Supporting Information

Synthesis and Polymerase Mediated Bypass Studies of the N^2-Deoxyguanosine DNA Damage Caused by a Lucidin Analogue

Pratibha P. Ghodke, S. Harikrishna; and P. I. Pradeepkumar*
Department of Chemistry, Indian Institute of Technology Bombay,
Powai, Mumbai 400076, India
Email: pradeep@chem.iitb.ac.in

TABLE OF CONTENTS

Synthetic strategy for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone

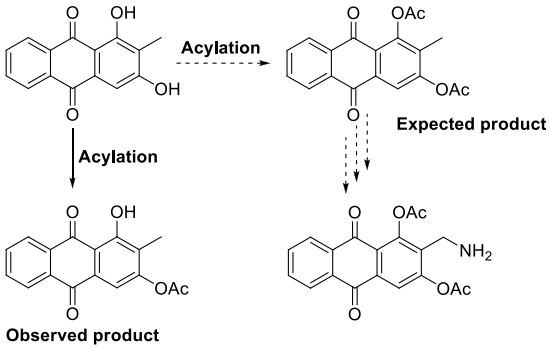


Figure S1. Synthetic strategy utilized for the 2-(aminomethyl)-1,3-diacetoxyanthraquinone

UV thermal melting curves of the unmodified and LdG modified DNAs

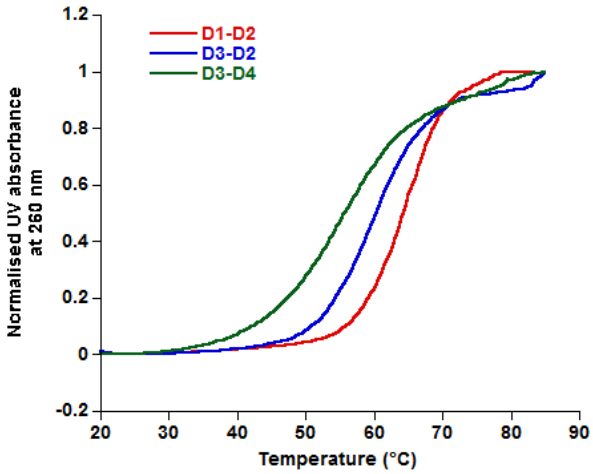


Figure S2. UV thermal melting of the unmodified and the LdG modified DNA duplexes performed using the phosphate buffer (100 mM NaCl, 20 mM phosphate buffer, 0.1 mM EDTA, pH 7.4). Melting curves were obtained using 5 μM concentration of ds DNA.

Full length extension reactions using the Klenow fragment exo⁻ (KF⁻)

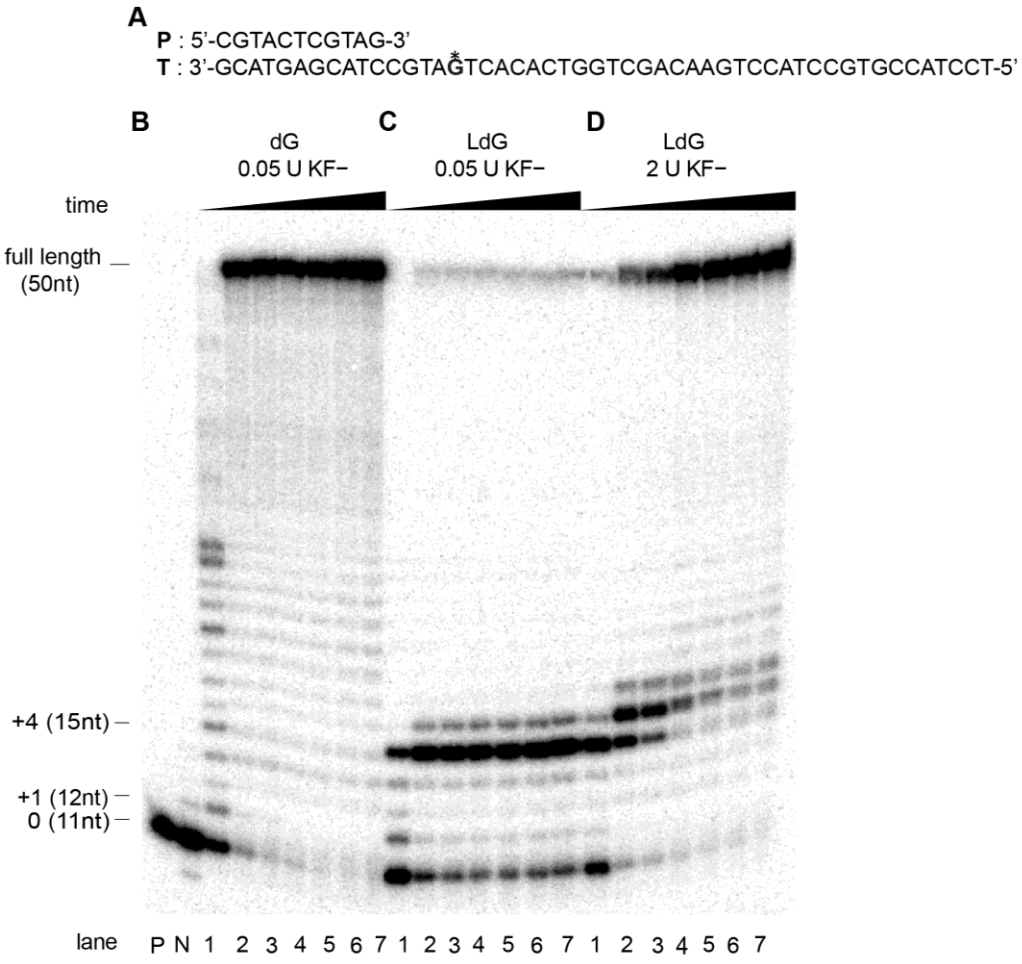
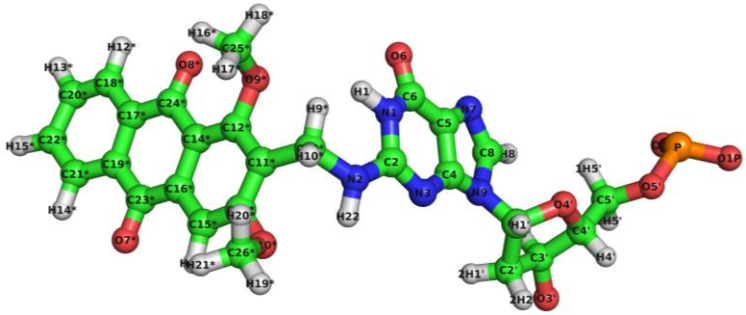


Figure S3. PAGE (20%, 7 M urea) of full length extension reactions using the KF⁻ with all dNTPs. A) Complete sequence of template, T (**G*** = LdG) and primer, P; B) unmodified template (dG) with 0.05 U of KF⁻; C) modified template (LdG) with 0.05 U of KF⁻; D) modified template (LdG) with 2 U of KF⁻. Lane P, primer; lane N, negative reaction control (no dNTPs); lanes 1 to 7: primer extension reactions with 100 μM of each dNTPs in different time course from 0.5, 5, 10, 30, 60, 120, and 240 min. All the reactions were carried out at 37 °C in the polymerase buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9).

Energy optimized structure of the LdG nucleotide



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

Figure S4. Energy optimized geometry and calculated RESP charges for the LdG using HF/6-31G* basis set in Gaussian 09 program.[1-3]

Time dependent RMSD map of the LdG modified ds DNA

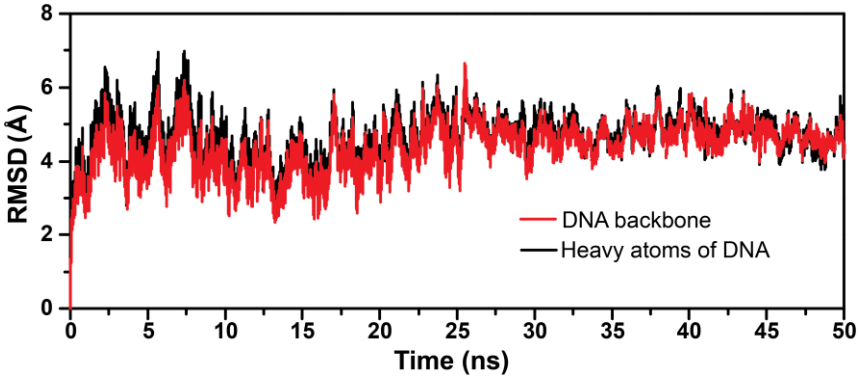


Figure S5. Root mean square deviation graphs of the trajectories obtained from MD simulation of the LdG modified duplex DNA. RMSD values from the MD trajectories were calculated with respect to the corresponding initial structure using ptraj module in AMBER 12.

CD spectra for the unmodified and LdG modified DNAs

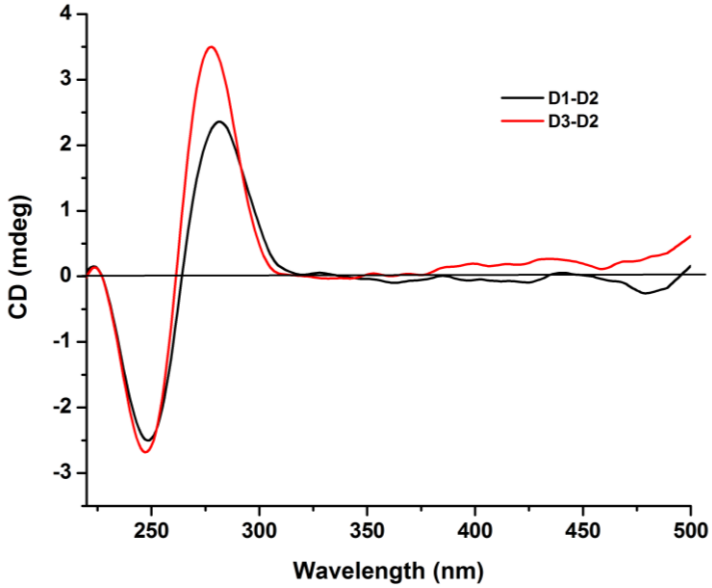


Figure S6. CD spectra for the unmodified (**D1-D2**) and the LdG modified (**D3-D2**) DNA duplexes (10 µM of each DNA strand in 100 mM NaCl, 20 mM sodium phosphate, 0.1 mM EDTA, pH 7.4). CD spectra show the characteristic peaks of B-form DNA. The enhanced CD intensity (280 nm) for the LdG modified DNA in comparison to the unmodified DNA may be attributed to the rigidly bound polycyclic aromatic adduct in the minor groove as revealed in our MD simulations. Similar CD spectrum is also observed for the N^2-CH_2(9-anthracenyl)-dG modified DNA.[4]

Time dependent RMSD map of the protein and DNA complex

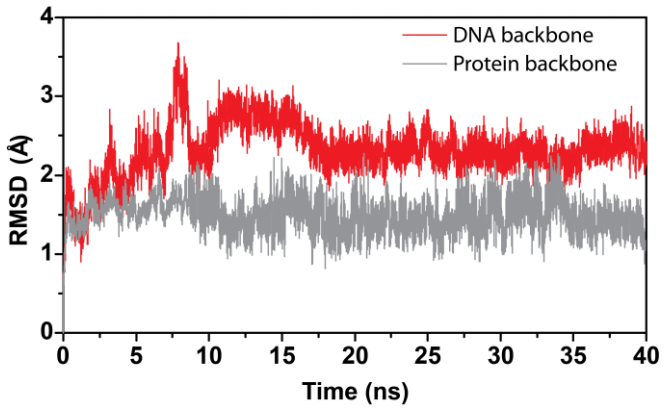


Figure S7. Root mean square deviation graphs of the trajectories obtained from MD simulation of protein and the LdG modified template and primer DNA. RMSD values from the MD trajectories were calculated with the corresponding initial structure using ptraj module in AMBER 12.

MD simulation snapshots from the conformational ensembles

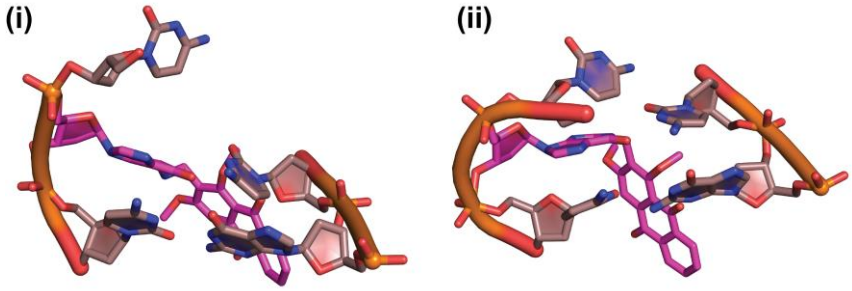


Figure S8. MD simulation snapshots around the LdG modification of template-primer DNA duplex obtained from ensemble analysis. The LdG modification is highlighted in magenta.

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

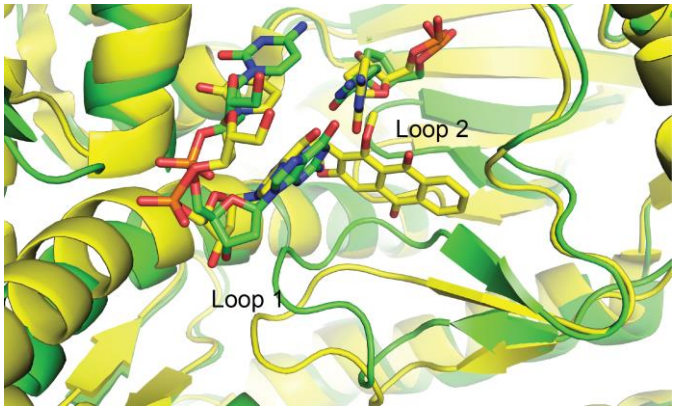


Figure S9. Superimposed structure of the *Bacillus* fragment (BF) in complex with the LdG modified DNA (yellow) and the unmodified DNA (green). The LdG modification (yellow) with the pairing dC and the dG with pairing dC is highlighted.

Watson-Crick H-bond occupancy of the LdG modified ds DNA

Duplex DNA (D2:D3)	Percentage occupancy
dC:dG	90 %
dC:LdG	45 %
dT:dA	75 %
dT:dA	83 %

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

Major and minor groove hydration of the LdG modified ds DNA

Duplex DNA (D2:D3)	Major groove	Minor groove
dC:dG	3 (95 %)	1 (60 %)
dC:LdG	1 (25 %)	0
dT:dA	2 (44 %)	0
dT:dA	2 (78 %)	1 (65 %)

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

Local base pair parameters of the template-primer ds DNA

Template: Primer	C1'-C1' distance	Shear	Stretch	Stagger	Buckle	Propeller	Opening
LdG:dC	11.95 (10.6)	0.79 (0.26)	−0.06 (0.03)	−0.75 (−0.53)	33.25 (29.37)	−14.27 (8.54)	2.15 (4.74)
dC:dG	10.82 (10.6)	−0.45 (0.22)	0.22 (−0.07)	0.14 (0.35)	16.59 (21.04)	−1.14 (8.72)	9.26 (7.12)
dA:dT	10.68 (10.7)	−0.24 (0.14)	0.06 (0.10)	0.52 (−0.17)	10.75 (18.54)	−24.58 (15.22)	8.47 (1.96)
dC:dG	10.58 (10.6)	−0.32 (0.22)	0.10 (0.07)	0.61 (0.02)	7.76 (6.22)	−7.53 (8.46)	5.68 (2.28)

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

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

Template: Primer	Shift	Slide	Rise	Tilt	Roll	Twist
LdG:dC	−0.57 (−0.21)	0.58 (−1.51)	3.28 (3.35)	16.52 (3.56)	21.94 (10.71)	26.35 (31.73)
dC:dG	0.29 (−0.35)	0.21 (−1.48)	3.06 (3.17)	0.27 (−5.22)	6.35 (11.82)	34.27 (25.81)
dA:dT	−0.48 (−0.14)	−0.43 (−0.58)	3.33 (3.89)	−2.45 (−2.37)	6.47 (7.59)	37.58 (35.72)
dC:dG	−0.33 (−0.26)	−0.16 (−0.18)	3.74 (3.27)	−5.21 (1.25)	10.58 (6.58)	31.24 (27.57)

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

Backbone torsion angles of the nucleotide in the template strand

Template Strand	α (O3-P-O5'-C5')	β (P-O5'-C5'-C4')	γ (O5'-C5'-C4'-C3')	δ (C5'-C4'-C3'-O3')	ε (C4'-C3'-O3'-P)	ζ (C3'-O3'-P-O5')
LdG	97.3 (93.2)	−176.1 (−170.3)	68.4 (61.4)	94.8 (82.5)	140.4 (171.5)	16.5 (−52.2)
dC	−95.5 (−90.5)	171.1 (172.4)	54.8 (52.1)	70.9 (62.4)	125.3 (176.8)	−87.7 (−20.8)
dA	−83.3 (−84.2)	179.2 (177.8)	54.1 (50.7)	93.0 (87.8)	−134.2 (−175.4)	−91.9 (−29.6)
dC	−64.2 (−66.8)	−177.9 (−179.1)	70.2 (60.4)	133.3 (117.5)	−173.3 (−175.3)	−87.6 (−71.4)

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

MALDI-TOF mass data of the DNA sequences

Code	DNA Sequences (5'-3')	MW (calc.)	MW (found)
D1	GCCGGAATAGCGCA	4299	4299
D2	TGCGCTATTCCGGC	4232	4233
D3	GCCG**G***AATAGCGCA	4579	4579
D4	TGC**G***CTATTCCGGC	4512	4513
T	TCCTACCGTGCCTACCTGAACAGCTGGTCACACT**G***AT GCCTACGAGTACG	15547	15547
P	CGTACTCGTAGGCAT	4569	4571

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

References

(1) Frisch, M. J.; Gaussian 09, Revision A.02: Gaussian, Inc.: Wallingford, CT, USA, **2009**.

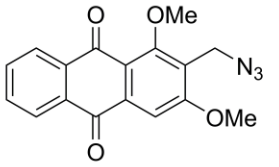
(2) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(3) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

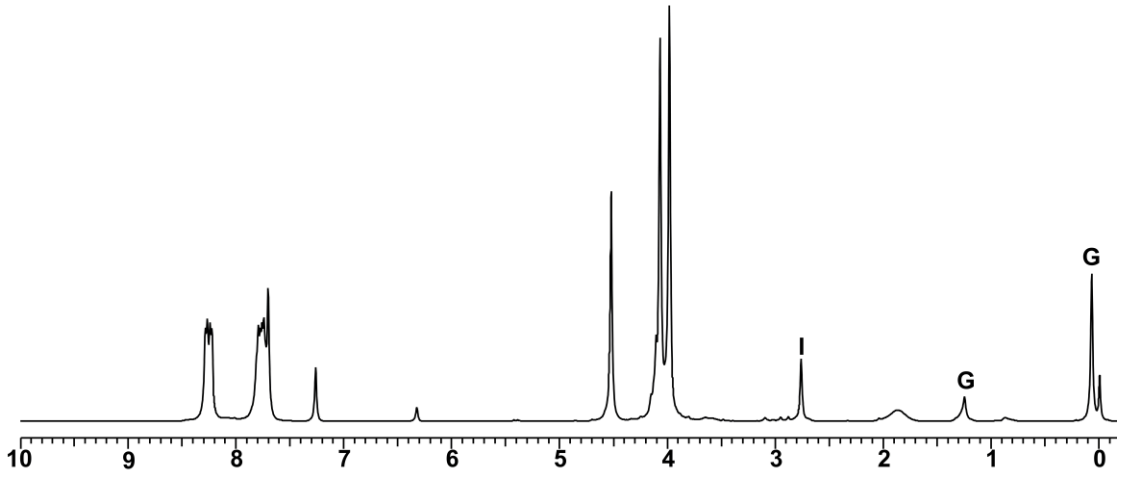
(4) Casale, R.; McLaughlin, L. W. *J. Am. Chem. Soc.***1990**, *112*, 5264–5271

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

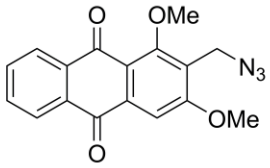
^{1}H NMR spectrum of compound **2**



2

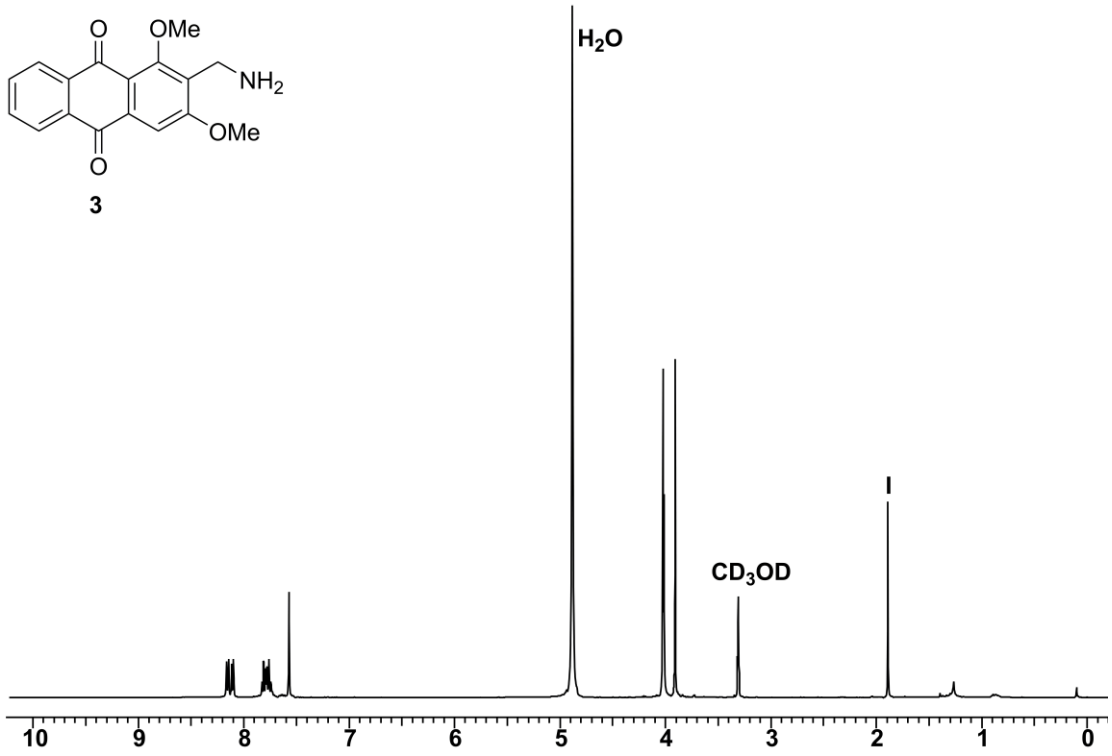


^{13}C NMR spectrum of compound **2**

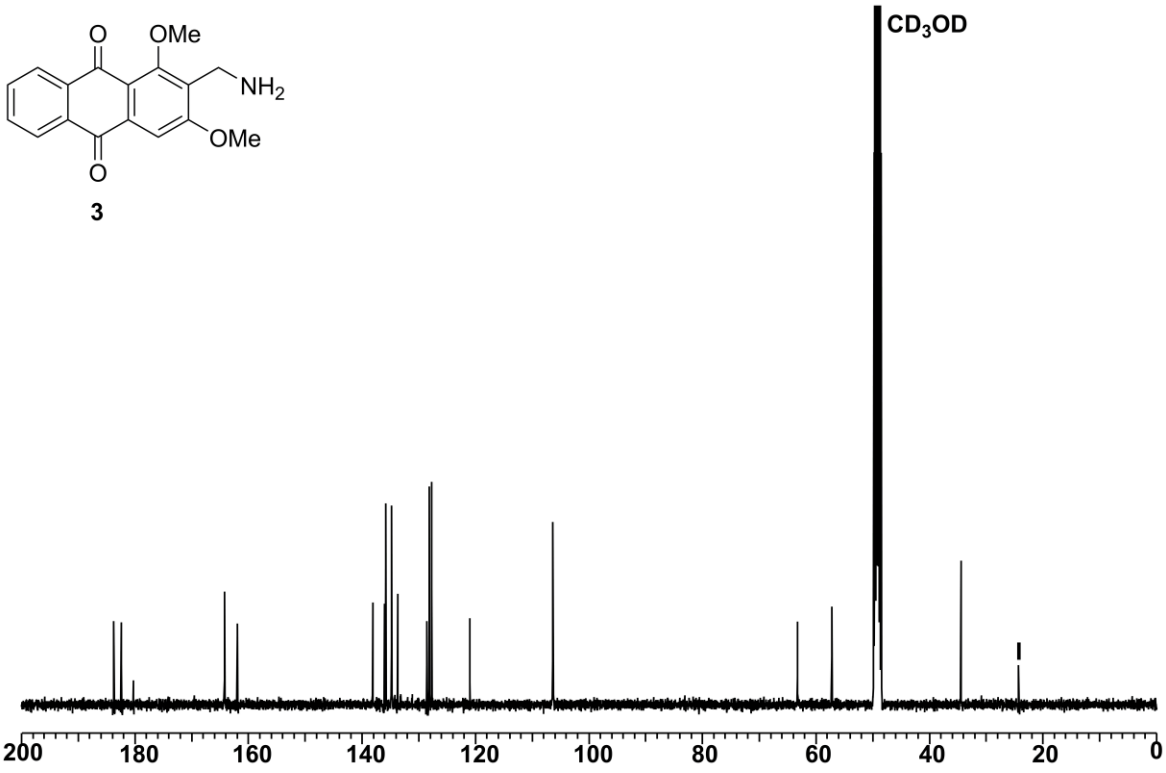


2

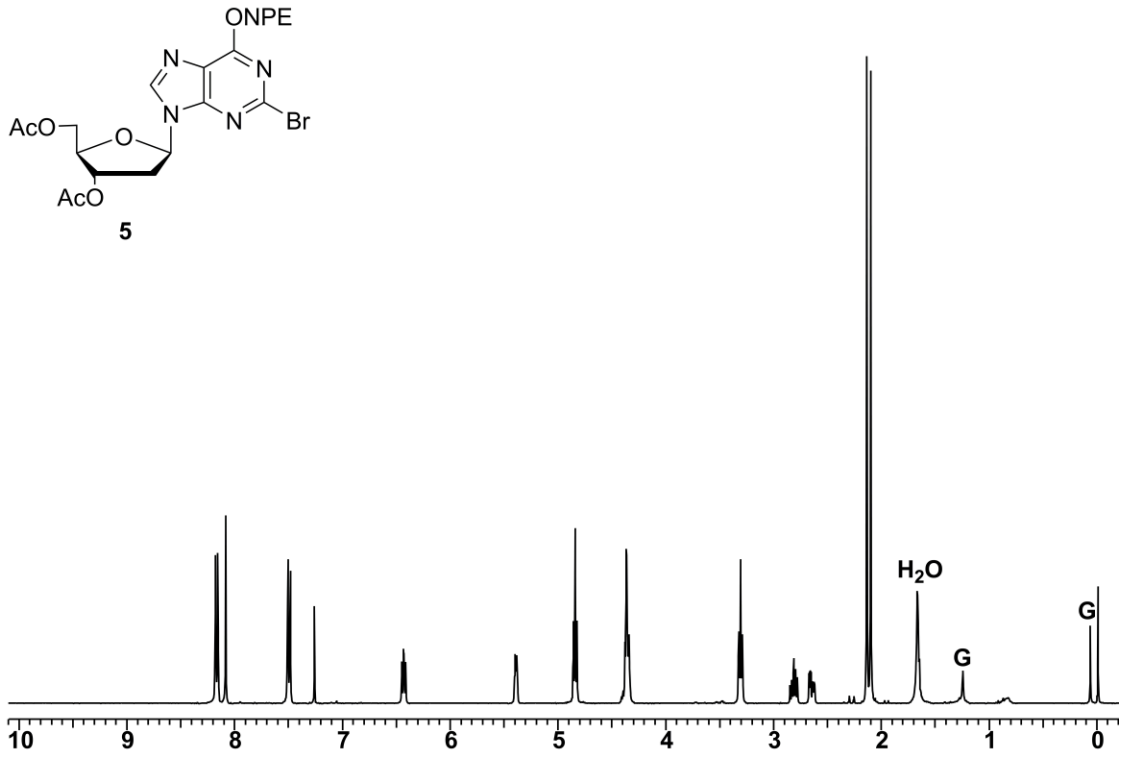
^{1}H NMR spectrum of compound **3**



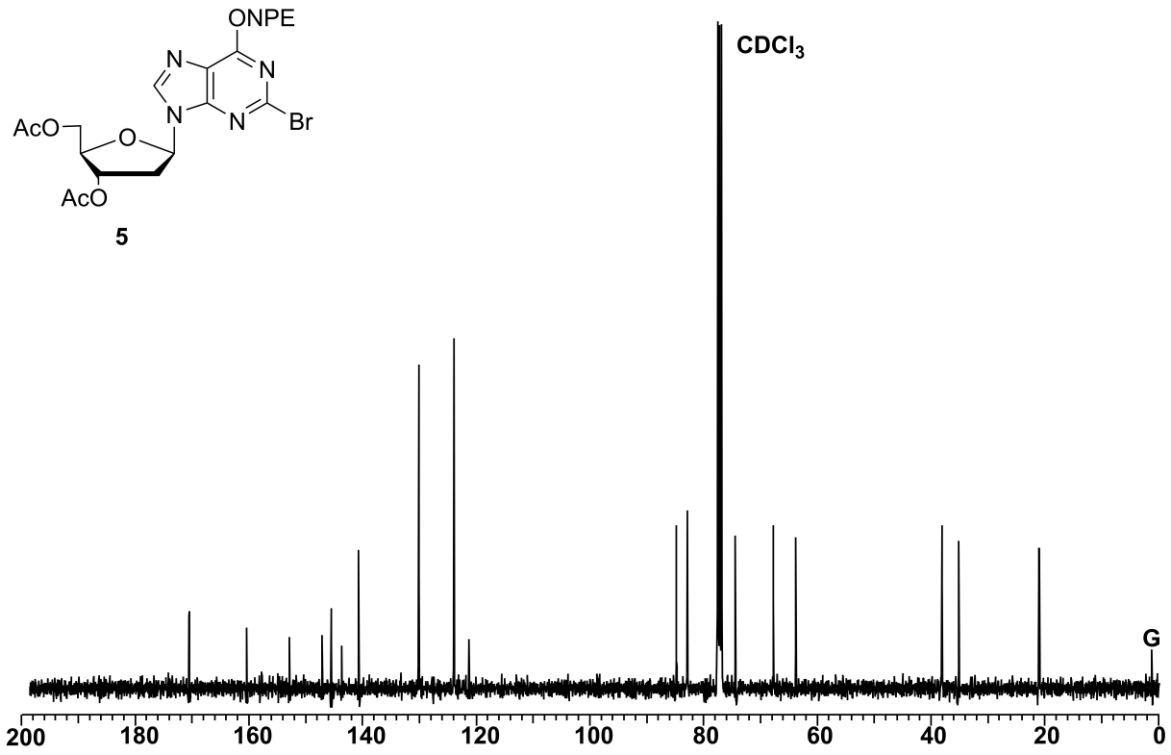
^{13}C NMR spectrum of compound **3**



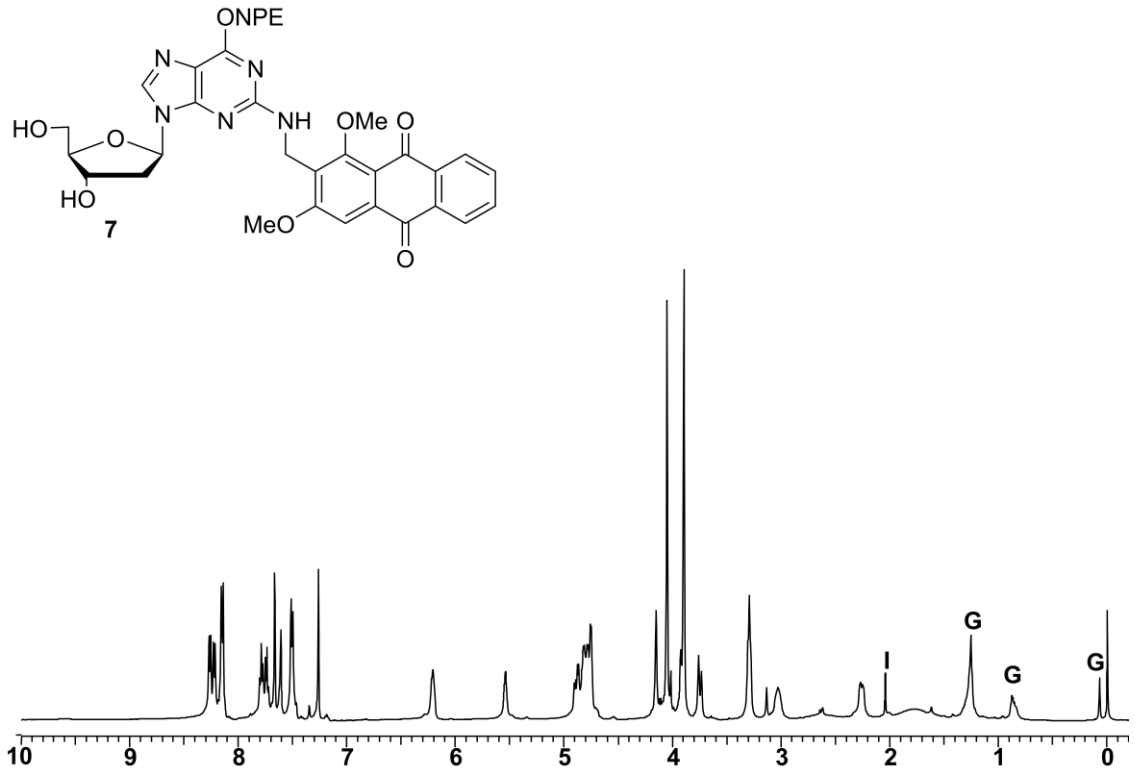
^{1}H NMR spectrum of compound **5**



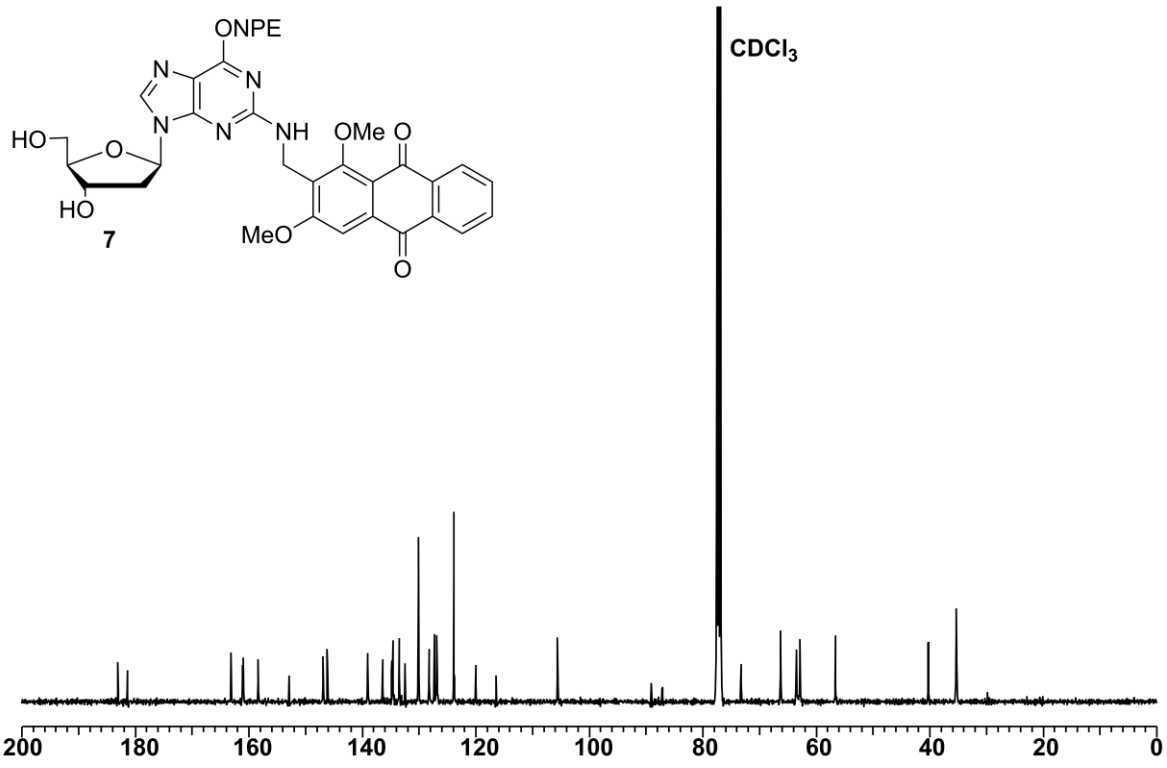
^{13}C NMR spectrum of compound **5**



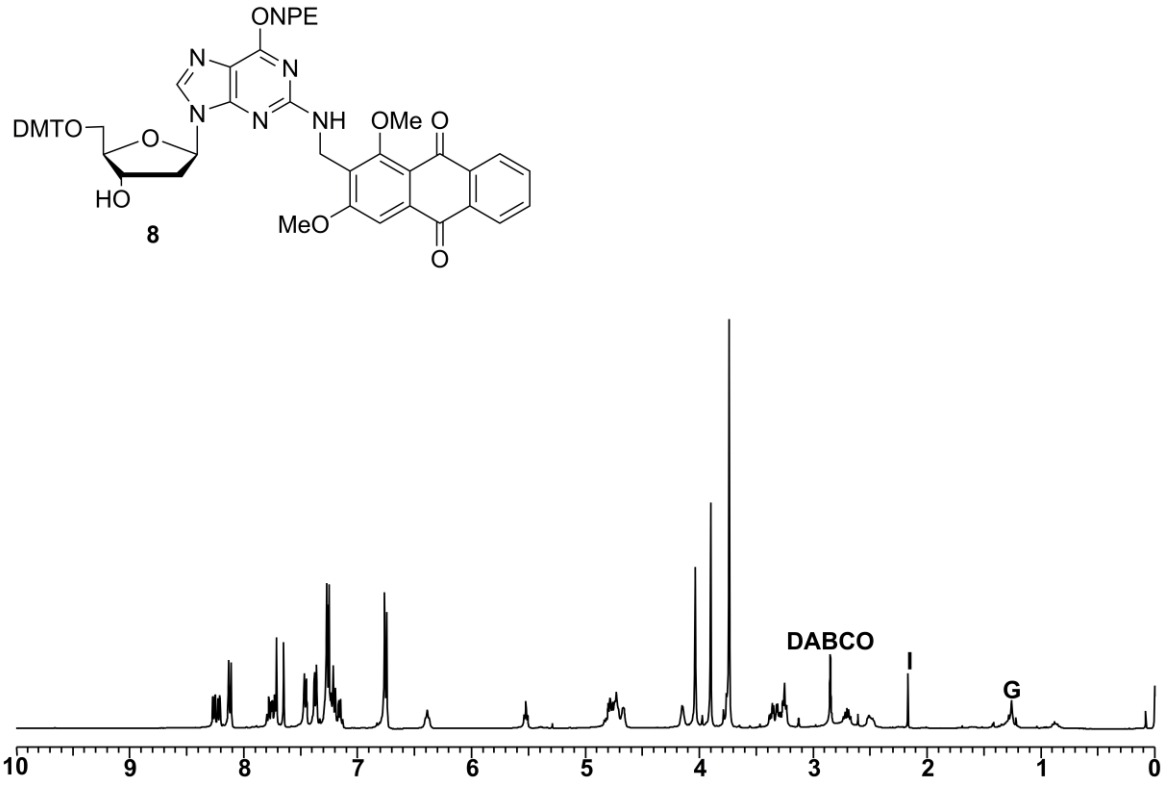
^{1}H NMR spectrum of compound **7**



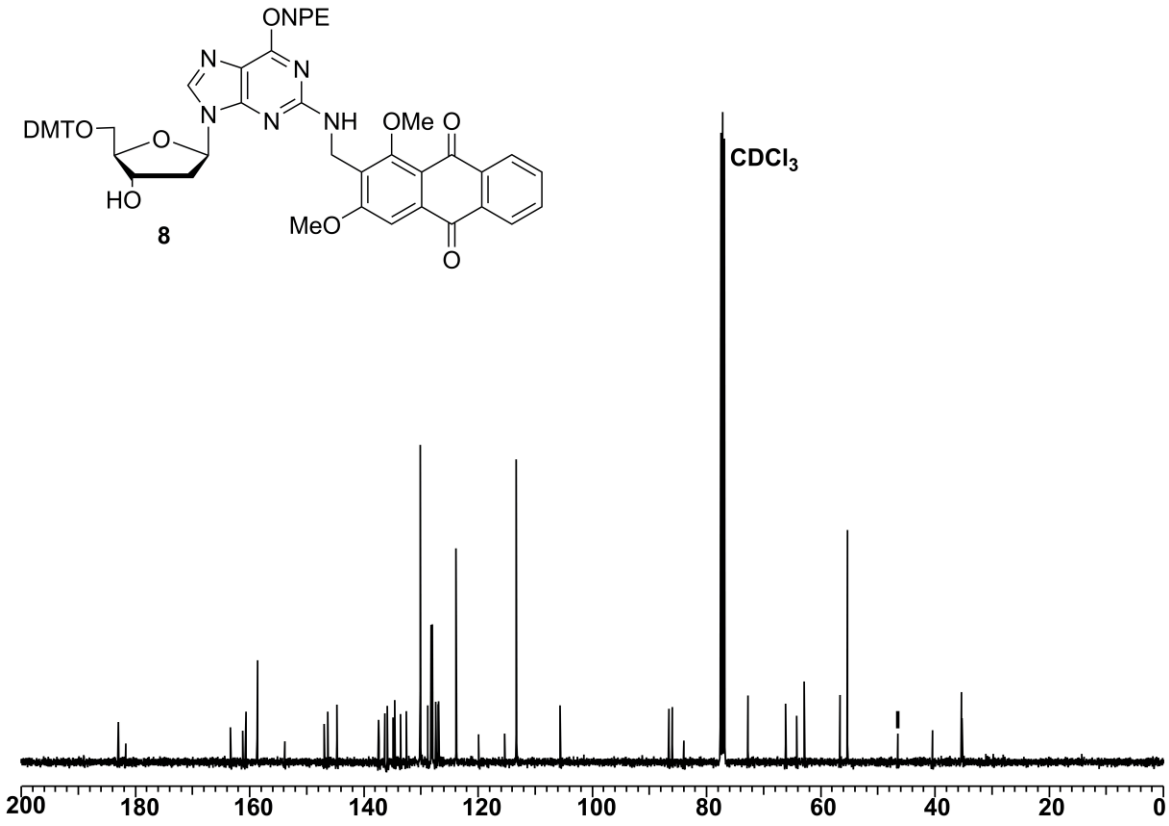
^{13}C NMR spectrum of compound **7**



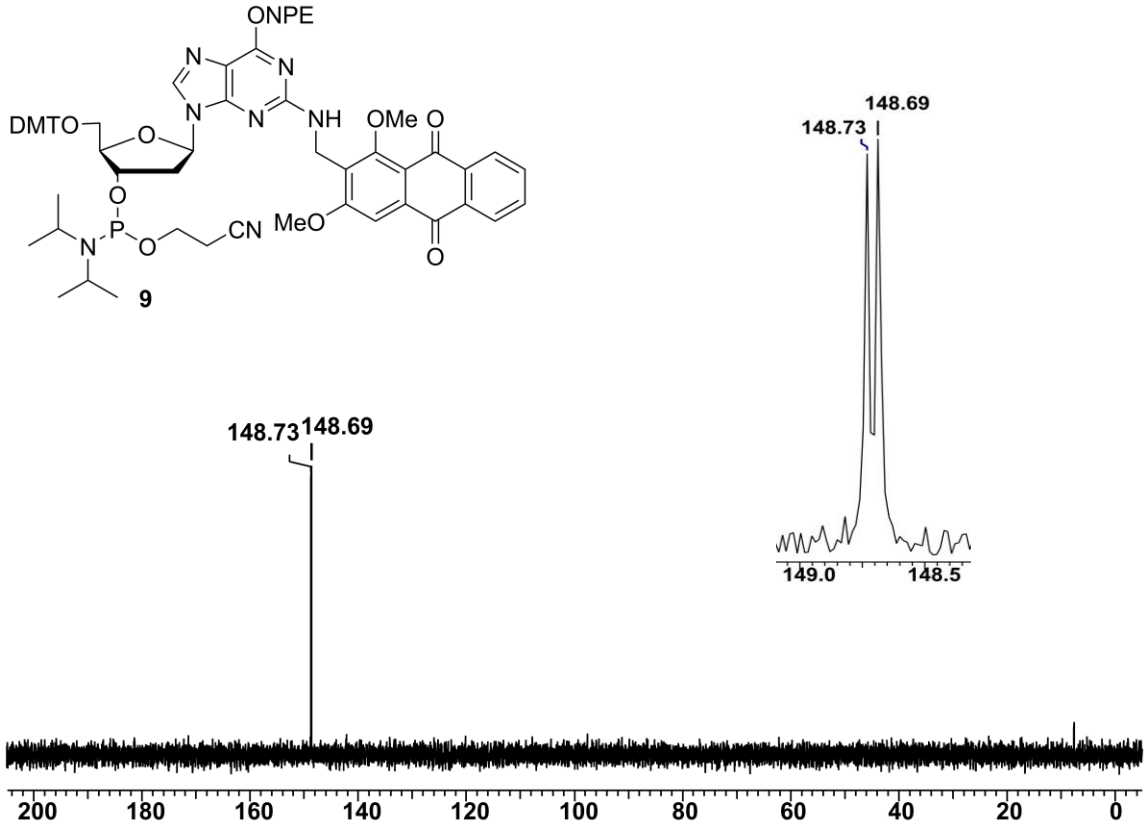
^{1}H NMR spectrum of compound **8**



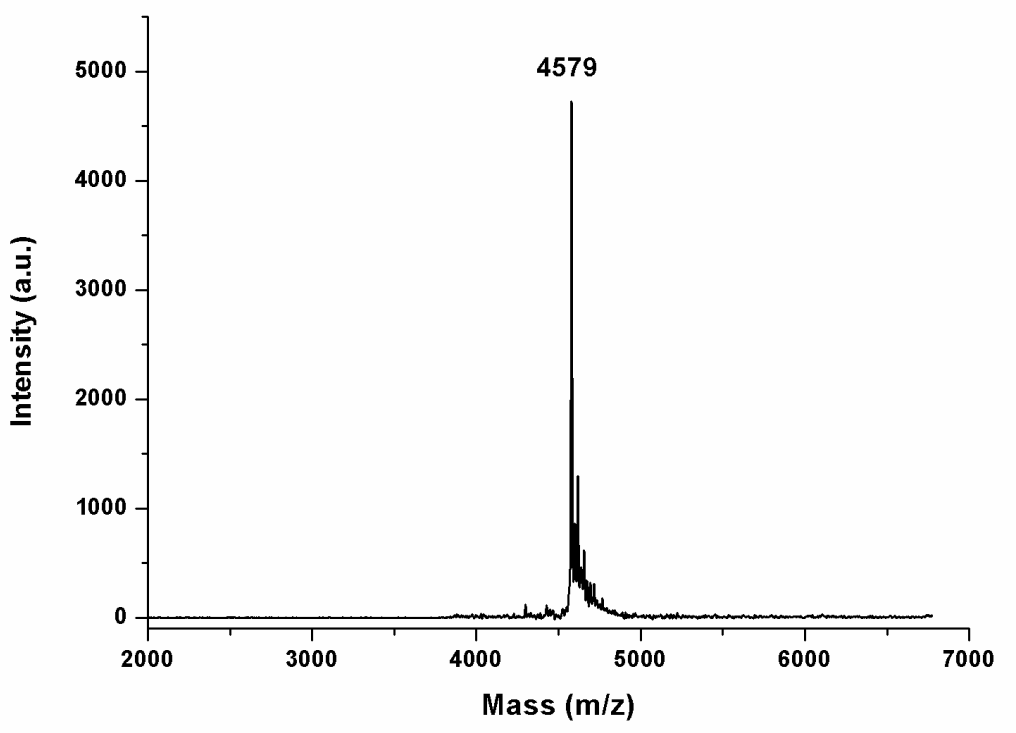
^{13}C NMR spectrum of compound **8**



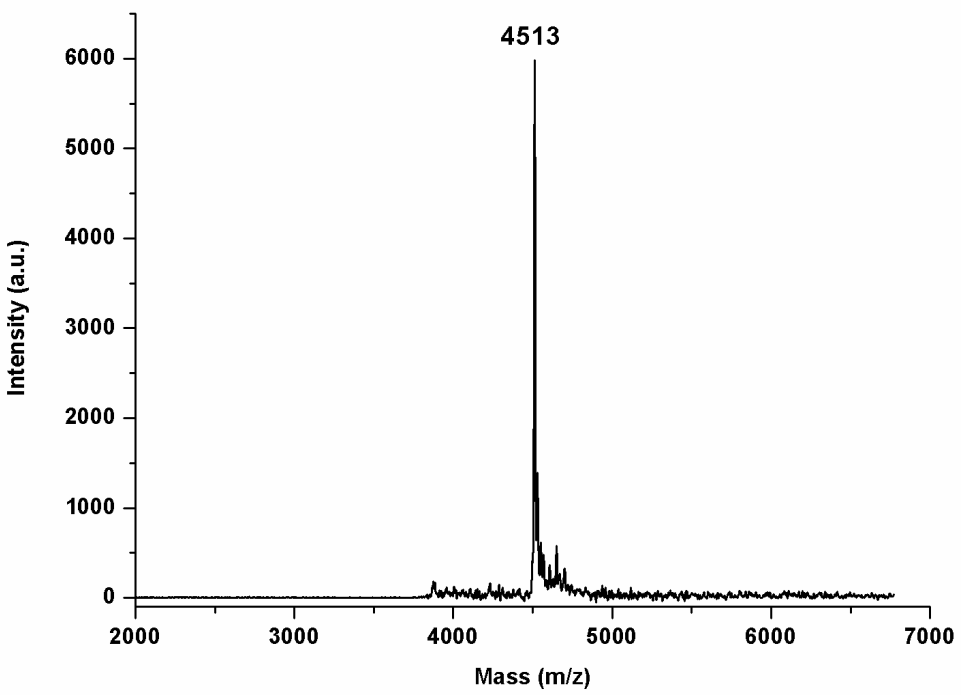
^{31}P NMR spectrum of compound **9**



MALDI spectrum of **D3** 5'-GCCG**G***AATAGCGCA-3'
Calc. mass, $[M + H]^+$ 4579; Obs. mass, $[M + H]^+$ 4579



MALDI spectrum of **D4** 5’-TGC**G***CTATTCCGGC-3’
Calc. mass, $[M + H]^+$ 4512; Obs. mass, $[M + H]^+$ 4513



MALDI spectrum of **T**, 5’-TCCTACCGTGCCTACCTGAACAGCTGGTCACACT**G***ATG CCTACGAGTACG-3’ Calc. mass, $[M + H]^+$ 15547; Obs. mass, $[M + H]^+$ 15547