Synthesis and Polymerase-Mediated Bypass Studies of the N²-Deoxyguanosine DNA Damage Caused by a Lucidin Analogue

Pratibha P. Ghodke, S. Harikrishna, and P. I. Pradeepkumar*

Department of Chemistry, Indian Institute of Technology Bombay, Mumbai 400076, India

Supporting Information

ABSTRACT: Lucidin is a genotoxic and mutagenic hydroxyanthraquinone metabolite, which originates from the roots of Rubia tinctorum L. (madder root). It reacts with exocyclic amino groups of DNA nucleobases and forms adducts/lesions leading to carcinogenesis. To study the effect of lucidin-induced DNA damage, herein, we report the first synthesis of a structural analogue of lucidin [N²-methyl-(1,3-dimethoxyanthraquinone)-deoxyguanosine, LdG] embedded DNAs utilizing phosphoramidite strategy. LdG modification in a DNA duplex imparts destabilization (ΔTm ~5°C/modification), which is attributed to the unfavorable contribution from the enthalpy. Primer extension studies using the Klenow fragment (exo−) of Escherichia coli DNA polymerase I demonstrate that bypass of LdG modification is error prone as well as slow compared to that across the unmodified sites. Molecular dynamics simulations of the binary complex of Bacillus fragment polymerase (homologue of the Klenow fragment) and LdG-DNA duplex elucidate the structural fluctuations imparted by the LdG lesion, as well as the molecular mechanism of bypass at the lesion site. Overall, the results presented here show that the lucidin adduct destabilizes DNA structure and reduces fidelity and processivity of DNA synthesis.

INTRODUCTION

The integrity of the genomic DNA is disturbed by continuous exposure to exogenous and endogenous agents, which leads to DNA damage. This is a major threat, and the first significant step of chemical carcinogenesis.¹ In general, the DNA damages are recognized by DNA polymerases during replication and are repaired by various DNA repair mechanisms.² If DNA damage sites are unrepaired, this can result in blockage, slippage, misinsertion, or mutations in daughter cells, leading to cancer.³ Therefore, it is essential to understand the interaction of polymerases with damaged DNAs at the molecular level. Each DNA polymerase has a unique role and shows varying fidelity toward individual DNA damage.⁴ Most of the replicative DNA polymerases are not able to tolerate the damages formed in the genomic DNA. However, DNA polymerases belonging to the Y-family can continue the DNA replication in the presence of damages employing a mechanism known as translesion synthesis (TLS) or bypass mechanism.⁵ These Y-family polymerases are devoid of intrinsic proofreading exonuclease activities, which are the characteristics of replicative polymerases.⁵,⁶

The DNA damages can be caused by chemical carcinogens and mutagens, which are electrophilic in nature, or they get converted to electrophiles through various metabolic pathways.⁷ These species react with nucleophilic sites of DNA, thereby altering the structural integrity by forming covalent bonds with the DNA bases known as adducts or lesions.⁸ Various reactive metabolites of carcinogens show different reactivity toward the nucleophilic centers of DNA bases.⁹ Among the various nucleophilic centers of DNA, the N² position of the guanine is vulnerable toward the attack by a wide variety of electrophiles.⁹ Importantly, the N²-dG DNA adducts are located in the minor groove of DNA, and depending upon their chemical and structural attributes, they have potential to disrupt standard Watson−Crick (W−C) base pairing, which lead to the stalling of the replication process.¹⁰,¹¹ For example, polycyclic aromatic hydrocarbons (PAH) such as the benzo(a)pyrene (BP) get converted to carcinogenic diol-epoxides (BPDE) and its o-quinone derivatives via different metabolic pathways to form N²-dG and N⁶-dA adducts.¹² Similarly, the number of other PAHs like the chrysene, benz(a)anthracene, aminofluorene (AF) etc., get activated enzymatically to their reactive metabolites, which can form N²-dG adducts.¹³,¹⁴

Received: November 18, 2014
The red colored dye extract from roots of the *Rubia tinctorum* L. (madder root) has been widely used as a dying agent, as well as an additive in a variety of food and drinks.\(^{15}\) The madder root contains various carcinogenic hydroxyanthraquinones and their glycoside conjugates like lucidin-O-primeveroside (LuP, Figure 1) and alizarinprimeveroside (AIP).\(^{16}\) The LuP and AIP glycosides are hydrolyzed into different derivatives of hydroxyanthraquinones named lucidin and alizarin, respectively.\(^{17}\) Lucidin is a potent carcinogen, and it is capable of forming N\(^2\)-dG and N\(^6\)-dA DNA adducts utilizing a possible sulfotransferase metabolic pathway or by a dehydration process (Figure 1).\(^{18,19}\) Lucidin is found to be mutagenic and genotoxic to bacteria and mammalian cells and forms DNA adducts in mice.\(^{20}\) Specific DNA adduct formation by lucidin under *in vivo* conditions has been studied by measuring the adduct levels in kidneys and livers of rats.\(^{21}\) To study the lucidin-induced chemical carcinogenesis and the effect of this adduct on DNA replication, it is necessary to develop robust chemical methods to synthesize the lucidin-modified DNA. To this end, herein, we report the synthesis of a structural analogue of *N*\(^2\)-Luc-dG-modified DNA using phosphoramidite chemistry (LdG, Figure 1).

**RESULTS AND DISCUSSION**

**Synthesis of the LdG-Modified Phosphoramidite and DNAs.** Our initial goal was to synthesize the *N*\(^2\)-Luc-dG (Figure 1)-modified phosphoramidite for its incorporation into DNA. Toward this end, it was planned to achieve the synthesis of 2-(aminomethyl)-1,3-diacetoxyanthraquinone, which is an acetyl-protected lucidin amine (Figure S1 of the Supporting Information) required to assemble the *N*\(^2\)-dG nucleoside via Buchwald-Hartwig coupling. This strategy was envisaged because the acetate group can easily be deprotected at the oligonucleotide level to get the lucidin-modified DNAs. But, all our efforts to synthesize the respective amine were not successful, perhaps due to steric hindrance encountered by one of the hydroxy group of the lucidin during acylation (Figure S1 of the Supporting Information).\(^{22}\) To circumvent this, a methyl-protected structural analogue of lucidin (LdG, Figure 1) was

![Figure 1. Proposed metabolic pathways for the formation of the *N*\(^2\)-Luc-dG adduct, and the structure of its methyl protected analogue (LdG).](image)

**Scheme 1. Synthesis of 2-(Aminomethyl)-1,3-dimethoxyanthraquinone 3**

\(\text{Reagents and conditions: (i) Sodium azide, DMF, rt, 30 h and (ii) PPh}_3\cdot\text{THF:H}_2\text{O, rt, 2 h.}\)
Scheme 2. Synthesis of the LdG Phosphoramidite 9×

Table 1. \( T_m \) and the Thermodynamic Parameters of the Unmodified and LdG Modified DNA Duplexes×

<table>
<thead>
<tr>
<th>code</th>
<th>duplex</th>
<th>( T_m ) (^{b,c} ) (°C)</th>
<th>( \Delta T_m ) (^{b,c} ) (°C)</th>
<th>( \Delta H^\circ ) ( (\Delta H^\circ)^c ) (kcal/mol)</th>
<th>( \Delta S^\circ ) ( (\Delta T\Delta S^\circ)^c ) (kcal/mol)</th>
<th>( \Delta G^\circ ) ( (\Delta G^\circ)^c ) (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1-D2</td>
<td>5′-GCCGAATAGCGCA-3′</td>
<td>64.5±0.50</td>
<td>-</td>
<td>-66.0±0.59</td>
<td>-195±1.4</td>
<td>-58.2</td>
</tr>
<tr>
<td></td>
<td>3′-CGGCCCTTATCGGCT-5′</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3-D2</td>
<td>5′-GCCGG*AATAGCGCA-3′</td>
<td>59.1±0.60</td>
<td>-5.4</td>
<td>-61.7±0.45</td>
<td>-185±0.7</td>
<td>-55.4</td>
</tr>
<tr>
<td></td>
<td>3′-CGGCCCTTATCGGCT-5′</td>
<td></td>
<td></td>
<td>(+4.3)</td>
<td></td>
<td>(+1.4)</td>
</tr>
<tr>
<td>D3-D4</td>
<td>5′-GCCGG*AATAGCGCA-3′</td>
<td>54.4±0.47</td>
<td>-4.7</td>
<td>-44.3±0.63</td>
<td>-135±1.8</td>
<td>-40.3</td>
</tr>
<tr>
<td></td>
<td>3′-CGGCCCTTATCGGCT-5′</td>
<td></td>
<td></td>
<td>(+21.7)</td>
<td></td>
<td>(+17.9)</td>
</tr>
</tbody>
</table>

The thermal denaturation was performed at 25 °C (298 K) using DNA duplex (5 μM) in phosphate buffer (100 mM NaCl, 20 mM sodium phosphate, 0.1 mM EDTA, pH 7.4). The LdG \( (G^\circ) \) modified nucleoside is highlighted as starred and shown in bold. \( b \)\( \Delta T_m \) represents the \[ T_m^{(\text{DNA}_{\text{unmod}})} - T_m^{(\text{DNA}_{\text{mod}})} \]. The \( T_m \) values reported are the average of 3 independent measurements with the estimated standard deviation. \( \Delta H^\circ, \Delta S^\circ, \text{and} \Delta G^\circ \) are average values emerged from 3 independent melting curve analysis using a two state model. The error propagation for \( \Delta H^\circ, \Delta S^\circ \), \( \text{and} \Delta G^\circ \) were calculated using a reported protocol.\(^{27} \) \( \Delta \Delta H^\circ, \Delta T \Delta S^\circ, \text{and} \Delta \Delta G^\circ \) represent the differences in the thermodynamic parameters of modified duplex with respect to those of the unmodified duplex.

The protected bromo nucleoside 5 required for the Buchwald-Hartwig coupling was synthesized utilizing 2-(aminomethyl)-1,3-dimethoxyanthraquinone 3, which was deprotected from a bromo precursor 1 as outlined in Scheme 1.\(^{23} \) A bromo group of 1 was displaced to form the azide 2 using sodium azide in DMF in a 46% yield. Staudinger reduction of azide compound 2 to the corresponding amine 3 was accomplished in a 78% yield using PPh₃ in the THF–water solvent system.

The protected bromo nucleoside 5 required for the Buchwald-Hartwig coupling was synthesized utilizing \( O^\circ \)-para-nitrophenylethyl (NPE)-protected deoxynosine 4 (Scheme 2).\(^{23} \) Nucleoside 4 was subjected to diazotization followed by bromination using SbBr₃ and t-butyl nitrite (TBN) in dibromomethane (DBM) to furnish bromo nucleoside 5 in 82% yield.\(^{24} \) Coupling between 5 and dimethoxyanthraquinone 3, in the presence of Pd(OAc)\(_{2}\) (R)-BINAP, and Cs\(_2\)CO\(_3\) in toluene afforded compound 6.\(^{25} \) The deacetylation of crude compound 6 was achieved using MeNH\(_2\) in ethanol to obtain the diol nucleoside 7 in 39% yield (after two steps from 5). The 5′-OH of 7 was protected using DMT-Cl in pyridine to afford tritylated nucleoside 8 in 93% yield. Nucleoside 8 was phosphorylated using 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (CEP-Cl) and N,N-diisopropylethylamine (DIPEA) in DCM to obtain the LdG phosphoramidite 9 in 66% yield.\(^{26} \) The \( O^\circ \)-NPE group is retained in the phosphoramidite to enhance the solubility during the solid phase synthesis.

The LdG-modified phosphoramidite building block 9 was successfully incorporated into the DNA sequences (Table S6 of the Supporting Information) employing an automated DNA synthesizer. The sequences D3 and D4 are the LdG-modified 14-mer DNAs, which were synthesized to evaluate the thermal stability of the modified duplexes. The sequence T is a LdG-
modified 50-mer DNA template utilized for the primer extension studies.

**Thermal Melting Studies.** To deduce the effect of the LdG modification on the DNA duplex stability, UV melting studies were carried out (Table 1 and Figure S2 of the Supporting Information). In comparison to the unmodified duplex (D1-D2), there was a significant decrease in $T_m$ ($\Delta T_m \sim -5$ °C/LdG unit) observed for LdG-modified duplexes (D3-D2 and D3-D4). To understand the factors responsible for these effects, all the thermodynamic parameters ($\Delta H^0$, $\Delta S^0$, and $\Delta G^0$ at 298 K) were calculated from the absorbance versus temperature melting curves, by fitting them into a two-state model (Figure S2 of the Supporting Information).27 The decrease in $T_m$ due to the presence of LdG modification was reflected in the +1.4 to 3.8 kcal/mol drop in $\Delta G^0$ values. Thermodynamic parameters show that the LdG provides an entropic advantage ($\Delta T \Delta S^0 = +2.8$ to 17.9 kcal/mol) to the duplex, while it had an adverse effect on enthalpy ($\Delta \Delta H = +4.3$ to 21.7 kcal/mol). Therefore, the net destabilization of the duplex can be attributed to the compensation of favorable contribution from entropy by the unfavorable contribution of enthalpy. Our molecular modeling and dynamics studies suggest that the LdG adduct imparts conformational flexibility due to which the base stacking, base pairing, and hydration were disturbed at the modified site (see the respective section for details). The observed thermodynamic parameters validate these findings.

**Primer Extension Studies.** To evaluate the bypass and the extension ability of the LdG-modified DNA by the KF−, single nucleotide incorporation as well as full length extension experiments were carried out. It has been reported that the KF− is able to bypass various N2-dG adducts, and it lacks proofreading activity, which results in the error prone replication.10,28,29 Single nucleotide incorporation experiments were performed using the dCTP and LdG-modified DNA template by varying concentration of enzyme to study the correct base incorporation opposite to the LdG adduct (Figure 2, panels B–D). The primer–template system used in these experiments is shown in Figure 2A. In the reactions employing 0.05 units of KF− with the unmodified template, both the incorporated (+1, 16nt) and extended (+2, 17nt) products were observed (Figure 2B). After 4 h, 10% of incorporation and 89% of extension were visible (lane 7, Figure 2B). These results validate the highly efficient incorporation as well as extension across the unmodified dG, even at a relatively low concentration of KF−. However, since the +2 site of the template contains dT, the extension of primer in the presence of dCTP suggests low fidelity of the KF− during the extension process. However, under similar reaction conditions and time, the LdG-modified template showed only 9% incorporation and 5% extension of the primer (lane 7, Figure 2C). This indicates that neither significant incorporation nor extension was achieved by the polymerase across LdG when compared to dG. It should be noted that by increasing the enzyme concentration to 2 units, 52% of incorporation and 14% of extension was observed across LdG after 4 h (lane 7, Figure 2D). These results clearly show that compared to correct base incorporation across LdG, furthering the extension is significantly slower. This may be attributed to the structural distortions at the modified site as well as the steric clashes caused by the bulky adduct with KF−.30 These findings are supported by our molecular modeling studies (see the respective section for details).

To study the miscoding potential opposite LdG, single nucleotide misincorporation experiments were performed using 2 units of KF− in the presence of dATP, dGTP, and dTTP (Figure 3, panels A–C). In the case of dATP, 84% of +2 and 16% of +3 extended products were observed after 4 h (lane 7, Figure 3A). This shows that dATP incorporation was quite rapid, and the incorporated product immediately gets extended due to the presence of dT at a +2 site. The dGTP was incorporated up to 60% along with the formation of 25% of +2 and 7% of +3 extended products (lane 7, Figure 3B). Similarly, dTTP...
incorporation across LdG was found to be 67% along with the formation of 8% of a +2 extended product (lane 7, Figure 3C). Overall, these results show that the efficiency of KF− for dATP misincorporation is higher than other dNTPs, and the rate of incorporation was found to be in the following order: dATP > dGTP > dTTP > dCTP. These results are in line with the reported findings for the widely studied BP adduct and others, where the KF− showed preference for dATP incorporation over other nucleotides.31–35 The reason for the better tolerance of dA across the adduct site has been attributed to the efficient stacking of purine bases with the polycyclic aromatic DNA adducts. This facilitates the formation of right conformation required for the base incorporation.32

To find the ability of KF− to polymerize the full length product, primer extensions of the LdG-modified and the unmodified DNA templates were carried out in the presence of a mixture of all four dNTPs (Figure 4, panels A–C). With the use of 0.05 units of the KF−, the unmodified template yielded 86% of the full length product after 4 h (lane 7, Figure 4A). However, under the identical conditions and time, the LdG-modified template showed only 2% of the full length product (lane 7, Figure 4B). Nevertheless, as expected, at a higher unit of enzyme (2 U), 84% of the fully extended primer was observed in the presence of the LdG-modified template (lane 7, Figure 4C). This observation is in contrast with the behavior of N2-dG adducts formed by BP and AF toward KF−. In the case of BP, the enzyme is unable to extend the primer after the adduct site, while for AF the extension beyond the +3 site is strongly blocked.35,36 To investigate the efficiency of full length extension reactions, additional experiments were carried out using the 11-mer primer and the LdG-modified template, in which the modification is five nucleotides away from the initiating point (Figure S3A of the Supporting Information). When 0.05 units of the KF− was used, after 4 h, the unmodified template yielded 96% of the full length product (lane 7, Figure S3B of the Supporting Information). While, under similar conditions, there was no significant product formation with the LdG-modified template, which clearly shows that DNA synthesis has been halted at the modification site (lane 7, Figure S3C of the Supporting Information). However, using a higher unit of enzyme (2 U KF−), the 92% of the fully extended product was observed (lane 7, Figure S3D, of the Supporting Information). These results clearly demonstrate that the incorporation and extension efficiency of the LdG-modified templates by the KF− is independent of the position of the lesion site. Overall, the primer extension studies reveal that KF− is able to perform TLS across LdG DNA adduct albeit at low fidelity and processivity.

**Molecular Modeling Studies.** Our first goal was to understand the possible structural effects of the LdG modification inside a DNA duplex. Toward this end, the energy optimized and RESP charge-fitted LdG-modified nucleotide using Gaussian 0937 (Figure S4 of the Supporting Information) was placed in a DNA duplex which is used for the thermal melting studies (Table 1, D3-D2).33,34 MD simulations of the LdG-modified DNA duplex was carried out for 50 ns using AMBER 12.40,41 The system reached the equilibrium after 30 ns as deduced from the RMSD graph (Figure S5 of the Supporting Information); thus, all the analyses were carried out from the 30–50 ns of MD trajectories. The averaged structure from the last 20 ns of the MD simulations is shown in Figure 5A, where the LdG adduct is positioned toward the 3’-end in the minor groove of the duplex DNA. The MD structure shows that the B-form of the LdG-modified DNA is maintained throughout the simulation. This is supported by CD spectra of the LdG-modified DNA duplex, where a positive peak is observed at 280 nm and a negative peak at 250 nm (Figure S6 of the Supporting Information).42

The distance between the 3’-dA nucleotide and the LdG adduct was found to be ∼4.2 ± 0.5 Å (Figure 5B). It should be noted that during the course of the simulations, the two consequent dA nucleotides at the 3’-end (Figure S, panels B and D) moved away from the original position to avoid the steric clashes with the LdG adduct. This movement could disturb the W−C hydrogen bonds between the pairing bases. This is reflected in the H-bond occupancy, where only one W−C H-bond between the LdG and dC was found to be stable and the other two H-bonds were absent (Figure 5C and Table S1 of the Supporting Information). Moreover, the base pairing of the neighboring nucleotides around the LdG adduct was also disturbed (Table S1 of the Supporting Information). The percentage of stacking interactions between the LdG and dC with the neighboring dA and dT were found to be ∼13% of the total simulation time (Figure 5D). Moreover, the hydration around the LdG adduct was severely affected both in the minor and major grooves of the DNA duplex (Table S2 of the Supporting Information). The average of the six sugar−phosphate backbone dihedral angles (α, β, γ, ε, δ, and ζ) were found to be −98.3°± 12, −178.5°± 14, 61.1°± 22, 71.1°± 18, −163.2°± 31, and −73.1°± 36, respectively, with maximum deviation of up to 32° with respect to those of the canonical B-form DNA.43 In particular, higher deviations were observed for α, δ, and ζ torsions. The γ angle of the glycosidic bond of LdG base was found to be −164.2°± 16, which is similar to the canonical B-DNA value (−150° to −180°).43 Overall, the MD simulations of the LdG-modified duplex suggest that due to the position of the bulky adduct in the minor groove, the stacking interactions,
W–C H-bond between bases, and the groove hydration were affected in and around the modified site.

To delineate the possible molecular mechanism of bypass of the LdG adduct by the KF−, a binary complex composed of a DNA duplex containing the LdG in the template and Bacillus fragment (BF; PDB entry: 1XC9) was investigated using MD simulations. The rationale to use BF structure is that, it is a homologue of KF−; additionally, the structure contains the BPdG adduct in the template DNA. Initially, the optimized and RESP charge-fitted LdG-modified nucleotide (Figure S4 of the Supporting Information) was placed at the post insertion site (Figure 6A), where the BPdG adduct is present in the BF complex. The modified binary complex was solvated, minimized, and then the unrestrained MD simulations were carried out for 40 ns using AMBER 12.

The final MD snapshot of the binary complex is shown in Figure 6B. The structure shows that the B-form of the template-primer duplex was maintained throughout the simulations. The LdG adduct was positioned in the minor groove of the DNA, in which the adduct makes van der Waals contact with the bases at the 5′-end of the template DNA. The orientation of the LdG adduct in the complex was found to be similar to that of the BPdG adduct. Here, it should be noted that the adducts placed in the major groove are well-tolerated by the polymerases in comparison to the adduct placed in the minor groove of the DNA, which makes key contacts with the polymerase active site.

In the binary complex, the LdG adduct was found to adopt a number of conformations (Figure 6C). This is reflected in the average RMSD (2.5 ± 1.4 Å) of DNA during the course of the simulations (Figure S7 of the Supporting Information). Moreover, the LdG adopts anti (ψ = −124°) conformation with higher deviation of up to 20° from the canonical conformation; and the C2′-endo conformation was found to be the most prominent sugar pucker. The RMSD analysis revealed that there are two major conformational ensembles for the LdG–DNA duplex in the binary complex: (i) the perpendicularly positioned adduct in the major groove are well-tolerated by the polymerases in comparison to the adduct placed in the minor groove of the DNA, which makes key contacts with the polymerase active site.

In Figure 6C, the averaged conformations from the ensemble analysis are shown. The conformer having widened groove size for the duplex and having weak T-shaped van der Waals contact between LdG with the dC and dA bases at the 5′-end of the template strand (Figure 6C and Figure 5).
S8 of the Supporting Information). The representative structures of the two conformational ensembles are shown in Figure 6C. These two major conformational ensembles were preferred for >82% of the total simulation time. The averaged structure of these two conformers having one W−C H-bond between the O6 of LdG and the hydrogen of N3 of dC is shown in Figure 6D. The other two W−C H-bonding sites of dC were stabilized by Tyr 714 of BF polymerase (Figure 7).

Figure 7. Active site orientation of the LdG adduct and pairing base dC in complex with the Bacillus fragment (BF). Polymerase is colored by domains (thumb in green, and finger in orange). The LdG adduct is shown in magenta, the hydrogen bonds between the atoms are represented in black dotted lines. The protein is rendered in cartoon, the atoms shown as sticks, and DNA as sticks along with filled rings. Key residues are labeled with numbers. Tyr 714 makes an H-bond with dC at a distance of 2.75 Å. Arg 629 and Arg 789 form an H-bond with LdG and dC, respectively, at a distance of ~2.73 Å. Ser 671 forms an H-bond with the carbonyl oxygen atom in the LdG at a distance of 2.63 Å. His 829 stacks on the polycyclic aromatic ring of LdG with a 3.0 Å distance. The W−C H-bond distance between LdG and dC is 3.1 Å. The conformer shown here is the averaged structure depicted in Figure 6C.

In the binary complex, the LdG appeared to be intact and its motions were stabilized by two different loops in the active site of the BF. Loops from the amino acids 607−618 and 828−831 constitute 16 residues, which form various noncovalent interactions with the adduct and they are highly flexible during the simulations as evident from their RMSD values (~5.2 ± 2.3 Å). Notably, the amino acid His-829 from the loop 2 stacks on the LdG adduct, which partly restricts the movement of the polycyclic aromatic ring during the course of simulation (Figure 7). In addition, Arg residues at 629 and 789 form hydrogen bonds with the phosphate backbone of both template and primer at the adducted site. Also, the oxygen atom in the carbonyl group of the LdG adduct forms a hydrogen bond with the Ser-617 hydroxyl group. The Arg-615 and Gln-797 in BF polymerase are known to make H-bond contact with the benzo[a]pyrene adduct at the post insertion site.34 To probe the interaction of these two amino acids with the LdG adduct, the distance between them has been computed. Those distances were found to be 3.5 to 4.0 Å (Figure 7), which corresponds to only weak van der Waals interactions. It has been reported that the efficiency of dNTP incorporation by KF is significantly lowered when the C1(dG)′−C1′(dC) distance is increased >12.5 Å.46,47 In the case of the LdG, the average C1′−C1′ distance was found to be 11.9 ± 2.9 Å. Due to this, the KF is not able to well-accommodate the LdG adduct, which is indeed reflected in the primer extension experiments.

To find the structural flexibility in BF induced by the LdG adduct, the BF crystal structure in complex with unmodified DNA (PDB entry: 1L3S) was superimposed on the averaged MD structure of BF in complex with LdG-modified DNA (Figure 8, and Figure S9 of the Supporting Information). It is clear from this analysis that the loops 1 and 2 of BF move away from its original position by ~4.5 Å when complexed with the LdG adduct. Thus, the polymerase requires the larger conformational rearrangements to accommodate the LdG. It results in the structural distortions, which create a high-energy barrier to bypass the adduct site. Additionally, to find the distortions in the DNA structure during the course of MD simulations, the backbone torsions and helical parameters of the LdG-modified duplex were calculated and the values were compared with those from the unmodified duplex (PDB entry: 1L3S). The results show that the base pair helical parameters like buckle, propeller, twist, and tilt are significantly deviated around the LdG adduct site (Tables S3−S4 of the Supporting Information). Backbone torsion angles such as α, β, γ, and δ were similar to those of the canonical values; however, ϵ and ζ were highly deviated (Table S5 of the Supporting Information).

Overall, the results from the molecular modeling and dynamics studies suggest that the LdG occupies the minor groove of the DNA duplex and imparts conformational flexibility due to which lengthening of the C1′ (LdG)−C1′ (dC) distance was observed at the template−primer interface. Therefore, to facilitate the primer extension, the polymerase undergoes structural rearrangements; also the H-bonding and van der Waals interactions with the amino acid residues at the active site stabilize the LdG:dC base pair.

■ CONCLUSIONS

In summary, we have achieved the synthesis of the LdG-modified DNAs. Our synthetic strategy involves Buchwald-Hartwig coupling as a key step to access the LdG phosphoramidite building block, which is utilized for the solid phase DNA synthesis. Thermal melting studies revealed that the LdG modification decreases the stability of the DNA duplex by ~5 °C. The disturbances in W−C H-bonding and base stacking lead to the destabilization of LdG-modified DNA duplexes, which were reflected in the unfavorable contribution of enthalpy. Primer extension studies showed that the KF was able to bypass the LdG in an error prone fashion. Adduct bypass by the KF has shown more preference for the incorporation of purines in
comparison to pyrimidines. The molecular modeling and MD simulation studies revealed the atomic interactions of the enzyme and LdG, which helped to unravel the possible molecular mechanism for the accommodation and nucleotide bypass across the adduct. It was evident from these studies that the increase in the 1H−C1’ distance and loss of two W−C H-bonding between LdG and dC, disturbances in the enzyme active site orientations, together may be responsible for the inefficient incorporation of the incoming nucleotide by KF. Overall, tools and insights emerging from these studies would pave the way to investigate the structural and functional requirements of possible TLS pathways across lucidin-modified DNAs by both the eukaryotic and prokaryotic Y-family polymerases.

**EXPERIMENTAL SECTION**

**General.** All the required chemicals and solvents were obtained from the commercial sources. Dioxane, pyridine, acetonitrile, DIPEA, DCM, and Et3N were dried using calcium hydride; and the toluene was dried using CaH2. Thin layer chromatography (TLC) was carried out on silica gel plates, which were precoated with the fluorescent indicator, visualized by UV light or by dipping into a solution of 5% (v/v) conc. H2SO4 in ethanol, and heating. Silica gel of 100–200 mesh size was used for column chromatography. 1H NMR (400 and 500 MHz), 13C NMR (100 MHz), and 31P NMR (162 MHz) were recorded on a 400 or 500 MHz NMR instrument. The chemical shifts are reported in parts per million (ppm) downfield from TMS and referenced to the TMS signal or residual proton signal of the deuterated solvent as follows: TMS (0 ppm) or CD3OD (3.31 ppm) for the 1H NMR spectra, and CDCl3 (77.2 ppm) or CD2OD (49.1 ppm) for the 13C NMR spectra. Multiplicities of 1H NMR spin couplings are reported as s for singlet, bs for broad singlet, d for doublet, dd for doublet of triplets, dd for doublet of doublets, dddd for doublet of doublet of doublets, or m for multiplet and overlapping spin systems. Values for the apparent coupling constants (J) are reported in Hz. High-resolution mass spectra (HRMS) were obtained in a positive ion electrospray ionization (ESI) mode using the Q-TOF analyzer. DNA sequences were synthesized using an automated DNA synthesizer. Mass spectra of the oligonucleotides were obtained by positive reflection mode in a MALDI-TOF spectrometer.

2-(Azidomethyl)-1,3-dimethoxyanthraquinone (2). 2-(Bromomethyl)-1,3-dimethoxyanthraquinone (987 mg, 2.73 mmol) was dissolved in sodium azide (284 mg, 4.37 mmol) in 13 mL of dry DMF and stirred for 60 h at room temperature. After completion of the reaction, product was extracted using ethyl acetate (2×30 mL), dried over Na2SO4, and filtered, and concentrated under reduced pressure. The crude compound 2 as a yellow solid (413 mg, 46%); Rf = 0.57 (5% ethyl acetate in petroleum ether); mp = 181−183 °C; 1H NMR (400 MHz, CDCl3): δ 8.26 (d, J = 7.9 Hz, 1H), 8.23 (d, J = 1.3, 7.5 Hz, 1H), 7.82−7.71 (m, 2H), 7.70 (s, 1H), 4.52 (s, 2H), 4.07 (s, 3H), 3.98 (s, 3H); 13C NMR (100 MHz, CDCl3): δ 183.1, 181.1, 163.1, 131.7, 134.9, 134.6, 133.5, 132.5, 127.4, 126.9, 125.2, 120.1, 105.4, 63.3, 56.6, 43.2; HRMS (ESI): calcd for C12H9BrN3O5Na [M + Na+] : 436.0878, found: [M + Na+] 436.0801 (Δm = +0.0007, error = +0.08 ppm).

2-(Azidomethyl)-1,3-dimethoxyanthraquinone (3). 2-(Azidomethyl)-1,3-dimethoxyanthraquinone (413 mg, 1.27 mmol) was dissolved in 3 mL of THF, followed by the addition of triphenylphosphine (502 mg, 1.91 mmol) at 0 °C. The reaction was removed from the ice bath, and water (0.06 mL) was added. The reaction mixture was stirred at room temperature for 3 h. After the completion of the reaction, the solvent was evaporated under reduced pressure. The crude compound was purified by column chromatography (5% ethyl acetate in petroleum ether) to give compound 3 as a pale yellow solid (98 mg, 82%); Rf = 0.55 (70% ethyl acetate in petroleum ether); mp: 127−129 °C; 1H NMR (400 MHz, CDCl3): δ 8.26 (d, J = 7.6 Hz, 1H), 8.22 (d, J = 7.3 Hz, 1H), 8.15 (d, J = 7.8 Hz, 2H), 7.82−7.71 (m, 2H), 7.76 (s, 1H), 7.61 (s, 1H), 7.51 (d, J = 7.9 Hz, 2H), 6.25−6.17 (m, 1H), 5.55 (bs, 1H), 4.93−4.71 (m, 3H), 4.15 (s, 1H), 4.05 (s, 3H), 3.92 (bs, 1H), 3.89 (s, 3H), 3.75 (d, J = 12.5 Hz, 1H), 3.32−3.27 (m, 2H), 3.04 (bs, 1H), 2.26 (d, J = 8.2 Hz, 1H); 13C NMR (100 MHz, CDCl3): δ 183, 181, 163.1, 161, 160.9, 158.3, 152.9, 146.9, 146.1, 139, 134.8, 134.6, 133.5, 132.4, 130.1, 130.8, 128.2, 127.3, 126.8, 123.8, 120.8, 116.5, 105.1, 89.7, 87.1, 73.2, 66.5, 62.8, 56.6, 40.2, 35.3, 35.3; HRMS (ESI): calcd for C12H9BrN3O5Na [M + Na+] : 719.2072; found: [M + Na+] 719.2079 (Δm = +0.0007 and error = +0.09 ppm).

5′-O-(4′,4′-dimethoxytrityl)methyl-(1,3-dimethoxyanthraquinone)-O2′-(2-(4-nitrophenyl)ethyl)-2′-deoxyoxynosine (5). The amine compound 4 (108 mg, 0.21 mmol) was dried by coevaporation using anhydrous toluene (12 mL) under reduced pressure. Antimony bromide (111 mg, 0.30 mmol) was added to the flask containing dried amine, and the system was purged with N2 gas. Anhydrous CH2Br2 (2.5 mL) was cooled to 0 °C and added to the reaction mixture with continuous stirring. Then the reaction mixture was kept stirring for around 2 h at the same temperature followed by stirring at 0 °C for an additional 2 h. After the completion of the reaction, the reaction mixture was poured into the saturated solution of NaHCO3 (50 mL) in crushed ice and extracted with DCM (3 × 30 mL). The combined organic layers are dried over Na2SO4, filtered, and concentrated under reduced pressure. The crude compound was purified by column chromatography (55% ethyl acetate in petroleum ether) to give compound 5 as a yellow solid (98 mg, 82%); Rf = 0.46 (70% ethyl acetate in petroleum ether); mp: 181−183 °C; 1H NMR (400 MHz, CDCl3): δ 8.21−8.13 (m, 2H), 8.08 (s, 1H), 7.49 (d, J = 8.5 Hz, 2H), 6.43 (dd, J = 6.0, 8.0 Hz, 1H), 5.39 (td, J = 2.4, 6.0 Hz, 1H), 4.84 (t, J = 6.7 Hz, 2H), 4.40−4.31 (m, 3H), 3.11 (t, J = 6.7 Hz, 2H), 2.81 (dd, J = 6.4, 8.0, 14.2 Hz, 1H), 2.65 (dd, J = 2.5, 5.9, 14.2 Hz, 1H), 2.14 (s, 3H), 2.10 (s, 3H); 13C NMR (100 MHz, CDCl3): δ 170.5, 170.4, 163.0, 152.8, 145.7, 144.5, 143.6, 140.7, 130.1, 123.9, 121.3, 84.8, 72.8, 74.7, 67.7, 63.7, 38.1, 35.1, 21, 20.9; HRMS (ESI): calcd for C22H19BrN3O5Na [M + Na+] : 864.0730, found: [M + Na+] 864.0734 (Δm = +0.0004; error = +0.08 ppm).
10 min before starting the experiments. All the experiments were replicated, and the \( T_m \) values reported are the average of 3 independent measurements. The absorbance versus temperature data was analyzed via a Marquadt algorithm for nonlinear curve fitting using Kaleidagraph (version 3.5) to decipher \( T_m \) and \( \Delta H^v \) values. The \( \Delta S^v \) and \( \Delta G^v \) values were calculated using following equations: \( \Delta S^v = \Delta H^v / T_m \) and \( \Delta G^v = \Delta H^v - T_m \Delta S^v \). All the thermodynamic parameters were derived from the analysis of three independent melting curves, and the average values are reported in Table 1. The error propagations for \( \Delta S^v \) and \( \Delta G^v \) were computed using the methods reported by Bevilaqua and co-workers.\(^{55} \)

**Circular Dichroism Experiments.** CD experiments were performed at wavelength in the range of 220–500 nm using a 1 mm path length quartz cuvette. The duplex DNAs (10 µM each strand) were annealed by heating at 95 °C for 5 min in phosphate buffer solution (100 mM NaCl, 20 mM NaPO₄, 0.1 mM EDTA, pH 7.4), then cooled to room temperature over a period of 3 h, and stored at 4 °C. The experiments were carried out at 25 °C, scanned at 100 nm/min using a response time of 2 s. Each spectrum was an average of 3 measurements; baseline corrected and analyzed using Origin (version 8.0).\(^{70} \)

**5′-Radiolabeling of Primer DNA.** The primer DNA (25 pmol) was mixed with T4 polynucleotide kinase (PNK) enzyme (5 U) and \([γ-^32P] \) ATP (20 µCi) in 1× PNK buffer (50 mM Tris-HCl, 10 mM MgCl₂, 0.1 mM spermidine, 0.1 mM EDTA, pH 7.6) in a total volume of 20 µL. The reaction mixture was incubated at 37 °C for 1 h followed by deactivation of the PNK enzyme by heating at 70 °C for 3 min. The radiolabeled primer was purified using a QIAquick nucleotide removal kit.

** Primer Extension Studies.** The primers were mixed with the unmodified and modified template in a molar ratio of 1:1.5 and annealed by heating to 95 °C for 3 min, followed by slow cooling to room temperature for 30 min. After annealing, the primer extension reactions were performed with a primer–template duplex in a total volume of 10 µL, containing 50 nM primer, traces of 5′-labeled primer, and 75 nM template, 100 µM concentration of all dNTPs in 1× polymerase buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM EDTA, pH 7.9). The primer extension reactions were initiated by incubation of the reaction mixture at 37 °C for 5 min, followed by the addition of a desired amount of KCl. The incorporation and misincorporation reactions were carried out using dCTP or dGTP or dATP or dTTP, and the extension was carried out using all dNTPs. The reactions were carried out at 37 °C for the indicated amount of time; aliquots were taken out of the reaction mixture and quenched by the addition of stop solution (80% formamide, 0.1 M EDTA), and a autoradiogram was generated and the bands were quantified by ImageQuant software. The percentage of each band was calculated by considering all the observed bands in the PAGE of the reaction using the following formula.

\[
\text{percentage incorporation/extension} = \frac{I_i}{(I_1 + I_2 + I_3)} \times 100
\]

\( I_i = \) intensity of primer band (nonshifted band), \( I_1 = \) intensity of shifted band 1, \( I_2 = \) intensity of shifted band 2 and so on. The same formula has been applied for calculating the percentage of each band in the same lane.

**Molecular Modeling Studies.** The force field parameters and RESP charges were developed for the LdG adduct using previously reported procedures.\(^{56,39} \) The adduct was placed in the identical ds DNA sequence, which was used in the thermal stability studies (Table 1, D3–D2). The B-form structure of the DNA was built using nucleic acid builder module in AMBER 12. The structure was initially minimized using 5000 steps of steepest descent minimization. Sodium ions were used to neutralize the backbone of the DNA; an octahedral TIP3P water box was extended up to 10 Å from any of the solute atoms. Consequently, a steepest descent minimization was carried out for 10000 steps, and then the system was heated from 10 to 300 K in 150 ps with the restraint force for 50 kcal/mol Å². The equilibration was carried
out for 700 ps with 1 atm pressure at constant temperature of 300 K. After this, the MD simulation of the duplex system was carried out for 50 ns. The RMSD, hydration, percentage of stacking and W–C H-bond occupancy of the duplex system was calculated using the PTRAJ module of AMBER 12.

The crystal structure of the BF (PDB entry: 1XC9) complexed with the template-containing BPDG adduct, and the primer was utilized as the model for MD simulations. The ff99SB with recent force field modifications41 was used for protein and DNA. The LdG adduct was energy optimized (HF/6-31G*) using Gaussian 0997, and RESP charges were calculated and fitted using a reported procedure.38,39 All the minimization and MD simulation were carried out using SANDER and PMEMD modules of the AMBER 12. The Leap module of AMBER 12 have been used to replace the BPDG with the LdG adduct at the template-primer junction. Hydrogen atoms were added to the crystal structure, and the structure was minimized using 500 steps of the steepest descent minimizations. Sodium ions were added to neutralize the net charge of the system using XLeap. A periodic octahedral TIP3P water box was extended up to 10 Å from any of the solute atoms which yielded around 10000 water molecules. Minimization of solvent and counterions for 5000 steps of the steepest descent and 5000 steps of conjugate gradient were carried out. After minimization, 50 ps of initial MD simulations were carried out to relax the water and Na+ ions, while the protein–DNA complex was detainted with a restraint force for 50 kcal/mol Å2. The system was then heated from 10 to 300 K at constant volume over 100 ps with the restraint force of 15 kcal/mol Å2 on solute atoms. The equilibrated system was then subjected to the production MD for 40 ns with constant 1 atm pressure and 300 K temperature. Long-range interactions were treated with the particle mesh Ewald method, and a 10 Å cutoff was applied for the nonbonded Lennard-Jones potential. The SHAKE algorithm was used to constrain all the hydrogen atoms. MD trajectories were visualized using UCSF Chimera. RMSD, ensemble, and superimposition analysis were carried out using the ptraj module in AMBER 12. Backbone torsions and base pair helical parameters were calculated using X3DNA.31 Figures were rendered using PyMOL.

**ASSOCIATED CONTENT**

1. Supporting Information

Synthetic strategy for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone; PAGE of full length extension reactions; energy optimized geometry and calculated RESP charges for the LdG modified nucleotide; time dependent RMSD maps of the protein and DNA complex; MD simulation snapshots from the conformational ensembles; superimposed structure of the BF complex with unmodified and LdG DNA; Watson-Crick H-bond occupancy of the LdG modified ds DNA; major and minor groove hydration of the LdG modified DNA; local base pair parameters of the template-primer ds DNA; local base pair step parameters of the template-primer ds DNA; backbone torsion angles of nucleotide in the template strand; MALDI-TOF mass data of the DNA sequences; the UV melting curves; and CD spectra for the unmodified and LdG modified DNA;1H, 13C, and 31P NMR spectra of all the new compounds; and MALDI spectra of the oligonucleotides. This material is available free of charge via the Internet at http://pubs.acs.org.

**AUTHOR INFORMATION**

**Corresponding Author**
*E-mail: pradeep@chem.iitb.ac.in*

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We are thankful to Prof. Deepak T. Nair for introducing us to the field of DNA adducts and polymerases and also for his critical reading of the manuscript; Prof. David A. Case for waiving the licensing fee for AMBER12; Prof. Philip C. Bevilacqua for the assistance in error analyses of the thermodynamic parameters; and Prof. I.N.N. Namboothiri for providing access to the melting point apparatus. We are also thankful to the central facility supported by IRCC-IIT Bombay for MALDI spectra and the Computer Centre, IIT Bombay, for use of the HPC facility. This work is financially supported by grants from the Department of Biotechnology (DBT)-Government of India. P.P.G. and S.H. thank the Council of Scientific and Industrial Research (CSIR) and Department of Atomic Energy-Board of Research in Nuclear Sciences (DAE-BRNS), respectively, for their fellowships.

**REFERENCES**


(37) Frisch, M. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.


Supporting Information

Synthesis and Polymerase Mediated Bypass Studies of the N²-Deoxyguanosine DNA Damage Caused by a Lucidin Analogue

Pratibha P. Ghodke, S. Harikrishna; and P. I. Pradeepkumar*

Department of Chemistry, Indian Institute of Technology Bombay,
Powai, Mumbai 400076, India
Email: pradeep@chem.iitb.ac.in

TABLE OF CONTENTS

Figure S1 Synthetic strategy for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone .................. Page S1
Figure S2 UV melting curves of the unmodified and LdG modified DNAs............................ Page S1
Figure S3 Full length extension reactions using the Klenow fragment exo⁻ (KF⁻) .................. Page S2
Figure S4 Energy optimized structure of the LdG nucleotide ........................................... Page S3
Figure S5 Time dependent RMSD map of the LdG modified ds DNA ............................... Page S3
Figure S6 CD spectra for the unmodified and LdG modified DNAs ................................. Page S4
Figure S7 Time dependent RMSD map of the protein and DNA complex .......................... Page S4
Figure S8 MD simulation snapshots from the conformational ensembles ........................... Page S5
Figure S9 Superimposed structure of the BF in complex with unmodified and LdG DNA ..... Page S5
Table S1 Watson-Crick H-bond occupancy of the LdG modified ds DNA ........................ Page S6
Table S2 Major and minor groove hydration of the LdG modified ds DNA ........................ Page S6
Table S3 Local base pair parameters of the template-primer ds DNA ............................... Page S7
Table S4 Local base pair step parameters of the template-primer ds DNA ........................ Page S7
Table S5 Backbone torsion angles of nucleotide in the template strand ............................ Page S8
Table S6 MALDI-TOF mass data of the DNA sequences ................................................ Page S8

1H NMR & 13C NMR spectrum of compound 2 ............................................................ Page S9
1H NMR & 13C NMR spectrum of compound 3 ............................................................ Page S10
1H NMR & 13C NMR spectrum of compound 5 ............................................................ Page S11
1H NMR & 13C NMR spectrum of compound 7 ............................................................ Page S12
1H NMR & 13C NMR spectrum of compound 8 ............................................................ Page S13
31P NMR spectrum of compound 9 ............................................................................. Page S14
MALDI spectrum of D3 .............................................................................................. Page S14
MALDI spectrum of D4 .............................................................................................. Page S15
MALDI spectrum of T ............................................................................................... Page S15
Synthetic strategy for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone

![Synthetic strategy diagram](image)

**Figure S1.** Synthetic strategy utilized for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone

UV thermal melting curves of the unmodified and LdG modified DNAs

![UV thermal melting curve](image)

**Figure S2.** UV thermal melting of the unmodified and the LdG modified DNA duplexes performed using the phosphate buffer (100 mM NaCl, 20 mM phosphate buffer, 0.1 mM EDTA, pH 7.4). Melting curves were obtained using 5 μM concentration of ds DNA.
Full length extension reactions using the Klenow fragment exo⁻ (KF⁻)

**Figure S3.** PAGE (20%, 7 M urea) of full length extension reactions using the KF⁻ with all dNTPs. A) Complete sequence of template, T (G* = LdG) and primer, P; B) unmodified template (dG) with 0.05 U of KF⁻; C) modified template (LdG) with 0.05 U of KF⁻; D) modified template (LdG) with 2 U of KF⁻. Lane P, primer; lane N, negative reaction control (no dNTPs); lanes 1 to 7: primer extension reactions with 100 μM of each dNTPs in different time course from 0.5, 5, 10, 30, 60, 120, and 240 min. All the reactions were carried out at 37 °C in the polymerase buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9).
Energy optimized structure of the LdG nucleotide

![Energy optimized structure of the LdG nucleotide](image)

<table>
<thead>
<tr>
<th>Atom Name</th>
<th>ESP Charge</th>
<th>Atom Name</th>
<th>ESP Charge</th>
<th>Atom Name</th>
<th>ESP Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1P</td>
<td>−0.576370</td>
<td>C6</td>
<td>0.550625</td>
<td>C14*</td>
<td>−0.163630</td>
</tr>
<tr>
<td>P</td>
<td>1.139459</td>
<td>O6</td>
<td>−0.556250</td>
<td>C24*</td>
<td>0.449486</td>
</tr>
<tr>
<td>O2P</td>
<td>−0.576370</td>
<td>N1</td>
<td>−0.588450</td>
<td>O8*</td>
<td>−0.454690</td>
</tr>
<tr>
<td>O5'</td>
<td>−0.247220</td>
<td>H1</td>
<td>0.323001</td>
<td>C16*</td>
<td>−0.181060</td>
</tr>
<tr>
<td>C5'</td>
<td>−0.122050</td>
<td>C2</td>
<td>0.717723</td>
<td>C23*</td>
<td>0.537457</td>
</tr>
<tr>
<td>1H5'</td>
<td>0.090022</td>
<td>N2</td>
<td>−0.472100</td>
<td>O7*</td>
<td>−0.513280</td>
</tr>
<tr>
<td>2H5'</td>
<td>0.090022</td>
<td>H22</td>
<td>0.296364</td>
<td>C19*</td>
<td>−0.081320</td>
</tr>
<tr>
<td>C4'</td>
<td>0.397914</td>
<td>C10*</td>
<td>0.075580</td>
<td>C17*</td>
<td>−0.035360</td>
</tr>
<tr>
<td>C3'</td>
<td>0.478016</td>
<td>H9*</td>
<td>0.039145</td>
<td>C18*</td>
<td>−0.082450</td>
</tr>
<tr>
<td>O3'</td>
<td>−0.805530</td>
<td>H10*</td>
<td>0.039145</td>
<td>H12*</td>
<td>0.107828</td>
</tr>
<tr>
<td>H3'</td>
<td>−0.055220</td>
<td>C11*</td>
<td>−0.119800</td>
<td>C20*</td>
<td>−0.101130</td>
</tr>
<tr>
<td>C2'</td>
<td>−0.431940</td>
<td>C13*</td>
<td>0.160285</td>
<td>H13*</td>
<td>0.105460</td>
</tr>
<tr>
<td>2H1'</td>
<td>0.097007</td>
<td>C15*</td>
<td>−0.111250</td>
<td>C22*</td>
<td>−0.115180</td>
</tr>
<tr>
<td>2H2'</td>
<td>0.097007</td>
<td>H11*</td>
<td>0.122802</td>
<td>H15*</td>
<td>0.105309</td>
</tr>
<tr>
<td>H4'</td>
<td>−0.031130</td>
<td>O10*</td>
<td>−0.158150</td>
<td>C21*</td>
<td>−0.075970</td>
</tr>
<tr>
<td>O4'</td>
<td>−0.571350</td>
<td>C26*</td>
<td>−0.205060</td>
<td>H14*</td>
<td>0.105668</td>
</tr>
<tr>
<td>C1'</td>
<td>0.481095</td>
<td>H19*</td>
<td>0.115604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1'</td>
<td>0.026656</td>
<td>H20*</td>
<td>0.115604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N9</td>
<td>−0.186200</td>
<td>H21*</td>
<td>0.115604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>0.393454</td>
<td>C12*</td>
<td>0.276178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>−0.778790</td>
<td>O9*</td>
<td>−0.224760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>0.269233</td>
<td>C25*</td>
<td>−0.035380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>0.047512</td>
<td>H16*</td>
<td>0.067443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N7</td>
<td>−0.558490</td>
<td>H17*</td>
<td>0.067443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>0.047314</td>
<td>H18*</td>
<td>0.067443</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure S4.** Energy optimized geometry and calculated RESP charges for the LdG using HF/6-31G* basis set in Gaussian 09 program.\(^1\)\(^3\)
Time dependent RMSD map of the LdG modified ds DNA

![RMSD graph](image)

**Figure S5.** Root mean square deviation graphs of the trajectories obtained from MD simulation of the LdG modified duplex DNA. RMSD values from the MD trajectories were calculated with respect to the corresponding initial structure using ptraj module in AMBER 12.

CD spectra for the unmodified and LdG modified DNAs

![CD spectra graph](image)

**Figure S6.** CD spectra for the unmodified (D1-D2) and the LdG modified (D3-D2) DNA duplexes (10 µM of each DNA strand in 100 mM NaCl, 20 mM sodium phosphate, 0.1 mM EDTA, pH 7.4). CD spectra show the characteristic peaks of B-form DNA. The enhanced CD intensity (280 nm) for the LdG modified DNA in comparison to the unmodified DNA may be attributed to the rigidly bound polycyclic aromatic adduct in the minor groove as revealed in our MD simulations. Similar CD spectrum is also observed for the N2-CH2(9-anthracenyl)-dG modified DNA.4
Time dependent RMSD map of the protein and DNA complex

![RMSD graph](image)

**Figure S7.** Root mean square deviation graphs of the trajectories obtained from MD simulation of protein and the LdG modified template and primer DNA. RMSD values from the MD trajectories were calculated with the corresponding initial structure using ptraj module in AMBER 12.

MD simulation snapshots from the conformational ensembles

(i) ![Snapshot](image)  (ii) ![Snapshot](image)

**Figure S8.** MD simulation snapshots around the LdG modification of template-primer DNA duplex obtained from ensemble analysis. The LdG modification is highlighted in magenta.

Superimposed structure of the BF in complex with unmodified and LdG DNA

![Superimposed structure](image)

**Figure S9.** Superimposed structure of the *Bacillus* fragment (BF) in complex with the LdG modified DNA (yellow) and the unmodified DNA (green). The LdG modification (yellow) with the pairing dC and the dG with pairing dC is highlighted.
**Watson-Crick H-bond occupancy of the LdG modified ds DNA**

<table>
<thead>
<tr>
<th>Duplex DNA (D2:D3)</th>
<th>Percentage occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>dC:dG</td>
<td>90 %</td>
</tr>
<tr>
<td>dC:LdG</td>
<td>45 %</td>
</tr>
<tr>
<td>dT:dA</td>
<td>75 %</td>
</tr>
<tr>
<td>dT:dA</td>
<td>83 %</td>
</tr>
</tbody>
</table>

Table S1. The Watson-Crick H-bond occupancy of the LdG modified duplex DNA during 50 ns of MD simulations. The percentage of H-bond occupancy was calculated using PTRAJ module in AMBER 12.

**Major and minor groove hydration of the LdG modified ds DNA**

<table>
<thead>
<tr>
<th>Duplex DNA (D2:D3)</th>
<th>Major groove</th>
<th>Minor groove</th>
</tr>
</thead>
<tbody>
<tr>
<td>dC:dG</td>
<td>3 (95 %)</td>
<td>1 (60 %)</td>
</tr>
<tr>
<td>dC:LdG</td>
<td>1 (25 %)</td>
<td>0</td>
</tr>
<tr>
<td>dT:dA</td>
<td>2 (44 %)</td>
<td>0</td>
</tr>
<tr>
<td>dT:dA</td>
<td>2 (78 %)</td>
<td>1 (65 %)</td>
</tr>
</tbody>
</table>

Table S2. Hydration of major and minor groove of the LdG modified duplex DNA during 50 ns of MD simulations. The table indicates the number of water molecules present in the minor and the major groove of the base pairs, and the percentage occupancy of the water molecules are shown in the brackets. Water molecule interaction with bases and their occupancy was calculated using PTRAJ module in AMBER 12.
### Local base pair parameters of the template-primer ds DNA

<table>
<thead>
<tr>
<th>Template: Primer</th>
<th>C1'-C1' distance</th>
<th>Shear</th>
<th>Stretch</th>
<th>Stagger</th>
<th>Buckle</th>
<th>Propeller</th>
<th>Opening</th>
</tr>
</thead>
<tbody>
<tr>
<td>LdG:dC</td>
<td>11.95 (10.6)</td>
<td>0.79</td>
<td>−0.06</td>
<td>−0.75</td>
<td>33.25</td>
<td>−14.27</td>
<td>2.15</td>
</tr>
<tr>
<td>dC:dG</td>
<td>10.82 (10.6)</td>
<td>−0.45</td>
<td>0.22</td>
<td>0.14</td>
<td>16.59</td>
<td>−1.14</td>
<td>9.26</td>
</tr>
<tr>
<td>dA:dT</td>
<td>10.68 (10.7)</td>
<td>−0.24</td>
<td>0.06</td>
<td>0.52</td>
<td>10.75</td>
<td>−24.58</td>
<td>8.47</td>
</tr>
<tr>
<td>dC:dG</td>
<td>10.58 (10.6)</td>
<td>−0.32</td>
<td>0.10</td>
<td>0.61</td>
<td>7.76</td>
<td>−7.53</td>
<td>5.68</td>
</tr>
</tbody>
</table>

Table S3. Six local base pair parameters for each base pair were calculated from the averaged structure obtained from MD simulation. The values shown in the bracket are for the unmodified nucleotide base pair in complex with the BF (PDB entry: 1L3S). All the base pair parameters for each step were calculated using X3DNA program. The shear, stretch and stagger values are reported in Å, buckle propeller, and opening values are reported in degrees.

### Local base pair step parameters of the template-primer ds DNA

<table>
<thead>
<tr>
<th>Template: Primer</th>
<th>Shift</th>
<th>Slide</th>
<th>Rise</th>
<th>Tilt</th>
<th>Roll</th>
<th>Twist</th>
</tr>
</thead>
<tbody>
<tr>
<td>LdG:dC</td>
<td>−0.57</td>
<td>0.58</td>
<td>3.28</td>
<td>16.52</td>
<td>21.94</td>
<td>26.35</td>
</tr>
<tr>
<td>dC:dG</td>
<td>0.29</td>
<td>0.21</td>
<td>3.06</td>
<td>0.27</td>
<td>6.35</td>
<td>34.27</td>
</tr>
<tr>
<td>dA:dT</td>
<td>−0.48</td>
<td>−0.43</td>
<td>3.33</td>
<td>−2.45</td>
<td>6.47</td>
<td>37.58</td>
</tr>
<tr>
<td>dC:dG</td>
<td>−0.33</td>
<td>−0.16</td>
<td>3.74</td>
<td>−5.21</td>
<td>10.58</td>
<td>31.24</td>
</tr>
</tbody>
</table>

Table S4. Six local base pair step parameters for each base pair were calculated from the averaged structure obtained from MD simulations. The values shown in the bracket corresponds to that of unmodified nucleotide base pair in complex with BF (PDB entry: 1L3S). All the base pair parameters for each step were calculated using the X3DNA program. The shift, slide and rise values are reported in Å, tilt, roll and twist are reported in degrees.
### Backbone torsion angles of the nucleotide in the template strand

<table>
<thead>
<tr>
<th>Template Strand</th>
<th>$\alpha$ (O3'-P-O5'-C5')</th>
<th>$\beta$ (P-O5'-C5'-C4')</th>
<th>$\gamma$ (O5'-C5'-C4'-C3')</th>
<th>$\delta$ (C5'-C4'-C3'-O3')</th>
<th>$\epsilon$ (C4'-C3'-O3'-P)</th>
<th>$\zeta$ (C3'-O3'-P-O5')</th>
</tr>
</thead>
<tbody>
<tr>
<td>LdG</td>
<td>97.3 (93.2)</td>
<td>-176.1 (-170.3)</td>
<td>68.4 (61.4)</td>
<td>94.8 (82.5)</td>
<td>140.4 (171.5)</td>
<td>16.5 (-52.2)</td>
</tr>
<tr>
<td>dC</td>
<td>-95.5 (-90.5)</td>
<td>171.1 (172.4)</td>
<td>54.8 (52.1)</td>
<td>70.9 (62.4)</td>
<td>125.3 (176.8)</td>
<td>-87.7 (-20.8)</td>
</tr>
<tr>
<td>dA</td>
<td>-83.3 (-84.2)</td>
<td>179.2 (177.8)</td>
<td>54.1 (50.7)</td>
<td>93.0 (87.8)</td>
<td>-134.2 (-175.4)</td>
<td>-91.9 (-29.6)</td>
</tr>
<tr>
<td>dC</td>
<td>-64.2 (-66.8)</td>
<td>-177.9 (-179.1)</td>
<td>70.2 (60.4)</td>
<td>133.3 (117.5)</td>
<td>-173.3 (-175.3)</td>
<td>-87.6 (-71.4)</td>
</tr>
</tbody>
</table>

Table S5. Nucleotide backbone torsion angles were calculated from the averaged structure obtained from the MD simulations using X3DNA. The values shown in the bracket are for the unmodified nucleotides (PDB entry: 1L3S). All the values are mentioned in degrees.

### MALDI-TOF mass data of the DNA sequences

<table>
<thead>
<tr>
<th>Code</th>
<th>DNA Sequences (5'-3')</th>
<th>MW (calc.)</th>
<th>MW (found)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>GCCGGAATAGCGCA</td>
<td>4299</td>
<td>4299</td>
</tr>
<tr>
<td>D2</td>
<td>TGCGCTATTCCGGC</td>
<td>4232</td>
<td>4233</td>
</tr>
<tr>
<td>D3</td>
<td>GCCCGG*AATAGCGCA</td>
<td>4579</td>
<td>4579</td>
</tr>
<tr>
<td>D4</td>
<td>TGCGG*CTATTCCGGC</td>
<td>4512</td>
<td>4513</td>
</tr>
<tr>
<td>T</td>
<td>TCCTACCGTGCTACCTGAACAGCTGGTCACACTG*AT</td>
<td>15547</td>
<td>15547</td>
</tr>
<tr>
<td></td>
<td>GCCCTACGAGTACG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>CGTACTCGTAGGCGAT</td>
<td>4569</td>
<td>4571</td>
</tr>
</tbody>
</table>

Table S6. Unmodified and the LdG modified DNAs were characterized using MALDI-TOF positive reflectron mode. (G* = LdG).

### References

1. Frisch, M. J.; Gaussian 09, Revision A.02: Gaussian, Inc.: Wallingford, CT, USA, **2009**.
NMR spectra ($^1$H, $^{13}$C, & $^{31}$P) (G-Grease, I-Impurity)

$^1$H NMR spectrum of compound 2

$^{13}$C NMR spectrum of compound 2
$^1$H NMR spectrum of compound 3

$^{13}$C NMR spectrum of compound 3
$^1$H NMR spectrum of compound 5

$^{13}$C NMR spectrum of compound 5
$^1$H NMR spectrum of compound 7

$^{13}$C NMR spectrum of compound 7
$^1$H NMR spectrum of compound 8

![H NMR spectrum](image)

$^{13}$C NMR spectrum of compound 8

![C NMR spectrum](image)
$^{31}$P NMR spectrum of compound 9

MALDI spectrum of D3 $5'\text{-GCCGG\text{*AATAGCGCA-3'}}$
Calc. mass, [M + H]$^+$ 4579; Obs. mass, [M + H]$^+$ 4579
MALDI spectrum of \textbf{D4} 5’-TGCG*CTATTCCGGC-3’
Calc. mass, [M + H]$^+$ 4512; Obs. mass, [M + H]$^+$ 4513

![MALDI spectrum of D4](image1.png)

MALDI spectrum of \textbf{T}, 5’-TCCTACCGTGCTACCTGAACAGCTGGTCACACTG*ATG CCTACGAGTACG-3’ Calc. mass, [M + H]$^+$ 15547; Obs. mass, [M + H]$^+$ 15547

![MALDI spectrum of T](image2.png)