adduct with θ = 23.8° and ϕ = 255.7°.

Calculated energy (in Hartrees) = -1436.96018716

	X	Y	Z
N	6.11677500	0.03593900	-0.61786600
C	6.77548100	-0.04498400	0.59805300
H	7.85212300	0.03846100	0.66423100
N	5.96134200	-0.22683300	1.60521800
C	4.70437800	-0.27022200	1.03203000
C	3.40975500	-0.44434100	1.64101800
O	3.08185000	-0.59272200	2.80401600
N	2.39516600	-0.42720000	0.61892600
H	1.44894800	-0.53575200	0.96871600
C	2.59760600	-0.25603200	-0.72173100
N	1.48715300	-0.30585100	-1.54556800
H	1.71299900	-0.01375900	-2.49069300
N	3.77863900	-0.08179500	-1.27035800
C	4.77985800	-0.10852400	-0.35265700
H	6.52161000	0.16869300	-1.53358000
C	0.16959500	-0.02784600	-1.06972000
C	-0.70115200	-1.05592700	-0.85441300
C	-0.27543100	1.33867700	-0.80604200
C	-2.02188000	-0.79060100	-0.37840800
C	-1.60355000	1.61676100	-0.33137400
C	0.60425100	2.42273700	-1.00944000
C	-2.71966600	-1.99035900	-0.23123100
C	-2.50551800	0.49903000	-0.10730800
C	-1.98197200	2.95251100	-0.09421800
H	1.61400100	2.22846000	-1.35713700
C	0.20490300	3.72730100	-0.76509700
C	-4.03592800	-1.99176400	0.22846200
C	-3.82852600	0.46719600	0.35583300
C	-1.09341300	3.99537900	-0.30751700
H	-2.98485600	3.15815100	0.26248500
H	0.90446100	4.54243500	-0.92731600
C	-4.54421600	-0.72658200	0.51117400
H	-4.62169600	-2.89421100	0.35871100
H	-1.40419800	5.01831400	-0.11611600
N	-0.59321200	-2.42861300	-0.99771800
C	-1.79504100	-3.08930700	-0.63970800
O	-1.97290400	-4.28962100	-0.68083400

H	0.22490400	-2.91883500	-1.32852600
O	-5.81369800	-0.44051700	0.95248600
O	-4.62389000	1.52627900	0.69961100
C	-5.82753900	0.95752600	1.24391100
H	-6.69351200	1.42471900	0.76731100
H	-5.83865900	1.11267100	2.33153500

Table B6: Cartesian coordinates of the nucleobase minimum for the ALII-N²-G adduct with θ = 336.1° and ϕ = 255.7°.

Calculated energy (in Hartrees) = -1436.96018717

	X	Y	Z
N	-6.11674000	0.03654800	-0.61774100
C	-6.77544400	-0.04483100	0.59815300
H	-7.85207500	0.03871900	0.66438200
N	-5.96130400	-0.22723000	1.60521700
C	-4.70435200	-0.27051000	1.03199300
C	-3.40974300	-0.44508200	1.64087600
O	-3.08183500	-0.59402200	2.80380100
N	-2.39516400	-0.42758800	0.61877900
H	-1.44896400	-0.53645900	0.96851000
C	-2.59759700	-0.25578200	-0.72180300
N	-1.48717000	-0.30535400	-1.54569600
H	-1.71303300	-0.01287400	-2.49069300
N	-3.77862600	-0.08116600	-1.27032200
C	-4.77983600	-0.10820400	-0.35262000
H	-6.52156600	0.16976200	-1.53339100
C	-0.16959800	-0.02752200	-1.06976800
C	0.70104600	-1.05570000	-0.85451200
C	0.27552200	1.33894100	-0.80589400
C	2.02178000	-0.79053100	-0.37844000
C	1.60366200	1.61686500	-0.33119200
C	-0.60408300	2.42309400	-1.00915700
C	2.71947900	-1.99035700	-0.23140600
C	2.50552700	0.49903400	-0.10723100
C	1.98219100	2.95255600	-0.09387800
H	-1.61385300	2.22894100	-1.35685300
C	-0.20463200	3.72759600	-0.76464400
C	4.03575500	-1.99191200	0.22824700
C	3.82854200	0.46705200	0.35587600
C	1.09371000	3.99551600	-0.30704700

H	2.98509600	3.15807900	0.26283500
H	-0.90412500	4.54281100	-0.92673800
C	4.54414600	-0.72679900	0.51108900
H	4.62145600	-2.89442000	0.35837600
H	1.40457300	5.01840600	-0.11553400
N	0.59301000	-2.42836400	-0.99799800
C	1.79479400	-3.08918800	-0.64006000
O	1.97259200	-4.28950400	-0.68137500
H	-0.22511300	-2.91847700	-1.32894900
O	5.81367700	-0.44085900	0.95235200
O	4.62401600	1.52604200	0.69969400
C	5.82748300	0.95710000	1.24420300
H	5.83824800	1.11191200	2.33188300
H	6.69361400	1.42439900	0.76801200

Table B7: Cartesian coordinates of the nucleoside minimum for the ALII-N^2-dG adduct with χ = 57.9 °, θ = 336.1° and ϕ = 255.7°.

Calculated energy (in Hartrees) = -1857.99374214

	X	Y	Z
N	-4.41165500	1.01999300	0.81913900
C	-4.81851800	1.87682400	1.83605600
H	-5.78978200	1.76112500	2.29963400
N	-3.93471300	2.79135300	2.13490200
C	-2.86503700	2.52210200	1.30068700
C	-1.56089200	3.13916700	1.22351700
O	-1.08632200	4.07901100	1.83062600
N	-0.75688100	2.45763700	0.23551500
H	0.17097000	2.85149600	0.11078400
C	-1.14113800	1.39859400	-0.52868400
N	-0.25073500	0.89919400	-1.47083800
H	-0.70556200	0.16972100	-2.01362300
N	-2.33002300	0.84256000	-0.43396500
C	-3.14211200	1.43243100	0.48417100
C	1.08540200	0.55386700	-1.06191300
C	1.34702600	-0.69506500	-0.57623500
C	2.16592800	1.52194600	-1.14889100
C	2.66140700	-1.02323000	-0.12892300
C	3.48747000	1.21010600	-0.67240900
C	1.93758100	2.80150300	-1.70364000
C	2.67559300	-2.33383500	0.35214700
C	3.73484300	-0.11875400	-0.13836300

C	4.49081800	2.19555300	-0.74659700
H	0.95476800	3.02405600	-2.10872200
C	2.94642600	3.75058400	-1.76211400
C	3.84585900	-2.87567100	0.88220700
C	4.89819400	-0.69379000	0.39255800
C	4.22726700	3.45031500	-1.27554900
H	5.48379500	1.95859100	-0.38136100
H	2.74162500	4.72670900	-2.19267100
C	4.93546900	-2.00784900	0.87385100
H	3.91202800	-3.88328300	1.27609600
H	5.01579100	4.19590700	-1.31793500
N	0.55168000	-1.81167700	-0.38212300
C	1.29057900	-2.87381900	0.18115700
O	0.83911000	-3.97030800	0.45585300
H	-0.42714200	-1.90879100	-0.67249300
O	6.19496100	-2.27883300	1.35658200
O	6.12638200	-0.11024900	0.55398600
C	7.00313500	-1.15331400	1.01330800
H	7.54455200	-0.80462800	1.89717900
C	-5.04475100	-0.22525700	0.42698900
O	-4.23119000	-1.29565500	0.88945300
H	-6.02460300	-0.24797900	0.92114300
C	-5.18949100	-0.44620000	-1.08577600
C	-4.31260400	-2.42058700	-0.00950700
C	-5.21100500	-1.97678300	-1.18340900
H	-4.32067500	-0.04245600	-1.60877600
H	-6.09502100	0.01666900	-1.48909800
C	-2.89327000	-2.82573500	-0.39133900
H	-4.80086500	-3.25594400	0.50736200
H	-4.82044500	-2.33526600	-2.14469300
H	-2.92061000	-3.73993600	-0.99369100
H	-2.32705500	-3.04383900	0.52570600
O	-2.23417200	-1.82956100	-1.17359100
H	-2.38758300	-0.94472800	-0.75598900
O	-6.50965200	-2.50879700	-0.91911900
H	-7.09157500	-2.25123100	-1.65117100
H	7.69567400	-1.42311400	0.20380100

Table B8: Cartesian coordinates of the nucleoside minimum for the ALII-N^2-dG adduct with χ = 203.0 °, θ = 188.6° and ϕ = 318.5°.

Calculated energy (in Hartrees) = - 1857.98599002

	X	Y	Z
N	3.62046600	0.62810200	0.40835800
C	4.87035700	1.22664200	0.44901700
H	5.74424600	0.66707200	0.74674100
N	4.84150000	2.50379300	0.15493400
C	3.51216000	2.76891000	-0.10171700
C	2.85373900	3.99014400	-0.48792000
O	3.28061000	5.11111800	-0.68847500
N	1.43434300	3.74837600	-0.63591900
H	0.92260200	4.59024700	-0.87626800
C	0.77787000	2.55616300	-0.44806800
N	-0.58401000	2.58298400	-0.63855900
H	-0.96809100	3.46484800	-0.94648600
N	1.40149700	1.45140700	-0.11086500
C	2.74237200	1.61801300	0.05611100
C	-1.54891100	1.58410100	-0.36803200
C	-1.38044800	0.25005400	-0.62355900
C	-2.84161300	2.02665300	0.16592800
C	-2.46028000	-0.65969500	-0.42218500
C	-3.94748700	1.11937900	0.31562800
C	-3.04185900	3.37046200	0.55493500
C	-2.05559300	-1.95219200	-0.77351900
C	-3.74263900	-0.28218800	0.00214300
C	-5.18317000	1.60594900	0.78497200
H	-2.21062600	4.06809700	0.53611800
C	-4.26730300	3.82224300	1.01923200
C	-2.94883800	-3.02227400	-0.70034100
C	-4.61697000	-1.37507900	0.06978300
C	-5.35017100	2.93940300	1.12315300
H	-6.01072900	0.91338600	0.88663100
H	-4.37887900	4.86185200	1.31373000
C	-4.22445700	-2.67683000	-0.26822000
H	-2.68114100	-4.03921800	-0.96254300
H	-6.31219200	3.29207900	1.48303800
N	-0.31844000	-0.49668600	-1.12822000
C	-0.62934900	-1.85932100	-1.17159300

O	0.15805400	-2.75198400	-1.47880600
H	0.62866000	-0.13538500	-1.05105500
O	-5.29834500	-3.52517100	-0.12981700
O	-5.93530800	-1.37460400	0.43454500
C	-6.33718800	-2.75603500	0.47426000
H	-7.26384200	-2.87633400	-0.09351500
C	3.26855000	-0.77140500	0.60039800
O	4.24462000	-1.36203900	1.44099200
H	2.27772500	-0.79331600	1.07005700
C	3.28202300	-1.62682300	-0.66896100
C	4.28670200	-2.78983400	1.17049600
C	3.36841100	-3.02556300	-0.04889400
H	4.18466300	-1.40708300	-1.24668800
H	2.40050700	-1.51100700	-1.30206900
C	5.74562000	-3.17408400	0.96108200
H	3.87326400	-3.32492300	2.03347400
H	3.80103000	-3.75947400	-0.74292100
H	5.81589300	-4.25208600	0.78283500
H	6.31314800	-2.94395300	1.87610200
O	6.31553000	-2.53785600	-0.17540000
H	6.28659000	-1.58502600	0.00500500
O	2.11373900	-3.45088500	0.45923900
H	1.45850600	-3.40443400	-0.26491800
H	-6.46920400	-3.06059000	1.52163500

Table B9: Cartesian coordinates of the constrained nucleoside minimum for the ALII-N^2-dG adduct with χ = 64.8 °, θ = 58.8° and ϕ = 77.3°.

Calculated energy (in Hartrees) = -1857.97491361

	X	Y	Z
N	-4.21444000	1.10343700	0.83513500
C	-4.56274400	1.84988500	1.95517400
H	-5.50733700	1.68056600	2.45640300
N	-3.66475000	2.73194400	2.30123800
C	-2.64339200	2.55455500	1.38689000
C	-1.34976100	3.19272400	1.31204800
O	-0.85440100	4.07212000	1.99313000
N	-0.59626300	2.62198900	0.22962100
H	0.32982600	3.01918100	0.10592500
C	-1.02357200	1.64111000	-0.61888100
N	-0.13228900	1.23723300	-1.62216300

H	-0.61985500	0.57392500	-2.21887500
N	-2.20058200	1.07515200	-0.54884300
C	-2.96087300	1.55334400	0.47193800
C	1.13856400	0.72792500	-1.16728300
C	1.23625700	-0.53776200	-0.66686400
C	2.32417700	1.56802000	-1.19588200
C	2.48677900	-1.01990100	-0.17894200
C	3.58553700	1.10217900	-0.68332500
C	2.25552200	2.87619900	-1.72464900
C	2.33058900	-2.32330700	0.29863300
C	3.65910500	-0.24807100	-0.15072200
C	4.69642900	1.96750200	-0.71158500
H	1.31173300	3.21567200	-2.14032200
C	3.36634300	3.70516000	-1.73607600
C	3.41133100	-2.99947700	0.86403500
C	4.72711100	-0.95497900	0.41997400
C	4.59163300	3.25174600	-1.22461900
H	5.64467400	1.61478300	-0.32157400
H	3.28520800	4.70757800	-2.14707200
C	4.59383100	-2.26444100	0.89778500
H	3.34703400	-4.00939900	1.25223100
H	5.46045200	3.90351600	-1.23179900
N	0.30963200	-1.55401900	-0.49853400
C	0.89702500	-2.69114600	0.08715300
O	0.30551800	-3.72552100	0.35067100
H	-0.67015000	-1.55138100	-0.79017800
O	5.81036900	-2.69952500	1.37260800
O	6.01932400	-0.53430100	0.58987000
C	6.66601900	-1.55716700	1.36607300
H	6.81100600	-1.19170100	2.39264400
C	-4.88865200	-0.10391800	0.41316300
O	-4.15839400	-1.22983400	0.89910800
H	-5.88452600	-0.07541000	0.87475600
C	-5.00361200	-0.33428000	-1.09575300
C	-4.35103800	-2.35016200	0.02535800
C	-5.24070100	-1.84225700	-1.14151300
H	-4.05197300	-0.08738600	-1.56783400
H	-5.82300000	0.22325800	-1.55382200
C	-2.99435500	-2.90784300	-0.39664900
H	-4.89403700	-3.13354000	0.57384600
H	-4.94936500	-2.27962600	-2.10588700
H	-3.11547900	-3.93060400	-0.77703400
H	-2.32576700	-2.95627200	0.46679300
O	-2.35159200	-2.08117800	-1.38497300
H	-2.45905100	-2.49260300	-2.25312900
O	-6.62506800	-2.06692900	-0.87144100

H	-6.87008500	-2.93899600	-1.21345900
H	7.62018800	-1.81418900	0.8985260

Table B10: Cartesian coordinates of the constrained nucleoside minimum for the ALII-N²-dG adduct with χ = 223.9°, θ = 177.4° and ϕ = 319.4°.

Calculated energy (in Hartrees) = -1857.97371221

	X	Y	Z
N	3.51723000	0.48322500	0.22348800
C	4.80495600	0.97835900	0.06582000
H	5.65734800	0.33395500	0.22358400
N	4.83053600	2.24215900	-0.28061600
C	3.49977500	2.60758200	-0.36266700
C	2.88718300	3.85765500	-0.72758900
O	3.36442900	4.93440500	-1.03622300
N	1.44645200	3.71749400	-0.70580200
H	0.96976200	4.57038300	-0.97693800
C	0.73743900	2.58114300	-0.40024500
N	-0.63044100	2.67097400	-0.48613600
H	-1.01030500	3.57427800	-0.73175500
N	1.31763500	1.45533700	-0.05437300
C	2.67527500	1.52783100	-0.05206000
C	-1.58704800	1.64785800	-0.27816300
C	-1.43706600	0.35831500	-0.71397400
C	-2.83955700	2.01566100	0.38449200
C	-2.50420500	-0.57685000	-0.55450700
C	-3.93628700	1.09169100	0.49313800
C	-3.01067600	3.30361500	0.94042500
C	-2.12622000	-1.81582800	-1.08225000
C	-3.75439200	-0.26051500	-0.00078400
C	-5.13859700	1.51338400	1.09383400
H	-2.17882200	4.00073700	0.94560600
C	-4.20330100	3.69095400	1.53122700
C	-3.01298700	-2.89173900	-1.07502100
C	-4.62369900	-1.35964100	-0.00490300
C	-5.27989700	2.79645900	1.59849300
H	-5.96059500	0.81015900	1.16350400
H	-4.29426900	4.68854700	1.95184300
C	-4.25701000	-2.60815900	-0.51992700
H	-2.76327800	-3.86927000	-1.47100900
H	-6.21646100	3.09948800	2.05747900
N	-0.41394000	-0.30784500	-1.37968100

C	-0.72160900	-1.67222800	-1.55895900
O	0.04572500	-2.50950100	-2.00232000
H	0.54251800	0.01927700	-1.29869800
O	-5.32198200	-3.47493400	-0.41236200
O	-5.91756100	-1.41164000	0.44437200
C	-6.30627900	-2.79334300	0.36318200
H	-7.27925500	-2.86445500	-0.13054700
C	3.11991000	-0.86188400	0.61895800
O	3.96074900	-1.27538000	1.69050200
H	2.08418500	-0.78731400	0.96501400
C	3.27035700	-1.91549500	-0.49028400
C	4.41663200	-2.61799200	1.48994100
C	3.59437200	-3.17498300	0.30457800
H	4.11376600	-1.65352300	-1.13222500
H	2.36332000	-2.04233700	-1.08791900
C	5.92873500	-2.63909300	1.29960100
H	4.19046900	-3.18882900	2.40207000
H	4.16626600	-3.90622000	-0.28326500
H	6.27005500	-3.68735100	1.32084600
H	6.39087800	-2.10941500	2.14553500
O	6.25321300	-2.01736500	0.05919000
H	7.21525100	-2.03576000	-0.05238900
O	2.35284100	-3.71556200	0.74865200
H	2.53619600	-4.50244400	1.28500400
H	-6.34347700	-3.21605100	1.37746100

Table B11: Cartesian coordinates of the nucleotide minimum for the ALII-N²-dG adduct with $\chi = 67.7^{\circ}$, $\theta = 21.8^{\circ}$ and $\phi = 99.5^{\circ}$.

Calculated energy (in Hartrees) = -2587.58678145

	X	Y	Z
N	4.21446300	1.99421100	-0.75246600
C	4.54109200	2.96386600	-1.69081900
H	5.56836400	3.10270700	-2.00187700
N	3.51336700	3.65115400	-2.11763400
C	2.43472600	3.10179400	-1.44241600
C	1.03748500	3.41073000	-1.53898600
O	0.46415000	4.25808600	-2.22069000
N	0.28120800	2.55042100	-0.69506800
H	-0.72143200	2.70562100	-0.72057400
C	0.79179400	1.57190000	0.12084800
N	-0.07671300	0.85133200	0.89804000

H	0.37533300	0.03689200	1.32152000
N	2.07966100	1.30391400	0.21266900
C	2.84476200	2.07218300	-0.59263900
C	-1.46841800	0.71080700	0.62870600
C	-1.93277200	-0.41930200	0.02103100
C	-2.43584900	1.71604400	1.05614100
C	-3.32956400	-0.62777100	-0.15555300
C	-3.84854400	1.53767500	0.83848100
C	-1.99955600	2.89324400	1.70338300
C	-3.54211700	-1.87760300	-0.74693400
C	-4.30740700	0.30498200	0.21914500
C	-4.73653100	2.55026800	1.25249200
H	-0.93990500	3.02037200	1.89890900
C	-2.89817700	3.87169500	2.09931900
C	-4.84261300	-2.31016300	-1.02166800
C	-5.60006200	-0.15358300	-0.07530300
C	-4.27240300	3.70377000	1.86790200
H	-5.79938000	2.41566900	1.08604300
H	-2.53424600	4.76766400	2.59372100
C	-5.83736300	-1.40366900	-0.66643400
H	-5.07599100	-3.26735100	-1.47373800
H	-4.97538400	4.47197000	2.17604700
N	-1.28462300	-1.53780200	-0.48057400
C	-2.19583300	-2.47891600	-0.94079200
O	-1.88687200	-3.58879100	-1.39494600
H	-0.25903200	-1.66161000	-0.56020100
O	-7.18971300	-1.58751700	-0.80954200
O	-6.78425500	0.47655300	0.15997800
C	-7.81468400	-0.36444000	-0.40078200
H	-8.25337300	0.13903400	-1.26933600
C	5.14417800	1.00224900	-0.25327900
O	4.89611600	-0.23935800	-0.91309800
H	6.14440800	1.37435500	-0.51065000
C	5.07373400	0.65977300	1.23419300
C	5.44350500	-1.30161000	-0.10781600
C	5.83707800	-0.66528700	1.25794200
H	4.03279800	0.50289100	1.51962900
H	5.53739700	1.41379800	1.87313000
C	4.44498000	-2.44475300	-0.01815400
H	6.35383900	-1.68027900	-0.59136600
H	5.54434500	-1.29859000	2.10053400
H	4.96127100	-3.35412700	0.31110300
H	4.00888600	-2.62148800	-1.00700400
O	3.42256800	-2.12419800	0.94013900
O	7.23428500	-0.36392800	1.29747100
H	7.69795600	-1.14015500	1.64738200

H	-8.56479400	-0.57118000	0.36595200
P	1.87780700	-2.55603200	0.64791300
O	1.63364800	-4.02434000	0.93076500
O	1.42365800	-2.02395000	-0.69846500
O	1.15410400	-1.64591600	1.81403200
Na	0.23186600	-4.61477000	-0.84239400
H	0.95961700	-2.18228600	2.60047700

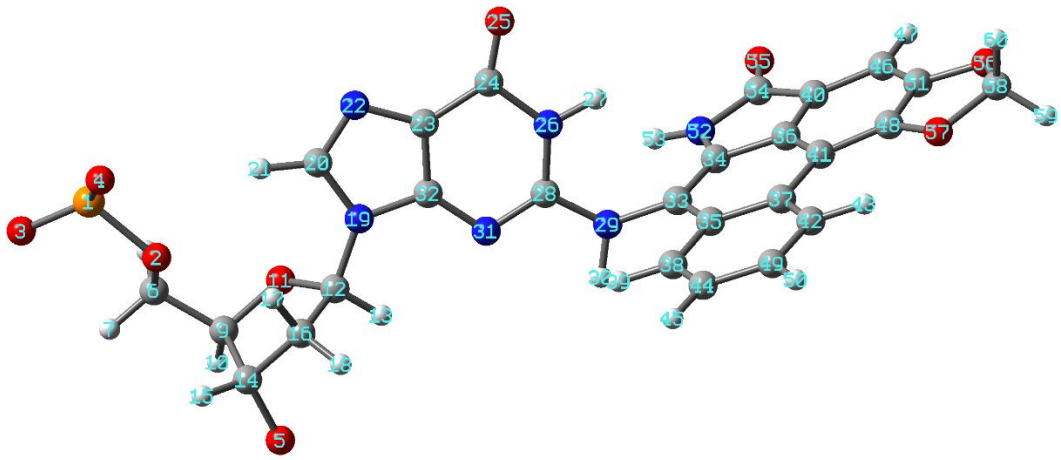
Table B12: Cartesian coordinates of the nucleotide minimum for the ALII-N^2-dG adduct with χ = 282.1°, θ = 185.7° and ϕ = 312.2°.

Calculated energy (in Hartrees) = -2587.57269630

	X	Y	Z
N	2.36174100	-0.12356000	-1.07093300
C	3.66626800	0.26847600	-1.29436400
H	4.49115500	-0.42521100	-1.30131300
N	3.79561900	1.56876400	-1.44544600
C	2.50985500	2.06807800	-1.30791000
C	2.03051400	3.41793900	-1.32621200
O	2.65327700	4.46688500	-1.47086900
N	0.62139000	3.44369600	-1.12731800
H	0.22224100	4.37692300	-1.11630100
C	-0.18424500	2.34828100	-0.90291700
N	-1.51291600	2.60290100	-0.72452500
H	-1.80228800	3.57102500	-0.77507700
N	0.27924300	1.11421200	-0.85990300
C	1.61378200	1.03524600	-1.06482500
C	-2.52544000	1.67748600	-0.37072000
C	-2.69050100	0.46223600	-0.97700900
C	-3.48232200	2.07661200	0.66452600
C	-3.80015700	-0.36398300	-0.62478500
C	-4.62645500	1.26692400	0.98587700
C	-3.30253700	3.27914900	1.38488700
C	-3.77479300	-1.53591900	-1.38678700
C	-4.78876700	-0.00412800	0.30445000
C	-5.53790300	1.71563600	1.96282000
H	-2.42010800	3.88514100	1.21088800
C	-4.21324900	3.69481100	2.34430200
C	-4.78082100	-2.49522100	-1.25113700
C	-5.78293500	-0.98536000	0.42394900

C	-5.34365400	2.91564100	2.62969100
H	-6.40037800	1.10155500	2.19489000
H	-4.04063000	4.62396800	2.87952200
C	-5.76579900	-2.17017800	-0.32488800
H	-4.80720900	-3.41787500	-1.81949000
H	-6.05899600	3.24182400	3.37878500
N	-2.00012900	-0.20370200	-1.98776900
C	-2.56075700	-1.46926700	-2.23969600
O	-2.09694100	-2.30051400	-3.00990200
H	-1.03026600	0.00515500	-2.19326800
O	-6.87488600	-2.91881500	-0.01599100
O	-6.89023300	-0.96279400	1.22065100
C	-7.51381700	-2.25609000	1.08160500
H	-8.57509700	-2.12112600	0.86390200
C	1.88088000	-1.40118200	-0.54640600
O	1.97035600	-1.36784000	0.87588000
H	0.82590300	-1.46038500	-0.82533900
C	2.67411300	-2.62635900	-1.02226000
C	2.67240800	-2.52030300	1.37476100
C	2.67196300	-3.53048500	0.20881700
H	3.70082300	-2.36682000	-1.27594800
H	2.20417500	-3.11097300	-1.88000300
C	4.04315600	-2.12667400	1.91135600
H	2.09259600	-2.91887700	2.21694200
H	3.54670800	-4.19213600	0.24127300
H	4.48779300	-2.97837200	2.44209200
H	3.91705500	-1.29772200	2.61540000
O	4.90312000	-1.73728700	0.83369600
O	1.45755400	-4.27956700	0.17169800
H	1.42801200	-4.84244800	0.96233200
H	-7.35976100	-2.82838300	2.00408900
P	6.00493700	-0.52928000	1.05799800
O	7.04366400	-1.17582400	2.15818400
O	6.66406600	-0.34307800	-0.28484000
O	5.35176700	0.65414200	1.74119400
H	7.59798300	-1.85030900	1.73023600
Na	5.55499300	2.30885600	0.13702900

Table B13: Mol2 file of the ALII-N²-dG adduct generated using the RED program (atom numbering provided in the figure below).



1 P	4.7990	2.3380	-0.2990 P	1 LIG	1.165900	
2 O5'	3.1440	2.4150	0.0000 OS	1 LIG	-0.506200	
3 O1P	5.0790	2.9190	-1.6510 O2	1 LIG	-0.767400	
4 O2P	5.4600	2.7480	0.9820 O2	1 LIG	-0.767400	
5 O3'	0.0000	0.0000	0.0000 OS	1 LIG	-0.545300	
6 C5'	2.2990	2.1870	-1.1060 CI	1 LIG	-0.007400	
7 H5'1	1.3770	2.7570	-0.9340 H1	1 LIG	0.076000	
8 H5'2	2.7400	2.5490	-2.0450 H1	1 LIG	0.076000	
9 C4'	1.8810	0.7330	-1.2880 CT	1 LIG	0.183300	
10 H4'	1.0270	0.7050	-1.9870 H1	1 LIG	0.096100	
11 O4'	2.9440	-0.0680	-1.8340 OS	1 LIG	-0.406400	
12 C1'	2.4870	-1.3980	-1.6180 CT	1 LIG	0.274800	
13 H1'	1.6060	-1.5780	-2.2540 H2	1 LIG	0.009200	
14 C3'	1.4260	-0.0000	0.0000 CT	1 LIG	0.139700	
15 H3'	1.8190	0.5190	0.8800 H1	1 LIG	0.071500	
16 C2'	2.0550	-1.4090	-0.1460 CT	1 LIG	-0.049800	
17 H2'1	2.9330	-1.5250	0.4930 HC	1 LIG	0.043100	
18 H2'2	1.3350	-2.2020	0.0760 HC	1 LIG	0.043100	
19 N9	3.4630	-2.3760	-2.0490 N*	1 LIG	-0.095500	
20 C8	3.3290	-3.1000	-3.2290 CK	1 LIG	0.058700	
21 H8	2.5050	-2.9010	-3.9020 H5	1 LIG	0.197700	
22 N7	4.2570	-3.9980	-3.4090 NB	1 LIG	-0.523000	
23 C5	5.0620	-3.8740	-2.2970 CB	1 LIG	0.207600	
24 C6	6.2310	-4.6320	-1.9340 C	1 LIG	0.415800	
25 O6	6.8130	-5.5370	-2.5030 O	1 LIG	-0.523200	

26 N1	6.7240	-4.1700	-0.6650	NA	1 LIG	-0.336000
27 H1	7.5710	-4.6480	-0.3760	H	1 LIG	0.314900
28 C2	6.1990	-3.1540	0.0970	CA	1 LIG	0.316700
29 N2	6.8590	-2.9210	1.2800	N2	1 LIG	-0.298400
30 H2	7.5870	-3.5820	1.5130	H	1 LIG	0.281100
31 N3	5.1390	-2.4620	-0.2570	NC	1 LIG	-0.449000
32 C4	4.6010	-2.8720	-1.4430	CB	1 LIG	0.212600
33 C10	6.7540	-1.8360	2.1780	CM	1 LIG	0.020700
34 C11	6.5420	-0.5370	1.8030	CN	1 LIG	0.088200
35 C12	7.0030	-2.1210	3.5930	CM	1 LIG	0.030300
36 C13	6.6750	0.5080	2.7670	CB	1 LIG	-0.028000
37 C14	7.1640	-1.0690	4.5580	CA	1 LIG	0.033800
38 C15	7.1000	-3.4550	4.0510	CA	1 LIG	-0.141800
39 H8	6.9100	-4.2720	3.3620	HA	1 LIG	0.123100
40 C16	6.5160	1.7470	2.1380	CB	1 LIG	-0.059500
41 C17	7.0110	0.3050	4.1150	CA	1 LIG	-0.100900
42 C18	7.4510	-1.3990	5.8970	CA	1 LIG	-0.058900
43 H9	7.5770	-0.5980	6.6160	HA	1 LIG	0.095800
44 C19	7.3780	-3.7520	5.3760	CA	1 LIG	-0.164200
45 H10	7.4370	-4.7900	5.6920	HA	1 LIG	0.161700
46 C20	6.6540	2.9350	2.8570	CA	1 LIG	-0.296800
47 H11	6.5400	3.9160	2.4090	HA	1 LIG	0.228300
48 C21	7.1330	1.5140	4.8140	CB	1 LIG	0.211300
49 C22	7.5660	-2.7200	6.3040	CA	1 LIG	-0.189100
50 H12	7.7880	-2.9500	7.3420	HA	1 LIG	0.162400
51 C23	6.9600	2.7610	4.2010	CB	1 LIG	0.266700
52 N5	6.2960	0.0750	0.5760	NA	1 LIG	-0.435400
53 H13	5.8540	-0.4190	-0.1930	H	1 LIG	0.321600
54 C24	6.2080	1.4670	0.7140	C	1 LIG	0.620200
55 O6	5.9220	2.2440	-0.1910	O	1 LIG	-0.547500
56 O7	7.1740	3.7560	5.1270	OS	1 LIG	-0.353300
57 O8	7.4510	1.6940	6.1340	OS	1 LIG	-0.287800
58 C25	7.3220	3.1040	6.3860	CT	1 LIG	0.181700
59 H14	8.2260	3.4630	6.8850	H2	1 LIG	0.104300
60 H15	6.4320	3.2770	7.0060	H2	1 LIG	0.104300

APPENDIX C

Effect of Base Sequence Context on the Conformational Heterogeneity of Aristolactam-I Adducted DNA

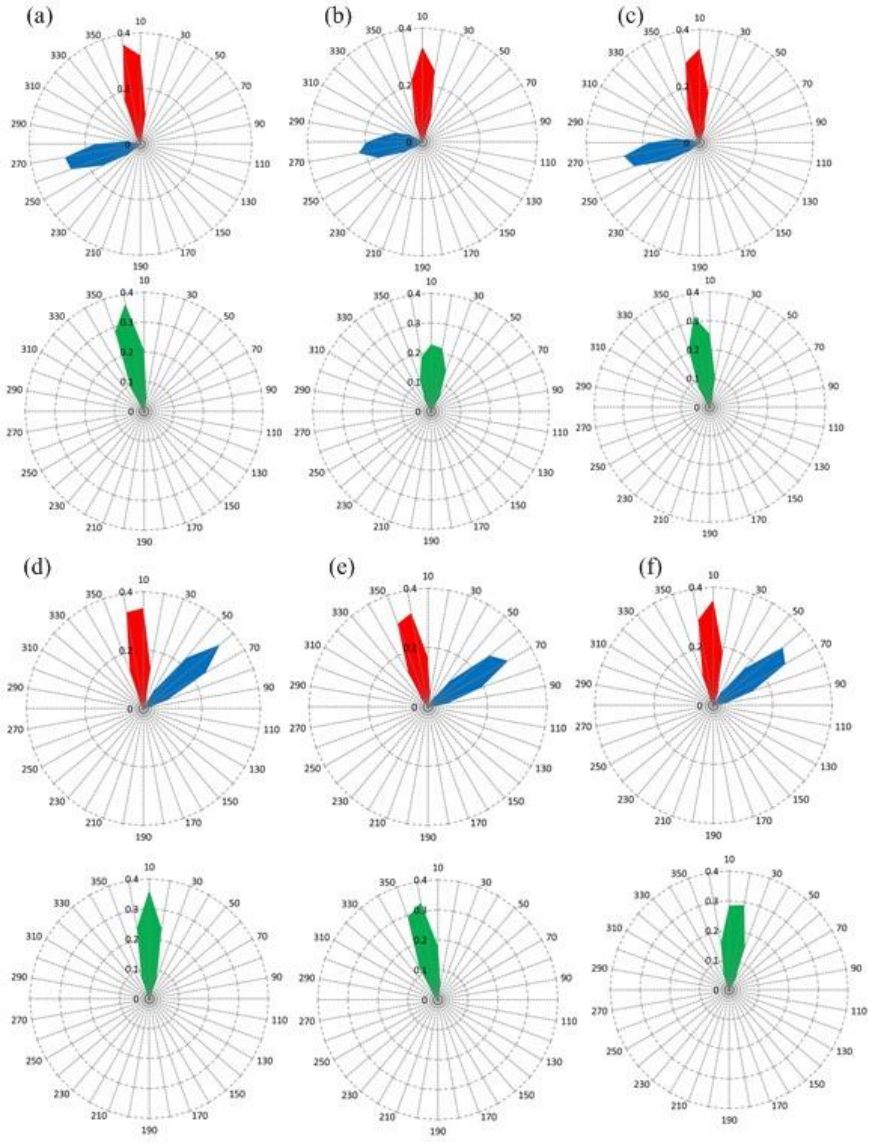


Figure C1: Radar plots for the probability distributions of χ (blue), θ (red) and ϕ (green) for* anti *(a, b and c) or* syn *(d, e and f) ALI-N^6-dA in the GXG sequence in the base-displaced intercalated (a and d), 5′–intercalated (b and e) or 3′–intercalated (c and f) adducted DNA conformations.

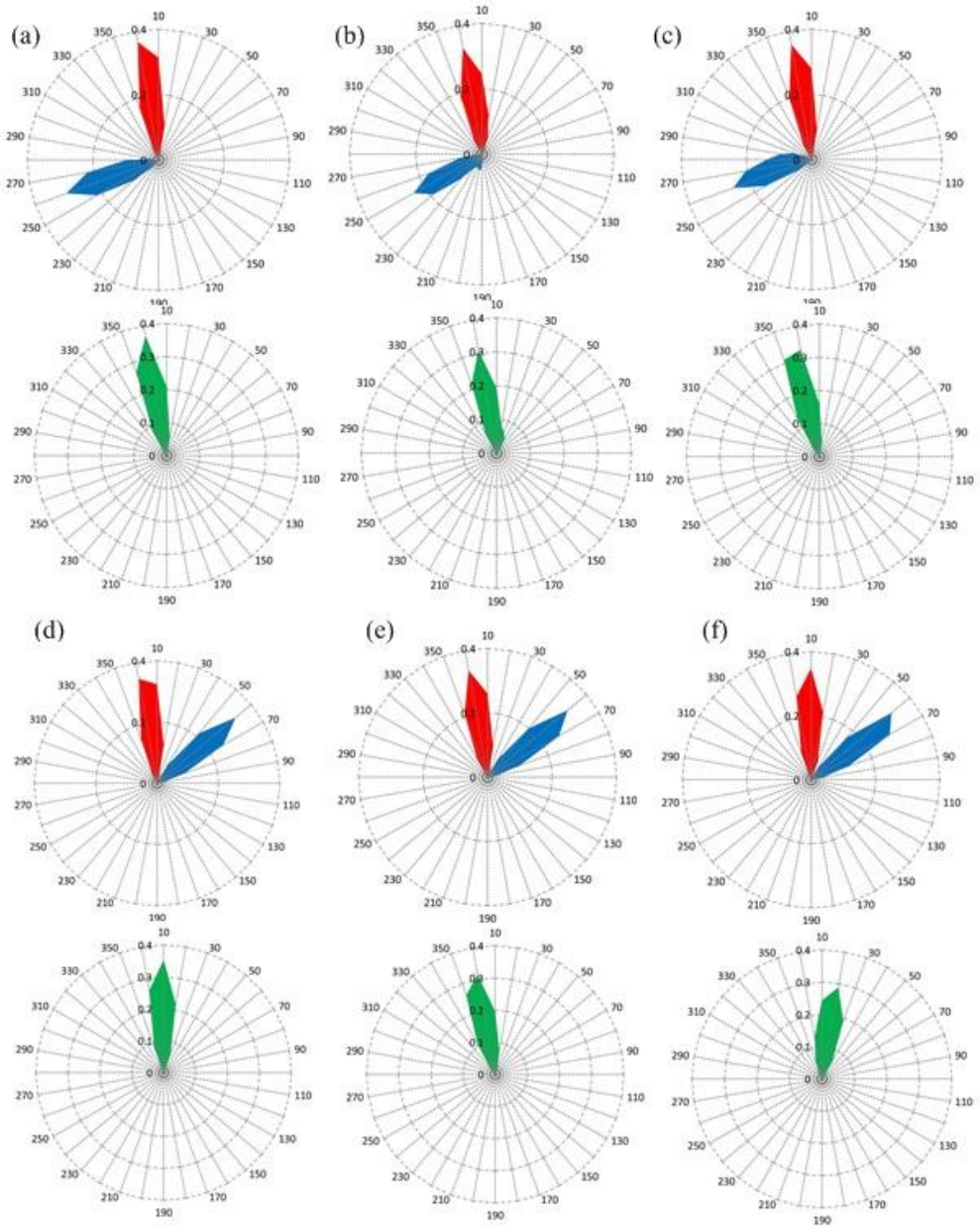


Figure C2: Radar plots for the probability distributions of χ (blue), θ (red) and ϕ (green) for* anti *(a, b and c) or* syn *(d, e and f) ALI-N^6-dA in the CXG sequence in the base-displaced intercalated (a and d), 5′–intercalated (b and e) or 3′–intercalated (c and f) adducted DNA conformations.

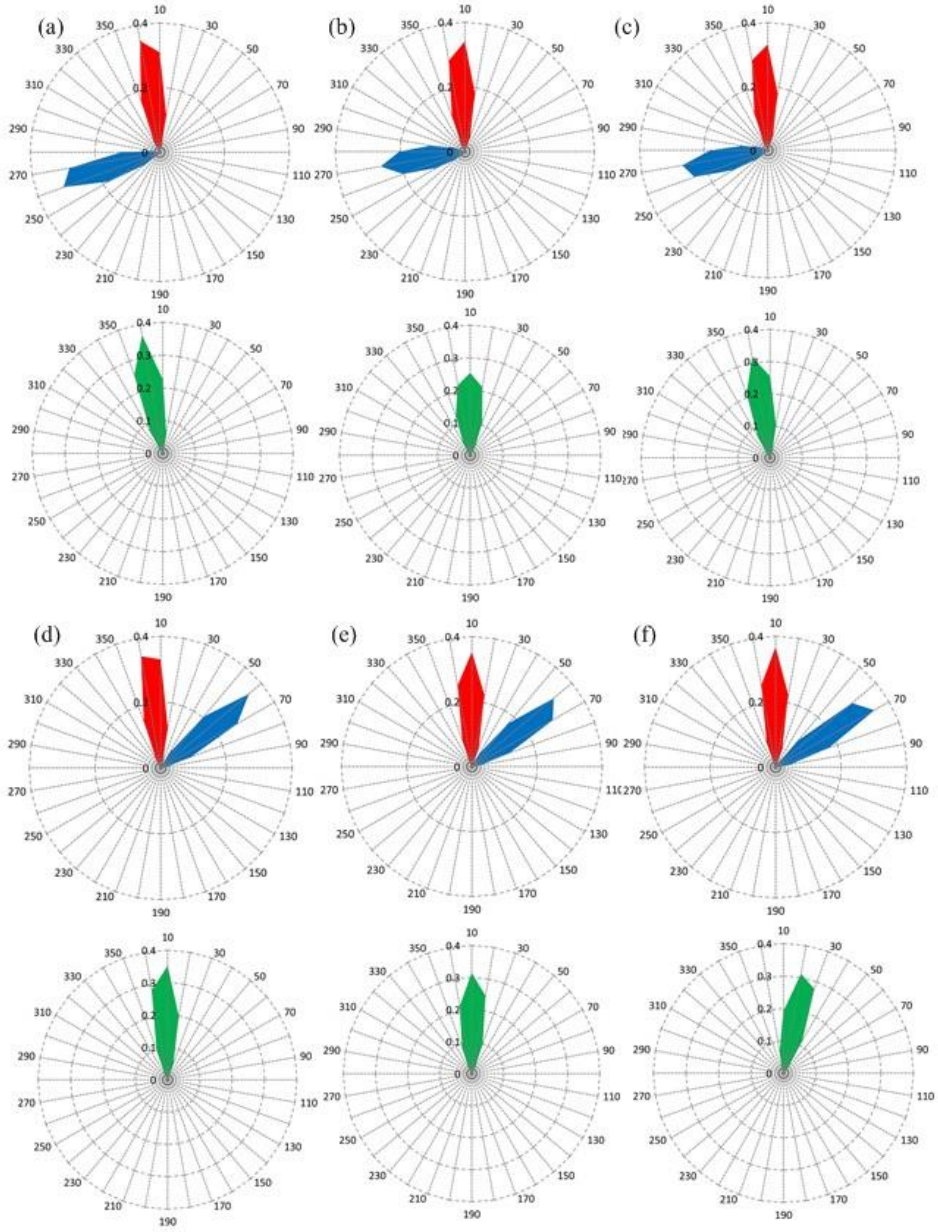


Figure C3: Radar plots for the probability distributions of χ (blue), θ (red) and ϕ (green) for* anti *(a, b and c) or* syn *(d, e and f) ALI-N^6-dA in the GXC sequence in the base-displaced intercalated (a and d), 5′–intercalated (b and e) or 3′–intercalated (c and f) adducted DNA conformations.

Table C1: RMSD of all heavy atoms in the trimer containing the damaged base pair and the 5′ and 3′–flanking bases in the base-displaced intercalated adducted DNA conformations during 20 and 320 ns simulations in the CXC sequence context.

Conformer	RMSD (Å)	
	20 ns	**320 ns**
anti base-displaced intercalated	0.82±0.2	0.96±0.3
syn base-displaced intercalated	1.09±0.2	1.21±0.3

Table C2: Backbone RMSD for each adducted DNA conformation in different sequence contexts.

Sequence Context	Conformers	RMSD (Å)
GXG	*anti* base-displaced	1.82±0.33
	anti 5′–intercalated	4.11±1.51
	anti 3′–intercalated	3.33±0.63
	syn base-displaced	2.05 ±0.37
	syn 5′–intercalated	3.73 ±0.77
	syn 3′–intercalated	2.21±0.53
CXG	*anti* base-displaced	1.90±0.39
	anti 5′–intercalated	2.24±0.46
	anti 3′–intercalated	3.11±0.75
	syn base-displaced	2.19±0.46
	syn 5′–intercalated	2.67±0.47
	syn 3′–intercalated	2.21±0.55
GXC	*anti* base-displaced	2.04 ±0.40
	anti 5′–intercalated	2.94±0.91
	anti 3′–intercalated	2.71±0.58
	syn base-displaced	2.04±0.41
	syn 5′–intercalated	3.01±0.76
	syn 3′–intercalated	2.89±0.72

Table C3: 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.[a,b]

Sequence Context	Conformation	Base Pair	H-Bond	Occupancy (%)
CXC	*anti* base-displaced	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.9
		3′–C(7):G(16)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.2
	syn base-displaced	5′–C(5):G(18)	O2•••H21–N2	99.7
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
		3′–C(7):G(16)	O2•••H21–N2	99.8
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
	anti 5′–intercalated	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	99.1
		3′–C(7):G(16)	N3•••H1–N1	99.9
			N4–H4•••O6	99.1
		3′–C(7):T(17)	O2•••H2–N2	96.9
			N4–H4•••O4	45.8
	syn 5′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.9
			N3•••H1–N1	99.9
			N4–H41•••O6	98.9
		3′–C(7):T(17)	O4•••H41–N4	83.7
			N3•••H3–N3	74.4
	anti 3′–intercalated	5′–C(5):T(17)	N4–H4•••O4	36.5
		3′–C(7):G(16)	O2•••H2–N2	99.7
			N3•••H1–N1	99.9
			N4–H4•••O6	98.6
	syn 3′–intercalated	5′–C(5):G(18)	O2•••H21–N2	99.5
			N3•••H1–N1	99.1
			N4–H41•••O6	97.1
		3′–C(7):G(16)	O2•••H21–N2	99.5
			N3•••H1–N1	99.9

			N4–H41•••O6	98.7
GXG	*anti* base-displaced	5′–G(5):C(18)	N2–H2•••O2	99.9
			N1–H1••• N3	99.9
			O6•••N4–H4	99.2
		3′–G(7):C(16)	N2–H2•••O2	99.7
			N1–H1••• N3	99.9
			O6•••N4–H4	98.9
	syn base-displaced	5′–G(5):C(18)	N2–H2•••O2	99.9
			N1–H1••• N3	99.9
			O6•••N4–H4	99.3
		3′–G(7):C(16)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	99.1
	anti 5′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.9
			N1–H1••• N3	99.9
			O6•••N4–H4	99.9
		3′–G(7):C(16)	N2–H2•••O2	91.6
			N1–H1••• N3	92.5
			O6•••N4–H4	93.5
	syn 5′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	98.4
		3′–G(7):T(17)	O2•••N1–H1	47.2
	anti 3′–intercalated	5′–G(5):C(18)	N2–H2•••O2	34.1
			N1–H1••• N3	34.0
			O6•••N4–H4	32.5
		3′–G(7):C(16)	N2–H2•••O2	99.7
			N1–H1••• N3	99.2
			O6•••N4–H4	97.9
	syn 3′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.6
			N1–H1••• N3	99.7
			O6•••N4–H4	96.1
		3′–G(7):C(16)	N2–H2•••O2	99.6
			N1–H1••• N3	99.9

			O6•••N4–H4	99.2
CXG	*anti* base-displaced	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.8
		3′–G(7):C(16)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	98.7
	syn base-displaced	5′–C(5):G(18)	O2•••H2–N2	99.7
			N3•••H1–N1	99.9
			N4–H4•••O6	98.9
		3′–G(7):C(16)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	99.1
	anti 5′–intercalated	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.7
		3′–G(7):C(16)	N2–H2•••O2	98.1
			N1–H1••• N3	98.7
			O6•••N4–H4	98.6
	syn 5′–intercalated	5′–C(5):G(18)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.9
		3′–G(7):C(16)	N2–H2•••O2	94.5
			N1–H1••• N3	95.6
			O6•••N4–H4	95.5
	anti 3′–intercalated	5′–C(5):G(18)	O2•••H2–N2	44.6
			N3•••H1–N1	44.5
			N4–H4•••O6	50.2
		3′–G(7):C(16)	N2–H2•••O2	99.9
			N1–H1••• N3	99.9
			O6•••N4–H4	98.8
	syn 3′–intercalated	5′–C(5):G(18)	O2•••H2–N2	99.5
			N3•••H1–N1	99.4

			N4–H4•••O6	98.8
		3′–G(7):C(16)	N2–H2•••O2	99.7
			N1–H1••• N3	99.9
			O6•••N4–H4	99.1
GXC	*anti* base-displaced	5′–G(5):C(18)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	99.3
		3′–C(7):G(16)	O2•••H2–N2	99.9
			N3•••H1–N1	99.9
			N4–H4•••O6	98.3
	syn base-displaced	5′–G(5):C(18)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	99.4
		3′–C(7):G(16)	O2•••H2–N2	99.8
			N3•••H1–N1	99.9
			N4–H4•••O6	99.1
	anti 5′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	99.0
		3′–C(7):G(16)	O2•••H2–N2	95.3
		3′–C(7):T(17)	N4–H4••• O4	52.1
			N3•••H2–N2	39.7
	syn 5′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.8
			N1–H1••• N3	99.9
			O6•••N4–H4	98.6
		3′–C(7):G(16)	O2•••H2–N2	80.3
		3′–C(7):T(17)	N4–H4••• O4	89.7
			N3•••H3–N3	80.6
	anti 3′–intercalated	5′–G(5):C(18)	N2–H2•••O2	99.2
			N1–H1••• N3	99.3
			O6•••N4–H4	97.9
		3′–C(7):G(16)	O2•••H2–N2	99.8
			N3•••H1–N1	99.9
			N4–H4•••O6	98.4

syn 3′–	5′–G(5):C(18)	N2–H2•••O2	99.9
intercalated		N1–H1••• N3	99.9
		O6•••N4–H4	99.5
	3′–C(7):G(16)	O2•••H2–N2	99.9
		N3•••H1–N1	99.9
		N4–H4•••O6	98.5

[a]Hydrogen-bond distance cut-off within 3.4 Å heavy atom separation and 120° X–H–X angle. [b]Nucleotides in the 11-mer DNA are numbered starting from the 5′ side of each strand, with the strand containing the lesion numbered first.

Table C4: Average pseudostep parameters,[a] minor groove width,[b] and standard deviations for each adducted DNA conformer in different sequence contexts.

Sequence Context	Conformer	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)	Minor Groove (Å)
GXG	unmodified	0.1±0.7	−1.6±1.3	6.8±0.4	−1.9±5.0	5.4±7.0	65.0±6.0	7.2±1.5
	anti base-displaced	−1.5±0.7	1.3±1.1	6.7±0.3	0.7±4.8	−4.4±5.5	57.1±6.0	7.7±1.2
	anti 5′–intercalated	−1.9±1.1	1.6±1.8	9.5±0.6	−19.3±6.5	−1.1±9.8	46.2±11.3	16.3±4.5
	anti 3′–intercalated	−1.6±0.9	2.1±1.0	10.1±0.4	−4.2±5.3	−4.4±8.2	46.3±7.1	10.5±1.8
	syn base-displaced	−1.4±0.9	0.7±1.1	6.7±0.3	2.5±4.9	6.1±6.1	51.4±6.3	9.1±1.7
	syn 5′–intercalated	−1.7±1.2	0.9±1.6	9.6±0.8	−9.1±9.3	−3.8±10.4	54.4±11.8	14.6±3.3
	syn 3′–intercalated	−1.5±1.0	0.9±1.2	10.0±0.7	−8.9±5.5	6.6±8.3	50.4±7.7	11.5±1.7
CXG	unmodified	0.0±0.8	−1.6±1.0	6.7±0.4	0.1±5.2	11.3±8.4	62.9±6.3	7.7±1.4
	anti base-displaced	−1.6±0.7	1.0±1.0	6.8±0.3	−1.5±5.0	−3.8±5.9	56.5±5.1	7.4±1.4
	anti 5′–intercalated	−2.1±1.0	0.5±1.5	9.4±0.5	−17.0±6.9	−2.0±9.6	38.7±14.6	10.2±3.1

	anti 3′–intercalated	–0.9±1.5	1.9±1.4	9.1±0.8	–6.2±7.9	–3.2±13.3	48.9±12.1	12.9±2.8
	syn base-displaced	–1.9±0.9	0.6±1.1	6.8±0.4	1.2±5.3	6.4±6.9	48.8±5.9	8.7±1.7
	syn 5′–intercalated	–2.2±1.1	–0.2±1.0	9.8±0.6	–16.6±6.9	–1.3±9.0	50.5±8.0	7.9±1.7
	syn 3′–intercalated	–1.6±1.2	1.4±1.5	10.0±0.7	–9.5±6.0	6.3±9.5	47.2±8.3	11.2±2.0
GXC								
	unmodified	0.2±0.7	–1.4±1.1	6.6±0.4	–0.9±5.1	3.1±7.0	66.9±6.1	6.9±1.5
	anti base-displaced	–1.4±0.7	0.9±0.9	6.8±0.3	–0.5±4.9	–5.1±5.6	57.8±4.7	7.8±1.3
	anti 5′–intercalated	–1.5±0.9	1.0±1.0	9.0±0.5	–14.5±6.3	–1.5±8.4	51.5±7.0	14.5±3.8
	anti 3′–intercalated	–1.6±1.2	1.3±1.2	10.0±0.5	–9.9±6.4	–2.6±9.4	48.3±9.4	10.1±1.8
	syn base-displaced	–1.6±0.8	–0.2±1.2	6.8±0.3	0.8±4.9	5.9±6.3	54.4±6.0	8.4±1.5
	syn 5′–intercalated	–1.5±1.0	–0.9±1.2	8.7±0.5	–7.7±6.5	7.4±10.6	45.3±10.3	10.1±2.3
	syn 3′–intercalated	–1.7±1.0	2.1±0.9	10.3±0.5	–9.3±5.0	1.7±8.8	48.6±8.0	10.4±1.8

CXC								
	unmodified	0.3±0.8	–1.3±1.0	6.5±0.4	1.9±5.2	8.6±7.8	63.1±6.0	7.2±1.5
	anti base-displaced	–1.6±0.7	0.8±0.8	6.9±0.3	–2.1±5.0	–5.2±5.9	56.8±4.8	7.6±1.2
	anti 5′–intercalated	–2.2±1.1	0.5±1.0	9.0±0.7	–14.9±6.7	–0.7±9.0	45.9±9.3	9.9±2.0
	anti 3′–intercalated	–0.5±1.2	1.4±1.2	9.1±1.0	–7.2±7.0	–1.5±9.1	54.5±9.2	11.5±2.7
	syn base-displaced	–1.9±0.9	0.0±1.2	6.9±0.4	0.7±5.1	7.7±8.1	51.2±6.1	7.7±1.7
	syn 5′–intercalated	–1.0±1.5	–1.0±1.7	8.9±0.7	–10.1±8.4	5.1±10.4	53.0±8.4	11.6±1.7
	syn 3′–intercalated	–1.7±1.1	1.5±1.4	10.1±0.6	–11.1±5.4	7.5±9.5	47.3±8.3	9.5±2.0

[a]Calculated using the 3′ and 5′–flanking base pairs. [c]Calculated as the distance between the P atom of the 7th and 20th residue (Figure 2.1b) minus 5.8 Å to account for the van der Waals radii of the two phosphate groups.

Table C5: Interaction energies between the adduct and the opposing thymine for the base-displaced intercalated conformations of ALI-N^6-dA adducted DNA in different sequence contexts derived from 20 ns MD simulations (kJ mol^{-1}).

Sequence Context	Conformation	$E_{vdW} + E_{elec}$
GXG	*anti* base-displaced	–18.9±3.8
	syn base-displaced	–10.5±4.4
CXG	*anti* base-displaced	–18.1±4.2
	syn base-displaced	–8.4±4.6
GXC	*anti* base-displaced	–17.6±4.2
	syn base-displaced	–10.1±3.3
CXC	*anti* base-displaced	–15.5±5.0
	syn base-displaced	–8.0±4.6