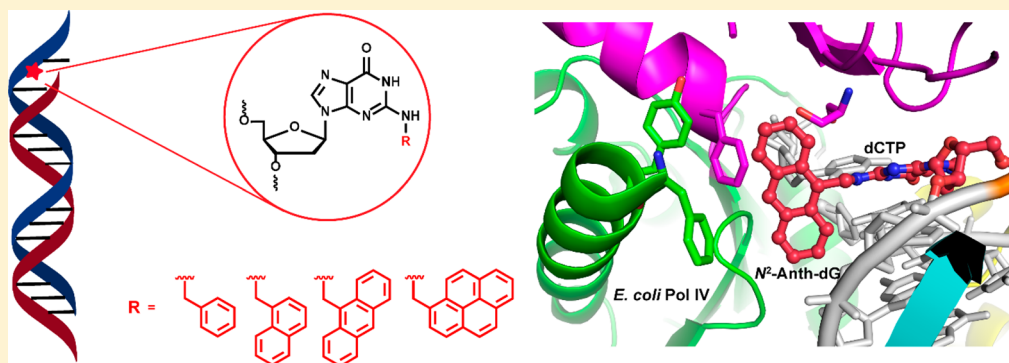


Synthesis of N^2 -Deoxyguanosine Modified DNAs and the Studies on Their Translesion Synthesis by the *E. coli* DNA Polymerase IV

Pratibha P. Ghodke,[†] Praneeth Bommiseti,[†] Deepak T. Nair,^{*,‡} and P. I. Pradeepkumar^{*,†}[†]Department of Chemistry, Indian Institute of Technology Bombay, Mumbai 400076, India[‡]Regional Centre for Biotechnology, NCR Biotech Science Cluster, third Milestone, Faridabad-Gurgaon Expressway, Faridabad 121001, India

Supporting Information



ABSTRACT: We report the synthesis of N^2 -aryl (benzyl, naphthyl, anthracenyl, and pyrenyl)-deoxyguanosine (dG) modified phosphoramidite building blocks and the corresponding damaged DNAs. Primer extension studies using *E. coli* Pol IV, a translesion polymerase, demonstrate that translesion synthesis (TLS) across these N^2 -dG adducts is error free. However, the efficiency of TLS activity decreases with increase in the steric bulkiness of the adducts. Molecular dynamics simulations of damaged DNA-Pol IV complexes reveal the van der Waals interactions between key amino acid residues (Phe13, Ile31, Gly32, Gly33, Ser42, Pro73, Gly74, Phe76, and Tyr79) of the enzyme and adduct that help to accommodate the bulky damages in a hydrophobic pocket to facilitate TLS. Overall, the results presented here provide insights into the TLS across N^2 -aryl-dG damaged DNAs by Pol IV.

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are fused benzenoid ring containing compounds that are well-known for their carcinogenic and mutagenic properties.¹ The exocyclic amino group of deoxyguanosine (dG) acts as a softer reaction center to form major carcinogenic N^2 -dG DNA adducts.^{2,3} These adducts can lead to genomic alterations via stalling replicative polymerases. To avoid any detrimental consequence, cells use specialized low-fidelity DNA polymerases belonging to the Y-family to bypass the damaged sites. This process is called translesion synthesis (TLS), an evolutionarily conserved DNA damage tolerance pathway, which can be error-free or error-prone.⁴ Among various TLS polymerases, Pol IV from *E. coli* is well-known for accurate and efficient bypass of N^2 -dG adducts.^{5–7} Pol IV is an ortholog of human TLS polymerase, Pol κ .⁵ A systematic study on the effects of the steric bulkiness of N^2 -dG adducts on TLS by *E. coli* Pol IV has yet to be reported.

Among the N^2 -dG DNA adducts, N^2 -benzyl-dG (N^2 -Bn-dG, Figure 1) is potentially formed from three key carcinogens. *N*-Nitroso-*N*-benzylurea (BnNU) and *N*-nitrosobenzylmethylamine (NBnMA) could form N^2 -Bn-dG

adducts via benzyldiazonium ion⁸ and methyl hydroxylation⁹ pathway, respectively. Whereas, benzyl halides could directly act as a benzylating agent.^{10,11} Among these carcinogens, NBnMA is believed to form N^7 -benzyl-dG adduct in vivo conditions, since the N^7 -position is one of the major site of DNA alkylation.¹¹ The 9-(sulfoxymethyl)anthracene is an electrophilic metabolite of 9-hydroxymethyl-anthracene, which was alluded to be responsible for the formation of N^2 -(9-anthracenylmethyl)-dG adduct (N^2 -Anth-dG, Figure 1).¹² Halomethyl-anthracenes such as bromo and chloromethyl derivatives have as well been suggested to alkylate the N^2 -position of dG.¹³ 1-Methylpyrene, a common environmental pollutant can enzymatically be activated to 1-(hydroxymethyl)pyrene, which in turn can be converted to 1-(sulfoxymethyl)pyrene (SMP). SMP on reaction with DNA leads to the formation of N^2 -(1-pyrenylmethyl)-dG adduct (N^2 -Pyre-dG, Figure 1).^{14,15} Though N^2 -Bn, N^2 -Naph, and N^2 -Anth (Figure 1) adducts are yet to be detected in cellular

Received: August 11, 2018

Published: January 10, 2019

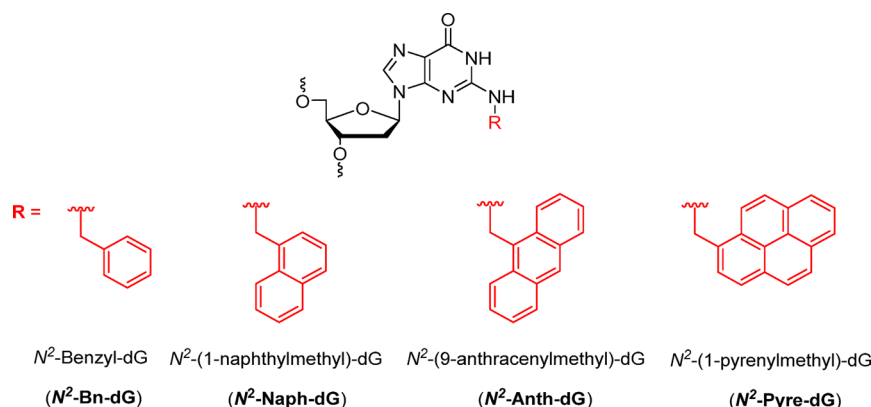
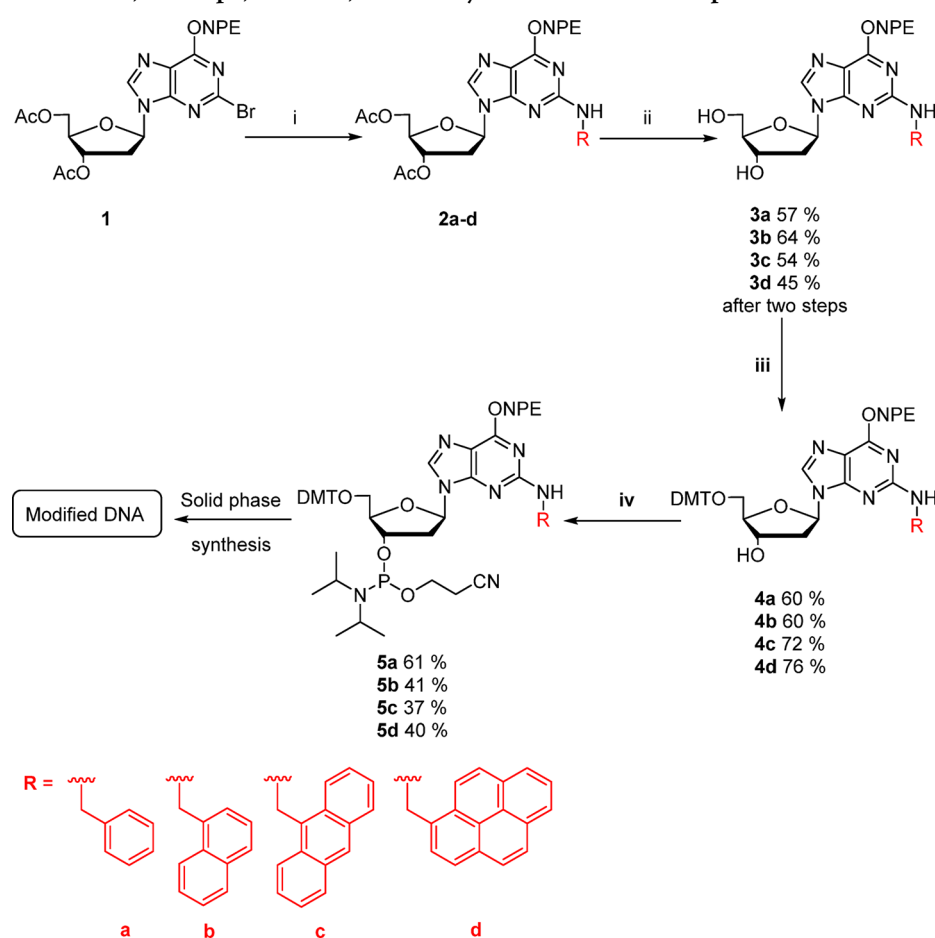


Figure 1. Structures of N^2 -dG adducts which originate from polycyclic aromatic hydrocarbons.

Scheme 1. Synthesis of N^2 -Bn, N^2 -Naph, N^2 -Anth, and N^2 -Pyre-dG Modified Phosphoramidites 5a–d^a



^aReagent and conditions: (i) Pd(OAc)₂, (R)-BINAP, respective amines, Cs₂CO₃, toluene, 85 °C; (ii) 33% MeNH₂ in EtOH, rt; (iii) DMT-Cl, pyridine, rt; (iv) CEP-Cl, DIPEA, DCM, rt.

DNA, the N^2 -Pyre-dG adduct was detected in DNA from lung, liver, and kidney tissues of mice.¹⁴

To unravel the molecular mechanism of TLS across N^2 -aryl-dG adducts formed by nondiol epoxide metabolic pathways by using various biochemical and biophysical methods, the corresponding modified DNA oligonucleotides are required. Herein we report a short and reliable synthesis of N^2 -dG adducts of varying steric bulkiness and the corresponding modified DNAs. Using primer extension assays, molecular dynamics studies, we were able to gain

insights on the TLS by *E. coli* Pol IV across N^2 -(Bn, Naph, Anth, Pyre)-dG damaged templates.

RESULTS AND DISCUSSION

To access various N^2 -dG damaged DNAs, either phosphoramidite or postoligomerization approaches were reported.^{16–18} Synthesis of the N^2 -Bn-dG adduct was achieved by different methods such as substitution of C2-trimethanesulfonate group bearing dG by benzylamine,¹⁹ aromatic nucleophilic substitution of 2-fluoro-2'-deoxyinosine (2-

Table 1. N^2 -dG Damaged Oligonucleotides and Their Molecular Weights^a

Code	DNA Sequences (5'-3')	Mol wt. (Calcd.)	Mol wt. (Found)	Error, % (found-calcd.)
D1	TCTA G1 GGTCCTAGGACCC	5566.7	5566.0	-0.01
D2	TCT G1 GGGTCCCTAGGACCC	5581.7	5581.9	0.004
D3	TCTAGG G1 TCCTAGGACCC	5566.7	5567.3	0.01
D4	TCCTACCGTGCCTACCTGAACAGCTG GTCACACT G1 ATGCCTACGAGTACG	15358	15354.8	-0.02
D5	TCTA G2 GGTCCTAGGACCC	5616.8	5616.8	0
D6	TCT G2 GGGTCCCTAGGACCC	5632.8	5635.1	0.04
D7	TCTAGG G2 TCCTAGGACCC	5616.8	5617.4	0.01
D8	TCCTACCGTGCCTACCTGAACAGCT GGTCACACT G2 ATGCCTACGAGTACG	15408.1	15409.1	0.007
D9	TCTA G3 GGTCCTAGGACCC	5666.9	5667.6	0.01
D10	TCT G3 GGGTCCCTAGGACCC	5682.9	5682.0	-0.02
D11	TCTAGG G3 TCCTAGGACCC	5666.9	5666.9	0.001
D12	TCCTACCGTGCCTACCTGAACAGCTG GTCACACT G3 ATGCCTACGAGTACG	15458.2	15453.8	-0.03
D13	TCTA G4 GGTCCTAGGACCC	5690.9	5690.2	-0.01
D14	TCT G4 GGGTCCCTAGGACCC	5706.9	5705.8	-0.02
D15	TCTAGG G4 TCCTAGGACCC	5690.9	5689.9	-0.02
D16	TCCTACCGTGCCTACCTGAACAGCTG GTCACACT G4 ATGCCTACGAGTACG	15482.2	15480.9	-0.01

^aStructures of modifications **G1** (N^2 -Bn-dG), **G2** (N^2 -Naph-dG), **G3** (N^2 -Anth-dG), and **G4** (N^2 -Pyre-dG) are shown in Figure 1. The modified DNAs were characterized by MALDI-TOF in the positive reflectron/linear mode or negative ion electrospray ionization (ESI, only for D2) technique (calculated and found).

fluoro-dI),²⁰ reaction of dG with benzyl bromide,^{10,21} reduction of N^2 -benzoyl group using reducing agents,²² reductive amination of dG with benzaldehyde,²³ and by Buchwald-Hartwig (B-H) coupling between TBDMS protected 2-bromo-2'-deoxyinosine (2-bromo-dI) with benzylamine.²⁴ The N^2 -Naph-dG adduct was synthesized using nucleophilic substitution of DMT protected 2-fluoro-dI with the corresponding amine.²⁰ Synthesis of N^2 -Anth-dG adduct was achieved using B-H coupling²⁴ or by the hydrolysis of 2'-deoxy-4-desmethylwyosine.¹³ The postoligomerization approach was also utilized for the synthesis of N^2 -Anth-dG adduct employing 2-fluoro-dI bearing oligonucleotide.²³ The N^2 -Pyre-dG adduct was accessed by fluoro substitution method using 2-fluoro-dI convertible nucleoside and 1-(aminomethyl)pyrene.²⁵

Our objective was to develop a robust protocol to synthesize N^2 -aryl-dG modified nucleoside and corresponding oligonucleotides (Figure 1) from a common precursor. To achieve this, first, we attempted various reported methods. However, in our hands, many of those methods were marred by lack of reproducibility, capricious yields, and poor solubility of various intermediates/amidites.²⁶ In the present work, we report a reliable and efficient method for the synthesis of various N^2 -aryl-dG modified phosphoramidite building blocks employing B-H coupling as the key step. These modified building blocks were synthesized in the presence of a versatile *para*-nitrophenylethyl (NPE) protect-

ing group at the O^6 -position of dG. This route is more convenient since all of the bulky modifications can be synthesized from a common precursor with good yields.

Synthesis of N^2 -Bn, N^2 -Naph, N^2 -Anth, and N^2 -Pyre-dG Modified Phosphoramidites 5a-d. The synthetic Scheme 1 was utilized to access N^2 -aryl-dG (N^2 -Bn, N^2 -Naph, N^2 -Anth, and N^2 -Pyre-dG) phosphoramidite building blocks. Using the protected 2-Br-dI, **1**, as a common precursor,²⁷ the C-N bond formation was carried out with different polycyclic aromatic amines in the presence of Pd(OAc)₂, (*R*)-BINAP, Cs₂CO₃ in toluene.²⁸ For N^2 -Bn and N^2 -Naph-dG modifications, B-H coupling were carried out using commercially available amine, while for N^2 -Anth-dG and N^2 -Pyre-dG modifications, the respective amines were prepared by reported methods.^{25,29,30} Deprotection of acyl protecting group was accomplished using 33% MeNH₂ in EtOH (v/v) to obtain N^2 -aryl-dG diols (**3a-d**). Final modified phosphoramidites **5a-d** were synthesized by the protection of the 5'-OH group using DMT-Cl, followed by phosphitylation reaction using CEP-Cl, DIPEA in DCM.

Solid Phase Synthesis of N^2 -dG Modified Oligonucleotides. The N^2 -dG modified phosphoramidite building blocks (**5a-5d**) were used to synthesize the desired modified oligonucleotides. The presence of versatile NPE protecting group at O^6 -position of dG provided enhanced solubility of respective phosphoramidites during solid phase DNA synthesis.³¹ We have chosen the NPE protecting group because

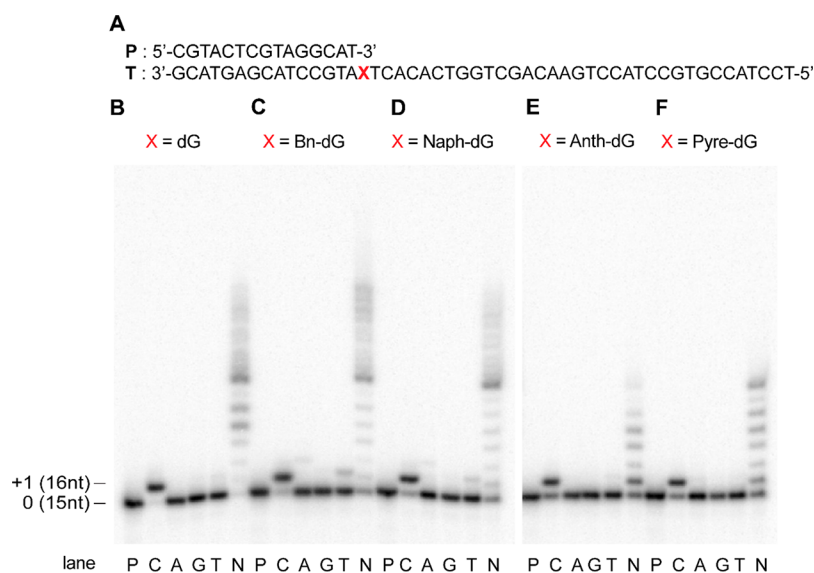


Figure 2. PAGE (20%, 7 M urea) of primer extension reactions with single and with the mixture of dNTPs using unmodified (dG) and modified (N^2 -dG adducts) templates employing *E. coli* Pol IV. (A) complete sequence of the primer, P (15-mer) and the template, T (50-mer); (B) reactions with dG template (X = dG; lanes: P, primer; C, dCTP; A, dATP; G, dGTP; T, dTTP; N, mixture of dNTPs); (C) reactions with N^2 -Bn-dG template; (D) reactions with N^2 -Naph-dG template; (E) reactions with N^2 -Anth-dG template; and (F) reactions with N^2 -Pyre-dG template. All reactions were carried out at 37 °C for 1.5 h time course.

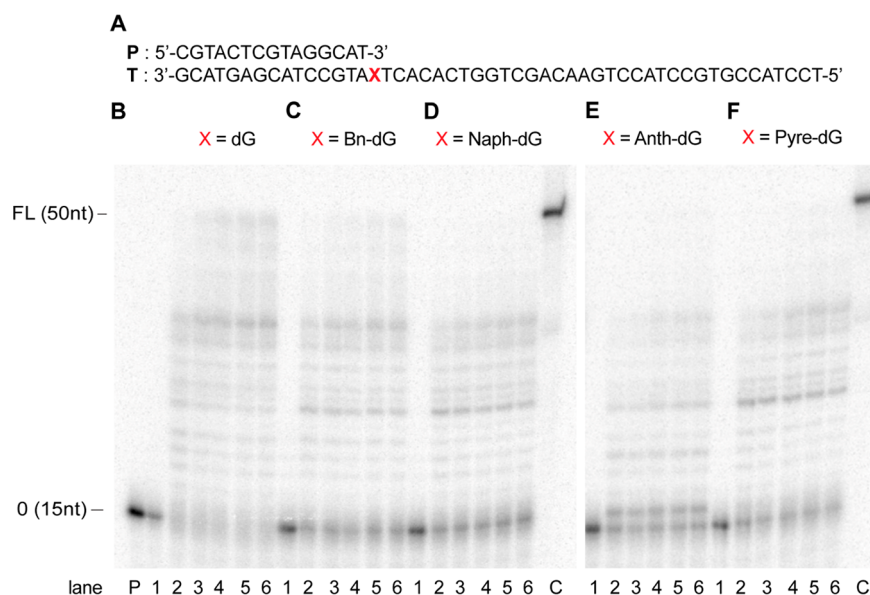


Figure 3. PAGE (20%, 7 M urea) of full-length extension reactions employing Pol IV from *E. coli* with all dNTPs. (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with dG template (X = dG); (C) reactions with N^2 -Bn-dG template; (D) reactions with N^2 -Naph-dG template; (E) reactions with N^2 -Anth-dG template; (F) reactions with N^2 -Pyre-dG template. Lane P, primer; lanes 1 to 6: primer extension reactions with 250 μ M of mixture of dNTPs in different time course from 30 s, 30 min, 1, 2, 4, and 8 h; lane C, 50-mer size standard. All of the reactions were carried out at 37 °C.

of its chemical inertness and its stability in the presence of mild acids and bases.³² The N^2 -aryl-dG damaged DNA sequences (Table 1) were synthesized using automated DNA synthesizer employing appropriate controlled pore glass (CPG) solid supports. The 18-mer DNA sequences (Table 1) were utilized for crystallization with complementary 14-mer primer, incoming nucleotide and *E. coli* Pol IV.³³ These sequences were designed for easy assembling of the three functional complexes of Pol IV with three different DNA duplexes containing the N^2 -dG adduct and incoming nucleotide. The 50-mer N^2 -dG DNA template sequences

were synthesized for the primer extension studies (Table 1). Deprotection of N^2 -dG modified DNAs was carried out in four different steps. The first step involved the selective deprotection of the cyanoethyl group using 10% diethylamine in ACN.³⁴ The deprotection of the NPE group was performed using 1 M DBU in ACN.²³ The cleavage of the CPG support and deprotection of the base protecting groups were carried out separately by treatment of aq. NH_3 . Modified DNAs were PAGE purified and the integrity of modified DNAs was confirmed by ESI-MS or MALDI-TOF (Table 1).

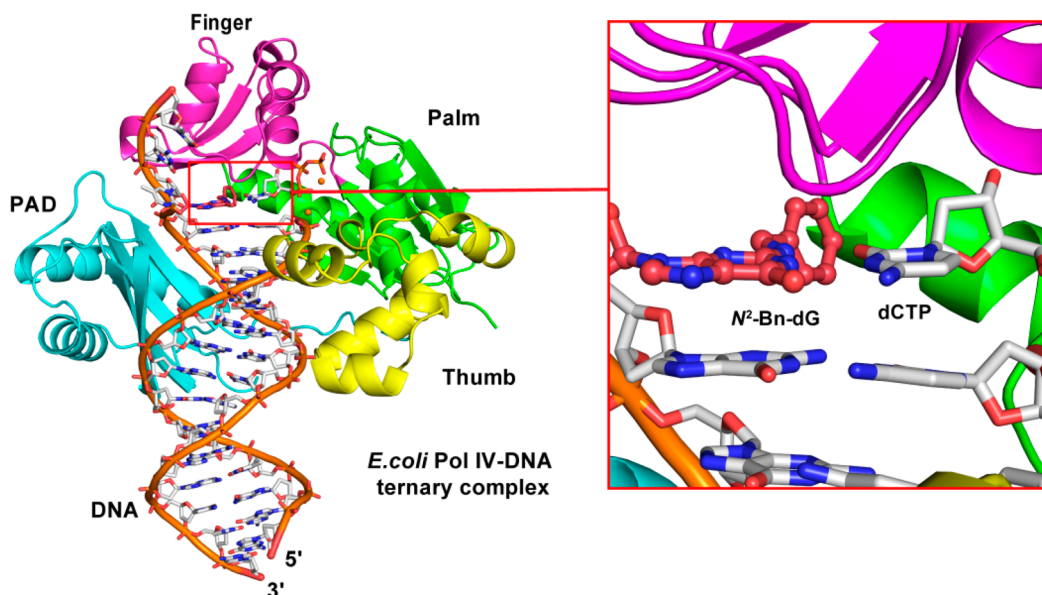


Figure 4. DNA-Pol IV ternary complex used for MD simulations. The initial structure of N^2 -Bn complex prepared from fdG containing crystal structure (PDB ID: 4Q43). The red colored rectangle is used to highlight the active site and zoomed view of the active site. Other complexes are shown in Figure S9, [Supporting Information](#). Cyan, magenta, green, and yellow are used for polymerase associated domain (PAD) and finger, palm, and thumb domains, respectively. DNA backbone is represented in orange-red color. The N^2 -modified residue is highlighted in desire red color. Carbon atoms are represented in white (desire red for N^2 -Bn-dG), nitrogen atoms in blue, oxygen atoms in red, and magnesium ions in orange color.

Primer Extension Studies Using *E. coli* DNA Pol IV.

Primer extension reactions with *E. coli* Pol IV were carried out using a primer-template system as shown in [Figure 2A](#). Reactions were carried out with individual dNTP or with a mixture of dNTPs employing unmodified and modified DNA templates and were monitored at various time courses from 30 s to 2 h ([Figure 2](#) and [Figures S1–S5](#) of the [Supporting Information](#)). In the case of single nucleotide incorporation reactions, we observed a prominent extension product of 16 nucleotide length, which corresponds to correct base (dCTP) incorporation across all of the N^2 -dG modified templates ([Figure 2B–F](#), lane C). With the use of 4 nM of Pol IV, ~94% incorporation of the correct base was observed opposite to unmodified dG template at 1.5 h time period ([Figure 2B](#), lane C). For the N^2 -Bn-dG modification ~80% ([Figure 2C](#), lane C), for the N^2 -Naph-dG modification ~75% ([Figure 2D](#), lane C), for the N^2 -Anth-dG modification ~62% ([Figure 2E](#), lane C), and for the N^2 -Pyre-dG modification ~64% ([Figure 2F](#), lane C) incorporation products were observed after 1.5 h. These results reveal that the TLS employing Pol IV is error free irrespective of the steric bulkiness of the N^2 -dG adducts.^{5,35} However, incorporation efficiency of Pol IV decreases with increasing steric bulkiness from N^2 -Bn to N^2 -Pyre-dG adducts. It should be noted that, in addition to the correct incorporation, a trace amount of dTTP misincorporation was also observed opposite to the N^2 -Bn-dG (~10%, [Figure 2C](#), lane T) and the N^2 -Naph-dG modifications (~4%, [Figure 2D](#), lane T).

To verify the efficiency of *E. coli* Pol IV toward full length extension of the primer, running start experiments were carried out using 15-mer and 11-mer primers with the modified DNA templates. Running start experiments using 15-mer primer showed that the DNA Pol IV is unable to synthesize full length products irrespective of the chemical nature of the damage ([Figure 3B–F](#)). Similar results were

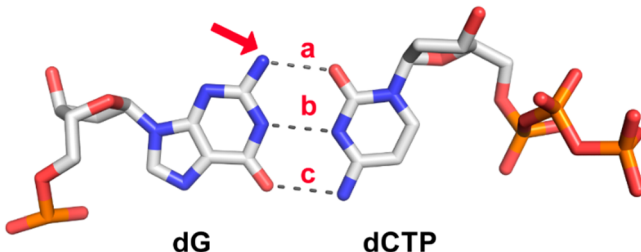
obtained in running start experiments with 11-mer primer ([Figures S6](#) and [S7](#) of the [Supporting Information](#)). As expected, in the case of an unmodified template, we observed traces of the full-length product. These results are in line with the low processivity reported for *E. coli* Pol IV, which is a hallmark of this family of DNA polymerases.^{5,36} It should be noted that Pol IV belongs to the Y-family of DNA polymerases, and members of this family are known to bypass DNA lesions and rescue replication stalled at damaged nucleotides. The majority of Y-family DNA polymerases serve only to extend the replication fork past the DNA lesion after which they are replaced by replicative DNA polymerases.⁴

Molecular Modeling and Dynamics Studies. Molecular dynamics (MD) simulations have been carried out to elucidate the factors that facilitate TLS across N^2 -dG DNA adducts by *E. coli* Pol IV. The reported X-ray crystal structure of furfuryl modified N^2 -dG DNA adduct in complex with the *E. coli* Pol IV (PDB ID: 4Q43⁶) was used to prepare the starting structures. The N^2 -dG adducts, which were used in the experimental studies ([Figure 1](#)) were incorporated into the DNA ([Figure S8](#), [Supporting Information](#)). A total of four DNA-Pol IV ternary complexes (enzyme, DNA, and incoming nucleotide components) were prepared and were subjected to 100 ns of unrestrained MD simulations ([Figures 4](#) and [S9](#), [Supporting Information](#)). Complexes were generated only for the insertion stage of dCTP. Trajectory analysis of each of these complexes illuminated the key interactions between the DNA and Pol IV at the active site.

Stability of the DNA–Pol IV Ternary Complexes. To confirm the adequate sampling and the stability of DNA-Pol IV ternary complexes, root-mean-square deviations (RMSD) of backbone atoms of DNA and protein were calculated separately with respect to the first frame. Protein backbone showed marginal deviations with an average RMSD value <1.8 Å for all of the complexes ([Figure S10](#), [Supporting](#)

Information). The DNA backbone showed considerable fluctuations with an average RMSD value >2.0 Å. In order to further investigate such large fluctuations in the DNA backbone, root-mean-square fluctuation (RMSF) values were calculated for DNA with respect to the first frame. Only a few nucleotides which do not interact with protein showed larger fluctuations, whereas the nucleotides around the active site showed fewer fluctuations (Figure S11, Supporting Information). RMSF values of protein residues showed only nominal fluctuations in all of the complexes (Figure S12, Supporting Information). Also, the distance between the α phosphate of dCTP and the O atom of 3'-OH of the primer terminal dG is <4.0 Å in all of the complexes, indicative of a productive conformation for the nucleotide incorporation at the insertion stage. The incoming nucleotide dCTP forms Watson–Crick (W–C) H-bonding with dG at the active site. These W–C H-bonds were monitored through the course of dynamics, and the percentage occupancy of each of H-bond was calculated (Table 2). The H-bond (a) formed by dG:N²-

Table 2. Percentage Occupancies of Each H-Bond at the Active Site of *E. coli* Pol IV^a



complex with adduct	a	b	c
N ² -Bn-dG	92.08%	95.26%	93.01%
N ² -Naph-dG	89.88%	95.59%	91.77%
N ² -Anth-dG	84.58%	88.16%	93.04%
N ² -Pyre-dG	88.18%	95.94%	92.27%

^aThe dotted lines indicate the H-bond present between the dG and incoming nucleotide dCTP. Red arrow indicates the position of modification.

H ... O2:dCTP is affected to some extent in all the complexes. The interaction between the surrounding amino acids (AAs) with the N²-modifications, made the orientation of the N²-modified dG unfavorable for W–C H-bonding at certain time frames during the dynamics. The H-bond labeled (a) in Table 2 is most affected in Anth complex in which the H-bond (b) was also disturbed (Table 2).

Clustering Analysis. MD snapshots (100 ns) of each of the complexes were clustered into 5 ensembles based on hierarchical agglomerative approach.³⁷ The percentage of each ensemble out of 100 ns of dynamics is tabulated in Table 3. To identify the differences in the major ensembles, representative structures of ensemble 1 and ensemble 2 were superimposed for all of the complexes. These representative structures differed primarily in the orientations of terminal nucleotides of DNA. The similarities in the orientations of the AAs surrounding the modifications are highlighted in Figure 5. The residues surrounding N²-Anth-dG adduct showed least differences compared to other adducts. Minor differences were observed in the orientations of Ser42 and Phe76 residues in the other complexes.^{5,6}

Table 3. Percentage Occupancies of Each Ensemble out of 100 ns of Dynamics^a

complex with adduct	ensemble 1	ensemble 2	ensemble 3
N ² -Bn-dG	43.3%	35.4%	14.1%
N ² -Naph-dG	59.9%	23.0%	7.2%
N ² -Anth-dG	56.4%	27.9%	8.2%
N ² -Pyre-dG	47.8%	31.8%	17.8%

^aEvery second frame was considered, and a total of 25 000 frames were considered for the calculation. Clustering was performed using hierarchical agglomerative approach, which was implemented by the CPPTRAJ³⁸ module of AMBER14³⁹

To find out the similarities between the four ternary complexes at the active site, the representative structures of ensemble 1 of all of the complexes were superimposed onto each other (Figure S13, Supporting Information). Orientations of the N²-modified moieties were found to be strikingly similar to that of the furfuryl adduct in the active site of DNA-Pol IV complex as revealed by its crystal structure (PDBID: 4Q43). This implies that the Pol IV orients N²-modified moieties into a particular conformation to facilitate TLS. To validate this, the dihedral angle labeled ξ (ξ) (Figure 6) was monitored through the course of dynamics. Bn, Naph, and Pyre complexes appeared to follow a pattern, but the Anth complex showed slight deviations. This difference might be a result of the unique position of linkage of the anthracene moiety to the nucleobase (Figure 1). The average dihedral angle values of 103.18°, 107.58°, 148.13°, and 96.31° were observed for Bn, Naph, Anth, and Pyre complexes, respectively. Based on the above results, few important AA residues were identified, and the trajectory was further analyzed to understand the interaction patterns between these adducts and AAs and also to probe the nature of the forces responsible for the accommodation of adducts in the active site.

Key Noncovalent Interactions and Interaction Energies. Average interaction energies of the N²-dG adducts with the surrounding AA residues were calculated to identify the nature of interactions in the active site. Since our focus is the modified nucleobases, only residues in close proximity to these modifications (within 6 Å) were chosen for the calculations. Interaction energies were calculated using the MM/GBSA method implemented in AmberTools package. The total average energy was calculated by summation of van der Waals, electrostatic, polar solvation, and nonpolar solvation values (internal = 0 for reference). The average energy values presented in Table S1, Supporting Information, helped to qualitatively identify the chemical nature of these interactions. The results suggest that Ser42 interacts with the dG adducts primarily through van der Waals interactions.³³ Upon visual inspection of trajectories, Ser42 was found to interact with π cloud of aromatic moiety through C–H... π interactions.⁴⁰ To verify this, the distance between C ^{β} of Ser42 and the centroid of the N²-modified aromatic moiety was plotted for all of the complexes (Figure S14, Supporting Information). The distances were in the appreciable range, i.e., <4 Å in all of the complexes, favorable for van der Waals interactions. In fact, there is a correlation between this distance and the dihedral ξ responsible for the particular orientation of the adducts (Figure S13, Supporting Information). When the distance is between 3 and 4 Å, the

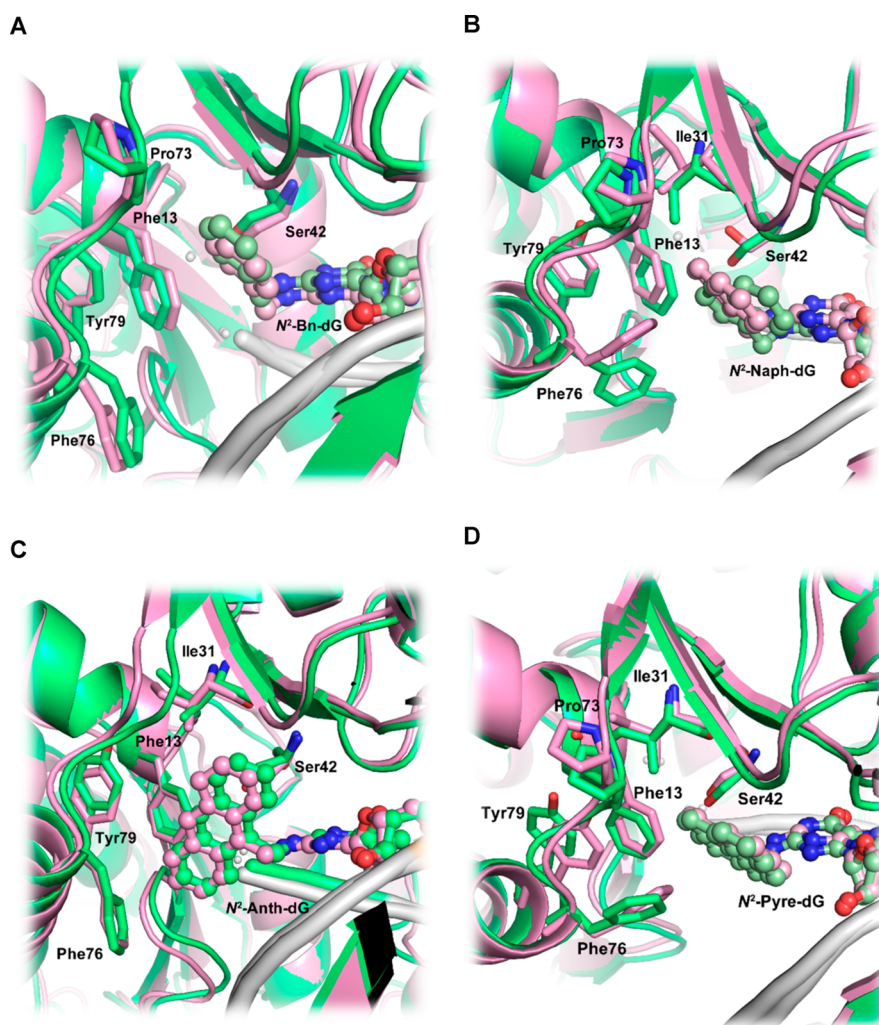


Figure 5. Superposition of representative structures of ensembles 1 and 2 of N^2 -dG-modified DNA-Pol IV ternary complexes. Only residues of interest are highlighted. (A) N^2 -Bn complex; (B) N^2 -Naph complex; (C) N^2 -Anth complex; and (D) N^2 -Pyre complex. Lime green colored cartoon representation is used for the ensemble 1 and light pink colored cartoon representation is used for the ensemble 2. DNA backbone is represented in white for both the representative structures. Pale green colored carbon framework is used for N^2 -adducts for ensemble 1 and pale pink colored carbon framework for ensemble 2. Nitrogen atoms are represented in blue and oxygen atoms in red. All of the hydrogens are removed for clarity.

orientation of aromatic moiety favors strong π -cloud interactions with the Ser42 residue.

Apart from Ser42, the other major interacting residues displayed similar noncovalent contacts. Phe76 played a crucial role in the complexes with high steric bulkiness, primarily by forming T-shaped $C-H\cdots\pi$ interactions with the adducts (Figures 5 and 7). The Phe13 also exhibited $C-H\cdots\pi$ interactions. This residue exhibited significant interactions in Anth complex compared to moderate interactions found in Bn, Anth, and Pyre complexes (Figures 5 and 7). Similarly, the $C-H$ bond of α carbon (C^α) from Gly32 showed $C-H\cdots\pi$ interactions in Naph and Pyre (Figure 7B,D) complexes.⁴⁰ The role of Tyr79 was found to be strategic as it interacts with Phe13 through parallel displaced stacking and with Phe76 through T-shaped interactions, which add strong hydrophobicity to the surroundings (Figures 5 and 7). Ile 31 and Pro73 were also found to form moderate to weak interactions in addition to weak interactions from Gly33 and Gly74 in all of the complexes. The RMSF values for these residues in each of the complexes are reported in Table S2.

The RMSF values were mostly $<1 \text{ \AA}$, indicating the importance of these residues in accommodating the adducts.

The representative structures of other ensembles of each complex had similar interaction patterns as discussed above. Overall, the van der Waals interactions are the major forces responsible for the accommodation of N^2 -modified moieties in the active site of Pol IV. The residues Phe13, Ile31, Gly32, Gly33, Ser42, Pro73, Gly74, Phe76, and Tyr79 were found to form a hydrophobic pocket⁶ into which the adducts are accommodated and our current MD results support the same. The role of Ser42 was elucidated before,^{6,41} and our results from the MD simulations underscore the importance of this residue in stabilizing nonpolar adducts through van der Waals interactions. General interaction patterns from Phe13, Phe76, Tyr79, and Gly32 were also identified. In summary, our MD studies provide insights on the molecular interactions that are responsible for the accommodation of N^2 -dG adducts and their bypass by *E. coli* Pol IV, which are in line with the results emerged from the experiments.

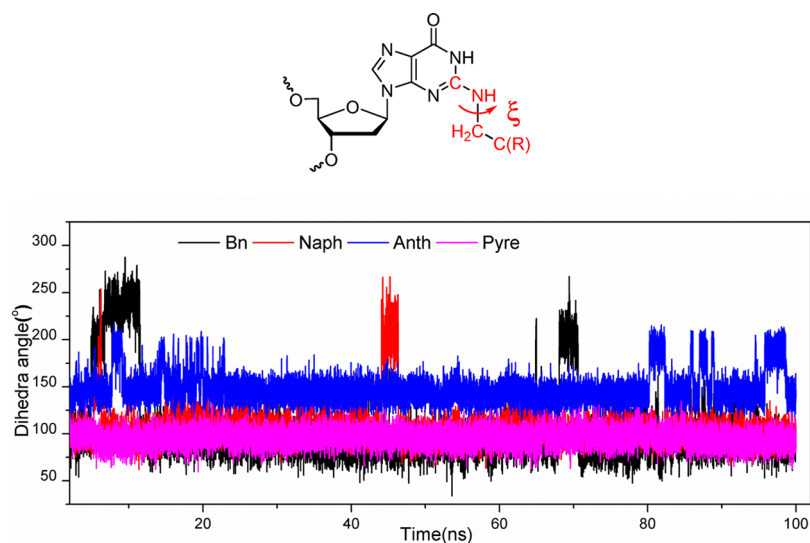


Figure 6. Plot of dihedral ξ (ξ , shown above using red arrow) vs time for all the complexes. Black, red, blue, and pink colors were used to represent N^2 -Bn, -Naph, -Anth, and -Pyre complexes, respectively.

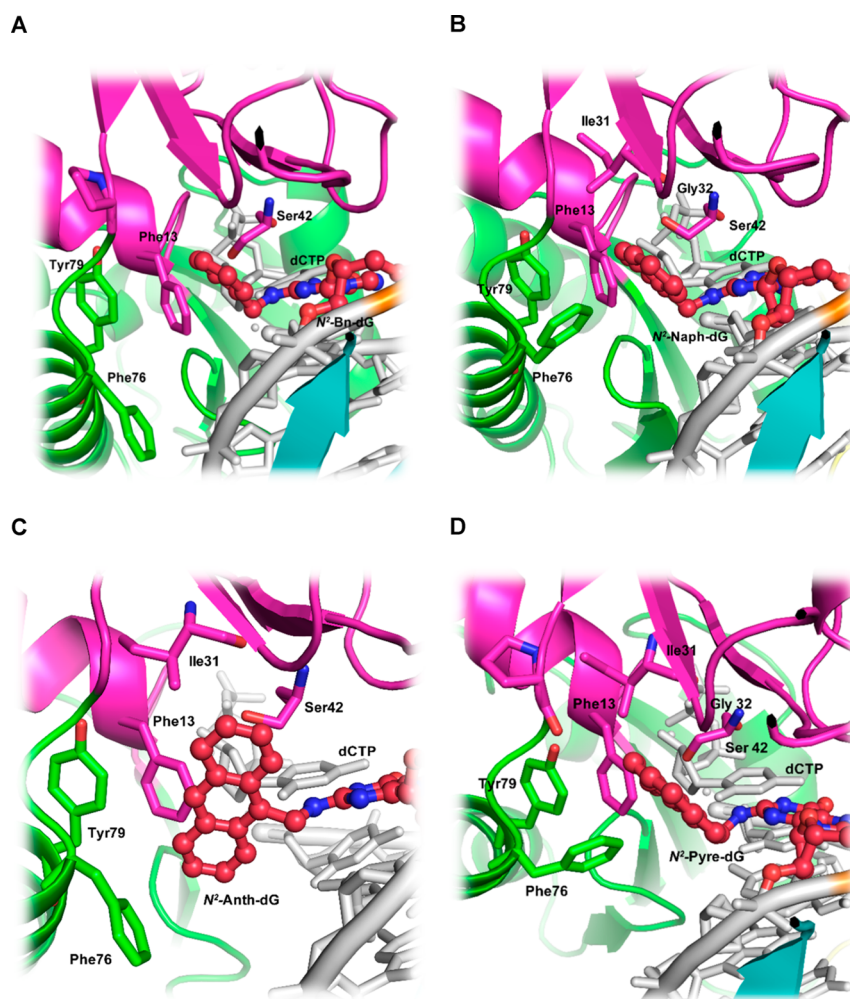


Figure 7. van der Waals interactions between AAs of Pol IV and N^2 -adducts found in DNA-Pol IV ternary complexes. (A) N^2 -Bn complex; (B) N^2 -Naph complex; (C) N^2 -Anth complex; and (D) N^2 -Pyre complex. Phe76 interacts with the N^2 -adduct through T-shaped C–H $\cdots\pi$ interactions in the Bn, Naph, Anth, and Pyre complexes. Carbon atom linked to O–H of Ser42 interacts with π cloud of Naph, Anth, and Pyre moieties. The backbone atom of Gly32 as well interacts with adducts through C–H $\cdots\pi$ interactions in Naph and Pyre complexes. Phe13 as well forms T-shaped interactions with Anth moiety.

SUMMARY AND CONCLUSIONS

In conclusion, we have developed a robust and reliable protocol for the synthesis of N^2 -aryl-dG modified DNAs. These modified DNAs were utilized to investigate the effect of steric bulkiness of the N^2 -dG adducts on the TLS by Pol IV from *E. coli*. The Pol IV was found to be efficient in bypassing these adducts with high fidelity but with low processivity. Molecular modeling and dynamics studies were also carried out by incorporating these N^2 -aryl-dG DNA adducts at the active site of *E. coli* Pol IV. The results revealed a common mechanism used by the enzyme that involves using its hydrophobic pocket at the active site to accommodate these modifications resulting in the efficient bypass of adducts. Further structural studies to validate these observations are currently being pursued.

EXPERIMENTAL SECTION

All chemicals and dry solvents were obtained from commercial sources and used without any further purification. Toluene, ACN, DCM, Et₃N, DIPEA, dioxane, and pyridine were dried using calcium hydride. Thin layer chromatography (TLC) was performed on silica gel plates precoated with fluorescent indicator with visualization by UV light (at 260 nm) or by dipping into a solution of 5% conc. H₂SO₄ in EtOH (v/v) and heating. Silica gel (100–200 mesh) was used for column chromatography. ¹H NMR (500 or 400 MHz), ¹³C NMR (125 or 100 MHz), and ³¹P NMR (202 or 162 MHz) were recorded on 500 or 400 MHz instruments. Chemical shifts in parts per million (δ) are reported downfield from TMS (0 ppm) and referenced to the TMS signal or residual proton signal of the deuterated solvent as follows: CD₃OD (3.31 ppm) for ¹H NMR spectra and CDCl₃ (77.2 ppm) or CD₃OD (49.1 ppm) for ¹³C NMR spectra. Multiplicities of ¹H NMR spin couplings are reported as s for singlet, br s for broad singlet, d for doublet, t for triplet, br t for broad triplet, dd for doublet of doublet, ddd for doublet of doublets, quint for quintet or m for multiplet and overlapping spin systems. Values for apparent coupling constants (J) are reported in Hz. High resolution mass spectra (HRMS) were obtained from Q-TOF analyzer in positive ion electrospray ionization (ESI) mode. All of the modified DNA sequences were synthesized using an automated solid phase synthesizer. Mass spectra of DNA oligonucleotides were obtained by MALDI (in positive linear or reflectron mode) or ESI (in negative mode). *E. coli* Pol IV was expressed and purified in house.⁴¹

General Procedure for the Buchwald-Hartwig Coupling and Deacetylation. In a screw capped tube, Pd(OAc)₂ (0.1 equiv) and (R)-BINAP (0.3 equiv) were taken and flushed with N₂ gas followed by addition of dry toluene (9.9 mL/mmol). Reaction mixture was stirred at room temperature for 5 min. Cs₂CO₃ (1.4 equiv), respective amines (1.1 equiv), and bromo nucleoside **1** (1 equiv) were added in the sequential manner. Further, the tube was purged with N₂ gas, sealed with a Teflon-lined cap, and heated in an oil bath at 85 °C. After completion of reaction, the reaction mixture was passed through a Celite pad and washed with EtOAc (250 mL). The filtrate was concentrated to give a crude B–H coupled product. Further, the crude compound was subjected directly to deacetylation using 33% MeNH₂ in EtOH (v/v, 22 mL/mmol) and stirred at room temperature. After completion of the reaction, the mixture was concentrated under reduced pressure. The oily residue was purified by silica gel column chromatography to obtain deprotected nucleosides.

N^2 -Benzyl- O^6 -(2-(4-nitrophenyl)ethyl)-2'-deoxyguanosine (3a). Pd(OAc)₂ (5.94 mg, 0.02 mmol), (R)-BINAP (50 mg, 0.07 mmol), dry toluene (3 mL), Cs₂CO₃ (131 mg, 0.37 mmol), benzylamine (0.03 mL, 0.29 mmol), and bromo nucleoside **1** (150 mg, 0.26 mmol) were used and heated for 18 h to give a brown crude compound **2a** (156 mg, 0.26 mmol), which was further dissolved in 33% MeNH₂ in EtOH (5 mL) and stirred at room temperature for 3 h. Silica gel column chromatography (4% MeOH

in DCM) yielded slightly brownish solid **3a** (76 mg, 57% after two steps). R_f = 0.47 (10% MeOH in DCM); mp: 60–63 °C; ¹H NMR (400 MHz, CDCl₃): δ 8.12–8.07 (m, 2H), 7.61 (s, 1H), 7.33 (d, J = 8.6 Hz, 2H), 7.31–7.21 (m, 5H), 6.19 (dd, J = 9.4, 5.6 Hz, 1H), 5.58 (br t, 1H), 4.70 (d, J = 5 Hz, 1H), 4.64 (t, J = 6.9 Hz, 2H), 4.60 (dd, J = 5.9, 2.4 Hz, 2H), 4.15 (s, 1H), 3.92 (dd, J = 12.6, 1.8 Hz, 1H), 3.71 (d, J = 11.3 Hz, 1H), 3.16 (t, J = 6.7 Hz, 2H), 2.97 (ddd, J = 14.2, 9.1, 5.1 Hz, 1H), 2.23 (dd, J = 12.6, 5.6 Hz, 1H); ¹³C{¹H} NMR (100 MHz, CDCl₃): δ 160.1, 158.5, 152.8, 146.9, 145.9, 139.3, 139.1, 130, 128.7, 127.3, 127.3, 123.8, 116.5, 89.1, 87.3, 73.3, 66.1, 63.5, 46.1, 40.2, 35.1; HRMS (ESI): Calcd for C₂₅H₂₇N₆O₆, [M + H]⁺ 507.1992; found, [M + H]⁺ 507.1980 (Δm = 0.0012, error –2.4 ppm).

N^2 -(1-Naphthylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-2'-deoxyguanosine (3b). The Pd(OAc)₂ (11 mg, 0.04 mmol), (R)-BINAP (88 mg, 0.14 mmol), dry toluene (7 mL), Cs₂CO₃ (214 mg, 0.65 mmol), 1-naphthylmethylamine (0.07 mL, 0.51 mmol), and bromo nucleoside **1** (269 mg, 0.47 mmol) were used and heated for 21 h to give a dark brown crude compound **2b** (260 mg, 0.40 mmol), which was further dissolved in 33% MeNH₂ in EtOH (9 mL) and stirred at room temperature for 2.5 h. Silica gel column chromatography (4% MeOH in DCM) yielded yellow solid **3b** (171 mg, 64%, after two steps). R_f = 0.49 (10% MeOH in DCM); mp: 67–70 °C; ¹H NMR (400 MHz, CDCl₃): δ 8.09–8.02 (m, 1H), 7.97 (d, J = 8.3 Hz, 2H), 7.92–7.85 (m, 1H), 7.77 (s, 1H), 7.60 (s, 1H), 7.57–7.36 (m, 6H), 7.16 (br s, 1H), 6.18 (dd, J = 9.2, 5.6 Hz, 1H), 5.51 (br s, 1H), 5.14–4.99 (m, 2H), 4.67 (d, J = 5.3 Hz, 1H), 4.60 (t, J = 7.0 Hz, 2H), 4.33 (s, 1H), 4.13 (s, 1H), 3.90 (d, J = 12 Hz, 1H), 3.68 (dd, J = 12.5, 1.8 Hz, 1H), 3.07 (t, J = 6.7 Hz, 2H), 3.00 (ddd, J = 14, 9.5, 5.5 Hz, 1H), 2.22 (dd, J = 13.4, 5.6 Hz, 1H); ¹³C{¹H} NMR (100 MHz, CDCl₃): δ 161, 158.4, 153, 146.8, 145.6, 139.2, 134.2, 133.9, 131.3, 129.8, 129, 128.1, 126.4, 126, 125.5, 125.1, 123.7, 123.1, 116.4, 89, 87, 73, 66.1, 63.3, 44, 40.1, 35; HRMS (ESI): Calcd for C₂₉H₂₉N₆O₆, [M + H]⁺ 557.2149; found, [M + H]⁺ 557.2156 (Δm + 0.0007, error +1.3 ppm).

N^2 -(9-Anthracenylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-2'-deoxyguanosine (3c). Pd(OAc)₂ (3.9 mg, 0.01 mmol), (R)-BINAP (33 mg, 0.05 mmol), dry toluene (2 mL), Cs₂CO₃ (87 mg, 0.24 mmol), 9-(aminomethyl)-anthracene²⁹ (41 mg, 0.19 mmol), and bromo nucleoside **1** (100 mg, 0.17 mmol) were used and heated for 8 h to give a dark brown crude compound **2c** (127 mg, 0.18 mmol), which was further dissolved in 33% MeNH₂ in EtOH (3.7 mL) and stirred at room temperature for 3 h. Column chromatography (3% MeOH in DCM) yielded slightly greenish compound **3c** (59 mg, 54% after two steps). R_f = 0.50 (10% MeOH in DCM); mp: 83–86 °C; ¹H NMR (400 MHz, CDCl₃): δ 8.47 (s, 1H), 8.32–8.27 (m, 2H), 8.12 (d, J = 8.5 Hz, 2H), 8.03 (dd, J = 8.5, 1.0 Hz, 2H), 7.64 (s, 1H), 7.54–7.42 (m, 6H), 6.18 (br t, 1H), 5.57–5.47 (m, 2H), 5.18 (t, J = 4.4 Hz, 1H), 4.90–4.71 (m, 2H), 4.63 (br s, 1H), 4.07 (br s, 1H), 3.88–3.77 (m, 1H), 3.63–3.53 (m, 1H), 3.29 (br s, 2H), 3.12–2.92 (m, 1H), 2.25–2.13 (m, 1H); ¹³C{¹H} NMR (100 MHz, CDCl₃): δ 161.1, 158.3, 153.1, 146.9, 145.9, 139.2, 131.6, 130.6, 130, 129.3, 128.7, 128.2, 126.6, 125.2, 124.1, 123.8, 116.5, 88.8, 87, 73, 66.3, 63.2, 40.1, 38.9, 35.3; HRMS (ESI): Calcd for C₃₃H₃₁N₆O₆, [M + H]⁺ 607.2300; found, [M + H]⁺ 607.2297 (Δm = 0.0003, error –0.5 ppm).

N^2 -(1-Pyrenylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-2'-deoxyguanosine (3d). Pd(OAc)₂ (16 mg, 0.07 mmol), (R)-BINAP (132 mg, 0.21 mmol), dry toluene (7 mL), Cs₂CO₃ (323 mg, 0.99 mmol), 1-(aminomethyl)pyrene³⁰ (182 mg, 0.78 mmol), and bromo nucleoside **1** (403 mg, 0.71 mmol) were used and heated in an oil bath for 21 h at 85 °C to get crude compound **2d** (547 mg, 0.76 mmol), which was further dissolved in 33% MeNH₂ in EtOH (15 mL) and stirred at room temperature for 1 h 15 min. Column chromatography on a silica gel (2.5% MeOH in DCM) yielded yellow compound **3d** (202 mg, 45% after two steps). R_f = 0.49 (10% MeOH in DCM); mp: 75–78 °C; ¹H NMR (400 MHz, MeOH-*d*₄ + CDCl₃): δ 8.24 (d, J = 9.1 Hz, 1H), 8.16–8.09 (m, 2H), 8.08–7.89 (m, 7H), 7.72 (s, 1H), 7.66 (br s, 2H), 6.86 (br s,

1H), 6.22 (dd, $J = 8.9, 5.9$ Hz, 1H), 5.23 (s, 2H), 4.55 (d, $J = 5.4$ Hz, 1H), 4.38 (t, $J = 6.6$ Hz, 2H), 4.09 (d, $J = 1.3$ Hz, 1H), 3.88 (d, $J = 12.4$ Hz, 1H), 3.73–3.66 (m, 2H), 3.58–3.49 (m, 1H), 2.88 (ddd, $J = 14.5, 9.1, 5.6$ Hz, 1H), 2.30–2.20 (m, 1H); $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, MeOH- d_4 + CDCl_3): δ 160.9, 158.5, 152.9, 146.4, 145.5, 139.1, 132.4, 131.3, 130.7, 129.4, 128.4, 127.9, 127.4, 127.2, 126.2, 125.4, 125.2, 124.9, 124.8, 123.4, 122.5, 115.9, 88.8, 87, 72.5, 65.9, 63.2, 44, 40.1, 34.7; HRMS (ESI): Calcd for $\text{C}_{35}\text{H}_{31}\text{N}_6\text{O}_8$ [M + H] $^+$ 631.2300; found, [M + H] $^+$ 631.2293 ($\Delta m - 0.0007$, error -1 ppm).

General Procedure for the DMT-Protection. The N^2 -modified-dG nucleosides (1 equiv) was coevaporated with dry pyridine (5 mL) and dissolved in the same solvent (10 mL/mmol). To this DMT-Cl (1.22 equiv) was added and stirred at room temperature. Further, the reaction mixture was diluted with DCM (100 mL) and washed with saturated NaHCO_3 (40 mL) and water (2×50 mL). The DCM layer was dried over Na_2SO_4 and evaporated. The crude DMT-protected compound was purified by silica gel column chromatography (using appropriate solvents along with 2% Et_3N) to afford pure compound.

N^2 -Benzyl- O^6 -(2-(4-nitrophenyl)ethyl)-5'-(4,4'-dimethoxytrityl)-2'-deoxyguanosine (4a). Compound 3a (251 mg, 0.49 mmol) in dry pyridine (5 mL) and DMT-Cl (332 mg, 0.99 mmol) was stirred for 24 h. Column chromatography (90% DCM in pet. ether + 2% Et_3N) yielded compound 4a as a pale yellow solid (237 mg, 60%). $R_f = 0.55$ (2% MeOH in DCM + 2% Et_3N); mp: 71–74 °C; ^1H NMR (400 MHz, CDCl_3): δ 8.13–8.09 (m, 2H), 7.68 (s, 1H), 7.39 (d, $J = 7.3$ Hz, 4H), 7.31–7.26 (m, 8H), 7.25–7.14 (m, 4H), 6.78 (d, $J = 8.0$ Hz, 4H), 6.27 (t, $J = 6.7$ Hz, 1H), 5.17 (t, $J = 5.5$ Hz, 1H), 4.66 (t, $J = 6.8$ Hz, 2H), 4.63–4.58 (m, 1H), 4.53 (t, $J = 5.8$ Hz, 2H), 4.09–4.04 (m, 1H), 3.76 (s, 6H), 3.38 (dd, $J = 9.8, 4.8$ Hz, 1H), 3.34 (dd, $J = 10.1, 5.4$ Hz, 1H), 3.20 (t, $J = 6.8$ Hz, 2H), 2.87–2.81 (m, 1H), 2.38 (ddd, $J = 13.2, 6.2, 4$ Hz, 1H); $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3): δ 160.6, 158.8, 158.7, 153.8, 146.9, 146.1, 144.7, 139.6, 137.9, 135.8, 130.1, 130, 128.6, 128.2, 128, 127.4, 127.2, 127, 123.8, 115.5, 113.3, 86.6, 85.9, 84, 72.8, 66, 64.1, 55.3, 46.9, 39.5, 35.3. HRMS (ESI): Calcd for $\text{C}_{46}\text{H}_{45}\text{N}_6\text{O}_8$ [M + H] $^+$ 809.3299; found, [M + H] $^+$ 809.3329 ($\Delta m + 0.003$, error +3.7 ppm).

N^2 -(1-Naphthylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-5'-(4,4'-dimethoxytrityl)-2'-deoxyguanosine (4b). Compound 3b (171 mg, 0.30 mmol) in dry pyridine (3 mL), DMT-Cl (126 mg, 0.37 mmol) were stirred for 10 h. Column chromatography (1% MeOH in DCM + 2% Et_3N) yielded compound 4b as a yellow solid (154 mg, 60%). $R_f = 0.50$ (2% MeOH in DCM + 2% Et_3N); mp: 85–88 °C; ^1H NMR (400 MHz, CDCl_3): δ 8.08–7.98 (m, 3H), 7.91–7.86 (m, 1H), 7.77 (d, $J = 8.0$ Hz, 1H), 7.70 (s, 1H), 7.55–7.46 (m, 2H), 7.44–7.35 (m, 4H), 7.30–7.26 (m, 5H), 7.26–7.13 (m, 4H), 6.80–6.72 (m, 4H), 6.29 (t, $J = 6.6$ Hz, 1H), 5.24–5.17 (m, 1H), 5.06–4.88 (m, 2H), 4.63 (t, $J = 6.9$ Hz, 2H), 4.56 (br s, 1H), 4.08–4.01 (m, 1H), 3.74 (d, $J = 1.3$ Hz, 6H), 3.68–3.63 (m, 1H), 3.39 (dd, $J = 10, 5$ Hz, 1H), 3.33 (dd, $J = 9.8, 4.8$ Hz, 1H), 3.14 (br s, 2H), 2.85–2.74 (m, 1H), 2.38 (ddd, $J = 13.5, 6.5, 4.5$ Hz, 1H); $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3): δ 160.7, 158.7, 158.6, 146.8, 146, 144.7, 137.9, 135.8, 134.9, 134.5, 133.9, 131.4, 130.1, 129.8, 129, 128.2, 128.1, 128, 127, 126.5, 126, 125.6, 123.8, 123.4, 115.7, 113.3, 86.6, 85.8, 84, 72.7, 66, 64.1, 55.3, 44.1, 39.6, 35.2; HRMS (ESI): Calcd for $\text{C}_{50}\text{H}_{47}\text{N}_6\text{O}_8$ [M + H] $^+$ 859.3455; found, [M + H] $^+$ 859.3478 ($\Delta m + 0.0023$, error +2.7 ppm).

N^2 -(9-Anthracenylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-5'-(4,4'-dimethoxytrityl)-2'-deoxyguanosine (4c). Compound 3c (162 mg, 0.26 mmol) in dry pyridine (3 mL) and DMT-Cl (180 mg, 0.53 mmol) was stirred for 10 h. Column chromatography (1% MeOH in DCM + 2% Et_3N) yielded compound 4c as a slightly greenish solid (177 mg, 72%). $R_f = 0.50$ (2% MeOH in DCM + 2% Et_3N); mp: 86–89 °C; ^1H NMR (500 MHz, CDCl_3): δ 8.46 (s, 1H), 8.29–8.24 (m, 2H), 8.08 (d, $J = 7.9$ Hz, 2H), 8.05–8.00 (m, 2H), 7.75 (s, 1H), 7.50–7.36 (m, 8H), 7.28 (dd, $J = 5.8, 2.7$ Hz, 3H), 7.20–7.15 (m, 3H), 7.12–7.08 (m, 1H), 6.73 (d, $J = 7.3$ Hz, 4H), 6.46–6.28 (m, 1H), 5.51–5.36 (m, 2H), 5.14 (t, $J = 4.0$ Hz, 1H), 4.74–4.59

(m, 2H), 4.09 (br s, 1H), 3.70 (d, $J = 1.5$ Hz, 6H), 3.46–3.40 (m, 1H), 3.38–3.30 (m, 1H), 3.26–3.15 (br s, 2H), 2.91–2.85 (m, 1H), 2.52–2.44 (m, 1H); $^{13}\text{C}\{^1\text{H}\}$ NMR (125 MHz, CDCl_3): δ 160.7, 158.6, 153.9, 146.8, 146.1, 144.6, 137.7, 135.8, 135.7, 131.6, 130.5, 130.1, 130, 130, 129.4, 129.3, 129.1, 128.3, 128.2, 128, 127, 126.5, 125.2, 124.2, 123.8, 115.5, 113.2, 86.6, 85.9, 83.9, 72.8, 66, 64.1, 55.3, 45.9, 38.7, 35.3; HRMS (ESI): Calcd for $\text{C}_{54}\text{H}_{49}\text{N}_6\text{O}_8$ [M + H] $^+$ 909.3606; found, [M + H] $^+$ 909.3608 ($\Delta m + 0.0002$, error +0.1 ppm).

N^2 -(1-Pyrenylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-5'-(4,4'-dimethoxytrityl)-2'-deoxyguanosine (4d). Compound 3d (151 mg, 0.24 mmol) in dry pyridine (3 mL) and DMT-Cl (102 mg, 0.30 mmol) was stirred for 12 h. Column chromatography (1% MeOH in DCM + 2% Et_3N) yielded compound 4d as a pale yellow solid (178 mg, 76%). $R_f = 0.55$ (2% MeOH in DCM + 2% Et_3N); mp: 77–80 °C; ^1H NMR (400 MHz, CDCl_3): δ 8.25 (d, $J = 9.3$ Hz, 1H), 8.22–8.16 (m, 2H), 8.10–7.86 (m, 7H), 7.71 (s, 1H), 7.42–7.36 (m, 2H), 7.32–7.26 (m, 5H), 7.25–7.10 (m, 5H), 6.78–6.72 (m, 4H), 6.30 (t, $J = 6.5$ Hz, 1H), 5.37–5.32 (m, 1H), 5.27–5.13 (m, 2H), 4.64–4.50 (m, 3H), 4.09–4.03 (m, 1H), 3.70 (d, $J = 2.5$ Hz, 6H), 3.39 (dd, $J = 9.9, 4.6$ Hz, 1H), 3.35 (dd, $J = 10, 5.2$ Hz, 1H), 3.04 (br s, 2H), 2.87–2.77 (m, 1H), 2.37 (ddd, $J = 13.4, 6.4, 4.0$ Hz, 1H); $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3): δ 160.7, 158.7, 158.6, 153.9, 146.6, 145.9, 144.7, 137.9, 135.8, 132.4, 131.4, 130.8, 130.1, 129.7, 128.2, 128, 127.5, 127.4, 127, 126.2, 125.5, 125.3, 125, 124.9, 124.8, 123.6, 122.8, 115.7, 113.3, 86.6, 85.9, 84, 72.8, 65.9, 64.1, 55.3, 44.3, 39.6, 35.1; HRMS (ESI): Calcd for $\text{C}_{56}\text{H}_{48}\text{N}_6\text{NaO}_8$ [M + Na] $^+$ 955.3426; found, [M + Na] $^+$ 955.3431 ($\Delta m + 0.0005$, error +0.6 ppm).

General Procedure for the Phosphitylation. DMT protected nucleoside (1 equiv) was dissolved in DCM (10 mL/mmol) followed by addition of DIPEA (8 equiv) and CEP-Cl (2 equiv). The reaction mixture was stirred at room temperature under N_2 atmosphere. After completion, the reaction was quenched by addition of a trace amount of MeOH and stirred for 5 to 10 min. Further, the reaction mixture was diluted with DCM (100 mL) and washed with NaHCO_3 (3×20 mL), dried over Na_2SO_4 and evaporated under reduced pressure. Crude compound was purified by silica gel column chromatography (using appropriate solvents along with 2% Et_3N) to obtain phosphitylated compound.

N^2 -Benzyl- O^6 -(2-(4-nitrophenyl)ethyl)-5'- O -(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl diisopropylphosphoramidite)-2'-deoxyguanosine (5a). Nucleoside 4a (150 mg, 0.18 mmol), DCM (2 mL), DIPEA (0.25 mL, 1.48 mmol), and CEP-Cl (127 mg, 0.54 mmol) were used and stirred for 2 h. Further, MeOH (0.5 mL) was added and stirred for an additional 10 min. Column chromatography (35% DCM in pet. ether + 2% Et_3N) yielded compound 5a as a pale yellow solid (111 mg, 61%). $R_f = 0.65$ (90% DCM in pet. ether + 2% Et_3N); mp: 58–61 °C; ^{31}P NMR (162 MHz, CDCl_3): δ 148.81, 148.63; HRMS (ESI): Calcd for $\text{C}_{55}\text{H}_{62}\text{N}_8\text{O}_9\text{P}$ [M + H] $^+$ 1009.4377; found, [M + H] $^+$ 1009.4348 ($\Delta m - 0.0029$, error -3 ppm).

N^2 -(1-Naphthylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-5'- O -(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl diisopropylphosphoramidite)-2'-deoxyguanosine (5b). The tritylated nucleoside 4b (150 mg, 0.17 mmol), DCM (2 mL), DIPEA (0.25 mL, 1.36 mmol), and CEP-Cl (83 mg, 0.34 mmol) were used and stirred for 2 h 15 min. Further, MeOH (0.5 mL) was added and stirred for an additional 10 min. Column chromatography (35% DCM in pet. ether + 2% Et_3N) yielded compound 5b as a pale yellow solid (74 mg, 41%). $R_f = 0.65$ (75% DCM in pet. ether + 2% Et_3N); mp: 70–73 °C; ^{31}P NMR (162 MHz, CDCl_3): δ 148.72, 148.57; HRMS (ESI): Calcd for $\text{C}_{59}\text{H}_{64}\text{N}_8\text{O}_9\text{P}$ [M + H] $^+$ 1059.4528; found, [M + H] $^+$ 1059.4529 ($\Delta m + 0.00009$, error +0.1 ppm).

N^2 -(9-Anthracenylmethyl)- O^6 -(2-(4-nitrophenyl)ethyl)-5'- O -(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl diisopropylphosphoramidite)-2'-deoxyguanosine (5c). Nucleoside 4c (170 mg, 0.18 mmol), DCM (2 mL), DIPEA (0.26 mL, 1.49 mmol), and CEP-Cl (88 mg, 0.37 mmol) were used and stirred for 1 h. Further, MeOH (0.5 mL) was added and stirred for an additional 10 min. Column chromatog-

raphy (40% DCM in pet. ether + 2% Et₃N) yielded compound **5c** as a slightly greenish solid (75 mg, 37%). $R_f = 0.60$ (75% DCM in pet. ether + 2% Et₃N); mp: 77–80 °C; ³¹P NMR (202 MHz, CDCl₃): δ 148.85, 148.72; HRMS (ESI): Calcd for C₆₃H₆₆N₈O₉P [M + H]⁺ 1109.4685; found, [M + H]⁺ 1109.4686 ($\Delta m + 0.00009$, error +0.1 ppm).

*N*²-(1-Pyrenylmethyl)-O⁶-(2-(4-nitrophenyl)ethyl)5'-O-(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl)diisopropylphosphoramidite)-2'-deoxyguanosine (**5d**). Nucleoside **4d** (200 mg, 0.21 mmol), DCM (2.5 mL), DIPEA (0.3 mL, 1.68 mmol), and CEP-Cl (100 mg, 0.42 mmol) were used and stirred for 1 h 15 min. Further, MeOH (0.5 mL) was added and stirred for an additional 10 min. Column chromatography (45% DCM in pet. ether + 2% Et₃N) yielded compound **5d** as a pale yellow solid (93 mg, 40%). $R_f = 0.66$ (75% DCM in pet. ether + 2% Et₃N); mp: 80–83 °C; ³¹P NMR (202 MHz, CDCl₃): δ 148.72; HRMS (ESI): Calcd for C₆₅H₆₆N₈O₉P, [M + H]⁺ 1133.4685; found, [M + H]⁺ 1133.4690 ($\Delta m + 0.0005$, error +0.4 ppm).

Solid Phase Synthesis, Deprotection, and Purification of N²-dG Modified Oligonucleotides. All DNA sequences (Table 1) were synthesized on 1 μ mol scale using appropriate CPG solid supports. Unmodified and modified phosphoramidite building blocks were dried under P₂O₅ for 24 h and then dissolved in anhydrous ACN and stored at 4 °C under molecular sieves (4 Å) for 10 h. Concentration used for the unmodified phosphoramidite was 0.067 M and for modified phosphoramidite was 0.1 M. Coupling time used for unmodified phosphoramidite was 2 min and for the N²-dG modified phosphoramidites was 6 min. After solid phase synthesis, deprotection of N²-Bn, N²-Naph, N²-Anth, and N²-Pyre-dG modified oligos were carried out in four different steps. Initially, the deprotection of the cyanoethyl group of the modified DNA sequences were carried out using 10% diethylamine in ACN (800 μ L, v/v) for 5 min. Supernatant layer was collected, and solid support was washed with ACN (2 \times 400 μ L). The CPG beads were air-dried (3 h) and treated with 1 M DBU in ACN (1 mL, v/v) for 1 h at room temperature to remove the NPE group.²³ After 1 h, the supernatant was removed and the CPG beads were washed with MeOH (2 \times 1 mL) and ACN (3 \times 1 mL). Furthermore, the modified DNAs intact with CPG was treated with 1 mL of 30% aq. NH₃ (v/v) for 3 h for the cleavage of solid support. To remove the base protecting groups, the supernatant solution was evaporated and again resuspended in 1 mL of 30% aq. NH₃ (v/v) and heated at 55 °C for 16 h. Further, the ammonia layer was evaporated to obtain the pellet of crude DNAs, which was purified by using 20% (7 M urea) denaturing PAGE (30 W, 3 h) with 1 X TBE running buffer (89 mM each Tris and boric acid and 2 mM EDTA, pH ~8.3). The gel thickness was 1 mm, and the gel dimension was 20 \times 30 cm. The desired oligo bands were visualized and marked under the UV lamp (260 nm). Further, the gel bands were cut and crushed into fine particles. The crushed gels were extracted three times with 15 mL of TEN buffer (10 mM Tris, 1 mM EDTA, 300 mM NaCl, pH ~8.0) at 37 °C. Finally, desalting of oligos were carried out using Sep-Pak column and the concentrations of N²-dG modified DNAs were measured at 260 nm in UV-vis spectrophotometer using the appropriate molar extinction coefficients (ϵ).

5'-Radiolabeling of Primers. The radiolabeling reaction mixture contains the DNA (25 pmol), T4 polynucleotide kinase (PNK) enzyme (5 U), and [γ -³²P] ATP (20 μ Ci) in 1 X PNK buffer (50 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, pH 7.6). The total reaction volume was 10 μ L. The reaction mixture was incubated at 37 °C for 1 h followed by deactivation of PNK enzyme by heating at 70 °C for 3 min. The radiolabeled DNAs (primers and 50-mer DNA standard) were purified using nucleotide removal kit using the protocol provided by the manufacturer.

Primer Extension Assay. The primer extension reactions were carried out using the mixture of 1:1.2 molar ratios of the primers with the unmodified and N²-dG modified DNA templates. Initially, the primer-template mixture was annealed by heating at 95 °C for 3 min, followed by slow cooling to room temperature for 30 min.

After annealing, the primer extension reactions were performed with primer–template duplex in a total volume of 20 μ L, containing 40 nM primer, traces of 5'-radiolabeled primer, 50 nM template, 50 μ g/mL of BSA, 0.1 mM of ammonium sulfate, 2.5 mM MgCl₂, 4 μ L of 5X assay buffer (125 mM Tris-Cl, 5 mM dithiothreitol, pH 8.0), Pol IV (4 nM), and individual dNTP or mixture of dNTPs (10 μ M). The primer extension reactions were incubated at 37 °C for 5 min, and the reactions were initiated by the addition of the desired amount of dNTPs. The reactions were conducted for the indicated amount of time, and aliquots were withdrawn and quenched by the addition of stop solution (80% formamide, 0.025% each bromophenol blue, xylene cyanol and 50 mM EDTA). The extended products were run on 20% denaturing PAGE (7 M urea), autoradiogram was generated, and the bands were quantified by ImageQuantTL software. The running start experiments were carried out using 15-mer and 11-mer primers employing a mixture of dNTPs (250 μ M). The reaction conditions were the same as described previously. The total reaction volume contains 40 nM primer, traces of 5'-radiolabeled primer, and 50 nM of each template. Finally, the reactions were terminated and analyzed as mentioned previously.

Molecular Modeling and Dynamics Studies. The modified nucleotides are constructed using GaussView and submitted to RED Server Development^{42–45} for the calculation of restraint electrostatic potential (RESP) charges and parameters (bond, angle, and dihedral) required for the simulation (Figure S15–S18, Supporting Information). The originally reported crystal structure (PDB ID: 4Q43) contains chemically modified dCTP at the active site which was required to prevent catalysis during crystallization. The modified dCTP is replaced with normal dCTP, and previously calculated bond, angle, and dihedral parameters were used (<http://upiv.q4md-forcefieldtools.org/REDDB/projects/F-90/>). The amber force fields ff99SB⁴⁶ and parmbsc0⁴⁷ were used for protein and DNA, respectively. The parameters for dCTP and modified nucleotides along with parameters for protein, DNA, and ions were fed to Leap program of AmberTools package to prepare input files. The Mg²⁺ ions reported in the original crystal structure were retained. Additional potassium ions were added to neutralize the system. Each system was then immersed in TIP3P water box with water molecules extending up to 10 Å. Then each complex was minimized in two stages to remove any atomic clashes.

Energy minimization was carried out using SANDER module of Amber 14. In the first step, the N²-dG adduct was left free and minimized for 10 000 steepest descent energy cycles, whereas in the second step the N²-dG adduct was restricted and minimized for 10 000 steepest descent energy minimizations. Later the system was heated from 0.1 K to close to 300 K over 100 ps with constant volume. Then the complexes were heated to final temperature (300 K) at constant pressure. During the heating, the DNA was restricted with a restraint of 10 kcal/mol Å². This restriction was slowly removed over 8 stages of equilibration (with restraints of 10, 8, 5, 4, 3, 2, 1, and 0.5 kcal/mol Å² 30 ps each). Each equilibration was carried under constant pressure and temperature (NPT) conditions. The pressure was maintained at 1 atm using Benderson weak-coupling barostat with a pressure relaxation time of 2 ps. The temperature was maintained at 300 K using Langevin dynamics. The equilibration was followed by energy minimization (10 000 steepest descent cycles) and the minimized structure was used as starting structure for the subsequent equilibration. After 8 stages, 100 ps of equilibration was run under NPT conditions without any restrictions before proceeding to the production run. 100 ns of unrestrained dynamics (under NPT conditions) was performed on each complex using GPU accelerated version of Particle Mesh Ewald Molecular Dynamics (PMEMD) implemented in Amber 14.^{39,48–50} A cutoff value of 10 Å is set for calculations of nonbonded interactions by Particle Mesh Ewald (PME) method. The SHAKE algorithm was used to constraint all of the bonds involving hydrogen atoms with an integration time step of 2 fs. The output coordinate files were saved for every 2 ps and a total of 50 000 snapshots (100 ns) were recorded for each complex.

The trajectories were visualized using UCSF Chimera.⁵¹ CPPTRAJ module of Amber 14 was used for calculations of dihedrals, RMSD values, RMSF values, H-bond occupancies, and clustering of dynamics. Only heavy atoms were considered for the calculation of RMSD values. Distances between Ser and N²-adducts were calculated using per frame analysis in UCSF Chimera. A distance of 3.0 Å and angle of 150° were used as cutoff values for H-bond calculations. A cut-off distance of 4.0 Å is used for van der Waals interactions. Clustering analysis was carried out employing hierarchical agglomerative approach using mass weighed RMSD. Interaction energies are calculated using MM/GB-SA⁵² method implemented as the program MMPBSA.py⁵³ in AmberTools package. Images are rendered using PyMol (www.pymol.org).

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.joc.8b02082](https://doi.org/10.1021/acs.joc.8b02082).

PAGEs of nucleotide incorporation and full length extension assays; additional figures and tables from molecular modeling and dynamics studies; ¹H, ¹³C, and ³¹P NMR of all of the new compounds; and representative MALDI spectra of N²-dG modified oligonucleotides (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: deepak@rcb.res.in.

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

ORCID

Pratibha P. Ghodke: 0000-0002-8402-3933

Deepak T. Nair: 0000-0002-0677-9444

P. I. Pradeepkumar: 0000-0001-9104-3708

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is financially supported by grants from Department of Biotechnology (DBT)-Government of India (Grant No. BT/PR8265/BRB/10/1228/2013). We are thankful to Dr. Claudia Höbartner (MPIIbc-Göttingen), Prof. K. V. R. Chary, and Ms. Gitanjali A. Dhotre (TIFR-Mumbai) and the central facility supported by IRCC-IIT Bombay for providing ESI/MALDI spectra of the modified DNAs. We also thank Dr. S. Harikrishna for his suggestions on MD studies and Saurja Dasgupta for critically reading the manuscript. Computer Centre, IIT Bombay is gratefully acknowledged for providing high performance computing facilities. P.P.G. thanks Ms. Sushree P. Pany for acquiring the mass spectra of phosphoramidites. P.P.G. also thanks the Council of Scientific and Industrial Research (CSIR) for the Ph.D. fellowship, and P.B. thanks Department of Science and Technology (DST) for the INSPIRE Fellowship.

■ DEDICATION

Dedicated to Professor Krishna N. Ganesh on the occasion of his 65th birthday.

■ REFERENCES

(1) Moorthy, B.; Chu, C.; Carlin, D. J. Polycyclic Aromatic Hydrocarbons: From Metabolism to Lung Cancer. *Toxicol. Sci.* **2015**, *145*, 5–15.

(2) Dipple, A. DNA Adducts of Chemical Carcinogens. *Carcinogenesis* **1995**, *16*, 437–441.

(3) Hemminki, K. Nucleic Acid Adducts of Chemical Carcinogens and Mutagens. *Arch. Toxicol.* **1983**, *52*, 249–285.

(4) Sale, J. E.; Lehmann, A. R.; Woodgate, R. Y-Family DNA Polymerases and their Role in Tolerance of Cellular DNA Damage. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 141–152.

(5) Jarosz, D. F.; Godoy, V. G.; Delaney, J. C.; Essigmann, J. M.; Walker, G. C. A Single Amino Acid Governs Enhanced Activity of DinB DNA Polymerases on Damaged Templates. *Nature* **2006**, *439*, 225–228.

(6) Kottur, J.; Sharma, A.; Gore, K. R.; Narayanan, N.; Samanta, B.; Pradeepkumar, P. I.; Nair, D. T. Unique Structural Features in DNA Polymerase IV Enable Efficient Bypass of the N2 Adduct Induced by the Nitrofurazone Antibiotic. *Structure* **2015**, *23*, 56–67.

(7) Ikeda, M.; Furukohri, A.; Philippin, G.; Loechler, E.; Akiyama, M. T.; Katayama, T.; Fuchs, R. P.; Maki, H. DNA Polymerase IV Mediates Efficient and Quick Recovery of Replication Forks Stalled at N2-dG Adducts. *Nucleic Acids Res.* **2014**, *42*, 8461–8472.

(8) Moschel, R. C.; Hudgins, W. R.; Dipple, A. Alkylation of Guanosine by the Carcinogen N-Nitroso-N-Benzylurea. *J. Org. Chem.* **1980**, *45*, 533–535.

(9) Peterson, L. A. N-Nitrosobenzylmethylamine Is Activated to a DNA Benzylating Agent in Rats. *Chem. Res. Toxicol.* **1997**, *10*, 19–26.

(10) Moon, K.-Y.; Kim, Y. S. Synthesis and Characterization of Oligonucleotides Containing Site-Specific Bulky N2-Aralkylated Guanines and N6-Aralkylated Adenines. *Arch. Pharmacol. Res.* **2000**, *23*, 139–146.

(11) Kou, Y.; Koag, M.-C.; Lee, S. Structural and Kinetic Studies of the Effect of Guanine-N7 Alkylation and Metal Cofactors on DNA Replication. *Biochemistry* **2018**, *57*, 5105–5116.

(12) Flesher, J. W.; Horn, J.; Lehner, A. F. 9-Sulfoxyanthracene is an ultimate electrophilic and carcinogenic form of 9-hydroxymethylanthracene. *Biochem. Biophys. Res. Commun.* **1998**, *251*, 239–243.

(13) Casale, R.; McLaughlin, L. W. Synthesis and Properties of an Oligodeoxynucleotide Containing a Polycyclic Aromatic Hydrocarbon Site Specifically Bound to the N2-Amino Group of a 2'-Deoxyguanosine Residue. *J. Am. Chem. Soc.* **1990**, *112*, 5264–5271.

(14) Bendadani, C.; Meinel, W.; Monien, B. H.; Dobbernack, G.; Glatt, H. The Carcinogen 1-Methylpyrene Forms Benzylic DNA Adducts in Mouse and Rat Tissues in Vivo via a Reactive Sulphuric Acid Ester. *Arch. Toxicol.* **2014**, *88*, 815–821.

(15) Bendadani, C.; Meinel, W.; Monien, B.; Dobbernack, G.; Florian, S.; Engst, W.; Nolden, T.; Himmelbauer, H.; Glatt, H. Determination of Sulfotransferase Forms Involved in the Metabolic Activation of the Genotoxicant 1-Hydroxymethylpyrene Using Bacterially Expressed Enzymes and Genetically Modified Mouse Models. *Chem. Res. Toxicol.* **2014**, *27*, 1060–1069.

(16) Guengerich, F. P. Interactions of Carcinogen-Bound DNA with Individual DNA Polymerases. *Chem. Rev.* **2006**, *106*, 420–452.

(17) Delaney, J. C.; Essigmann, J. M. Biological Properties of Single Chemical-DNA Adducts: A Twenty Year Perspective. *Chem. Res. Toxicol.* **2008**, *21*, 232–252.

(18) DeCorte, B. L.; Tsarouhtsis, D.; Kuchimanchi, S.; Cooper, M. D.; Horton, P.; Harris, C. M.; Harris, T. M. Improved Strategies for Postoligomerization Synthesis of Oligodeoxynucleotides Bearing Structurally Defined Adducts at the N2 Position of Deoxyguanosine. *Chem. Res. Toxicol.* **1996**, *9*, 630–637.

(19) Steinbrecher, T.; Wameling, C.; Oesch, F.; Seidel, A. Activation of the C2 Position of Purine by the Trifluoromethanesulfonate Group: Synthesis of N2-Alkylated Deoxyguanosines. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 404–406.

(20) He, G. X.; Krawczyk, S. H.; Swaminathan, S.; Shea, R. G.; Dougherty, J. P.; Terhorst, T.; Law, V. S.; Griffin, L. C.; Coutré, S.; Bischofberger, N. N2- and C8-Substituted Oligodeoxynucleotides with Enhanced Thrombin Inhibitory Activity in Vitro and in Vivo. *J. Med. Chem.* **1998**, *41*, 2234–2242.

- (21) Moon, K. Y.; Moschel, R. C. Effect of Ionic State of 2'-Deoxyguanosine and Solvent on Its Alkylation by Benzyl Bromide. *Chem. Res. Toxicol.* **1998**, *11*, 696–702.
- (22) Sergueeva, Z. A.; Sergueev, D. S.; Shaw, B. R. Rapid and Selective Reduction of Amide Group by Borane-Amine Complexes in Acyl Protected Nucleosides. *Nucleos. Nucleot. Nucl.* **2000**, *19*, 275–282.
- (23) Choi, J.; Guengerich, F. P. Analysis of the Effect of Bulk at N2-Alkylguanine DNA Adducts on Catalytic Efficiency and Fidelity of the Processive DNA Polymerases Bacteriophage T7 Exonuclease and HIV-1 Reverse Transcriptase. *J. Biol. Chem.* **2004**, *279*, 19217–19229.
- (24) Bonala, R. R.; Shishkina, I. G.; Johnson, F. Synthesis of Biologically Active N2-Amine Adducts of 2'-Deoxyguanosine. *Tetrahedron Lett.* **2000**, *41*, 7281–7284.
- (25) Lee, H.; Luna, E.; Hinz, M.; Stezowski, J. J.; Kiselyo, A. S.; Harvey, R. G. Synthesis of Oligonucleotide Adducts of the Bay Region Diol Epoxide Metabolites of Carcinogenic Polycyclic Aromatic Hydrocarbons. *J. Org. Chem.* **1995**, *60*, 5604–5613.
- (26) Kath, J. E.; Jergic, S.; Heltzel, J. M. H.; Jacob, D. T.; Dixon, N. E.; Sutton, M. D.; Walker, G. C.; Loparo, J. J. Polymerase Exchange on Single DNA Molecules Reveals Processivity Clamp Control of Translesion Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7647–7652.
- (27) Ghodke, P. P.; Harikrishna, S.; Pradeepkumar, P. I. Synthesis and Polymerase-Mediated Bypass Studies of the N2-Deoxyguanosine DNA Damage Caused by a Lucidin Analogue. *J. Org. Chem.* **2015**, *80*, 2128–2138.
- (28) Champeil, E.; Pradhan, P.; Lakshman, M. K. Palladium-Catalyzed Synthesis of Nucleoside Adducts from Bay- and Fjord-Region Diol Epoxides. *J. Org. Chem.* **2007**, *72*, 5035–5045.
- (29) Lu, D.; Lei, J.; Tian, Z.; Wang, L.; Zhang, J. Cu²⁺ Fluorescent Sensor Based on Mesoporous Silica Nanosphere. *Dyes Pigm.* **2012**, *94*, 239–246.
- (30) Abiraj, K.; Gowda, D. C. Zinc/Ammonium Formate: A New Facile System for the Rapid and Selective Reduction of Oximes to Amines. *J. Chem. Res.* **2003**, *2003*, 332–334.
- (31) Dai, Q.; Zheng, G.; Schwartz, M. H.; Clark, W. C.; Pan, T. Selective Enzymatic Demethylation of N²,N²-Dimethylguanosine in RNA and Its Application in High-Throughput tRNA Sequencing. *Angew. Chem., Int. Ed.* **2017**, *56*, 5017–5020.
- (32) Himmelsbach, F.; Schulz, B. S.; Trichtinger, T.; Charubala, R.; Pfeleiderer, W. The p-nitrophenylethyl (NPE) group: A versatile new blocking group for phosphate and aglycone protection in nucleosides and nucleotides. *Tetrahedron* **1984**, *40*, 59–72.
- (33) Kottur, J.; Ghodke, P. P.; Nair, D. T.; Pradeepkumar, P. I. Manuscript in preparation.
- (34) Alam, R.; Dixit, V.; Kang, H.; Li, Z. B.; Chen, X.; Trejo, J.; Fisher, M.; Juliano, R. L. Intracellular Delivery of an Anionic Antisense Oligonucleotide via Receptor-Mediated Endocytosis. *Nucleic Acids Res.* **2008**, *36*, 2764–2776.
- (35) Sholder, G.; Creech, A.; Loechler, E. L. How Y-Family DNA Polymerase IV Is More Accurate than Dpo4 at dCTP Insertion Opposite an N²-dG Adduct of Benzo[a]Pyrene. *DNA Repair* **2015**, *35*, 144–153.
- (36) Ghodke, P. P.; Gore, K. R.; Harikrishna, S.; Samanta, B.; Kottur, J.; Nair, D. T.; Pradeepkumar, P. I. The N2-Furfuryl-Deoxyguanosine Adduct Does Not Alter the Structure of B-DNA. *J. Org. Chem.* **2016**, *81*, 502–511.
- (37) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
- (38) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (39) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R. B.; Wu, X.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, CA, 2014.
- (40) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-H... π -Interactions in Proteins. *J. Mol. Biol.* **2001**, *307*, 357–377.
- (41) Sharma, A.; Kottur, J.; Narayanan, N.; Nair, D. T. A Strategically Located Serine Residue Is Critical for the Mutator Activity of DNA Polymerase IV from Escherichia Coli. *Nucleic Acids Res.* **2013**, *41*, 5104–5114.
- (42) Vanqualef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerek, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39*, W511–7.
- (43) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821.
- (44) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, P. 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.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, Ö.; Foresman, J. B.; Fox, D. J. *Gaussian 09*, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2013.
- (46) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (47) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Lughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (48) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (49) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- (50) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (52) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(53) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

Supporting Information

Synthesis of N^2 -deoxyguanosine modified DNAs and the studies on their translesion synthesis by the *E. coli* DNA polymerase IV

Pratibha P. Ghodke,[†] Praneeth Bommiseti,[†] Deepak T. Nair,^{*,§} P. I. Pradeepkumar^{*,†}

[†]*Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076*

[§]*Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurgaon Expressway, Faridabad-121001, India*

Email: deepak@rcb.res.in or pradeep@chem.iitb.ac.in

TABLE OF CONTENTS

Figure S1	Nucleotide incorporation using DNA Pol IV with unmodified template.....	Page S1
Figure S2	Nucleotide incorporation using DNA Pol IV with N^2 -Bn-dG template.....	Page S2
Figure S3	Nucleotide incorporation using DNA Pol IV with N^2 -Naph-dG template.....	Page S3
Figure S4	Nucleotide incorporation using DNA Pol IV with N^2 -Anth-dG template.....	Page S4
Figure S5	Nucleotide incorporation using DNA Pol IV with N^2 -Pyre-dG template.....	Page S5
Figure S6	Full length extension reactions using DNA Pol IV with unmodified, N^2 -Bn-dG, N^2 -Naph-dG templates.....	Page S6
Figure S7	Full length extension reactions using DNA Pol IV with unmodified, N^2 -Anth-dG, N^2 -Pyre-dG templates.....	Page S7
Figure S8	Energy optimized structures of N^2 -modified nucleotides.....	Page S8
Figure S9	N^2 -modified DNA – Pol IV ternary complexes before the start of simulation ...	Page S9
Figure S10	RMSD graphs of DNA and protein backbones for N^2 -modified DNA Pol IV ternary complexes.....	Page S10
Figure S11	RMSF values of DNA nucleotides.....	Page S12
Figure S12	RMSF values of Pol IV residues.....	Page S13
Figure S13	Similarity in the orientation of N^2 -modified entities in the protein environment during the course of dynamics.....	Page S15
Table S1	Pairwise interaction energies of N^2 modified entities with AAs of Pol IV.....	Page S16
Table S2	RMSF values of residues interacting with the N^2 -modified entities.....	Page S17
Figure S14	Distance between Ser42 of Pol IV and centroid of aromatic entity of N^2 -modified dG incorporated into DNA-Pol IV complex.....	Page S18

Figure S15	Cartesian coordinates and RESP charges calculated for <i>N</i> ² -Bn-dG nucleotide	Page S20
Figure S16	Cartesian coordinates and RESP charges calculated for <i>N</i> ² -Naph-dG nucleotide	Page S23
Figure S17	Cartesian coordinates and RESP charges calculated for <i>N</i> ² -Anth-dG nucleotide....	Page S27
Figure S18	Cartesian coordinates and RESP charges calculated for <i>N</i> ² -Pyre-dG nucleotide.....	Page S31
	¹ H NMR & ¹³ C NMR spectrum of compound 3a	Page S35
	¹ H NMR & ¹³ C NMR spectrum of compound 4a	Page S36
	³¹ P NMR spectrum of compound 5a	Page S37
	¹ H NMR & ¹³ C NMR spectrum of compound 3b	Page S37-S38
	¹ H NMR & ¹³ C NMR spectrum of compound 4b	Page S38-S39
	³¹ P NMR spectrum of compound 5b	Page S39
	¹ H NMR & ¹³ C NMR spectrum of compound 3c	Page S40
	¹ H NMR & ¹³ C NMR spectrum of compound 4c	Page S41
	³¹ P NMR spectrum of compound 5c	Page S42
	¹ H NMR & ¹³ C NMR spectrum of compound 3d	Page S42-S43
	¹ H NMR & ¹³ C NMR spectrum of compound 4d	Page S43-S44
	³¹ P NMR spectrum of compound 5d	Page S44
	ESI-MS spectrum of D2	Page S45
	MALDI spectrum of D6	Page S45
	MALDI spectrum of D10	Page S46
	MALDI spectrum of D14	Page S46

Nucleotide incorporation using DNA Pol IV with unmodified template

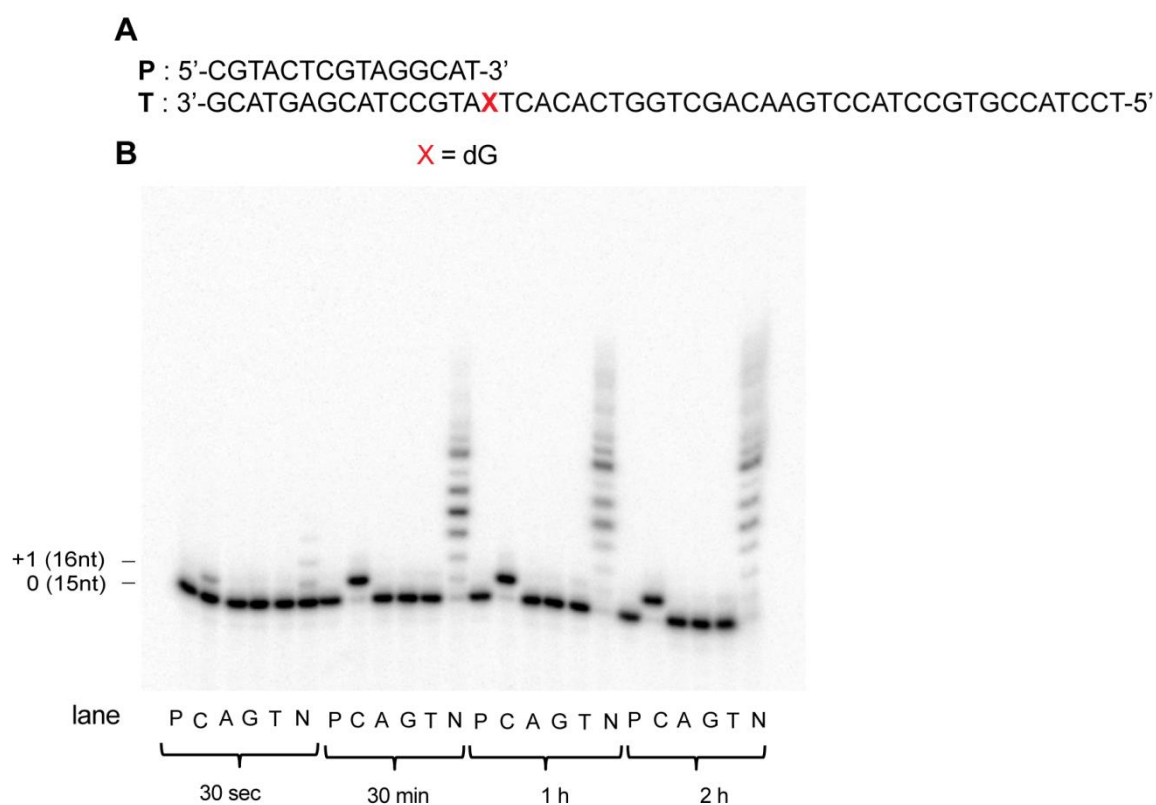


Figure S1. PAGE (20%, 7 M urea) of primer extension reactions with *E. coli* DNA polymerase IV. (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with dG template (X = dG): Lanes: P:primer, C:dCTP, A:dATP, G:dGTP, T:dTTP, N:mixture of dNTPs (10 μ M). All reactions were carried out at 37 $^{\circ}$ C for 30 sec, 30 min (~ 93% incorporation of dCTP), 1 h (~ 94% incorporation of dCTP) and 2 h (~ 93% incorporation of dCTP) time course.

Nucleotide incorporation using DNA Pol IV with N^2 -Bn-dG template

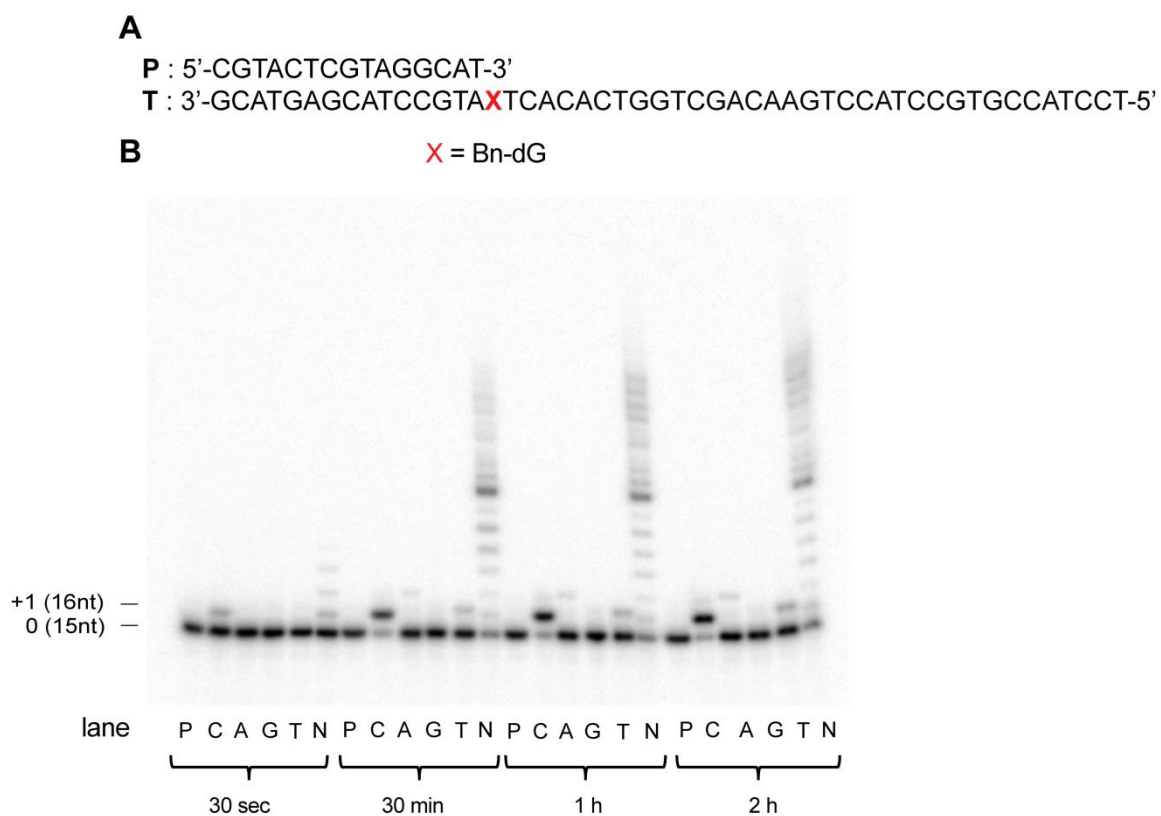


Figure S2. PAGE (20%, 7 M urea) of primer extension reactions with *E. coli* DNA polymerase IV. (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with N^2 -Bn-dG template ($\color{red}{X}$ = N^2 -Bn-dG): Lanes: P:primer, C:dCTP, A:dATP, G:dGTP, T:dTTP, N:mixture of dNTPs (10 μ M). All reactions were carried out at 37 °C for 30 sec, 30 min (~ 78% incorporation of dCTP), 1 h (~ 77% incorporation of dCTP) and 2 h (~ 81% incorporation of dCTP) time course.

Nucleotide incorporation using DNA Pol IV with N^2 -Naph-dG template

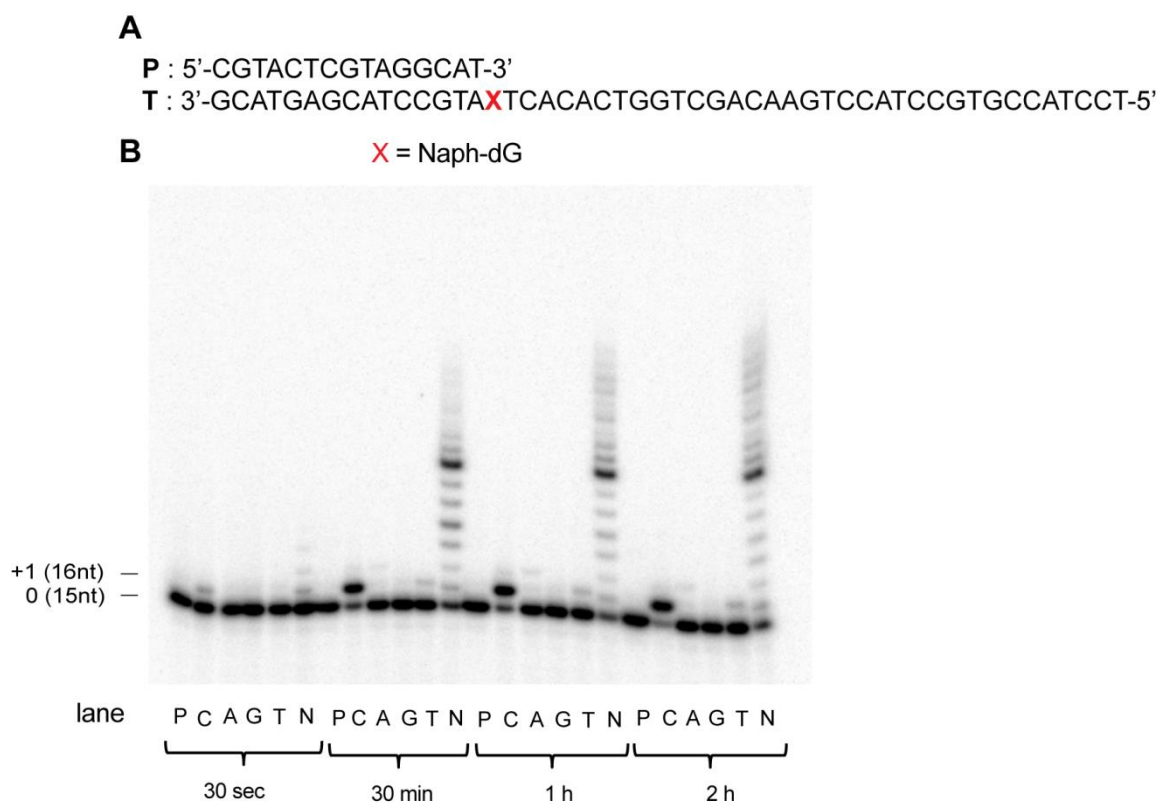


Figure S3. PAGE (20%, 7 M urea) of primer extension reactions with *E. coli* DNA polymerase IV. (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with N^2 -Naph-dG template (X = N^2 -Naph-dG): Lanes: P:primer, C:dCTP, A:dATP, G:dGTP, T:dTTP, N:mixture of dNTPs (10 μ M). All reactions were carried out at 37 °C for 30 sec, 30 min (~ 66% incorporation of dCTP), 1 h (~ 73% incorporation of dCTP) and 2 h (~ 75% incorporation of dCTP) time course.

Nucleotide incorporation using DNA Pol IV with N^2 -Anth-dG template

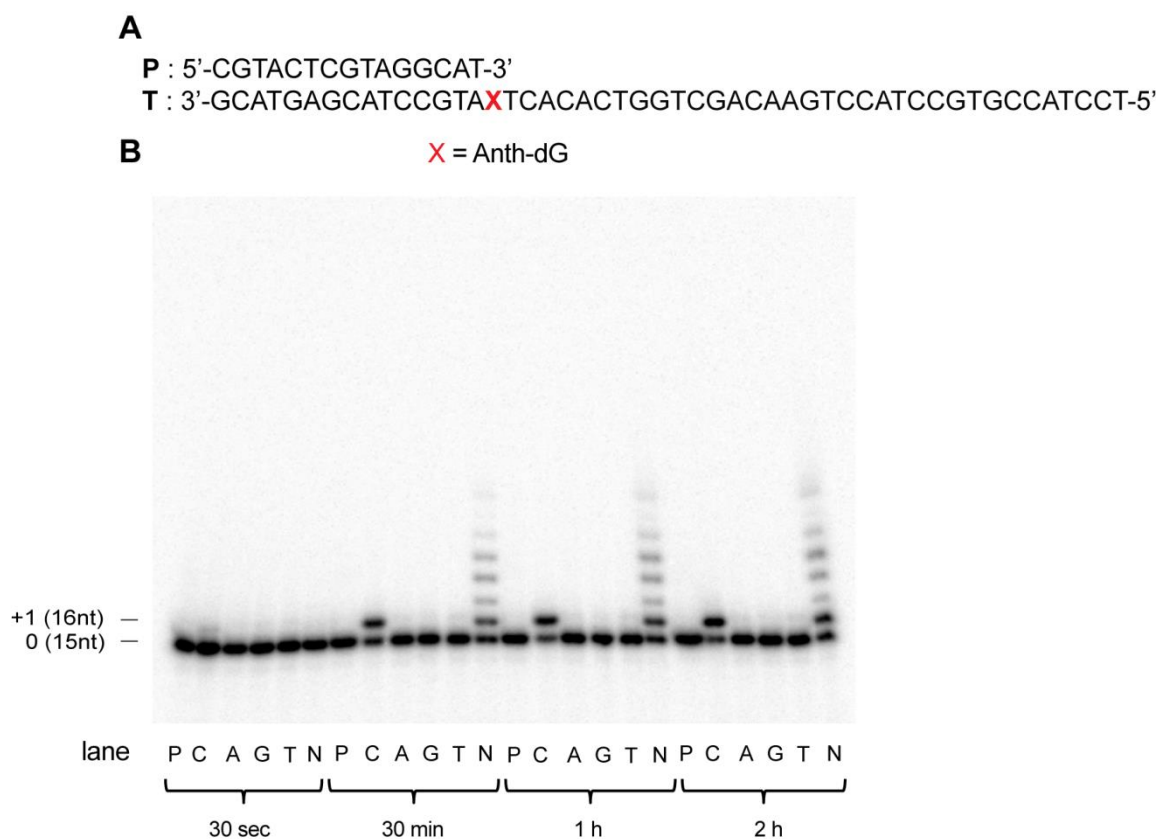


Figure S4. PAGE (20%, 7 M urea) of primer extension reactions with *E. coli* DNA polymerase IV . (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with N^2 -Anth-dG template ($\color{red}{X}$ = N^2 -Anth-dG): Lanes: P:primer, C:dCTP, A:dATP, G:dGTP, T:dTTP, N:mixture of dNTPs (10 μ M). All reactions were carried out at 37 °C for 30 sec, 30 min (~ 49% incorporation of dCTP), 1 h (~ 58% incorporation of dCTP) and 2 h (~ 62% incorporation of dCTP) time course.

Nucleotide incorporation using DNA Pol IV with N^2 -Pyre-dG template

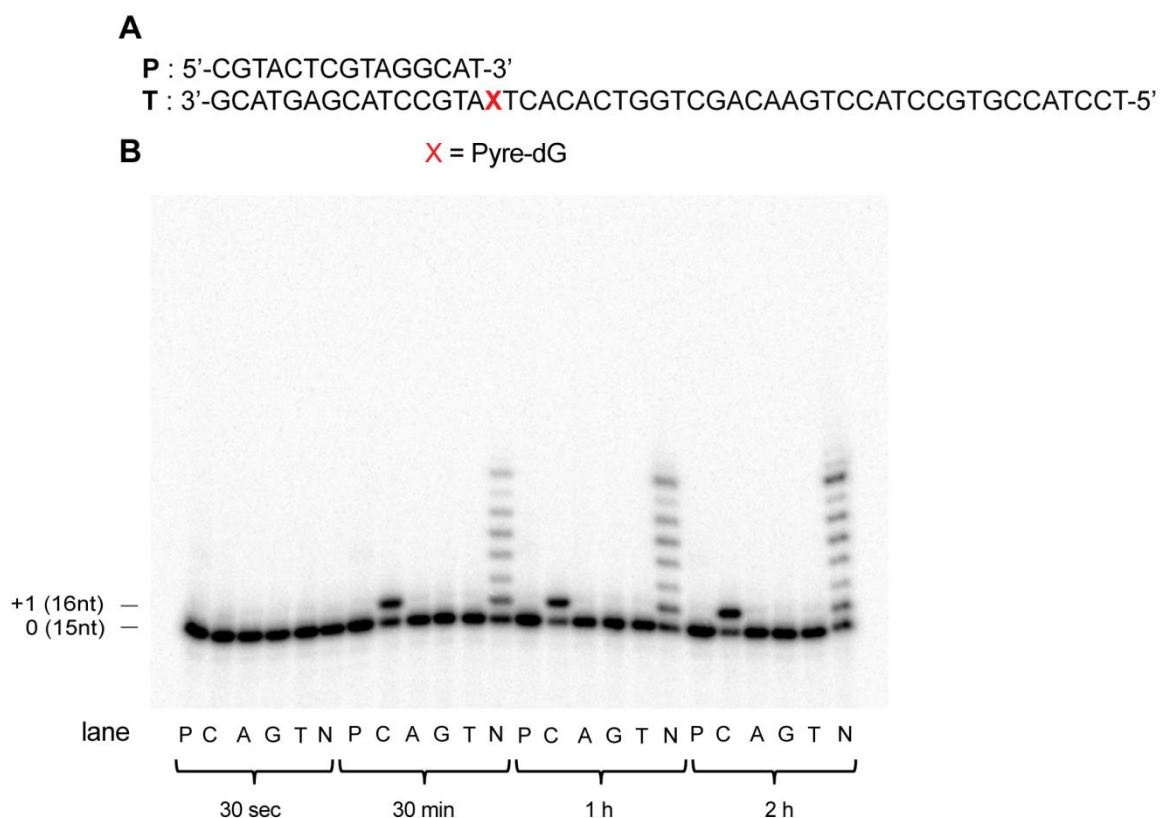


Figure S5. PAGE (20%, 7 M urea) of primer extension reactions with *E. coli* DNA polymerase IV. (A) Complete sequence of template, T and primer, P (15-mer); (B) reactions with N^2 -Pyre-dG template (X = N^2 -Pyre-dG): Lanes: P:primer, C:dCTP, A:dATP, G:dGTP, T:dTTP, N:mixture of dNTPs (10 μ M). All reactions were carried out at 37 °C for 30 sec, 30 min (~ 51% incorporation of dCTP), 1 h (~ 61% incorporation of dCTP) and 2 h (~ 66% incorporation of dCTP) time course.

Full length extension reactions using DNA Pol IV with unmodified, N^2 -Bn-dG, N^2 -Naph-dG templates

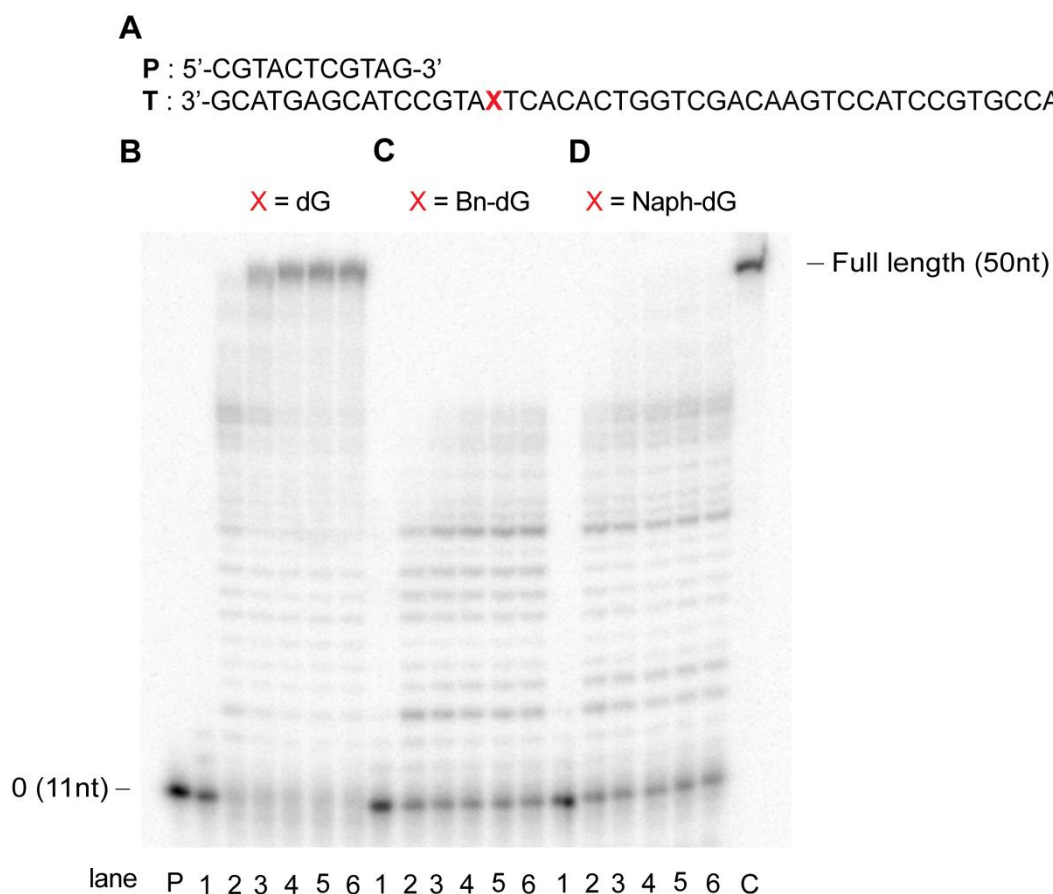


Figure S6. PAGE (20%, 7 M urea) of full length extension reactions using the *E. coli* DNA polymerase IV with all dNTPs. (A) Complete sequence of template, T and primer, P (11-mer); (B) reactions with dG template ($\color{red}{X}$ = dG); (C) reactions with N^2 -Bn-dG template ($\color{red}{X}$ = N^2 -Bn-dG); (D) reactions with N^2 -Naph-dG template ($\color{red}{X}$ = N^2 -Naph-dG). Lane P, primer; lanes 1 to 6: primer extension reactions with 250 μ M of mixture of dNTPs in different time course from 30 sec, 30 min, 1, 2, 4 and 8 h; lane C, 50-mer size standard. All the reactions were carried out at 37 $^{\circ}$ C.

Full length extension reactions using DNA Pol IV with unmodified, N^2 -Anth-dG, N^2 -Pyre-dG templates

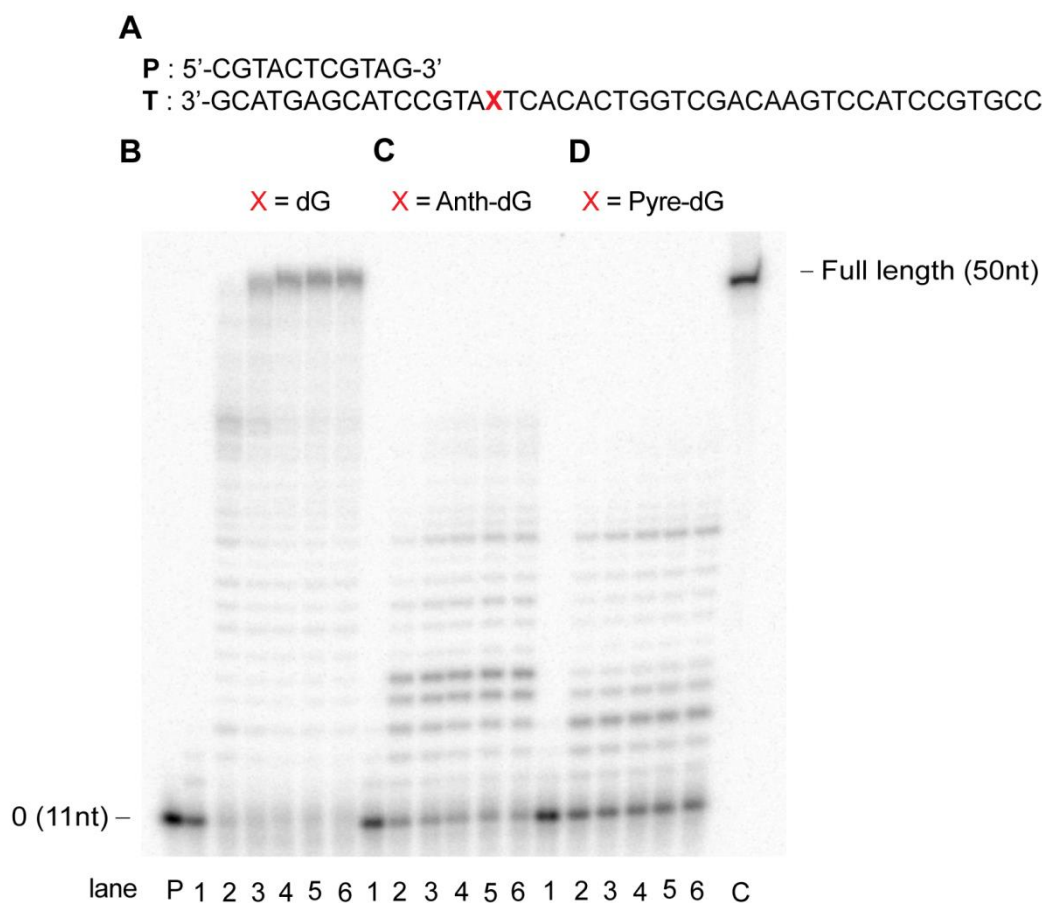


Figure S7. PAGE (20%, 7 M urea) of full length extension reactions using the *E. coli* DNA polymerase IV with all dNTPs. (A) Complete sequence of template, T and primer, P (11-mer); (B) reactions with dG template ($\color{red}{X}$ = dG); (C) reactions with N^2 -Anth-dG template ($\color{red}{X}$ = N^2 -Anth-dG); (D) reactions with N^2 -Pyre-dG template ($\color{red}{X}$ = N^2 -Pyre-dG). Lane P, primer; lanes 1 to 6: primer extension reactions with 250 μ M of mixture of dNTPs in different time course from 30 sec, 30 min, 1, 2, 4 and 8 h; lane C, 50-mer size standard. All the reactions were carried out at 37 $^{\circ}$ C.

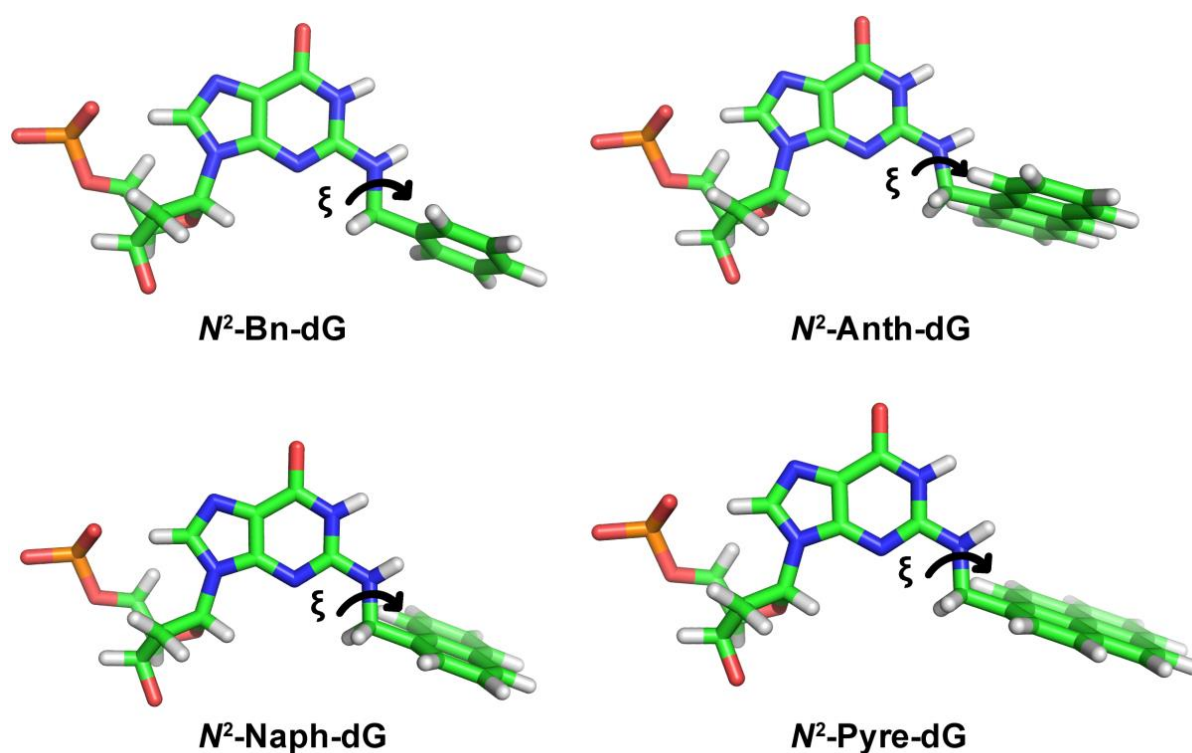
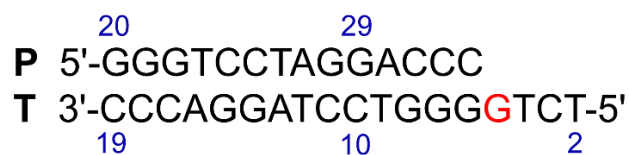
Energy optimized structures of N^2 -modified nucleotides

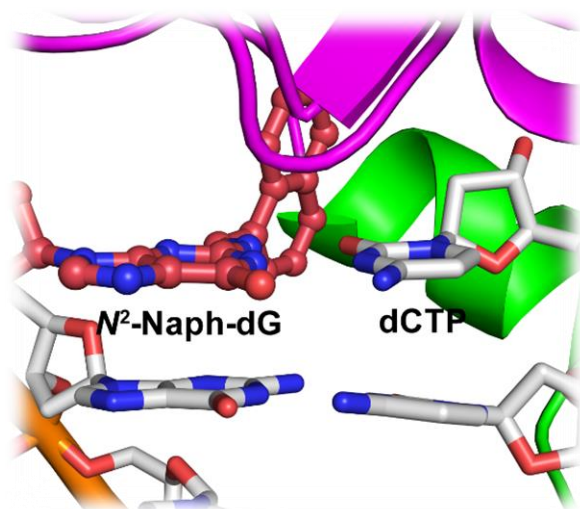
Figure S8. The energy optimized structures of N^2 -dG adducts that were incorporated into DNA at the active site of DNA-Pol IV ternary complex. Theory level of HF/6-31G* is used for optimization, and calculation of RESP charges. Carbon atoms were represented in green, nitrogen atoms in blue, oxygen atoms using in red, phosphorous atoms using reddish brown and hydrogen atoms using gray colors.

***N*²-modified DNA-Pol IV ternary complexes before the start of simulation**

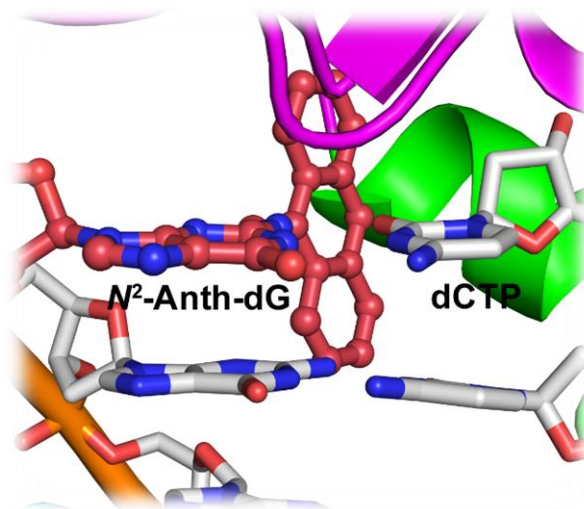
A.



B.



C.



D.

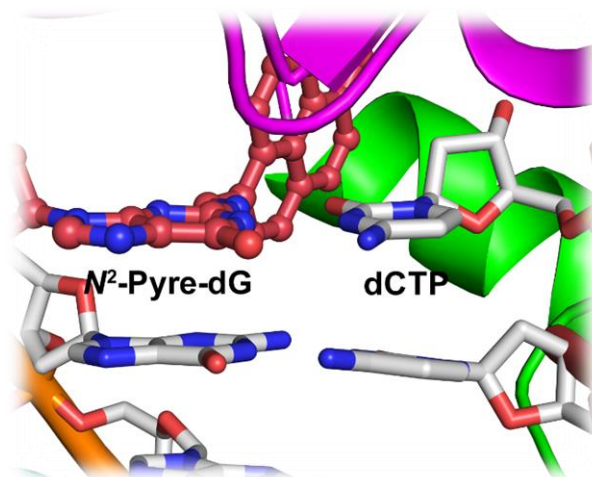
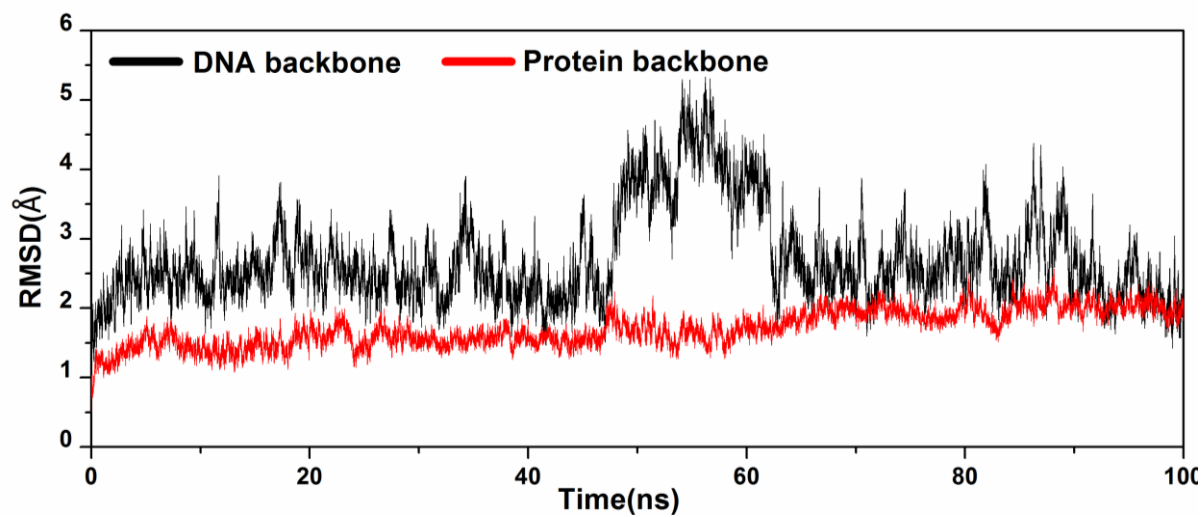


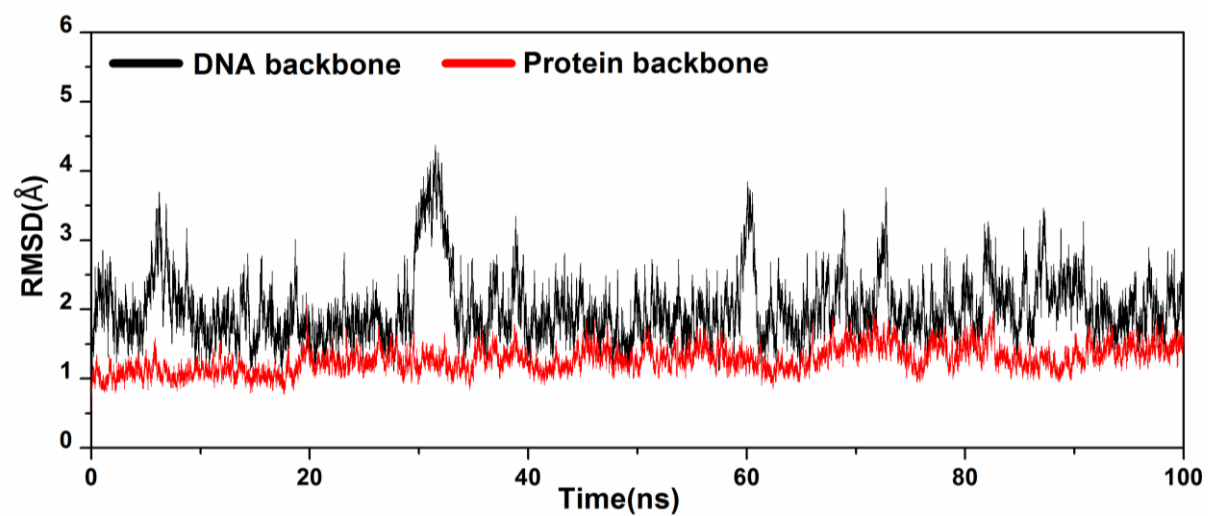
Figure S9. (A) The sequence of primer and template that was used for dynamics. The modified residue is highlighted in red. (B-D) Zoomed view of active site after incorporating the *N*²-Naph, -Anth and -Pyre dG modifications (starting structures). Cyan, magenta and green colours are used for polymerase associated domain (PAD), finger, palm and thumb domains respectively. Carbon atoms are represented in grey, nitrogen atoms in blue, oxygen atoms using in red, phosphorous atoms in reddish brown colours.

RMSD graphs of DNA and protein backbones for N^2 -modified DNA- Pol IV ternary complexes

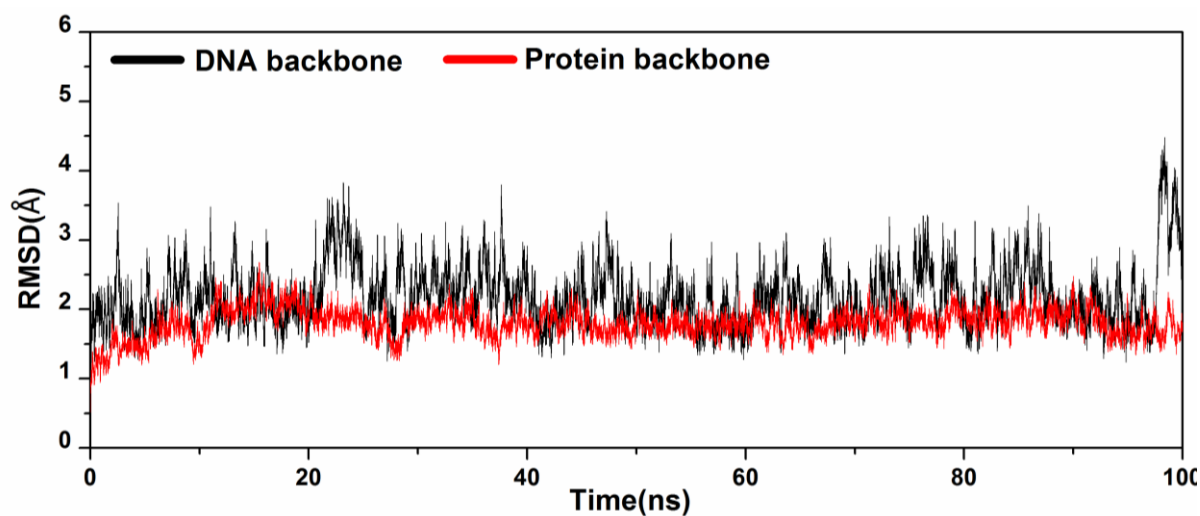
A



B



C



D

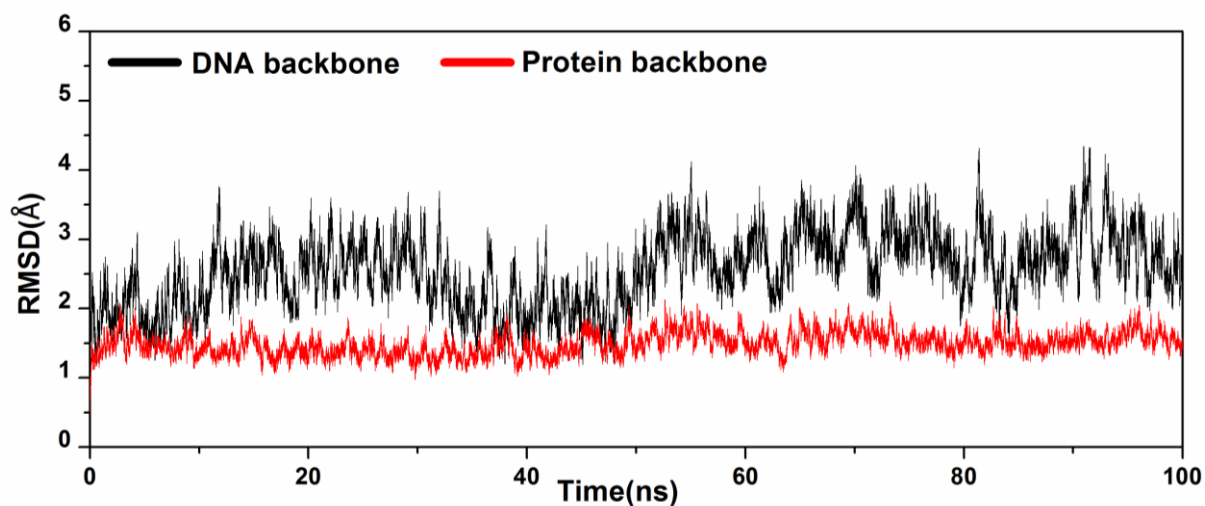


Figure S10. Root mean square deviation (RMSD) plots of DNA and protein backbone of Pol-IV-DNA complexes. (A) N^2 -Bn; (B) N^2 -Naph; (C) N^2 -Anth; (D) N^2 -Pyre complexes. DNA backbone is plotted using black and protein backbone using red colours. A total of 25000 frames were used for the calculations. CPPTRAJ of Amber14 is used for calculating values.

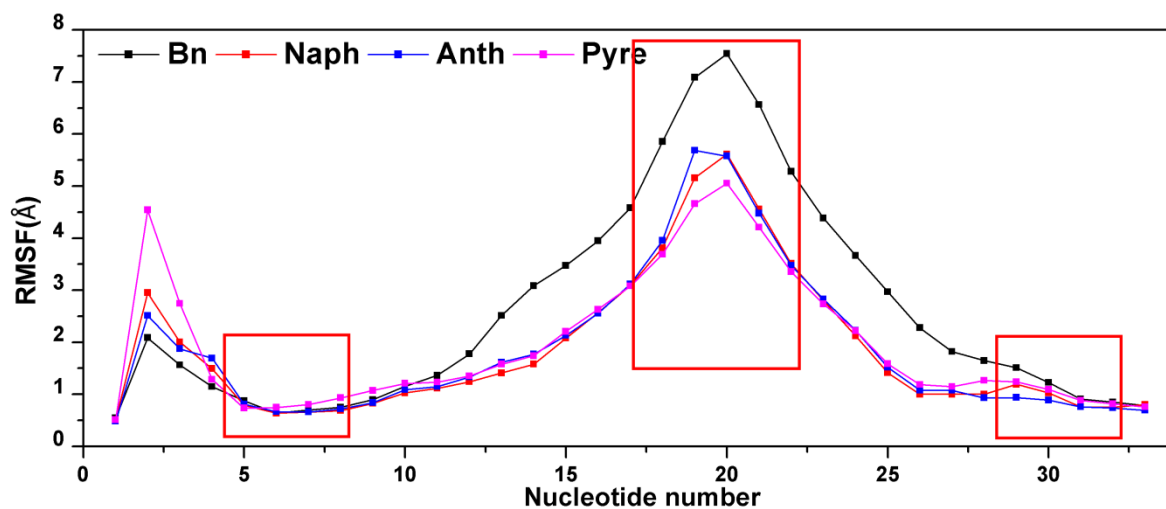
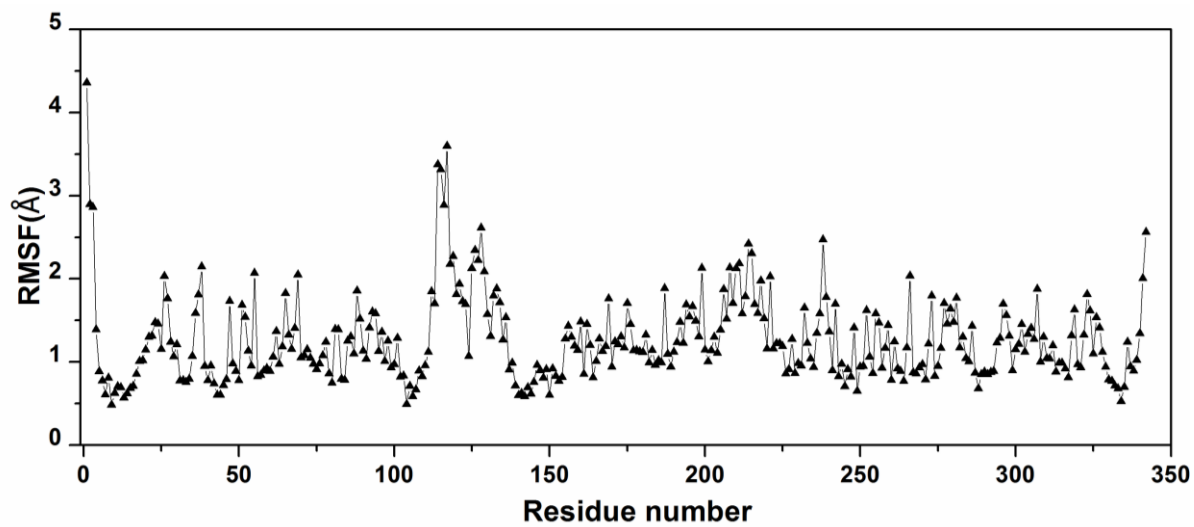
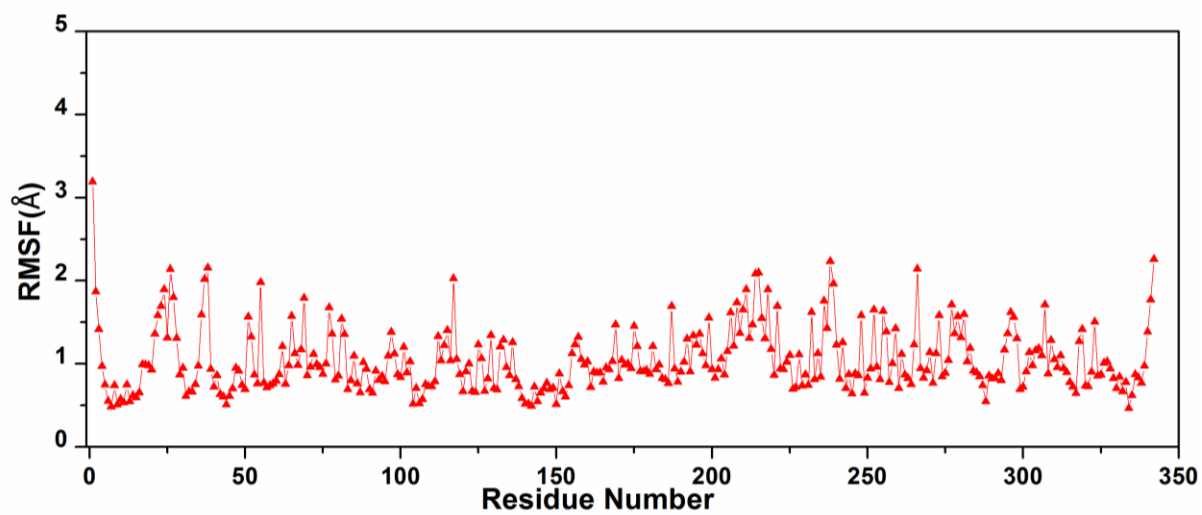
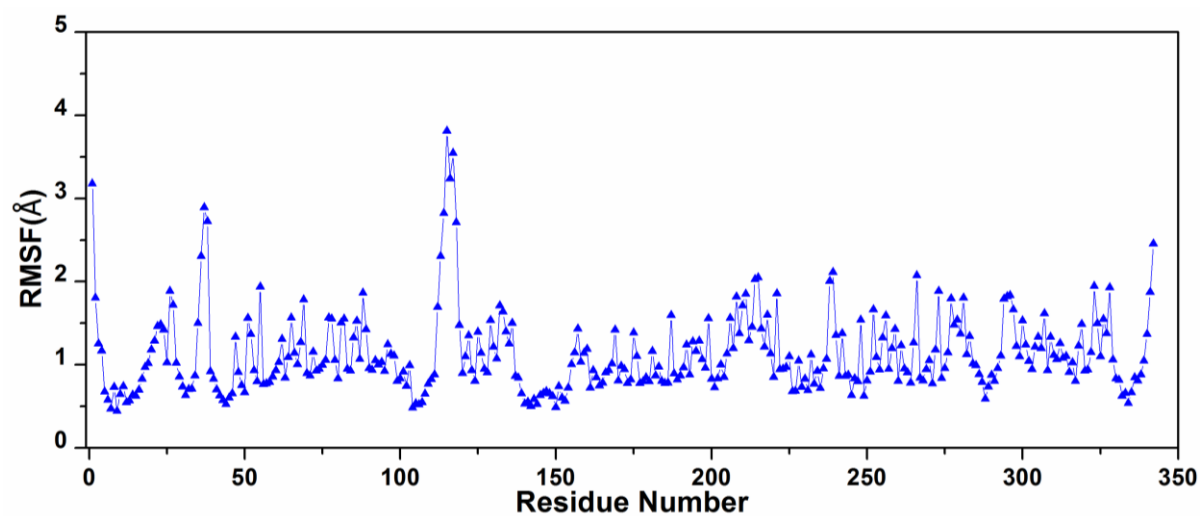
RMSF values of DNA nucleotides

Figure S11. Root mean square fluctuations (RMSF) plotted against nucleotide number of DNA (refer to S9A for numbering). The bigger rectangle highlights the terminal nucleotides of DNA and the smaller rectangle highlights the residues around the active site. Black, red, blue and pink colors are used to represent the values of N^2 -Bn, -Naph, -Anth and -Pyre complexes respectively. A total of 25000 frames were used for the calculation. CPPTRAJ of Amber14 is used for the calculations.

RMSF values of Pol IV residues**A****B**

C



D

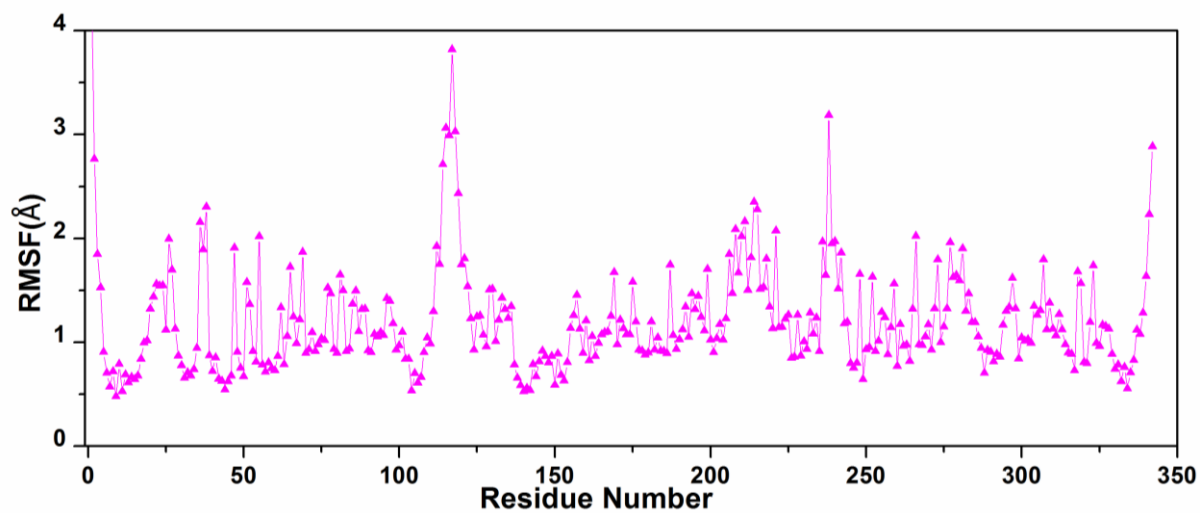


Figure S12. Root mean square fluctuations (RMSF) plotted against residue number of protein. Black, red, blue and pink were used to represent the values of N^2 -Bn, N^2 -Naph, N^2 -Anth and N^2 -Pyre complexes respectively. A total of 25000 frames were used for the calculation. CPPTRAJ of Amber14 was used for the calculations.

Similarity in the orientations of N^2 -modified entities in the protein environment during the course of dynamics

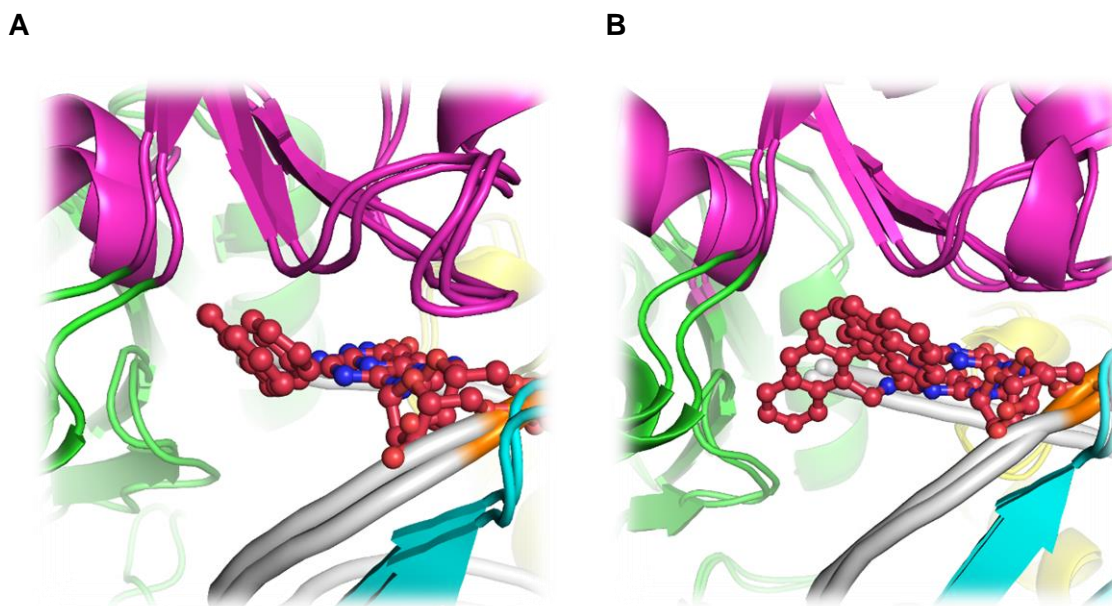


Figure S13. Similarity in the orientation of adduct moieties. (A) The representative structures of ensemble 1 of Bn and Naph complexes; (B) The representative structures of ensemble 1 of Anth and Pyre complexes. The orientations of N^2 -Bn, -Naph, and -Pyre are similar and resemble the orientation of furfuryl moiety in the crystal structure (PDB: 4Q43) of N^2 -fdG DNA – Pol IV complex.

Table S1. Pairwise interaction energies of N^2 -modified entities with AAs of Pol IV

Residue 1	Residue 2	X=Bn	X=Naph	X=Anth	X=Pyre
		Avg.	Avg.	Avg.	Avg.
N^2 -X-dG	Phe13	-0.699±0.234	-0.736±0.165	-1.710±0.373	-0.732±0.160
N^2 -X-dG	Ile31	-1.020±0.240	-1.033±0.234	-1.178±0.374	-1.227±0.310
N^2 -X-dG	Gly32	-1.656±0.571	-2.121±0.411	-0.878±0.323	-2.805±0.389
N^2 -X-dG	Gly33	-0.864±0.292	-1.173±0.233	-0.556±0.250	-1.842±0.334
N^2 -X-dG	Ser42	-1.984±0.433	-2.164±0.357	-2.217±0.471	-2.080±0.356
N^2 -X-dG	Leu71	-0.094±0.030	-0.091±0.027	-0.064±0.023	-0.146±0.0430
N^2 -X-dG	Leu72	-0.118±0.053	-0.138±0.050	-0.126±0.058	-0.272±0.104
N^2 -X-dG	Pro73	-0.642±0.1946	-0.975±0.203	-0.705±0.212	-0.831±0.240
N^2 -X-dG	Gly74	-0.378±0.160	-0.373±0.155	-0.385±0.147	-0.893±0.394
N^2 -X-dG	Phe76	-0.263±0.343	-0.937±0.62	-1.775±0.540	-1.669±0.537
N^2 -X-dG	Tyr79	-0.146±0.107	-0.195±0.09	-0.562±0.163	-0.167±0.127

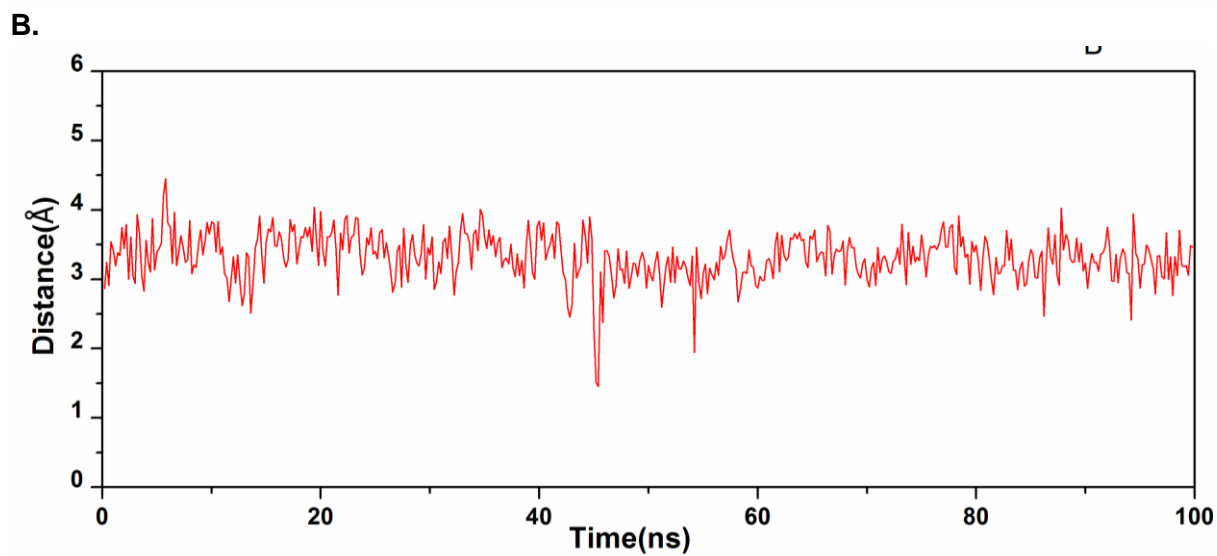
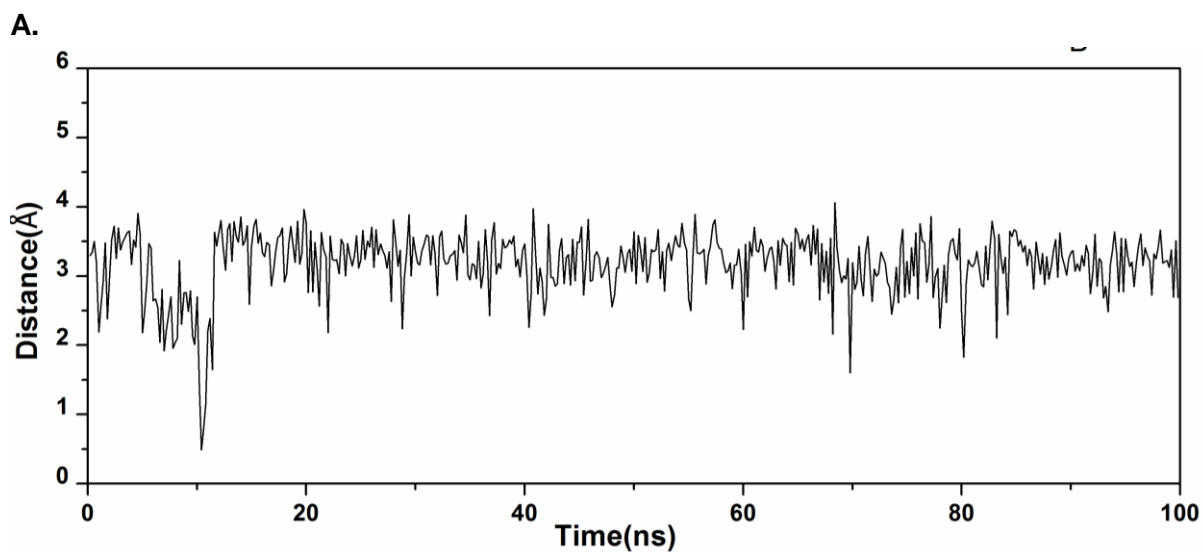
All values are reported in kcal/mol. Avg = Σ Internal + Electrostatic + vdW + polar solvation + non polar solvation. Internal comprises of bond, angle and dihedral energies and is taken as zero for reference. Electrostatic comprises of long range electrostatic interactions. vdW comprises of vander Waals interaction energies. Polar and non polar solvation energies are solvation energies. Energies are calculated over last 20 ns for N^2 -Naph, -Anth and -Pyre complexes, and over last 40 ns for Bn complex. A total of 500 snapshots are used for the calculations. The major contributions of interaction energies in each complexes are highlighted in bold. Calculations are performed using MM/GB-SA method implemented as the program MMPBSA.py in AmberTools package.

Table S2. RMSF values of residues interacting with the N^2 -modified entities in Pol IV-DNA complexes.

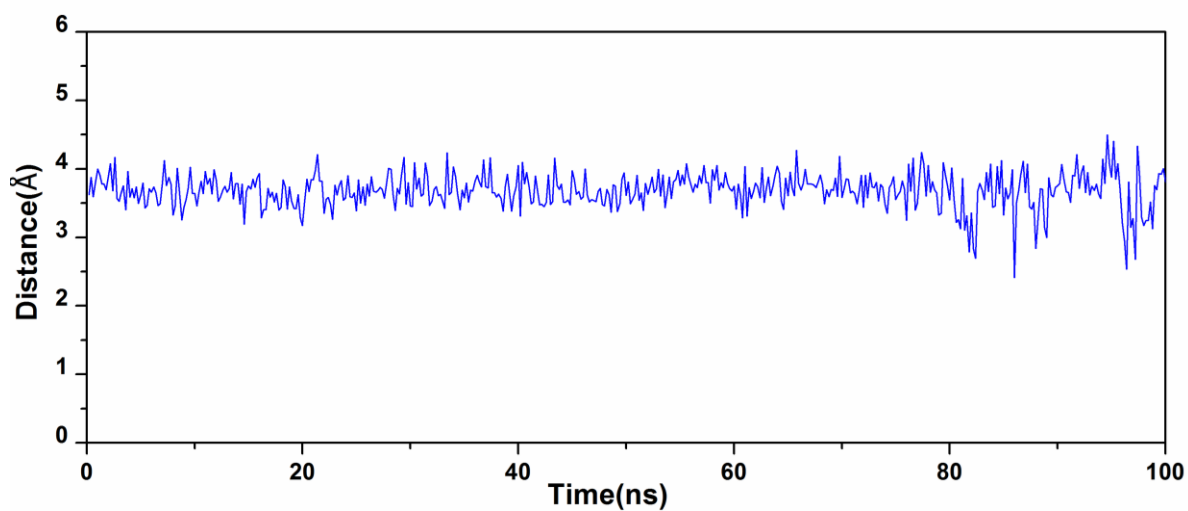
Residue	RMSF (Å)			
	Bn	Naph	Anth	Pyre
Phe13	0.569	0.548	0.569	0.612
Ile31	0.775	0.610	0.635	0.656
Gly32	0.777	0.662	0.704	0.708
Gly33	0.759	0.664	0.713	0.740
Ser42	0.729	0.632	0.63	0.647
Pro73	1.05	0.991	0.928	0.912
Gly74	0.974	0.956	0.948	0.980
Phe76	0.981	1.002	1.055	1.019
Tyr79	0.858	0.806	1.057	0.932

A total of 25000 frames were considered for this calculation. Calculations were performed using CPPTRAJ of Amber 14.

Distance between Ser42 of Pol IV and aromatic entity of N^2 -modified dG incorporated into DNA – Pol IV complex



C.



D.

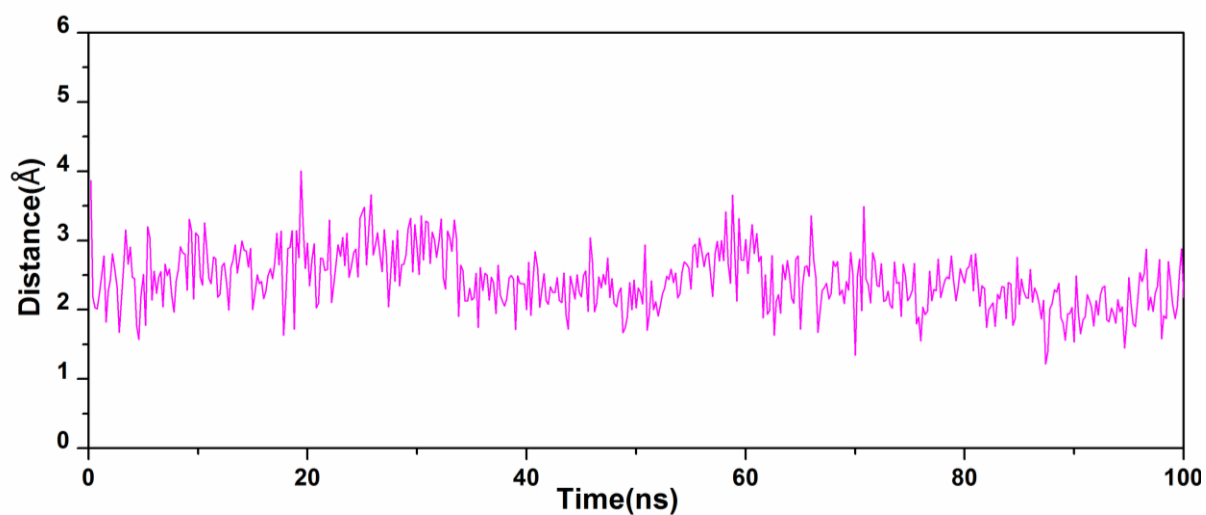
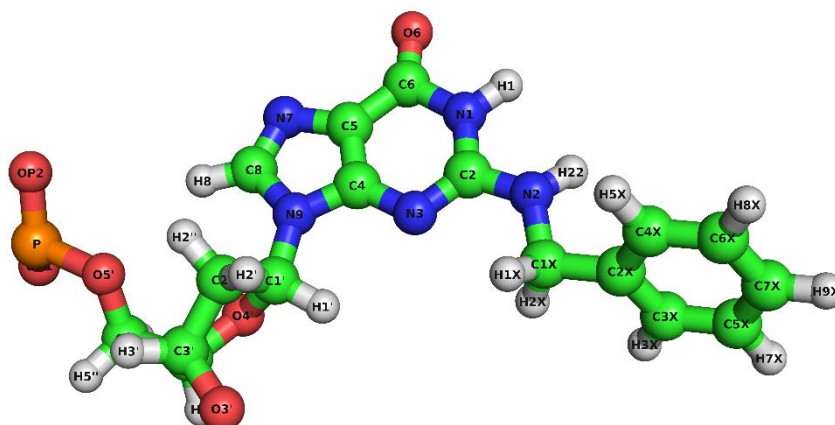


Figure S14. Distance between carbon atom attached to O-H of Ser42 and centroid of N^2 -modified aromatic entity. Black, red, blue and pink were used to represent the values of N^2 -Bn, -Naph, -Anth and -Pyre complexes respectively. Every 1000th frame and a total of 500 frames were used for the calculation.

Cartesian coordinates and RESP charges calculated for the N^2 -Bn-dG nucleotide

```

@<TRIPOS>MOLECULE
bD
      46      49      1      0      1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
  
```

1	O3'	-2.68159	-0.79023	-0.6172	OS	1	bD	-0.489	0	****
2	P	-4.11968	-1.56167	-0.43549	P	1	bD	1.2124	0	****
3	O1P	-4.4011	-2.32398	-1.66085	O2	1	bD	-0.7918	0	****
4	O2P	-5.03515	-0.58797	0.177407	O2	1	bD	-0.7918	0	****
5	O5'	1.407465	0.00000	0.00000	OS	1	bD	-0.5225	0	****
6	C1'	-0.49443	1.787454	-1.47939	CT	1	bD	0.2473	0	****
7	H1'	0.470563	2.198764	-1.73299	H2	1	bD	0.0872	0	****
8	C2'	-0.5707	1.412072	0.00000	CT	1	bD	-0.0893	0	****
9	H2'	-0.01303	2.097897	0.627461	HC	1	bD	0.0529	0	****
10	H2''	-1.6014	1.396675	0.327058	HC	1	bD	0.0529	0	****
11	C3'	0.00000	0.00000	0.00000	CT	1	bD	0.0508	0	****
12	H3'	-0.37979	-0.60417	0.815823	H1	1	bD	0.1138	0	****
13	C4'	-0.44398	-0.5386	-1.36727	CT	1	bD	0.1806	0	****
14	H4'	0.373549	-1.0964	-1.8072	H1	1	bD	0.1058	0	****
15	C5'	-1.66851	-1.42999	-1.32806	CI	1	bD	-0.0525	0	****
16	H5'	-1.97909	-1.63388	-2.348	H1	1	bD	0.0882	0	****
17	H5''	-1.40386	-2.37344	-0.85475	H1	1	bD	0.0882	0	****
18	O4'	-0.70962	0.590758	-2.17137	OS	1	bD	-0.4114	0	****
19	N2	1.038051	6.545553	-3.07064	N2	1	bD	-0.6342	0	****
20	H22	0.991314	7.313596	-3.70269	H	1	bD	0.3582	0	****
21	C6	-2.55399	5.868911	-3.16616	C	1	bD	0.6356	0	****
22	O6	-3.49312	6.535066	-3.48579	O	1	bD	-0.5689	0	****
23	C5	-2.47457	4.527739	-2.66253	CB	1	bD	0.0927	0	****
24	N7	-3.46325	3.602563	-2.40798	NB	1	bD	-0.5591	0	****
25	C8	-2.84085	2.576323	-1.96793	CK	1	bD	0.1323	0	****

26	H8	-3.28412	1.647364	-1.67872	H5	1	bD	0.1736	0	****
27	N9	-1.47547	2.753719	-1.90357	N*	1	bD	-0.0242	0	****
28	C4	-1.24528	4.010561	-2.35953	CB	1	bD	0.1857	0	****
29	N3	-0.02701	4.596997	-2.47477	NC	1	bD	-0.5839	0	****
30	C2	-0.08823	5.803639	-2.92744	CA	1	bD	0.763	0	****
31	N1	-1.25514	6.420772	-3.26294	NA	1	bD	-0.6607	0	****
32	H1	-1.24303	7.373498	-3.55834	H	1	bD	0.358	0	****
33	C1X	2.355464	5.945939	-2.90572	CT	1	bD	0.0307	0	****
34	H1X	2.382914	5.471428	-1.93513	H1	1	bD	0.076	0	****
35	H2X	2.52053	5.168887	-3.64442	H1	1	bD	0.076	0	****
36	C2X	3.42568	7.009283	-3.01202	CA	1	bD	0.0498	0	****
37	C3X	4.350824	6.984631	-4.0446	CA	1	bD	-0.1413	0	****
38	H3X	4.308226	6.200903	-4.78082	HA	1	bD	0.1348	0	****
39	C4X	3.498247	8.032289	-2.06959	CA	1	bD	-0.1413	0	****
40	H5X	2.785555	8.061588	-1.26404	HA	1	bD	0.1348	0	****
41	C5X	5.335037	7.959441	-4.13773	CA	1	bD	-0.1687	0	****
42	H7X	6.04603	7.925301	-4.94371	HA	1	bD	0.146	0	****
43	C6X	4.47538	9.005894	-2.16102	CA	1	bD	-0.1687	0	****
44	H8X	4.521362	9.788959	-1.42541	HA	1	bD	0.146	0	****
45	C7X	5.398449	8.971241	-3.19758	CA	1	bD	-0.1091	0	****
46	H9X	6.159404	9.727699	-3.26769	HA	1	bD	0.1351	0	****

@<TRIPOS>BOND

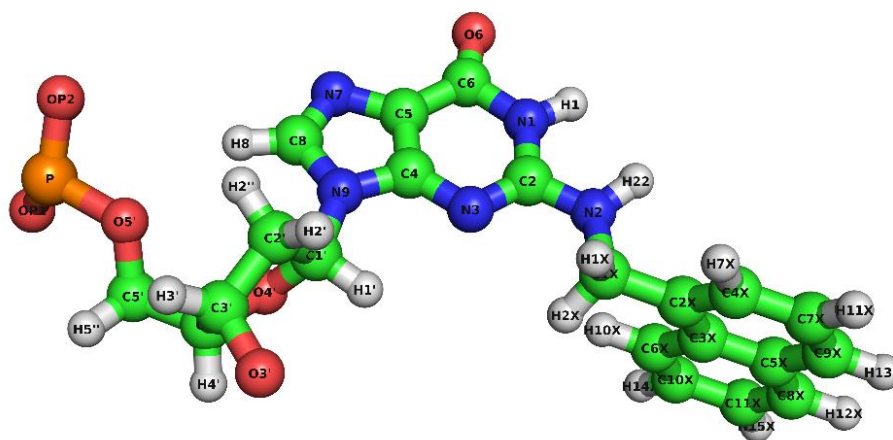
1	1	2	1
2	1	15	1
3	2	3	1
4	2	4	1
5	5	11	1
6	6	7	1
7	6	8	1
8	6	18	1
9	6	27	1
10	8	9	1
11	8	10	1
12	8	11	1
13	11	12	1
14	11	13	1
15	13	14	1
16	13	15	1
17	13	18	1
18	15	16	1
19	15	17	1
20	19	20	1
21	19	30	1
22	19	33	1
23	21	22	1
24	21	23	1

```

25      21      31      1
26      23      24      1
27      23      28      1
28      24      25      1
29      25      26      1
30      25      27      1
31      27      28      1
32      28      29      1
33      29      30      1
34      30      31      1
35      31      32      1
36      33      34      1
37      33      35      1
38      33      36      1
39      36      37      1
40      36      39      1
41      37      38      1
42      37      41      1
43      39      40      1
44      39      43      1
45      41      42      1
46      41      45      1
47      43      44      1
48      43      45      1
49      45      46      1
@<TRIPOS>SUBSTRUCTURE
      1      bD      1      ****      0      ****      ****
@<TRIPOS>HEADTAIL
P      1
O5'    1
@<TRIPOS>RESIDUECONNECT
1      P      O5'      0      0      0      0

```

Figure S15. Cartesian coordinates and RESP charges calculated for the N^2 -Bn-dG nucleotide in SYBL Mol2 format. Since it was not feasible to directly optimize and calculate charges for the molecular fragments, RESP charges were derived via two-step process. The N^2 -Bn modified dG and dimethyl phosphate were optimized and partial charges were calculated separately for both molecules by the RED Dev. Server (<http://upjv.q4md-forcefieldtools.org/REDServer-Development/>). The charges were fitted into single molecular fragment by using PyRED program on the RED Dev. Server. A theory level of HF/6-31G(d) and Gaussian 09 Revision E.01. was used for all calculations by the server. Carbon atoms are represented using green, nitrogen atoms using blue, oxygen atoms using red, phosphorous using brown red and hydrogens using grey colour. Electronic energy for the modified nucleoside: -1226.53639067 (RHF). Electronic energy for the dimethylphosphate: -719.520866767 (RHF). Detailed procedure for the derivation of partial charges for modified nucleotide fragments can be found in the tutorial for the server at: <http://upjv.q4md-forcefieldtools.org/Tutorial/Tutorial-4.php>

Cartesian coordinates and RESP charges calculated for the N²-Naph-dG nucleotide

```

@<TRIPOS>MOLECULE
nD
      52      56      1      0      1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
  1      O3'      -2.68118      -0.79113      -0.61595      OS      1      nD      -0.4868      0      ****
  2      P      -4.11899      -1.56297      -0.43373      P      1      nD      1.2118      0      ****
  3      O1P      -4.40096      -2.32483      -1.65925      O2      1      nD      -0.792      0      ****
  4      O2P      -5.03432      -0.58973      0.180106      O2      1      nD      -0.792      0      ****
  5      O5'      1.407427      0.00000      0.00000      OS      1      nD      -0.5143      0      ****
  6      C1'      -0.49515      1.787383      -1.47947      CT      1      nD      0.116      0      ****
  7      H1'      0.469737      2.198828      -1.73325      H2      1      nD      0.1374      0      ****
  8      C2'      -0.57076      1.412083      0.00000      CT      1      nD      -0.1038      0      ****
  9      H2'      -0.01274      2.097974      0.627084      HC      1      nD      0.061      0      ****
  10     H2''      -1.60128      1.396693      0.327642      HC      1      nD      0.061      0      ****
  11     C3'      0.00000      0.00000      0.00000      CT      1      nD      0.0512      0      ****
  12     H3'      -0.3798      -0.60412      0.815858      H1      1      nD      0.1228      0      ****
  13     C4'      -0.44405      -0.53871      -1.36722      CT      1      nD      0.1165      0      ****
  14     H4'      0.373551      -1.09626      -1.8073      H1      1      nD      0.1285      0      ****
  15     C5'      -1.66836      -1.43037      -1.32767      CI      1      nD      -0.0208      0      ****
  16     H5'      -1.97948      -1.63387      -2.34751      H1      1      nD      0.0819      0      ****
  17     H5''      -1.40331      -2.37397      -0.85488      H1      1      nD      0.0819      0      ****
  18     O4'      -0.71025      0.590551      -2.17123      OS      1      nD      -0.3645      0      ****
  19     N2      1.042264      6.506583      -3.17002      N2      1      nD      -0.5673      0      ****
  20     H22     0.950325      7.493819      -3.25845      H      1      nD      0.3432      0      ****
  21     C6      -2.55583      5.859564      -3.18729      C      1      nD      0.6188      0      ****

```


22	O6	-3.49383	6.517158	-3.52709	O	1	nD	-0.5653	0	****
23	C5	-2.47668	4.525331	-2.66572	CB	1	nD	0.1074	0	****
24	N7	-3.4652	3.601379	-2.40566	NB	1	nD	-0.5702	0	****
25	C8	-2.84229	2.576527	-1.96342	CK	1	nD	0.1284	0	****
26	H8	-3.28529	1.64885	-1.6697	H5	1	nD	0.1797	0	****
27	N9	-1.47648	2.753371	-1.90339	N*	1	nD	0.0063	0	****
28	C4	-1.24664	4.007927	-2.36589	CB	1	nD	0.2103	0	****
29	N3	-0.02951	4.597839	-2.47442	NC	1	nD	-0.588	0	****
30	C2	-0.08967	5.795621	-2.95226	CA	1	nD	0.7074	0	****
31	N1	-1.2572	6.412052	-3.28676	NA	1	nD	-0.6489	0	****
32	H1	-1.23661	7.321517	-3.69617	H	1	nD	0.3612	0	****
33	C1X	2.337621	6.027024	-2.70644	CT	1	nD	0.0008	0	****
34	H1X	2.351775	5.958598	-1.62444	H1	1	nD	0.0783	0	****
35	H2X	2.479771	5.025295	-3.08465	H1	1	nD	0.0783	0	****
36	C2X	3.439481	6.953373	-3.17072	CA	1	nD	-0.0073	0	****
37	C3X	3.833532	7.010665	-4.54849	CB	1	nD	0.1181	0	****
38	C4X	4.070903	7.746335	-2.26213	CA	1	nD	-0.1564	0	****
39	H7X	3.781496	7.702907	-1.22667	HA	1	nD	0.1532	0	****
40	C5X	4.871092	7.889122	-4.9255	CB	1	nD	0.1046	0	****
41	C6X	3.227359	6.211661	-5.55665	CA	1	nD	-0.2861	0	****
42	H10X	2.429149	5.544484	-5.29533	HA	1	nD	0.1857	0	****
43	C7X	5.110132	8.629725	-2.64287	CA	1	nD	-0.1957	0	****
44	H11X	5.585107	9.23988	-1.89601	HA	1	nD	0.1582	0	****
45	C8X	5.275071	7.948136	-6.28482	CA	1	nD	-0.1946	0	****
46	H12X	6.067593	8.622068	-6.55914	HA	1	nD	0.1405	0	****
47	C9X	5.497944	8.698181	-3.93907	CA	1	nD	-0.192	0	****
48	H13X	6.287265	9.364377	-4.2399	HA	1	nD	0.1531	0	****
49	C10X	3.637858	6.288241	-6.85064	CA	1	nD	-0.0949	0	****
50	H14X	3.168246	5.675422	-7.59895	HA	1	nD	0.1347	0	****
51	C11X	4.677517	7.170204	-7.22402	CA	1	nD	-0.1364	0	****
52	H15X	4.990124	7.220662	-8.25147	HA	1	nD	0.1391	0	****

@<TRIPOS>BOND

1	1	2	1
2	1	15	1
3	2	3	1
4	2	4	1
5	5	11	1
6	6	7	1
7	6	8	1
8	6	18	1
9	6	27	1
10	8	9	1
11	8	10	1
12	8	11	1
13	11	12	1
14	11	13	1

15	13	14	1
16	13	15	1
17	13	18	1
18	15	16	1
19	15	17	1
20	19	20	1
21	19	30	1
22	19	33	1
23	21	22	1
24	21	23	1
25	21	31	1
26	23	24	1
27	23	28	1
28	24	25	1
29	25	26	1
30	25	27	1
31	27	28	1
32	28	29	1
33	29	30	1
34	30	31	1
35	31	32	1
36	33	34	1
37	33	35	1
38	33	36	1
39	36	37	1
40	36	38	1
41	37	40	1
42	37	41	1
43	38	39	1
44	38	43	1
45	40	45	1
46	40	47	1
47	41	42	1
48	41	49	1
49	43	44	1
50	43	47	1
51	45	46	1
52	45	51	1
53	47	48	1
54	49	50	1
55	49	51	1
56	51	52	1

@<TRIPOS>SUBSTRUCTURE

	1	nD	1	****	0	****	****
--	---	----	---	------	---	------	------

@<TRIPOS>HEADTAIL

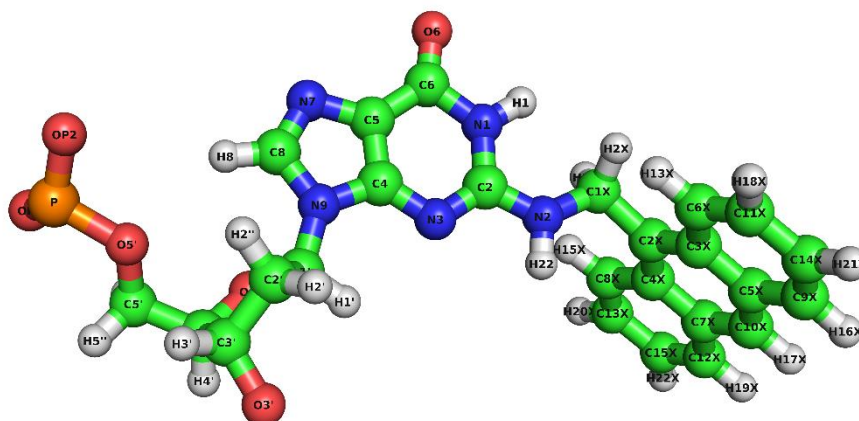
P	1
---	---

```

O5'      1
@<TRIPOS>RESIDUECONNECT
1        P      O5'      0      0      0      0

```

Figure S16. Cartesian coordinates and RESP charges calculated for the *N*²-Naph-dG nucleotide in SYBL Mol2 format. Since it was not feasible to directly optimize and calculate charges for the molecular fragments, RESP charges were derived via two-step process. The *N*²-Naph modified dG and dimethyl phosphate were optimized and partial charges were calculated separately for both molecules by the RED Dev. Server (<http://upjv.q4md-forcefieldtools.org/REDServer-Development/>). The charges were fitted into single molecular fragment by using PyRED program on the RED Dev. Server. A theory level of HF/6-31G(d) and Gaussian 09 Revision E.01. was used for all calculations by the server. Carbon atoms are represented using green, nitrogen atoms using blue, oxygen atoms using red, phosphorous using brown red and hydrogens using grey colour. Electronic energy for the modified nucleoside: -1379.18647701 (RHF). Electronic energy for the dimethyl phosphate: -719.520866767 (RHF). Detailed procedure for the derivation of partial charges for modified nucleotide fragments can be found in the tutorial at: <http://upjv.q4md-forcefieldtools.org/Tutorial/Tutorial-4.php>

Cartesian coordinates and RESP charges calculated for the N^2 -Anth-dG nucleotide

```
@<TRIPOS>MOLECULE
aD
      58      63      1      0      1
SMALL
USER_CHARGES
@<TRIPOS>ATOM
  1      O3'      -2.68048      -0.796807      -0.616424      OS      1      aD      -0.4874      0      ****
  2      P      -4.116314      -1.572411      -0.434573      P      1      aD      1.2122      0      ****
  3      O1P      -4.396031      -2.334899      -1.660214      O2      1      aD      -0.7919      0      ****
  4      O2P      -5.03431      -0.60161      0.179148      O2      1      aD      -0.7919      0      ****
  5      O5'      1.407456      0.0000      0.00000      OS      1      aD      -0.5209      0      ****
  6      C1      -0.496378      1.787308      -1.479338      CT      1      aD      0.1603      0      ****
  7      H2      0.468453      2.198479      -1.733654      H2      1      aD      0.1453      0      ****
  8      C7      -0.571496      1.411666      0.00000      CT      1      aD      -0.0925      0      ****
  9      H2'      -0.014285      2.098231      0.627092      HC      1      aD      0.0572      0      ****
  10     H2''     -1.602188      1.395463      0.327278      HC      1      aD      0.0572      0      ****
  11     C3      0.00000      0.00000      0.00000      CT      1      aD      0.0474      0      ****
  12     H3      -0.379507      -0.604443      0.815835      H1      1      aD      0.1193      0      ****
  13     C9      -0.443662      -0.538755      -1.367407      CT      1      aD      0.1408      0      ****
  14     H4      0.374998      -1.094563      -1.807881      H1      1      aD      0.12      0      ****
  15     C10     -1.665852      -1.433341      -1.327977      CI      1      aD      -0.0281      0      ****
  16     H5'     -1.976331      -1.637572      -2.34792      H1      1      aD      0.0829      0      ****
  17     H5''     -1.398537      -2.376399      -0.855284      H1      1      aD      0.0829      0      ****
  18     O4      -0.712711      0.589994      -2.170772      OS      1      aD      -0.3903      0      ****
  19     C6      -2.557545      5.871019      -3.147496      C      1      aD      0.455      0      ****
  20     O6      -3.491538      6.540451      -3.477678      O      1      aD      -0.5393      0      ****
  21     C5      -2.478646      4.531223      -2.65084      CB      1      aD      0.1472      0      ****
  22     N7      -3.466591      3.599525      -2.409566      NB      1      aD      -0.556      0      ****
  23     C8      -2.843703      2.572301      -1.975448      CK      1      aD      0.0931      0      ****
  24     H8      -3.285115      1.639536      -1.695758      H5      1      aD      0.1851      0      ****
```

25	N9	-1.477721	2.752954	-1.902825	N*	1	aD	-0.0138	0	****
26	C4	-1.247004	4.01412	-2.345967	CB	1	aD	0.2866	0	****
27	N3	-0.033559	4.603832	-2.442752	NC	1	aD	-0.7132	0	****
28	C2	-0.093444	5.817161	-2.892301	CA	1	aD	0.5584	0	****
29	N1	-1.256934	6.430869	-3.231813	NA	1	aD	-0.3592	0	****
30	H1	-1.252228	7.370226	-3.563279	H	1	aD	0.2829	0	****
31	N2	1.049163	6.524658	-3.023231	N2	1	aD	-0.4567	0	****
32	H22	1.873492	5.999192	-2.837078	H	1	aD	0.3339	0	****
33	C1X	1.189213	7.811594	-3.684593	CT	1	aD	-0.0168	0	****
34	H1X	0.779144	7.773609	-4.68574	H1	1	aD	0.0913	0	****
35	H2X	0.618841	8.562178	-3.15318	H1	1	aD	0.0913	0	****
36	C2X	2.645297	8.231107	-3.739998	CA	1	aD	-0.1341	0	****
37	C3X	3.203566	8.951948	-2.673335	CB	1	aD	0.1123	0	****
38	C4X	3.428019	7.901453	-4.856718	CB	1	aD	0.1123	0	****
39	C5X	4.551287	9.408753	-2.761371	CB	1	aD	0.0891	0	****
40	C6X	2.483521	9.25348	-1.46253	CA	1	aD	-0.2138	0	****
41	H13X	1.483758	8.888576	-1.333554	HA	1	aD	0.1424	0	****
42	C7X	4.77602	8.362799	-4.931047	CB	1	aD	0.0891	0	****
43	C8X	2.948239	7.095399	-5.951015	CA	1	aD	-0.2138	0	****
44	H15X	1.955598	6.691258	-5.921265	HA	1	aD	0.1424	0	****
45	C9X	5.114161	10.16676	-1.680062	CA	1	aD	-0.1696	0	****
46	H16X	6.128945	10.51078	-1.774711	HA	1	aD	0.1341	0	****
47	C10X	5.299348	9.113311	-3.890388	CA	1	aD	-0.2771	0	****
48	H17X	6.314014	9.466101	-3.955551	HA	1	aD	0.1763	0	****
49	C11X	3.052968	9.960476	-0.465515	CA	1	aD	-0.1102	0	****
50	H18X	2.495026	10.16527	0.430464	HA	1	aD	0.1397	0	****
51	C12X	5.572155	8.042811	-6.082222	CA	1	aD	-0.1696	0	****
52	H19X	6.581767	8.412042	-6.119911	HA	1	aD	0.1341	0	****
53	C13X	3.733088	6.808742	-7.009055	CA	1	aD	-0.1102	0	****
54	H20X	3.350896	6.198443	-7.807404	HA	1	aD	0.1397	0	****
55	C14X	4.396042	10.43806	-0.574438	CA	1	aD	-0.1449	0	****
56	H21X	4.823612	11.0046	0.232774	HA	1	aD	0.1422	0	****
57	C15X	5.07447	7.297557	-7.086167	CA	1	aD	-0.1449	0	****
58	H22X	5.676389	7.05913	-7.944143	HA	1	aD	0.1422	0	****

@<TRIPOS>BOND

1	1	2	1
2	1	15	1
3	2	3	1
4	2	4	1
5	5	11	1
6	6	7	1
7	6	8	1
8	6	18	1
9	6	25	1
10	8	9	1
11	8	10	1

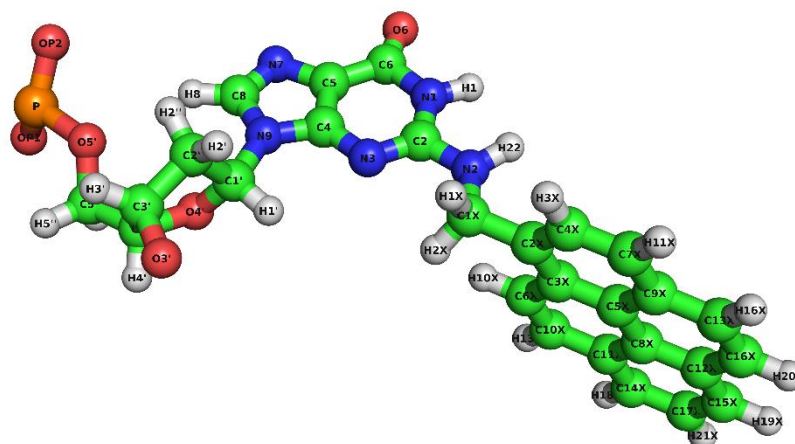
12	8	11	1
13	11	12	1
14	11	13	1
15	13	14	1
16	13	15	1
17	13	18	1
18	15	16	1
19	15	17	1
20	19	20	1
21	19	21	1
22	19	29	1
23	21	22	1
24	21	26	1
25	22	23	1
26	23	24	1
27	23	25	1
28	25	26	1
29	26	27	1
30	27	28	1
31	28	29	1
32	28	31	1
33	29	30	1
34	31	32	1
35	31	33	1
36	33	34	1
37	33	35	1
38	33	36	1
39	36	37	1
40	36	38	1
41	37	39	1
42	37	40	1
43	38	42	1
44	38	43	1
45	39	45	1
46	39	47	1
47	40	41	1
48	40	49	1
49	42	47	1
50	42	51	1
51	43	44	1
52	43	53	1
53	45	46	1
54	45	55	1
55	47	48	1
56	49	50	1
57	49	55	1

```

58      51      52      1
59      51      57      1
60      53      54      1
61      53      57      1
62      55      56      1
63      57      58      1
@<TRIPOS>SUBSTRUCTURE
      1      aD      1      ****      0      ****      ****
@<TRIPOS>HEADTAIL
P      1
O5'    1
@<TRIPOS>RESIDUECONNECT
1      P      O5'    0      0      0      0

```

Figure S17. Cartesian coordinates and RESP charges calculated for the *N*²-Anth-dG nucleotide in SYBL Mol2 format. Since it was not feasible to directly optimize and calculate charges for the molecular fragments, RESP charges were derived via two-step process. The *N*²-Anth-dG and dimethyl phosphate were optimized and partial charges were calculated separately for both molecules by the RED Dev. Server (<http://upjv.q4md-forcefieldtools.org/REDServer-Development/>). The charges were fitted into single molecular fragment by using PyRED program on the RED Dev. Server. A theory level of HF/6-31G(d) and Gaussian 09 Revision E.01. was used for all calculations by the server. Carbon atoms are represented using green, nitrogen atoms using blue, oxygen atoms using red, phosphorous using brown red and hydrogens using grey colour. Electronic energy for the modified nucleoside: -1531.82187224 (RHF). Electronic energy for the dimethyl phosphate: -719.520866767 (RHF). Detailed procedure for the derivation of partial charges for modified nucleotide fragments can be found in the tutorial at: <http://upjv.q4md-forcefieldtools.org/Tutorial/Tutorial-4.php>

Cartesian coordinates and RESP charges calculated for the N^2 -Pyre-dG nucleotide

@<TRIPOS>MOLECULE

pD

60 66 1 0 1

SMALL

USER_CHARGES

@<TRIPOS>ATOM

1	O3'	-2.681156	-0.791215	-0.615972	OS	1	pD	-0.4898	0	****
2	P	-4.118943	-1.563114	-0.433784	P	1	pD	1.2125	0	****
3	O1P	-4.400874	-2.324942	-1.659327	O2	1	pD	-0.7918	0	****
4	O2P	-5.034311	-0.589923	0.180078	O2	1	pD	-0.7918	0	****
5	O5'	1.407421	0.00000	0.00000	OS	1	pD	-0.5215	0	****
6	C1'	-0.495142	1.787407	-1.479458	CT	1	pD	0.1472	0	****
7	H1'	0.469754	2.19884	-1.733225	H2	1	pD	0.1324	0	****
8	C2'	-0.570782	1.412075	0.00000	CT	1	pD	-0.1132	0	****
9	H2'	-0.012779	2.09796	0.627113	HC	1	pD	0.0599	0	****
10	H2''	-1.601304	1.396675	0.327615	HC	1	pD	0.0599	0	****
11	C3'	0.0000	0.00000	0.00000	CT	1	pD	0.0717	0	****
12	H3'	-0.379802	-0.604131	0.815848	H1	1	pD	0.1157	0	****
13	C4'	-0.444029	-0.538697	-1.367228	CT	1	pD	0.1242	0	****
14	H4'	0.373593	-1.09623	-1.807292	H1	1	pD	0.123	0	****
15	C5'	-1.668316	-1.430394	-1.327702	CI	1	pD	-0.0162	0	****
16	H5'	-1.979423	-1.633877	-2.347556	H1	1	pD	0.0801	0	****
17	H5''	-1.403234	-2.373996	-0.854941	H1	1	pD	0.0801	0	****
18	O4'	-0.71023	0.590557	-2.171236	OS	1	pD	-0.3762	0	****
19	N2	1.042706	6.50573	-3.168799	N2	1	pD	-0.5248	0	****
20	H22	0.953403	7.492275	-3.265402	H	1	pD	0.3304	0	****
21	C6	-2.555656	5.859584	-3.187515	C	1	pD	0.6265	0	****
22	O6	-3.493706	6.51716	-3.527235	O	1	pD	-0.5682	0	****
23	C5	-2.476509	4.525317	-2.666029	CB	1	pD	0.0963	0	****
24	N7	-3.465074	3.601329	-2.406141	NB	1	pD	-0.5624	0	****
25	C8	-2.842247	2.576514	-1.963734	CK	1	pD	0.1123	0	****
26	H8	-3.285296	1.648837	-1.670071	H5	1	pD	0.1858	0	****

27	N9	-1.476443	2.753393	-1.903398	N*	1	pD	-0.006	0	****
28	C4	-1.246511	4.007956	-2.365859	CB	1	pD	0.2401	0	****
29	N3	-0.02938	4.597766	-2.474402	NC	1	pD	-0.6008	0	****
30	C2	-0.089597	5.79572	-2.952118	CA	1	pD	0.6682	0	****
31	N1	-1.257066	6.412189	-3.286827	NA	1	pD	-0.6425	0	****
32	H1	-1.236697	7.322463	-3.694458	H	1	pD	0.3601	0	****
33	C1X	2.338341	6.022359	-2.70864	CT	1	pD	0.0208	0	****
34	H1X	2.341212	5.920149	-1.629324	H1	1	pD	0.0716	0	****
35	H2X	2.492147	5.033049	-3.114031	H1	1	pD	0.0716	0	****
36	C2X	3.437346	6.973133	-3.131617	CA	1	pD	0.0423	0	****
37	C3X	3.859147	7.074292	-4.466645	CB	1	pD	0.0353	0	****
38	C4X	4.042056	7.766681	-2.168766	CA	1	pD	-0.2291	0	****
39	H3X	3.720456	7.686508	-1.145126	HA	1	pD	0.1699	0	****
40	C5X	4.89267	7.979067	-4.799794	CB	1	pD	0.0435	0	****
41	C6X	3.279369	6.280369	-5.530464	CA	1	pD	-0.2718	0	****
42	H10X	2.48625	5.597875	-5.295772	HA	1	pD	0.1872	0	****
43	C7X	5.052243	8.65583	-2.489022	CA	1	pD	-0.1985	0	****
44	H11X	5.501573	9.25646	-1.718126	HA	1	pD	0.1573	0	****
45	C8X	5.343342	8.093684	-6.156764	CB	1	pD	0.0908	0	****
46	C9X	5.492249	8.776747	-3.799049	CB	1	pD	0.0455	0	****
47	C10X	3.70436	6.388535	-6.796348	CA	1	pD	-0.1532	0	****
48	H13X	3.25805	5.78701	-7.568452	HA	1	pD	0.1456	0	****
49	C11X	4.757118	7.302921	-7.165732	CB	1	pD	0.0351	0	****
50	C12X	6.376147	8.998342	-6.488165	CB	1	pD	0.048	0	****
51	C13X	6.547083	9.694036	-4.170043	CA	1	pD	-0.1617	0	****
52	H16X	6.992321	10.29429	-3.396438	HA	1	pD	0.1415	0	****
53	C14X	5.205707	7.425931	-8.478153	CA	1	pD	-0.1512	0	****
54	H18X	4.758116	6.822604	-9.24812	HA	1	pD	0.1456	0	****
55	C15X	6.798085	9.09404	-7.809276	CA	1	pD	-0.1491	0	****
56	H19X	7.584514	9.782507	-8.06397	HA	1	pD	0.1442	0	****
57	C16X	6.963623	9.798798	-5.437118	CA	1	pD	-0.2098	0	****
58	H20X	7.749415	10.48524	-5.698695	HA	1	pD	0.1522	0	****
59	C17X	6.216887	8.313789	-8.794928	CA	1	pD	-0.2065	0	****
60	H21X	6.55439	8.399435	-9.812137	HA	1	pD	0.1617	0	****

@<TRIPOS>BOND

1	1	2	1
2	1	15	1
3	2	3	1
4	2	4	1
5	5	11	1
6	6	7	1
7	6	8	1
8	6	18	1
9	6	27	1
10	8	9	1
11	8	10	1

12	8	11	1
13	11	12	1
14	11	13	1
15	13	14	1
16	13	15	1
17	13	18	1
18	15	16	1
19	15	17	1
20	19	20	1
21	19	30	1
22	19	33	1
23	21	22	1
24	21	23	1
25	21	31	1
26	23	24	1
27	23	28	1
28	24	25	1
29	25	26	1
30	25	27	1
31	27	28	1
32	28	29	1
33	29	30	1
34	30	31	1
35	31	32	1
36	33	34	1
37	33	35	1
38	33	36	1
39	36	37	1
40	36	38	1
41	37	40	1
42	37	41	1
43	38	39	1
44	38	43	1
45	40	45	1
46	40	46	1
47	41	42	1
48	41	47	1
49	43	44	1
50	43	46	1
51	45	49	1
52	45	50	1
53	46	51	1
54	47	48	1
55	47	49	1
56	49	53	1
57	50	55	1


```

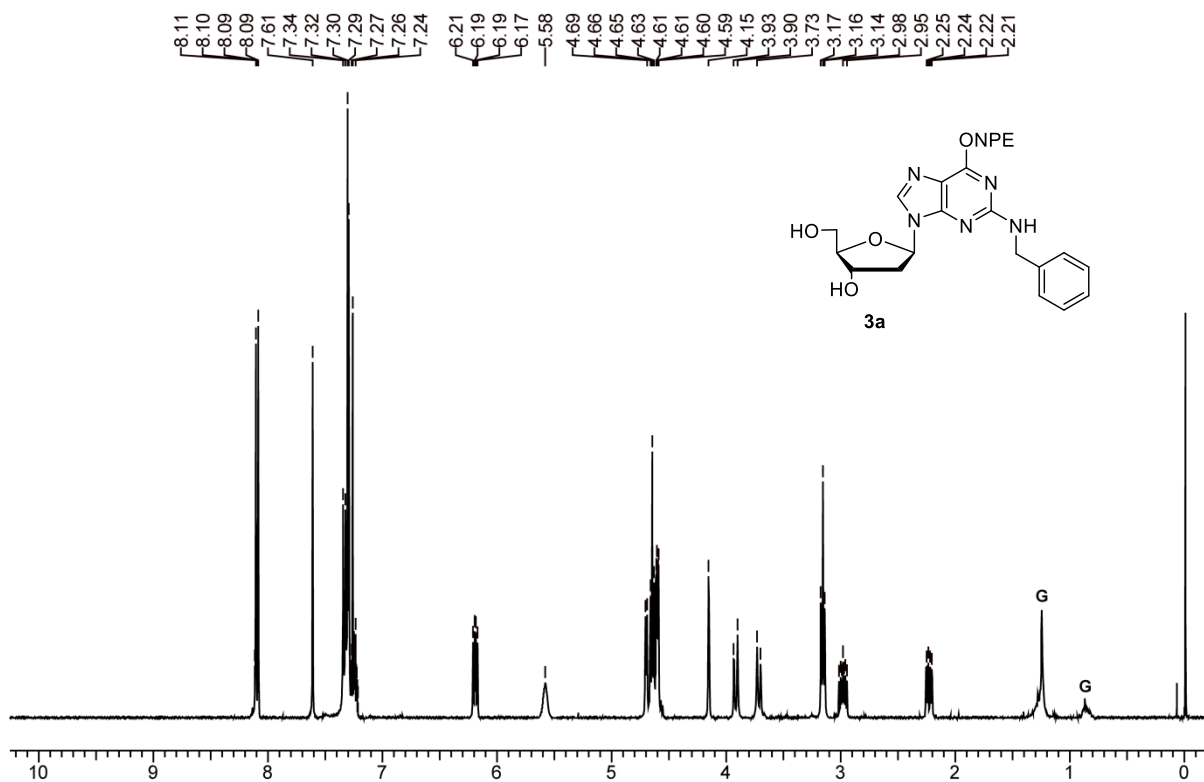
58      50      57      1
59      51      52      1
60      51      57      1
61      53      54      1
62      53      59      1
63      55      56      1
64      55      59      1
65      57      58      1
66      59      60      1
@<TRIPOS>SUBSTRUCTURE
          1  pD          1  ****          0  ****  ****
@<TRIPOS>HEADTAIL
P          1
O5'       1
@<TRIPOS>RESIDUECONNECT
          1  P          O5'          0          0          0          0

```

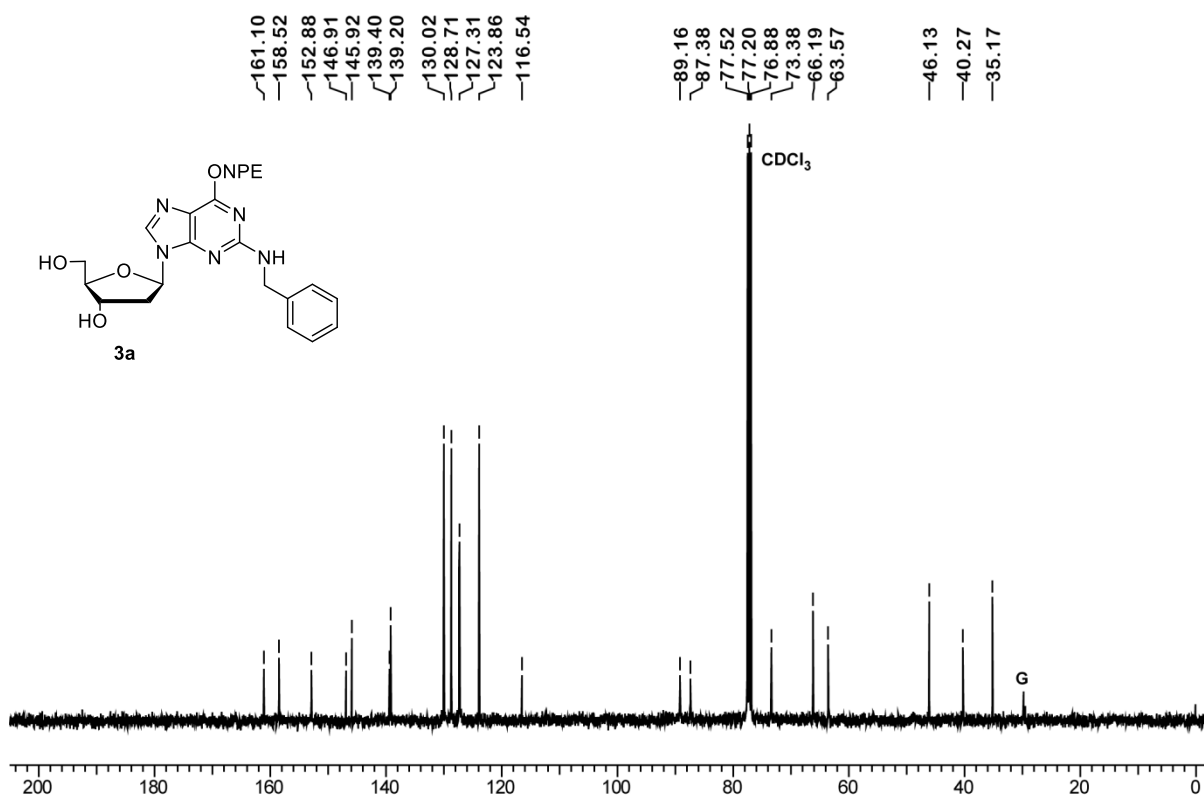
Figure S18. Cartesian coordinates and RESP charges calculated for the N^2 -Pyre-dG nucleotide in SYBL Mol2 format. Since it was not feasible to directly optimize and calculate charges for the molecular fragments, RESP charges were derived via two-step process. The N^2 -Pyre-dG and dimethyl phosphate were optimized and partial charges were calculated separately for both molecules by the RED Dev. Server (<http://upjv.q4md-forcefieldtools.org/REDServer-Development/>). The charges were fitted into single molecular fragment by using PyRED program on the RED Dev. Server. A theory level of HF/6-31G(d) and Gaussian 09 Revision E.01. was used for all calculations by the server. Carbon atoms are represented using green, nitrogen atoms using blue, oxygen atoms using red, phosphorous using brown red and hydrogens using grey colour. Electronic energy for the modified nucleoside: -1607.59876044 (RHF). Electronic energy for the dimethyl phosphate: -719.520866767 (RHF). Detailed procedure for the derivation of partial charges for modified nucleotide fragments can be found in the tutorial at: <http://upjv.q4md-forcefieldtools.org/Tutorial/Tutorial-4.php>

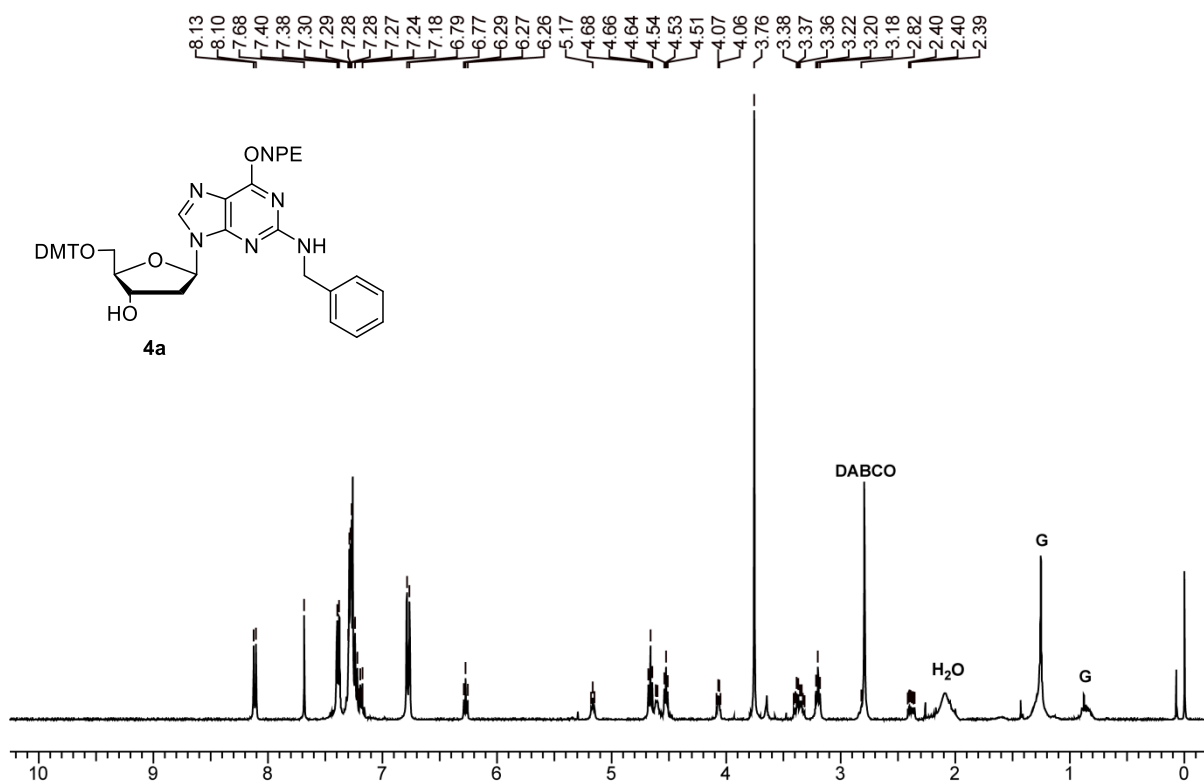
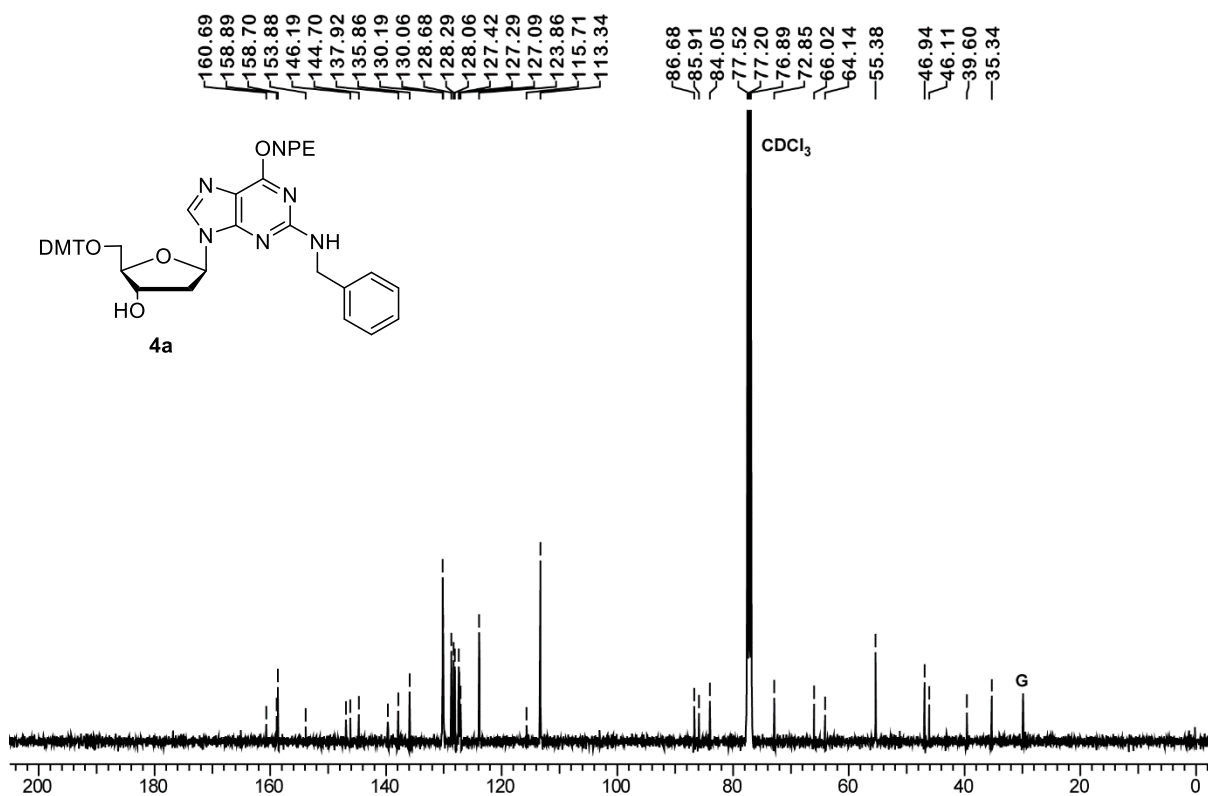
NMR spectra (^1H , ^{13}C , & ^{31}P) (G-Grease, Im-Impurity)

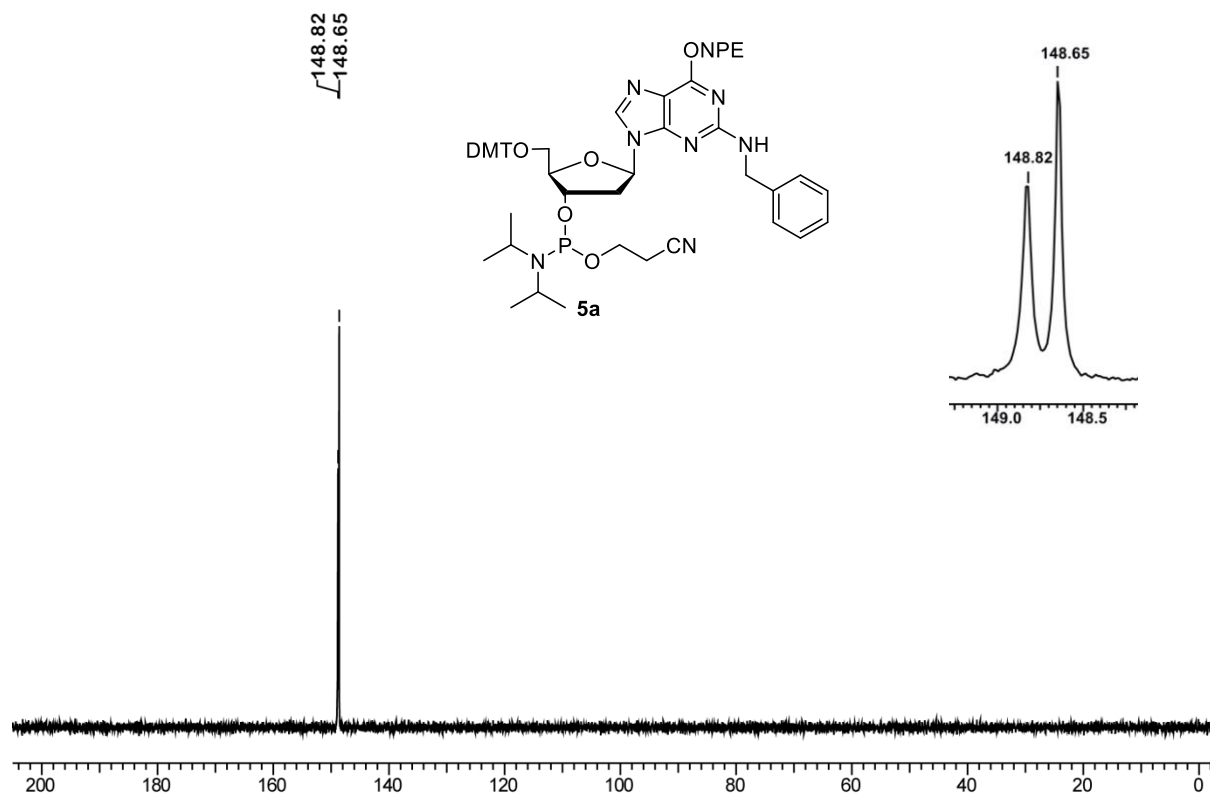
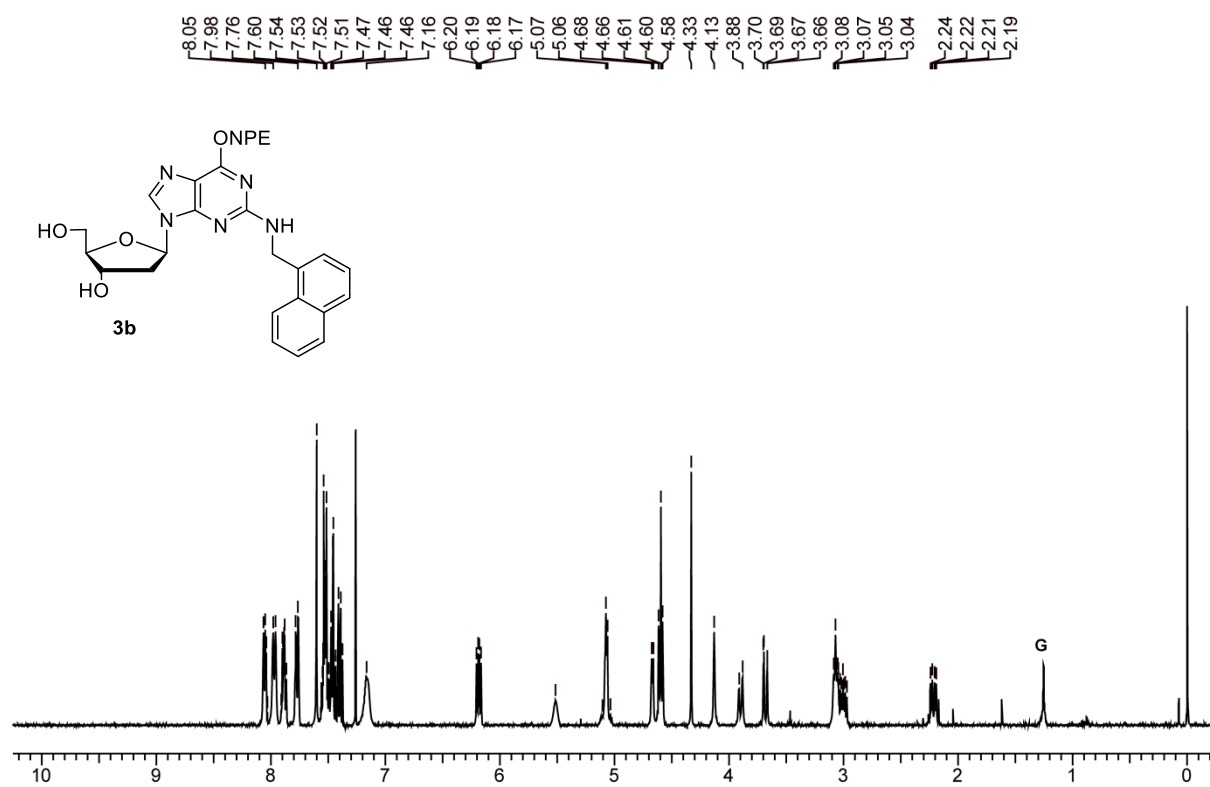
^1H NMR for Compound **3a**

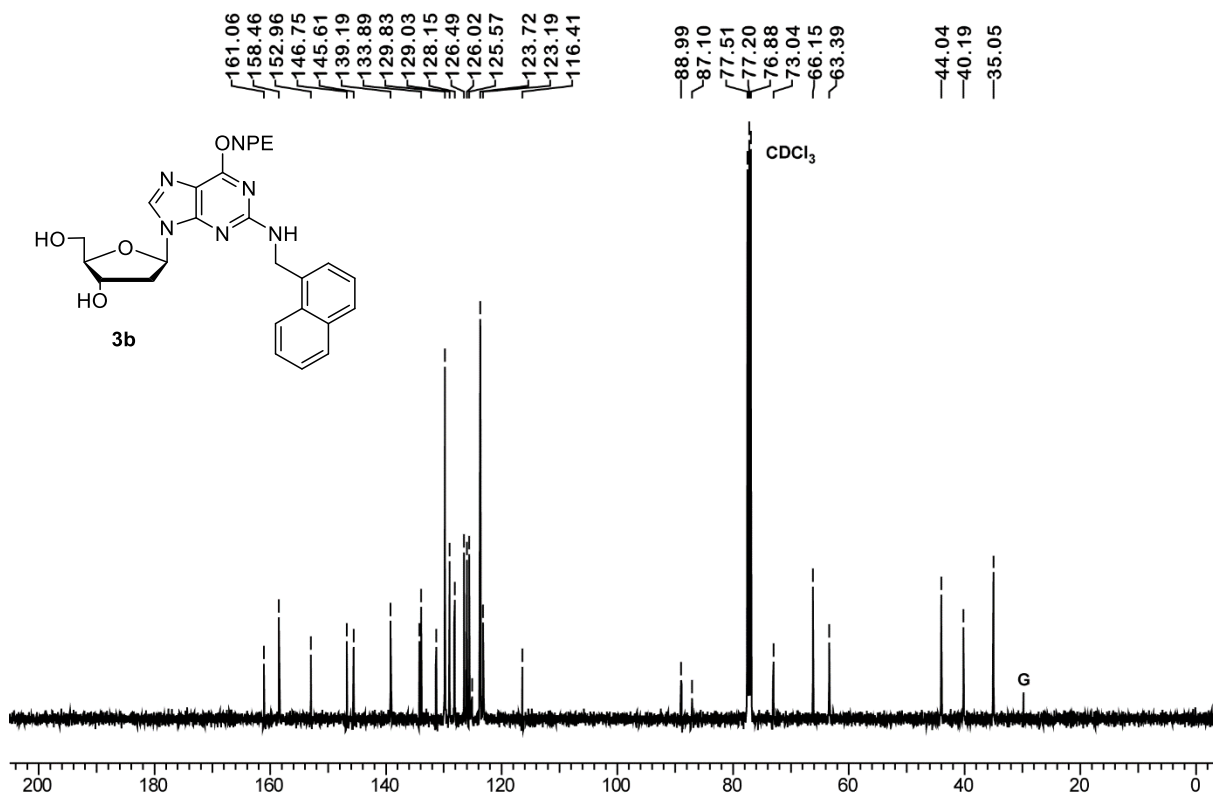
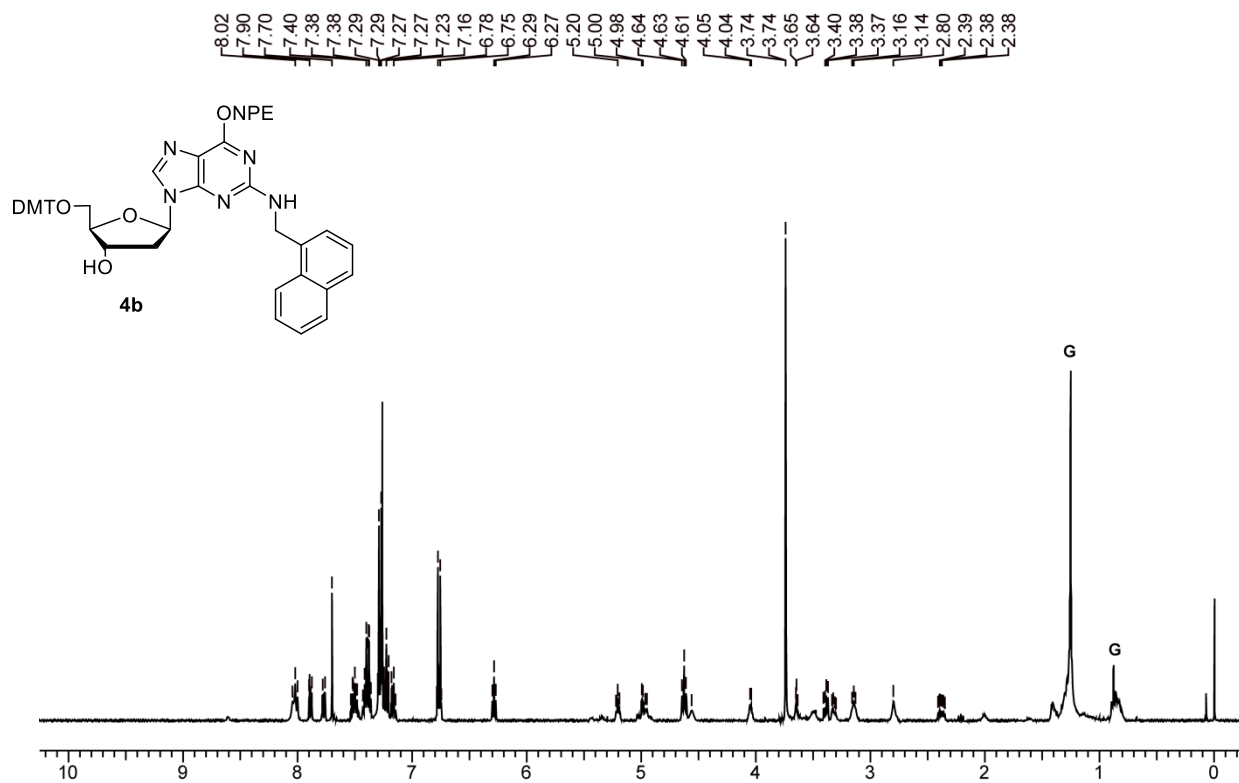


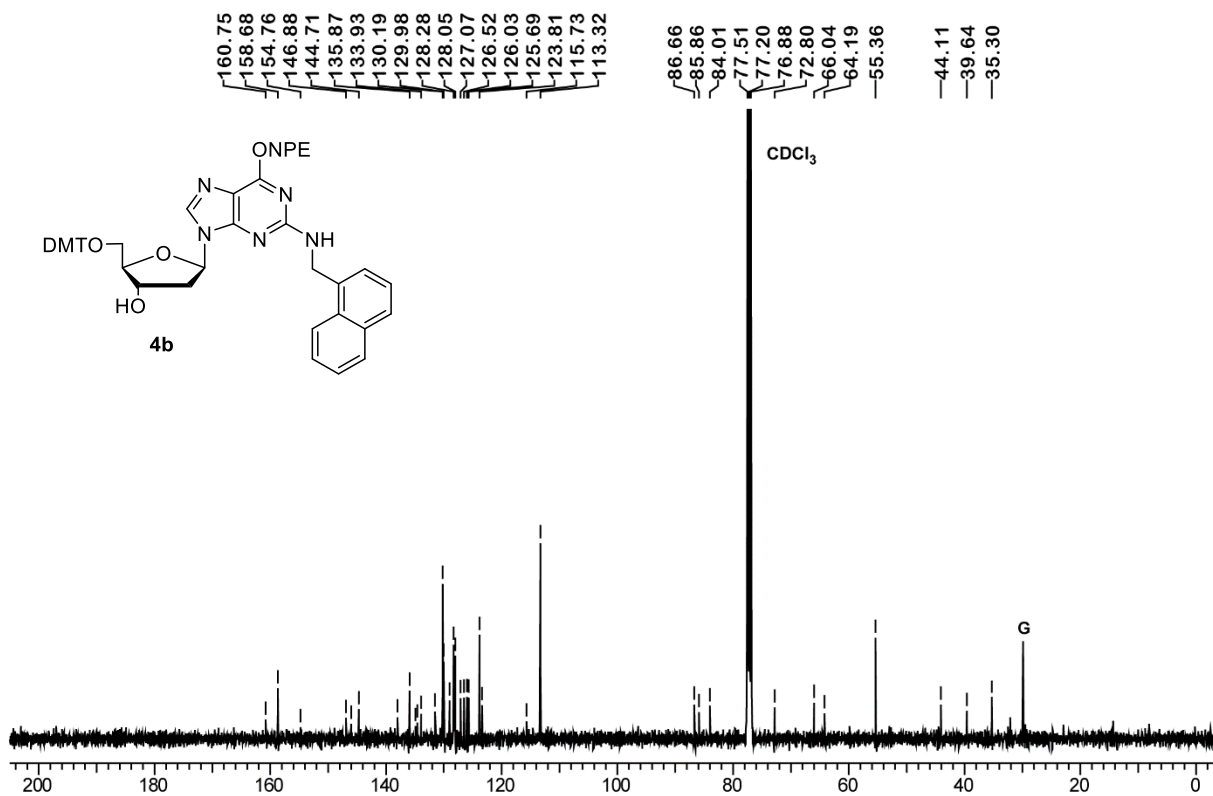
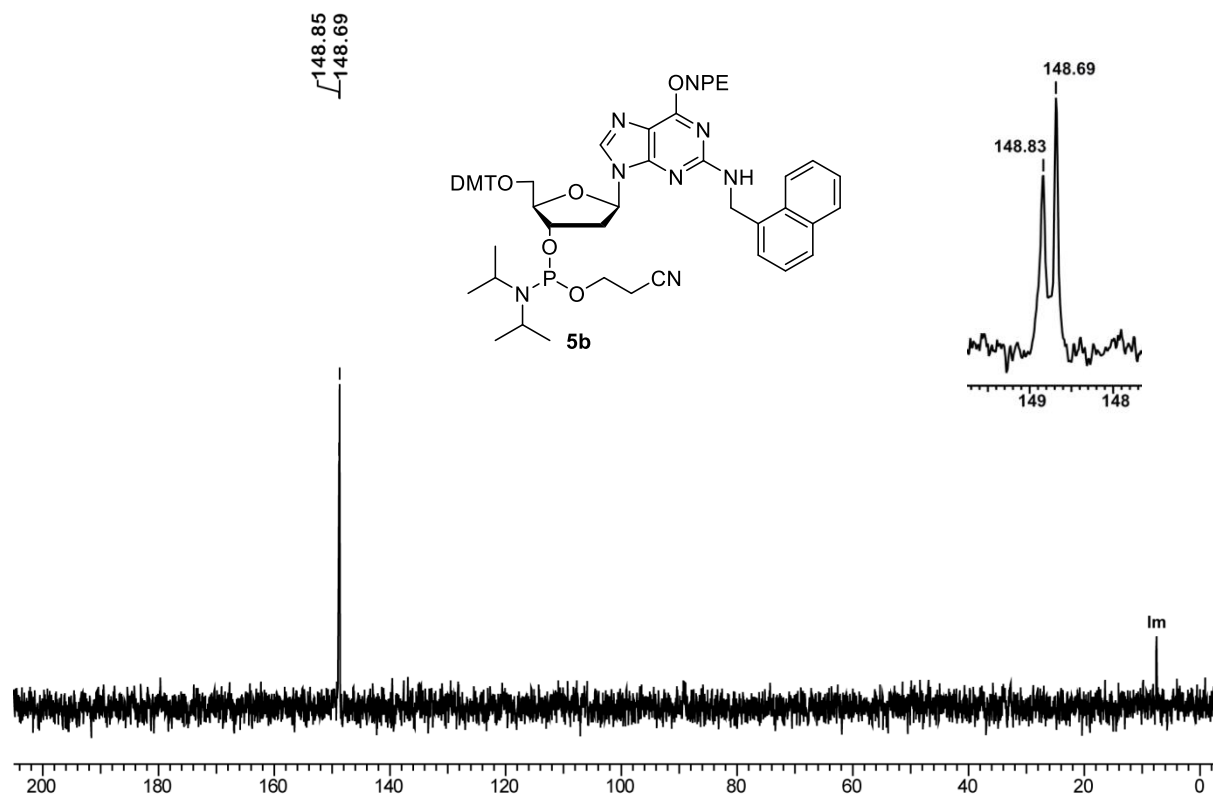
^{13}C NMR for Compound **3a**

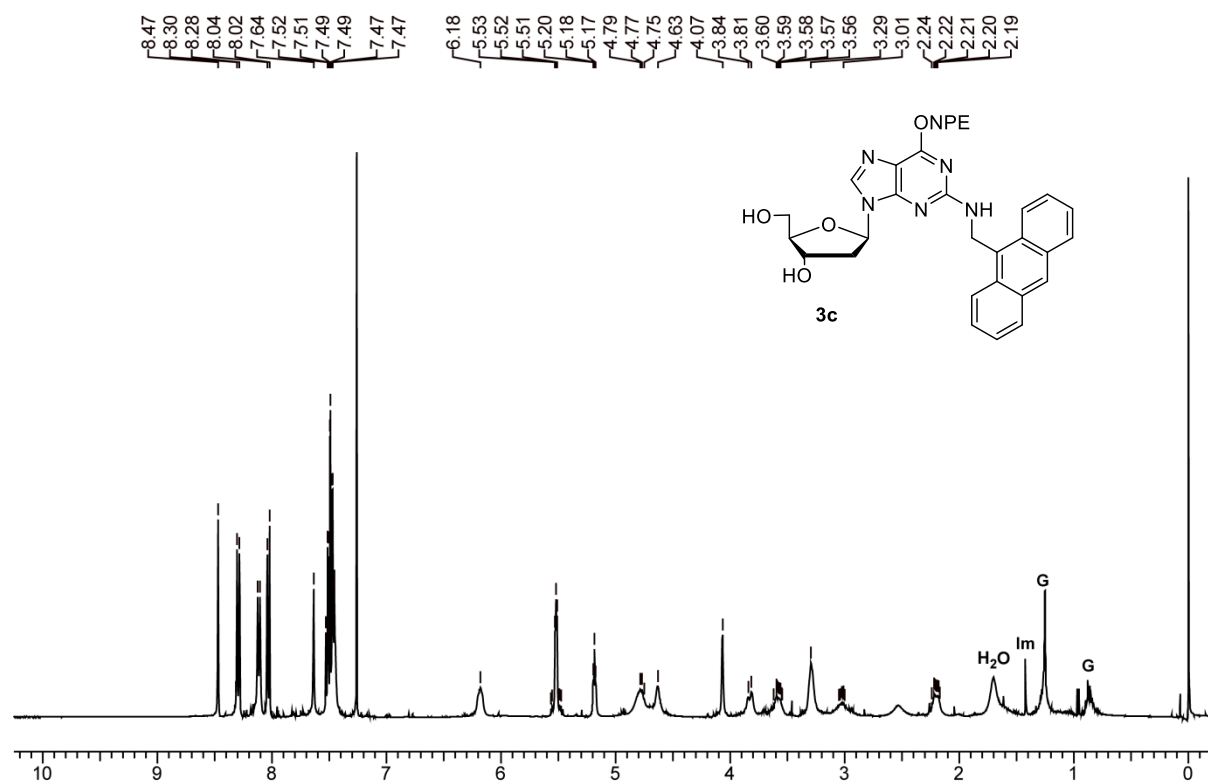
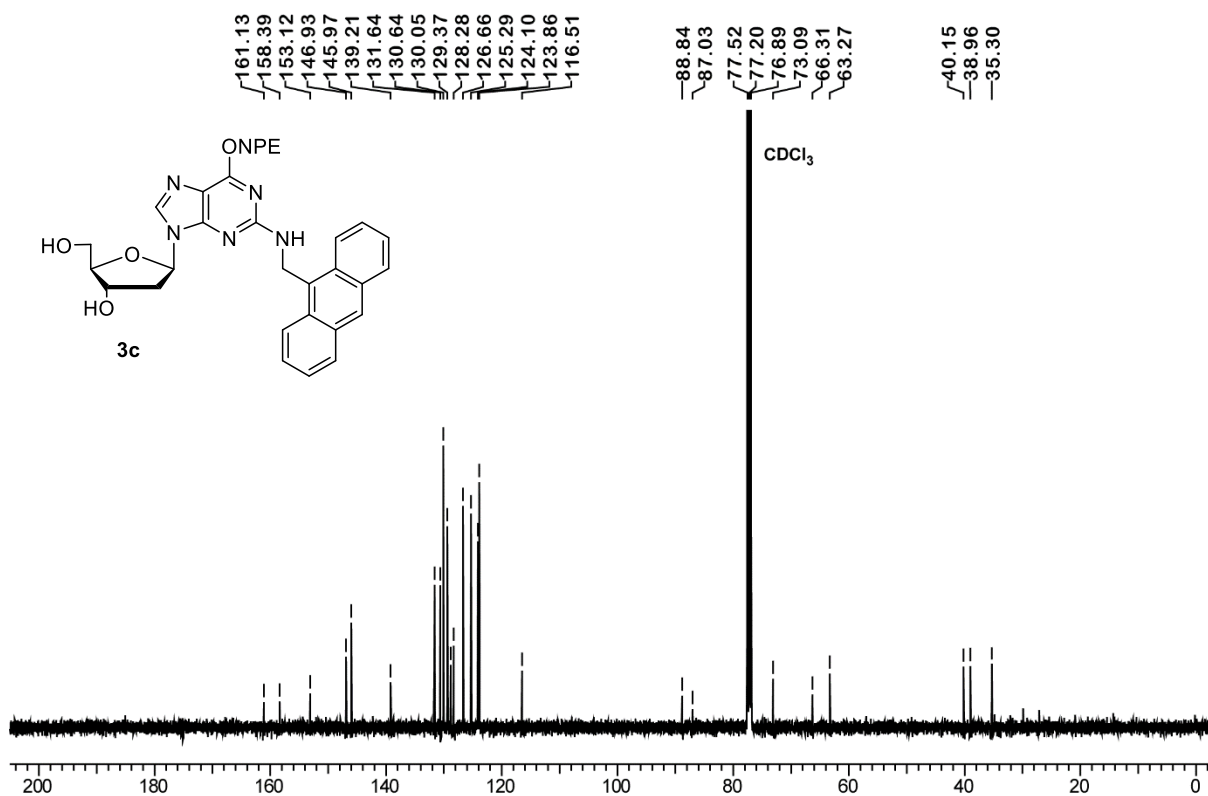


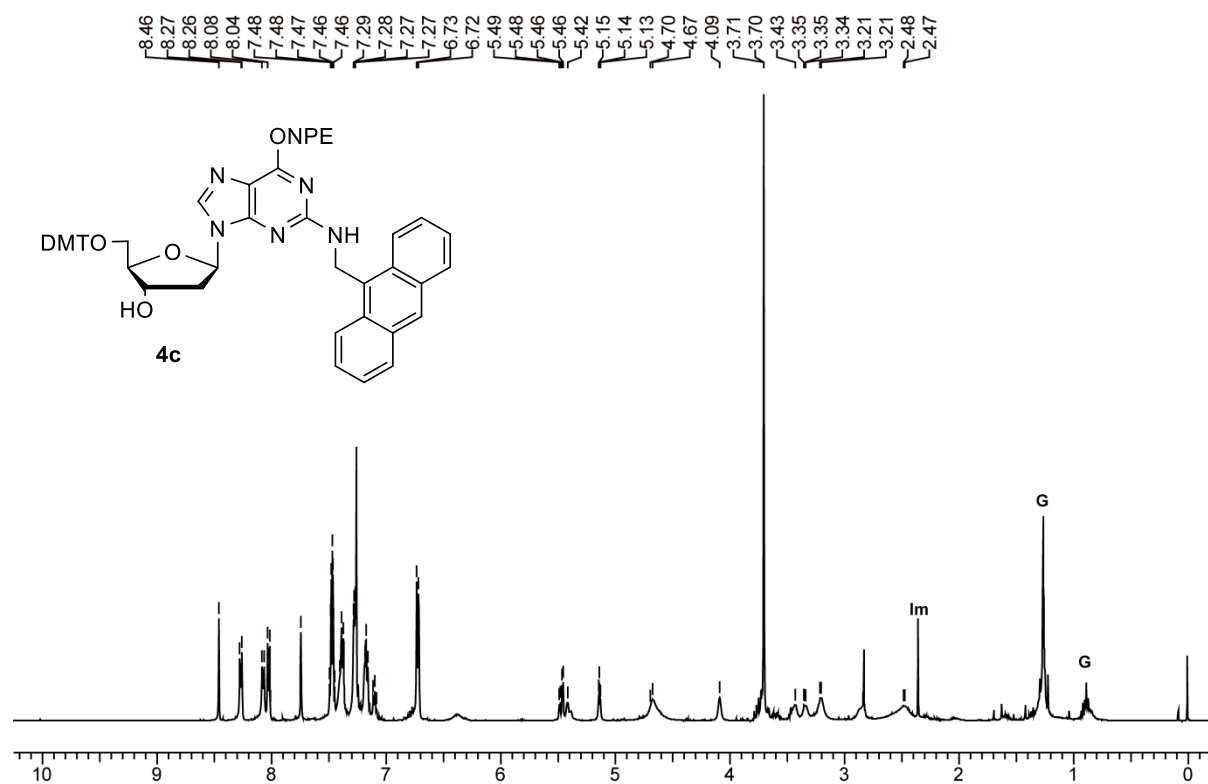
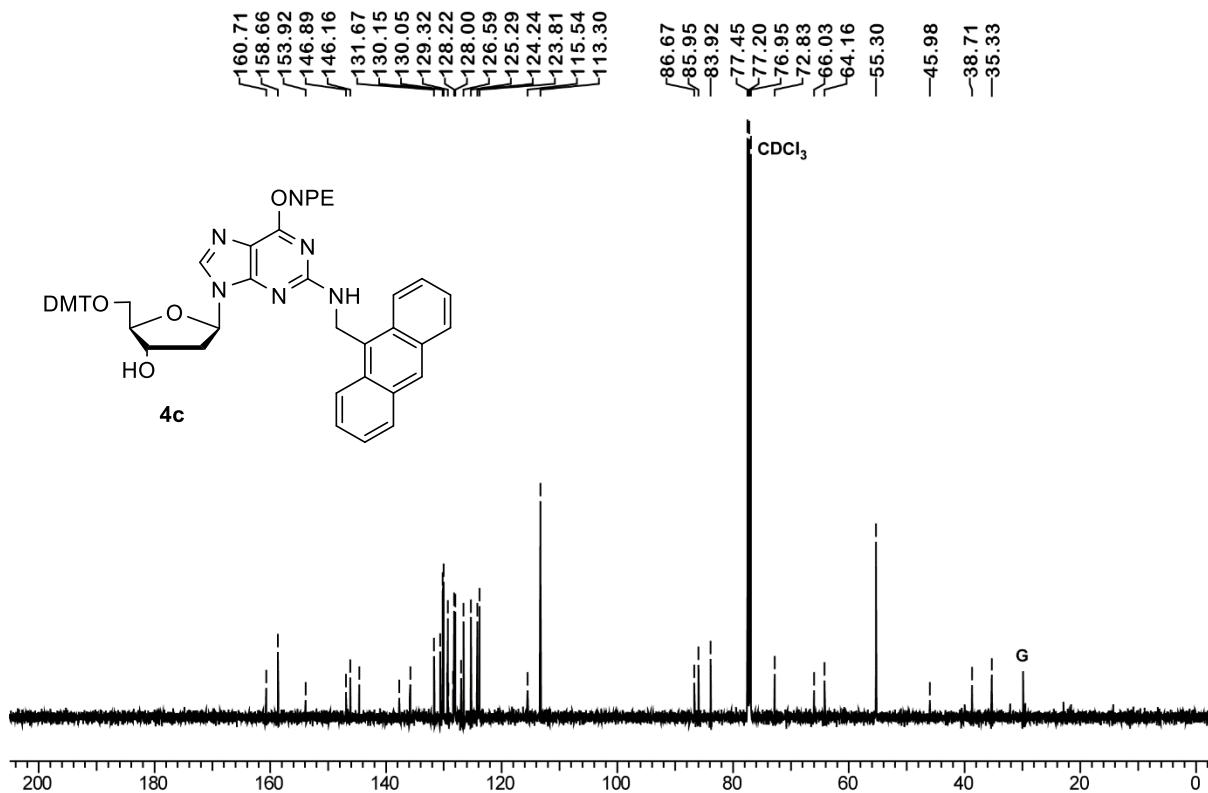
¹H NMR for Compound **4a**¹³C NMR for Compound **4a**

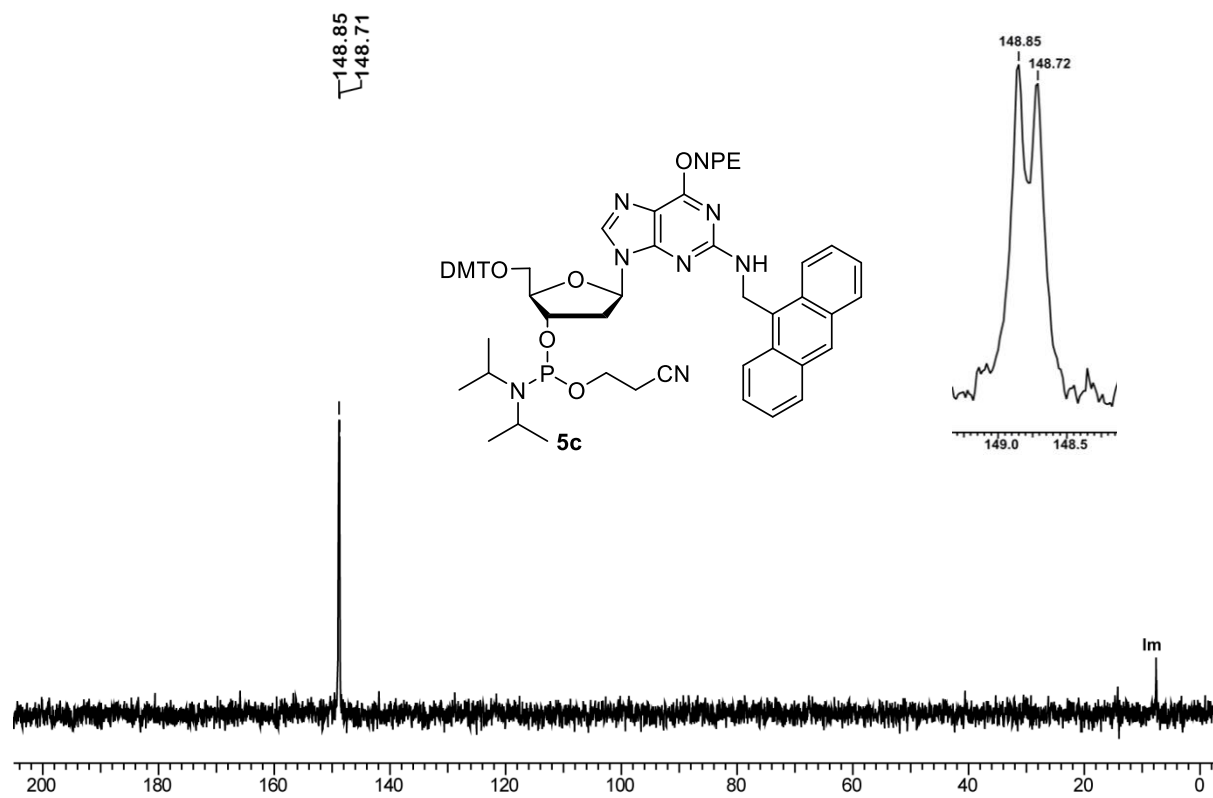
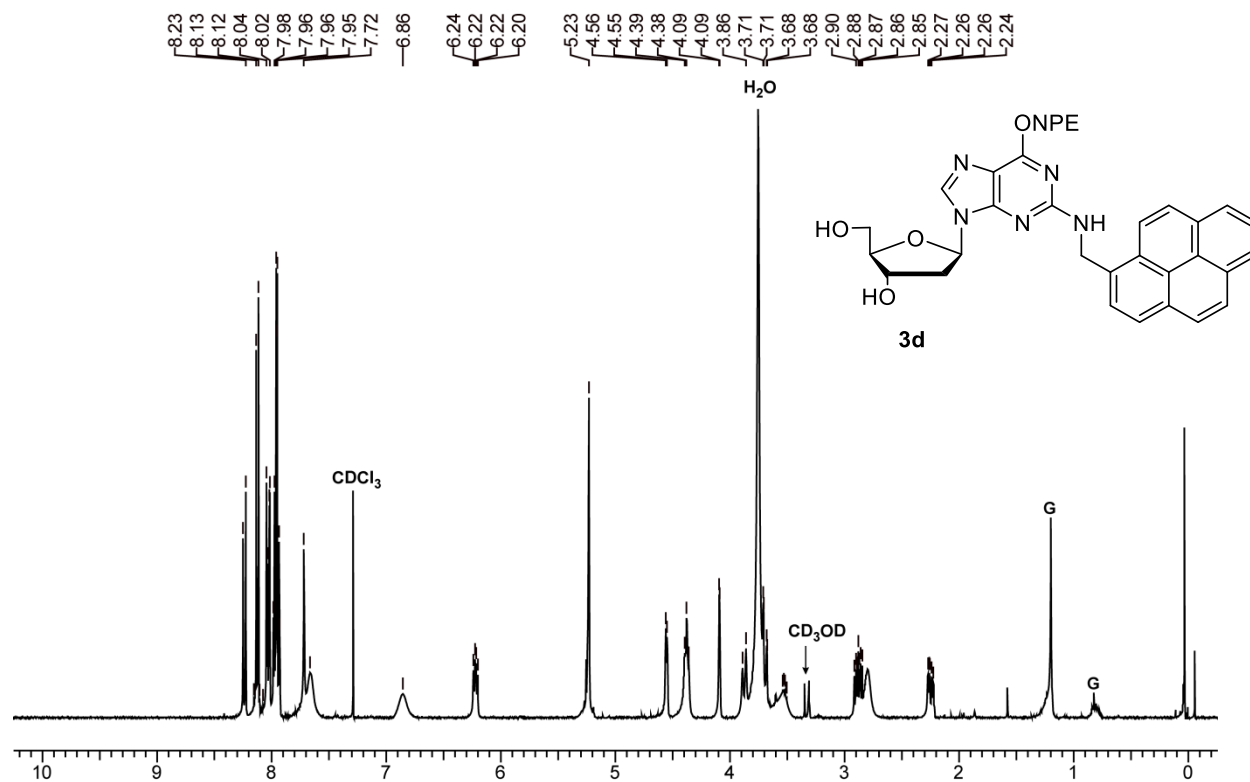
^{31}P NMR for Compound **5a** ^1H NMR for Compound **3b**

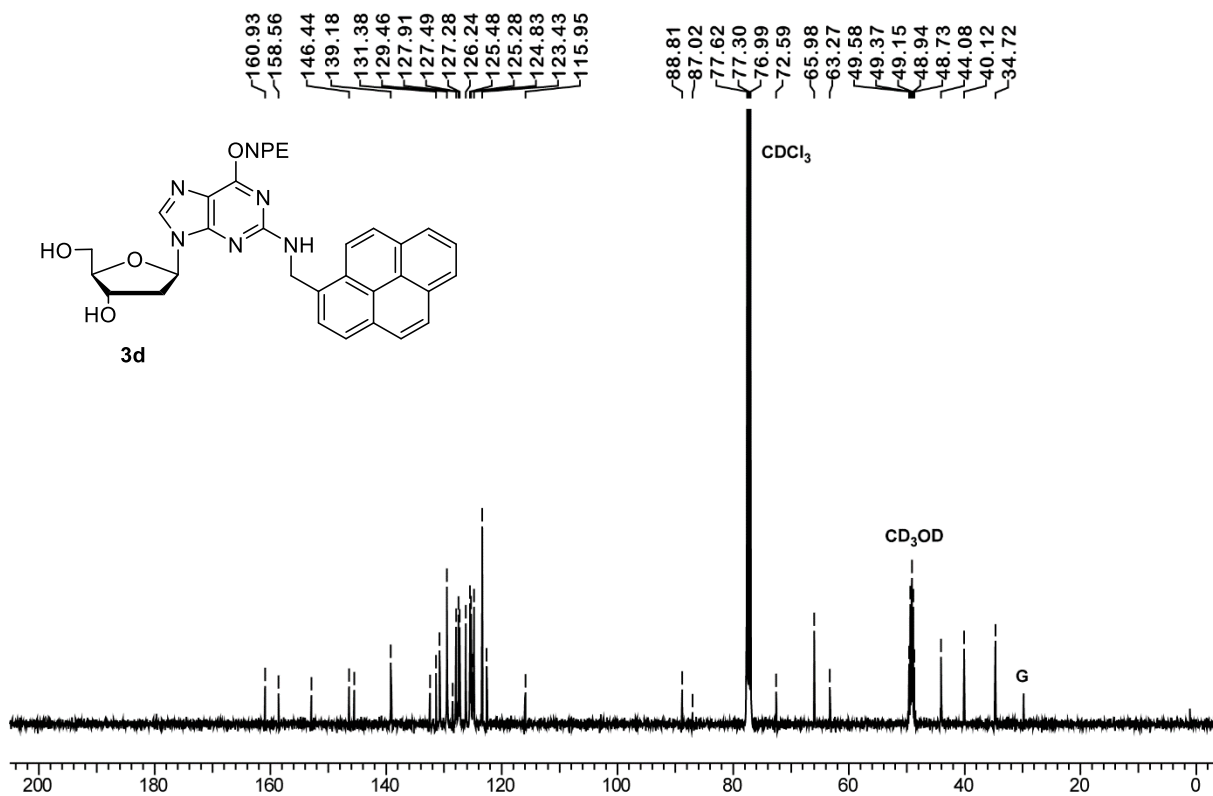
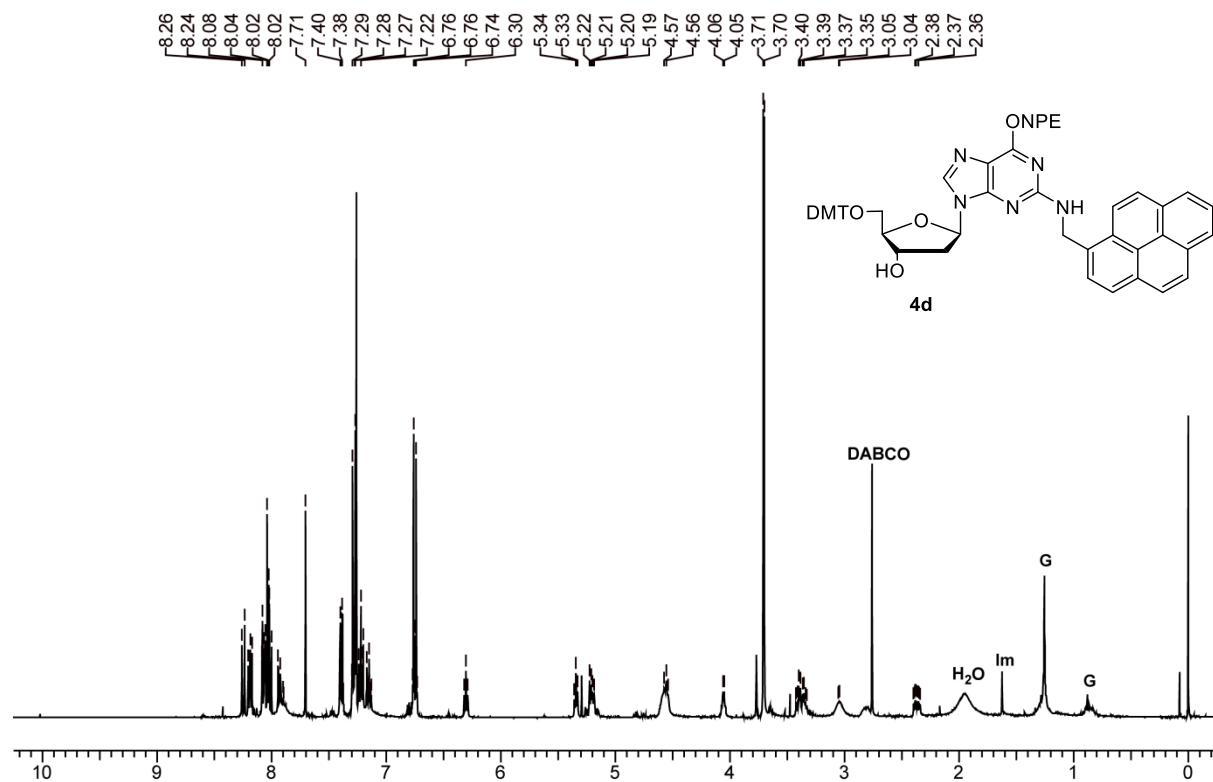
¹³C NMR for Compound **3b**¹H NMR for Compound **4b**

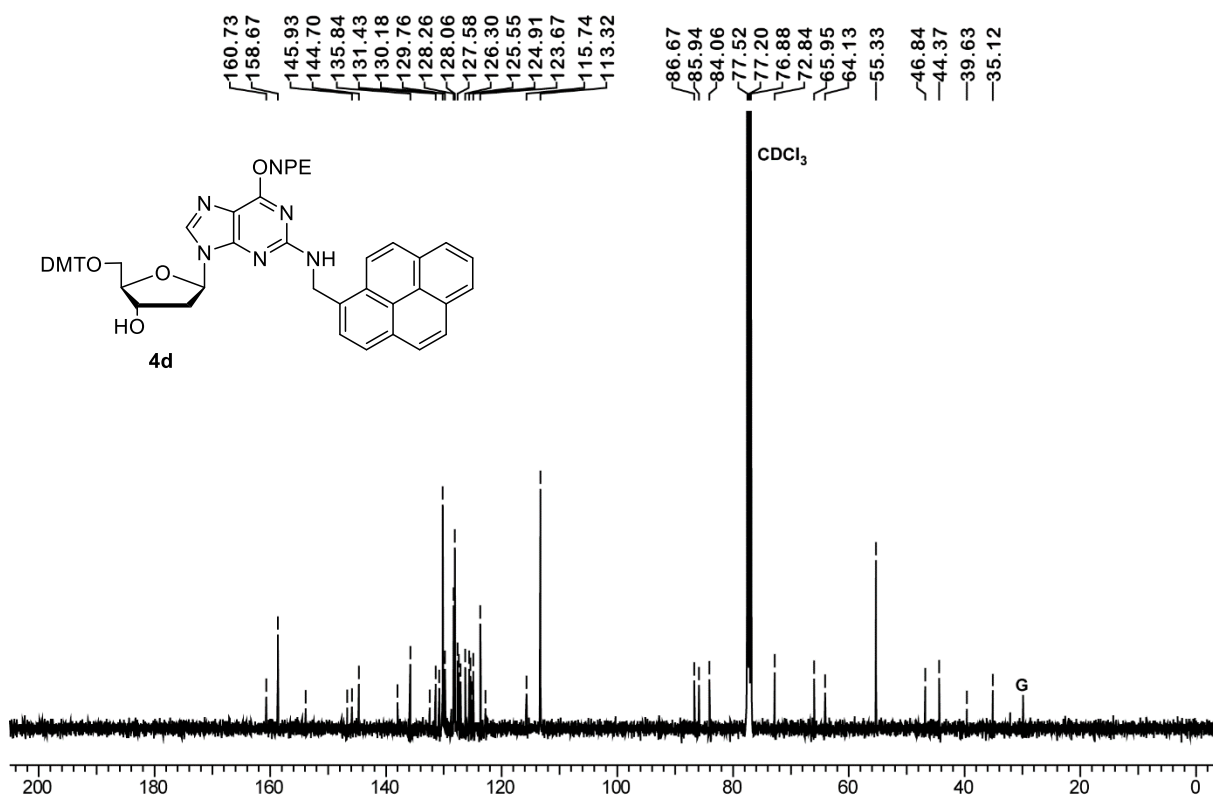
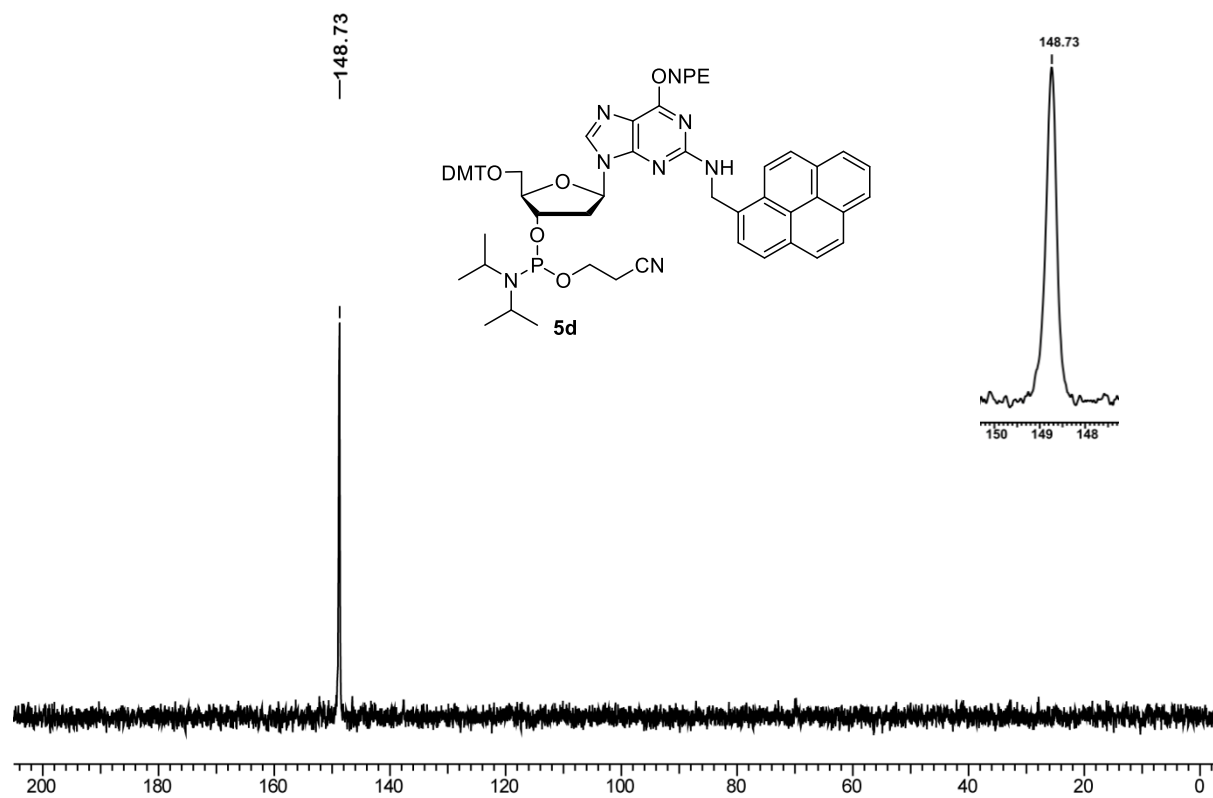
^{13}C NMR for Compound **4b** ^{31}P NMR for Compound **5b**

¹H NMR for Compound **3c**¹³C NMR for Compound **3c**

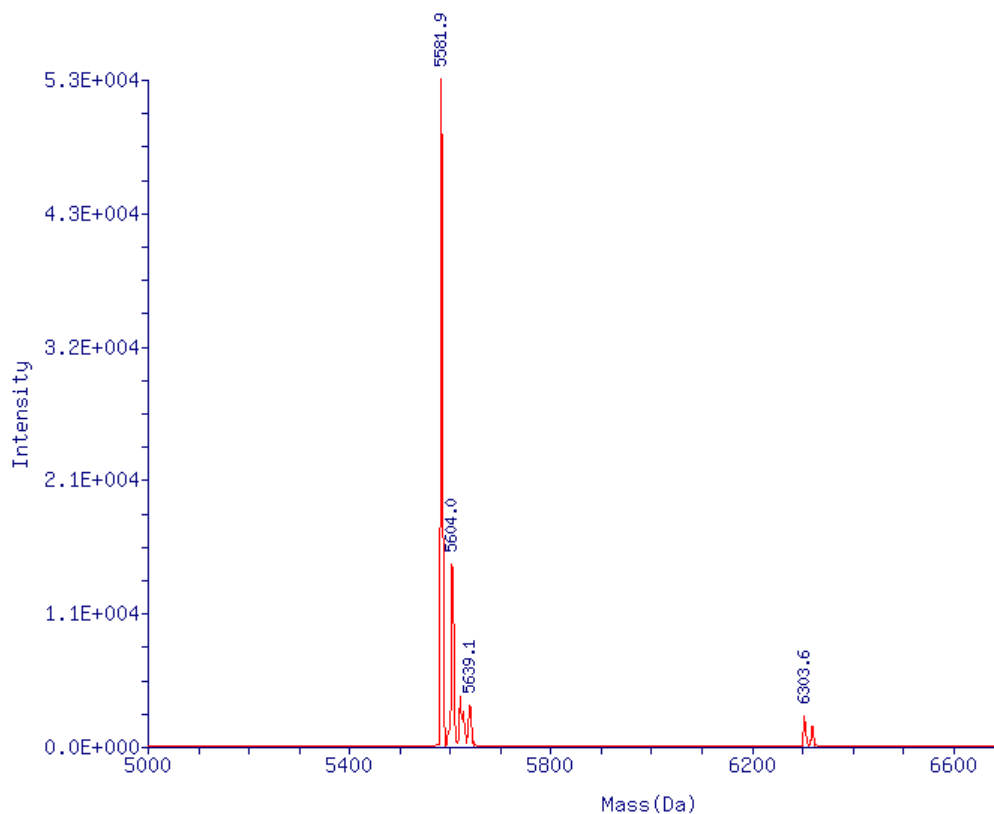
^1H NMR for Compound **4c** ^{13}C NMR for Compound **4c**

^{13}P NMR for Compound **5c** ^1H NMR for Compound **3d**

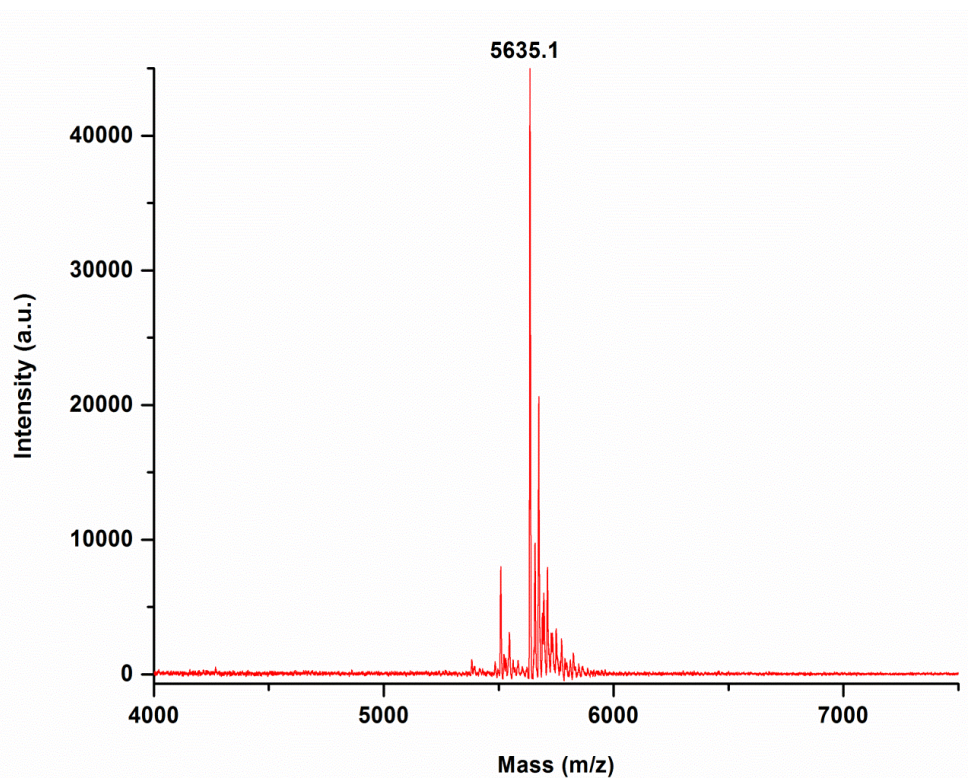
^{13}C NMR for Compound **3d** ^1H NMR for Compound **4d**

¹³C NMR for Compound **4d**³¹P NMR for Compound **5d**

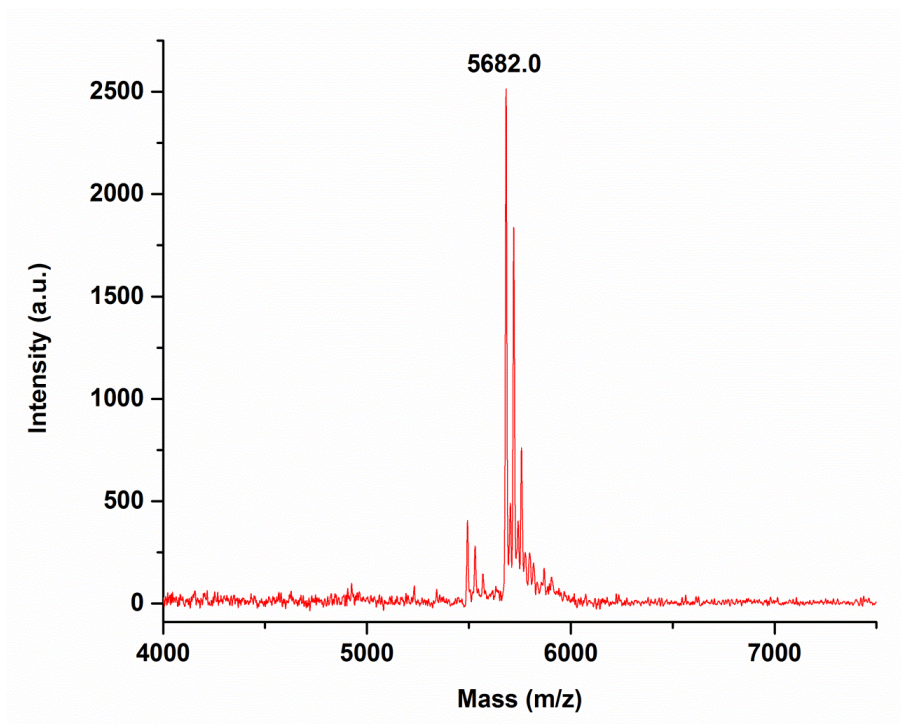
ESI-MS of **D2**, 5'-TCT**G1**GGGTCCTAGGACCC-3' *N*²-Bn-dG
Calc. mass, [M-H]⁻ 5581.7; Obs. mass, [M-H]⁻ 5581.9



MALDI of **D6**, 5'-TCT**G2**GGGTCCTAGGACCC-3' *N*²-Naph-dG
Calc. mass, [M+H]⁺ 5632.8; Obs. mass, [M+H]⁺ 5635.1



MALDI of **D10** 5'-TCT**G3**GGGTCCTAGGACCC-3' *N*²-Anth-dG
Calc. mass, [M+H]⁺ 5682.9; Obs. mass, [M+H]⁺ 5682.0



MALDI of **D14** 5'-TCT**G4**GGGTCCTAGGACCC-3' *N*²-Pyre-dG
Calc. mass, [M+H]⁺ 5706.9; Obs. mass, [M+H]⁺ 5705.8

