Supporting Information

Synthesis and Base Pairing Studies of Geranylated 2-thiothymidine, a Natural Variant of Thymidine

Rui Wang,[a,b] Srivathsan V. Ranganathan,[b] Maria Basanta-Sanchez,[b] Fusheng Shen,[a,b] Alan Chen[a,b] and Jia Sheng*[a,b]

[a]Department of Chemistry and [b]The RNA Institute, University at Albany, State University of New York, 1400 Washington Avenue, 12222, Albany, NY 12203 United States. Email: jsheng@albany.edu

Table of Contents

I. Materials and general procedures of synthesis..……...….....….....…….…S2

II. Synthesis of geranylated phosphoramidite ……....…….….....….S2

III. ^{1}H and ^{13}C NMR and HRMS spectra of synthesized compounds.....S5

IV. Synthesis, HPLC and Characterization of geranylated oligonucleotides.....S13

V. Enzymatic hydrolysis and UHPLC-MS/MS study……..S15

VI. UV-melting temperature (T_m) study……....S17

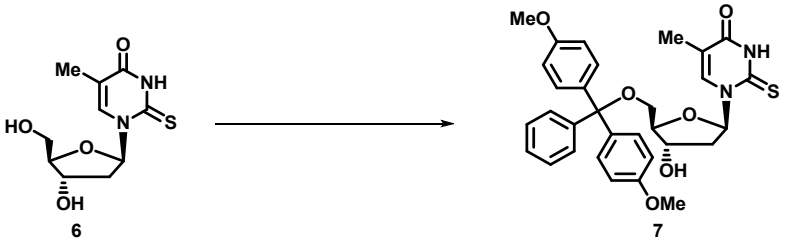
VII. Molecular simulation study……..……...……...……..S17

VIII. References.....S18

Part I. Materials and general procedures of synthesis

Anhydrous solvents were used and redistilled using standard procedures. All solid reagents were dried under a high vacuum line prior to use. Air sensitive reactions were carried out under argon. RNase-free water, tips and tubes were used for RNA purification, crystallization and thermodynamic studies. Analytical TLC plates pre-coated with silica gel F_{254} (Dynamic Adsorbents) were used for monitoring reactions and visualized by UV light. Flash column chromatography was performed using silica gel (32-63 μm). All ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a Brucker 400 spectrometer. Chemical shift values are in ppm. ^{13}C NMR signals were determined by using APT technique. High-resolution MS were achieved by ESI at University at Albany, SUNY.

Part II. Synthesis of geranylated phosphoramidite



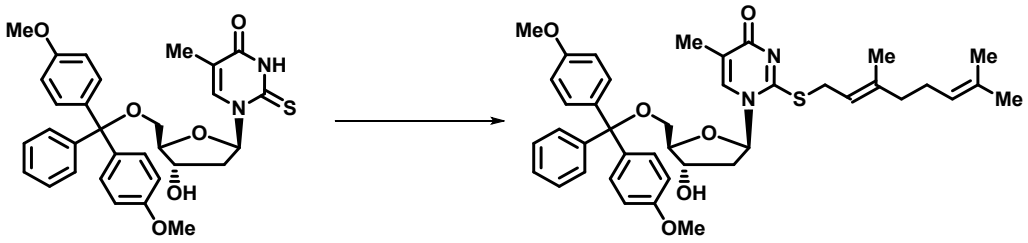
1. 1-((*2R,4S,5R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-5-methyl-2-thioxo-2,3-dihydropyrimidin-4(1H)-one

The 2-thiothymidine **6** (100 mg, 0.39 mmol) was dissolved in pyridine (0.4 mL) and removed of pyridine and repeated 3 times. The crude **1** was placed under vacuum for 12 hours before re-dissolved in the pyridine. Then the pyridine solution of dimethoxytrityl chloride (152 mg, 0.45 mmol) was slowly added at room temperature. The reaction mixture was kept stirring for four hours before being quenched with anhydrous methanol (1.0 mL) and evaporated in *vacuo*. The residue was partitioned between dichloromethane and water. The organic layer was washed with saturated sodium bicarbonate solution, dried over magnesium sulfate, filtered and the solvent was removed in vacuum. The residue was purified by silica gel flash chromatography (eluent 1% methanol in CH_2Cl_2) to afford **7** as a white solid (184.4 mg, 85 %). TLC R_f = 0.60 (1% MeOH in CH_2Cl_2 with 1% triethylamine).

^{1}H NMR (400 HMz, $CDCl_3$) δ 7.89 (s, 1H), 7.69 (t, J = 7.84 Hz, 1H), 7.40 (t, J = 7.48 Hz, 2H), 7.30-7.26 (m, 4H), 7.23 (d, J = 7.16 Hz, 1H), 6.97 (t, J = 6.16 Hz, 1H), 6.83 (dd, J_1 = 8.88 Hz, J_2 = 1.72 Hz, 4H), 4.62-4.59 (m, 1H), 4.13 (d, J = 3.08 Hz, 1H), 3.77 (s, 6H), 3.55 (dd, J_1 = 10.92 Hz, J_2 = 2.72 Hz, 1H), 3.48 (dd, J_1 = 10.56 Hz, J_2 = 2.72 Hz, 1H), 2.68 (dd, J_1 = 14.36 Hz, J_2 = 2.65 Hz, 1H), 2.63 (dd, J_1 = 6.16 Hz, J_2 = 4.12 Hz, 1H), 2.33-2.26 (m, 1H), 1.45 (s, 3H);

^{13}C NMR (100 HMz, $CDCl_3$) δ 174.26, 161.41, 158.61, 149.34, 144.20, 136.43, 136.24, 135.27, 129.98, 128.02, 127.89, 127.05, 123.78, 116.22, 113.17, 89.79, 86.80, 86.56, 71.07, 62.92, 55.15, (45.56), 41.13, 12.04 (10.55);

ESI-MS: $[M+Na]^+$: 583.1892 (calculated MS: 583.1879).



7 8

2. 1-((*2R,4S,5R*)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-2-(((E)-3,7-dimethylocta-2,6-dien-1-yl)thio)-5-methylpyrimidin-4(1H)-one

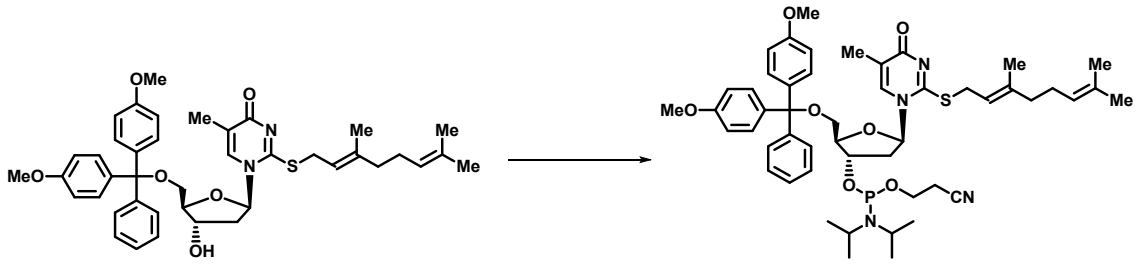
A solution of compound **7** (134.5 mg, 0.24 mmol), geranyl bromide (114.0 uL, 0.50 mmol) and *N,N*-diisopropylethylamine (175.0 uL, 1.0 mmol) in MeOH (3.0 mL) was stirred at 25 °C for 12 h. The resulting reaction mixture was quenched with water and washed with brine (8 X 10 mL), dried over Na_2SO_4. The crude product was concentrated in vacuum and directly subjected to silica gel chromatography. A white solid **8** (153.8 mg, 92% yield) was obtained after silica gel flash chromatography. TLC R_f = 0.70 (1% MeOH in CH_2Cl_2 with 1% triethylamine).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (s, 1H), 7.39 (d, *J* = 7.16 Hz, 2H), 7.30-7.16 (m, 6H), 6.83 (s, 2H), 6.81 (s, 2H), 6.24 (t, *J* = 5.80 Hz, 1H), 5.31 (t, *J* = 7.88 Hz, 1H), 5.06 (t, *J* = 7.88 Hz, 1H), 4.64 (t, *J* = 3.08 Hz, 1H), 4.12 (d, *J* = 2.72 Hz, 1H), 3.96-3.86 (m, 2H), 3.77 (s, 6H), 3.50 (dd, J_1 = 10.60 Hz, J_2 = 3.08 Hz, 1H), 3.41-3.37 (m, 1H), 2.51-2.46 (m, 1H), 2.37-2.30 (m, 1H), 2.10-2.00 (m, 5H), 1.70 (s, 3H), 1.67 (s, 3H), 1.59 (s, 3H), 1.52 (s, 3H);

^{13}C NMR (100 MHz, $CDCl_3$) δ 169.59, 160.71, 158.63, 144.12, 142.58, 135.22, 134.43, 131.68, 129.96, 129.02, 127.99, 127.90, 127.70 (127.65), 127.06, 126.65, 123.64, 118.93, 116.32, 113.19, 113.04 (113.00), 105.43, 87.95, 86.93, 86.67, 71.90, 63.23, 55.13, 41.62, 39.46, 30.67, 26.26, 25.58, 17.61, 16.26, 13.27;

ESI-MS: $[M+H]^+$: 697.3341 (calculated MS: 697.3311).

3. (*2R,3S,5R*)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-5-(2-(((E)-3,7-dimethylocta-2,6-dien-1-yl)thio)-5-methyl-4-oxopyrimidin-1(4H)-yl)tetrahydrofuran-3-yl-(2-cyanoethyl)-diisopropyl phosphoramidite



8 9

A solution of **8** (153.8 mg, 0.22 mmol), *N,N*-diisopropylethylamine (175.0 uL, 0.5 mmol) and Cyanoethyl N, N-diisopropylchlorophosphoramidite (75.7 mg, 0.32 mmol) in anhydrous CH_2Cl_2 (3.0 mL) was stirred at 25 °C for 6 h. The resulting reaction mixture as the two diastereomers was concentrated in vacuum and directly subjected to silica gel chromatography. A sticky yellowish liquid **9** (143.0 mg, 72% yield) was obtained using silica gel flash chromatography. TLC R_f = 0.43 and 0.45, respectively (1% MeOH in CH_2Cl_2 with 3% triethylamine).

^{1}H NMR (400 MHz, CD_3OD) δ 7.81-7.74 (m, 1H), 7.39-7.37 (m, 1H), 7.29-7.11 (m, 7H), 6.83-6.80 (m, 4H), 6.22-6.17 (m, 1H), 5.34-5.29 (m, 1H), 5.06-5.04 (m, 1H), 4.20-4.10 (m, 1H), 3.77 (s, 6H),

3.67-3.46 (m, 5H), 3.42-3.30 (m, 2H), 3.05 (m, 3H), 2.74 (t, *J* = 6.12 Hz, 1H), 2.66-2.59 (m, 2H), 2.45-2.32 (m, 1H), 2.05-2.03 (m, 3H), 1.71 (s, 2H), 1.66 (s, 3H), 1.51-1.45 (m, 1H), 1.37 (t, *J* = 7.16 Hz, 3H), 1.26 (t, *J* = 6.12 Hz, 6H), 1.87-1.43 (m, 9H);

^{13}C NMR (100 MHz, $CDCl_3$) δ 169.33, 160.50, 158.64, 143.99, 142.55, 135.14 (135.08, 135.03), 134.11 (134.07), 131.63, 130.02 (129.98, 129.97, 129.94), 129.01, 128.06, 128.00, 127.89, 127.68, 127.65, 127.09, 127.06 (126.90), 123.62, 119.10, 116.46 (116.41), 113.16 (113.15), 113.00 (112.92, 112.91, 112.89), 87.82, 86.93, 86.10 (86.06), 85.73 (85.73), (73.59, 73.42, 73.24), 62.86 (62.77), (58.29) 58.05 (58.00, 57.98), 55.15, 45.94, 45.20 (45.14), 43.25 (43.18, 43.13, 43.06), (42.88), (42.76), 39.46 (39.45), 30.66 (29.56, 28.96, 28.60), 26.25, 25.56, 24.50 (24.49, 24.46, 24.44, 24.41, 24.39, 24.33), 22.86, 22.84, 22.78, 22.75, (20.31, 20.24, 20.13, 20.06, 20.01, 19.95), 17.58, 16.26 (16.23), 13.22;

^{31}P NMR (200 MHz, $CDCl_3$) δ 149.16, 148.77;

ESI-MS: $[M+H]^+$: 897.4419 (calculated MS: 897.4390).

Part III. ^{1}H and ^{13}C NMR and HRMS spectra of synthesized compounds

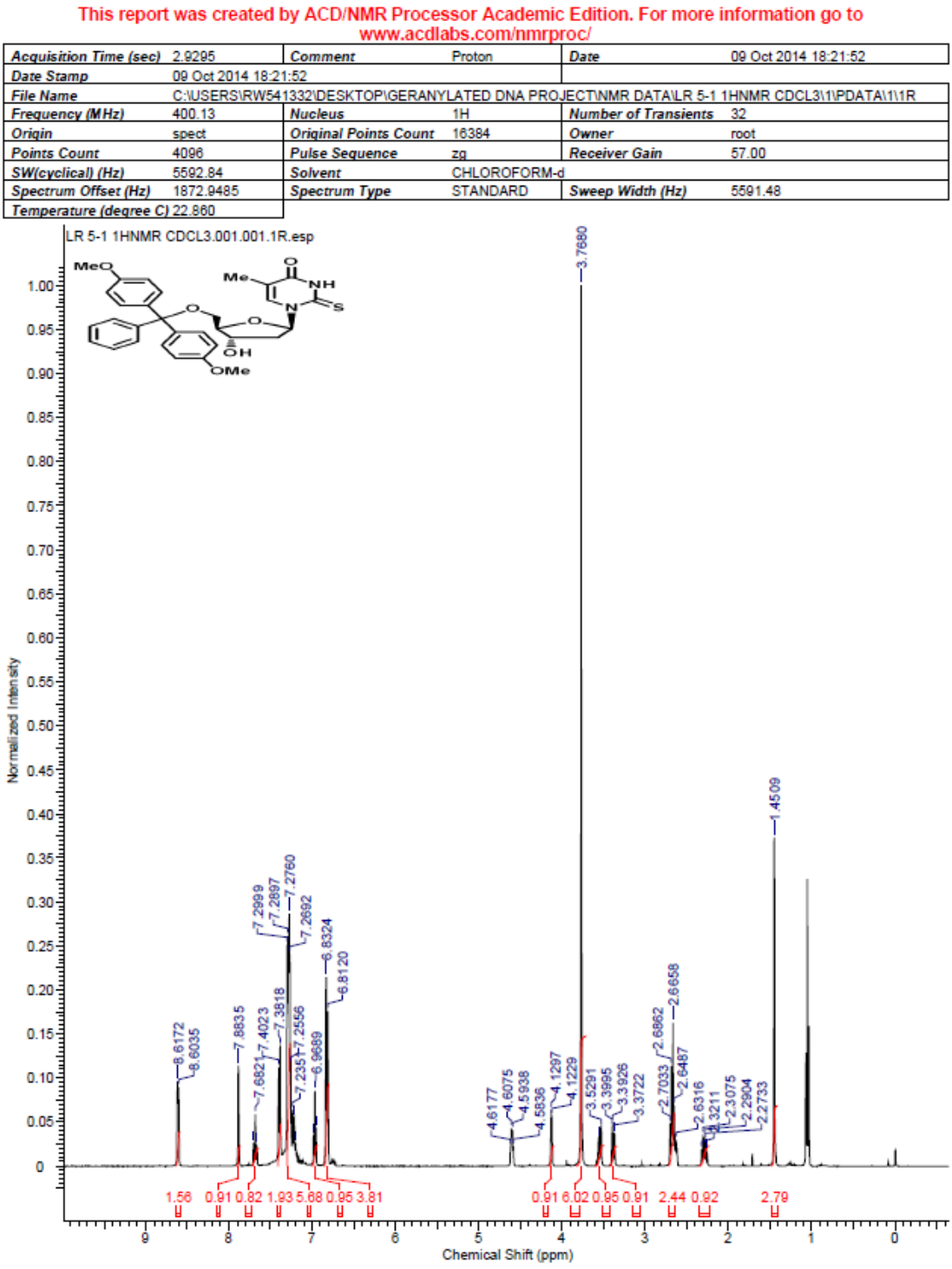


Fig. S1 ^{1}H NMR of compound **7**.

Acquisition Time (sec)	1.4008	*Comment*	carbon	*Date*	09 Oct 2014 22:12:16
Date Stamp	09 Oct 2014 22:12:16				
File Name	C:\USERS\RW541332\DESKTOP\GERANYLATED DNA PROJECT\NMR DATA\LR 5-1 13CNMR CDCL3\1\PDATA\1\1R				
Frequency (MHz)	100.61	*Nucleus*	13C	*Number of Transients*	5002
Origin	spect	*Original Points Count*	40960	*Owner*	root
Points Count	65536	*Pulse Sequence*	zgpg	*Receiver Gain*	16384.00
SW(cyclical) (Hz)	29239.77	*Solvent*	CHLOROFORM-d		
Spectrum Offset (Hz)	8072.6572	*Spectrum Type*	STANDARD	*Sweep Width (Hz)*	29239.32
Temperature (degree C)	23.360				

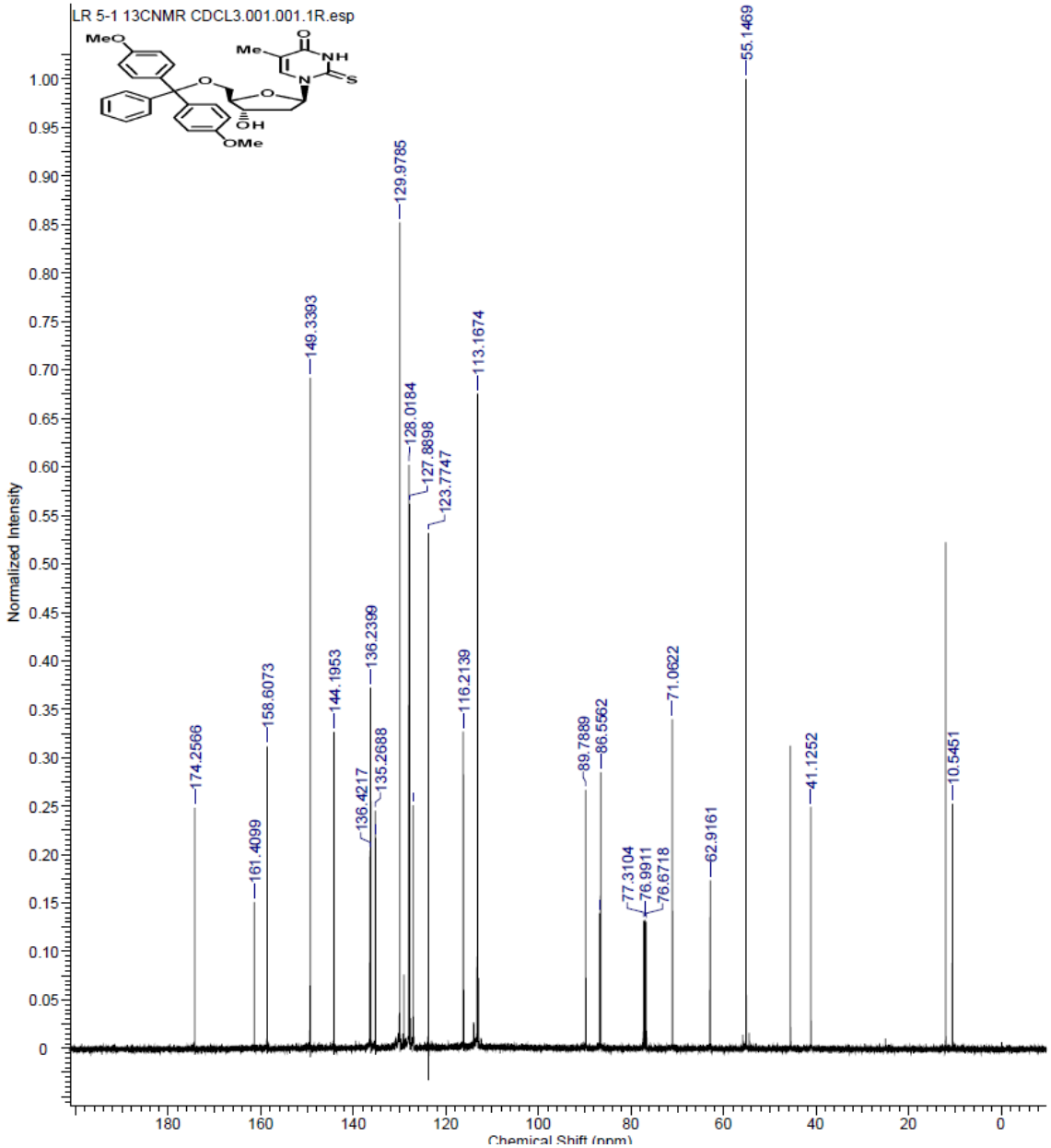


Fig. S2 ^{13}C NMR of compound **7**.

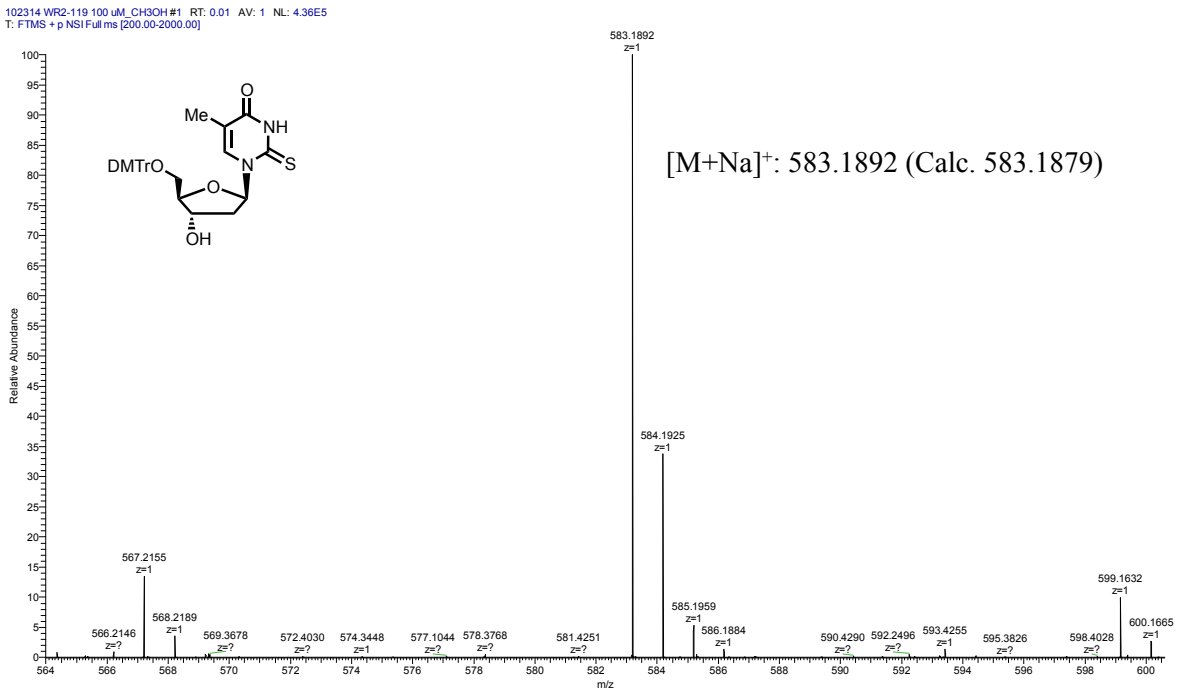


Fig. S3 High-resolution MS of compound **7**.

This report was created by ACD/NMR Processor Academic Edition. For more information go to www.acdlabs.com/nmrproc/

Acquisition Time (sec)	2.9295	*Comment*	Proton	*Date*	13 Oct 2014 07:24:48
Date Stamp	13 Oct 2014 07:24:48				
File Name	C:\USERS\RW541332\DESKTOP\GERANYLATED DNA PROJECT\NMR DATA\LR-1-GER-STEP2-1HNMR-CDCL3\1\1\PDATA\1\1R				
Frequency (MHz)	400.13	*Nucleus*	1H	*Number of Transients*	32
Origin	spect	*Original Points Count*	16384	*Owner*	root
Points Count	4096	*Pulse Sequence*	zg	*Receiver Gain*	143.70
SW(cyclical) (Hz)	5592.84	*Solvent*	CHLOROFORM-d		
Spectrum Offset (Hz)	1872.9485	*Spectrum Type*	STANDARD	*Sweep Width (Hz)*	5591.48
Temperature (degree C)	22.660				

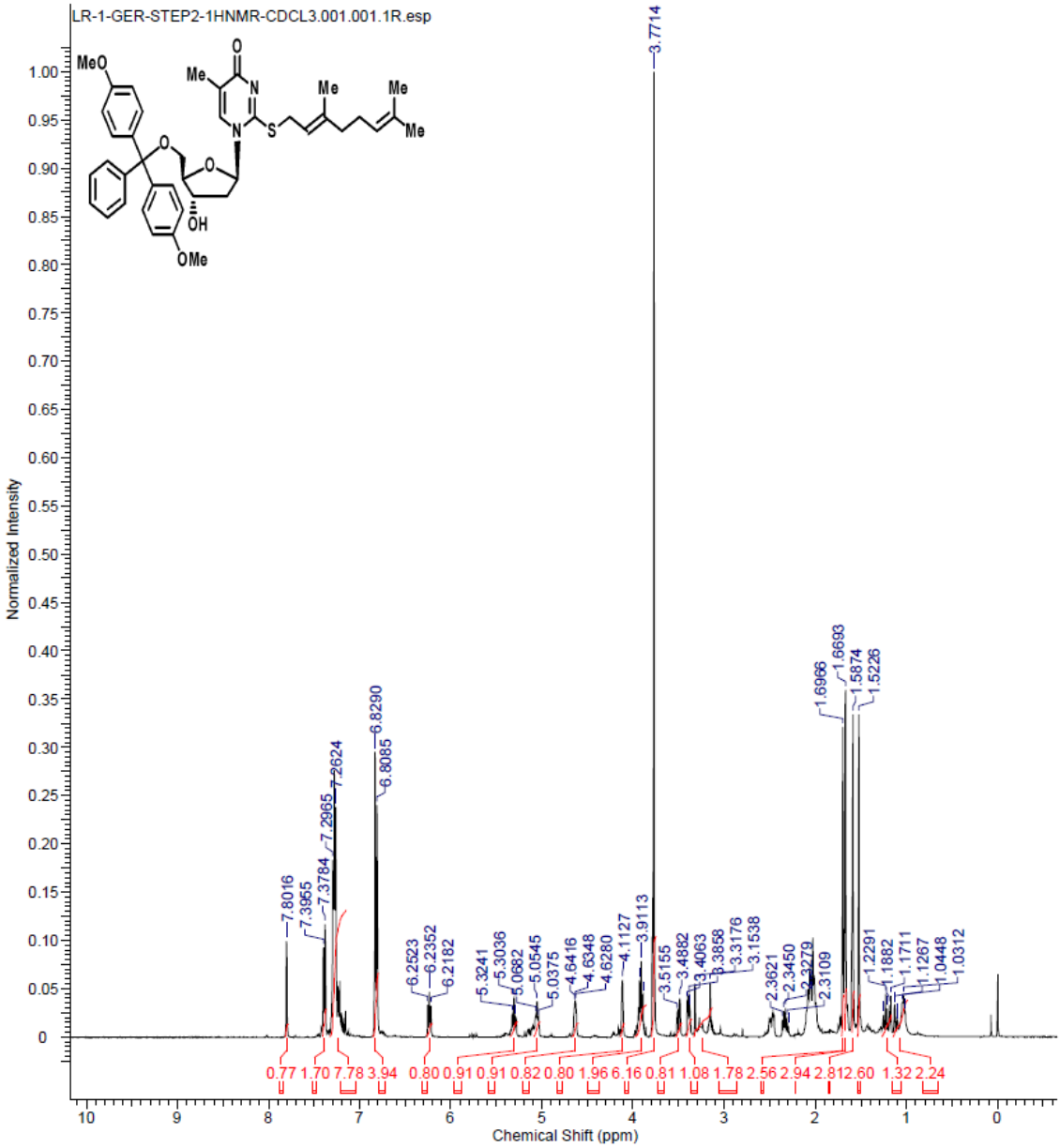


Fig. S4 ^{1}H NMR of compound **8**.

This report was created by ACD/NMR Processor Academic Edition. For more information go to www.acdlabs.com/nmrproc/

Acquisition Time (sec)	1.4008	*Comment*	carbon	*Date*	13 Oct 2014 09:41:20
Date Stamp	13 Oct 2014 09:41:20				
File Name	C:\USERS\RW541332\DESKTOP\GERANYLATED DNA PROJECT\NMR DATA\LR-1-GER-13CNMR-CDCL3\1\PDATA\1\1R				
Frequency (MHz)	100.61	*Nucleus*	13C	*Number of Transients*	2882
Origin	spect	*Original Points Count*	40960	*Owner*	root
Points Count	65536	*Pulse Sequence*	zgpg	*Receiver Gain*	16384.00
SW(cyclical) (Hz)	29239.77	*Solvent*	CHLOROFORM-d		
Spectrum Offset (Hz)	8072.6572	*Spectrum Type*	STANDARD	*Sweep Width (Hz)*	29239.32
Temperature (degree C)	23.260				

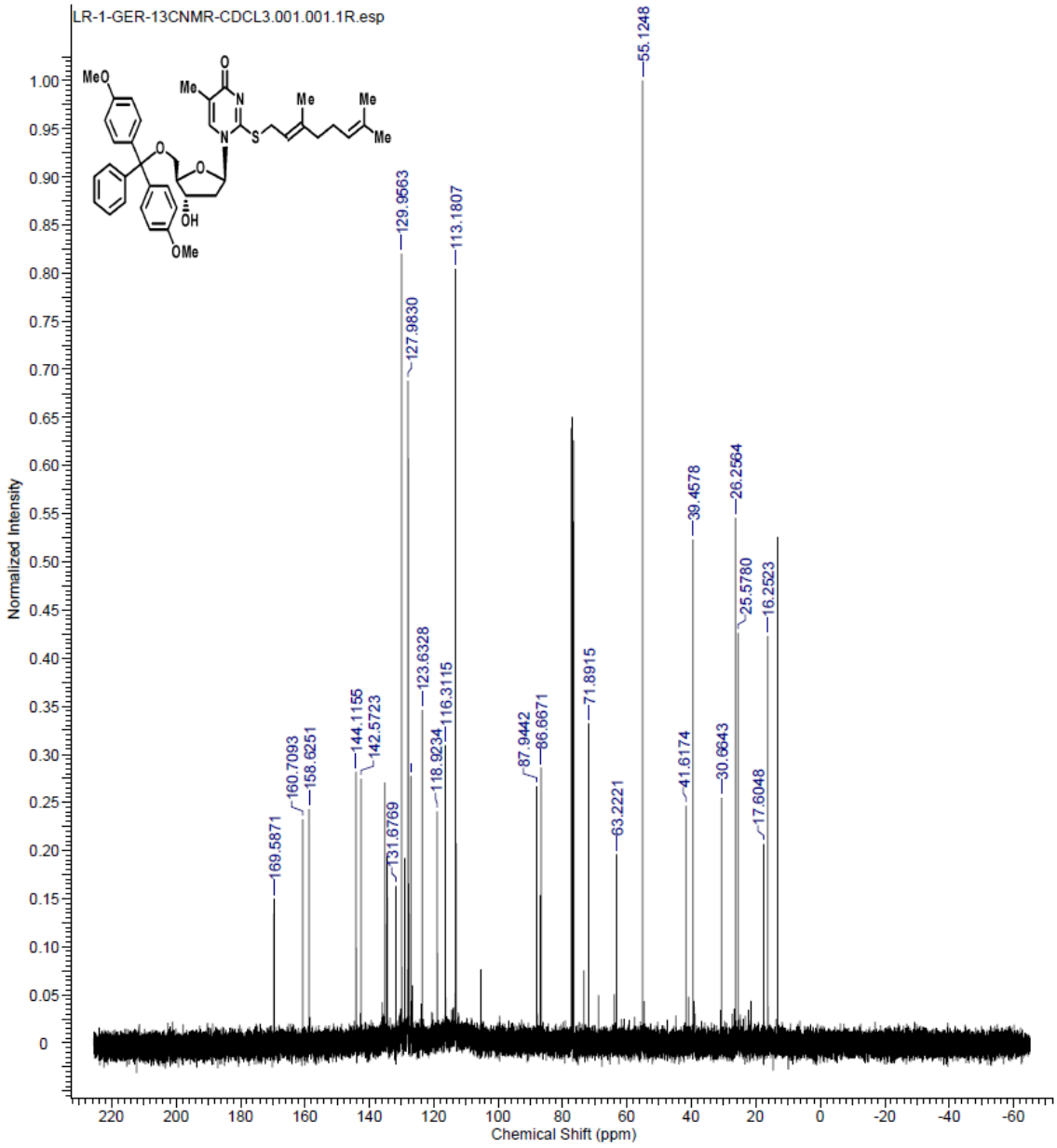


Fig. S5 ^{13}C NMR of compound **8**.

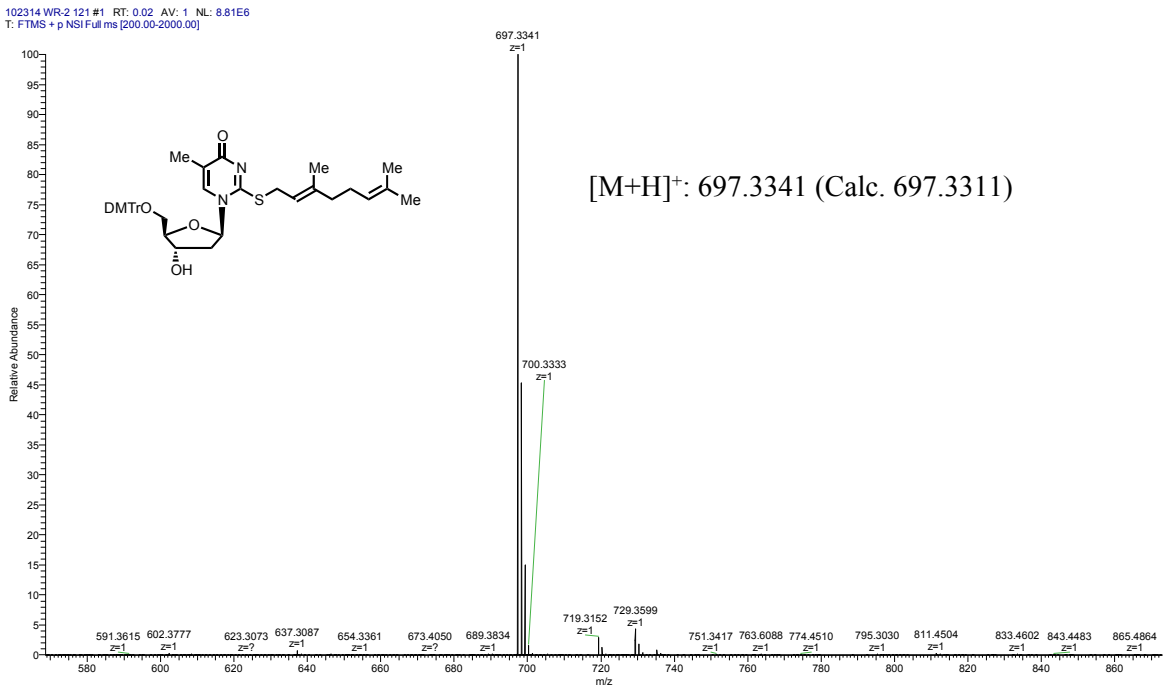


Fig. S6 High-resolution MS of compound **8**.

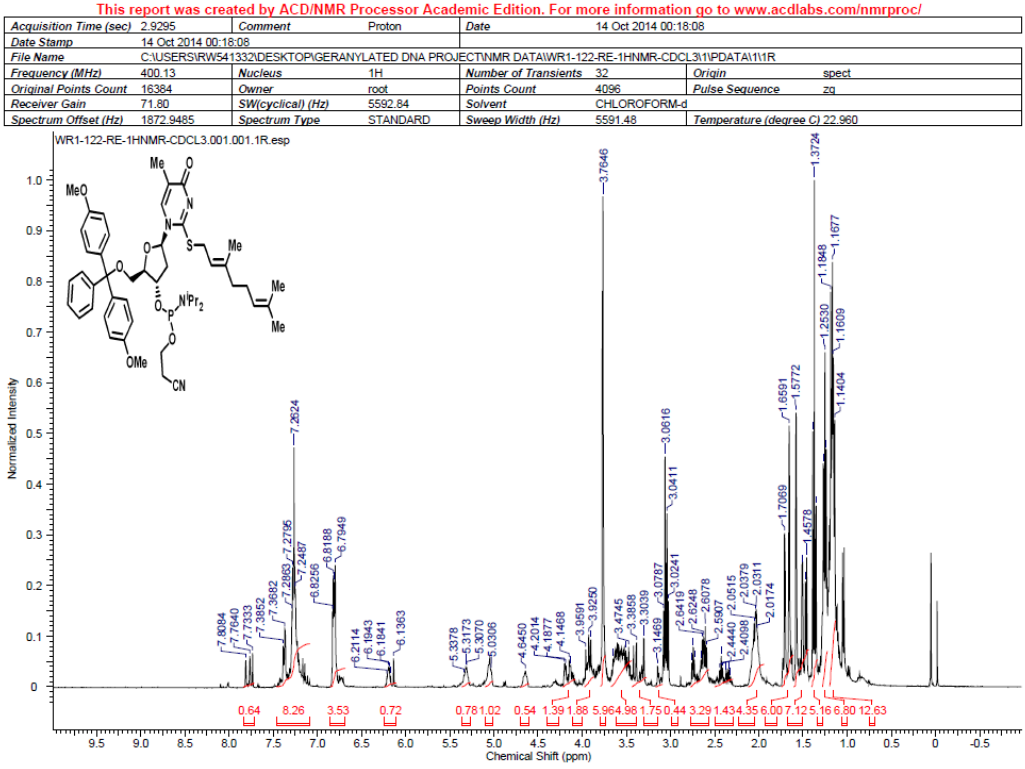


Fig. S7 ^{1}H NMR of compound **9**.

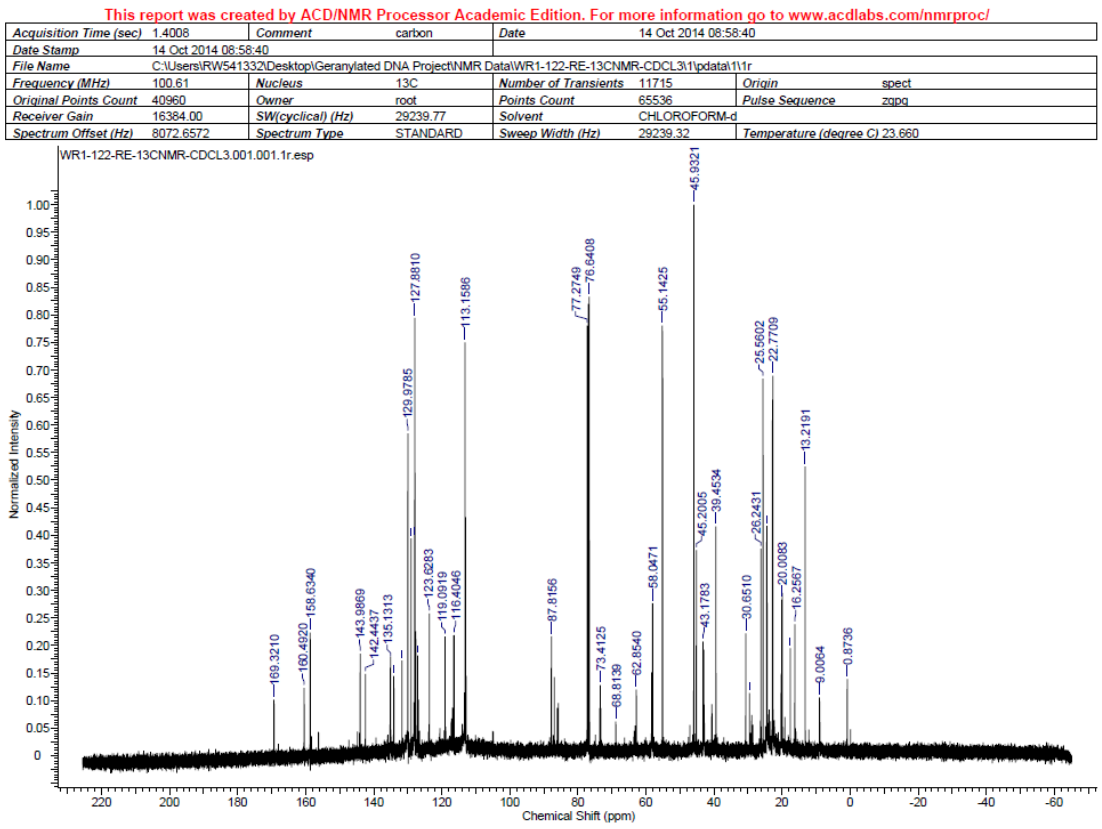


Fig. S8 ^{13}C NMR of compound **9**.

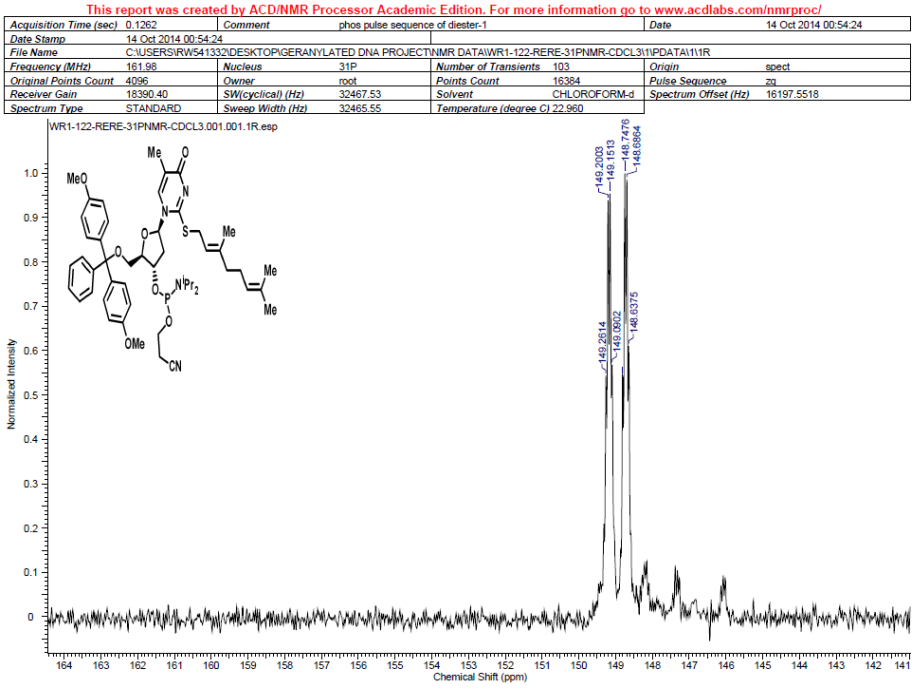


Fig. S9 ^{31}P NMR of compound **9**.

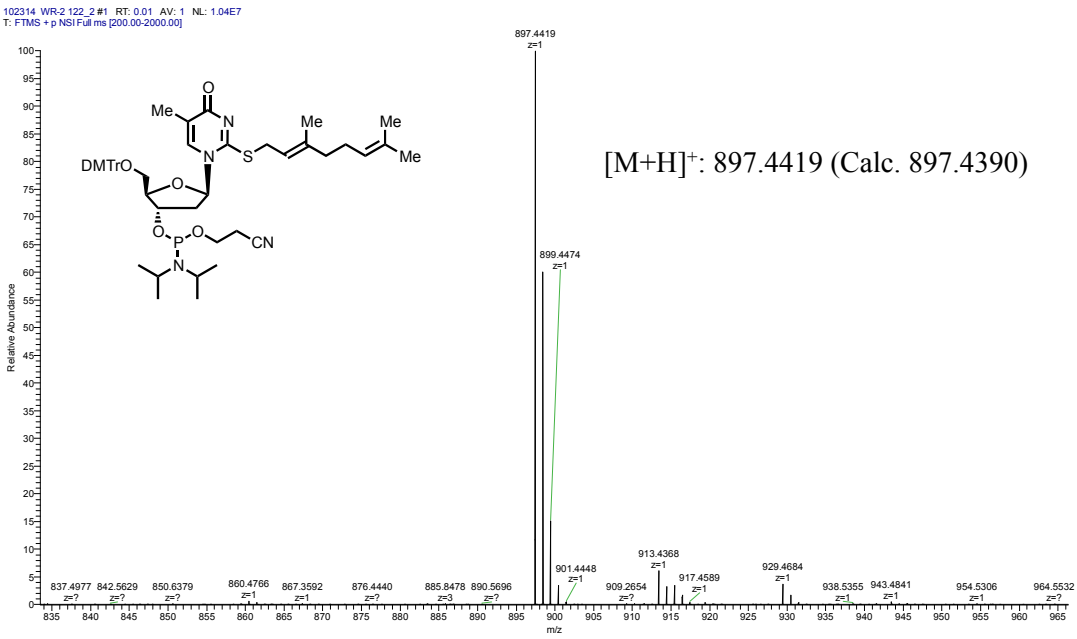


Fig. S10 High-resolution MS of compound **9**.

Part IV. Synthesis, HPLC and Characterization of geranylated oligonucleotides

Synthesis: All the oligonucleotides were chemically synthesized at 1.0-μmol scales by solid phase synthesis using a MerMade MM8 synthesizer. The geranyl thymidine phosphoramidite was dissolved in dichloromethane to a concentration of 0.07 M. 0.02 M I2 in THF/Py/H2O solution was used as oxidizing reagent. All the other reagents are standard solutions obtained from ChemGenes Corporation. Synthesis was performed on the appropriate nucleoside immobilized via a succinate linker to control-pore glass (CPG-500). All oligonucleotides were prepared in DMTr-on form. After synthesis, the oligos were cleaved from the solid support and fully deprotected with AMA (ammonium hydroxide:methylamine=1:1) at 65 °C for 30min. The amines were removed by Speed-Vac concentrator before HPLC purification. After the DMTr-on purification, the detritylation was carried out by the treatment of 3% trichloroacetice acid for 3min, followed by the neutralizationwith triethylamine to pH 7.0. The DMTr-off oligonucleotides were purified again by HPLC.

HPLC purification and analysis. The oligonucleotides were purified by reverse phase HPLC using a Zorbax SB-C18 column at a flow rate of 6 mL/min. Buffer A was 20 mM triethylammonium acetate, pH 7.1; buffer B contains 50% acetonitrile in 20 mM triethylammonium acetate, pH 7.1. A linear gradient from buffer A to 80% buffer B in 25 min was used to elute the oligos. The analysis was carried out by using the same type of analytical column with the same eluent gradient. All the geranyl-oligos were checked by high-resolution MS, as summarized in Table S1 and Fig. S11-13.

Table S1. MS analysis of geranylated oligonucleotides

Entry	Oligonucleotide molecular formula	Measured $[M+H]^+$ (calc.) *m/z*
1	5'-ATGG(ges2T)GCTC-3'	2881.59 (2881.94)
2	5'-CTTCT(ges2T)GTCCG-3'	3426.67 (3426.28)
3	5'-G(2'-OMe-U)G(ges2T)ACAC-3'	2577.54 (2577.59)

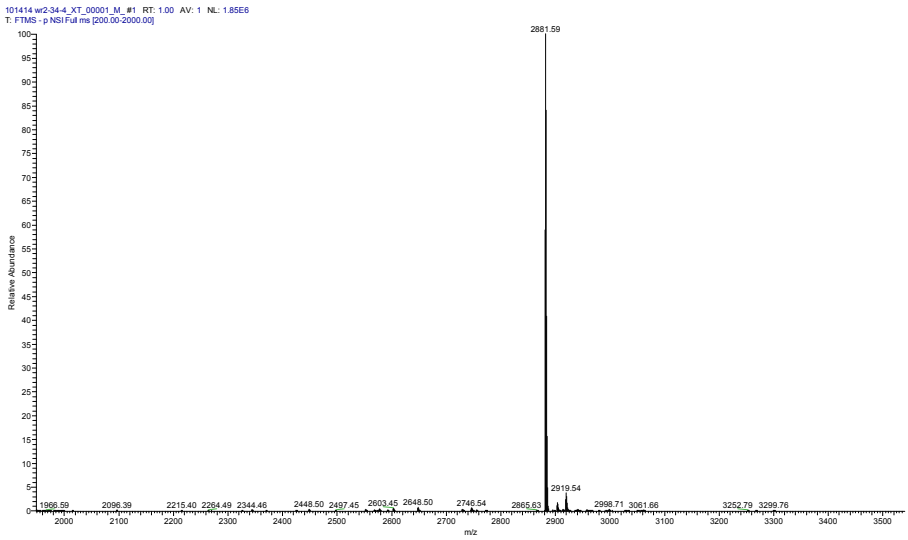


Fig. S11 Mass-spec of Oligo-1: 5'-ATGG(ges2T)GCTC-3'.

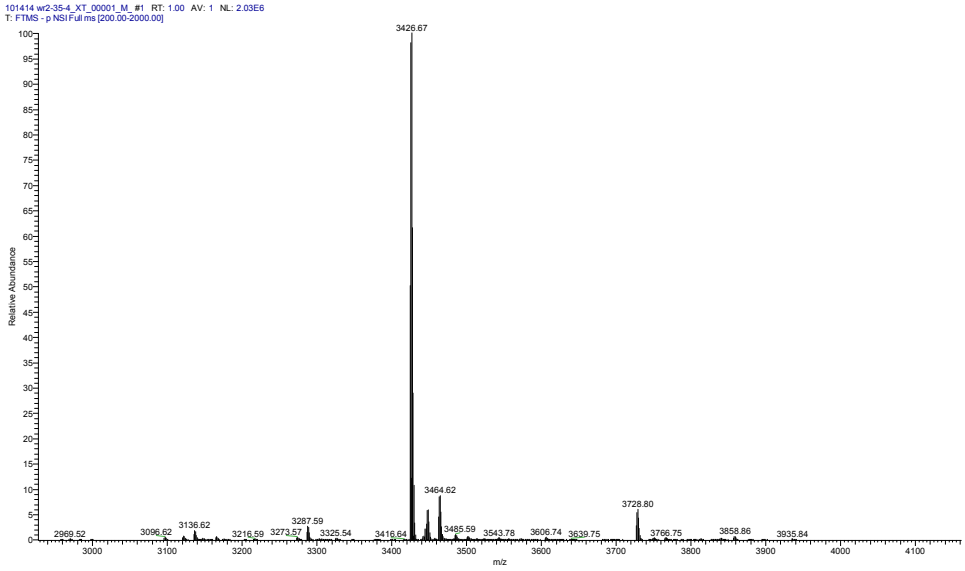


Fig. S12 Mass-spec of Oligo-2: 5'-CTTCT(ges2T)GTCCG-3'

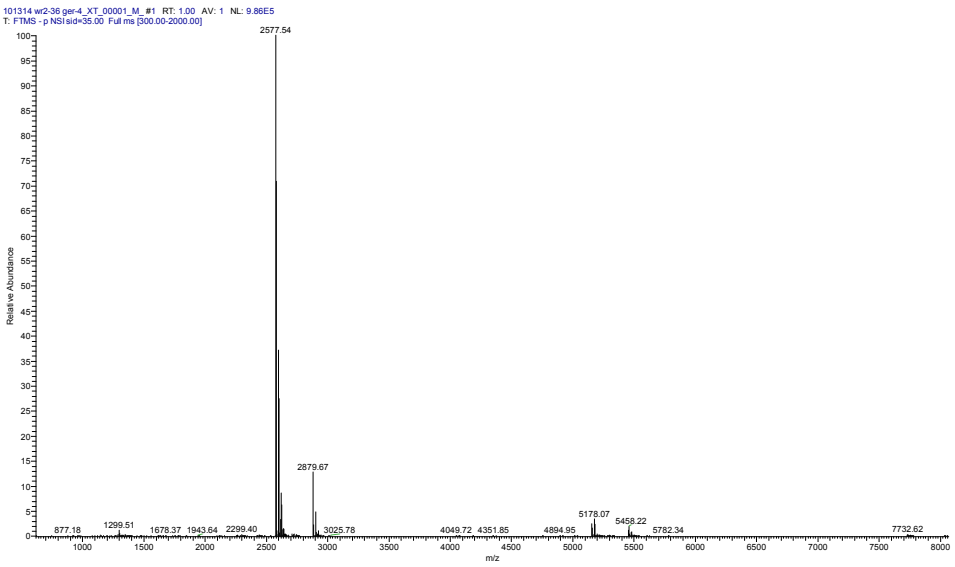


Fig. S13 Mass-spec of Oligo-3: 5'-G(2'-OMe-U)G(ges2T)ACAC-3'

Part V. Enzymatic hydrolysis and UHPLC-MS/MS study

Enzymatic hydrolysis of DNA

DNA oligos, 100 ng, were hydrolyzed to the composite mononucleosides via a two-step enzymatic hydrolysis.[1-3] The first step of phosphodiester bond cleavage was accomplished with nuclease P1 resulting in nucleoside-5'-monophosphates. Optimum nuclease P1 activity was achieved at pH 5.5 by addition of 1/10 volume of 1.0 M ammonium acetate pH 5.5 and the pH change confirmed. For each 0.5 absorbance unit of DNA, 2 units of nuclease P1 were added and incubated overnight at 37 °C. The second step of nucleoside preparation uses bacterial alkaline phosphatase (BAP) to cleave the 5'-phosphate from the nucleosides resulting in individual nucleosides and phosphoric acid. For optimum BAP activity, the pH was adjusted to pH 8.3 by adding a 1/10 volume of 1.0 M ammonium bicarbonate pH 8.3. One unit of BAP was added for each 0.5 absorbance units of DNA and incubated at 37 °C for 2 hours. The nucleosides were lyophilized and stored at -20°C. Samples were reconstituted in water prior to UHPLC-MS/MS analysis.

UHPLC – MS/ MS

Hydrolyzed oligos were subjected to chromatography using a Waters ACQUITY I-Class UPLC™ (Waters, USA) liquid chromatographic system equipped with a binary pump and autosampler that was maintained at 4 °C. A Waters ACQUITY UPLC™ HSS T3 column (2.1 x 50mm 1.7µm) and a HSS T3 guard column (2.1 x 5mm, 1.8µm) were used for the separation. The assay was completed at a flow rate of 0.2 ml/min and column temperature of 25 °C. Mobile phases included RNase-free water (18.0 MΩcm^{-1}) containing 0.01 % formic acid (Buffer A) and 50 % acetonitrile in aqueous 0.01% formic acid pH 3.5 (Buffer B). A 20-minute gradient was developed to obtain optimum separation of modified nucleosides, as shown in Table S2.

Table S2. UHPLC gradient conditions for analysis of DNA nucleosides.

Time (min)	Gradient (% B)
0	0
1	0
2.4	0.2
3.8	0.8
5.2	1.8
6.6	3.2
10.0	5.0
13.5	8.0
18.0	30.0
18.5	100.0
20.0	100.0
21.0	0.0
25.0	0.0

Tandem MS analysis of DNA nucleosides was performed on a Waters XEVO TQ-S™ (Waters, USA) triple quadrupole mass spectrometer equipped with an electrospray ionization (ESI) source maintained at 150 °C and the capillary voltage was set at 1kV. The desolvation gas, nitrogen was maintained at 500L/hr and desolvation temperature at 500 °C. The cone gas flow was set to 150L/hr and nebulizer pressure to 7 bars. Composition analysis was performed in ESI positive-ion mode using multiple-reaction monitoring (MRM). The ion transitions, cone voltage and collision energy used for UHPLC-MS/MS were determined using MassLynx V4.1Intellistart software. Retention times and the corresponding protonated molecular and product ion pairs [MH^+/BH_2^+] were obtained for each individual nucleoside including the synthetic geranyl-thymidine nucleoside as the standard (Fig. S14 and Table S3).

The incorporation of Geranyl-thymidine nucleoside in the sequence of [CTTCT(ges2T)GTCCG] was confirmed using MS/MS where the molecular ions [MH^+] are fragmented at the glycosidic bond providing the product ion [BH_2^+] and neutral sugar residue. [4] The resultant [MH^+]/[BH_2^+] ion-pair (Table S3) are then monitored using multiple reaction monitoring (MRM). Individual MRM selected to verify the oligonucleotide composition and successful incorporation of geranyl-nucleoside included Ger-T 395.21→279.22; dG 268.14→152.08; T 243.10→127.12; dC 228.12→112.11 (Fig. S14). Geranyl-thymidine-nucleoside was successfully incorporated showing the presence of ger-nucleoside only on the target oligo (Fig. S14-A) compare to the native DNA as negative control (Fig. S14-B) and the analysis of the individual synthetic geranyl-thymidine as our positive control (Fig. S14-C).

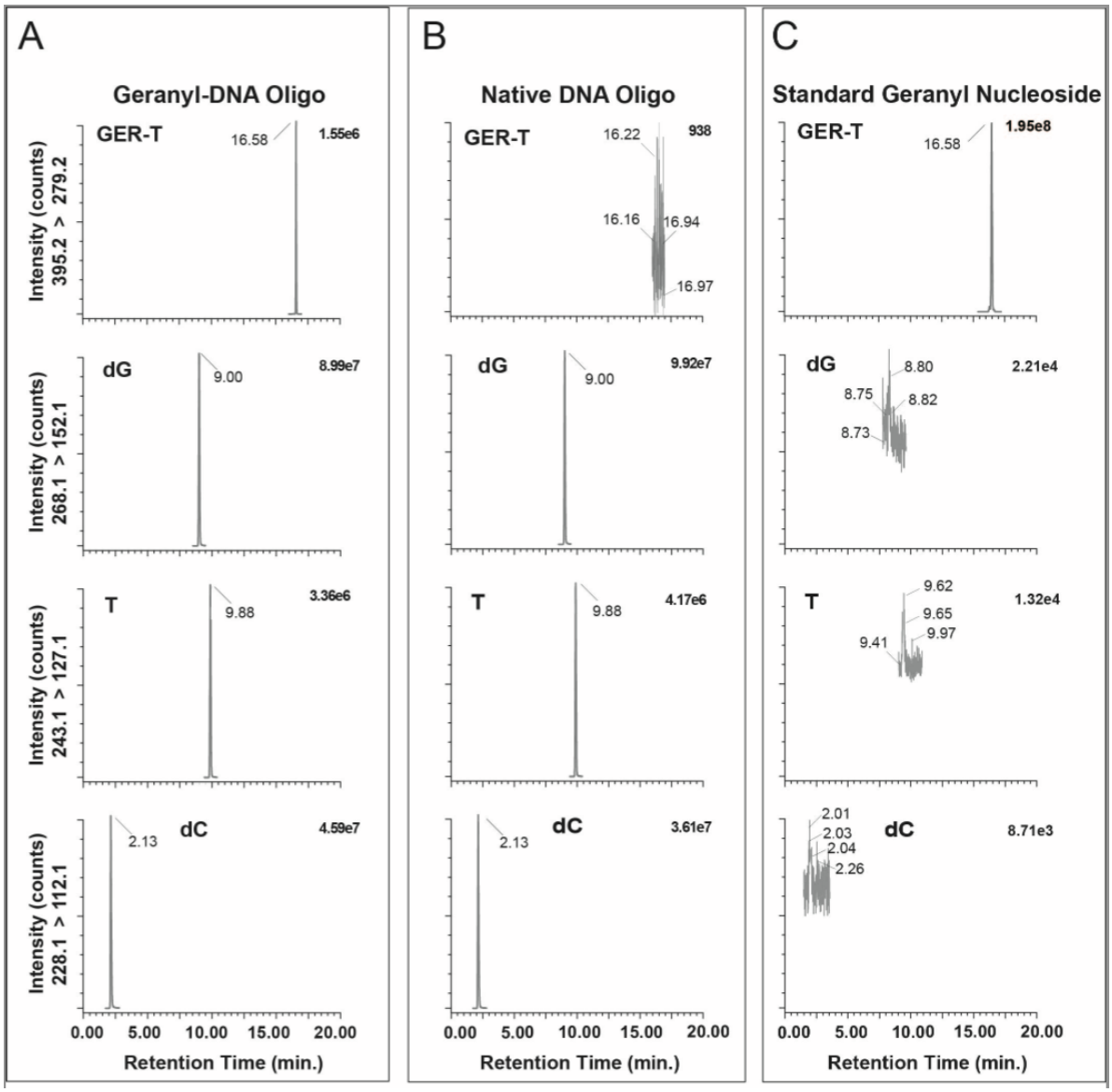


Fig S14. Composition analysis of DNA oligos were attained using tandem mass spectrometry in conjunction with ultra-high performance liquid chromatography. Molecular ion/product ion pairs, $[MH^+]/[BH_2^+]$, of geranyl-DNA oligo (A), native DNA oligo (B) and standard geranyl-thymidine nucleoside were monitored using Multiple reaction monitoring (MRM). Individual MRM selected included Ger-T 395.2→279.2; dG 268.1→152.1; T 243.1→127.1; dC 228.1→112.1. The retention times (min) are: 16.58, 9.00, 9.88 and 2.13 respectively as indicated in each trace. The native DNA sequence is [CTTCTTGTCCG], geranyl-DNA sequence is [CTTCT(ges2T)GTCCG].

Table S3. Mass spectrometric and UHPLC parameters for DNA nucleosides and geranyl thymidine. From left to right: molecular weight (MW); tandem MS fragmentation molecular/product ion pair $[MH^+]/[BH_2^+]$; collision energy (CE); and retention time (Rt).

Nucleosides	MW (g/mol)	$[MH^+]/[BH_2^+]$	CE (V)	Rt (min)
Geranyl-thymidine (GerT)	394.5	395.2/279.2	6	16.58

Deoxyguanosine (dG)	267.2	268.1/152.1	10	9.00
Thymidine (T)	242.2	243.1/127.1	12	9.88
Deoxycytidine (dC)	227.2	228.12/112.1	10	2.13

Part VI. UV-melting temperature (T_m) study

Solutions of the duplex DNAs (0.5 μM) were prepared by dissolving the purified DNAs in sodium phosphate (10 mM, pH 6.5) buffer containing 100 mM NaCl. The solutions were heated to 85 °C for 3 min, then cooled down slowly to room temperature, and stored at 4°C for 2h before Tm measurement. Thermal denaturation was performed in a Cary 300 UV-Visible Spectrophotometer with a temperature controller. The temperature reported is the block temperature. Each denaturizing curves were acquired at 260 nm by heating and cooling from 5 to 80°C for four times in a rate of 0.5 °C/min. All the melting curves were repeated for at least four times. The thermaldynamic parameter of each strand was obtained by fitting the melting curves in the Meltwin software.

Part VII. Molecular simulation study

To study the gesT modification in the context of the duplex in MD simulations, we developed AMBER[5] type force-field parameters for the atoms of the modified nucleoside. For obtaining the partial charges on the atoms, we used the online RESP/ESP charge-fitting server, REDS.[6] The geometry of the modified nucleoside was energy minimized, and Hartree-Fock level theory and 6-31G* basis-sets were employed to arrive at a set of partial charges.[7] AMBER-99 force-field parameters were used for bonded interactions,[5] and AMBER-99 parameters with Chen-Garcia corrections were used for LJ interactions.[8]

The unmodified DNA duplex was constructed in B-form using Nucleic Acid Builder (NAB) suite of AMBER 11 package. Using the WebMO graphical editor, we performed mutations such as, T → gesT and A → G, to get the two different DNA duplexes for MD studies.

Molecular dynamics simulations were performed using GROMACS-4.6.3 package.[9] The simulation system included the DNA duplex in a solution of 1M NaCl solution in a 3D periodic box. The box size was 6 x 6 x 6 nm^3 containing DNA duplex, 152 Na^+ ions, 130 Cl^- ions and 6,600 water molecules. The system was subjected to energy minimization to prevent any overlap of atoms, followed by a 10 ns MD simulation. The MD simulations incorporated leap-frog algorithm with a 2 fs timestep to integrate the equations of motion. The system was maintained at 300K and 1 bar, using the velocity rescaling thermostat[10] and Parrinello-Rahman barostat[11] respectively. The long-ranged electrostatic interactions were calculated using particle mesh Ewald (PME)[12] algorithm with a real space cut-off of 1.2 nm. LJ interactions were also truncated at 1.2 nm. TIP3P model[13] was used represent the water molecules, and LINCS[14] algorithm was used to constrain the motion of

hydrogen atoms bonded to heavy atoms. Coordinates of the DNA molecule were stored every 1 ps for further analysis.

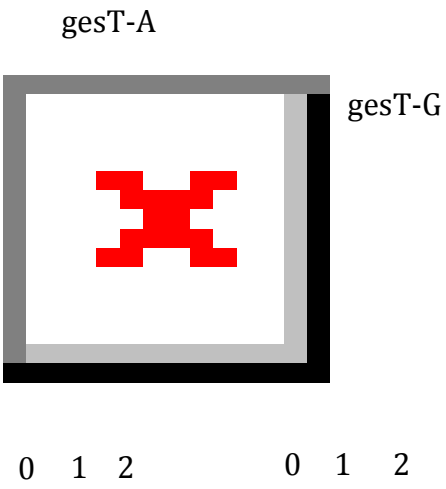


Fig. S15. Probability of hydrogen bond numbers during the whole simulation calculated from Fig. 5E: most of the ges2T-A pairs are in unbounded state without any hydrogen bonds (left bars), but ~65% of the ges2T-G patterns show two stable H-bonds (right bars). (The first 3 ns of the trajectory is assumed as equilibration time, and the rest of the 7 ns data is used to construct this plot).

Part VIII. References

1 C. W. Gehrke, K. C. Kuo, R. A. McCune, K. O. Gerhardt and P. F. Agris, *J. Chromatogr.*, 1982, **230**, 297.
2 D. Su, C. T. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley and P. C. Dedon, *Nat. Protoc.*, 2014, **9**, 828.
3 D. Dominissini, S. Moshitch-Moshkovitz, M. Salmon-Divon, N. Amariglio and G. Rechavi, *Nat. Protoc.*, 2013, **8**, 176.
4 R. Quinn, M. Basanta-Sanchez, R. E. Rose and D. Fabris, *J. Mass. Spectrom.*, 2013, **48**, 703.
5 W. Cornell, *Nurs. Qual. Connect.*, 1995, **5**, 1.
6 F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7821.
7 W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *J Am Chem Soc*, 1993, **115**, 9620.
8 A. A. Chen and A. E. Garcia, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 16820.
9 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory. Comput.*, 2008, **4**, 435.
10 G. Bussi, D. Donadio and M. Parrinello, *J Chem Phys*, 2007, **126**, 014101.
11 H. J. C. P. Berendsen, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *J. Chem. Phys.*, 1984, **81**, 3684.
12 T. Darden, York, D., and Pedersen, L., *J Chem Phys,* , 1993, **98**, 10089.
13 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.

14 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.